



(51) International Patent Classification:

C12N 15/62 (2006.01) A01H 5/00 (2018.01)

C12N 15/113 (2010.01)

(21) International Application Number:

PCT/IL2020/050254

(22) International Filing Date:

04 March 2020 (04.03.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/813,805 05 March 2019 (05.03.2019) US

(71) Applicant: RAMOT AT TEL-AVIV UNIVERSITY LTD. [IL/IL]; P.O. Box 39296, 6139201 Tel-Aviv (IL).

(72) Inventors: ZEMACH, Assaf; c/o Ramot at Tel-Aviv University Ltd., P.O. Box 39296, 6139201 Tel Aviv (IL). OHAD, Nir; c/o Ramot at Tel-Aviv University Ltd., P.O. Box 39296, 6139201 Tel Aviv (IL). YAARI, Rafael; c/o Ramot at Tel-Aviv University Ltd., P.O. Box 39296, 6139201 Tel Aviv (IL). KATZ, Aviva; c/o Ramot at Tel-Aviv University Ltd., P.O. Box 39296, 6139201 Tel Aviv

(IL). DOMB, Katherine; c/o Ramot at Tel-Aviv University Ltd., P.O. Box 39296, 6139201 Tel Aviv (IL).

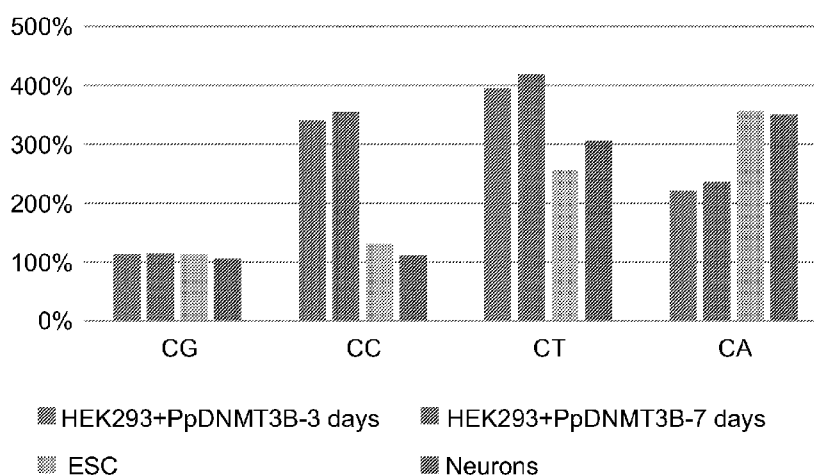
(74) Agent: EHRLICH, Gal et al.; G.E. EHRLICH (1995) LTD., 11 Menachem Begin Road, 5268104 Ramat Gan (IL).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: PLANT DNA METHYLTRANSFERASES AND USES THEREOF

FIG. 13



(57) Abstract: An isolated polynucleotide encoding a fusion protein which comprises a DNA targeting moiety linked to a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein is disclosed. Uses thereof are also disclosed.



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

PLANT DNA METHYLTRANSFERASES AND USES THEREOF

RELATED APPLICATIONS

This application claims the benefit of priority of U.S. Provisional Patent Application
5 No. 62/813,805 filed 5 March 2019, the contents of which are incorporated herein by reference in
their entirety.

SEQUENCE LISTING STATEMENT

The ASCII file, entitled 81636 Sequence Listing.txt, created on 3 March 2020,
10 comprising 143,498 bytes, submitted concurrently with the filing of this application is
incorporated herein by reference.

The project leading to this application has received funding from the European Research
Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme
15 (Grant Agreement No. 679551).

FIELD AND BACKGROUND OF THE INVENTION

The present invention, in some embodiments thereof, relates to plant DNA
methyltransferases (DNMTs) and more particularly to plant DNMT3 uses thereof.

20 DNA methylation, the addition of a methyl group to a cytosine base, is a prominent
epigenetic modification in many eukaryotes. It is catalyzed by distinct DNA methyltransferase
(DNMT) families of proteins that share a conserved methyl-transferase domain (MTD). In plants,
DNMTs evolved to methylate cytosines located in specific contexts (CG, CHG, and CHH; H= A,
C, or T), distinct genetic elements (e.g. transposons and genes), various chromatin configurations
25 (hetero- and eu-chromatin), as well as to establish methylation *de novo* at newly unmethylated
sites or to maintain methylation upon DNA replication. Plants encode four types of DNMTs:
Methyltransferase 1 (*MET1*), DNA methyltransferase 3 (*DNMT3*), chromomethylase (*CMT*), and
domain rearranged methyltransferase (*DRM*). *MET1*s are homologs of mammalian DNMT1 and
maintain CG methylation. *CMT*s are plant specific DNMTs first to appear in charophytes.
30 *Arabidopsis thaliana* (*Arabidopsis*) *CMT2* and *CMT3* orthologs utilize their chromodomain (CD)
to bind to histone H3 lysine 9 dimethylation (H3K9me2) heterochromatin and to methylate CHH
and CHG sites, respectively. DNMT3s are ancient DNMTs that exist in animals, plants, and other
eukaryotes^{1,23}. Mammalian DNMT3s function primarily as *de novo* CG methylases and in
specific tissues also at CH sites^{1,6,24}. However, despite their significant role in mammals, non-

animal DNMT3s have not been investigated thus far. DNMT3s were overlooked in plants probably due to their deficiency in angiosperms (flowering plants) and the discovery of their close homologs, DRMs, which function in *de novo* methylation. DRMs are plant specific DNMTs with a rearranged DNMT3-MTD. Angiosperm DRMs are a part of the RNA directed DNA methylation (RdDM) pathway that utilizes small RNA to establish *de novo* methylation within euchromatic transposons, that is enriched with active histone marks such as H3K4me3 and depleted of repressive marks as H3K9me2. So far, the function of plant DNMTs was comprehensively investigated in *Arabidopsis thaliana* and partially explored in a few additional angiosperms, all which lack DNMT3 in their genome.

SUMMARY OF THE INVENTION

According to an aspect of the present invention there is provided an isolated polynucleotide encoding a fusion protein which comprises a DNA targeting moiety linked to a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein.

According to embodiments of the present invention, the DNA targeting moiety comprises a DNA endonuclease protein.

According to embodiments of the present invention, the DNA endonuclease protein is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 and Cpf1 endonuclease.

According to embodiments of the present invention, the DNA endonuclease protein comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.

According to an aspect of the present invention there is provided an isolated polynucleotide encoding a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein having a codon usage optimized for expression in an organism which is not a gymnosperm or a bryophyte.

According to embodiments of the present invention, the plant DNMT3 protein is a gymnosperm or a bryophyte DNMT3 protein.

According to embodiments of the present invention, the organism is a mammal.

According to embodiments of the present invention, the mammal is a human.

According to embodiments of the present invention, the organism is an angiosperm.

According to embodiments of the present invention, the DNMT3 protein is fused to a DNA targeting moiety.

According to embodiments of the present invention, the DNA targeting moiety comprises a DNA endonuclease protein.

According to embodiments of the present invention, the DNA endonuclease protein is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 and Cpf1 endonuclease.

According to embodiments of the present invention, the DNA endonuclease protein comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.

According to embodiments of the present invention, the fusion protein comprises a single copy of said DNMT3 protein.

According to embodiments of the present invention, the catalytic domain of the DNMT3 protein comprises an amino acid sequence as least 70 % similar or identical to at least one of the sequences as set forth in SEQ ID NOs: 1-11.

According to embodiments of the present invention, the catalytic domain of the DNMT3 protein comprises an amino acid sequence 100 % similar or identical to at least one of the sequences as set forth in SEQ ID NO: 1-11.

According to embodiments of the present invention, the catalytic domain is linked directly to said DNA endonuclease protein.

According to embodiments of the present invention, the catalytic domain is linked to said endonuclease protein via a peptide linker.

According to embodiments of the present invention, the catalytically inactive Cas9 protein comprises mutations at a site selected from the group consisting of D10, E762, H983, D986, H840 and N863.

According to embodiments of the present invention, the mutations are: (i) D10A or D10N, and (ii) H840A, H840N, or H840Y.

According to embodiments of the present invention, the mutations are D10A and H840A.

According to embodiments of the present invention, the dCAS9 comprises the sequence as set forth in SEQ ID NO: 23.

According to embodiments of the present invention, the DNMT3 protein is linked to the C terminus of said endonuclease protein.

According to embodiments of the present invention, the DNMT3 protein is linked to the N terminus of said endonuclease protein.

According to embodiments of the present invention, the DNMT3 methylates a target DNA at a CHH site.

According to embodiments of the present invention, the DNMT3 additionally methylates a target DNA at a CpG site.

5 According to embodiments of the present invention, the DNMT3 methylates a target DNA at a CC site and/or a CT site to a greater extent than a human DNMT3 methylates the target DNA under identical conditions.

According to an aspect of the present invention there is provided a polypeptide comprising a DNA targeting moiety linked to a DNMT3 protein.

10 According to embodiments of the present invention, the DNA targeting moiety comprises a DNA endonuclease protein.

According to embodiments of the present invention, the DNA endonuclease protein is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, 15 Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 and Cpf1 endonuclease.

According to embodiments of the present invention, the DNA endonuclease protein comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.

An expression vector comprising the polynucleotide described herein.

20 According to an aspect of the present invention there is provided an expression vector comprising a polynucleotide encoding a catalytic domain of a species of a plant DNA methyltransferase 3 (DNMT3) protein operatively linked to a transcriptional regulatory sequence which is not of said species.

According to embodiments of the present invention, the transcriptional regulatory 25 sequence is not a gymnosperm transcriptional regulatory sequence or a bryophyte regulatory sequence.

According to embodiments of the present invention, the transcriptional regulatory sequence comprises a mammalian transcriptional regulatory sequence.

According to embodiments of the present invention, the transcriptional regulatory 30 sequence comprises an angiosperm transcriptional regulatory sequence.

According to an aspect of the present invention there is provided a cell which expresses the polynucleotide described herein.

According to an aspect of the present invention there is provided a cell which comprises the expression vector described herein.

According to embodiments of the present invention, the cell is a mammalian cell.

According to embodiments of the present invention, the cell is a plant cell.

According to embodiments of the present invention, the plant cell is an angiosperm cell.

According to an aspect of the present invention there is provided a kit comprising the
5 polynucleotide described herein and at least one guide RNA which is directed to a predetermined target gene.

According to an aspect of the present invention there is provided a method of increasing methylation of DNA in a cell, the method comprising expressing a polynucleotide encoding a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein in the cell, thereby
10 increasing methylation of DNA in the cell, wherein the cell is not of a gymnosperm plant.

According to an aspect of the present invention there is provided a method of increasing methylation of DNA in a cell, the method comprising expressing the polynucleotide described herein in the cell, thereby increasing methylation of DNA in the cell.

According to embodiments of the present invention, the method further comprises
15 expressing one or more guide RNA directed to a target gene of the cell.

According to embodiments of the present invention, the cell is a mammalian cell.

According to embodiments of the present invention, the cell is a plant cell.

According to embodiments of the present invention, the cell is a diseased cell.

According to embodiments of the present invention, the mammalian cell is a human cell.

Unless otherwise defined, all technical and/or scientific terms used herein have the same
20 meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of the invention, exemplary methods and/or materials are described below. In case of conflict, the patent specification, including definitions,
25 will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be necessarily limiting.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Some embodiments of the invention are herein described, by way of example only, with
30 reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of embodiments of the invention. In this regard, the description taken with the drawings makes apparent to those skilled in the art how embodiments of the invention may be practiced.

In the drawings:

FIGs. 1A-C. PpDNMT3b and PpCMT establish DNA methylation and maintain the entire non-CG methylome.

A. Sequences of DNMT3 and DRM MTD regions were aligned using MUSCLE ⁷¹. The phylogenetic tree was constructed by IQ-TREE⁶⁷⁻⁷⁰ and illustrated by FigTree. DRM MTDs were reorganized to fit the linear motif order as in canonical DNMTs. DNMT1 homologs were added as an outgroup. Clades having bootstrap value above 70 % are marked with a circle. Protein accessions are listed in Table 5. Colors depict taxonomic groups: red- animals; blue- charophytes; green- non-seed land plants; brown- gymnosperms; magenta- basal angiosperms; orange – monocots; purple – dicots.

B. Averaged genomic cytosine methylation of WT and DNMT mutants in three sequence contexts, CG, CHG, and CHH. See Table 7 for detailed information.

C. RPS methylation level in WT and indicated mutants. Red, yellow, and green represent CG, CHG, and CHH methylation, respectively.

FIGs. 2A-E. Genomic CG methylation regulation by PpCMT and PpDNMT3b.

A. Patterns of TE CG methylation in WT and indicated mutants. *P. patens* TEs were aligned at the 5' end and average methylation for all cytosines within each 100 bp interval is plotted. The dashed lines represent the points of alignment.

B. Box plot of NCG methylation difference in TEs between WT and indicated mutants (N= any nucleotide).

C. Box plot of the residual NCG methylation in TEs in *met* mutant.

D. Patterns of TE CHG methylation in WT and indicated mutants (similar to B).

E. Averaged genomic CHG methylation level in WT and DNMT mutants separated to CWG (i.e. CAG or CTG) and CCG.

FIGs. 3A-G. PpCMT and PpDNMT3 methylate heterochromatin.

A. Pearson correlation coefficients between CG/CHG/CHH methylation, GC content, and indicated histone modifications of TEs in 50 bp windows.

B. Box plots showing GC content, H3K9me2, and H3K4me3 levels in 50 bp windows within five quantile TE sizes.

C. Box plots of averaged DNA methylation in 50 bp windows of WT protonoma over five quantiles of TE sizes.

D. Box plots of percent-methylation-change between WT and indicated mutants 50 bp windows with a minimum 10% methylation in either of the samples, over TE size.

E. Patterns of TE CHH methylation in WT and indicated mutants as described in Figure 2A.

F. Box plots showing the distribution of percent-methylation-change per 50 bp windows between WT and cmt mutant over H3K9me2, GC content, and TE size quantiles.

G. CHH methylation level (red WT, blue mutant) CHH methylation difference (cmt minus WT), H3K9me2, and gene/TE annotations of a representative region. Genes and TEs oriented 5' to 3' and 3' to 5' are shown above and below the line, respectively. Open black box marks a cmt hypo-methylated region enriched for H3K9me2.

FIGs. 4A-E. PpDRMs methylate active-euchromatic TEs.

A. Box plots of percent-methylation-change between indicated samples within differentially CHH methylated 50 bp windows, separated based on the level of various genomic/chromatin attributes. Note the hypo-methylation trend in protonema *drm12* sample (top track) in genomic regions with high siRNA counts, low GC content, absent H3K9me2 signal, high H3K4me3 signal, short TEs, LTR annotations, and TE expression. B. Venn diagram showing abundance and overlap between siRNA, CHH methylation, and TE annotation, in *Arabidopsis* and *P. patens*. C. siRNA abundance over increased quantiles of indicated chromatin features in *A. thaliana* (c) and *P. patens* (c). D. Patterns of TE integration in *Arabidopsis* and *P. patens* upstream to gene TSS. *Arabidopsis* or *P. patens* genes were aligned at the 5' end (0 at x axis) and percentage of the number of TEs (first and closest nucleotide of TE to TSS) within each 25 bp is plotted. E. LOWESS fit of DNA methylation distribution averaged in 100 kb bins across chromosome 1 in *Arabidopsis* and *P. patens*.

FIGs. 5A-B. Mechanisms and evolution of plant DNMTs.

A. DNMT methylation mechanisms are illustrated based on current knowledge. Black line represents the DNA with different cytosine subcontexts embedded in it. Lollipops represent methylation. Arrows width are corresponding qualitatively to the relative level of methylation mediated by indicated DNMTs. HeC.- heterochromatin, EuC.- euchromatin. De novo and maintenance methylation activities are shown above and below the DNA, respectively.

B. Schematic illustration of the evolution of plant DNMTs and their function based on previous and the present studies.

FIG. 6. Genotyping of PpDNMT mutants. BS-seq reads coverage of *PpDNMT* genes in WT and *dnmt* mutants.

FIGs. 7A-U. Mutagenesis of PpDNMT3 and PpDRM does not disrupt *P. patens* development. Morphological analysis of protonema and gametophore development in WT and PpDNMT3 and DRM deletion mutants. a-g, Seven days old protonemata of WT (A) and mutants

(B-G). Scale bar: 50 μ m. H-N, Three weeks old plants bearing gametophores of WT (H) and mutants (I-N). Scale bar: 250 μ m. O-U, Six weeks old gametophores of WT (O) and mutants (P-U). Scale bar: 1 mm.

FIG. 8. Preferences in CHH methylation subcontexts. Averaged genomic CHH methylation level, in wild type and DNMT mutants, separated to its subcontexts.

FIG. 9. Regulation of CHH methylation by PpCMT. Box plots showing the distribution of percent-methylation-change per 50 bp windows between wild type and cmt mutant over H3K9me2, GC content within TEs shorter than 500 bp long.

FIGs. 10A-E. DNA methylation in *drm* and *rdr2* mutants. A. DNA methylation difference between WT and indicated mutants. B. Number of hypo- and hyper-methylated CG, CHG, and CHH DMRs in each of the mutants. C-D. Percent-methylation-change between WT and *rdr2* mutant within *rdr2*-CHH-DMRs (C) or *drm12*-CHH-DMRs (D) over five centiles of indicated genomic or chromatin attributes. The WT plant is genetically unrelated to the *rdr2* one, thus comparison between these two plants could contribute to a noise level that could mask a weak hypo-methylation signal. Therefore, the CHH methylation change in *rdr2* was analyzed within *drm12*-CHH-DMRs (d). Note the change from global hypermethylation in *rdr2* in (C) to a slight hypo-methylation at low GC regions, short TEs, LTRs, and expressed TEs in (D).

FIG. 11. Global patterns of CG, CHG and CHH methylation (H= A, C, or T) in genes and transposons (TEs) in HEK293 cells expressing PpDNMT3b-GFP or GFP (control). DNA methylation was profiled in 84 Mega bases in the human genome using SureSelectXT Human Methyl-Seq Target Enrichment Panel (Agilent) and Illumina high throughput sequencing. Genes and TEs were aligned at either the 5' or 3' end and average methylation for all cytosines within each 50 bp interval was plotted. The dashed lines represent the points of alignment. These graphs show a specific CHH hypermethylation in cells transfected with PpDNMT3b.

FIG. 12. Localized patterns of CG, CHG and CHH methylation in a particular gene and transposon in HEK293 cells expressing PpDNMT3b-GFP or GFP (control). A snapshot of methylation patterns in a representative gene (top panel) and a TE (bottom panel) regions is presented. Tracks order from the top to the bottom is as following. The three top tracks display CG, CHG and CHH methylation levels of control sample, the three bottom tracks display the corresponding PpDNMT3b methylation. CG, CHG and CHH methylation levels are represented as color scale of blue, green and red, respectively (white bar means zero methylation). These graphs show a specific CHH hypermethylation (red bars) in PpDNMT3b transfected cells, either in a region depleted of methylated CGs (top panel) or containing methylated CGs (bottom panel).

FIG. 13. Comparison of methylation levels between HEK293 cells expressing PpDNMT3b and two other human cell types with heightened non-CG methylation. DNA methylation was profiled for 84 Mega bases of the human genome using SureSelectXT Human Methyl-Seq Target Enrichment kit (Agilent) and Illumina high throughput sequencing. This figure shows global average methylation levels separated to CG/CC/CT/CA sequence context groups normalized with fetal tissue (having low non-CG methylation as most human tissues/cell-types and specifically HEK293 cells) for HEK293 cells expressing PpDNMT3b, 3 or 7 days following transfection as well as human tissues having significant non-CG methylation levels: neurons from adult and embryonic stem cells (ESC). PpDNMT3b expression in human HEK293 cells resulted in non-CG hypermethylation genome wide.

FIG. 14. Several PpDNMT3b upregulated genes in HEK293 cells show increase in non-CG methylation. For each gene, the difference in methylation between sites having at least 10% methylation in either PpDNMT3b or control lines is plotted in boxplots.

FIG. 15. Expression of PpDNMT3b in Arabidopsis induces CHH methylation. BS-seq data of an Arabidopsis plant expressing PpDNMT3b in the background of ddcc (drm1 drm2 cmt2 cmt3 quadruple mutant which has trivial non-CG methylation levels) was analyzed with methylpy. The difference in methylation genome wide between sites having at least 10% methylation in either PpDNMT3b/ddcc or control (ddcc) lines is plotted in boxplots.

FIG. 16 is the DNA sequence of FLAG-NLS-dcas9-NLS-PpDNMT3b_MTD-T2A-PuroR (SEQ ID NO: 64). PpDNMT3b methyltransferase domain (MTD) (marked in red within the DNA sequence was expressed in fusion with dcas9 (marked in blue) along with FLAG-tag, protein nuclear localization sequences (NLS) and poly-Gly linkers separating dcas9, NLS and PpDNMT3b-MTD. Additionally, this open reading frame continues following the PpDNMT3b-MTD sequence with a T2A protein separation sequence (marked in green) to allow expression of Puromycin resistance gene (One reading frame allowing expression of dcas9-PpDNMT3b-MTD and PuroR as separate proteins).

FIG. 17 is the protein sequence of FLAG-NLS-dcas9-NLS-PpDNMT3b_MTD-T2A-PuroR (SEQ ID NO: 65). PpDNMT3b methyltransferase domain (MTD) (marked in red within the amino acid sequence was expressed in fusion with dcas9 (marked in blue) along with FLAG-tag, protein nuclear localization sequences (NLS) and poly-Gly linkers separating dcas9, NLS and PpDNMT3b-MTD.

DESCRIPTION OF SPECIFIC EMBODIMENTS OF THE INVENTION

The present invention, in some embodiments thereof, relates to plant DNA methyltransferases (DNMTs) and more particularly to plant DNMT3 uses thereof. Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not necessarily limited in its application to the details set forth in the following description or exemplified by the Examples. The invention is capable of other embodiments or of being practiced or carried out in various ways.

To properly regulate the genome, cytosine methylation is established by animal DNA methyltransferase 3s (DNMT3s). While altered DNMT3 homologs, Domains rearranged methyltransferases (DRMs), have been shown to establish methylation via the RNA directed DNA methylation (RdDM) pathway, the role of true-plant DNMT3 orthologs have so far remained elusive.

The present inventors have now profiled de novo (RPS transgene) and genomic methylation in the basal plant, *Physcomitrella patens*, mutated in each of its PpDNMTs. The present inventors have shown that PpDNMT3b mediates CG and CHH de novo methylation, independently of PpDRMs (Figure 1B).

Whilst further reducing the present invention to practice, the present inventors have shown that the novel plant derived DNMT3 has a methyltransferase activity when expressed in mammalian cells (see Figures 11-14). The expressed DNMT3 had a higher preference to methylate CC or CT sites than human DNMT3s under identical conditions. As a result of the methyltransferase activity, the present inventors showed that expression of numerous genes was upregulated (see Table 8).

In addition, the present inventors expressed the novel plant DNMT3 in a heterologous plant system (*Arabidopsis*) and showed that it carried out CHH methylation (Figure 15).

According to a first aspect of the present invention there is provided an isolated polynucleotide encoding a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein.

As used herein the term “polynucleotide” refers to a single or double stranded nucleic acid sequence which is isolated and provided in the form of an RNA sequence, a complementary polynucleotide sequence (cDNA), a genomic polynucleotide sequence and/or a composite polynucleotide sequences (e.g., a combination of the above).

Thus, some embodiments of the invention encompasses nucleic acid sequences described herein; fragments thereof, sequences hybridizable therewith, sequences homologous thereto, sequences orthologous thereto, sequences encoding similar polypeptides with different codon

usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion.

The term “isolated” refers to at least partially separated from its natural environment.

5 The polynucleotide sequence may be a DNA or RNA sequence encoding a catalytic domain of a plant DNA methyltransferase (DNMT) protein, capable of methylating target DNA at a CHH site.

The DNA methyltransferase may be derived from any plant.

10 In one embodiment, the DNA methyltransferase is derived from a monocotyledonous plant.

Monocotyledonous plants belong to the orders of the Alismatales, Arales, Arecales, Bromeliales, Commelinales, Cyclanthales, Cyperales, Eriocaulales, Hydrocharitales, Juncales, Lilliales, Najadales, Orchidales, Pandanales, Poales, Restionales, Triuridales, Typhales, and Zingiberales. Plants belonging to the class of the Gymnospermae are Cycadales, Conifers, 15 Ginkgoales, Gnetales, and Pinales.

In another embodiment, the DNA methyltransferase is derived from a gymnosperm including but not limited to *Encephalartos barteri*, *Stangeria eriopus*, *Welwitschia mirabilis*, *Welwitschia mirabilis*, *Pinus taeda*, *Pinus sylvestris*, *Manoao colensoi*, *Sundacarpus amarus* and *Pinus jeffreyi*.

20 In another embodiment, the DNA methyltransferase is derived from a bryophytes (e.g. a moss or a liverworts), specific examples of such including but not limited to *Marchantia polymorpha*, *Physcomitrella patens* and *Sphagnum fallax*.

In particular, the DNA methyltransferase is derived from *Physcomitrella patens*.

25 In still another embodiment, the DNA methyltransferase is derived from a Charophytes, including for example *Klebsormidium flaccidum*.

In yet another embodiment, the DNA methyltransferase is derived from a lycophytes including for example *Selaginella moellendorffii*.

In another embodiment, the DNA methyltransferase is derived from a dicotyledonous plant. Such plants include those belonging to the orders of the Aristochiales, Asterales, Batales, 30 Campanulales, Capparales, Caryophyllales, Casuarinales, Celastrales, Cornales, Diapensales, Dilleniales, Dipsacales, Ebenales, Ericales, Eucomiales, Euphorbiales, Fabales, Fagales, Gentianales, Geraniales, Haloragales, Hamamelidales, Middles, Juglandales, Lamiales, Laurales, Lecythidales, Leitneriales, Magniolales, Malvales, Myricales, Myrtales, Nymphaeales, Papeverales, Piperales, Plantaginales, Plumbaginales, Podostemales, Polemoniales, Polygalales,

Polygonales, Primulales, Proteales, Rafflesiales, Ranunculales, Rhamnales, Rosales, Rubiales, Salicales, Santales, Sapindales, Sarraceniaceae, Scrophulariales, Theales, Trochodendrales, Umbellales, Urticales, and Violales.

In one embodiment, the plant DNA methyltransferase protein is DNMT3. The plant DNA methyltransferase is capable of methylating target DNA at a CG site as well as at a CHH site.

In one embodiment, the plant DNMT3 methylates a target DNA at a CC site and/or a CT site to a greater extent than a human DNMT3 methylates the target DNA under identical experimental conditions.

The plant DNA methyltransferase of this aspect of the present invention is not a DRM and does not require siRNA to bring about methylation.

The phrase “catalytic domain” as used herein refers to part of the DNMT3 protein (*i.e.*, a polypeptide) which exhibits functional properties of the enzyme such as methylating target DNA (the functional domain). According to preferred embodiments of the invention the catalytic domain of a plant DNMT3 is a polypeptide sequence which comprises a sequence at least 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or 100 % identical (or similar) to one of the sequences set forth in SEQ ID NOs: 1-11.

According to a particular embodiment, the catalytic domain of a plant DNMT3 is a polypeptide sequence which comprises a sequence at least 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or 100 % identical (or similar) to one of the sequences set forth in SEQ ID NOs: 5 or 6, as determined using the BestFit software of the Wisconsin sequence analysis package, utilizing the Smith and Waterman algorithm, where gap weight equals 50, length weight equals 3, average match equals 10 and average mismatch equals -9.

To determine the percent identity of two sequences, the sequences are aligned for optimal comparison purposes (gaps are introduced in one or both of a first and a second amino acid or nucleic acid sequence as required for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 50% (in some embodiments, about 50%, 55%, 60%, 65%, 70%, 75%, 85%, 90%, 95%, or 100% of the length of the reference sequence) is aligned. The nucleotides or residues at corresponding positions are then compared. When a position in the first sequence is occupied by the same nucleotide or residue as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the

number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For purposes of the present application, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package, using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The isolated polynucleotide of this aspect of the present invention may encode the full length DNMT3 (i.e. the catalytic domain and the regulatory domain). Thus, the isolated polynucleotide may encode proteins comprising amino acid sequences which are at least 50 %, at least 55 %, at least 60 %, at least 65 %, at least 70 %, at least 75 %, at least 80 %, at least 81 %, at least 82 %, at least 83 %, at least 84 %, at least 85 %, at least 86 %, at least 87 %, at least 88 %, at least 89 %, at least 90 %, at least 91 %, at least 92 %, at least 93 %, at least 94 %, at least 95 %, at least 96 %, at least 97 %, at least 98 %, at least 99 % or 100 % identical to any one of SEQ ID NOS: 12-22, or at least 70 %, at least 75 %, at least 80 %, at least 81 %, at least 82 %, at least 83 %, at least 84 %, at least 85 %, at least 86 %, at least 87 %, at least 88 %, at least 89 %, at least 90 %, at least 91 %, at least 92 %, at least 93 %, at least 94 %, at least 95 %, at least 96 %, at least 97 %, at least 98 %, at least 99 % or 100 % identical to any one of SEQ ID NOS: 1-11, as determined using the BestFit software of the Wisconsin sequence analysis package, utilizing the Smith and Waterman algorithm, where gap weight equals 50, length weight equals 3, average match equals 10 and average mismatch equals -9.

According to a particular embodiment, the plant DNMT3 is a polypeptide sequence which comprises a sequence at least 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or 100 % identical (or similar) to SEQ ID NO: 16.

To express the plant derived DNMT3 in a heterologous system, the codon usage of the nucleic acid sequence which encodes the DNMT3 may be optimized.

Nucleic acid sequences encoding the enzymes of some embodiments of the invention may be optimized for expression in a particular system. Examples of such sequence modifications include, but are not limited to, an altered G/C content to more closely approach that typically found in the species of interest, and the removal of codons atypically found in the species commonly referred to as codon optimization.

The phrase "codon optimization" refers to the selection of appropriate DNA nucleotides for use within a structural gene or fragment thereof that approaches codon usage within the

species of interest. Therefore, an optimized gene or nucleic acid sequence refers to a gene in which the nucleotide sequence of a native or naturally occurring gene has been modified in order to utilize statistically-preferred or statistically-favored codons within the species. The nucleotide sequence typically is examined at the DNA level and the coding region optimized for expression in the species determined using any suitable procedure, for example as described in Sardana *et al.* (1996, Plant Cell Reports 15:677-681). In this method, the standard deviation of codon usage, a measure of codon usage bias, may be calculated by first finding the squared proportional deviation of usage of each codon of the native gene relative to that of highly expressed genes, followed by a calculation of the average squared deviation. The formula used is: $1 \text{ SDCU} = n = 1/N \sum [(X_n - Y_n) / Y_n]^2$, where X_n refers to the frequency of usage of codon n in highly expressed genes, where Y_n to the frequency of usage of codon n in the gene of interest and N refers to the total number of codons in the gene of interest.

One method of optimizing the nucleic acid sequence in accordance with the preferred codon usage for a particular cell type is based on the direct use, without performing any extra statistical calculations, of codon optimization tables such as those provided on-line at the Codon Usage Database through the NIAS (National Institute of Agrobiological Sciences) DNA bank in Japan ([www\(dot\)kazusator\(dot\)jp/codon/](http://www.kazusator.jp/codon/)). The Codon Usage Database contains codon usage tables for a number of different species, with each codon usage table having been statistically determined based on the data present in Genbank.

By using the above tables to determine the most preferred or most favored codons for each amino acid in a particular species (for example, human), a naturally-occurring nucleotide sequence encoding a protein of interest can be codon optimized for that particular species. This is effected by replacing codons that may have a low statistical incidence in the particular species genome with corresponding codons, in regard to an amino acid, that are statistically more favored. However, one or more less-favored codons may be selected to delete existing restriction sites, to create new ones at potentially useful junctions (5' and 3' ends to add signal peptide or termination cassettes, internal sites that might be used to cut and splice segments together to produce a correct full-length sequence), or to eliminate nucleotide sequences that may negatively affect mRNA stability or expression.

The naturally-occurring encoding nucleotide sequence may already, in advance of any modification, contain a number of codons that correspond to a statistically-favored codon in a particular species. Therefore, codon optimization of the native nucleotide sequence may comprise determining which codons, within the native nucleotide sequence, are not statistically-favored with regards to a particular plant or mammalian species, and modifying these codons in

accordance with a codon usage table of the particular species to produce a codon optimized derivative. A modified nucleotide sequence may be fully or partially optimized for the particular species codon usage provided that the protein encoded by the modified nucleotide sequence is produced at a level higher than the protein encoded by the corresponding naturally occurring or native gene. Construction of synthetic genes by altering the codon usage is described in for example PCT Patent Application No. 93/07278.

In one embodiment, the nucleic acid sequence encoding the DNMT3 of this aspect of the present invention is codon-optimized for expression in human cells.

An example of a human codon optimized nucleic acid sequence encoding DNMT3 contemplated by the present invention is set forth in SEQ ID NO: 66.

According to a specific embodiment, the nucleic acid sequence encoding the DNMT3 of the present invention is not codon optimized for expression in gymnosperm or a bryophyte.

In addition, the nucleic acid sequence encoding the DNMT3 of the present invention is in a particular embodiment, not codon optimized for expression in a lycophyte or a charophyte.

To express the exogenous DNMT3 in a heterologous system (e.g. mammalian cells or plant cells), a polynucleotide sequence encoding the DNMT3 is preferably ligated into a nucleic acid construct suitable for cell expression in that system. Such a nucleic acid construct includes a promoter sequence for directing transcription of the polynucleotide sequence in the cell in a constitutive or inducible manner.

According to a specific embodiment, the expression vector comprises a polynucleotide encoding a catalytic domain of a species of a plant DNA methyltransferase 3 (DNMT3) protein operatively linked to a transcriptional regulatory sequence which is not of that species. Thus, for example if the expression vector encodes a DNMT3 of a gymnosperm or a bryophyte, then the present invention contemplates that the transcriptional regulatory sequence is not one which is naturally found in the gymnosperm or bryophyte (i.e. it is heterologous to the gymnosperm or bryophyte). In one embodiment, the transcriptional regulatory sequence is a sequence that induces expression in mammalian cells (e.g. CMV, SV40 or EF-1). In one embodiment, the transcriptional regulatory sequence is a sequence that induces expression in angiosperms.

Constitutive promoters suitable for use with some embodiments of the invention are promoter sequences which are active under most environmental conditions and most types of cells such as the cytomegalovirus (CMV) and Rous sarcoma virus (RSV). Inducible promoters suitable for use with some embodiments of the invention include for example the tetracycline-inducible promoter (Zabala M, et al., Cancer Res. 2004, 64(8): 2799-804).

The polynucleotides of the present invention can be inserted into nucleic acid constructs to direct expression thereof in plant cells. In one embodiment, the plant cells are not gymnosperm or bryophyte cells. In another embodiment, the plant cells are angiosperm cells.

The term "plant" as used herein encompasses whole plants, a grafted plant, ancestors and progeny of the plants and plant parts, including seeds, shoots, stems, roots (including tubers), rootstock, scion, and plant cells, tissues and organs. The plant may be in any form including suspension cultures, embryos, meristematic regions, callus tissue, leaves, gametophytes, sporophytes, pollen, and microspores. Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including a fodder or forage legume, ornamental plant, food crop, tree, or shrub selected from the list comprising *Acacia* spp., *Acer* spp., *Actinidia* spp., *Aesculus* spp., *Agathis australis*, *Albizia amara*, *Alsophila tricolor*, *Andropogon* spp., *Arachis* spp., *Areca catechu*, *Astelia fragrans*, *Astragalus cicer*, *Baikiaea plurijuga*, *Betula* spp., *Brassica* spp., *Bruguiera gymnorhiza*, *Burkea africana*, *Butea frondosa*, *Cadaba farinosa*, *Calliandra* spp., *Camellia sinensis*, *Canna indica*, *Capsicum* spp., *Cassia* spp., *Centroema pubescens*, *Chacoomeles* spp., *Cinnamomum cassia*, *Coffea arabica*, *Colophospermum mopane*, *Coronillia varia*, *Cotoneaster serotina*, *Crataegus* spp., *Cucumis* spp., *Cupressus* spp., *Cyathea dealbata*, *Cydonia oblonga*, *Cryptomeria japonica*, *Cymbopogon* spp., *Cynthea dealbata*, *Cydonia oblonga*, *Dalbergia monetaria*, *Davallia divaricata*, *Desmodium* spp., *Dicksonia squarosa*, *Dibeteropogon amplexans*, *Dioclea* spp., *Dolichos* spp., *Dorycnium rectum*, *Echinochloa pyramidalis*, *Ehaffia* spp., *Eleusine coracana*, *Eragrestis* spp., *Erythrina* spp., *Eucalyptus* spp., *Euclea schimperii*, *Eulalia villosa*, *Pagopyrum* spp., *Feijoa sellowiana*, *Fragaria* spp., *Flemingia* spp., *Freyinetia banksii*, *Geranium thunbergii*, *Ginkgo biloba*, *Glycine javanica*, *Gliricidia* spp., *Gossypium hirsutum*, *Grevillea* spp., *Guibourtia coleosperma*, *Hedysarum* spp., *Hemaphysalis altissima*, *Heteropogon contortus*, *Hordeum vulgare*, *Hyparrhenia rufa*, *Hypericum erectum*, *Hypochaeris dissoluta*, *Indigo tinctoria*, *Iris* spp., *Leptarrhena pyrolifolia*, *Lespedeza* spp., *Lettuca* spp., *Leucaena leucocephala*, *Loudetia simplex*, *Lotonotus bainesii*, *Lotus* spp., *Macrotyloma axillare*, *Malus* spp., *Manihot esculenta*, *Medicago sativa*, *Metasequoia glyptostroboides*, *Musa sapientum*, *Nicotiana* spp., *Onobrychis* spp., *Ornithopus* spp., *Oryza* spp., *Peltophorum africanum*, *Pennisetum* spp., *Persea gratissima*, *Petunia* spp., *Phaseolus* spp., *Phoenix canariensis*, *Phormium cookianum*, *Photinia* spp., *Picea glauca*, *Pinus* spp., *Pisum sativum*, *Podocarpus totara*, *Pogonarthria fleckii*, *Pogonaffhia squarrosa*, *Populus* spp., *Prosopis cineraria*, *Pseudotsuga menziesii*, *Pterolobium stellatum*, *Pyrus communis*, *Quercus* spp., *Rhaphiolepis umbellata*, *Rhopalostylis sapida*, *Rhus natalensis*, *Ribes grossularia*, *Ribes* spp.,

Robinia pseudoacacia, Rosa spp., Rubus spp., Salix spp., Schyzachyrium sanguineum, Sciadopitys vefficillata, Sequoia sempervirens, Sequoiadendron giganteum, Sorghum bicolor, Spinacia spp., Sporobolus fimbriatus, Stiburus alopecuroides, Stylosanthos humilis, Tadehagi spp, Taxodium distichum, Themeda triandra, Trifolium spp., Triticum spp., Tsuga heterophylla, 5 Vaccinium spp., Vicia spp., Vitis vinifera, Watsonia pyramidata, Zantedeschia aethiopica, Zea mays, amaranth, artichoke, asparagus, broccoli, Brussels sprouts, cabbage, canola, carrot, cauliflower, celery, collard greens, flax, kale, lentil, oilseed rape, okra, onion, potato, rice, soybean, straw, sugar beet, sugar cane, sunflower, tomato, squash tea, trees. Alternatively algae and other non-Viridiplantae can be used for the methods of some embodiments of the invention.

10 In other embodiments, the DNA methyltransferase of this aspect of the present invention is expressed in an agricultural plant. Agricultural plants include monocotyledonous species such as: maize (Zea mays), common wheat (Triticum aestivum), spelt (Triticum spelta), einkorn wheat (Triticum monococcum), emmer wheat (Triticum dicoccum), durum wheat (Triticum durum), Asian rice (Oryza sativa), African rice (Oryza glaberrima), wild rice (Zizania 15 aquatica, Zizania latifolia, Zizania palustris, Zizania texana), barley (Hordeum vulgare), Sorghum (Sorghum bicolor), Finger millet (Eleusine coracana), Proso millet (Panicum miliaceum), Pearl millet (Pennisetum glaucum), Foxtail millet (Setaria italica), Oat (Avena sativa), Triticale (Triticosecale), rye (Secale cereal), Russian wild rye (Psathyrostachys juncea), bamboo (Bambuseae), or sugarcane (e.g., Saccharum arundinaceum, Saccharum barberi, 20 Saccharum bengalense, Saccharum edule, Saccharum munja, Saccharum officinarum, Saccharum procerum, Saccharum ravennae, Saccharum robustum, Saccharum sinense, or Saccharum spontaneum); as well as dicotyledonous species such as: soybean (Glycine max), canola and rapeseed cultivars (Brassica napus), cotton (genus Gossypium), alfalfa (Medicago sativa), cassava (genus Manihot), potato (Solanum tuberosum), tomato (Solanum lycopersicum), 25 pea (Pisum sativum), chick pea (Cicer arietinum), lentil (Lens culinaris), flax (Linum usitatissimum) and many varieties of vegetables.

Constructs useful in the methods according to some embodiments of the invention may be constructed using recombinant DNA technology well known to persons skilled in the art.

30 Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include PMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin

promoter, or other promoters shown effective for expression in eukaryotic cells. Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with the fusion protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra).

The gene constructs may be inserted into vectors, which may be commercially available, suitable for transforming into plants and suitable for expression of the gene of interest in the transformed cells. The genetic construct can be an expression vector wherein said nucleic acid sequence is operably linked to one or more regulatory sequences allowing expression in the plant cells.

In a particular embodiment of some embodiments of the invention the regulatory sequence is a plant-expressible promoter.

As used herein the phrase "plant-expressible" refers to a promoter sequence, including any additional regulatory elements added thereto or contained therein, is at least capable of inducing, conferring, activating or enhancing expression in a plant cell, tissue or organ, preferably a monocotyledonous or dicotyledonous plant cell, tissue, or organ. Examples of preferred promoters useful for the methods of some embodiments of the invention are presented in Tables 1-4.

Table 1

Exemplary constitutive promoters for use in the performance of some embodiments of the invention

<i>Gene Source</i>	<i>Expression Pattern</i>	<i>Reference</i>
Actin	constitutive	McElroy et al, Plant Cell, 2: 163-171, 1990
CAMV 35S	constitutive	Odell et al, Nature, 313:

		810-812, 1985
CaMV 19S	constitutive	Nilsson et al., Physiol. Plant 100:456-462, 1997
GOS2	constitutive	de Pater et al, Plant J Nov;2(6):837-44, 1992
ubiquitin	constitutive	Christensen et al, Plant Mol. Biol. 18: 675-689, 1992
Rice cyclophilin	constitutive	Bucholz et al, Plant Mol Biol. 25(5):837-43, 1994
Maize H3 histone	constitutive	Lepetit et al, Mol. Gen. Genet. 231: 276-285, 1992
Actin 2	constitutive	An et al, Plant J. 10(1);107-121, 1996

Table 2

Exemplary seed-preferred promoters for use in the performance of some embodiments of the invention

5

Gene Source	Expression Pattern	Reference
Seed specific genes	seed	Simon, et al., Plant Mol. Biol. 5. 191, 1985; Scofield, et al., J. Biol. Chem. 262: 12202, 1987.; Baszczynski, et al., Plant Mol. Biol. 14: 633, 1990.
Brazil Nut albumin	seed	Pearson' et al., Plant Mol. Biol. 18: 235- 245, 1992.
legumin	seed	Ellis, et al., Plant Mol. Biol. 10: 203-214, 1988
Glutelin (rice)	seed	Takaiwa, et al., Mol. Gen.

		Genet. 208: 15-22, 1986; Takaiwa, et al., FEBS Letts. 221: 43-47, 1987
Zein	seed	Matzke et al Plant Mol Biol, 143).323-32 1990
napA	seed	Stalberg, et al, Planta 199: 515-519, 1996
wheat LMW and HMW, glutenin-1	endosperm	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2,
Wheat SPA	seed	Albanietal, Plant Cell, 9: 171- 184, 1997
wheat a, b and g gliadins	endosperm	EMBO3:1409-15, 1984
Barley ltrl promoter	endosperm	
barley B1, C, D hordein	endosperm	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750- 60, 1996
Barley DOF	endosperm	Mena et al, The Plant Journal, 116(1): 53- 62, 1998
Biz2	endosperm	EP99106056.7
Synthetic promoter	endosperm	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998
rice prolamin NRP33	endosperm	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice -globulin Glb-1	endosperm	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	emryo	Sato et al, Proc. Nati. Acad. Sci. USA, 93: 8117-8122
rice alpha-globulin REB/OHP-1	endosperm	Nakase et al. Plant Mol. Biol. 33: 513-S22, 1997

rice ADP-glucose PP	endosperm	Trans Res 6:157-68, 1997
maize ESR gene family	endosperm	Plant J 12:235-46, 1997
sorgum gamma- kafirin	endosperm	PMB 32:1029-35, 1996
KNOX	emryo	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
rice oleosin	Embryo and aleuton	Wu et at, J. Biochem., 123:386, 1998
sunflower oleosin	Seed (embryo and dry seed)	Cummins, et al., Plant Mol. Biol. 19: 873- 876, 1992

Table 3***Exemplary flower-specific promoters for use in the performance of the invention***

<i>Gene Source</i>	<i>Expression Pattern</i>	<i>Reference</i>
AtPRP4	flowers	www(dot)salus(dot) medium(dot)edu/m mg/tierney/html
chalcone synthase (chsA)	flowers	Van der Meer et al., Plant Mol. Biol. 15, 95- 109, 1990.
LAT52	anther	Twell et al., Mol. Gen Genet. 217:240-245 (1989)
apetala- 3	flowers	

Table 4***Alternative rice promoters for use in the performance of the invention***

<i>PRO #</i>	<i>gene</i>	<i>expression</i>
PR00001	Metallothionein Mte	transfer layer of embryo + calli
PR00005	putative beta-amylase	transfer layer of embryo
PR00009	Putative cellulose synthase	Weak in roots
PR00012	lipase (putative)	
PR00014	Transferase (putative)	
PR00016	peptidyl prolyl cis-trans isomerase (putative)	
PR00019	unknown	
PR00020	prp protein (putative)	
PR00029	noduline (putative)	
PR00058	Proteinase inhibitor Rgpi9	seed
PR00061	beta expansine EXPB9	Weak in young flowers
PR00063	Structural protein	young tissues+calli+embryo
PR00069	xylosidase (putative)	
PR00075	Prolamine 10Kda	strong in endosperm
PR00076	allergen RA2	strong in endosperm
PR00077	prolamine RP7	strong in endosperm
PR00078	CBP80	
PR00079	starch branching enzyme I	
PR00080	Metallothioneine-like ML2	transfer layer of embryo + calli
PR00081	putative caffeoyl- CoA 3-0 methyltransferase	shoot
PR00087	prolamine RM9	strong in endosperm
PR00090	prolamine RP6	strong in endosperm
PR00091	prolamine RP5	strong in endosperm
PR00092	allergen RA5	
PR00095	putative methionine	embryo

	aminopeptidase	
PR00098	ras-related GTP binding protein	
PR00104	beta expansin EXPB1	
PR00105	Glycine rich protein	
PR00108	metallothionein like protein (putative)	
PR00110	RCc3 strong root	
PR00111	uclacyanin 3-like protein	weak discrimination center / shoot meristem
PR00116	26S proteasome regulatory particle non-ATPase subunit 11	very weak meristem specific
PR00117	putative 40S ribosomal protein	weak in endosperm
PR00122	chlorophyll a/lb-binding protein precursor (Cab27)	very weak in shoot
PR00123	putative protochlorophyllide reductase	Strong leaves
PR00126	metallothionein RiCMT	strong discrimination center shoot meristem
PR00129	GOS2	Strong constitutive
PR00131	GOS9	
PR00133	chitinase Cht-3	very weak meristem specific
PR00135	alpha- globulin	Strong in endosperm
PR00136	alanine aminotransferase	Weak in endosperm
PR00138	Cyclin A2	
PR00139	Cyclin D2	
PR00140	Cyclin D3	
PR00141	Cyclophyllin 2	Shoot and seed
PR00146	sucrose synthase SS1 (barley)	medium constitutive
PR00147	trypsin inhibitor ITR1	weak in endosperm

	(barley)	
PR00149	ubiquitine 2 with intron	strong constitutive
PR00151	WSI18	Embryo and stress
PR00156	HVA22 homologue (putative)	
PR00157	EL2	
PR00169	aquaporine	medium constitutive in young plants
PR00170	High mobility group protein	Strong constitutive
PR00171	reversibly glycosylated protein RGP1	weak constitutive
PR00173	cytosolic MDH	shoot
PR00175	RAB21	Embryo and stress
PR00176	CDPK7	
PR00177	Cdc2-1	very weak in meristem
PR00197	sucrose synthase 3	
PRO0198	OsVP1	
PRO0200	OSH1	very weak in young plant meristem
PRO0208	putative chlorophyllase	
PRO0210	OsNRT1	
PRO0211	EXP3	
PRO0216	phosphate transporter OjPT1	
PRO0218	oleosin 18kd	aleurone + embryo
PRO0219	ubiquitine 2 without intron	
PRO0220	RFL	
PRO0221	maize UBI delta intron	not detected
PRO0223	glutelin-1	
PRO0224	fragment of prolamin RP6 promoter	
PRO0225	4xABRE	
PRO0226	glutelin OSGLUA3	
PRO0227	BLZ-2_short (barley)	

PR00228	BLZ-2_long (barley)	
---------	---------------------	--

In one embodiment, the promoter is not a gymnosperm or a bryophyte promoter. In another embodiment, the promoter is an angiosperm promoter.

Cells of the heterologous system (e.g. mammalian cells, or plant cells) may be transformed stably or transiently with the nucleic acid constructs of some embodiments of the invention. In stable transformation, the nucleic acid molecule of some embodiments of the invention is integrated into the genome of the organism and as such it represents a stable and inherited trait. In transient transformation, the nucleic acid molecule is expressed by the cell transformed but it is not integrated into the genome and as such it represents a transient trait.

There are various methods of introducing foreign genes into both monocotyledonous and dicotyledonous plants (Potrykus, I., *Annu. Rev. Plant. Physiol., Plant. Mol. Biol.* (1991) 42:205-225; Shimamoto et al., *Nature* (1989) 338:274-276).

The principle methods of causing stable integration of exogenous DNA into plant genomic DNA include two main approaches:

(i) *Agrobacterium*-mediated gene transfer: Klee et al. (1987) *Annu. Rev. Plant Physiol.* 38:467-486; Klee and Rogers in *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 6, *Molecular Biology of Plant Nuclear Genes*, eds. Schell, J., and Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) p. 2-25; Gatenby, in *Plant Biotechnology*, eds. Kung, S. and Arntzen, C. J., Butterworth Publishers, Boston, Mass. (1989) p. 93-112.

(ii) direct DNA uptake: Paszkowski et al., in *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 6, *Molecular Biology of Plant Nuclear Genes* eds. Schell, J., and Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) p. 52-68; including methods for direct uptake of DNA into protoplasts, Toriyama, K. et al. (1988) *Bio/Technology* 6:1072-1074. DNA uptake induced by brief electric shock of plant cells: Zhang et al. *Plant Cell Rep.* (1988) 7:379-384. Fromm et al. *Nature* (1986) 319:791-793. DNA injection into plant cells or tissues by particle bombardment, Klein et al. *Bio/Technology* (1988) 6:559-563; McCabe et al. *Bio/Technology* (1988) 6:923-926; Sanford, *Physiol. Plant.* (1990) 79:206-209; by the use of micropipette systems: Neuhaus et al., *Theor. Appl. Genet.* (1987) 75:30-36; Neuhaus and Spangenberg, *Physiol. Plant.* (1990) 79:213-217; glass fibers or silicon carbide whisker transformation of cell cultures, embryos or callus tissue, U.S. Pat. No. 5,464,765 or by the direct incubation of DNA with germinating pollen, DeWet et al. in *Experimental Manipulation of Ovule Tissue*, eds. Chapman, G. P. and Mantell, S. H. and Daniels, W. Longman, London, (1985) p. 197-209; and Ohta, *Proc. Natl. Acad. Sci. USA* (1986) 83:715-719.

The *Agrobacterium* system includes the use of plasmid vectors that contain defined DNA segments that integrate into the plant genomic DNA. Methods of inoculation of the plant tissue vary depending upon the plant species and the *Agrobacterium* delivery system. A widely used approach is the leaf disc procedure which can be performed with any tissue explant that provides a good source for initiation of whole plant differentiation. Horsch et al. in *Plant Molecular Biology Manual A5*, Kluwer Academic Publishers, Dordrecht (1988) p. 1-9. A supplementary approach employs the *Agrobacterium* delivery system in combination with vacuum infiltration. The *Agrobacterium* system is especially viable in the creation of transgenic dicotyledonous plants.

There are various methods of direct DNA transfer into plant cells. In electroporation, the protoplasts are briefly exposed to a strong electric field. In microinjection, the DNA is mechanically injected directly into the cells using very small micropipettes. In microparticle bombardment, the DNA is adsorbed on microprojectiles such as magnesium sulfate crystals or tungsten particles, and the microprojectiles are physically accelerated into cells or plant tissues.

Following stable transformation plant propagation is exercised. The most common method of plant propagation is by seed. Regeneration by seed propagation, however, has the deficiency that due to heterozygosity there is a lack of uniformity in the crop, since seeds are produced by plants according to the genetic variances governed by Mendelian rules. Basically, each seed is genetically different and each will grow with its own specific traits. Therefore, it is preferred that the transformed plant be produced such that the regenerated plant has the identical traits and characteristics of the parent transgenic plant. Therefore, it is preferred that the transformed plant be regenerated by micropropagation which provides a rapid, consistent reproduction of the transformed plants.

Micropropagation is a process of growing new generation plants from a single piece of tissue that has been excised from a selected parent plant or cultivar. This process permits the mass reproduction of plants having the preferred tissue expressing the fusion protein. The new generation plants which are produced are genetically identical to, and have all of the characteristics of, the original plant. Micropropagation allows mass production of quality plant material in a short period of time and offers a rapid multiplication of selected cultivars in the preservation of the characteristics of the original transgenic or transformed plant. The advantages of cloning plants are the speed of plant multiplication and the quality and uniformity of plants produced.

Micropropagation is a multi-stage procedure that requires alteration of culture medium or growth conditions between stages. Thus, the micropropagation process involves four basic

stages: Stage one, initial tissue culturing; stage two, tissue culture multiplication; stage three, differentiation and plant formation; and stage four, greenhouse culturing and hardening. During stage one, initial tissue culturing, the tissue culture is established and certified contaminant-free. During stage two, the initial tissue culture is multiplied until a sufficient number of tissue samples are produced to meet production goals. During stage three, the tissue samples grown in stage two are divided and grown into individual plantlets. At stage four, the transformed plantlets are transferred to a greenhouse for hardening where the plants' tolerance to light is gradually increased so that it can be grown in the natural environment.

Although stable transformation is presently preferred, transient transformation of leaf cells, meristematic cells or the whole plant is also envisaged by some embodiments of the invention.

Transient transformation can be effected by any of the direct DNA transfer methods described above or by viral infection using modified plant viruses.

Viruses that have been shown to be useful for the transformation of plant hosts include CaMV, TMV and BV. Transformation of plants using plant viruses is described in U.S. Pat. No. 4,855,237 (BGV), EP-A 67,553 (TMV), Japanese Published Application No. 63-14693 (TMV), EPA 194,809 (BV), EPA 278,667 (BV); and Gluzman, Y. et al., Communications in Molecular Biology: Viral Vectors, Cold Spring Harbor Laboratory, New York, pp. 172-189 (1988). Pseudovirus particles for use in expressing foreign DNA in many hosts, including plants, is described in WO 87/06261.

Construction of plant RNA viruses for the introduction and expression of non-viral exogenous nucleic acid sequences in plants is demonstrated by the above references as well as by Dawson, W. O. et al., Virology (1989) 172:285-292; Takamatsu et al. EMBO J. (1987) 6:307-311; French et al. Science (1986) 231:1294-1297; and Takamatsu et al. FEBS Letters (1990) 269:73-76.

When the virus is a DNA virus, suitable modifications can be made to the virus itself. Alternatively, the virus can first be cloned into a bacterial plasmid for ease of constructing the desired viral vector with the foreign DNA. The virus can then be excised from the plasmid. If the virus is a DNA virus, a bacterial origin of replication can be attached to the viral DNA, which is then replicated by the bacteria. Transcription and translation of this DNA will produce the coat protein which will encapsidate the viral DNA. If the virus is an RNA virus, the virus is generally cloned as a cDNA and inserted into a plasmid. The plasmid is then used to make all of the constructions. The RNA virus is then produced by transcribing the viral sequence of the

plasmid and translation of the viral genes to produce the coat protein(s) which encapsidate the viral RNA.

Construction of plant RNA viruses for the introduction and expression in plants of non-viral exogenous nucleic acid sequences such as those included in the construct of some
5 embodiments of the invention is demonstrated by the above references as well as in U.S. Pat. No. 5,316,931.

In one embodiment, a plant viral nucleic acid is provided in which the native coat protein coding sequence has been deleted from a viral nucleic acid, a non-native plant viral coat protein coding sequence and a non-native promoter, preferably the subgenomic promoter of the non-
10 native coat protein coding sequence, capable of expression in the plant host, packaging of the recombinant plant viral nucleic acid, and ensuring a systemic infection of the host by the recombinant plant viral nucleic acid, has been inserted. Alternatively, the coat protein gene may be inactivated by insertion of the non-native nucleic acid sequence within it, such that a protein is produced. The recombinant plant viral nucleic acid may contain one or more additional non-
15 native subgenomic promoters. Each non-native subgenomic promoter is capable of transcribing or expressing adjacent genes or nucleic acid sequences in the plant host and incapable of recombination with each other and with native subgenomic promoters. Non-native (foreign) nucleic acid sequences may be inserted adjacent the native plant viral subgenomic promoter or the native and a non-native plant viral subgenomic promoters if more than one nucleic acid
20 sequence is included. The non-native nucleic acid sequences are transcribed or expressed in the host plant under control of the subgenomic promoter to produce the desired products.

In a second embodiment, a recombinant plant viral nucleic acid is provided as in the first embodiment except that the native coat protein coding sequence is placed adjacent one of the non-native coat protein subgenomic promoters instead of a non-native coat protein coding
25 sequence.

In a third embodiment, a recombinant plant viral nucleic acid is provided in which the native coat protein gene is adjacent its subgenomic promoter and one or more non-native subgenomic promoters have been inserted into the viral nucleic acid. The inserted non-native subgenomic promoters are capable of transcribing or expressing adjacent genes in a plant host and are incapable of recombination with each other and with native subgenomic promoters.
30 Non-native nucleic acid sequences may be inserted adjacent the non-native subgenomic plant viral promoters such that said sequences are transcribed or expressed in the host plant under control of the subgenomic promoters to produce the desired product.

In a fourth embodiment, a recombinant plant viral nucleic acid is provided as in the third embodiment except that the native coat protein coding sequence is replaced by a non-native coat protein coding sequence.

The viral vectors are encapsidated by the coat proteins encoded by the recombinant plant viral nucleic acid to produce a recombinant plant virus. The recombinant plant viral nucleic acid or recombinant plant virus is used to infect appropriate host plants. The recombinant plant viral nucleic acid is capable of replication in the host, systemic spread in the host, and transcription or expression of foreign gene(s) (isolated nucleic acid) in the host to produce the desired protein.

The present application further contemplates fusion proteins comprising the plant DNMT3s linked to DNA targeting moieties.

According to a particular embodiment, the DNA targeting moiety is a DNA endonuclease protein.

Contemplated endonuclease proteins include RNA-guided DNA endonuclease enzyme including for example zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR associated protein.

In particular embodiments, the RNA-guided DNA endonuclease is Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 or Cpf1 endonuclease.

In one embodiment, the DNA targeting moiety comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.

Cas9

Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While the *S. pyogenes* and *S. thermophilus* Cas9 molecules are exemplified herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed in US Patent Application No. 20160010076 can be used as well. Additional Cas9 proteins are described in Esvelt et al., Nat Methods. 2013 November; 10(11):1116-21 and Fonfara et al., "Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems." Nucleic Acids Res. 2013 Nov. 22. doi:10.1093/nar/gkt1074.

The constructs and methods described herein can include the use of any of those Cas9 proteins, and their corresponding guide RNAs or other guide RNAs that are compatible. The Cas9 from *Streptococcus thermophilus* LMD-9 CRISPR1 system has been shown to function in

human cells in Cong et al (Science 339, 819 (2013)). Additionally, Jinek et al. showed in vitro that Cas9 orthologs from *S. thermophilus* and *L. innocua*, can be guided by a dual *S. pyogenes* gRNA to cleave target plasmid DNA, albeit with slightly decreased efficiency.

In some embodiments, the present system utilizes the Cas9 protein from *S. pyogenes*, either as encoded in bacteria or codon-optimized for expression in mammalian cells (e.g. human cells) or plant cells, containing mutations at D10, E762, H983, or D986 and H840 or N863, e.g., D10A/D10N and H840A/H840N/H840Y, to render the nuclease portion of the protein catalytically inactive; substitutions at these positions could be alanine (as they are in Nishimasu al., Cell 156, 935-949 (2014)) or they could be other residues, e.g., glutamine, asparagine, tyrosine, serine, or aspartate, e.g., E762Q, H983N, H983Y, D986N, N863D, N863S, or N863H. The sequence of the catalytically inactive *S. pyogenes* Cas9 that can be used in the methods and compositions described herein is as set forth in SEQ ID NO: 23.

In some embodiments, the Cas9 nuclease used herein is at least about 50% identical to the sequence of *S. pyogenes* Cas9, i.e., at least 50% identical to SEQ ID NO: 23. In some embodiments, the nucleotide sequences are about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO: 23.

In some embodiments, the catalytically inactive Cas9 used herein is at least about 50% identical to the sequence of the catalytically inactive *S. pyogenes* Cas9, i.e., at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO:24, wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

In some embodiments, any differences from SEQ ID NO:23 are in non-conserved regions, as identified by sequence alignment of sequences set forth in Chylinski et al., RNA Biology 10:5, 1-12; 2013; Esvelt et al., Nat Methods. 2013 November; 10(11):1116-21 and Fonfara et al., Nucl. Acids Res. (2014) 42 (4): 2577-2590, and wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

An exemplary nucleic acid sequence of human-codon optimized Cas9 is set forth in SEQ ID NO: 67.

The catalytic domain of the DNMT3 (or the full length DNMT3) may be linked directly to the DNA endonuclease protein or via a peptide linker.

The linker may comprise amino acids linked together by peptide bonds which serve as spacers such that the linker does not interfere with the biological activity of the fusion protein. The linker is preferably made up of amino acids linked together by peptide bonds. Thus, in preferred embodiments, the linker is made up of from 1 to 10 amino acids linked by peptide bonds, wherein the amino acids are selected from the 20 naturally occurring amino acids. Some

of these amino acids may be glycosylated, as is well understood by those in the art. In a particular embodiment, the amino acids in the linker are selected from glycine, alanine, proline, asparagine, glutamine, and lysine. Even more preferably, the linker is made up of a majority of amino acids that are sterically unhindered, such as glycine and alanine.

5 In one embodiment, the peptide linker is between 2 and 60 amino acids, between 2 and 50 amino acids, between 2 and 40 amino acids, between 2 and 30 amino acids, between 2 and 20 amino acids or even between 2 and 10 amino acids.

The DNMT3 may be linked to the C terminus of the endonuclease protein or the N terminus of said endonuclease protein.

10 The fusion proteins described herein may be provided as a kit together with particular guide RNAs (gRNAs).

The gRNA comprises a "gRNA guide sequence" or "gRNA target sequence" which corresponds to the target sequence on a target polynucleotide gene sequence.

The gRNA may comprise a "G" at the 5' end of the polynucleotide sequence. The presence of a "G" in 5' is preferred when the gRNA is expressed under the control of the U6 promoter. The gRNA may be of varying lengths. The gRNA may comprise at least a 10 nts, at least 11 nts, at least a 12 nts, at least a 13 nts, at least a 14 nts, at least a 15 nts, at least a 16 nts, at least a 17 nts, at least a 18 nts, at least a 19 nts, at least a 20 nts, at least a 21 nts, at least a 22 nts, at least a 23 nts, at least a 24 nts, at least a 25 nts, at least a 30 nts, or at least a 35 nts of the target caspase 6 DNA sequence which is followed by a PAM sequence. The "gRNA guide sequence" or "gRNA target sequence" may be at least 17 nucleotides (17, 18, 19, 20, 21, 22, 23), preferably between 17 and 30 nts long, more preferably between 18-22 nucleotides long. In an embodiment, gRNA guide sequence is between 10-40, 10-30, 12-30, 15-30, 18-30, or 10-22 nucleotides long. The PAM sequence may be "NGG", where "N" can be any nucleotide. gRNA may target any region of a target gene which is immediately upstream (contiguous, adjoining, in 5') to a PAM (e.g., NGG) sequence.

Although a perfect match between the gRNA guide sequence and the DNA sequence on the targeted gene is preferred, a mismatch between a gRNA guide sequence and target sequence on the gene sequence of interest is also permitted as long as it still allows hybridization of the gRNA with the complementary strand of the gRNA target polynucleotide sequence on the targeted gene. A seed sequence of between 8-12 consecutive nucleotides in the gRNA, which perfectly matches a corresponding portion of the gRNA target sequence is preferred for proper recognition of the target sequence. The remainder of the guide sequence may comprise one or more mismatches. In general, gRNA activity is inversely correlated with the number of

mismatches. Preferably, the gRNA of the present invention comprises 7 mismatches, 6 mismatches, 5 mismatches, 4 mismatches, 3 mismatches, more preferably 2 mismatches, or less, and even more preferably no mismatch, with the corresponding gRNA target gene sequence (less the PAM). Preferably, the gRNA nucleic acid sequence is at least 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % and 99 % identical to the gRNA target polynucleotide sequence in the gene of interest. Of course, the smaller the number of nucleotides in the gRNA guide sequence the smaller the number of mismatches tolerated. The binding affinity is thought to depend on the sum of matching gRNA-DNA combinations.

Any gRNA guide sequence can be selected in the target nucleic acid sequence, as long as it allows introducing at the proper location, the patch/donor sequence of the present invention. Accordingly, the gRNA guide sequence or target sequence of the present invention may be in coding or non-coding regions a gene (i.e., introns or exons).

In one embodiment, the gRNA is a sgRNA.

As used herein, the term "sgRNA" refers to single guide RNA used in conjunction with CRISPR associated systems (Cas). sgRNAs are a fusion of crRNA and tracrRNA and contain nucleotides of sequence complementary to the desired target site. Jinek et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity" Science 337(6096):816-821 (2012) Watson-Crick pairing of the sgRNA with the target site permits R-loop formation, which in conjunction with a functional PAM permits DNA cleavage or in the case of nuclease-deficient Cas9 allows binds to the DNA at that locus.

Modified RNA oligonucleotides such as locked nucleic acids (LNAs) have been demonstrated to increase the specificity of RNA-DNA hybridization by locking the modified oligonucleotides in a more favorable (stable) conformation. For example, 2'-O-methyl RNA is a modified base where there is an additional covalent linkage between the 2' oxygen and 4' carbon which when incorporated into oligonucleotides can improve overall thermal stability and selectivity.

Thus, the gRNAs disclosed herein may comprise one or more modified RNA oligonucleotides. For example, the truncated guide RNAs molecules described herein can have one, some or all of the region of the guide RNA complementary to the target sequence are modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methylpseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In other embodiments, one, some or all of the nucleotides of the gRNA sequence may be modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-

pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In some embodiments, the single guide RNAs and/or crRNAs and/or tracrRNAs can include one or more Adenine (A) or Uracil (U) nucleotides on the 3' end.

5 The guide RNA may be provided per se or in an expression vector. The vectors for expressing the guide RNAs can include RNA Pol III promoters to drive expression of the guide RNAs, e.g., the H1, U6 or 7SK promoters. These human promoters allow for expression of gRNAs in mammalian cells following plasmid transfection. Alternatively, a T7 promoter may be used, e.g., for in vitro transcription, and the RNA can be transcribed in vitro and purified. Vectors
10 suitable for the expression of short RNAs, e.g., siRNAs, shRNAs, or other small RNAs, can be used.

According to another aspect of the present invention there is provided a method of increasing methylation of DNA in a cell, the method comprising expressing a polynucleotide encoding a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein in the cell,
15 thereby increasing methylation of DNA in the cell. In one embodiment, the cell is not of a gymnosperm plant.

In one embodiment, the catalytic domain is introduced into the cell as a fusion protein (e.g. linked to a DNA targeting moiety, as described herein above).

Together with the fusion proteins of the present invention (or polynucleotides encoding
20 same), the gRNAs may be introduced into a wide variety of cell types, embryos at different developmental stages, tissues and species may be targeted, including somatic and embryonic stem cells of human and animal models. In one embodiment, the cell is a stem cell (e.g. a pluripotent stem cell such as an embryonic stem cell or an induced pluripotent stem cell), a mesenchymal stem cell, a tissue stem cell (e.g. a neuronal stem cell or muscle stem cell). In
25 another embodiment, the cell is a healthy cell. In another embodiment, the cell is a diseased cell (e.g., a cancer cell).

In other embodiments the fusion protein (and gRNA) may be injected into the cell. This is particularly relevant for editing of single cells, eggs or embryonic stem cells.

Following introduction of the fusion protein and gRNA described herein, the gene (at the
30 targeted site) may be analyzed to ensure (i.e. confirm) that methylation has occurred. Thus, for example bisulfite sequencing may be carried out to determine the extent of methylation prior to and/or following the treatment.

Bisulfite sequencing (also known as bisulphite sequencing) is the use of bisulfite treatment of DNA to determine its pattern of methylation.

Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected. Therefore, DNA that has been treated with bisulfite retains only methylated cytosines. Thus, bisulfite treatment introduces specific changes in the DNA sequence that depend on the methylation status of individual cytosine residues, yielding single-nucleotide resolution information about the methylation status of a segment of DNA. Various analyses can be performed on the altered sequence to retrieve this information. The objective of this analysis is therefore reduced to differentiating between single nucleotide polymorphisms (cytosines and thymidine) resulting from bisulfite conversion.

As described in Example 3, introduction of the DNMT3s of the present invention can lead to alteration in expression levels of particular genes. Thus, the present inventors further contemplate analyzing the expression level of relevant genes to uncover the effect methylation has on gene expression.

It is envisaged by the present inventors that enhancement of methylation at particular sites may aid in treating diseases which are associated with hypo-methylation. Such diseases include for example autoimmune diseases (multiple sclerosis, rheumatoid arthritis, lupus, metabolic disorders (diabetes, lipid related disorders, obesity), neurological disorders (autism, Parkinson's disease) and aging (see for example Jin and Liu [Genes Dis. 2018 Mar; 5(1): 1–8], the contents of which are incorporated herein by reference.

It is expected that during the life of a patent maturing from this application many relevant plant DNMT3s will be uncovered and the scope of the term plant DNMT3 is intended to include all such new technologies *a priori*.

As used herein the term "about" refers to $\pm 10\%$

The terms "comprises", "comprising", "includes", "including", "having" and their conjugates mean "including but not limited to".

The term "consisting of" means "including and limited to".

The term "consisting essentially of" means that the composition, method or structure may include additional ingredients, steps and/or parts, but only if the additional ingredients, steps and/or parts do not materially alter the basic and novel characteristics of the claimed composition, method or structure.

As used herein, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a compound" or "at least one compound" may include a plurality of compounds, including mixtures thereof.

Throughout this application, various embodiments of this invention may be presented in a range format. It should be understood that the description in range format is merely for

convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

Whenever a numerical range is indicated herein, it is meant to include any cited numeral (fractional or integral) within the indicated range. The phrases “ranging/ranges between” a first indicate number and a second indicate number and “ranging/ranges from” a first indicate number “to” a second indicate number are used herein interchangeably and are meant to include the first and second indicated numbers and all the fractional and integral numerals therebetween.

As used herein the term "method" refers to manners, means, techniques and procedures for accomplishing a given task including, but not limited to, those manners, means, techniques and procedures either known to, or readily developed from known manners, means, techniques and procedures by practitioners of the chemical, pharmacological, biological, biochemical and medical arts.

As used herein, the term “treating” includes abrogating, substantially inhibiting, slowing or reversing the progression of a condition, substantially ameliorating clinical or aesthetical symptoms of a condition or substantially preventing the appearance of clinical or aesthetical symptoms of a condition.

When reference is made to particular sequence listings, such reference is to be understood to also encompass sequences that substantially correspond to its complementary sequence as including minor sequence variations, resulting from, e.g., sequencing errors, cloning errors, or other alterations resulting in base substitution, base deletion or base addition, provided that the frequency of such variations is less than 1 in 50 nucleotides, alternatively, less than 1 in 100 nucleotides, alternatively, less than 1 in 200 nucleotides, alternatively, less than 1 in 500 nucleotides, alternatively, less than 1 in 1000 nucleotides, alternatively, less than 1 in 5,000 nucleotides, alternatively, less than 1 in 10,000 nucleotides.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination or as suitable in any other described embodiment of the invention. Certain

features described in the context of various embodiments are not to be considered essential features of those embodiments, unless the embodiment is inoperative without those elements.

Various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below find experimental support in the following examples.

EXAMPLES

Reference is now made to the following examples, which together with the above descriptions illustrate some embodiments of the invention in a non-limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Culture of Animal Cells - A Manual of Basic Technique" by Freshney, Wiley-Liss, N. Y. (1994), Third Edition; "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA (1990); Marshak et al., "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are

incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

EXAMPLE 1

MATERIALS AND METHODS

Biological Materials

All mutant plants were generated in the background of 'Gransden 2004' strain of *P. patens*^{50,59} and were propagated on BCD or BCDAT media⁶⁰ at 25 °C under a 16 h light and 8 h dark regime⁶¹. Plant morphology was documented as previously described⁶².

Generation of transgenic mutant lines

P. patens single deletion mutant lines for the following genes: *PpDRM1* (Pp3c15_14360V1.1), *PpDRM2* (Pp3c15_21430V1.1), *PpDNMT3a* (Pp3c3_3540V1.1) and *PpDNMT3b* (Pp3c13_8320V1.1) were generated by replacing the genomic region coding for the methyltransferase domain with either the hygromycin resistance cassette (*hptII*) or the G418 resistance cassette (*nptII*) *via* homologous recombination (illustrated in Figure 6). Genomic fragments corresponding to the 5' and 3' flanking regions of the deleted sequence were amplified using KOD hot start DNA polymerase (Novagen), cloned into the pTZ57 vector (Fermentas) and sequenced to validate their integrity. Next, the 5' and 3' fragments were subcloned into either the pMBL5 vector ([GenBank](#): DQ228130.1) or the pMBL5 Nos Hyg vector³⁶. Constructs were introduced into protoplasts *via* PEG-mediated transformation as described⁶⁰ using 15 µg of plasmid restricted to linearize the construct. Six days after regeneration, transformants were selected on BCDAT medium containing 25 µg/ml hygromycin (Duchefa) or 25 µg/ml G418 (calbiochem). Resistant plants were further tested by tissue PCR³⁶ to verify correct integration of the construct into the genome, by amplifying the junction regions between the insert and the sequence flanking the deleted fragment at both the 5' and 3' ends (primers listed in Table 6). In addition, loss of the endogenous targeted loci was correlated with lack of amplification of the targeted sequence as compared to a positive control. $\Delta Ppdrm2$ and $\Delta Ppdnmt3a$ single deletion mutant protoplasts were used to generate $\Delta Ppdrm\Delta Ppdrm2$ and $\Delta Ppdnmt3a\Delta Ppdnmt3b$ double deletion mutant lines, respectively, as described above.

Generation of RPS transgenic lines

The RPS transgene was introduced into the genome of WT and mutant plants *via* non-homologous recombination. To this end, a pMBL5+Zeo vector was constructed by subcloning

the Zeocin resistance cassette (*Sh ble* gene) from pRT101-Zeo⁶³ replacing the G418 resistance cassette (*nptII* gene) of the pMBL5 vector (GenBank: DQ228130.1). The RPS fragment was subcloned from the p35 GUS/RPS vector³⁸ into pMBL5+Zeo vector. Both the RPS and Zeocin resistance cassettes were sequenced in the final pMBL5+Zeo+RPS construct to ensure integrity.

5 Following transformation (as described above) and selection on BCDAT medium containing 50 µg/ml Zeocin (Invivogen), resistant plants were tested to verify insertion of the construct into the genome by tissue PCR³⁶ amplifying an internal transgene sequence spanning both the RPS sequence and the selection cassette (primers listed in Table 6).

Search of RPS homologues sequences in *P. patens* genome and sRNAome

10 The RPS sequence (GenBank: X92381.1) was used for homology search (blastn) in the *P. patens* V3.0 genome⁶⁴. Additionally, it was used to search for corresponding small RNAs by NCBI SRA-Blast⁶⁵ using small RNA-seq data of *P. patens* protonema⁴⁰ (SRX247005-SRX247008 and SRX327325-SRX327330).

Published Genomic Data

15 Data for sRNA were derived from⁴⁰, for mRNA from⁵, and for histone modifications from⁴³.

Bisulfite sequencing of the RPS transgene

A fragment of RPS was PCR amplified from bisulfite treated genomic DNA, extracted from protonema tissue, using primers RPS-top-R-new and RPS-top-F (primers listed in Table 6) and KAPA HiFi Uracil+ polymerase (kappa biosystems), then cloned into pJET1.2 (Thermo Fisher Scientific). The methylation status of individual clones was determined by Sanger sequencing.

Phylogenetic Analysis

PpDNMT3b and PpDRM2 protein sequences were used to search for homologs by blastp versus NCBI Non-redundant protein database⁶⁶ and by tblastn versus the 1000 plants (1kp) transcriptome database⁶⁷⁻⁷⁰. Alignment of selected DNMT3, DRM and DNMT1 MTD protein sequences was performed using MUSCLE v3.8.31⁷¹. The motif order was rearranged in DRM sequences to match the linear organization of canonical DNMTs. Protein accessions are listed in Table 5. MTDs of animal and plant DNMT1/MET1 homologs were added as an outgroup. The phylogenetic tree was constructed by IQ-TREE v1.6.4⁷²⁻⁷⁴ using default parameters and illustrated by FigTree v1.4.3 ([www\(dot\)tree\(dot\)bio\(dot\)edac\(dot\)uk/software/figtree/](http://www(dot)tree(dot)bio(dot)edac(dot)uk/software/figtree/)).

BS-seq library preparation

About 500 ng of genomic DNA isolated from protonema was fragmented by sonication, end repaired, and ligated to custom synthesized methylated adapters (Eurofins MWG Operon)

according to the manufacturer's (Illumina) instructions for gDNA library construction. Adaptor-ligated libraries were subjected to two successive treatments of sodium bisulfite conversion using the EpiTect Bisulfite kit (QIAGEN) as outlined in the manufacturer's instructions. The bisulfite-converted libraries were then amplified by PCR using the following conditions: 2.5 U of ExTaq DNA polymerase (Takara Bio), 5 µl of 10x Extaq reaction buffer, 25 mM dNTPs, 1 µl Primer 1.1, 1 µl Primer 2.1 (50 µl final). PCR reactions were carried out as follows: 95°C for 3 min, then 12–14 cycles of 95°C for 30 s, 65°C for 30 s, and 72°C for 60 s. The enriched libraries were either gel purified (~300 bp band) or purified with the solid-phase reversible immobilization method using AM-Pure beads (Beckman Coulter) prior to quantification with a Bioanalyzer (Agilent). Deep sequencing was performed on Illumina Hi-Seq 2000.

BS-seq data analysis

BS-seq data processing was performed as described⁵⁴. In short, custom Perl scripts were used to convert all the Cs in the 'forward' reads (and in the scaffold) to Ts, and all the Gs in the 'reverse' reads and scaffold to As. The converted reads were aligned to the converted scaffold using Bowtie2 aligner v2.3.2⁷⁵. Perl scripts were used to recover the original sequence information and, for each C (on either strand), count the number of times it was sequenced as a C or a T. For each sequence context (CG, CHG, CHH) the genomic averaged fractional methylation was calculated (Table 7 and Figures 3A-G), as well as fractional methylation within a 50 bp sliding window that were used in downstream analyses.

TE frequency meta-analysis

The abundance of TEs near TSSs of *P. patens* and *A. thaliana* genes was assessed using publicly available genes and TEs annotations and a custom Perl script, which creates a histogram of scores relative to edges of entries from one annotation file based on the presence of entries from another annotation file. Gene annotations (v3.3 for *P. patens*, Araport11 for *Arabidopsis*) and *A. thaliana* TE annotation (TAIR10) were downloaded from [www\(dot\)phytozome\(dot\)org](http://www.phytozome.org). *P. patens* TE annotation⁴² was downloaded from [www\(dot\)genomeevolution\(dot\)org](http://www.genomeevolution.org). TE annotations were reformatted to contain separate entries for start and end positions of each TE, and to assign each entry a score of 1. For each gene, the presence of TE edge was tested in a 25 bp sliding window up to 500 bp upstream to TSS, assigning "positive" windows with scores. In order to count only one edge of a TE closest to each gene, this analysis was performed separately on TEs "ends" against genes on the plus strand, and vice versa. Then, genes were aligned at TSS, and the percentage of genes with a TE ending in each 25 bp window were calculated.

Percent methylation change

This number was calculated by dividing the difference in methylation level between two samples by the level of methylation in the sample with the higher methylation level. For example, percent-methylation-change between WT and cmt was calculated as follows:

$$\begin{aligned} & \frac{WT\ mCHH - cmt\ mCHH}{WT\ mCHH} * 100\ if\ WT\ mCHH > cmt\ mCHH \\ & - \frac{cmt\ mCHH - WT\ mCHH}{cmt\ mCHH} * 100\ if\ WT\ mCHH < cmt\ mCHH \end{aligned}$$

Box plots

Box plots compare percent-methylation-change within 50-bp windows with CHH methylation level of at least 0.1 in either of the samples, and with at least 20 informative sequenced cytosines. To examine the correlation between methylation change and chromatin structure, TE windows are separated into centiles in ascending order according to siRNA (24nt sRNA), GC ratio, H3K9me2, H3K4me3, TE size, and TE LTR/INT annotations. GC ratio and TE size were divided into five centiles. siRNA counts were divided into 10 centiles, which due to the high abundance of score 1 windows, only centile 1, 7, 9, and 10 are showing. H3K9me2 and H3K4me3 are Log2 ratio over total H3 that were divided into four centiles. For H3K9me2 and H3K4me3, an additional category, ND, was added, that corresponded to windows that did not have any signal in either H3K9me2, H3K4me3 or H3.

Identification of DMRs

Fractional methylation in 50 bp windows across the genome was compared between WT and each of the DNMT mutants. DMRs were called for windows with at least 0.1 fractional methylation, 10 informative sequenced cytosines, and Fisher's exact test p-value < 0.05.

Accession Numbers

Sequencing data have been deposited in Gene Expression Omnibus under accession number GSE116837.

RESULTS

DNMT3s are persistent in plants and are evolutionary distinct from DRMs

P. patens encodes two DNMT3s, designated here as PpDNMT3a and PpDNMT3b, which are composed of a DNMT3-type N-terminal MTD and a C-terminal domain of unknown function 3444 (DUF3444)¹⁸. Genome and transcriptome searches revealed that the protein organization is conserved among non-flowering streptophytes DNMT3s. The existence of two-full length DNMT3 homologs in two distantly-related gymnosperm subclasses that were separated around

300 million years ago implies that DNMT3 persists in gymnosperms. DNMT3 was not detected in any available angiosperm genomes or transcriptomes, supporting the notion that DNMT3 completely disappeared from this plant lineage. Phylogenetic analysis of the MTD show that plant DNMT3s form a monophyletic clade together with animal DNMT3s which is separated from the DRM clade (Figure 1A), suggesting the functional conservation of DNMT3s among plants and animals and/or functional speciation between plant DNMT3 and DRM proteins. Additionally, while DRM paralogs are common along plant evolution, they diverged into distinct orthologs only in seed plants, e.g. DRM2 and DRM3 in angiosperm (Figure 1A), implying further functional diversification of DRMs in this plant lineage. Paralogs of plant DNMT3s are also common, however based on our evolutionary analysis, these duplications did not evolve into conserved DNMT3 ortholog families across multiple species (Figure 1A). Of note, PpDNMT3a and PpDNMT3b are not orthologs of mammalian DNMT3a and PpDNMT3b, respectively (Figure 1A). Similarly, PpDRM1 and PpDRM2 are not orthologs of angiosperm DRM1 and DRM2, respectively (Figure 1A). In summary, while DRMs are commonly considered as the plant homologs of eukaryotic DNMT3, here it is shown that DRMs are evolutionary distinct from DNMT3, and that true DNMT3 plant homologs exist throughout the plant kingdom, except in angiosperm.

PpDNMT3b and PpCMT establish *de novo* methylation and maintain the non-CG methylome

To determine the role of *P. patens* DNMTs in DNA methylation, the methylomes of *P. patens* DNMT deletion mutant plants, namely *met*, *cmt*, *dnmt3a*, *dnmt3b*, *drm1*, and *drm2* single deletion mutants, as well as in *drm1/drm2* (*drm12*) and *dnmt3a/dnmt3b* (*dnmt3ab*) double deletion mutants^{35,36}; (Figure 6) were profiled. All single and double *DRM* and *DNMT3* mutants were viable and developed similarly to wild type (WT) (Figure 7). Genomic methylation averages clearly showed that CG, CHG, and CHH sites were nearly eliminated and specifically disrupted in *met*, *cmt*, and *dnmt3b* mutants, respectively (Figure 1B). More precisely, *met* mutant lost 93% of CG methylation, *cmt* mutant lost 97% of CHG methylation, and *dnmt3b* mutant lost 95% of CHH methylation (Figure 1B and Table 7). The *dnmt3ab* double mutant lost 95% of CHH methylation, which is comparable to the CHH loss in *dnmt3b* single mutant. Neither *dnmt3a*, *drm1*, *drm2* single mutants nor *drm12* double mutant showed any significant global hypo-methylation in any of the sequence contexts (Figure 1B). These complete and specific hypo-methylations in *P. patens* DNMT mutants led to the conclusion that CG, CHG, and CHH

contexts in *P. patens* are directly and primarily methylated by *PpMET*, *PpCMT*, and *PpDNMT3b*, respectively.

Testing for the loss of preexisting methylation in mutant backgrounds accounts for maintenance methylation

To evaluate the activity of *P. patens* DNMTs in *de novo* methylation, the repetitive DNA sequence (RPS) from *Petunia hybrida*^{37–39}, uncommon to moss, was introduced into *P. patens*. DNA methylation analysis of RPS was conducted in the first transgenic generation (T1) and within the same transformed plant tissue. Using bisulfite sequencing, it was found that RPS is methylated in WT cells in all three methylation contexts, CG, CHG, and CHH (Figure 1C), implying that it can be *de novo* methylated in *P. patens*. By introducing and examining RPS methylation in the various DNMT mutants, it was found that CG methylation is significantly reduced in *met*, *dnmt3b* and *dnmt3ab* mutants (paired t-test p-value < 0.0016, 0.0023, 0.0022, respectively), CHG methylation is specifically and significantly reduced in *cmt* mutants (paired t-test p-value < 10⁻⁵), and CHH methylation is eliminated in *dnmt3b* and *dnmt3ab* mutants (paired t-test p-value < 10⁻⁵ for both) while unchanged in *dnmt3a*. In *drm12* mutant, RPS was methylated same as in WT. In angiosperms, DRMs are directed to the DNA by 24nt small interfering RNA (siRNA)²⁷. Accordingly, the ability of RPS to undergo *de novo* methylation in *P. patens* plants mutated in the *RNA Directed RNA polymerase 2* (*PpRDR2*) and subsequently depleted of siRNA⁴⁰ was tested. Similarly to *drm12*, it was found that RPS is regularly methylated in *P. patens rdr2* mutant plants.

Altogether, these context-specific RPS methylation phenotypes in each of the mutants suggest that *de novo* methylation in *P. patens* can be mediated by DNMT3b at CG and CHH sites and by CMT at CHG sites without the involvement of DRMs or the canonical RdDM pathway. The reduction of CG methylation in RPS DNA in *met* T1 plants suggests that *de novo* CG methylation of RPS is relied also on *PpMET*. Alternatively, CG hypomethylation in *met* mutant could suggest that CG methylation in RPS is dependent on *PpMET* maintenance activity within just a few rounds of somatic cell generations.

PpDNMT3b and PpCMT regulate genomic CG methylation

The near-complete elimination of CG methylation in the *met* genome (Figure 1B) suggests that unlike animal DNMT3, *PpDNMT3s* do not play a role in maintaining genomic CG methylation. However, by focusing on transposable elements (TEs), a consistent decrease of 13% in CG methylation in both single *dnmt3b* and double *dnmt3ab* mutants was found (Figure 2A), suggesting that DNMT3b is partially involved in maintaining the CG methylome. Further dissection of CG methylation based on their neighboring 5' nucleotides, i.e. NCG sites (N = any

nucleotide), revealed that ACG sites are preferentially hypomethylated in *dnmt3b* and *dnmt3ab* (Figure 2B). In association with the particular ACG hypo-methylation in *dnmt3b* plants, it was found that in *met* mutants, ACG sites exhibit the highest residual CG methylation levels (Figure 2c).

Among the four NCG sites, CCGs had the lowest CG-hypomethylated effect in *dnmt3b* mutants (Figure 2B). CCG is one form of CHG for which it was previously shown that its methylation (mCCG) in the entire *Arabidopsis* genome and a couple of examined sequences in *P. patens*, is dependent on the methylation of the internal CG site (CmCG) regulated by MET1 genes³⁶. Here, this observation was extended to the entire *P. patens* genome by showing that CHG methylation, specifically at CCG sites, is diminished in the *met* mutant (Figure 2E). This contributed to a 13% reduction in CHG methylation at TE sequences (Figure 2D). Interestingly, it was found that the reciprocate effect also exists, i.e. CmCG dependency on mCCG. Out of the four NmCG methylation contexts, CmCG is particularly reduced in the *cmt* mutant (Figure 2B), while in *met* mutant CmCG residual level is second to ACG (Figure 2C). Accordingly, along with their *de novo* methylation activities these results demonstrate the ability of PpCMT and PpDNMT3b to bring about CG methylation at genomic CCG and DCGs (D= A, G, or T) sites, respectively.

Non-CG methylation by mammalian DNMT3 is targeted preferentially to CW sites (W= A or T), such as CAC and CAG²⁴. Herein, it was found that CHH methylation (mediated by PpDNMT3b) preferentially targets CWH sites (Figure 8), suggesting functional conservation of CW methylation between mammalian and moss DNMT3s. However, the particular regulation of CHG methylation (including CWG) by PpCMT (Figure 2E), infers diversification of PpDNMT3b by avoiding methylating CWG sites that are controlled solely by PpCMT.

PpDNMT3b-dependent CHH methylation preferentially targets heterochromatin and is partially regulated by PpCMT

DNA methylation in *P. patens* is specifically targeted to transposable elements (TEs) (Figure 9) and segregated away from genes⁵. Only about 0.5% of the methylated cytosines reside within genic sequences, which are mostly transcriptionally silenced⁴² and are controlled by PpDNMTs similarly to the way TE methylation is regulated by PpDNMTs (Figure 9). In agreement, DNA methylation in *P. patens* is positively associated with heterochromatic (i.e. H3K9me2) and negatively associated with euchromatic (e.g. H3K4me3) marks (Figure 3A)^{42,43}. It is further shown, that similar to *Arabidopsis*, long TEs in *P. patens* tend to be more heterochromatic, whereas short TEs are more euchromatic (Figure 3B)²². Consistent with the relationship with heterochromatin, it was found that DNA methylation levels associate with TE

size, i.e. they accumulate at relatively longer TEs (Figure 3C). These correlations of DNA methylation with heterochromatin, together with the complete or near complete elimination of CG, CHG, and CHH methylation in *met*, *cmt*, and *dnmt3b* mutants (Figure 1B, 3D), respectively, suggest that PpMET, PpCMT and PpDNMT3b function preferentially within heterochromatic TE sequences.

Interestingly, while genome wide CHH methylation in *P. patens cmt* mutant was similar to levels in WT (Figure 1B), when profiling methylation along TEs, it was found that CHH methylation in *cmt* mutant is substantially altered, i.e. increased closer to TE-edges and gradually decreased inward into the elements (Figure 3E). In the TE meta-analysis short and long TEs are relatively enriched and depleted closer and away to the points of TE-alignment, respectively (Figure 10A). Consequently, it was found that CHH methylation in *cmt* is preferentially hypomethylated in long TEs and hyper-methylated in short ones (Figure 3F). In accordance with the association of TE size with chromatin configuration (Figure 3B), it was found that CHH methylation in *cmt* mutant is preferentially depleted at genomic regions enriched for GC nucleotides and H3K9me2, and particularly increased within low GC and H3K9me2 TE regions (Figures 3F, G). When focusing on short TEs (<500 bps), it was found that hyper- and hypomethylation in *cmt* background continue to associate with eu- and hetero-chromatic regions, respectively (Figure 10B), suggesting that the chromatin structure, rather than TE size, determines the CHH methylation effect in *cmt* mutant. Overall, the results suggest that PpMET, PpCMT and PpDNMT3b function preferentially at H3K9me2-heterochromatic regions, and that PpDNMT3b CHH methylation activity is regulated to some extent by PpCMT.

PpDRMs target transcribed euchromatic TEs

Neither *drm1*, *drm2*, nor *drm12* mutants showed reduction of global genomic methylation (Figure 1B and Figure 10A). Similar to *drm* mutants, no effect on methylation was recently reported for *P. patens rdr2* mutant⁴⁰, which is validated herein. The analysis was expanded to a larger genomic portion (80% vs. 20%; Figure 10A and Table 7). These results, together with the complete CG, CHG, and CHH hypo-methylation in *met*, *cmt*, and *dnmt3b* mutants, respectively (Figure 1B), imply a trivial methylation activity of DRMs and RDR2 in *P. patens*.

As opposed to a global phenotype, the present inventors next checked for a localized methylation effect in *drm* mutants within statistically supported differentially methylated regions (DMRs) separated into distinct chromatin configurations. While hypo-methylated DMRs were not significantly enriched over hyper-methylated DMRs in neither the *drm* nor the *rdr2* mutants (Figure 10B), it was found that CHH-DMRs of *drm12* double mutant are particularly hypo-

5 methylated within genomic regions enriched for siRNA, low GC content, low histone H3 abundance, high H3K4me3, short TEs, and long-terminal-repeat (LTR) regions of retrotransposons (Figure 4A – top panel). *drm1* and *drm2* CHH-DMRs were mostly hyper-methylated and did not associate with any chromatin or DNA features (Figure 4A lower panels), thus implying an unrelated noise, which is a common feature of asymmetric methylation. Under this assumption, the particular hypomethylation effect in *drm12* (Figure 4A top panel) suggests some functional redundancy between DRM1 and DRM2. Intriguingly, it was found that CHH-DMRs of single and double *drm* mutants as well as of *rdr2* are gradually hypo-methylated within a small number of windows (≤ 5610) of expressed TEs (Figure 4A right panels and Figure 10C and D), which are also abundant in H3K4me3 and depleted of H3K9me2 (Figure 10E). Overall, these results associate PpDRMs methylation activity with RDR2 generated siRNA as well as with actively-transcribed euchromatic TE sequences, both of which are signatures of RdDM activity in angiosperm.

15 The weak genomic methylation activity of DRMs in *P. patens* could be explained by the exceptional high efficiency of PpCMT and PpDNMT3b. PpCMT targets CHG methylation as strongly as PpMET targets CG methylation (Figures 2A, D, 4E) while PpDNMT3b targets CHH methylation with more than the double the level of *Arabidopsis* CHH methylation activity (Figure 4E)⁴⁴. Consequently, together with their ability to *de novo* methylate DNA, it is possible that PpCMT and PpDNMT3b target and maintain non-CG methylation even within euchromatic regions that have a weak heterochromatic signal.

20 In support of a trivial role for RdDM in *P. patens*, it was found that siRNA in *P. patens* overlap with only 5% of methylated TEs, in comparison to 65% in *Arabidopsis* (Figure 4B). Moreover, similar to *Arabidopsis*, siRNA in *P. patens* was found to be enriched within long-heterochromatic TEs (Figure 4C). In *Arabidopsis*, RdDM functions mostly in euchromatic TEs, while heterochromatic siRNAs are hardly involved in maintaining DNA methylation^{22,45,46}. If the same is true in *P. patens*, then the exceptionally low abundance of siRNA in euchromatic TEs (0.9%) could further explain the minor role of PpDRMs in genomic methylation.

30 In addition to actively transcribed TEs, another source for euchromatic TEs could be those located in gene promoters^{22,47}. Notably, the frequency of TE integration within the first 200 bp upstream to transcription start site (TSS) of genes, was found to be lower by up to 2.3 times in *P. patens* than in *Arabidopsis* (Figure 4D). This result is counterintuitive, considering that the *P. patens* genome contains eight times more TEs than that of *Arabidopsis*, which are also spread more evenly along the chromosomes in comparison to the centric concentration of TEs in *Arabidopsis*⁴². Hence, the particular depletion of TEs in *P. patens* from gene promoters, which

are known to be the main target of DRMs and RdDM in angiosperms^{22,47-49}, could contribute to the weak genomic methylation effect of PpDRMs and RDR2 in *P.patens*.

To date, functional analyses of plant DNMTs were focused primarily on *Arabidopsis* and a few additional angiosperms. *P. patens* is a basal land plant that diverged from angiosperms about 400 million years ago⁵⁰ and encodes homologs of all four plant DNMT protein families¹⁸, including DNMT3 which has been lost during angiosperms evolution. Thus, the present comprehensive analysis of the entire PpDNMT proteins under *de novo* and homeostasis methylation conditions allowed the present inventors to reveal their function as well as to infer on the evolutionary mechanisms of DNA methylation in plants (Figure 5).

Mammalian DNMT3s function primarily as *de novo* methylases of CG sites and in some tissues also of CH sites^{24,51}. The present inventors show here that PpDNMT3s are required for *de novo* methylation of CG and CHH sites (Figure 1B). As PpDNMT3b is the first non-animal DNMT3 to be functionally characterized, the results imply that *de novo* methylation of CG and non-CG sites is an ancient feature of eukaryotic DNMT3 that predates the divergence of plant and animal DNMT3s. Additionally, the data demonstrate the ability of DNMT3 to be specialized in their hosts, such as the preference of mammalian DNMT3 towards CG sites and that of moss DNMT3 towards CHH sites. Conservation and diversification between mammalian and moss DNMT3s would provide the basis for further structure-function interactions of eukaryotic DNMT3. The narrow overlap of siRNA with DNA methylation (Figure 4B) and the trivial methylation effect in *Pprdr2* (Figure 1C and Figure 10A), suggest that the robust genomic methylation of PpDNMT3 does not involve the RdDM pathway. In comparison, the association between PpDRMs methylation effect, siRNA signal (Figure 4A), and PpRDR2 methylation profile (Figure 10D) link basal DRMs with RdDM. Consequently, these results suggest that since its emergence RdDM included DRMs rather than DNMT3s as its methylase component (Fig. 5B).

RPS methylation by PpCMT (Figure 1C) is the first *in vivo* evidence for *de novo* methylation by a CMT protein. *In vitro* studies have shown that *Arabidopsis* CMT2 and CMT3 can methylate unmethylated-DNA templates^{20,21}. Thus, it is possible that CMTs in *Arabidopsis* and other angiosperms are capable of mediating *de novo* methylation, as well (Figure 5A). CMT *de novo* methylation activity would help in resolving how DNA methylation is targeted to regions that are normally not regulated by RdDM, such as heterochromatic TEs and intra-genic sequences (gene bodies), methylation of the latter was recently genetically linked to CMT3⁵².

The antagonistic CHH methylation changes in *Ppcmt*, from hypomethylation in heterochromatin to hypermethylation in euchromatin (Figures 3E-G), resembles the methylation

phenotype of *Arabidopsis* histone *h1* mutation²². Similar to *Arabidopsis h1*, the elimination of CHG methylation in *Ppcmt* could disturb the chromatin in a way that affected regular CHH methylation activities as well as demethylation ones^{22,53}. The present data suggests that the role of H3K9me2 in targeting non-CG methylation in angiosperms^{4,10} has already been established in basal plants. However, unlike many angiosperms that utilize two CMT orthologues to methylate distinct non-CG contexts, i.e. CMT2 for CHH and CMT3 for CHG, basal plants use CMT for CHG and DNMT3 for CHH sites. Similar to angiosperm-CMTs, early diverged CMTs, such as PpCMT, probably also utilize their chromodomain to be targeted to H3K9me2-chromatin. Plant DNMT3s are missing a chromodomain, thus it is likely that the association of their CHH methylation with H3K9me2 is indirect. Mammalian DNMT3s were found to bind H3K9-methylated chromatin via attachment to chromodomain proteins or via unmethylated-H3K4 residues (H3K4me0), a histone mark associated with H3K9me2⁷. The partial dependency of CHH methylation on PpCMT/CHG methylation (Figure 3F) and the absence of the reverse effect, i.e. control of CHG methylation by DNMT3/CHH methylation (Figure 1B), suggest a hierarchy between CHG and CHH methylation. In this hierarchy, PpCMT is positioned on the higher level, possibly by recruiting the DNMT3 protein itself or by regulating the level of DNMT3 substrates, e.g. H3K9me2 and/or H3K4me0. An alternative explanation for the CHH hypomethylation in *cmt* mutants (Figure 3E), could be that PpCMT is involved in establishing CHH methylation that is subsequently maintained by PpDNMT3b. This hypothesis is supported by the ability of *Arabidopsis* CMTs to establish CHH methylation *in vitro*^{20,21}, and by the residual of CWA methylation in *Ppdnmt3b* mutants (Figure 8) that resembles the preference of some angiosperm CMTs toward such CHH subcontext⁴¹.

Table 5- DNMT gene model ID and source

DN MT group	Name	Gene Model	Taxonomic Group1	Taxonomic Group2	Organism	Protein Length (AA)	Source
DN MT1	NveDN MT1	125496	Animals	Cnidaria	Nematostella vectensis	1263	Metazome
DN	HsDNM	ENSG00000130816	Animals	Mammalia	Homo	163	Metaz

MT1	T1			ls	sapiens	2	ome
DN MT1	PpMET	Pp3c11_20540	Bryophy tes	Mosses	Physcom itrella patens	157 9	Phytoz ome
DN MT1	ZmMET	GRMZM2G333916_T01	Angiosp erms	Monoco ts	Zea mays	153 6	Phytoz ome
DN MT1	AtMET1	AT5G49160	Angiosp erms	Eudicot s	Arabidop sis thaliana	153 5	Phytoz ome
DN MT3	NveDN MT3	127267	Animals	Cnidari a	Nematost ella vectensis	624	Metaz ome
DN MT3	CiDNM T3	270939	Animals	Tunicat a	Ciona intestinal is	618	Metaz ome
DN MT3	MmDN MT3a	ENSMUSP00000020991	Animals	Mamma ls	Mus musculus	910	Metaz ome
DN MT3	MmDN MT3b	ENSMUSP00000051830	Animals	Mamma ls	Mus musculus	860	Metaz ome
DN MT3	HsDNM T3a	ENSG00000119772	Animals	Mamma ls	Homo sapiens	914	Metaz ome
DN MT3	HsDNM T3b	ENSG00000088305	Animals	Mamma ls	Homo sapiens	853	Metaz ome
DN MT3	KfDNM T3	kfl00080_0350_v1.1	Charoph ytes		Klebsor midium flaccidu m	893	K. nitens genom e project
DN MT3	MpDNM T3i	Mapoly0043s0021.1	Bryophy tes	Liverw orts	Marchant ia polymorp ha	654	Phytoz ome
DN MT3	MpDNM T3ii	Mapoly0095s0030.1	Bryophy tes	Liverw orts	Marchant ia	687	Phytoz ome

					polymorpha		
DN MT3	PpDNM T3a	Pp3c3_3540V3.1	Bryophytes	Mosses	Physcomitrella patens	736	Phytozome
DN MT3	PpDNM T3b	Pp3c13_8320V3.1	Bryophytes	Mosses	Physcomitrella patens	711	Phytozome
DN MT3	SfDNM T3i	Sphfalx0068s0040.1	Bryophytes	Mosses	Sphagnum fallax	739	Phytozome
DN MT3	SfDNM T3ii	Sphfalx0127s0004.1	Bryophytes	Mosses	Sphagnum fallax	767	Phytozome
DN MT3	SmDNM T3i	fgenes2_pg.C_scaffold_6000551	Lycophytes		Selaginella moellendorffii	630	Phytozome
DN MT3	SmDNM T3ii	fgenes2_pg.C_scaffold_17000071	Lycophytes		Selaginella moellendorffii	685	Phytozome
DN MT3	EbDNM T3	GNQG-2005968	Gymnosperms	Cycadales	Encephalartos barteri	551	1KP
DN MT3	SeDNM T3_F1	KAWQ-2009204	Gymnosperms	Cycadales	Stangeria eriopus	248	1KP
DN MT3	SeDNM T3_F2	KAWQ-2054998	Gymnosperms	Cycadales	Stangeria eriopus	244	1KP
DN MT3	WmDN MT3	TOXE-2012823	Gymnosperms	Gnetales	Welwitschia mirabilis	562	1KP
DN MT3	PtaDNM T3	PTA00084430	Gymnosperms	Conifers	Pinus taeda	426	PLAZA
DN MT3	PsyDN MT3	PSY00011178	Gymnosperms	Conifers	Pinus sylvestris	426	PLAZA

DN MT3	McDNM T3	CDFR-2008757	Gymnos perms	Conifer s	Manoao colensoi	405	1KP
DN MT3	SaDNM T3_F1	KLGF-2012241	Gymnos perms	Conifer s	Sundacar pus amarus	344	1KP
DN MT3	SaDNM T3_F2	KLGF-2078876	Gymnos perms	Conifer s	Sundacar pus amarus	105	1KP
DN MT3	PjDNM T3	MFTM-2078885	Gymnos perms	Conifer s	Pinus jeffreyi	298	1KP
DR M	KfDRM a	kfl00018_0570_v1.1	Charoph ytes		Klebsor midium flaccidu m	697	K. nitens genom e project
DR M	KfDRM b	kfl00253_0110_v1.1	Charoph ytes		Klebsor midium flaccidu m	139 5	K. nitens genom e project
DR M	MpDRM a	Mapoly0103s0053.3	Bryophy tes	Liverw orts	Marchant ia polymorp ha	723	Phytoz ome
DR M	MpDRM b	Mapoly0109s0015.2	Bryophy tes	Liverw orts	Marchant ia polymorp ha	758	Phytoz ome
DR M	PpDRM 1	Pp3c15_14362	Bryophy tes	Mosses	Physcom itrella patens	103 6	Phytoz ome
DR M	PpDRM 2	Pp3c15_21430	Bryophy tes	Mosses	Physcom itrella patens	842	Phytoz ome

DR M	SfDRMa	Sphfalx0006s0094	Bryophy tes	Mosses	Sphagnu m fallax	551	Phytoz ome
DR M	SfDRMb	Sphfalx0011s0183	Bryophy tes	Mosses	Sphagnu m fallax	836	Phytoz ome
DR M	SfDRMc	Sphfalx0209s0004	Bryophy tes	Mosses	Sphagnu m fallax	790	Phytoz ome
DR M	SmDRM a	411110	Lycophy tes		Selaginel la moellend orffii	566	Phytoz ome
DR M	SmDRM b	fgenes2_pg.C_scaffold_5 000240	Lycophy tes		Selaginel la moellend orffii	551	Phytoz ome
DR M	GbiDR M	GBI00021934	Gymnos perms	Ginkgo ales	Ginkgo biloba	655	PLAZ A
DR M	PabDR Ma	PAB00006266	Gymnos perms	Conifer s	Picea abies	356	PLAZ A
DR M	PabDR Mb	PAB00012689	Gymnos perms	Conifer s	Picea abies	482	PLAZ A
DR M	PabDR Mc	PAB00046707	Gymnos perms	Conifer s	Picea abies	596	PLAZ A
DR M	PabDR Md	PAB00048154	Gymnos perms	Conifer s	Picea abies	532	PLAZ A
DR M	PabDRe	PAB00049465	Gymnos perms	Conifer s	Picea abies	398	PLAZ A
DR M	PpiDRM a	PPI00065449	Gymnos perms	Conifer s	Pinus pinaster	747	PLAZ A
DR M	PpiDRM b	PPI00073543	Gymnos perms	Conifer s	Pinus pinaster	423	PLAZ A
DR M	PsyDRM a	PSY00004978	Gymnos perms	Conifer s	Pinus sylvestris	747	PLAZ A
DR M	PsyDRM b	PSY00005220	Gymnos perms	Conifer s	Pinus sylvestris	473	PLAZ A

DR M	PsyDRM c	PSY00006869	Gymnos perms	Conifer s	Pinus sylvestris	662	PLAZ A
DR M	PsyDRM d	PSY00008531	Gymnos perms	Conifer s	Pinus sylvestris	645	PLAZ A
DR M	PtaDRM a	PTA00004821	Gymnos perms	Conifer s	Pinus taeda	687	PLAZ A
DR M	PtaDRM b	PTA00013841	Gymnos perms	Conifer s	Pinus taeda	581	PLAZ A
DR M	PtaDRM c	PTA00015283	Gymnos perms	Conifer s	Pinus taeda	367	PLAZ A
DR M	PtaDRM d	PTA00023209	Gymnos perms	Conifer s	Pinus taeda	461	PLAZ A
DR M	PtaDRM e	PTA00046027	Gymnos perms	Conifer s	Pinus taeda	435	PLAZ A
DR M	PtaDRM f	PTA00067039	Gymnos perms	Conifer s	Pinus taeda	349	PLAZ A
DR M	AtrDRM 3	evm_27.TU.AmTr_v1.0_sc affold00003.192	Angiosp perms	Basel Angios perms	Amborell a trichopod a	719	Phytoz ome
DR M	AtrDRM 2	evm_27.TU.AmTr_v1.0_sc affold00096.44	Angiosp perms	Basel Angios perms	Amborell a trichopod a	503	Phytoz ome
DR M	BdDRM 2a	Bradi1g11420	Angiosp perms	Monoco ts	Brachypo dium distachyo n	518	Phytoz ome
DR M	BdDRM 2b	Bradi1g77873	Angiosp perms	Monoco ts	Brachypo dium distachyo n	610	Phytoz ome
DR M	BdDRM 3	Bradi2g38577	Angiosp perms	Monoco ts	Brachypo dium	706	Phytoz ome

					distachyon		
DR M	BdDRM 2c	Bradi4g05680	Angiosperms	Monocots	Brachypodium distachyon	588	Phytome
DR M	PhDRM 3	Pahal.C01225	Angiosperms	Monocots	Panicum hallii	660	Phytome
DR M	PhDRM 2a	Pahal.I00079	Angiosperms	Monocots	Panicum hallii	603	Phytome
DR M	PhDRM 2b	Pahal.I00969	Angiosperms	Monocots	Panicum hallii	588	Phytome
DR M	SbDRM 2a	Sobic.001G458100	Angiosperms	Monocots	Sorghum bicolor	577	Phytome
DR M	SbDRM 2b	Sobic.001G535800	Angiosperms	Monocots	Sorghum bicolor	608	Phytome
DR M	SbDRM 2c	Sobic.003G124000	Angiosperms	Monocots	Sorghum bicolor	609	Phytome
DR M	SbDRM 3	Sobic.009G032200	Angiosperms	Monocots	Sorghum bicolor	657	Phytome
DR M	ZmDRM 3	GRMZM2G065599	Angiosperms	Monocots	Zea mays	461	Phytome
DR M	ZmDRM 2a	GRMZM2G092497	Angiosperms	Monocots	Zea mays	604	Phytome
DR M	ZmDRM 2b	GRMZM2G137366	Angiosperms	Monocots	Zea mays	610	Phytome
DR M	ZorDRM 3	Zosma21g00360.1	Angiosperms	Monocots	Zostera marina	597	Phytome
DR M	ZorDRM 2	Zosma67g00380.1	Angiosperms	Monocots	Zostera marina	570	Phytome
DR M	AtDRM 3	AT3G17310	Angiosperms	Eudicots	Arabidopsis thaliana	711	Phytome
DR	AtDRM	AT5G14620	Angiosp	Eudicot	Arabidop	627	Phytome

M	2		erms	s	sis thaliana		ome
DR M	AtDRM 1	AT5G15380	Angiosp erms	Eudicot s	Arabidop sis thaliana	625	Phytoz ome
DR M	BrDRM 2a	Brara.B00561	Angiosp erms	Eudicot s	Brassica rapa	406	Phytoz ome
DR M	BrDRM 3b	Brara.E02454	Angiosp erms	Eudicot s	Brassica rapa	703	Phytoz ome
DR M	BrDRM 2b	Brara.G02881	Angiosp erms	Eudicot s	Brassica rapa	599	Phytoz ome
DR M	BrDRM 3a	Brara.I04643	Angiosp erms	Eudicot s	Brassica rapa	502	Phytoz ome
DR M	BrDRM 2c	Brara.J02008	Angiosp erms	Eudicot s	Brassica rapa	602	Phytoz ome
DR M	GmDR M2a	Glyma.02G035700	Angiosp erms	Eudicot s	Glycine max	538	Phytoz ome
DR M	GmDR M2b	Glyma.05G005600	Angiosp erms	Eudicot s	Glycine max	591	Phytoz ome
DR M	GmDR M3a	Glyma.07G233200	Angiosp erms	Eudicot s	Glycine max	695	Phytoz ome
DR M	GmDR M3b	Glyma.17G038300	Angiosp erms	Eudicot s	Glycine max	730	Phytoz ome
DR M	GmDR M2c	Glyma.19G006100	Angiosp erms	Eudicot s	Glycine max	581	Phytoz ome
DR M	MeDRM 3	Manes.03G210200	Angiosp erms	Eudicot s	Manihot esculenta	780	Phytoz ome
DR M	MeDRM 2a	Manes.15G149300	Angiosp erms	Eudicot s	Manihot esculenta	639	Phytoz ome
DR M	MeDRM 2b	Manes.17G113600	Angiosp erms	Eudicot s	Manihot esculenta	638	Phytoz ome
DR M	PrpDRM 3	Prupe.1G109800	Angiosp erms	Eudicot s	Prunus persica	718	Phytoz ome
DR	PrpDRM	Prupe.3G287400	Angiosp	Eudicot	Prunus	585	Phytoz

M	2a		erms	s	persica		ome
DR	PrpDRM	Prupe.8G038800	Angiosp	Eudicot	Prunus	583	Phytoz
M	2b		erms	s	persica		ome

Table 6

	Primer Name	Sequence (5' -> 3')	SEQ ID NO:
Cloning	Ppdrml 5'UTR BamHI Fw	GGATCCTGGAACGCAAAACAAGACCG	24
	Ppdrml 5'UTR BamHI Rv	GGAGGTGGATCCAATTGTTCTTTC	25
	Ppdrml 3'UTR SphI Fw	GCATGCTTCCCCTGGCAGAAATTGG	26
	Ppdrml 3'UTR NotI SphI Rv	GCATGCGGCCGCGAGCATCAAATTAGAGCTTCAGGG	27
	Ppdrml 2 5'UTR BamHI Fw	GGATCCGTGGTGAACCTAGTTGTCCATTGG	28
	Ppdrml 2 5'UTR BamHI Rv	GGATCCTGGCGTGTAAGCTCACACTAA	29
	Ppdrml 2 3'UTR SphI Fw	GCATGCGTTTGCCCTGCCTTGTTCTTTC	30
	Ppdrml 2 3'UTR NotI SphI Rv	GCATGCGGCCGCCATCCTTTTGCAACAATCCTCC	31
	Ppdrml3a HindIII 5'KO Fw	AAGCTTGTCGTGCTGAGTATTCAGATAATCGTAGC	32
	Ppdrml3a HindIII 5'KO Rv	AAGCTTAATCCAACCTGTTCCAATTCCGC	33
	Ppdrml3a SphI 3'KO Fw	GCATGCTAGCTCTCTTGAAGTATCCG	34
	Ppdrml3a 3'KO Rv	AATCGTGCTTTCTACCATAGTATGCC	35
	Ppdrml3b HindIII 5'KO Fw	AAGCTTCGGGTTTCGGAGTTCTGGGTT	36
	Ppdrml3b HindIII 5'KO Rv	AAGCTTGCAGGCCAGAGGAAAGAGCG	37
	Ppdrml3b SphI 3'KO Fw	GCATGCCCATGTTCCAATCTTTGACTTGCC	38
	Ppdrml3b 3'KO Rv	ACATTCCGTTTACCAGTAGCATCTGG	39
	KpnI Zeo Fw	GGTACCGTCAACATGGTGGAGCACGACA	40
	SphI Zeo Rv	GCATGCCAGGTCAGTGGATTTTGGTTTATAGG	41
Deletion mutant screening			
	Ppdrml 2240 5' Fw	GGAACACGGTGGATGTATTCTTCT	42
	Ppdrml 5550 3' Rv	AGGCGGTATGGTTGTGCCACC	43
	Ppdrml 3209 e1 Fw	GGTCAAGGTGCAATCATCTCAACG	44
	Ppdrml 3779 e1 Rv	GCGTTGGGATGTTTGGAGCA	45
	Ppdrml 2292 5' Fw	GACAATTTCATTTCATGCGAGTTGTC	46
	Ppdrml 5324 3' Rv	CAAGCCATGCCTATTGTTATCACTGTTC	47
	Ppdrml 23202 e1 Fw	ATTGGCTTTGGTCTTCCTGGTCA	48
	Ppdrml 23581 e1 Rv	TGTGGGAATTGCAGTGGCGT	49
	Ppdrml3a 5046 Fw	GCTGCAAGCGTGAGCGATTC	50
	Ppdrml3a 10651 Rv	GGGTTGGATATCACTAAGCTCCACC	51
	Ppdrml3a 6377 Fw	GCTGACCAATCTAGGCATCCCG	52
	Ppdrml3a 8427 Rv	TGGAGGGCTTGATTTAGGCAGAG	53
	Ppdrml3b 5678 Fw	GCTGATGACTGCTTGAGCCTTCG	54
	Ppdrml3b 10445 Rv	TCCACTCGTCTACTTCTTTGAGATAGG	55

	PpDnmt3b 7159 Fw	GGTCGGGTGAACGGCTGG	56
	PpDnmt3b e11 Rv	AAGGCTATCCTGTCGAGTTGGCTT	57
	35S Rv	TGGGACCACTGTCGGCAGAG	58
	35S-Ter-R-Fw	GCCCCCGCTTAAAAATTGGT	59
Validation of RPS insertion	RPS-top-R_new	AAGTAGAGAAAGGAAAGAGAAAAGGGG	60
	35S Rv	TGGGACCACTGTCGGCAGAG	61
Bisulfite assay	RPS-top-R_new	AAGTAGAGAAAGGAAAGAGAAAAGGGG	62
	RPS-top-F	CTGTATTTTCTCCCTTCA	63

Table 7 - BS-seq summary

Median coverage and averaged methylation levels in wild type and PpDNMT mutant genomes. Substantial losses in methylation (>90%) are marked in bold and underlined.

5

Genetic background	Mapping			Median site coverage	Average genomic methylation levels				Average organelle methylation levels	
	Total reads	Mapped reads	Ratio		CG	CWG	CCG	CHH	Mitochondria	Chloroplast
wild type	41145982	34088890	83%	3	19.56%	21.12%	8.44%	15%	0.17%	0.17%
<i>drm1</i>	41372443	33415124	81%	3	20.90%	22.26%	9.57%	16.24%	0.17%	0.17%
<i>drm2</i>	52331188	42380753	81%	3	19.79%	21.25%	8.87%	15.29%	0.17%	0.17%
<i>drm12</i>	39059321	31104455	80%	3	19.43%	21.18%	8.84%	15.38%	0.17%	0.16%
<i>dnmt3a</i>	46395247	34361843	74%	3	20.14%	21.40%	8.83%	14.88%	0.17%	0.17%
<i>dnmt3b</i>	46037700	34325065	75%	3	19.94%	22.79%	10.68%	<u>0.84%</u>	0.18%	0.17%
<i>dnmt3ab</i>	51456246	40940817	80%	3	19.62%	22.59%	10.57%	<u>0.83%</u>	0.17%	0.16%
<i>met</i>	48204209	37218804	77%	3	<u>1.28%</u>	21.74%	<u>0.65%</u>	15.35%	0.17%	0.17%
<i>cmt</i>	47855012	37225542	78%	3	20.68%	<u>0.49%</u>	<u>0.39%</u>	15.50%	0.17%	0.17%
<i>rdr2</i>	51728794	40797730	79%	3	24.80%	25.83%	13.20%	17.82%	0.53%	0.59%

EXAMPLE 2***Expression of PpDNMT3b in human cells*****MATERIALS AND METHODS****Biological material**

10 HEK 293T cells were maintained in a complete medium (Dulbecco's modified Eagle medium [DMEM], Gibco, supplemented with 10 % [vol/vol] fetal calf serum [FCS], 4mM L-glutamin, 40 V/ml Penicillin, 40 mg/ml streptomycin and 5 V/ml hystatin [Biological Industries, Israel] at 37 °C and 5% CO₂.

PpDNMT3b open reading frame was codon-optimized for expression in human, synthesized and cloned into pCDNA3.1+N-eGFP by GENSCRIPT. For control, pCDNA3.1+N-eGFP was used. The transfection was done in 60 mm tissue culture plate with cells reaching approximately 80% confluency, using PolyJet™ reagent (SignaGen) according to manufacturer's instructions. The cells were maintained for 3 or 7 days prior to DNA extraction.

Genomic DNA was extracted using GenElute™ Mammalian Genomic DNA Miniprep Kit (Qiagen) according to the manufacturer's instructions.

Targeted bisulfite sequencing and data analysis

Bisulfite library was prepared using SureSelectXT Methyl-Seq Library Preparation Kit for targeted methylation sequencing (Agilent). Deep sequencing was performed on Illumina Hi-Seq 2000, yielding approximately 80 million 150-bp paired-end reads per sample, covering 84 Mb of the human genome.

The reads were aligned to Hg38 reference genome using Bismark v0.19.1 with Bowtie2.2.6. Methylation counting for each cytosine in CG, CHG and CHH contexts was performed with bismark_methylation_extractor.

DNA methylation of genes and transposable elements was calculated using a custom perl script. Genes and TEs were aligned at either the 5' or 3' end and average methylation for all cytosines in CG, CHG or CHH context was calculated in 50 bp sliding window within 2 kb upstream and downstream to the alignment point.

Integrative Genomics Viewer 2.4.19 was used for visualization of cytosine methylation data.

Differential expression analysis

RNA was extracted using RNeasy Mini Kit (Qiagen) and enriched for poly-T and processed for sequencing by the Weizmann institute sequencing unit on Hiseq2500 as SE. RNA-seq data was aligned to the human genome (version GRCh38.p13) using STAR (2.7.3a) and differential expression per gene was analyzed using DeSeq2. The human disease database was used to list diseases related to miss-regulated genes found in the analyses.

RESULTS

As can be seen from Figures 11 and 12, human cells genetically modified to express PpDNMT3b show CHH hypermethylation.

PpDNMT3b differs from human DNMT3a and DNMT3b in sequence specificity as well as the ability to produce non-CG methylation in human cells which express DNMT3a and DNMT3b. Non-CG methylation is performed by DNMT3s and can be found in human embryonic stem cells and neurons. In both cell types, non-CG methylation occurs mainly in CA sequences.

Physcomitrella patens PpDNMT3b has a similar preference, however, is more efficient in methylating CT and CC sequences while limited in methylating CHG sequences (CAG/CTG) (Figure 13). Expression of PpDNMT3b in HEK293 cells resulted in upregulation of 14 genes, most of which belong to heat shock response genes and are involved in various human diseases (Table 8). Indeed, several of these genes had hyper non-CG methylation within their gene bodies (Figure 14).

Table 8

#	Name	Expression fold change	Type	Major Disease effect
1	HSPA1A	48	heat shock protein	Carotid Artery Occlusion and Transient Cerebral Ischemia
2	HSPH1	4	heat shock protein	Aggressive B-Cell Non-Hodgkin Lymphoma and Cystic Fibrosis.
3	HSPA7	32	heat shock protein	
4	HSPA6	35	heat shock protein	Hepatocellular Carcinoma
5	DNAJB1	4	heat shock protein	Fibrolamellar Carcinoma
6	ZFAND2 A	3	zinc finger	
7	HSPB1	2	heat shock protein	Charcot-Marie-Tooth Disease and Neuronopathy
8	HSP90AA1	3	heat shock protein	Hypersensitivity Reaction Disease and Hand-Foot-Genital Syndrome.
9	CRYAB	6	crystallin alpha B	Cardiomyopathy and Cataract
10	SERPINH1	2	serpin family	Osteogenesis Imperfecta
11	BAG3	2	BCL2 associated	Myopathy, Myofibrillar and Cardiomyopathy
12	CHORDC1	2	cys/his rich domain	Involved in stress response. Prevents tumorigenesis
13	HSPA4L	2	heat shock	

			protein	59
14	ZNF79	2	zinc finger	Prostate Carcinoma In Situ and Kabuki Syndrome

EXAMPLE 3

Expression of PpDNMT3b in plant cells

MATERIALS AND METHODS

Plants

Arabidopsis ddcc (drm1 drm2 cmt2 cmt3 quadruple mutant) plants were grown in a controlled growth room under long-day photoperiod (16-h light and 8-h dark, light intensity 200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) at 22 °C $\pm$ 2 and 70% humidity.

Cloning

pEGAD-hyg was generated by replacing the BASTA resistance cassette in pEGAD (GenBank: AF218816.1) with hygromycin resistance cassette from pcambia1300 via restriction with VspI (Thermo Fisher Scientific). In-Fusion® HD Cloning (Takara bio) was used to clone the *Arabidopsis* CMT3 promotor and PpDNMT3b ORF in frame with EGFP in the pEGAD-hyg vector. The final plasmid was verified by sanger sequencing.

Generation of transgenic mutant lines

ddcc mutant plants which have low non-CG methylation levels were used as background for PpDNMT3b expression which was introduced into plants via *Agrobacterium tumefaciens*-mediated transformation.

BS-seq and analysis

Genomic DNA was extracted from leaves using DNeasy Plant Mini Kit (Qiagen) according to manufacturer's instructions. WGBS was performed by BGI. The reads were aligned using methylpy and analyzed via python scripts.

RESULTS

As illustrated in Figure 15, expression of PpDNMT3b in *Arabidopsis*, even under the weak AtCMT3 promotor, induced genome wide hyper CHH methylation.

EXAMPLE 4

Generation of pdCas9-DNMT3 fusion protein

Material and methods

The human codon-optimized PpDNMT3b methyltransferase domain (MTD) DNA region was cloned in frame with dCas9 replacing HsDNMT3a-MTD in pdCas9-DNMT3A-

PuroR_BACH2-sgRNA8 (Addgene plasmid #71828) by BamHI and FseI. The transfection was done in 60 mm tissue culture plate with cells reaching approximately 80% confluency, using PolyJet™ reagent (SignaGen) according to manufacturer's instructions. The cells were maintained for 7 days prior to DNA extraction.

5 The plasmid was transformed into HEK293 as described in Example 2. Cells were harvested 7 days following transfection. Bisulfite sequencing of the dCAS9 targeted region will be analyzed to determine methylation by the fusion protein.

RESULTS

10 The DNA sequence encoding the fusion protein is set forth in SEQ ID NO: 64 and illustrated in Figure 16.

 The amino acid sequence of the fusion protein is set forth in SEQ ID NO: 65 and illustrated in Figure 17.

15 Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims.

20 All publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention. To the extent that section headings are used, they should not be construed as necessarily limiting. In addition, any priority document(s) of this application is/are hereby incorporated herein by reference in its/their entirety.

25 In addition, any priority documents of this application are hereby incorporated herein by reference in their entirety.

REFERENCES

1. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**, 481–514 (2005).
2. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8689–8694 (2010).
3. Niederhuth, C. E. *et al.* Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* **17**, 194 (2016).
4. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16**, 519–32 (2015).
5. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science (80-.).* **328**, 916–919 (2010).
6. Cedar, H. & Bergman, Y. Programming of DNA methylation patterns. *Annu. Rev. Biochem.* **81**, 97–117 (2012).
7. Jurkowska, R. Z. & Jeltsch, A. Enzymology of Mammalian DNA Methyltransferases. *Adv. Exp. Med. Biol.* **945**, 87–122 (2016).
8. Feng, W. & Michaels, S. D. Accessing the Inaccessible: The Organization, Transcription, Replication, and Repair of Heterochromatin in Plants. *Annu. Rev. Genet.* **49**, 439–459 (2015).
9. Springer, N. M., Lisch, D. & Li, Q. Creating Order from Chaos: Epigenome Dynamics in Plants with Complex Genomes. *Plant Cell* **28**, 314–325 (2016).
10. Wendte, J. M. & Schmitz, R. J. Specifications of Targeting Heterochromatin Modifications in Plants. *Mol. Plant* **11**, 381–387 (2018).
11. Zhang, H., Lang, Z. & Zhu, J.-K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).
12. Sotelo-Silveira, M., Chávez Montes, R. A., Sotelo-Silveira, J. R., Marsch-Martínez, N. & de Folter, S. Entering the Next Dimension: Plant Genomes in 3D. *Trends Plant Sci.* **23**,

598–612 (2018).

13. Song, X. & Cao, X. Context and Complexity: Analyzing Methylation in Trinucleotide Sequences. *Trends Plant Sci.* **22**, 351–353 (2017).
14. Seymour, D. K. & Becker, C. The causes and consequences of DNA methylome variation in plants. *Curr. Opin. Plant Biol.* **36**, 56–63 (2017).
15. Liu, C. *et al.* Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution. *Genome Res.* **26**, 1057–1068 (2016).
16. Satyaki, P. R. V. & Gehring, M. DNA methylation and imprinting in plants: machinery and mechanisms. *Crit. Rev. Biochem. Mol. Biol.* **52**, 163–175 (2017).
17. Quadrana, L. & Colot, V. Plant Epigenetics. *Annu. Rev. Genet.* (2016). doi:10.1146/annurev-genet-120215-035254
18. Malik, G., Dangwal, M., Kapoor, S. & Kapoor, M. Role of DNA methylation in growth and differentiation in *Physcomitrella patens* and characterization of cytosine DNA methyltransferases. *FEBS J.* **279**, 4081–4094 (2012).
19. Bewick, A. J. *et al.* The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. *Genome Biol.* **18**, 65 (2017).
20. Du, J. *et al.* Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167–180 (2012).
21. Stroud, H. *et al.* Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2013).
22. Zemach, A. *et al.* The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
23. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
24. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).

25. Cao, X. *et al.* Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proc Natl Acad Sci U S A* **97**, 4979–4984 (2000).
26. Tamiru, M., Hardcastle, T. J. & Lewsey, M. G. Regulation of genome-wide DNA methylation by mobile small RNAs. *New Phytol.* **217**, 540–546 (2018).
27. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev.* **15**, 394–408 (2014).
28. Cuerda-Gil, D. & Slotkin, R. K. Non-canonical RNA-directed DNA methylation. *Nat. plants* **2**, 16163 (2016).
29. Underwood, C. J., Henderson, I. R. & Martienssen, R. A. Genetic and epigenetic variation of transposable elements in Arabidopsis. *Curr. Opin. Plant Biol.* **36**, 135–141 (2017).
30. Zhou, M., Palanca, A. M. S. & Law, J. A. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nat. Genet.* **50**, 865–873 (2018).
31. Daccord, N. *et al.* High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **49**, 1099–1106 (2017).
32. Schmid, M. W. *et al.* Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. *Genome Biol.* **19**, 9 (2018).
33. Richards, C. L. *et al.* Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecol. Lett.* **20**, 1576–1590 (2017).
34. Anderson, S. N. *et al.* Subtle Perturbations of the Maize Methylome Reveal Genes and Transposons Silenced by Chromomethylase or RNA-Directed DNA Methylation Pathways. *G3 (Bethesda)*. **8**, 1921–1932 (2018).
35. Noy-Malka, C. *et al.* A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of *Physcomitrella patens*. *Plant Mol. Biol.* **84**, 719–35 (2014).
36. Yaari, R. *et al.* DNA METHYLTRANSFERASE 1 is involved in mCG and mCCG DNA

- methylation and is essential for sporophyte development in *Physcomitrella patens*. *Plant Mol. Biol.* **88**, 387–400 (2015).
37. Gentry, M. & Meyer, P. An 11bp region with stem formation potential is essential for de novo DNA methylation of the RPS element. *PLoS One* **8**, e63652 (2013).
 38. Muller, A., Marins, M., Kamisugi, Y. & Meyer, P. Analysis of hypermethylation in the RPS element suggests a signal function for short inverted repeats in de novo methylation. *Plant Mol. Biol.* **48**, 383–399 (2002).
 39. Singh, A., Zubko, E. & Meyer, P. Cooperative activity of DNA methyltransferases for maintenance of symmetrical and non-symmetrical cytosine methylation in *Arabidopsis thaliana*. *Plant J.* **56**, 814–823 (2008).
 40. Coruh, C. *et al.* Comprehensive Annotation of *Physcomitrella patens* Small RNA Loci Reveals That the Heterochromatic Short Interfering RNA Pathway Is Largely Conserved in Land Plants. *Plant Cell* **27**, 2148–2162 (2015).
 41. Gouil, Q. & Baulcombe, D. C. DNA Methylation Signatures of the Plant Chromomethyltransferases. *PLoS Genet.* **12**, 1–17 (2016).
 42. Lang, D. *et al.* The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).
 43. Widiez, T. *et al.* The chromatin landscape of the moss *Physcomitrella patens* and its dynamics during development and drought stress. *Plant J.* **79**, 67–81 (2014).
 44. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
 45. Huettel, B. *et al.* Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. **25**, 2828–2836 (2006).
 46. Zhong, X. *et al.* DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19**, 870–875 (2012).

47. Stroud, H., Greenberg, M. & Feng, S. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* **152**, 352–364 (2013).
48. Li, Q. *et al.* RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14728–33 (2015).
49. Tan, F. *et al.* Analysis of Chromatin Regulators Reveals Specific Features of Rice DNA Methylation Pathways. *Plant Physiol.* (2016). doi:10.1104/pp.16.00393
50. Rensing, S. A. *et al.* The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64–69 (2008).
51. Law, J. A. & Jacobsen, S. E. Establishing , maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
52. Bewick, A. J. *et al.* On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9111–6 (2016).
53. Frost, J. M. *et al.* FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4720–E4729 (2018).
54. Ibarra, C. A. *et al.* Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. *Science* (80-.). **337**, 1360–1364 (2012).
55. Roudier, F. *et al.* Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J.* **30**, 1928–1938 (2011).
56. Matzke, M. A., Kanno, T. & Matzke, A. J. M. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu. Rev. Plant Biol.* **66**, 243–267 (2015).
57. Ma, L. *et al.* Angiosperms Are Unique among Land Plant Lineages in the Occurrence of Key Genes in the RNA-Directed DNA Methylation (RdDM) Pathway. *Genome Biol. Evol.* **7**, 2648–2662 (2015).

58. Law, J. A., Vashisht, A. A., Wohlschlegel, J. A. & Jacobsen, S. E. 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 (2011).
59. Ashton, N. W. & Cove, D. J. The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. *Mol. Gen. Genet. MGG* **154**, 87–95 (1977).
60. Nishiyama, T., Hiwatashi, Y., Sakakibara, I., Kato, M. & Hasebe, M. Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. *DNA Res.* **7**, 9–17 (2000).
61. Frank, W., Decker, E. L. & Reski, R. Molecular tools to study *Physcomitrella patens*. *Plant Biol. (Stuttg.)* **7**, 220–227 (2005).
62. Mosquna, A. *et al.* Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. *Development* **136**, 2433–2444 (2009).
63. Parsons, J. *et al.* Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol. J.* **10**, 851–861 (2012).
64. Zimmer, A. D. *et al.* Reannotation and extended community resources for the genome of the non-seed plant *Physcomitrella patens* provide insights into the evolution of plant gene structures and functions. *BMC Genomics* **14**, 498 (2013).
65. Leinonen, R., Sugawara, H., Shumway, M. & Collaboration, I. N. S. D. The sequence read archive. *Nucleic Acids Res.* **39**, D19–21 (2011).
66. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
67. Wickett, N. J. *et al.* Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4859–68 (2014).
68. Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**, 17 (2014).

69. Xie, Y. *et al.* SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**, 1660–1666 (2014).
70. Johnson, M. T. J. *et al.* Evaluating Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. *PLoS One* **7**, e50226 (2012).
71. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
72. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
73. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
74. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
75. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

WHAT IS CLAIMED IS:

1. An isolated polynucleotide encoding a fusion protein which comprises a DNA targeting moiety linked to a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein.
2. The isolated polynucleotide of claim 1, wherein said DNA targeting moiety comprises a DNA endonuclease protein.
3. The isolated polynucleotide of claim 2, wherein said DNA endonuclease protein is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 and Cpf1 endonuclease.
4. The isolated polynucleotide of claim 2, wherein said DNA endonuclease protein comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.
5. An isolated polynucleotide encoding a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein having a codon usage optimized for expression in an organism which is not a gymnosperm or a bryophyte.
6. The isolated polynucleotide of claims 1 or 5, wherein said plant DNMT3 protein is a gymnosperm or a bryophyte DNMT3 protein.
7. The isolated polynucleotide of claim 5, wherein said organism is a mammal.
8. The isolated polynucleotide of claim 7, wherein said mammal is a human.
9. The isolated polynucleotide of claim 5, wherein said organism is an angiosperm.
10. The isolated polynucleotide of claim 5, wherein said DNMT3 protein is fused to a DNA targeting moiety.

11. The isolated polynucleotide of claim 10, wherein said DNA targeting moiety comprises a DNA endonuclease protein.

12. The isolated polynucleotide of claim 11, wherein said DNA endonuclease protein is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 and Cpf1 endonuclease.

13. The isolated polynucleotide of claim 11, wherein said DNA endonuclease protein comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.

14. The isolated polynucleotide of claim 1, wherein said fusion protein comprises a single copy of said DNMT3 protein.

15. The isolated polynucleotide of any one of claims 1-14, wherein said catalytic domain of the DNMT3 protein comprises an amino acid sequence as least 70 % similar or identical to at least one of the sequences as set forth in SEQ ID NOs: 1-11.

16. The isolated polynucleotide of any one of claims 1-14, wherein said catalytic domain of the DNMT3 protein comprises an amino acid sequence 100 % similar or identical to at least one of the sequences as set forth in SEQ ID NO: 1-11.

17. The isolated polynucleotide of claim 2, wherein said catalytic domain is linked directly to said DNA endonuclease protein.

18. The isolated polynucleotide of claim 2, wherein said catalytic domain is linked to said endonuclease protein via a peptide linker.

19. The isolated polynucleotide of claims 4 or 13, wherein the catalytically inactive Cas9 protein comprises mutations at a site selected from the group consisting of D10, E762, H983, D986, H840 and N863.

20. The isolated polynucleotide of claim 19, wherein the mutations are: (i) D10A or D10N, and (ii) H840A, H840N, or H840Y.
21. The isolated polynucleotide of claim 20 wherein said mutations are D10A and H840A.
22. The isolated polynucleotide of claims 4 or 13, wherein said dCAS9 comprises the sequence as set forth in SEQ ID NO: 23.
23. The isolated polynucleotide of claim 2, wherein said DNMT3 protein is linked to the C terminus of said endonuclease protein.
24. The isolated polynucleotide of claim 2, wherein said DNMT3 protein is linked to the N terminus of said endonuclease protein.
25. The isolated polynucleotide of any one of claims 1-24, wherein said DNMT3 methylates a target DNA at a CHH site.
26. The isolated polynucleotide of any one of claims 1-24, wherein said DNMT3 methylates a target DNA at a CC site and/or a CT site to a greater extent than a human DNMT3 methylates said target DNA under identical conditions.
27. The isolated polynucleotide of claim 25, wherein said DNMT3 additionally methylates a target DNA at a CpG site.
28. A polypeptide comprising a DNA targeting moiety linked to a DNMT3 protein.
29. The polypeptide of claim 28, wherein said DNA targeting moiety comprises a DNA endonuclease protein.
30. The polypeptide of claim 29, wherein said DNA endonuclease protein is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5,

Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 and Cpf1 endonuclease.

31. The polypeptide of claim 29, wherein said DNA endonuclease protein comprises a catalytically inactive CRISPR associated 9 (dCas9) protein.

32. An expression vector comprising the polynucleotide of any one of claims 1-27.

33. An expression vector comprising a polynucleotide encoding a catalytic domain of a species of a plant DNA methyltransferase 3 (DNMT3) protein operatively linked to a transcriptional regulatory sequence which is not of said species.

34. The expression vector of claim 33, wherein said transcriptional regulatory sequence is not a gymnosperm transcriptional regulatory sequence or a bryophyte regulatory sequence.

35. The expression vector of claim 34, wherein said transcriptional regulatory sequence comprises a mammalian transcriptional regulatory sequence.

36. The expression vector of claim 34, wherein said transcriptional regulatory sequence comprises an angiosperm transcriptional regulatory sequence.

37. A cell which expresses the polynucleotide of any one of claims 1-27.

38. A cell which comprises the expression vector of any one of claims 32-36.

39. The cell of claim 37, wherein the cell is a mammalian cell.

40. The cell of claim 37, wherein the cell is a plant cell.

41. The cell of claim 40, wherein said plant cell is an angiosperm cell.

42. A kit comprising the polynucleotide of any one of claims 1-27 and at least one guide RNA which is directed to a predetermined target gene.

43. A method of increasing methylation of DNA in a cell, the method comprising expressing a polynucleotide encoding a catalytic domain of a plant DNA methyltransferase 3 (DNMT3) protein in the cell, thereby increasing methylation of DNA in the cell, wherein the cell is not of a gymnosperm plant.

44. A method of increasing methylation of DNA in a cell, the method comprising expressing the polynucleotide of any one of claims 1-27 in the cell, thereby increasing methylation of DNA in the cell.

45. The method of claim 44, further comprising expressing one or more guide RNA directed to a target gene of the cell.

46. The method of claims 43 or 44, wherein the cell is a mammalian cell.

47. The method of claim 44, wherein the cell is a plant cell.

48. The method of claims 43 or 44, wherein the cell is a diseased cell.

49. The method of claim 46, wherein said mammalian cell is a human cell.

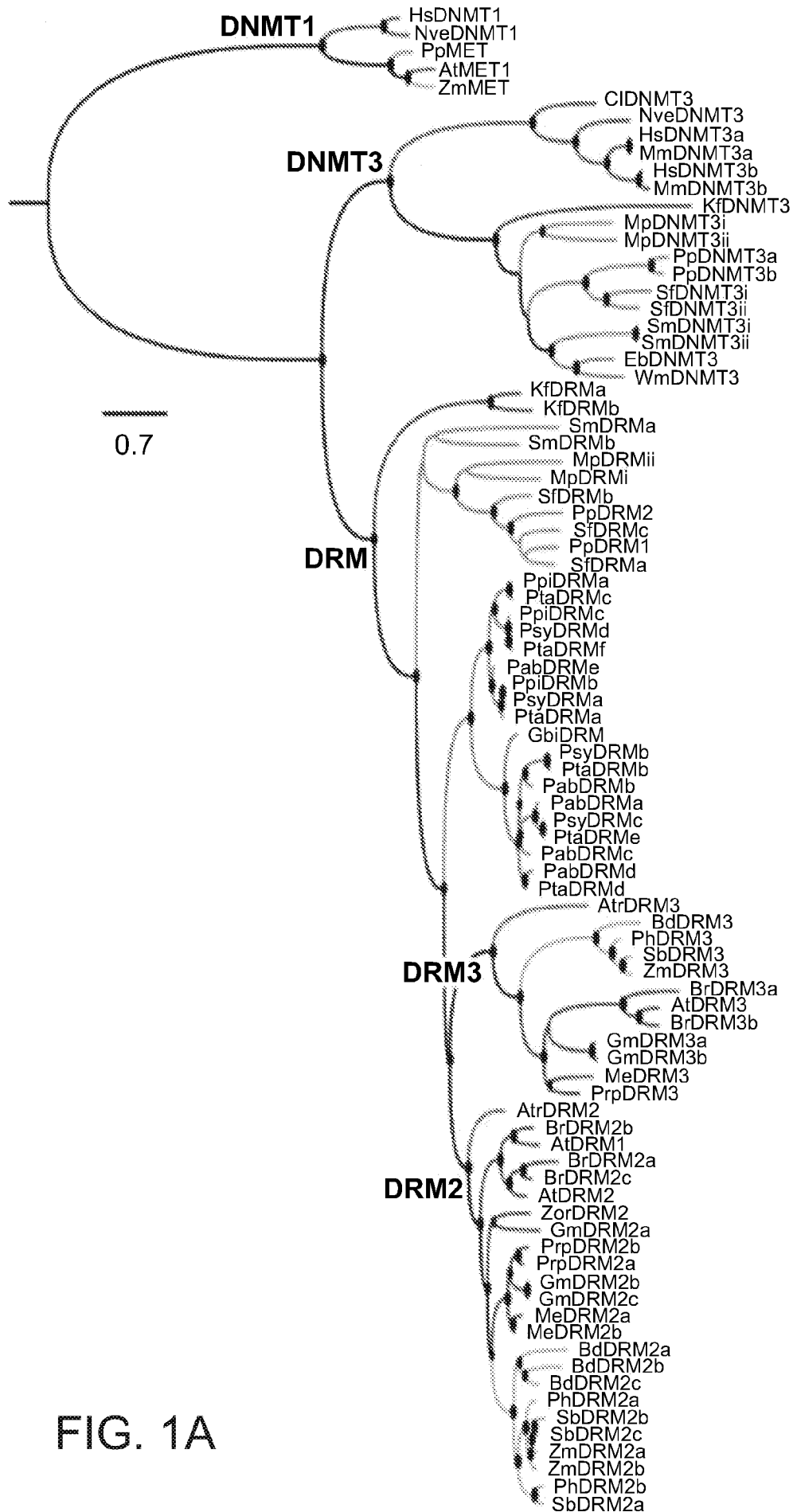


FIG. 1A

FIG. 1B

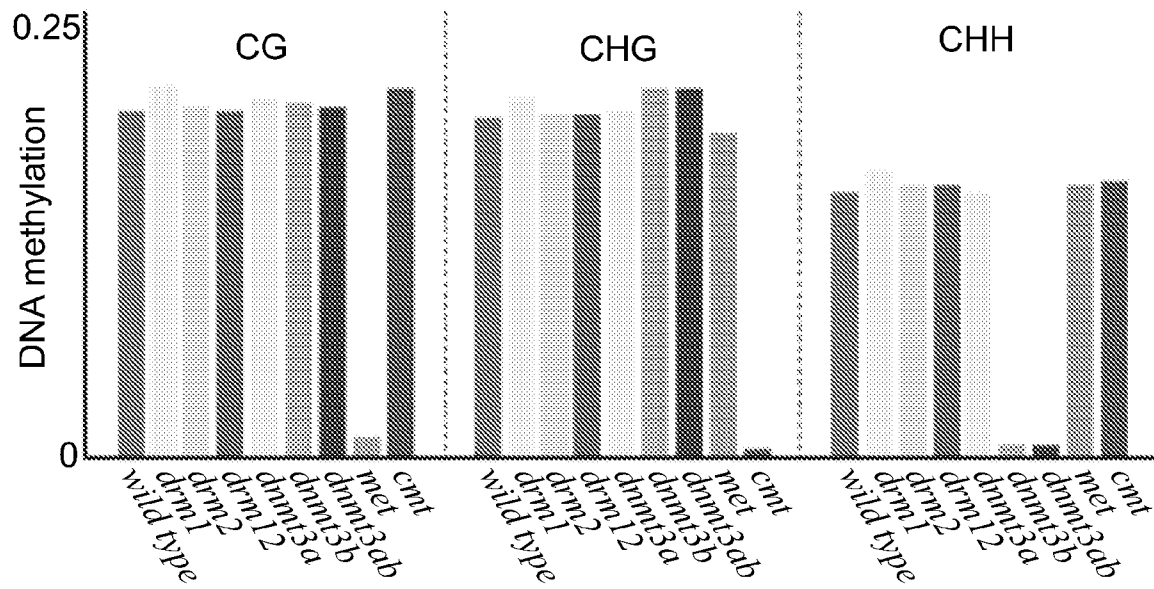
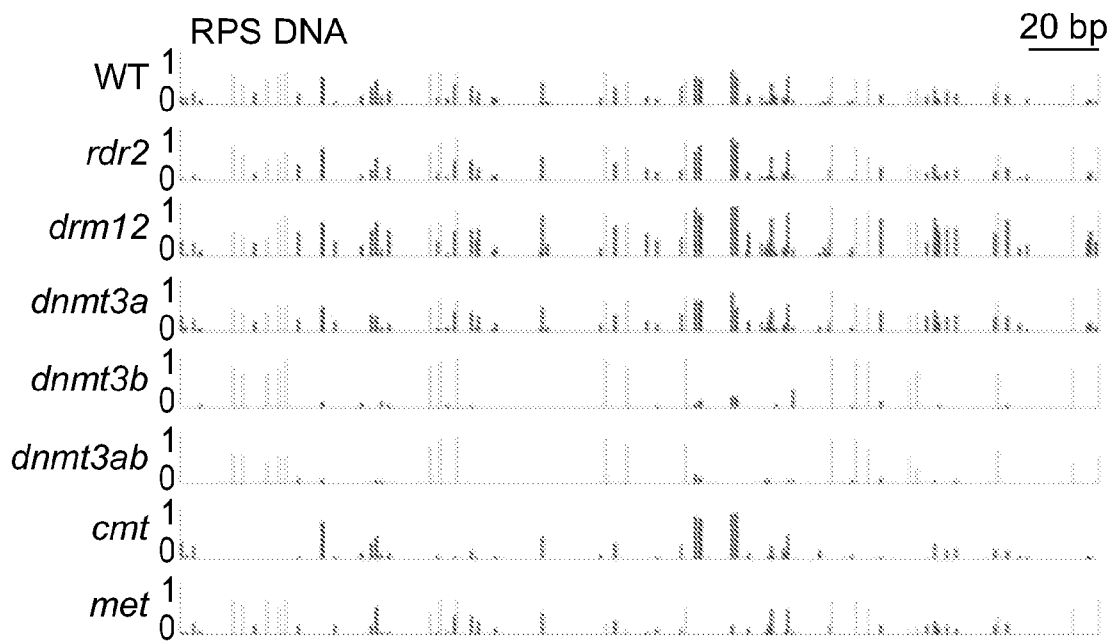


FIG. 1C



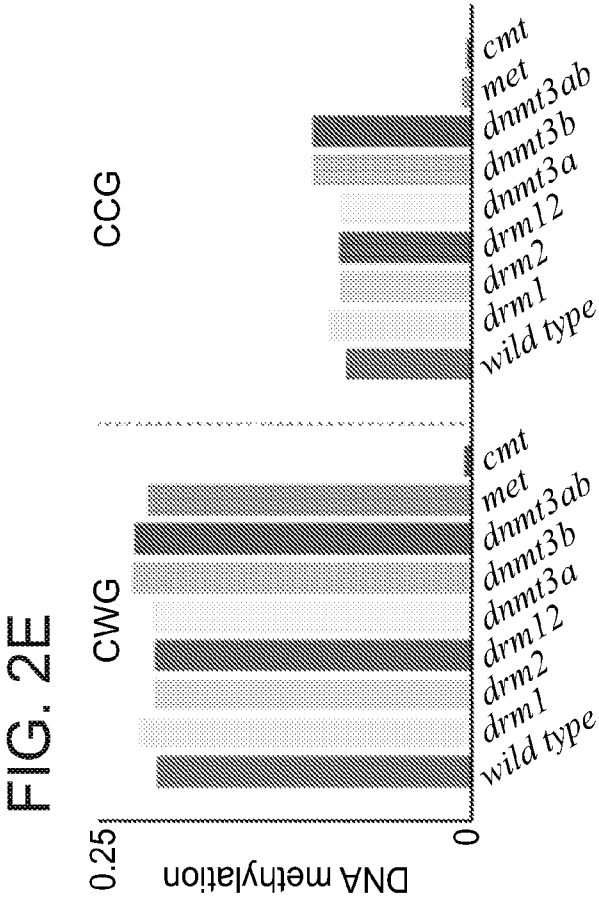
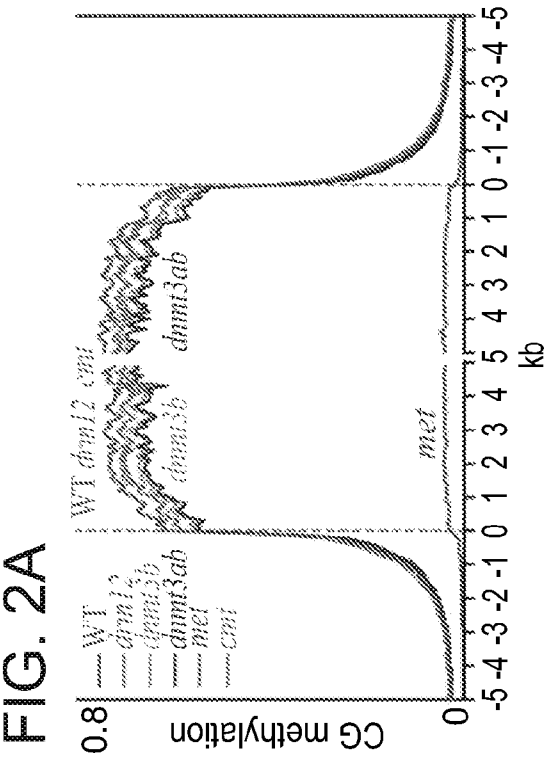
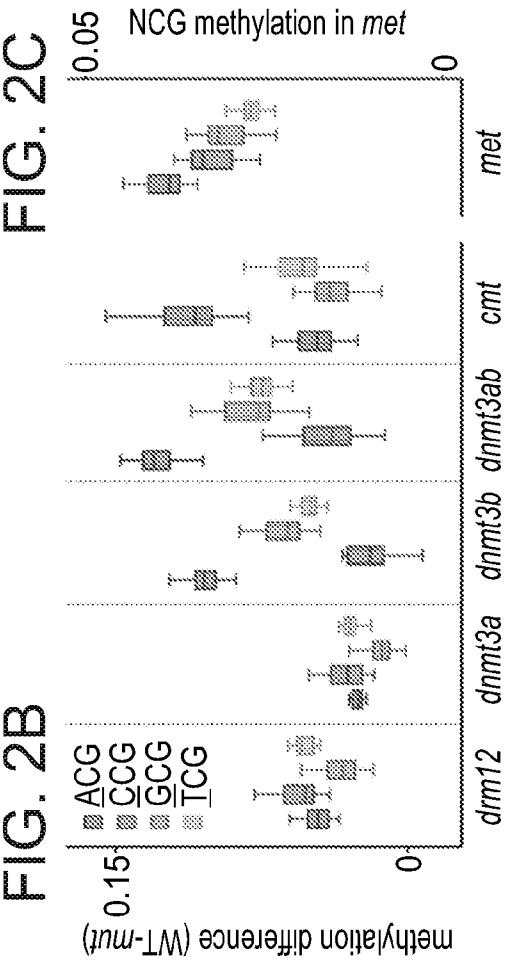


FIG. 3A

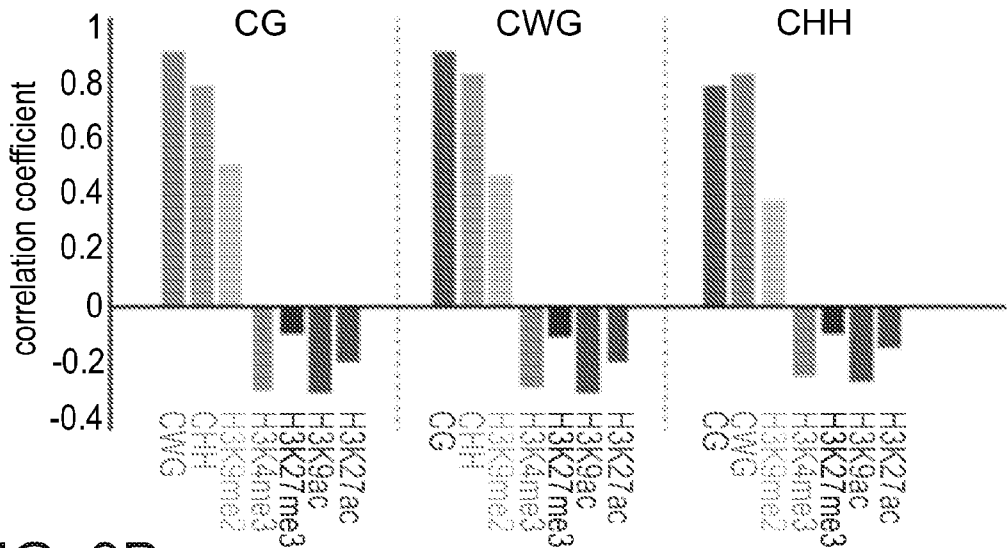


FIG. 3B

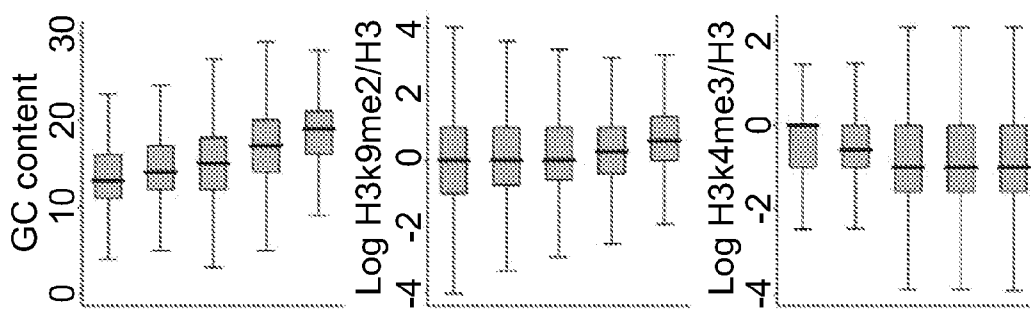


FIG. 3C

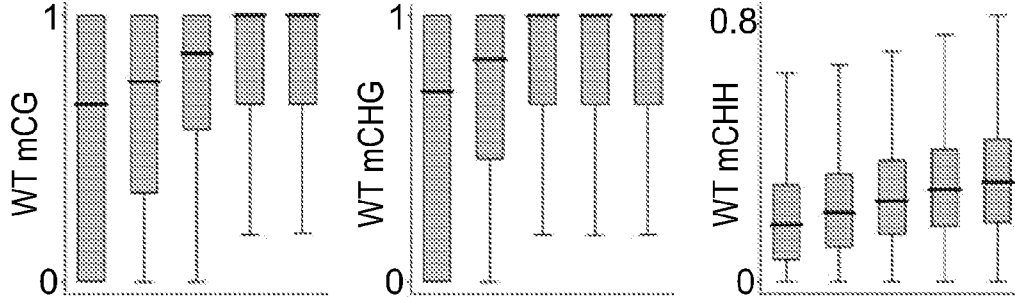


FIG. 3D

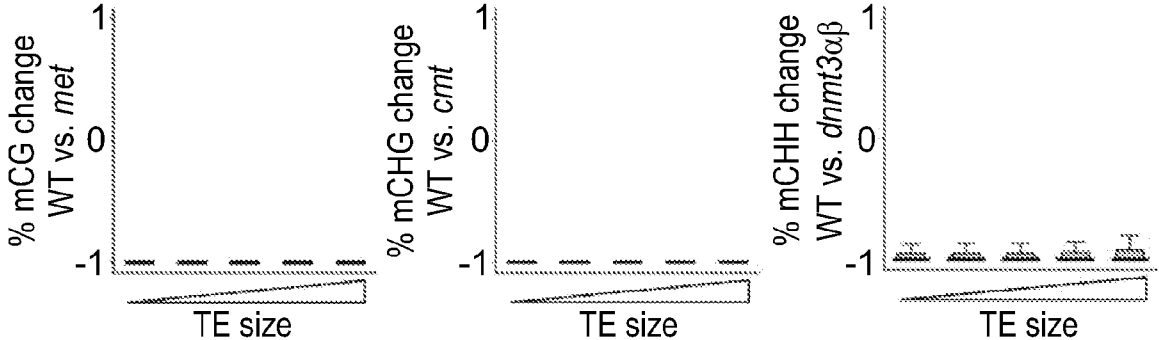


FIG. 3E

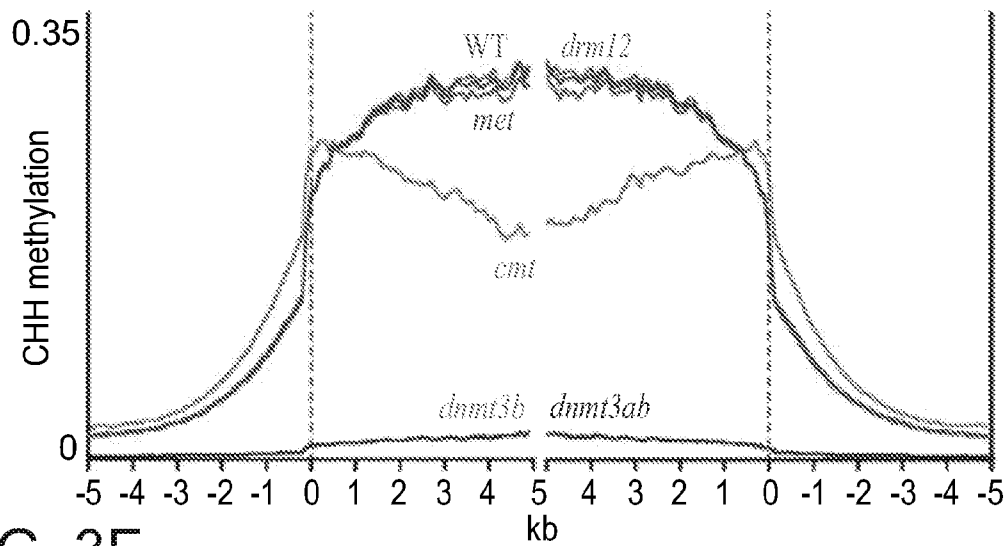


FIG. 3F

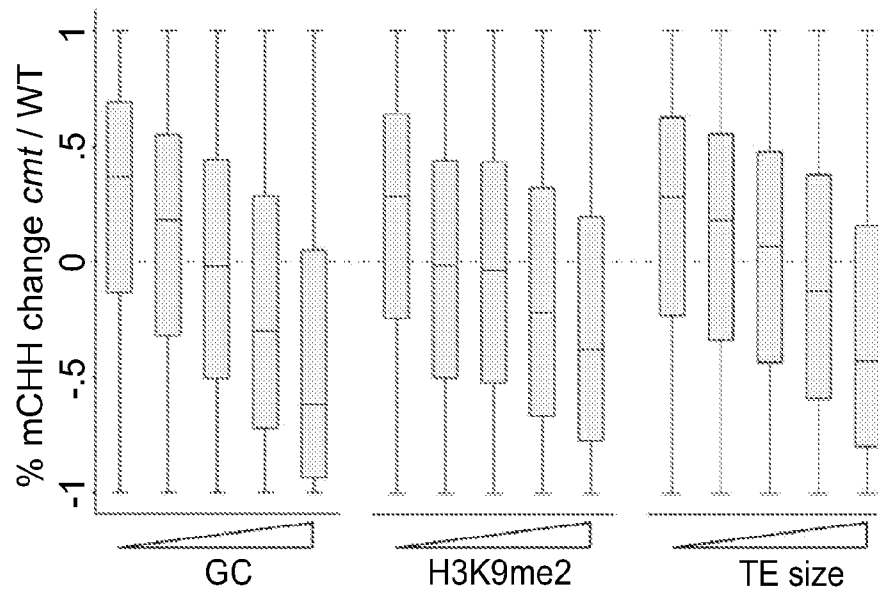


FIG. 3G

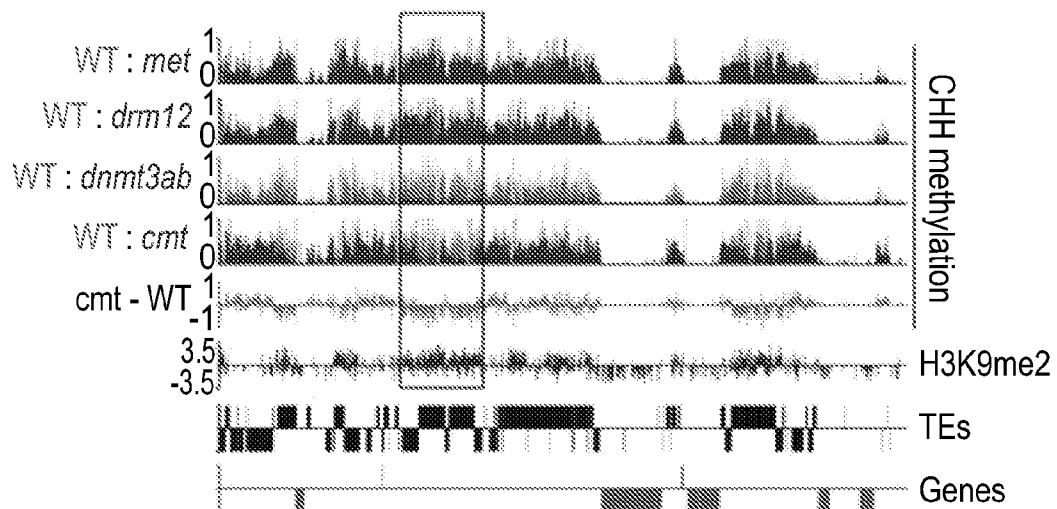


FIG. 4A

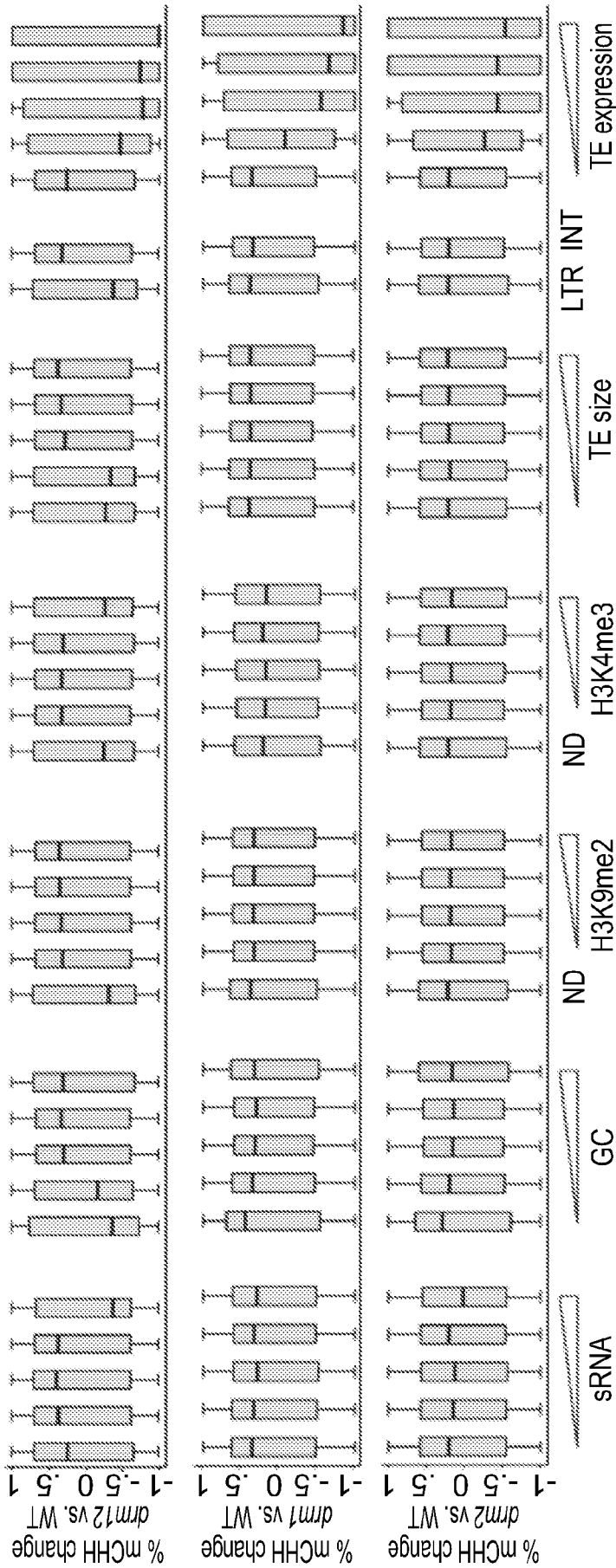


FIG. 4B

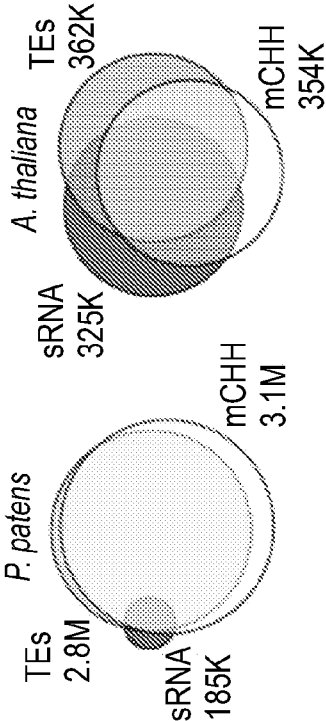


FIG. 4C

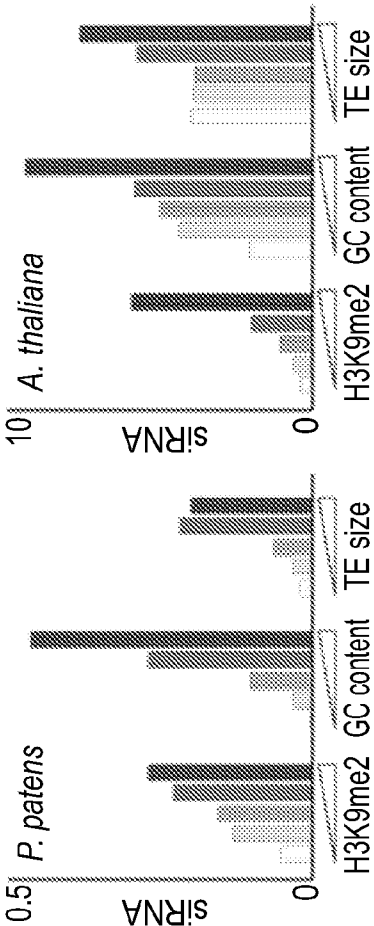


FIG. 4D

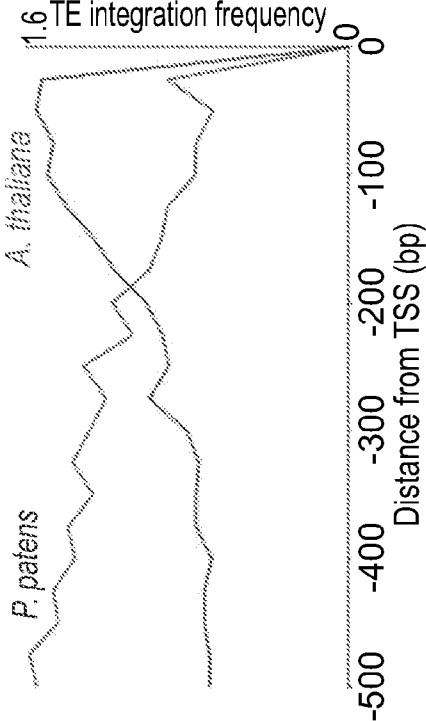
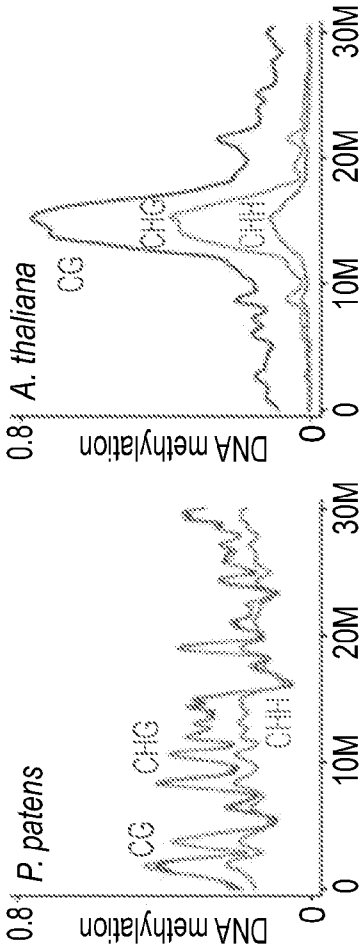


FIG. 4E



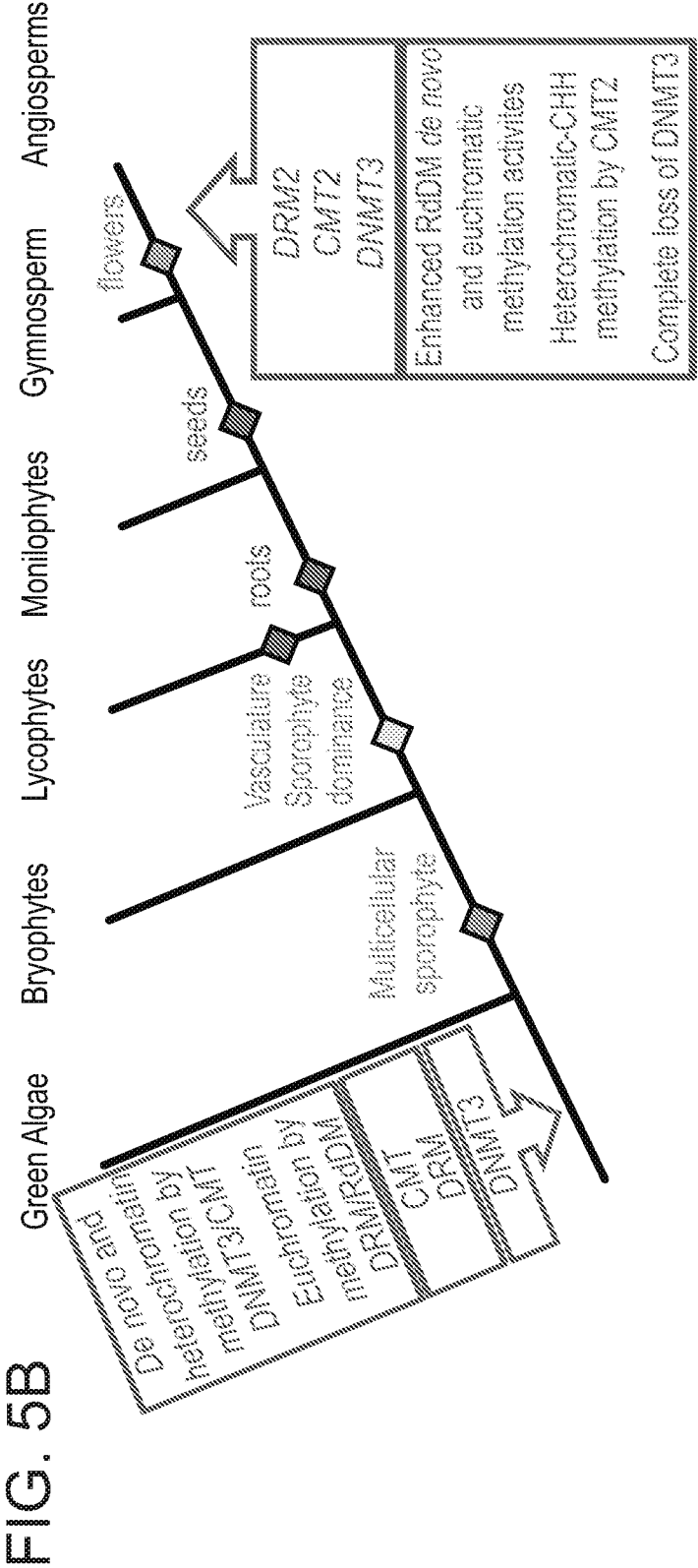
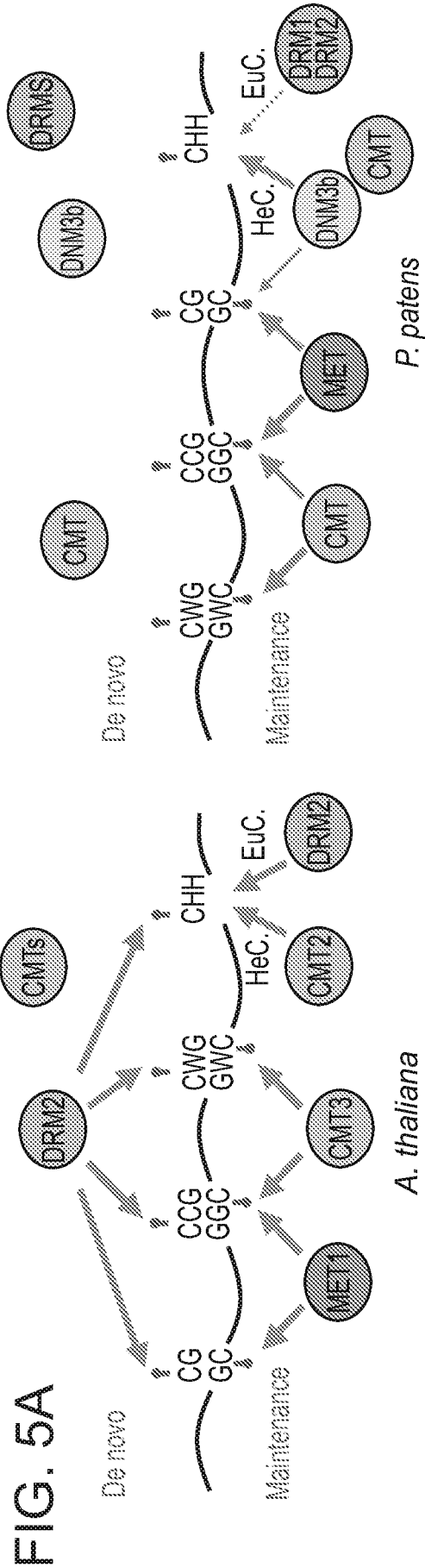


FIG. 6

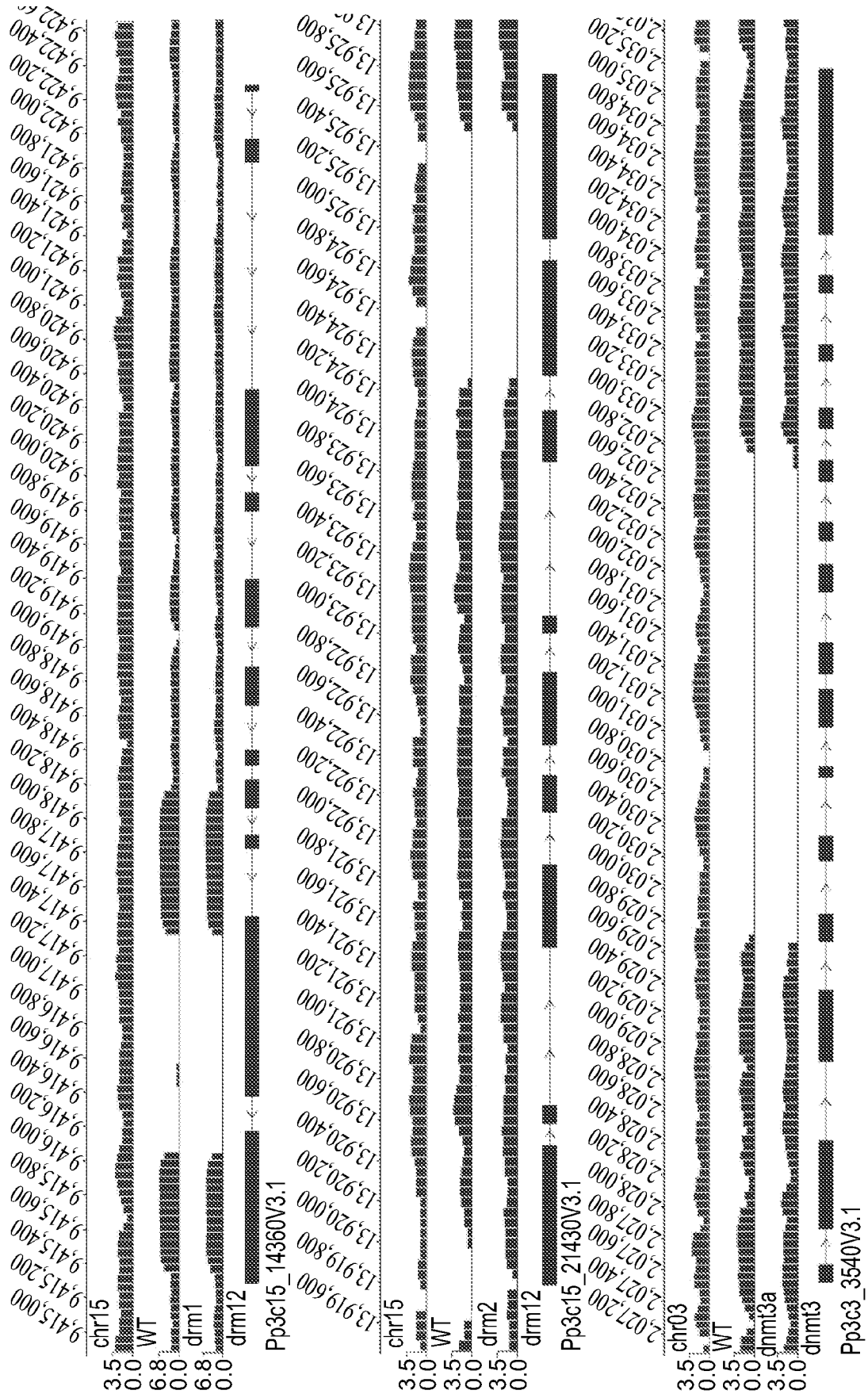


FIG. 6 continued

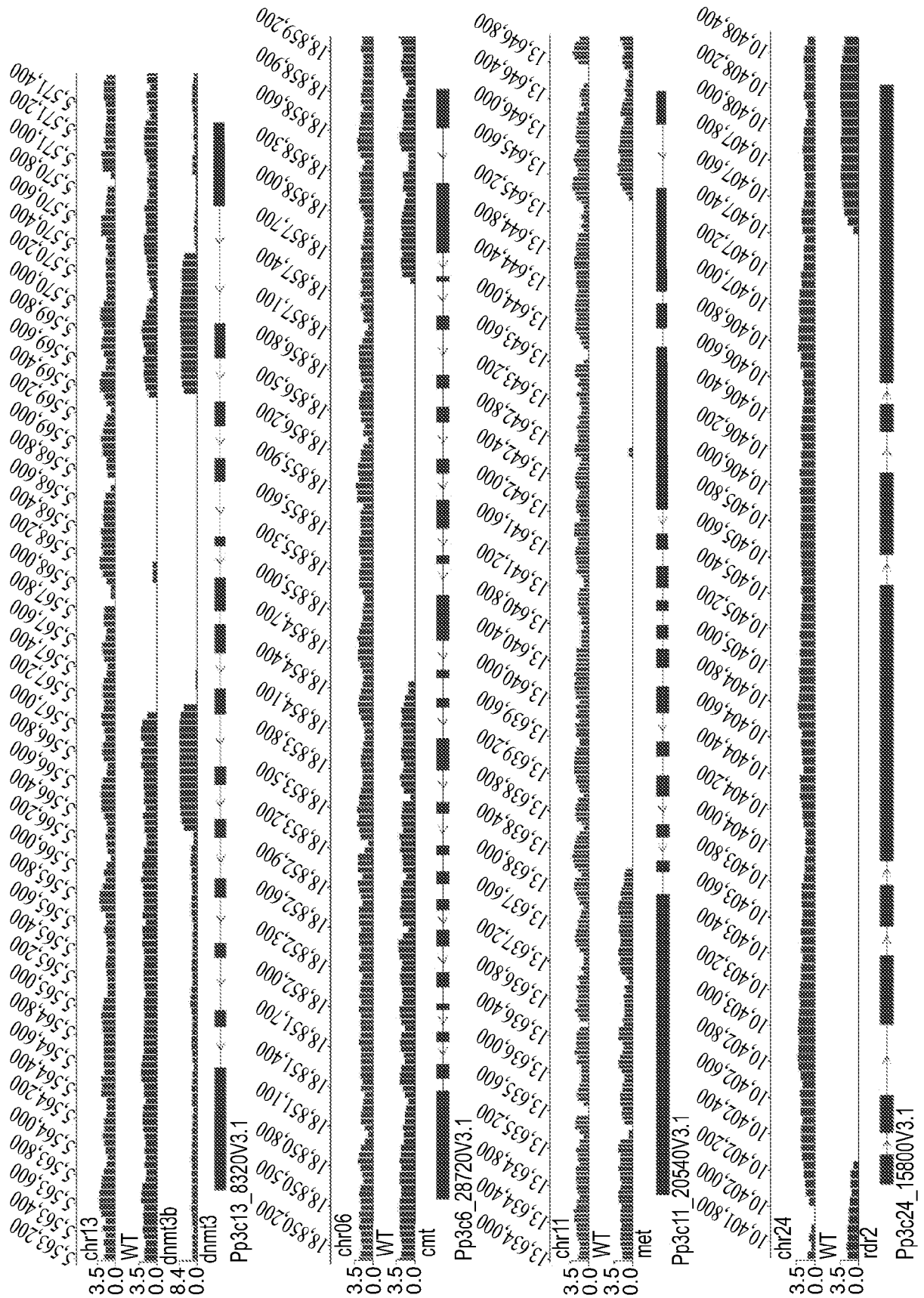


FIG. 7A Wild type

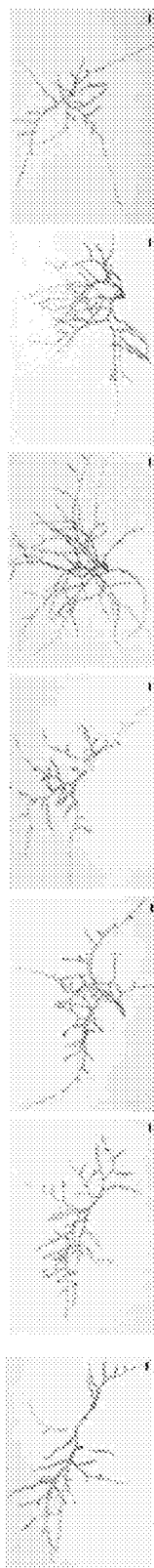


FIG. 7B *drm1*

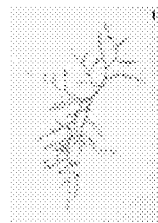


FIG. 7C *drm2*

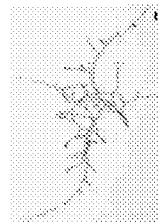


FIG. 7D *drm12*

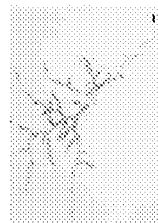


FIG. 7E *dnmt3a*

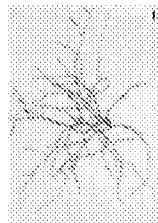


FIG. 7F *dnmt3b*



FIG. 7G *dnmt3ab*

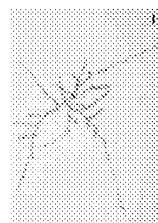


FIG. 7H FIG. 7I FIG. 7J FIG. 7K FIG. 7L FIG. 7M FIG. 7N



FIG. 7O FIG. 7P FIG. 7Q FIG. 7R FIG. 7S FIG. 7T FIG. 7U

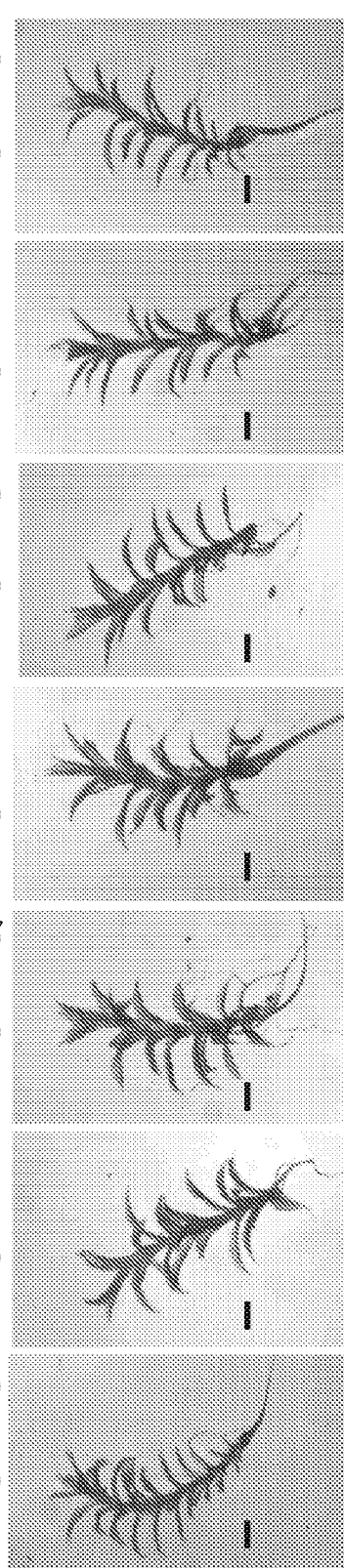


FIG. 8

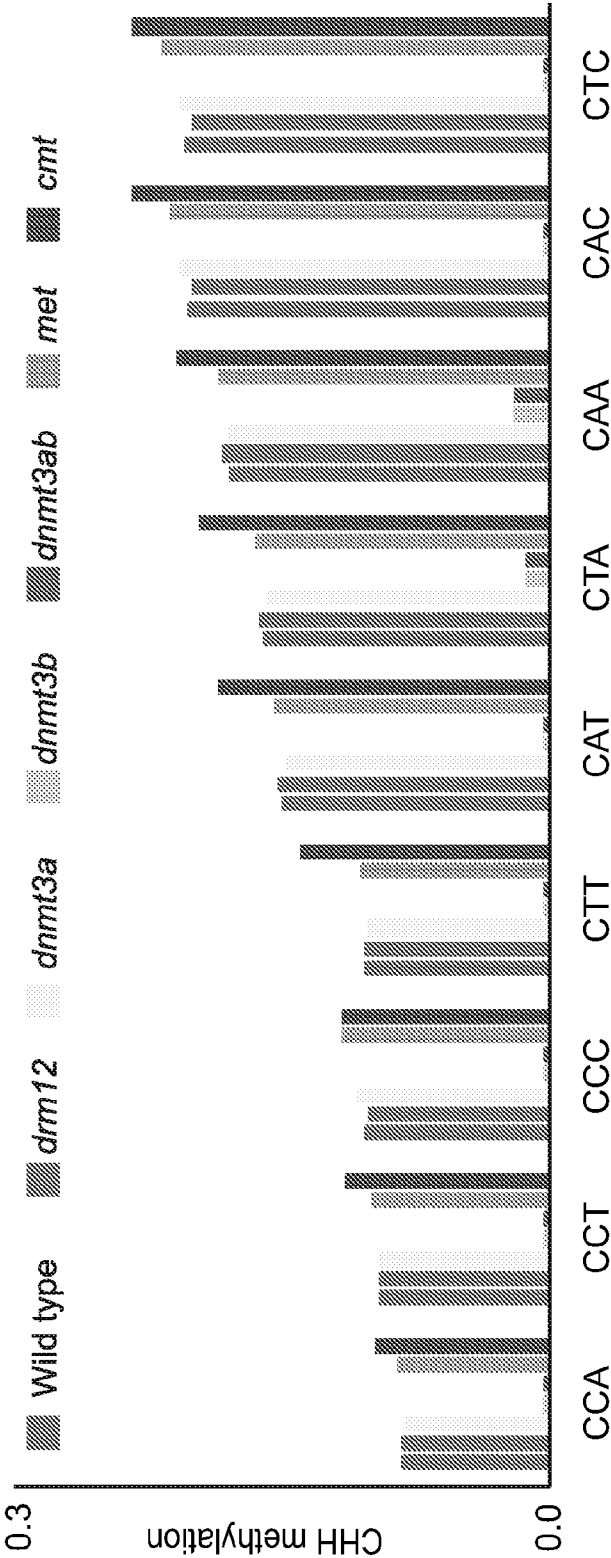


FIG. 9

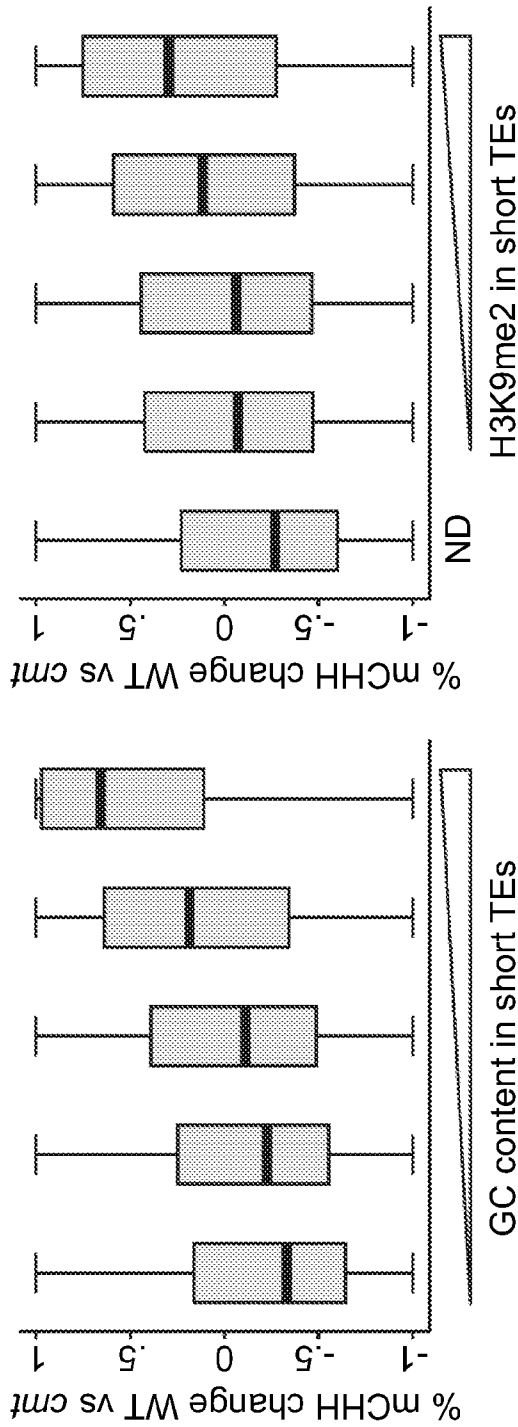


FIG. 10A

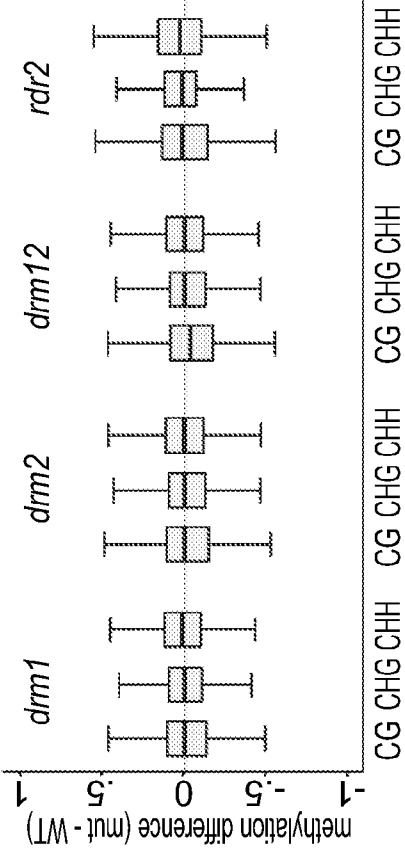


FIG. 10B

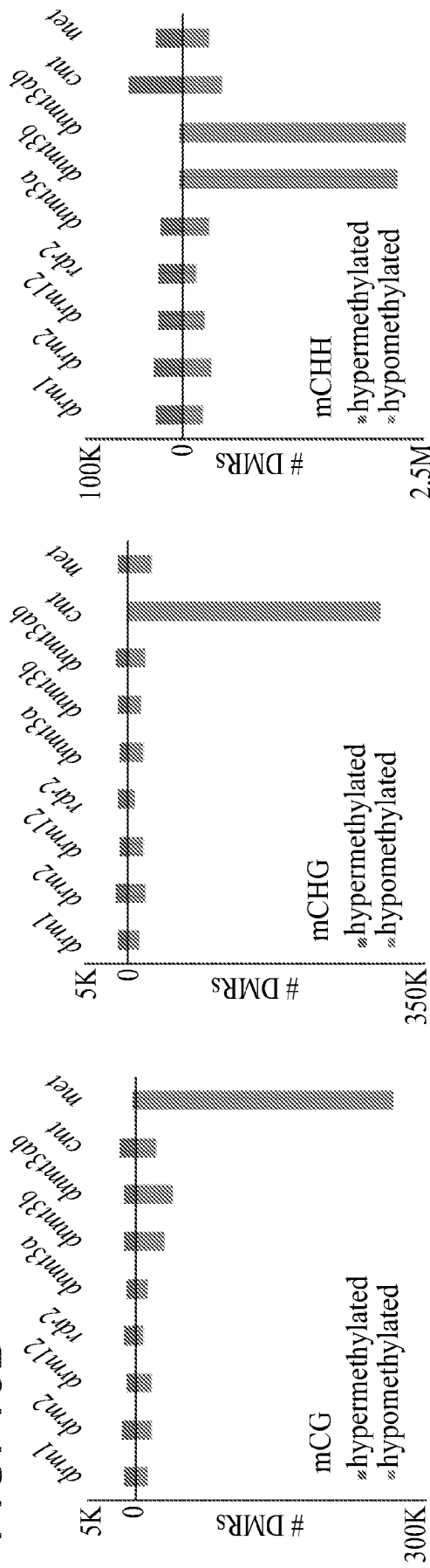


FIG. 10C

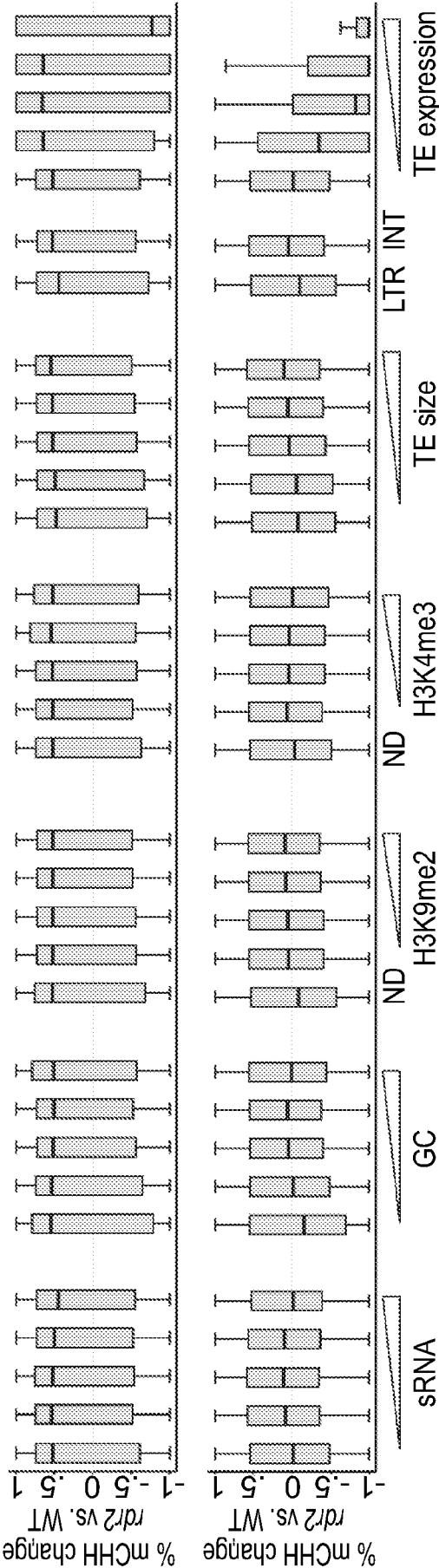


FIG. 10D

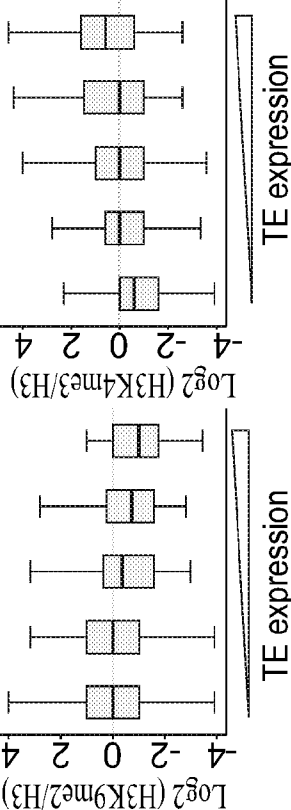


FIG. 10E

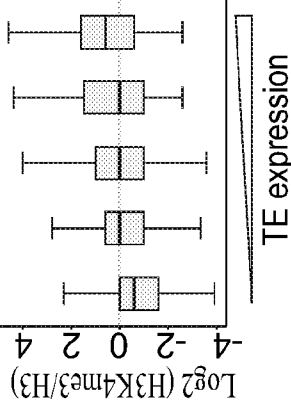


FIG. 11

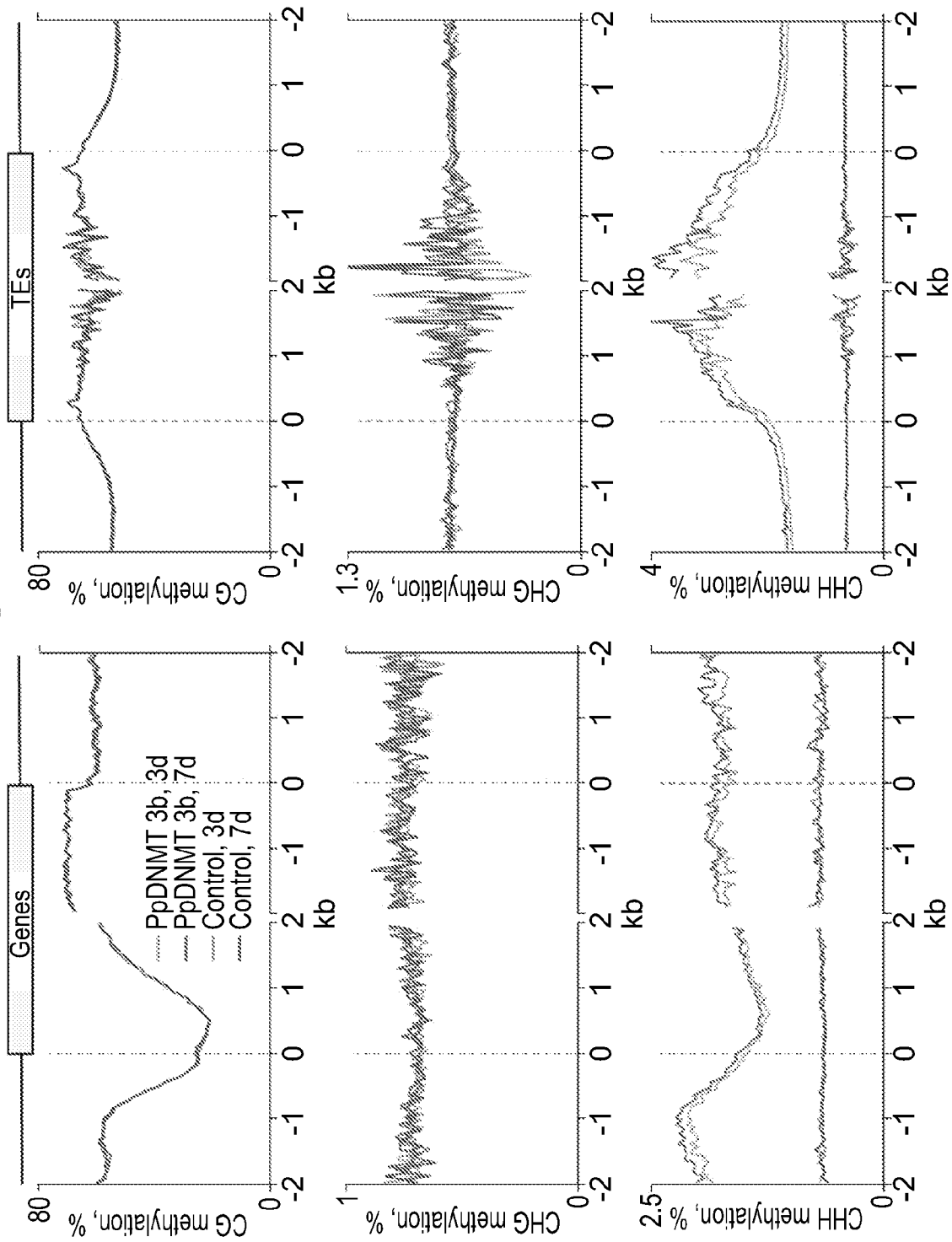


FIG. 12

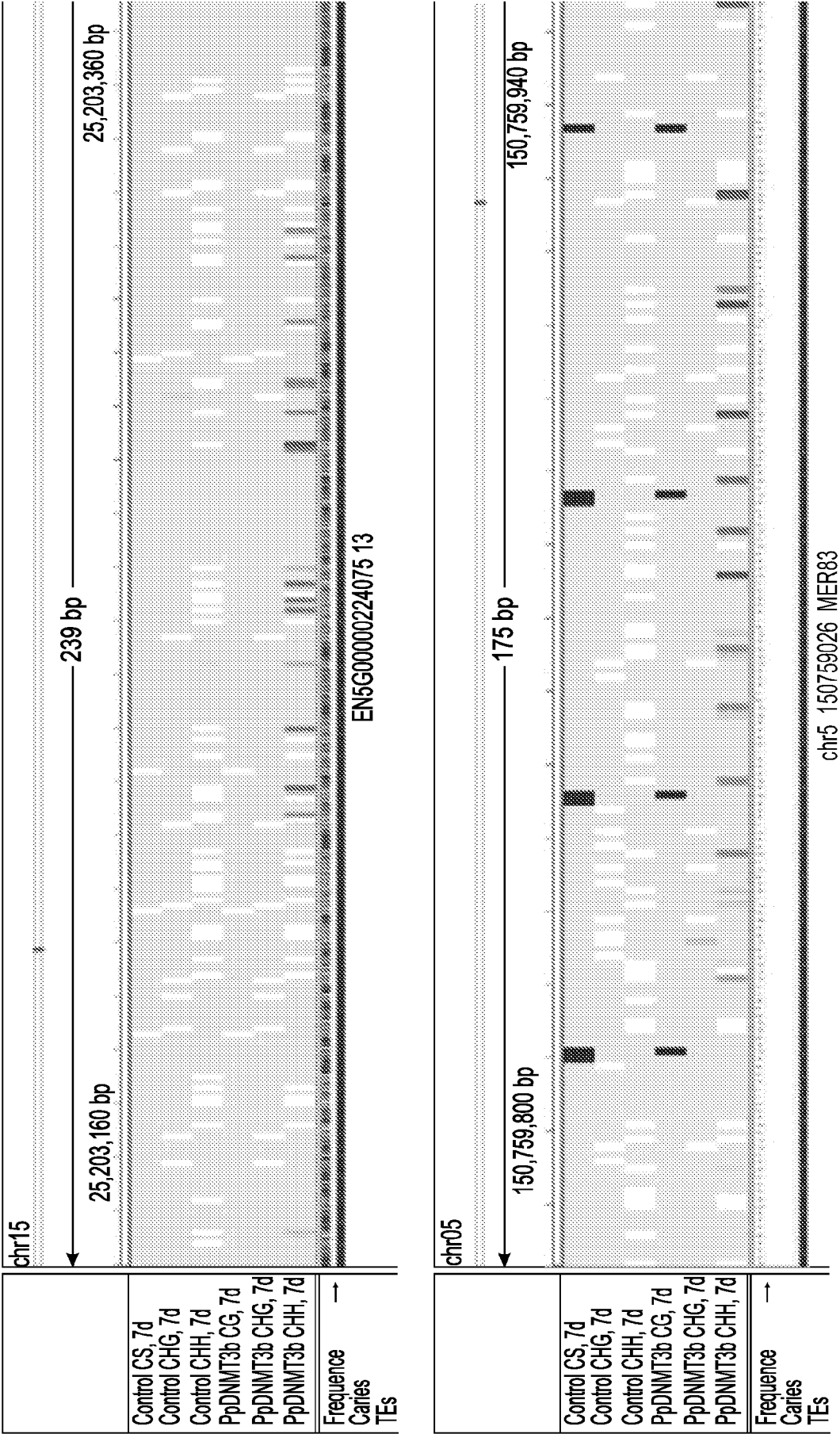


FIG. 13

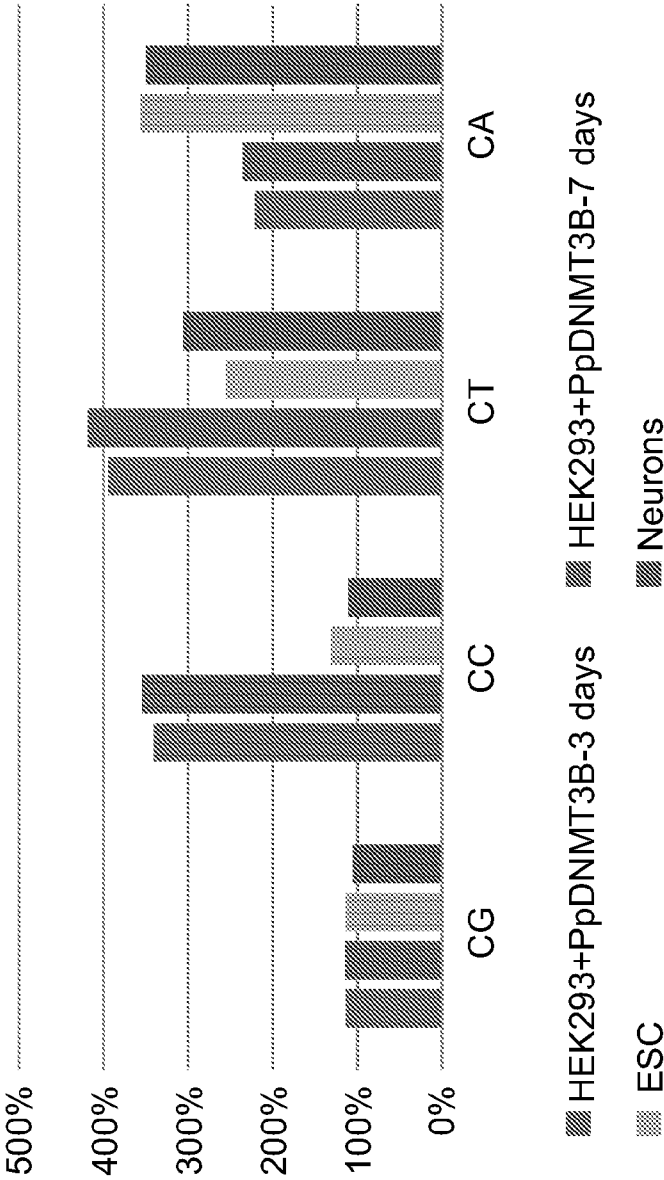


FIG. 14

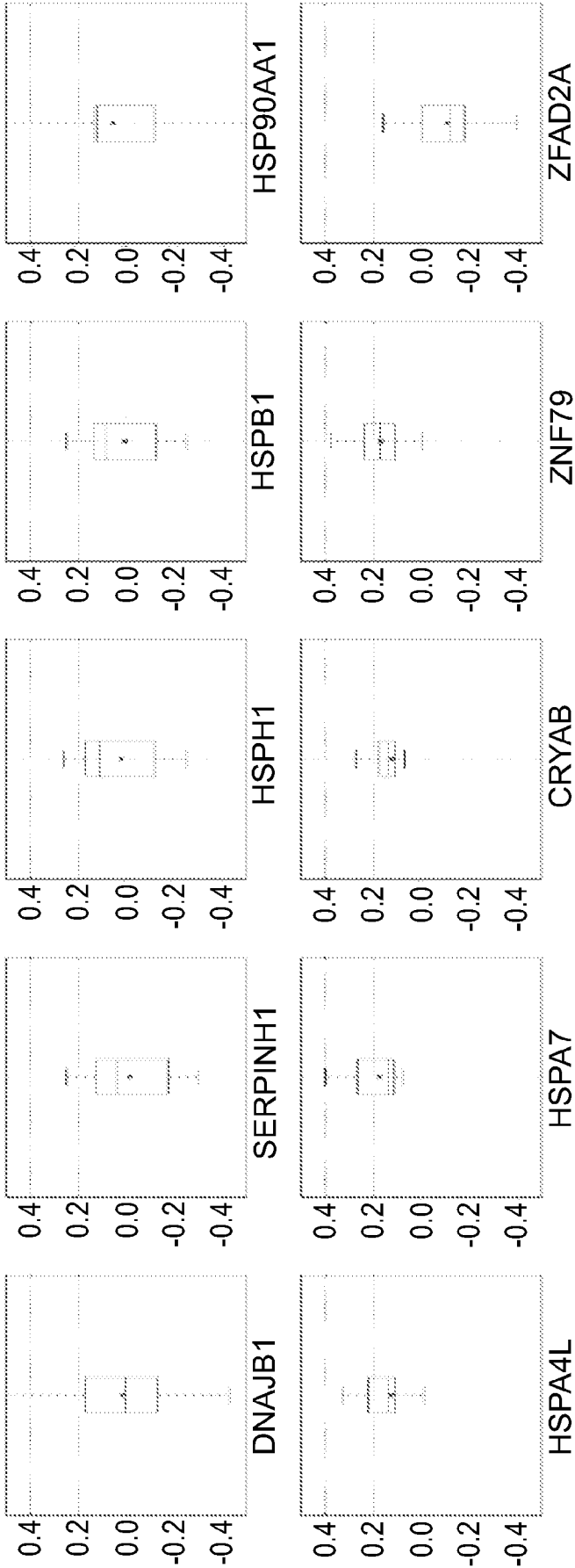
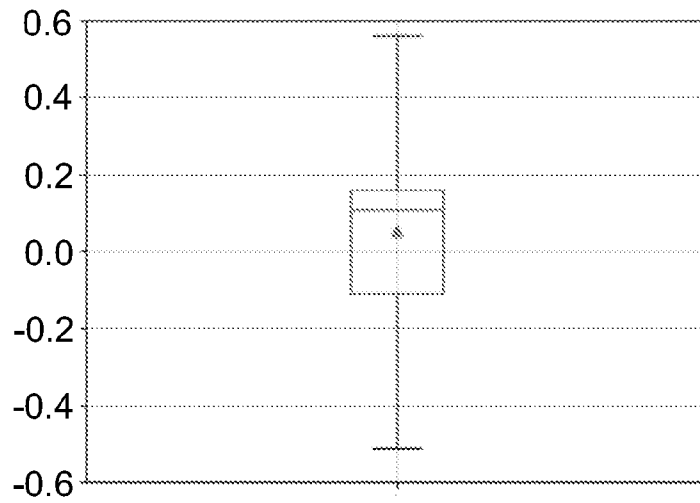


FIG. 15



21/24

FIG. 16

atggactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataaga
tggccccaagaagaagcggaaggctcggtatccacggagtcccagcagcc gacaagaagtacagcatcg
gcctggccatcggcaccacactctgtgggctggggccgtgatcaccgacgagtacaagggtgcccagcaagaaa
ttcaagggtgctgggcaacaccgaccggcacagcatcaagaagaacctgatcggagccctgctgttcgacag
cggcgaaacagccgagggccacccggctgaagagaaccgccagaagaagatacaccagacggaagaacc
ggatctgctatctgcaagagatcttca gcaacgagatggccaagggtggacgacagcttctccacagactgg
aagagtccttctggtggaagaggataagaagcacgagcggcaccctcttcggcaacatcgtggacga
gggtggcctaccacgagaagtagcccccacatctaccacctgagaaagaaaactgggtggacagcaccgacaag
gccgacctgcggtgatctatctggccctggcccacatgatcaagttccggggccacttctgatcgaggg c
gacctgaaccccgacaacagcgacgtggacaagctgttcatccagctgggtgcagacctacaaccagctgtt
cgaggaaaaccccatcaacgccagcggcgtggacgccaaggccatcctgtctgccagactgagcaagagc
agacggctggaaaaatctgatcgcccagctgccggcgagaagaagaatggcctgttcggcaacctgattgc
cctgagcctgggctgacccccaacttcaagagcaacttcgac ctggccgaggatgccaaaactgcagctga
gcaaggacacctacgacgacgacctggacaacctgtctggcccagatcggcgaccagtacgccgacctgtt
ctggccccaagaacctgtccgacgccatcctgtctgagcgacatcctgagagtgaaacaccgagatcaccaa
ggccccctgagcgcctctatgatcaagagatacagcagcaccaccaggacctgacctgctgaaagctc
tcgtgcggcagcagctgcctgagaagtacaaagagattttcttcgaccagagcaagaacggctacgccggc
tacattgacggcggagccagccaggaagagttctacaagttcatcaagcccatcctggaaaagatggacgg
caccgaggaactgctcgtgaagctgaacagagaggacctgctgcggaagcagcggaccttcgacaacggc
agcatccccaccagatccacctgggagagctgcacgccattctgcggcggcaggaaga ttttaccattc
ctgaaggacaaccgggaaaagatcgagaagatcctgaccttcgcatcccctactacgtggggccctctggc
caggggaaacagcagattcgcttgatgaccagaaagagcagggaaaccatcacccctggaaacttcgag
gaagtgggtggacaaggggcgcttcgccccagagcttcatcgagcggatgaccaacttcgataagaacctgcc
caacgagaagggtgctgcccagcacagcctg ctgtacgagtacttcaccgtgtataacgagctgaccaaag
tgaaatacgtgaccgaggggaatgagaaaagcccgccctcctgagcggcagcagaaaaaggccatcgtgga
cctgctgttcaagaccaaccggaaagtgaacctgaagcagctgaaagaggactacttcaagaaaatcgag
tgcttcgactccgtggaaatctccggcgtgggaagatcgggtcaacgcctcctgggacacataccacgatctgc
tgaaaattatcaaggacaaggacttcctggacaatgaggaaaacgaggacattctggaagatatcgtgctg
accctgacactgtttgaggacagagagatgatcgaggaaacggctgaaaacctatgccacctgttcgacga
caaagtgatgaagcagctgaagcggcggagatacaccggctggggcaggctgagccggaagctgatcaa

22/24

FIG. 16continued1

cggcatccgggacaagcagtcgggcaagacaatcctggatttctgaa gtccgacggcttcgccaacagaa
acttcatgcagctgatccacgacgacagcctgaccttaaaagaggacatccagaaagcccagggtgtccggc
cagggcgatagcctgcacgagcacatlgccaatctggccggcagccccgccattaagaaggggcatcctgca
gacagtgaaggtggtagcagctcgtgaaagtgatgggcccggcacaagcccagagaacatcgtgatcgaa
atggccagagagaaccagaccacccagaaggggacagaagaacagccgcgagagaatgaagcggatcga
agagggcatcaaagagctgggagccagatcctgaaagaacaccccgtaggaaaacacccagctgcagaa
cgagaagctgtacctgtactacctgcagaatgggagggatgtacgtggaccaggaactggacatcaacc
ggctgtccgactacgatgtggacgctatcgtgcctcagagcttctgaaggagcagctccatcgacaaca agg
tgctgaccagaagcgacaagaaccgggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaaga
tgaagaactactggcggcagctgctgaacgccaagctgattaccagagaaagtgcacaatctgaccaag
gccgagagagggcgccctgagcgaactggataaggccggcttcatcaagagacagctggtaggaaacccggc
agatcacaaagcacgtggcacagatcctggactcccgatgaa cactaagtacgacgagaatgacaagct
gatccgggaagtgaagtgatcacctgaagtccaagctgggtgtccgatttccggaaggatttccagttta
caaagtgcgcgagatcaacaactaccaccacgcccacgacgcctacctgaacgccgtcgtgggaaccgcc
ctgatcaaaaagtaccctaagctggaaagcgagttcgtgtacggcgactacaagggtgtacgacgtgcggaa
gatgatcgccaagagc gagcaggaaatcggcaaggctaccgccaagtacttcttctacagcaacatcatga
acttttcaagaccgagattaccctggccaacggcgagatccggaagcggcctctgatcgagacaaacggc
gaaaccggggagatcgtgtgggataaggggccgggattttgccaccgtgcggaaagtgtgctgagcatgcccc
agtgaatatcgtgaaaaagaccgaggtgcagacaggcggcttcagcaaagagictatcc tgcccaagagg
aacagcgataagctgatcgccagaaagaaggactgggaccctaagaagtacggcggttcgacagcccca
ccgtggcctattctgtgtggtgggccaagtggaaaaggggcaagtccaagaaactgaagagtgtagaa
gagctgctggggatcacatcatggaaagaagcagcttcgagaagaatcccatcgactttctggaagccaa
gggctacaaagaagtgaaaaaggacctgatcat caagctgcctaagtactccctgttcgagctggaaaacg
gccggaagagaatgctggcctctgccggcgaactgcagaagggaacgaactggccctgccctccaaata
tgtgaacttctgtacctggccagccactatgagaagctgaagggtctccccgaggataatgagcagaaac
agctgtttgtggaacagcacaaagcactacctggacgagatcatcgagcagatcagcaggttctccaagaga
gtgatcctggccgacgctaactctggacaaagtgtgtccgcctacaacaagcaccgggataagcccatcag
agagcaggccgagaatatcatccacctgtttacctgaccaatctgggagcccctgccgccttcaagtacttt
gacaccaccatcgaccggaagaggtacaccagcaccaaagaggtgctggacgccacctgatccaccaga
gcatcacccggcctgtacgagacacggatcgacctgtctcagctgggag gcgacaaaaggccggcgccac
gaaaaaggccggacaggccaaaaaagaaaagctcgaggggcgaggcgggagcgGATCCGCAGAG
GCAGGCCCTGGTCCGGAGAGCGCCTGGTGGTGTCTGTCTCTGTTTCGACGGCCTGG

23/24

FIG. 16continued2

GAGGCATCTGGCAGGCCCTGACCAAGCTGGGCATCCCTTTTTCTGGATACTCTAGC
GAGGTGCTGGCCCCAGCAATCCAGGTGGTGAAGAGCCGCCACCCTCGGGTGAAG
CACGTGGGCGACATCCGGAAGCTGAACCTGAGCGCCGTGCCAGAGAAAGGTGGAC
CTGGTGGTGGGAGGATTCCCATGCCAGGATCTGTCCATCATGGGCAAGAAGGAG
GGCCTGCACGGCTCCCGGTCTAAGCTGTTCTTTGACCTGCTGAGAGTGCTGAAGG
TGTTCAAGCCTAAGTGGTTTCTGGTGGAGAATGTGGCCAGCATGTCCTGGGTGGA
CAGGGAGGAGATCACACGCCACCTGAAGGTGGCCCCAATGGAGCTGGATTCTCA
GGAGATCACCGCCAGCAAGCGGAGAAGGCTGTATTGGACAAACATCCCACACCC
ACCTAGACTGCCCCGCCTGCGGGATCACCCAGCACCCAGCCTCCAGTCCTGTCTGG
AGGGCGCCCTGGCCCTGGAGCAGAAGTGCGGCGTGATCCTGTGCAGCAATCTGT
ACAAGGGCTCTACCGCACGGCTGGAGCTGGTGCTGGACAACAAGACCAATAAGC
TGAGATATATCAAGCAGACAGAGGTGG AGGTGCTGATGGGCTACCCAAAGGATT
ATACCAACGTGGTGGCCACGAGACAAAGGGCAGGACAGAGCAGGCCGAGAAG
GTGCTGAAAACCCCCGTGCGCGCAAAGAGCGTGGAGCCTAAGCCATCCTCTGTGA
CACCACCCTCCGGCCGGcccgaattcggcagtggagagggcagagggaagtcgtgtaacatgcggtg
acgtcgaggagagaatcctggccca atgaccgagtacaagcccacgggtgcgcctcgc caccgcgacgacgt
ccccagggccgtacgcaccctcgccgccggttcgccgactacccgccacgcgccacaccgtcgatccgg
accgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaag
gtgtgggtcgcggaacgacggcgccggtggcggtctggaccacgccggagagcgtcgaagcggggcg
tgttcgccgagatcgccccgcgcattggcc gagttgagcgggtcccggtggccgcgcagcaacagatggaa
ggcctcctggcgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctgcccaccac
cagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcgccgagcgcgcgggggtgccc
ccttcctggagacctccgcgccccacaacctccccttctacgagcggctcggcttcaccgtcaccgccgacg t
cgaggtgcccgaaggaccgcgcacctggtgcatgaccgcgaagcccgggtgcctga

24/24

FIG. 17

MDYKDHGDKDHDIDYKDDDDKMAPKKRKVGIHGVPAA DKKYSIGLAIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLLYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQL
FEENPINASGVDAKILSARLSKSRRENLIQAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAQLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP
LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKD
NREKIEKILTRIPYVVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQAQSFIER
MTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENE
DILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDK
QSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGK
ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPOSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
DFRKDFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD KRPAATKKAGQAK
KKKLEGGGGSGSAEAGPWSEGERLVVLSLFDGLGGIWQALT KLGI PFSGYSSEVLAPAI
QVVKS RHPRVKHVGDIRKLNLSAVPEKVDLVVGGFPCQDLSIMGKKEGLHGSRSKLFF
DLLRVLKVFKPKWFLVENVASMSWVDREEITRHLKVAPMELDSQEITASKRRRLYWT
NIPHPRLPRLRDHPSTSLQSCLEGALALEQKCGVILCSNLYKGSTARLELVLDNKTNKL
RYIKQTEVEVLMGYPKDYTNVVAHETKGRTEQA EKVLKTPVRAKSVEPKPSSVTPPS G
RPEFGSGEGRGSLLTCGDVEENPG

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL2020/050254

A. CLASSIFICATION OF SUBJECT MATTER IPC (20200101) C12N 15/62, C12N 15/113, A01H 5/00 CPC (20130101) C12N 15/62, C07K 2319/80, C12N 15/1137, A01H 5/00 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC (20200101) C12N 15/62, C12N 15/113, A01H 5/00 CPC (20130101) C12N 15/62, C07K 2319/80, C12N 15/1137, A01H 5/00 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Databases consulted: NCBI, Google Patents, CAPLUS, BIOSIS, EMBASE, MEDLINE, Google Scholar Search terms used: plant DNMT3 DNA targeting CRISPR Cas		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2019032049 A1 Ospedale San Raffaele S.R.L. ; Fondazione Telethon 31 Jan 2019 (2019/01/31) claims 1, 3 and 8; description paragraphs [0233], [0236], [0244] and [0261]	28-31
Y	whole document, especially claims 1, 3, 8, 23 and 39; description paragraphs [0233], [0236], [0244], [0261], [0355-0363], [0414] and [0422]; figures 2A and 19A	1-6,10-27,32-34,37, 38,40,42-45,47,48
Y	DATABASE NCBI [Online] 12 April 2018 (2018/04/18). Uncharacterized protein LOC9640092 isoform X2 [Selaginella moellendorffii]. GeneBank accession number: XP_002971634.2. 12 Apr 2018 (2018/04/12) whole document	1-6,10-27,32-34,37, 38,40,42-45,47,48
A	Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science, 328(5980), 916-919. 15 Apr 2010 (2010/04/15) whole document	1-49
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 09 Jun 2020		Date of mailing of the international search report 09 Jun 2020
Name and mailing address of the ISA: Israel Patent Office Technology Park, Bldg.5, Malcha, Jerusalem, 9695101, Israel Email address: pctoffice@justice.gov.il		Authorized officer MAZEL Alexander Telephone No. 972-73-3927174

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL2020/050254

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem., 74, 481-514. 31 Dec 2005 (2005/12/31) whole document	1-49

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IL2020/050254

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
US 2019032049 A1	31 Jan 2019	US 2019032049 A1	31 Jan 2019
		AU 2015334469 A1	25 May 2017
		CA 2965591 A1	28 Apr 2016
		CN 107109433 A	29 Aug 2017
		EP 3209783 A1	30 Aug 2017
		GB 201418965 D0	10 Dec 2014
		JP 2017537611 A	21 Dec 2017
		WO 2016063264 A1	28 Apr 2016