



(51) International Patent Classification:

A01H 1/02 (2006.01) C12N 15/113 (2010.01)

A01H 4/00 (2006.01) C12N 15/82 (2006.01)

(21) International Application Number:

PCT/US2017/029354

(22) International Filing Date:

25 April 2017 (25.04.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/327,662 26 April 2016 (26.04.2016) US

(71) Applicant: UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. [US/US]; Boyd Graduate Studies Research Center, Athens, GA 30602-7411 (US).

(72) Inventor: SCHMITZ, Robert, J.; 928 Hill Street, Athens, GA 30606 (US).

(74) Agent: PROVENCE, David, L.; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581336, Minneapolis, MN 55458-1336 (US).

(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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: PLANTS HAVING REDUCED METHYLATION OF CYTOSINE NUCLEOTIDES AND METHODS OF USE

(57) Abstract: Provided herein are non-natural plants, such as plants having at least one 5- hydroxymethylcytosine (5-hmC) in the genome; transgenic plants having a catalytic domain of a Tet protein, a polynucleotide encoding a catalytic domain of a Tet protein, or a combination thereof; or plants having a genome with reduced methylation of cytosine residues compared to a control plant. Also provided are methods for making and using such plants.



PLANTS HAVING REDUCED METHYLATION OF CYTOSINE NUCLEOTIDES AND
METHODS OF USE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 62/327,662, filed April 26, 2016, which is incorporated by reference herein.

GOVERNMENT FUNDING

This invention was made with government support under grant number R00GM100000, awarded by the NIH. The government has certain rights in the invention.

BACKGROUND

Cytosine DNA methylation is critical to maintaining transposon and gene silencing in plant genomes and failure to do so often result in lethality. Prior studies using the model plant *Arabidopsis thaliana* have demonstrated that using mutations that fail to maintain proper DNA methylation states results in a variety of novel morphologies; however, this approach used in *Arabidopsis thaliana* has not been successfully implemented in crop genomes. This is likely a result of the much larger size of crop genomes and that failure to maintain proper DNA methylation states is catastrophic to maintenance of genome integrity, and results in lethality.

SUMMARY

Provided herein are non-natural plants. In one embodiment, a plant includes at least one 5-hydroxymethylcytosine (5-hmC) in the genome. A 5-hmC can be immediately adjacent to and 5' of a guanine. In other embodiments, a 5-hmC is the cytosine nucleotide in CHG or CHH, where H is A, C, or T. In one embodiment, the plant is a transgenic plant that includes a catalytic

domain of a Tet protein, a polynucleotide encoding a catalytic domain of a Tet protein, or a combination thereof. Also provided is a cell of the plant described herein, where the cell includes the catalytic domain of the Tet protein, the polynucleotide encoding the catalytic domain of the Tet protein, or both.

5 Also provided herein is a non-natural plant that includes a genome with reduced methylation of cytosine residues compared to a control plant. In one embodiment, the reduction in methylation includes a reduction of the methylation of cytosine residues at CG dinucleotides. In another embodiment, the reduction in methylation includes a reduction of the methylation of cytosine residues at CG dinucleotides, a reduction of the methylation of cytosine residues at
10 CHH trinucleotides, where H is A, C, or T, a reduction of the methylation of cytosine residues at CHG trinucleotides, or a combination thereof. The reduction may be statistically significant. In one embodiment, at least one cell of the plant having a genome with reduced methylation of cytosine residues includes a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase (Tet) protein, a polynucleotide encoding a catalytic domain of a Tet protein, or a
15 combination thereof. In another embodiment, the plant having a genome with reduced methylation of cytosine residues does not include a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase (Tet) protein or a polynucleotide encoding a catalytic domain of a Tet protein, or a combination thereof. In one embodiment, such a non-transgenic plant is the progeny of a plant described herein that is transgenic for the Tet protein.

20 A plant described herein can be a monocot or a dicot. Also provided are the progeny of a plant described herein. In one embodiment, the progeny is a hybrid plant. In one embodiment, the plant is a crop plant, such as soybean, corn, tomato, rice, common bean, sugarcane, pumpkin, wheat, cassava, potato, cotton, a grass such as ryegrass or tall fescue, or a forage legume such as clover or alfalfa. In one embodiment, the plant is an algae. In one embodiment, the plant is a
25 mosaic. The plant can be homozygous or hemizygous.

 Also provided herein are methods. In one embodiment, a method includes providing a plant described herein (e.g., a plant includes at least one 5-hydroxymethylcytosine (5-hmC) in the genome; a transgenic plant that includes a catalytic domain of a Tet protein, a polynucleotide encoding a catalytic domain of a Tet protein, or a combination thereof; or a plant that includes a
30 genome with reduced methylation of cytosine residues compared to a control plant), and screening the plant for any change compared to a control plant. In one embodiment, the change

includes at least one phenotypic variation. In one embodiment, the change includes an alteration in flowering time, an alteration in root length, or an alteration in resistance to bacterial infection.

In one embodiment, a method includes providing a plant described herein (e.g., a plant includes at least one 5-hydroxymethylcytosine (5-hmC) in the genome; a transgenic plant that
5 includes a catalytic domain of a Tet protein, a polynucleotide encoding a catalytic domain of a Tet protein, or a combination thereof; or a plant that includes a genome with reduced methylation of cytosine residues compared to a control plant), and screening the plant for a change in methylation of the genome compared to a control plant. The change in methylation can be, for instance, a change in the amount of methylation of the genome, or a change in the location of
10 methylated nucleotides of the genome. In one embodiment, the method includes screening the plant for a redistribution of heterochromatin.

As used herein, the term “transgenic plant” refers to a plant that has been transformed to contain at least one coding region. For example, a coding region not normally present in a plant may be introduced and optionally expressed. A transformed plant described herein can include,
15 but is not limited to (i) direct transfectant, meaning that the DNA construct was introduced directly into the plant, such as through *Agrobacterium*, (ii) a progeny of a transfected plant, or (iii) a plant that is the result of regeneration from callus. The second or subsequent generation plant can be produced by sexual reproduction, i.e., fertilization. A transgenic plant may, and in certain embodiment does, have a phenotype that is different from a plant that has not been
20 transformed.

As used herein, the term “protein” refers broadly to a polymer of two or more amino acids joined together by peptide bonds. The term “protein” also includes molecules which contain more than one protein joined by a disulfide bond, or complexes of proteins that are joined together, covalently or noncovalently, as multimers (e.g., dimers, tetramers). Thus, the
25 terms peptide, oligopeptide, and polypeptide are all included within the definition of protein and these terms are used interchangeably.

As used herein, “heterologous amino acids” refers to amino acids that are not normally or naturally found flanking the sequence depicted at, for instance, SEQ ID NO:1, 2, or 3, or a catalytically active domain of SEQ ID NO:1, 2, or 3.

30 As used herein, the term “polynucleotide” refers to a polymeric form of nucleotides of any length, either ribonucleotides, deoxynucleotides, peptide nucleic acids, or a combination

thereof, and includes both single-stranded molecules and double-stranded duplexes. A polynucleotide can be obtained directly from a natural source, or can be prepared with the aid of recombinant, enzymatic, or chemical techniques. A polynucleotide described herein may be isolated. An “isolated” polynucleotide is one that has been removed from its natural
5 environment. Polynucleotides that are produced by recombinant, enzymatic, or chemical techniques are considered to be isolated and purified by definition, since they were never present in a natural environment.

A “regulatory sequence” is a nucleotide sequence that regulates expression of a coding sequence to which it is operably linked. Nonlimiting examples of regulatory sequences include
10 promoters, enhancers, transcription initiation sites, translation start sites, translation stop sites, transcription terminators, and poly(A) signals. The term “operably linked” refers to a juxtaposition of components such that they are in a relationship permitting them to function in their intended manner. A regulatory sequence is “operably linked” to a coding region when it is joined in such a way that expression of the coding region is achieved under conditions
15 compatible with the regulatory sequence.

Conditions that are “suitable” for an event to occur, or “suitable” conditions are conditions that do not prevent such events from occurring. Thus, these conditions permit, enhance, facilitate, and/or are conducive to the event.

The term “and/or” means one or all of the listed elements or a combination of any two or
20 more of the listed elements.

The words “preferred” and “preferably” refer to embodiments of the invention that may afford certain benefits, under certain circumstances. However, other embodiments may also be preferred, under the same or other circumstances. Furthermore, the recitation of one or more preferred embodiments does not imply that other embodiments are not useful, and is not intended
25 to exclude other embodiments from the scope of the invention.

The terms “comprises” and variations thereof do not have a limiting meaning where these terms appear in the description and claims.

It is understood that wherever embodiments are described herein with the language “include,” “includes,” or “including,” and the like, otherwise analogous embodiments described
30 in terms of “consisting of” and/or “consisting essentially of” are also provided.

Unless otherwise specified, "a," "an," "the," and "at least one" are used interchangeably and mean one or more than one.

Also herein, the recitations of numerical ranges by endpoints include all numbers subsumed within that range (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.80, 4, 5, etc.).

5

BRIEF DESCRIPTION OF THE FIGURES

10 FIG. 1A-1I shows overexpression of TET1 induced global CG demethylation in *A. thaliana*. FIG. 1A is a bar graph of global methylation levels in two Col-0 WT replicates, two 35S:TET1 transgenic individuals and *met1*. FIG. 1B shows a metaplot of CG methylation levels (100 kb windows) across five *A. thaliana* chromosomes. Methylation level differences were defined relative to Col-0 WT-1, and Col-0 WT-2 was used to assess background interference.

15 Genome browser view of methylome profile of two regions (FIG. 1C and FIG. 1D) of the *A. thaliana* genome (purple vertical lines = CG methylation, blue vertical lines = CHG methylation and gold vertical lines = CHH methylation). Metagene plots of CG methylation level across (FIG. 1E) gene bodies and (FIG. 1F) transposable elements. FIG. 1G shows a heat map of CG methylation level of CG DMRs. Bar plots of CHG (FIG. 1H) and CHH (FIG. 1I) methylation

20 levels of CG DMRs that possess non-CG methylation in wild-type individuals.

FIG. 2A-2G shows widespread redistribution of heterochromatin in 35S:TET1 plants. FIG. 2A is a genome browser view of *IBM1* (AT3G07610) in Col-0 WT, three 35S:TET1 transgenic plants and *met1*. A decrease in CG methylation from coding regions was accompanied by an increase in non-CG methylation. Both CG and non-CG methylation was lost from the large

25 intron (purple vertical lines = CG methylation, blue vertical lines = CHG methylation and gold vertical lines = CHH methylation). FIG. 2B shows the amount of the genome affected by differential CHG methylation. These DMRs were defined relative to Col-0 WT-1, as Col-0 WT-2 DMRs were used to assess background interference. FIG. 2C is a heat map of CHG methylation displaying CHG DMRs. Corresponding CG and CHH methylation levels are shown

30 in FIG. 2D and FIG. 2E. Heat maps showing log₂ transformed FPKM profiles of up-regulated genes (FIG. 2F) and down-regulated genes (FIG. 2G) in two 35S:TET1 transgenic individuals, *met1* and *ibm1* mutants relative to WT.

FIG. 3A-3D shows 35S:TET1 plants have a delayed flowering phenotype. FIG 3A shows photographs of one 35S:TET1-1 transgenic plant and Col-0 WT plant and (FIG 3B) corresponding number of rosette leaves. FIG 3C is a genome browser view of *FWA* (AT4G25530). Both CG and non-CG DNA methylation are depleted from the 5'UTR in 35S:TET1-2 plants (purple vertical lines = CG methylation, blue vertical lines = CHG methylation and gold vertical lines = CHH methylation). FIG 3D shows expression level (FPKM) of *FWA*.

FIG. 4A-4H shows global methylation (FIG 4A) and 5hmC (FIG 4B) levels of Col-0 WT plants and 35S:TET1 transgenic plants. Metagene plots of CHG (FIG 4C and FIG 4D) and CHH (FIG 4E and FIG 4F) methylation levels across gene bodies and transposable elements. Frequency plots of mCHG (FIG 4G) and mCHH (FIG 4H) differences in CG DMRs that possess non-CG methylation in wild type individuals. The average value of two Col-0 WT plants was defined as the Col-0 WT.

FIG. 5-01 through FIG. 5-03 shows the amino acid sequences of SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3.

FIG. 6-01 through FIG. 6-04 is a multiple sequence alignment of a Tet1, a Tet2, and a Tet3 protein based on the Clustal Omega algorithm.

FIG. 7A shows the vector map of the plasmid hTet1-CDinpMDC32 used for plant transformation. FIG. 7B-01 through FIG. 7B-03 shows the sequence of the plasmid hTet1-CDinpMDC32. FIG. 7B-04 shows a nucleotide sequence (SEQ ID NO:4) that encodes amino acids 1418 – 2136 of SEQ ID NO:1.

FIG 8 shows nucleotide sequences of promoters. SEQ ID NO:5 (AT3G18780); SEQ ID NO:6 (pAT4G40020); SEQ ID NO:7 (AT3G22880_DMC1 Promoter); SEQ ID NO:8 (AT4G20900_ms5 promoter); SEQ ID NO:9(GmuCoreIN); SEQ ID NO:10 (Gmupri); SEQ ID NO:11 (GmuCore); SEQ ID NO:12 (ZmRbcs1 Promoter); SEQ ID NO:13 (OsAnt1 Promoter); SEQ ID NO:14 (Nos promoter).

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Provided herein are non-natural plants. Also provided herein are non-natural plant cells. A non-natural plant cell can be an individual cell or a cell culture. A non-natural plant cell can be part of a plant. A non-natural plant cell can be a portion of an embryo, endosperm, sperm or egg cell, or a fertilized egg. The term plant includes whole plant, plant parts (stems, roots, leaves, fruit, etc.) or organs, plant cells, seeds, and progeny of same. A plant can be a gametophyte (haploid stage) or a sporophyte (diploid stage).

As used herein, a “plant” refers to eukaryotic algae and to members of the kingdom Plantae. Examples of eukaryotic algae include, but are not limited to, members of the genus *Chlamydomonas*, such as *C. reinhardtii*, members of the genus *Botryococcus*, such as *B. braunii*, members of the genus *Chlorella*, members of the genus *Dunaliella*, such as *D. tertiolecta*, members of the genus *Gracilaria*, members of the genus *Pleurochrysis*, such as *P. carterae* and members of the genus *Sargassum*, to name a few examples. Eukaryotic algae include unicellular and multicellular forms. As used herein, reference to “plant” includes both multicellular plants and unicellular plants, e.g., a plant cell.

A plant that is a member of the kingdom Plantae can be a monocot or a dicot. Examples of plants that are a member of the kingdom Plantae include, but are not limited to, flowering plants, also referred to as angiosperms. Examples of flowering plants include crop plants. A crop plant is a cultivated plant that is harvested for food, clothing, livestock fodder, biofuel, medicine, or other uses. Examples of crop plants include, but are not limited to, soybean (e.g., *Glycine max*), corn (*Zea mays*), tomato, rice (e.g., *Oryza sativa*), common bean (e.g., *Phaseolus vulgaris*), sugarcane, pumpkin, wheat (e.g., *Triticum* spp.), cassava, potato, cotton, or peanut. An additional example of a crop plant includes those found in hay, such as forage grasses (e.g., ryegrass and tall fescue), forage legumes (e.g., clover and alfalfa), or other herbaceous plants. Another example of a flowering plant is a tree. A tree can be a crop plant. Examples of trees include, but are not limited to, members of the *Populus* genus. Other plants that are a member of the kingdom Plantae include, but are not limited to, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Setaria viridis*, *Setaria italica*. In one embodiment, a non-natural plant is not an *A. thaliana*. In one embodiment, the plant is an epigenetic recombinant inbred line (epiRIL). epiRILs are plants that have been derived by using a wild-type plant as one parent and another

parent described herein having reduced DNA methylation. The resulting F1 progeny from this cross can be used to create large F2 populations for which plants can be selected and self fertilized for numerous generations to create a recombinant inbred line. Optionally, a recombinant inbred line created in this way does not include a Tet protein, or a polynucleotide encoding a Tet protein, e.g., the plants of the recombinant inbred line are not transgenic.

In one embodiment, a plant is a transgenic plant. In one embodiment, a plant cell is a transgenic plant cell. A transgenic plant or transgenic plant cell includes a Ten-eleven translocation methylcytosine dioxygenase (Tet) protein, a coding region encoding a Tet protein, or a combination thereof. A transgenic plant or a transgenic plant cell can be homozygous or hemizygous for a coding region encoding a Tet protein. Such a plant or plant cell can have a genome with reduced methylation of cytosine nucleotides as described herein.

A Tet protein useful herein catalyzes the conversion of the modified DNA base 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) (Tahiliani et al., 2009, Science 324 (5929): 930–5). A Tet protein that catalyzes the conversion of the modified DNA base 5-mC to 5-hmC has methylcytosine dioxygenase activity. Methods for measuring the ability of a Tet protein to convert the modified DNA base 5-mC to 5-hmC are routine and well known. Examples of methods include, but are not limited to, mass spectrometry, dot blot using anti-5hmC antibodies, and Tet-assisted bisulfite sequencing (TAB-seq). In vertebrate cells a Tet protein can also catalyze the conversion of 5-hmC to 5-formylcytosine, and the conversion of 5-formylcytosine to 5-carboxylcytosine.

In one embodiment, a Tet protein modifies a 5-methylcytosine that is present in the dinucleotide CG. The dinucleotide CG is also referred to herein as CpG, where the “p” refers to the phosphate present between the C and the G in a single strand of DNA. In one embodiment, a Tet protein does not modify a 5-methylcytosine that is present in a trinucleotide CHG or CHH, where H is A, C, or T. Tet proteins, and the coding regions that encode them, are not present in a naturally occurring plant. Likewise, 5-hydroxymethylcytosine is not present in a naturally occurring plant. Thus, also provided herein is a non-natural plant that includes at least one 5-hydroxymethylcytosine (5-hmC) present in the genome of the plant. In one embodiment, the 5-hmC is immediately adjacent to and 5' of a guanine, e.g., the 5-methylcytosine is present in the dinucleotide CG.

A Tet protein can be a member of the group having EC number 1.14.11.n2. An example of a Tet protein is Tet1. A specific example of a Tet1 protein is depicted at SEQ ID NO:1 (Genbank accession number NP_085128) (FIG. 5). Another example of a Tet protein is Tet2. A specific example of a Tet2 protein is depicted at SEQ ID NO:2 (Genbank accession number NP_001120680). Another example of a Tet protein is Tet3. A specific example of a Tet3 protein is depicted at SEQ ID NO:3 (Genbank accession number NP_001274420).

In one embodiment, the Tet protein present in a plant is a fragment of a Tet protein that has methylcytosine dioxygenase activity. The fragment includes the catalytic domain. It is known that all Tet enzymes contain a C-terminal catalytic domain that belongs to the dioxygenase superfamily and oxidizes 5mC in a 2-oxoglutarate-(2-OG) and Fe(II)-dependent manner (Hu et. al., 2013, Cell, 155:1545-1555). Hu et al. shows that truncation of amino terminal residues of Tet2 before residue 1129 or of carboxy terminal residues after residue 1936 had a minor effect on Tet2 activity. Further truncation of Tet2 (residues before 1156 or after 1913) significantly decreased its activity. According to Hu et al., these results indicate that Tet2 (1129–1936) is the minimum, catalytically active fragment. Residues 1129–1936 of the Tet2 analyzed by Hu et al. correspond to residues 1129–1936 of the Tet2 depicted at SEQ ID NO:2. Furthermore, residues 1129–1936 of SEQ ID NO:2 correspond to residues 1418-2086 of SEQ ID NO:1 and residues 824-1731 of SEQ ID NO:3. Thus, in one embodiment, a Tet protein useful herein includes the catalytic domain of residues 1418-2086 of SEQ ID NO:1, residues 1129–1936 of SEQ ID NO:2, or residues 824-1731 of SEQ ID NO:3, and optionally additional amino acids, such as heterologous amino acids. An example of a Tet catalytic domain with additional amino acids is amino acids 1418-2136 of SEQ ID NO:1, which was used in Example 1.

In one embodiment, a Tet protein having methylcytosine dioxygenase activity is, or is structurally similar to, a reference protein that includes the amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3. In one embodiment, a Tet protein having methylcytosine dioxygenase activity is, or is structurally similar to, a reference protein that includes the amino acid sequence of a catalytic domain of a Tet1, Tet2, or Tet3 protein (e.g., amino acids 1418-2086 of SEQ ID NO:1, amino acids 1129–1936 of SEQ ID NO:2, or amino acids 824-1731 of SEQ ID NO:3).

As used herein, a protein is “structurally similar” to a reference protein if the amino acid sequence of the protein possesses a specified amount of similarity and/or identity compared to

the reference protein. Structural similarity of two proteins can be determined by aligning the residues of the two proteins (for example, a candidate protein and the protein of, for example, any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, or a portion of one of those proteins) to optimize the number of identical amino acids along the lengths of their sequences; gaps in
5 either or both sequences are permitted in making the alignment in order to optimize the number of identical amino acids, although the amino acids in each sequence must nonetheless remain in their proper order.

A pair-wise comparison analysis of amino acid sequences can be carried out using the BESTFIT algorithm in the GCG package (version 10.2, Madison, WI). Alternatively, proteins
10 may be compared using the Blastp program of the BLAST 2 search algorithm, as described by Tatiana et al., (1999 *FEMS Microbiol Lett*, 174:247-250), and available on the National Center for Biotechnology Information (NCBI) website. The default values for all BLAST 2 search parameters may be used, including matrix = BLOSUM62; open gap penalty = 11, extension gap penalty = 1, gap x_dropoff = 50, expect = 10, wordsize = 3, and filter on.

15 In the comparison of two amino acid sequences, structural similarity may be referred to by percent "identity" or may be referred to by percent "similarity." "Identity" refers to the presence of identical amino acids. "Similarity" refers to the presence of not only identical amino acids but also the presence of conservative substitutions. A conservative substitution for an amino acid in a protein described herein may be selected from other members of the class to
20 which the amino acid belongs. For example, it is well-known in the art of protein biochemistry that an amino acid belonging to a grouping of amino acids having a particular size or characteristic (such as charge, hydrophobicity, and hydrophilicity) can be substituted for another amino acid without altering the activity of a protein, particularly in regions of the protein that are not directly associated with biological activity. For example, nonpolar (hydrophobic) amino
25 acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and tyrosine. Polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Conservative substitutions include, for example, Lys for Arg and vice versa to maintain a
30 positive charge; Glu for Asp and vice versa to maintain a negative charge; Ser for Thr so that a free -OH is maintained; and Gln for Asn to maintain a free -NH₂. Likewise, biologically active

analogues of a protein containing deletions or additions of one or more contiguous or noncontiguous amino acids that do not eliminate a functional activity of the protein are also contemplated.

Guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie et al. (1990, *Science*, 247:1306-1310), wherein the authors indicate proteins are surprisingly tolerant of amino acid substitutions. For example, Bowie et al. disclose that there are two main approaches for studying the tolerance of a polypeptide sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selects or screens to identify sequences that maintain functionality. As stated by the authors, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require non-polar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie et al, and the references cited therein.

Guidance on how to modify the amino acid sequences of proteins disclosed herein and maintain catalytic activity is also provided at FIG. 6. FIG. 6 depicts a Clustal Omega amino acid alignment of SEQ ID NOs:1, 2, and 3 (Sievers et al., 2011, *Molecular Systems Biology* 7: 539, doi:10.1038/msb.2011.75; Goujon et al., 2010, *Nucleic acids research* 38 (Suppl 2):W695-9, doi:10.1093/nar/gkq313). Identical amino acids are marked with an asterisk ("*"), strongly conserved amino acids are marked with a colon (":"), and weakly conserved amino acids are marked with a period ("."). Guidance on how to modify the amino acid sequences of proteins disclosed herein is also available to the skilled person. Hu et al. (Hu et. al., 2013, *Cell*, 155:1545-1555) discloses a structure based protein sequence alignment of human and mouse Tet1, Tet2, and Tet3 proteins, and other related proteins (see Figure S2 of Hu et al.). The figure shows identical and highly conserved residues highlighted in black and conserved residues in gray. Secondary structural elements are indicated above the sequences. Residues that are involved in zinc coordination are connected by solid lines. Residues involved in Fe (II) coordination and N-oxalylglycine (NOG, a 2-OG analogue) interaction are indicated by triangles and circles,

respectively. Residues involved in DNA binding and CpG recognition are indicated by squares and hexagons, respectively. Loops L1 and L2 that are involved in DNA interaction are indicated. Wu et al. (Wu et al., 2011, Genes & Development, 25:2436-2452) teaches that mutation of the residues involved in Fe (II) coordination abolishes enzymatic activity.

5 By reference to these figures, the skilled person can predict which alterations to an amino acid sequence having structural similarity with a protein disclosed herein are likely to modify enzymatic activity, as well as which alterations are unlikely to modify enzymatic activity.

Also available to the skilled person is the crystal structure of a Tet protein, Tet2
10 (Hu et. al., 2013, Cell, 155:1545-1555). The crystal structure shows the interaction between the protein and a DNA substrate, and critical residues for TET2-DNA interaction (see Figure 4 of Hu et al.). The skilled person will recognize that the structure of TET2 can be used to help predict which amino acids may be substituted, and which sorts of substitutions (e.g., conservative or non-conservative) can be made to a Tet protein
15 disclosed herein without altering the methylcytosine dioxygenase activity of the Tet protein.

Thus, as used herein, reference to a protein as described herein and/or reference to the amino acid sequence of one or more SEQ ID NOs can include a protein with at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at
20 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%, or at least 99% sequence similarity to a reference amino acid sequence.

Alternatively, as used herein, reference to a protein as described herein and/or reference
25 to the amino acid sequence of one or more SEQ ID NOs can include a protein with 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%, or at least 99% sequence identity to a reference amino acid
30 sequence.

A Tet protein may include heterologous amino acids. Heterologous amino acids can be at the N-terminal end, the C-terminal end, or the combination thereof. The number of heterologous amino acids can be, for instance, at least 10, at least 100, or at least 1000. Any series of amino acids can be present provided they do not completely inhibit the methylcytosine dioxygenase activity of the Tet protein. Such a protein may be referred to as a fusion protein. Examples of amino acid sequences include a detectable marker, e.g., a protein that is easily detected by various methods. Examples include fluorescent polypeptides such as green, yellow, blue, or red fluorescent proteins, and a tag, such as, but not limited to, a polyhistidine-tag (His-tag). In another embodiment, the heterologous amino acids can have a specific DNA binding activity. Examples of proteins having DNA binding activity that can be fused to a Tet protein include, but are not limited to, Cas9d of the CRISPR system and TALE of the TAL system (Maeder, 2013, Nat. Biotechnol., 31:1137-1142). The use of a fusion between a Tet protein and a domain that binds a specific DNA sequence can result in demethylation of a specific target DNA.

Another non-natural plant provided herein has a genome with reduced methylation of cytosine nucleotides. In one embodiment, the methylated cytosine nucleotides are present in the dinucleotide CG. In one embodiment, the methylated cytosine nucleotides are not present in the trinucleotide CHG or CHH, where H is A, C, or T.

In one embodiment, such a plant includes a catalytic domain of a Tet protein, such as a full length Tet protein. In one embodiment, such a plant includes a polynucleotide encoding catalytic domain of a Tet protein, such as a full length Tet protein. In one embodiment, such a plant does not include a catalytic domain of a Tet protein or a full length Tet protein, or a polynucleotide encoding any part of a Tet protein, e.g., the plant is not transgenic. Such a plant can occur by breeding out the transgenic polynucleotide encoding a Tet protein while maintaining a reduced methylation status. Such a non-transgenic plant produced as described herein is not a natural plant. The level of methylation in the non-transgenic plant can be different than the level of methylation in the transgenic parent from which the non-transgenic plant was derived; however, the non-transgenic plant has a reduced methylation level compared to the control plant.

In the Example, CG methylation in wild type plants was 18.2%, and was reduced to 8.9% and 6.9% in two transgenic plants expressing Tet, a reduction of greater than 50%. The

reduction of methylation of a genome of a plant or plant cell described herein can be from at least 1% to 100%. In one embodiment, the reduction can be at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90%. In one embodiment, the reduction can be no greater than 90%, greater than 80%, greater than 70%,
5 greater than 60%, greater than 50%, greater than 40%, greater than 30%, greater than 20%, or greater than 10%. In one embodiment, the reduction of methylation of a genome of a plant or plant cell is a range between at least 1% and no greater than 100%, or any combination of concentrations selected from the numbers listed above (e.g., from at least 10% to no greater than 90%, from at least 20% to no greater than 60%, etc.). The reduction of methylation can occur at
10 the cytosine residue of CG, CHG, CHH, or a combination thereof. While the inventor has identified the reduction of methylation in plants expressing a Tet protein, increases in methylation of CHG and CHH in some regions has also been observed in plants having reduced overall methylation. Thus, in some embodiments a plant having reduced methylation can include regions having increased methylation; however, the overall methylation level in the plant
15 is still reduced. Without intending to be limiting, it is expected that extreme reduction of methylation, for instance 100%, may be lethal in some plants having a high number of transposons due to reactivation of transposons that are typically silenced by methylation. In one embodiment, the reduction is a statistically significant difference. The reduction is relative to the methylation of the same species. In one embodiment, the reduction is relative to the methylation
20 of the plant or the plant cell before introduction of the polynucleotide encoding the Tet protein, e.g., a control plant. As used herein, the term “control plant” refers to a plant that is the same species as a transgenic plant, but has not been transformed with the same polynucleotide used to make the transgenic plant. Methods for determining the level of methylation is routine and known in the art. An example of such a method is Tet-assisted whole genome bisulfite
25 sequencing.

A methylome of a plant refers to the set of methylated cytosines in the genome of the plant. A non-natural plant described herein has a methylome that is different when compared to a plant of the same species that does not include the Tet protein or has not had the Tet protein expressed, e.g., a naturally occurring plant. A non-natural plant described herein can be
30 homozygous or heterozygous for the altered methylome. For instance, in a heterozygous plant cell one allele has reduced methylation compared to a naturally occurring plant cell of the same

species and a second allele of the coding region does not have reduced methylation. A non-natural plant can be a mosaic, e.g., some of the cells in the plant have altered methylation compared to a naturally occurring plant of the same species and other cells in the plant do not have altered methylation.

5 Also provided is the biomass, such as lignocellulosic material including wood and wood pulp, derived from the plants described herein. Cells of the biomass have reduced CG methylation, and optionally, a Tet protein described herein.

 Also provided herein are polynucleotides that encode a Tet protein or a portion thereof (e.g., a catalytic domain). Examples of polynucleotides encoding SEQ ID NO:1, SEQ ID NO:2, 10 or SEQ ID NO:3 or a portion thereof are known and readily available. It should be understood that a polynucleotide encoding one of the Tet proteins disclosed at SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 is not limited to a nucleotide sequence known to encode one of those proteins, but also includes the class of polynucleotides encoding the Tet proteins as a result of the degeneracy of the genetic code. For example, the naturally occurring nucleotide sequence 15 encoding SEQ ID NO:1 is but one member of the class of nucleotide sequences encoding a Tet protein having the amino acid sequence SEQ ID NO:1. The class of nucleotide sequences encoding a selected protein sequence is large but finite, and the nucleotide sequence of each member of the class may be readily determined by one skilled in the art by reference to the standard genetic code, wherein different nucleotide triplets (codons) are known to encode the 20 same amino acid. In one embodiment, a polynucleotide used herein is naturally occurring in a vertebrate cell and expressed in a plant cell. In another embodiment, the nucleotide sequence of a polynucleotide used herein is altered to optimize the codon usage bias of the plant in which the polynucleotide is to be expressed. Such codon optimized sequences can have structural similarity with a naturally occurring polynucleotide. An example of a polynucleotide encoding a 25 Tet catalytic domain is depicted at SEQ ID NO:4, which encodes amino acids amino acids 1418-2136 of SEQ ID NO:1.

 Structural similarity of two polynucleotides can be determined by aligning the residues of the two polynucleotides (for example, a candidate polynucleotide and any appropriate reference polynucleotide, for instance a polynucleotide sequence encoding SEQ ID NO:1) to optimize the 30 number of identical amino acids along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of identical

nucleotides, although the nucleotides in each sequence must nonetheless remain in their proper order. A reference polynucleotide may be a naturally occurring polynucleotide. A candidate polynucleotide is the polynucleotide being compared to the reference polynucleotide. A candidate polynucleotide may be isolated, for example, from an animal cell, or can be produced using recombinant techniques, or chemically or enzymatically synthesized. A candidate polynucleotide may be present in a vertebrate cell and predicted to encode a Tet protein.

Unless modified as otherwise described herein, a pair-wise comparison analysis of nucleotide sequences can be carried out using the Blastn program of the BLAST search algorithm, available through the World Wide Web, for instance at the internet site maintained by the National Center for Biotechnology Information, National Institutes of Health. Preferably, the default values for all Blastn search parameters are used. Alternatively, sequence similarity may be determined, for example, using sequence techniques such as GCG FastA (Genetics Computer Group, Madison, Wisconsin), MacVector 4.5 (Kodak/IBI software package) or other suitable sequencing programs or methods known in the art.

Thus, as used herein, a candidate polynucleotide useful in the methods described herein includes those with at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, 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%, or at least 99% polynucleotide sequence identity to a reference polynucleotide sequence.

A polynucleotide may be present in a vector. A vector is a replicating polynucleotide, such as a plasmid, phage, or cosmid, to which another polynucleotide may be attached so as to bring about the replication of the attached polynucleotide. Construction of vectors containing a polynucleotide of the invention employs standard ligation techniques known in the art. See, e.g., Sambrook et al, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989). A vector can provide for further cloning (amplification of the polynucleotide), i.e., a cloning vector, or for expression of the polynucleotide, i.e., an expression vector. The term vector includes, but is not limited to, plasmid vectors, viral vectors, cosmid vectors, transposon vectors, and artificial chromosome vectors. A vector may result in integration into a cell's genomic DNA. A vector may be capable of replication in a bacterial host, for instance *E. coli*. Preferably the vector is a plasmid.

Selection of a vector depends upon a variety of desired characteristics in the resulting construct, such as a selection marker, vector replication rate, and the like. Suitable host cells for cloning or expressing the vectors herein are prokaryotic or eukaryotic cells. Suitable eukaryotic cells include plant cells, such as *Agrobacterium tumefaciens*. Suitable prokaryotic cells include eubacteria, such as gram-negative organisms, for example, *E. coli*.

A selection marker is useful in identifying and selecting transformed plant cells or plants. Examples of such markers include, but are not limited to, a neomycin phosphotransferase (nptII) gene (Potrykus et al., 1985, Mol. Gen. Genet., 199:183-188), which confers kanamycin resistance. Cells expressing the nptII gene can be selected using an appropriate antibiotic such as kanamycin or G418. Other commonly used selectable markers include a mutant EPSP synthase gene (Hinchee et al., 1988, Bio/Technology 6:915-922), which confers glyphosate resistance; and a mutant acetolactate synthase gene (ALS), which confers imidazolinone or sulphonylurea resistance (Conner and Santino, 1985, European Patent Application 154,204).

Polynucleotides described herein can be produced *in vitro* or *in vivo*. For instance, methods for *in vitro* synthesis include, but are not limited to, chemical synthesis with a conventional DNA/RNA synthesizer. Commercial suppliers of synthetic polynucleotides and reagents for *in vitro* synthesis are well known. Methods for *in vitro* synthesis also include, for instance, *in vitro* transcription using a circular or linear expression vector in a cell free system. Expression vectors can also be used to produce a polynucleotide of the present invention in a cell, and the polynucleotide may then be isolated from the cell.

Polynucleotides described herein, including nucleotide sequences which are a portion of a coding region encoding a Tet protein, may be operably linked to a regulatory sequence. An example of a regulatory region is a promoter. A promoter is a nucleic acid, such as DNA, that binds RNA polymerase and/or other transcription regulatory elements. A promoter facilitates or controls the transcription of DNA to generate an RNA molecule from a nucleic acid molecule that is operably linked to the promoter. Promoters useful in the invention include constitutive promoters, inducible promoters, and/or tissue preferred promoters for expression of a polynucleotide in a particular tissue or intracellular environment, examples of which are known to one of ordinary skill in the art.

A constitutive promoter refers to a promoter that is transcriptionally active during most, but not necessarily all, phases of growth and development and under most environmental

conditions, in at least one cell, tissue or organ. Examples of useful constitutive plant promoters include, but are not limited to, the cauliflower mosaic virus (CaMV) 35S promoter, (Odel et al., 1985, *Nature*, 313:810), the nopaline synthase promoter (An et al., 1988, *Plant Physiol.*, 88:547), and the octopine synthase promoter (Fromm et al., 1989, *Plant Cell* 1: 977).

5 An inducible promoter has induced or increased transcription initiation in response to a chemical, environmental, or physical stimulus. Examples of inducible promoters include, but are not limited to, auxin-inducible promoters (Baumann et al., 1999, *Plant Cell*, 11:323-334), cytokinin-inducible promoters (Guevara-Garcia, 1998, *Plant Mol. Biol.*, 38:743-753), and gibberellin-responsive promoters (Shi et al., 1998, *Plant Mol. Biol.*, 38:1053-1060).

10 Additionally, promoters responsive to heat, light, wounding, pathogen resistance, and chemicals such as methyl jasmonate or salicylic acid, can be used, as can tissue or cell-type specific promoters such as xylem-specific promoters (Lu et al., 2003, *Plant Growth Regulation* 41:279-286).

15 A tissue preferred promoter is one that is capable of preferentially initiating transcription in certain organs, tissues, or cells types such as the leaves, roots, seed tissue, columella, cortex, endodermis, etc. For example, a root-specific promoter is a promoter that is transcriptionally active predominantly in plant roots, substantially to the exclusion of any other parts of a plant, while still allowing for any leaky expression in these other plant parts. Promoters able to initiate transcription in certain cells only are referred to herein as cell-specific.

20 A seed-specific promoter is transcriptionally active predominantly in seed tissue, but not necessarily exclusively in seed tissue (in cases of leaky expression). The seed-specific promoter may be active during seed development and/or during germination. The seed specific promoter may be endosperm/aleurone/embryo specific. Examples of seed-specific promoters (endosperm/aleurone/embryo specific) are described in Russinova and Reuzeau (US Patent
25 Application 20120331584). Further examples of seed-specific promoters are given in Qing Qu and Takaiwa (2004, *Plant Biotechnol. J.*, 2:113-125).

A green tissue-specific promoter is a promoter that is transcriptionally active predominantly in green tissue, substantially to the exclusion of any other parts of a plant, while still allowing for any leaky expression in these other plant parts.

30 Another example of a tissue-specific promoter is a meristem-specific promoter, which is transcriptionally active predominantly in meristematic tissue, substantially to the exclusion of

any other parts of a plant, while still allowing for any leaky expression in these other plant parts. A further example of a tissue-specific promoter is the RuBisCo promoter, which is transcriptionally active predominantly in the leaf or cotyledon.

In one embodiment, the promoter used can influence the amount of demethylation. For instance, the use of a relatively strong promoter to drive Tet protein expression results in a greater reduction of methylated cytosine nucleotides than a weaker promoter. Relatively stronger and weaker promoters are known to the person skilled in the art.

Examples of promoters useful herein are depicted at FIG. 8 and by De La Torre and Finer (2015, Plant Cell Rep., 34:111-120). The promoter depicted at SEQ ID NO:5 is useful in dicots, and is typically broadly expressed. The promoter depicted at SEQ ID NO:6 is useful in dicots and is typically meiosis I specific. SEQ ID NO:7 is useful in dicots and is typically broadly expressed in meiosis and in vegetative cells. The promoter depicted at SEQ ID NO:8 is useful in dicots and is typically anther specific. The promoters depicted at SEQ ID NO:9 (GmuCoreIN), SEQ ID NO:10 (Gmupri), and SEQ ID NO:11 (GmuCore) are useful in dicot plants. The promoters depicted at SEQ ID NO:12 (ZmRbcs1 promoter), SEQ ID NO:13 (OsAnt1 promoter), and SEQ ID NO:14 (Nos promoter) are useful in monocot plants. The skilled person will recognize that each promoter can be inserted upstream of a coding region to drive expression of the coding region in a plant.

Another example of a regulatory region is a transcription terminator. Suitable transcription terminators are known in the art and include, for instance, a stretch of 5 consecutive thymidine nucleotides.

Provided herein are methods for making a plant described herein. In one embodiment, a method includes making a transgenic plant. Transgenic plants described herein may be produced using routine methods. Expression of a Tet protein or a portion thereof results in a reduced number of methylated cytosine nucleotides, including CG, CHG, and CHH. The method of making a plant described herein can be referred to as an epimutagenesis method that introduces random methylation variations. Methods for transformation and regeneration are known to the skilled person. Transformation of a plant cell with a polynucleotide described herein may be achieved by any known method for the insertion of nucleic acid sequences into a prokaryotic or eukaryotic host cell, including *Agrobacterium*-mediated transformation protocols, viral infection,

whiskers, electroporation, microinjection, polyethylene glycol-treatment, heat shock, lipofection, particle bombardment, and chloroplast transformation.

Transformation techniques for dicotyledons are known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-

5 *Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This may be accomplished by PEG or electroporation mediated-uptake, particle bombardment-mediated delivery, or microinjection. In each case the transformed cells may be regenerated to whole plants using standard techniques known in the art.

10 Techniques for the transformation of monocotyledon species include direct gene transfer into protoplasts using PEG or electroporation techniques, particle bombardment into callus tissue or organized structures, as well as *Agrobacterium*-mediated transformation.

In one embodiment, a plant or plant cell used for the production of a transgenic plant includes a mutation that reduces or eliminates mobile 24 nucleotide small RNAs. 24 nucleotide small RNAs are typically associated with actively silenced regions of plant genomes in both
15 monocots and dicots (Melnik et al., 2011, Current Biol., 21: 1678-1683). In general there are plant specific RNA polymerases (poliv) that produce long-non-coding RNAs that are processed into 24 nucleotides by DICER enzymes. These small RNAs play a role in directing a *de novo* DNA methyltransferase to target sequences from which the small RNAs were derived. This creates a feedback loop that ensures certain sequences in the genome are consistently silenced,
20 typically by modifying cytosine residues to include methyl groups. Mutations that reduce or eliminate mobile 24 nucleotide small RNAs are known and include rdr2 or poliv.

The cells that have been transformed may be grown into plants in accordance with conventional techniques. See, for example, McCormick et al. (1986, Plant Cell Reports, 5:81-84). These plants may then be grown and evaluated for expression of the Tet transgene. In one
25 embodiment, the plants are evaluated for the level of cytosine methylation. In one embodiment, the plants are evaluated for desired phenotypic characteristics. These plants may be either pollinated with the same transformed strain or different strains, and the resulting hybrid having the Tet transgene, and optionally desired phenotypic characteristics, identified. In one embodiment, two or more generations may be grown to ensure that the desired Tet transgene is
30 stably maintained and inherited and then seeds harvested to ensure stability of the transgene has been achieved.

In one embodiment, a method includes making a non-natural plant with reduced methylation that is also non-transgenic. This method can include starting with a plant that includes a polynucleotide encoding the Tet protein and reduced methylation, and then removing the polynucleotide encoding the Tet protein from the genome of the plant. Methods for
5 removing the polynucleotide typically involve standard breeding practices to segregate the polynucleotide encoding the Tet protein out of a desired genetic background, e.g., a genome that includes the reduced methylation.

In one embodiment, the method includes making an epigenetic recombinant inbred line. Methods for making such a plant include routine breeding (Johannes et al., PLoS Genet 5,
10 e1000530 (2009), Reinders et al., Genes Dev, 23, 939-950 (2009), and Cortijo et al., Science, (2014); published online EpubFeb 6 (10.1126/science.1248127)). Also provided is the epigenetic recombinant inbred line.

Also provided herein are methods of using the non-natural plants described herein. In plants DNA methylation silences the expression of genes. The plants described herein include
15 the loss of DNA methylation, and the inventor has observed the increased expression of coding regions after loss of methylation. The inventor has also observed non-natural plants described herein as having phenotypic variation.

In contrast to routine methods of obtaining phenotypic variants in plants, e.g., traditional mutagenesis, in the plants described herein new phenotypes are the result of turning on or
20 waking up coding regions already present. Since the changes in phenotypic variation in the plants described herein are generally the result of changes in expression of coding regions and not the result of changes in DNA sequence, the changes in the plants described herein include epigenetic changes, and the generation of epigenome controlled phenotypic variation, such as morphological, biochemical, or physiological variation. A plant described herein that has
25 phenotypic variation compared to a plant of the same species that does not have the reduced methylation can be referred to as an epimutant, e.g., mutants that have variation in DNA methylation states and no or limited changes in the genome sequence.

In one embodiment, a method includes providing a plant described herein and screening the plant for a phenotypic variation. The phenotypic variation is relative to a plant of the same
30 species that does not have the reduced methylation. Any phenotypic variation can be screened.

Examples include useful traits include, but are not limited to, differences in flowering time (longer or shorter), root length, resistance to bacterial infection, drought resistance, and the like.

In one embodiment, a method includes identifying a coding region that is silent in a natural plant and is predicted to alter one or more phenotypes if expressed. A coding region can be silent due, for instance, to methylation of one or more cytosine nucleotides present in the coding region or methylation of the regulatory region of the coding region. Examples of such coding regions include, but are not limited to, coding regions predicted to encode proteins that have structural similarity with proteins known to have a specific function. One example of such a protein is a transcription factor. In another example, alterations of methylation sites of specific coding regions can be determined. A population of coding regions in flowering plants having suppressed expression are known, and these coding regions have cytosine methylation at the CG dinucleotide that overlap the transcriptional start site (Niederhuth et al., 2016, Widespread natural variation of DNA methylation within angiosperms, BioRxiv, doi:

<http://dx.doi.org/10.1101/045880>). Plants having reduced methylation of members of this population of coding regions can be identified and screened. A plant described herein that has reduced methylation of that specific coding region can be identified using standard methods (e.g., whole genome bisulfite sequencing). The plant can then be grown in normal or specific conditions (e.g., stress conditions) and the phenotype of the plant determined and compared to the natural plant.

After identification of a change in phenotype, the epigenetic basis or bases for that change can be identified, and optionally moved into other natural or non-natural plants using standard plant breeding techniques.

In the preceding description, particular embodiments may be described in isolation for clarity. Unless otherwise expressly specified that the features of a particular embodiment are incompatible with the features of another embodiment, certain embodiments can include a combination of compatible features described herein in connection with one or more embodiments.

For any method disclosed herein that includes discrete steps, the steps may be conducted in any feasible order. And, as appropriate, any combination of two or more steps may be conducted simultaneously.

The present invention is illustrated by the following examples. It is to be understood that the particular examples, materials, amounts, and procedures are to be interpreted broadly in accordance with the scope and spirit of the invention as set forth herein.

5

Example 1

TET-mediated epimutagenesis of the *Arabidopsis thaliana* methylome

DNA methylation in the promoters of plant genes often leads to transcriptional
10 repression, and the wholesale removal of DNA methylation in methyltransferase mutants results in severe gene expression and developmental defects. However, many cases of naturally-occurring DNA methylation variations have been reported, where the differential expression of differentially methylated genes are responsible for agronomically important traits. The ability to manipulate plant methylomes to generate populations of epigenetically distinct plants could
15 provide invaluable resources for breeding and research. Here we describe a novel “epimutagenesis” method to rapidly generate methylation variations through random demethylation of the *Arabidopsis thaliana* genome. This method involves the expression of a human Ten-eleven translocation (TET) enzyme and results in widespread hypomethylation and the redistribution of heterochromatin, mimicking mutants in the maintenance DNA
20 methyltransferase *met1*. Application of TET-mediated epimutagenesis to agriculturally significant plants may result in differential expression of alleles typically silenced by DNA methylation, uncovering previously hidden traits.

Our ability to develop novel beneficial crop traits has significantly improved over the last
100 years, although the ability to maintain this trajectory is limited by allelic diversity. While
25 genetic variation has been heavily exploited for crop improvement, utility of epigenetic variation has yet to be efficiently implemented. Epigenetic variation arises not from a change in the DNA sequence, but by changes in modifications to DNA such as DNA methylation that can result in stably inherited changes of both gene expression and phenotypes.

In plant genomes, cytosine methylation occurs at three major sequence contexts: CG,
30 CHG and CHH (where H = A, C or T)¹. Methylation at these different contexts is coordinated by distinct maintenance mechanisms during DNA replication. The methylation of DNA in all three

contexts is essential for transcriptional silencing of transposons, repeat sequences and certain genes. Genes regulated by this mechanism are stably repressed throughout the soma and represent an untapped source of hidden genetic variation if transcriptionally re-activated, as revealed from pioneering studies in the model plant *A. thaliana*²⁻⁴. However, the impact of this variation is not observed in wild-type plants, as genes silenced by DNA methylation are not expressed. This novel source of genetic variation was uncovered by creating epigenetic recombinant inbred lines (epiRILs) from crosses between a wild-type individual and a mutant defective in maintenance of DNA methylation²⁻⁴. EpiRILs, while genetically wild type, contain mosaic DNA methylomes dependent on chromosomal inheritance patterns, as DNA methylation is meiotically inherited in *A. thaliana*^{2,5-7}. Phenotypic characterization of epiRILs has revealed extensive morphological variation with respect to traits such as flowering time, root length and resistance to bacterial infection²⁻⁴. The morphological variation generated by the creation of epiRILs has revealed extensive hidden genetic variation in plant genomes that can be observed due to expression of newly unmethylated regions. However, the creation of epiRILs requires that one founding parent to be a null mutant in the maintenance DNA methylation pathway. Unfortunately, unlike in *A. thaliana*, the loss of DNA methylation maintenance activity often results in lethality in crops^{8,9}. Therefore, novel methodologies are required to realize the potential of these hidden epialleles in crop genomes.

Epimutagenesis is an alternative method to generate epiRILs. Instead of relying on the genome-wide demethylation of one of the two funding parents, epimutagenesis introduces random methylation variations. Here we describe a novel epimutagenesis approach in *A. thaliana* using a human Ten-eleven translocation (TET1) methylcytosine dioxygenase¹⁰⁻¹², which catalyzes the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). Although TET enzymes or their primary product, 5hmC, are not found in plant genomes¹³, ectopic expression of a human TET enzyme resulted in widespread DNA demethylation, redistribution of heterochromatin and induced phenotypic variations in *A. thaliana*.

RESULTS

Overexpressing Tet1 in Arabidopsis hypomethylates the genome

Transgenic *A. thaliana* plants were generated expressing the catalytic domain of the human TET1 protein (hTET1cd) under the control of the CaMV35S promoter. To assess the

impact of hTET1cd expression on the *A. thaliana* methylome, whole genome bisulfite sequencing (WGBS) was performed on two independently derived transgenic plants (35S:TET1-1 and 35S:TET1-2; Table 1). The results revealed a global reduction of CG methylation levels from 18.2% in two wild-type individuals to 8.9% in 35S:TET1-1 and 6.9% in 35S:TET1-2 (compared to 0.5% in *met1-3*). The effects of hTET1cd expression on *A. thaliana* CHG and CHH methylation were not as strong as it was for CG methylation (Fig. 1A). Importantly, different degrees of CG hypomethylation were observed in different independent transgenic plants. This result has important implications for epimutagenesis in economically and agriculturally significant plant species, as it appears feasible to control the degree of DNA hypomethylation by screening for plants with desired levels of demethylation activity. Taken together, these results showed that the expression of hTET1cd resulted in intermediate CG methylation levels when compared to wild-type and *met1* individuals.

Table 1. Methylome sequencing summary statistics

Sample	Uniquely mapped reads	Non-conversion (%)	Genome coverage	Coverage per-strand
<i>Col-0 WT-1</i>	25,858,836	0.30%	32.5	16.3
<i>Col-0 WT-2</i>	20,708,437	0.17%	26.0	13.0
<i>35S:TET1-1</i>	11,483,070	0.17%	14.4	7.2
<i>35S:TET1-2</i>	25,300,675	0.12%	31.8	15.9
<i>35S:TET1-2.5</i>	10,149,189	0.19%	12.8	6.4
<i>35S:TET1-3 rep1</i>	15,239,451	0.16%	19.2	9.6
<i>35S:TET1-3 rep2</i>	13,976,359	0.13%	17.6	8.8

The primary product of TET1 oxidation is 5hmC, which is indistinguishable from 5mC by WGBS. We therefore performed Tet-assisted bisulfite sequencing (TAB-seq) to profile 5hmC levels in 35S:TET1 plants¹⁴. No detectable levels of 5hmC were found in either of the transgenic lines assayed (Figs. 4A-4B). Thus, the widespread loss of CG DNA methylation observed may
 5 result from a failure to maintain methylation at CG sites that possess 5hmC, or through active removal of 5hmC or further oxidized products via the base excision repair pathway.

To better understand the effects of hTET1cd expression, we determined changes in the *A. thaliana* methylome at the chromosomal and local levels. Plotting methylation levels across all five chromosomes revealed a strong depletion of CG methylation at the pericentromeric region
 10 (Fig. 1B). CG hypomethylation occurred at both gene body methylated (gbM) and select RNA-directed DNA methylated (RdDM) loci. (Figs. 1C and 1D). To further quantify the observed hypomethylation, metaplots were created for genes and transposons, respectively (Figs. 1E and 1F, Figs. 4C-4F). Strong reduction of mCG and mild reduction of mCHG/mCHH were observed at both genes and transposons. On average, 97.9% of gbM genes and 56.7% of methylated
 15 transposons (where these regions have at least 50% mCG in wild type) have lost at least half of their CG methylation in epimutagenized lines. Collectively, these results indicate hypomethylation was more severe in genes than transposons, possibly the result of *de novo* methylation by the RdDM pathway, which is primarily active at transposons.

20 Tet1-mediated DNA demethylation mimics *met1* mutants

An analysis of Differentially Methylated Regions (DMRs) was then carried out to assess the genome-wide impact of hTET1cd expression. 56,283 CG DMRs ranging in size from 6 - 20,286 base pairs (bp) were identified (Fig. 1G). Of these, 38.7% were located in intergenic sequences, 53.7% overlapped with genes and 7.6% were located in promoter regions (1kb
 25 upstream of a gene). Like the *met1* mutant, the predominant effect of hTET1cd expression is CG hypomethylation (12,641 and 20,601 DMRs lost more than 50% mCG in 35S:TET1-1 and 35S:TET1-2, respectively; no region gained more than 50% mCG). However, the extent of CG methylation caused by hTETcd expression is lower than in *met1*: 31.8 Mb of the genome significantly lost CG methylation in *met1*, whereas 9.9 Mb and 18.0 Mb were lost in 35S:TET1-1
 30 and 35S:TET1-2, respectively.

Previous studies of the *met1* methylome have revealed a loss of mCHG/mCHH methylation in a subset of CG-hypomethylated regions. At these loci, DNA methylation is stably lost, in contrast to regions where DNA methylation is re-established by *de novo* methylation pathways. These loci are ideal targets of epimutagenesis as the co-existence of all three types of methylation is more frequently correlated with transcriptional repression of genes than CG methylation alone. This, coupled with the long-term stability of hypomethylation, may facilitate inheritable transcriptional changes. An analysis of the interdependence of the loss of CG methylation on non-CG methylation levels revealed that 39.7 Kb and 931.5 Kb of CHG methylated sequences lost significant amounts of methylation in two epimutagenized lines, compared to 4.0 Mb of sequence in *met1* mutants. A similar analysis for the loss of CHH methylation revealed losses of 23.3 Kb and 492.5 Kb in epimutagenized individuals, compared to 1.1 Mb lost in *met1* mutants. Of the 56,283 identified CG DMRs, 10,491 overlapped regions that contained at least 20% CHG methylation and 7,214 overlapped regions that contained at least 10% CHH methylation in wild-type individuals. To determine how many of these regions are susceptible to losing non-CG methylation if CG methylation is first depleted, we created a frequency distribution of mCHG and mCHH levels in wild-type and epimutagenized individuals (Figs. 1H and 1I). 2,341 and 3,447 regions lost more than 10% CHG methylation in 35S:TET1-1 and 35S:TET1-2, respectively, whereas 2,475 and 3,379 regions lost more than 5% CHH methylation in 35S:TET1-1 and 35S:TET1-2, respectively. Regions that are susceptible to losses of CG and non-CG methylation in lines expressing hTET1cd share a substantial overlap with regions that lose non-CG methylation in *met1* (Figs. 4G and 4H). 1,708 (73.0%) and 2,386 (69.2%) regions that have lost more than 10% mCHG in 35S:TET1-1 and 35S:TET1-2 have reduced levels in *met1*, whereas 2,013 (81.6%) and 2,563 (75.9%) regions that have lost more than 5% mCHH in 35S:TET1-1 and 35S:TET1-2 have reduced levels in *met1*. As crop genomes have a greater number of loci targeted for silencing by CG, CHG and CHH methylation when compared to *A. thaliana*, ectopic expression of hTET1cd is likely a viable approach for the creation epiRILs¹⁵.

Tet1-mediated redistribution of heterochromatin

Mutations in *met1* also leads to hypermethylation of CHG sites in gene bodies due to the loss of CG methylation in the 7th intron of the histone 3 lysine 9 (H3K9) demethylase,

*INCREASED IN BONSAI METHYLATION 1 (IBM1)*¹⁶⁻¹⁹. This results in alternative splicing of *IBM1*, ultimately producing a non-functional gene product and resulting in ectopic accumulation of di-methylation of H3K9 (H3K9me2) throughout the genome. As in *met1*, the 7th intron of *IBM1* was hypomethylated in 35S:TET1-1, 35S:TET1-2 and 35S:TET1-2.5, a line that was propagated for an additional two generations (Fig. 2A). A subtle increase in global CHG methylation was observed in 35S:TET1-1 and 35S:TET1-2 (Fig. 1A). However, extensive variation in genome-wide gains and losses of CHG methylation was observed in line 35S:TET1-2.5 as it displayed approximately 1.8 Mb of additional CHG methylation (Fig. 2B). This indicates that the expression of hTET1cd over a longer period of time not only leads to loss of methylation, but also results in redistribution of DNA methylation genome-wide.

To further characterize regions of differential CHG methylation, identified CHG DMRs were categorized into discrete groups based on their DNA methylation status in wild-type individuals. Of the 9,917 CHG DMRs identified, 1,460 were in loci that are defined as gbM in wild-type individuals, 584 were in unmethylated regions, and 6,940 of them were in RdDM-like regions (Figs. 2C-2E). Interestingly, in line 35S:TET1-2.5, 1,408 (96.4%) of the CHG DMRs in gbM-like loci gained CHG hypermethylation, whereas 2,156 (31.1%) of the CHG DMRs in RdDM-like regions lost CHG, in contrast to 595 (8.6%) RdDM-like regions that gained CHG methylation. Lastly, there were 502 (86.0%) loci that are unmethylated in wild-type individuals that gain CHG methylation as well as CG and CHH methylation in the epimutagenized lines (Figs. 2C-2E). These results reveal that methods for epimutagenesis can result in both losses and gains in DNA methylation genome-wide.

To characterize the effect of hTET1cd-induced methylome changes on gene expression, we performed RNA-sequencing (RNA-seq) on leaf tissue of wild-type, 35S:TET1-1 and 35S:TET1-2 plants and compared the results to *met1* and *ibm1* transcriptomes. Compared to wild-type plants, 629 and 736 up-regulated genes and 1,277 and 1,428 down-regulated genes were identified in 35S:TET1-1 and 35S:TET1-2, respectively. There was a high level of overlapping in transcriptome changes with 35S:TET1-1 and 35S:TET1-2 compared to *met1* and *ibm1* (Figs. 2F and 2G). Of the genes up-regulated in *met1*, 36.8% and 21.5% overlapped with 35S:TET1-1 and 35S:TET1-2, respectively. An even greater overlap was observed with down-regulated genes as 60.1% and 72.9% of down-regulated genes in 35S:TET1-1 and 35S:TET1-2 overlapped with down-regulated genes in *met1*, respectively. These results reveal that expressing

the catalytic domain of human *TET1* in *A. thaliana* is a viable approach to access hidden sources of allelic variation by inducing expression variation.

Tet1 expression leads to a delay in the floral transition

5 In the transgenic plants that were used for WGBS, we observed a delay in the developmental transition from vegetative growth to flowering (Figs. 3A and 3B). We hypothesized that the observed late-flowering phenotype was associated with the demethylation of the *FWA* (*FLOWERING WAGENINGEN*) locus, as is observed in *met1* mutants^{20,21}. A closer inspection of the DNA methylation status of this locus revealed that DNA methylation was
10 completely abolished, as was methylation at adjacent CHG and CHH sites (Fig. 3C). As in *met1*, the loss of methylation at the *FWA* locus was associated with an increase in *FWA* expression (Fig. 3D), which is known to cause a delay in flowering by restricting the movement of the florigen signal, *FT*, to the shoot apex²². These results demonstrate that expression of hTET1cd leads to phenotypic variation by abolishing methylation at some regions in all sequence contexts
15 (CG, CHG and CHH sites).

DISCUSSION

The discovery that expression of the catalytic domain of the human TET1 protein in *A. thaliana* leads to widespread loss of CG methylation and redistribution of heterochromatin
20 makes it possible to create epimutants without the need for methyltransferase mutants, which often causes lethality in crops. In addition to epimutagenesis, TET1cd could be used in combination with sequence-specific targeting machinery such as CRISPR-dCas9 to direct DNA demethylation in plant genomes, as has been demonstrated in mammalian systems²³⁻²⁷. The stable meiotic inheritance of DNA methylation states in flowering plant genomes provides a
25 stark contrast to the inheritance of DNA methylation in mammalian genomes, where genome-wide erasure of DNA methylation and reprogramming occurs each generation²⁸. This property of flowering plant genomes makes them ideal targets of induced-epialleles, as once a new methylation state occurs it is often inherited in subsequent generations. Application of epimutagenesis and the use of TET-mediated engineering of DNA methylation states in
30 economically and agriculturally significant plant species will be an interesting area of future investigation.

METHODS

Synthesis and cloning of the human TET1 catalytic domain. A human TET1 catalytic domain (hTET1-CD) sequence was synthesized by GenScript, and moved to a plant transformation compatible vector (pMDC32) using LR clonase from Life Technologies per the manufacturer's instructions (Catalog #11791100). The resulting plasmid was designated Tet1CDinpMDC32, and is shown in FIG. 7.

Plant transformation and screening. The hTET1-CD sequence in the pMDC32 vector was transformed into *Agrobacterium tumefaciens* strain C58C1 and plated on LB-agar supplemented with kanamycin (50 µg/mL), gentamicin (25 µg/mL), and rifampicin (15 µg/mL). A single kanamycin resistant colony was selected and used to start a 250-mL culture in LB Broth Miller liquid media supplemented with gentamicin (25 µg/mL), kanamycin (50 µg/mL), and rifampicin (15 µg/mL), which was incubated for two days at 30°C. Bacterial cells were pelleted by centrifugation at 4,000 RPM for 30 minutes and the supernatant decanted. The remaining bacterial pellet was re-suspended in 200 mL of 5% sucrose with 0.05% Silwet L77. Plant transformation was performed using the floral dip method described by Clough and Bent²⁹. Seeds were harvested upon senescence and transgenic plants were identified via selection on ½ LS plates supplemented with Hygromycin B (15 µg/mL). 35S:TET1-1 is a T1 individual, whereas 35S:TET1-2 is a T3 plant. All transgenic individuals chosen for analysis contain independent insertions of hTET1cd and are not the result of single-seed decent unless otherwise noted.

DNA and RNA isolation. *A. thaliana* leaf tissue was flash-frozen and finely ground to a powder using a mortar and pestle. DNA extraction was carried out on all samples using the DNeasy Plant Mini Kit (QIAGEN), and the DNA was sheered to approximately 200 bp by sonication. RNA was isolated from finely ground flash-frozen leaf tissue using Trizol (Thermo Scientific).

Library construction. Genomic DNA libraries were prepared following the MethylC-seq protocol without use of the bisulfite conversion step. MethylC-seq libraries were prepared as previously described in³⁰. RNA-seq libraries were constructed using Illumina TruSeq Stranded RNA LT Kit (Illumina, San Diego, CA) following the manufacturer's instructions with limited

modifications. The starting quantity of total RNA was adjusted to 1.3 µg, and all volumes were reduced to a third of the described quantity. TAB-seq libraries were prepared as previously described in¹⁴.

Sequencing. Illumina sequencing was performed at the University of Georgia Genomics Facility using an Illumina NextSeq 500 instrument. Methylomes, and 5-hydroxymethylomes were sequenced to 150 bp whereas transcriptomes were sequenced to 75 bp. For MethylC-seq and TAB-seq, raw reads were trimmed for adapters and preprocessed to remove low quality reads using cutadapt 1.9.dev1³¹. For RNA-seq, these processes were carried out by Trimmomatic v0.32³².

MethylC-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome as described in³³. Chloroplast DNA (which is fully unmethylated) was used as a control to calculate the sodium bisulfite reaction non-conversion rate of unmodified cytosines. All conversion rates were >99% (Table 1). The list of gbM genes used in this study was previously curated¹⁵. All methylation levels reported in all analyses are presented as differences in absolute values, including defining DMRs and calculating hyper/hypomethylated regions. The only exception is in the comparison of mCG loss between gbM, where we used a percentage difference..

RNA-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome using TopHat v2.0.13³⁴ (Table 2). Gene expression values were computed using Cufflinks v2.2.1³⁵. Genes determined to have at least 2-fold log₂ expression changes by Cufflinks were identified as differentially expressed genes.

Table 2. Transcriptome sequencing summary statistics

Sample	mapped reads	Percent Mapped
<i>Col-0 WT rep1</i>	18,065,032	95.39%
<i>Col-0 WT rep2</i>	24,156,251	96.63%
<i>Col-0 WT rep3</i>	20,987,335	96.48%
<i>35S:TET1-1</i>	19,919,035	95.38%
<i>35S:TET1-2</i>	15,977,645	96.51%
<i>met1</i>	45,615,533	94.61%
<i>ibm1 rep1</i>	20,810,937	94.16%
<i>ibm1 rep2</i>	22,011,997	96.17%
<i>ibm1 rep3</i>	25,241,389	97.05%

TAB-seq data processing. Qualified reads were aligned to the *A. thaliana* TAIR10 reference genome using Methylypy as described in³³. The control 5mC modified lambda DNA sequence was used to calculate the 5mC non-conversion rate upon TET and bisulfite treatment. Non-CG dinucleotide sites were used to compute the non-conversion rate of unmodified cytosines upon bisulfite treatment (Table 3).

5

Table 3. TAB-seq sequencing summary statistics

	Uniquely mapped reads	5mC non-conversion (%)	Non-conversion (%)	Genome coverage
<i>Col-0 WT</i>	15,044,614	3.97%	0.28%	18.9
<i>35S::TET1-3 rep1</i>	14,293,516	4.69%	0.28%	18.0
<i>35S::TET1-3 rep2</i>	18,145,140	4.68%	0.29%	20.3

Metaplot analysis. For metaplot analyses, twenty 50 bp bins were created for both upstream and downstream regions of gene bodies/TEs. Gene bodies/TE regions were evenly divided into 20 bins. Weighted methylation levels were computed for each bin as described previously³⁶.

5

DMR analysis. Identification of DMRs was performed as described in³⁷. Only DMRs with at least 5 DMSs (Differential Methylated Sites) and a 10% absolute methylation level difference within each DMR were reported and used for subsequent analysis. For coverage calculations, each sample was combined with two Col-0 WT replicates to identify DMRs. Each sample was compared with both Col-0 WT replicates separately and for a DMR to be identified it must have been identified in both comparisons. Absolute methylation differences of +/- (50% for CG, 10% for CHG and CHH) were defined as hyper/hypo methylation, respectively. DMRs overlapping regions with mCG $\geq$ 5%, mCHG and mCHH $\geq$ 1% in both two Col-0 WT replicates were defined as RdDM-like regions. DMRs overlapping regions with mCG $\geq$ 5%, mCHG and mCHH $<$ 1% in both two Col-0 WT replicates were defined as gbM regions. DMRs overlapping regions with all three contexts less methylated at less than 1% in both Col-0 WT replicates were defined as unmethylated regions. Overlap comparisons were performed using bedtools v2.26.0³⁸.

10

15

20 References

1. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11, 204-20 (2010).
2. Johannes, F. et al. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* 5, e1000530 (2009).
3. Reinders, J. et al. Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev* 23, 939-50 (2009).
4. Cortijo, S. et al. Mapping the Epigenetic Basis of Complex Traits. *Science* (2014).
5. 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).

30

6. Cortijo, S., Wardenaar, R., Colome-Tatche, M., Johannes, F. & Colot, V. Genome-Wide Analysis of DNA Methylation in Arabidopsis Using MeDIP-Chip. *Methods Mol Biol* 1112, 125-49 (2014).
7. Reinders, J. et al. Genome-wide, high-resolution DNA methylation profiling using
5 bisulfite-mediated cytosine conversion. *Genome Res* 18, 469-76 (2008).
8. Li, Q. et al. Genetic Perturbation of the Maize Methyloome. *Plant Cell* (2014).
9. Hu, L. et al. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc Natl Acad Sci U S A* 111, 10642-7 (2014).
10. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-5 (2009).
11. Pastor, W.A. et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394-7 (2011).
12. Ito, S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and
15 inner cell mass specification. *Nature* 466, 1129-33 (2010).
13. Erdmann, R.M., Souza, A.L., Clish, C.B. & Gehring, M. 5-hydroxymethylcytosine is not present in appreciable quantities in Arabidopsis DNA. *G3 (Bethesda)* 5, 1-8 (2014).
14. Yu, M. et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc* 7, 2159-70 (2012).
- 20 15. Niederhuth, C.E. et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* 17, 194 (2016).
16. Saze, H., Shiraishi, A., Miura, A. & Kakutani, T. Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science* 319, 462-5 (2008).
17. Miura, A. et al. An Arabidopsis jmjC domain protein protects transcribed genes from
25 DNA methylation at CHG sites. *EMBO J* 28, 1078-86 (2009).
18. Rigal, M., Kevei, Z., Pelissier, T. & Mathieu, O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* 31, 2981-93 (2012).
19. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in
30 Arabidopsis. *Cell* 133, 523-36 (2008).

20. Soppe, W.J. et al. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* 6, 791-802 (2000).
21. Finnegan, E.J., Peacock, W.J. & Dennis, E.S. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc Natl Acad Sci U S A* 93, 8449-54 (1996).
- 5 22. Ikeda, Y., Kobayashi, Y., Yamaguchi, A., Abe, M. & Araki, T. Molecular basis of late-flowering phenotype caused by dominant epi-alleles of the *FWA* locus in *Arabidopsis*. *Plant Cell Physiol* 48, 205-20 (2007).
23. Choudhury, S.R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at *BRCA1* promoter. *Oncotarget* 10 (2016).
24. Maeder, M.L. et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* 31, 1137-42 (2013).
25. Vojta, A. et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* (2016).
- 15 26. Mendenhall, E.M. et al. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* 31, 1133-6 (2013).
27. Liu, X.S. et al. Editing DNA Methylation in the Mammalian Genome. *Cell* 167, 233-247 e17 (2016).
28. Heard, E. & Martienssen, R.A. Transgenerational epigenetic inheritance: myths and 20 mechanisms. *Cell* 157, 95-109 (2014).
29. Clough, S.J. & Bent, A.F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-43 (1998).
30. Urich, M.A., Nery, J.R., Lister, R., Schmitz, R.J. & Ecker, J.R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* 10, 475-83 25 (2015).
31. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17, pp. 10-12 (2011).
32. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-20 (2014).
- 30 33. Schmitz, R.J. et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res* 23, 1663-74 (2013).

34. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36 (2013).
35. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511-5 (2010).
- 5 36. Schultz, M.D., Schmitz, R.J. & Ecker, J.R. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet* 28, 583-5 (2012).
37. Schultz, M.D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523, 212-6 (2015).
- 10 38. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-2 (2010).

The complete disclosure of all patents, patent applications, and publications, and electronically available material (including, for instance, nucleotide sequence submissions in, e.g., GenBank and RefSeq, and amino acid sequence submissions in, e.g., SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq) cited herein are incorporated by reference in their entirety. Supplementary materials referenced in publications (such as supplementary tables, supplementary figures, supplementary materials and methods, and/or supplementary experimental data) are likewise incorporated by reference in their entirety. In the event that any inconsistency exists between the disclosure of the present application and the disclosure(s) of any document incorporated herein by reference, the disclosure of the present application shall govern. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.

Unless otherwise indicated, all numbers expressing quantities of components, molecular weights, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless otherwise indicated to the contrary, the numerical parameters set forth in the specification and claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques.

Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. All numerical values, however, inherently contain a range necessarily resulting from the standard deviation found in their respective testing measurements.

All headings are for the convenience of the reader and should not be used to limit the meaning of the text that follows the heading, unless so specified.

What is claimed is:

1. A transgenic plant comprising a catalytic domain of a Tet protein.
2. A transgenic plant comprising a polynucleotide encoding a catalytic domain of a Tet protein.
3. A cell of the plant of claim 1 or 2 wherein the cell comprises the catalytic domain of the Tet protein, the polynucleotide encoding the catalytic domain of the Tet protein, or a combination thereof.
4. The plant of claim 1 or 2 wherein the plant is a monocot.
5. The plant of claim 1 or 2 wherein the plant is a dicot.
6. The progeny of the plant of claim 1 or 2.
7. The progeny of claim 6 wherein the progeny is a hybrid plant.
8. The plant of claim 1 or 2 wherein the plant is a crop plant.
9. The plant of claim 8 wherein the crop plant is soybean, corn, tomato, rice, common bean (*Phaseolus vulgaris*), sugarcane, pumpkin, wheat, cassava, potato, cotton, a forage grass, or a forage legume.
10. The plant of claim 1 or 2 wherein the plant is an algae.
11. The plant of claim 1 or 2 wherein the plant is a mosaic.
12. The plant of claim 1 or 2 wherein the plant is homozygous.

13. The plant of claim 1 or 2 wherein the plant is hemizygous.
14. A cell of the plant of claim 1 or 2.
15. A non-natural plant comprising a genome with reduced methylation of cytosine, wherein the reduction of cytosine methylation is reduced compared to a control plant.
16. The plant of claim 15 wherein the reduction is statistically significant.
17. The plant of claim 15 wherein the reduction is at least 10%.
18. The plant of claim 15 wherein the methylation of cytosine at CG dinucleotides is reduced compared to the control plant.
19. The plant of claim 15 wherein the methylation of cytosine at CHG and CHH, where H is A, C, or T, is reduced compared to the control plant.
20. The plant of any one of claims 15 to 19 wherein at least one cell of the plant comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase (Tet) protein, a polynucleotide encoding a catalytic domain of a Tet protein, or a combination thereof
21. The transgenic plant of any one of claims 15 to 19 wherein the plant is a monocot.
22. The transgenic plant of any one of claims 15 to 19 wherein the plant is a dicot.
23. The progeny of the plant of any one of claims 15 to 19.
24. The progeny of claim 23 wherein said progeny is a hybrid plant.
25. The plant of any one of claims 15 to 19 wherein the plant is a crop plant.

26. The plant of claim 25 wherein the crop plant is soybean, corn, tomato, rice, common bean (*Phaseolus vulgaris*), sugarcane, pumpkin, wheat, cassava, hay, potatoes or cotton.
27. The plant of any one of claims 15 to 19 wherein the plant is an algae.
28. The plant of any one of claims 15 to 19 wherein the plant is a mosaic.
29. The plant of any one of claims 15 to 19 wherein the plant is homozygous.
30. The plant of any one of claims 15 to 19 wherein the plant is hemizygous.
31. A cell of the plant of any one of claims 15 to 19.
32. A method comprising:
providing a plant of any one of claims 1 or 2 or 15 to 19;
screening the plant for a phenotypic variation compared to a control plant.
33. The method of claim 32 wherein the phenotypic variation comprises an alteration in flowering time, an alteration in root length, or an alteration in resistance to bacterial infection.
34. A method comprising:
providing a plant of any one of claims 1 or 2 or 15 to 19;
screening the plant for a change in methylation of the genome compared to a control plant.

FIG. 1

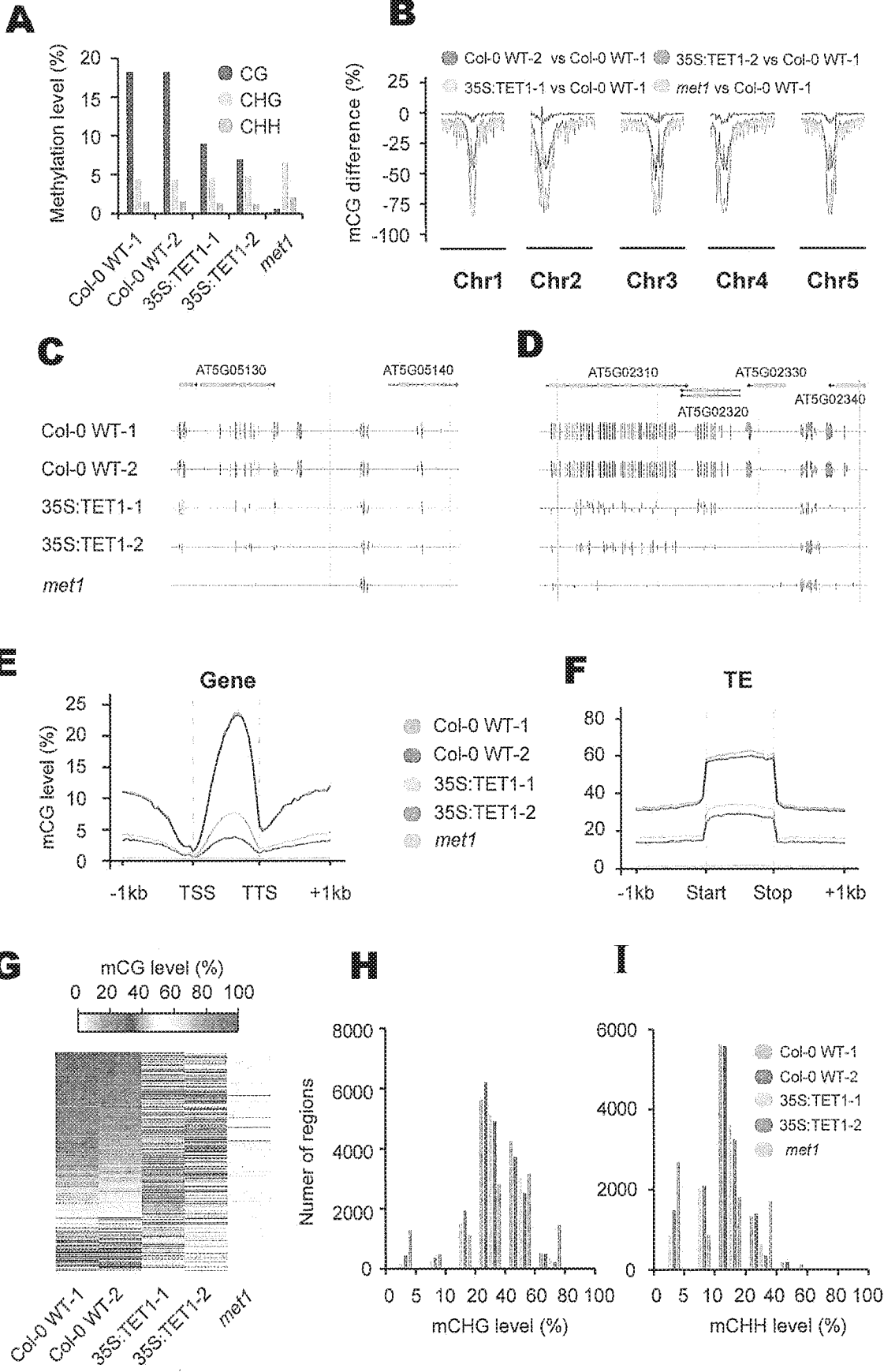


FIG. 2

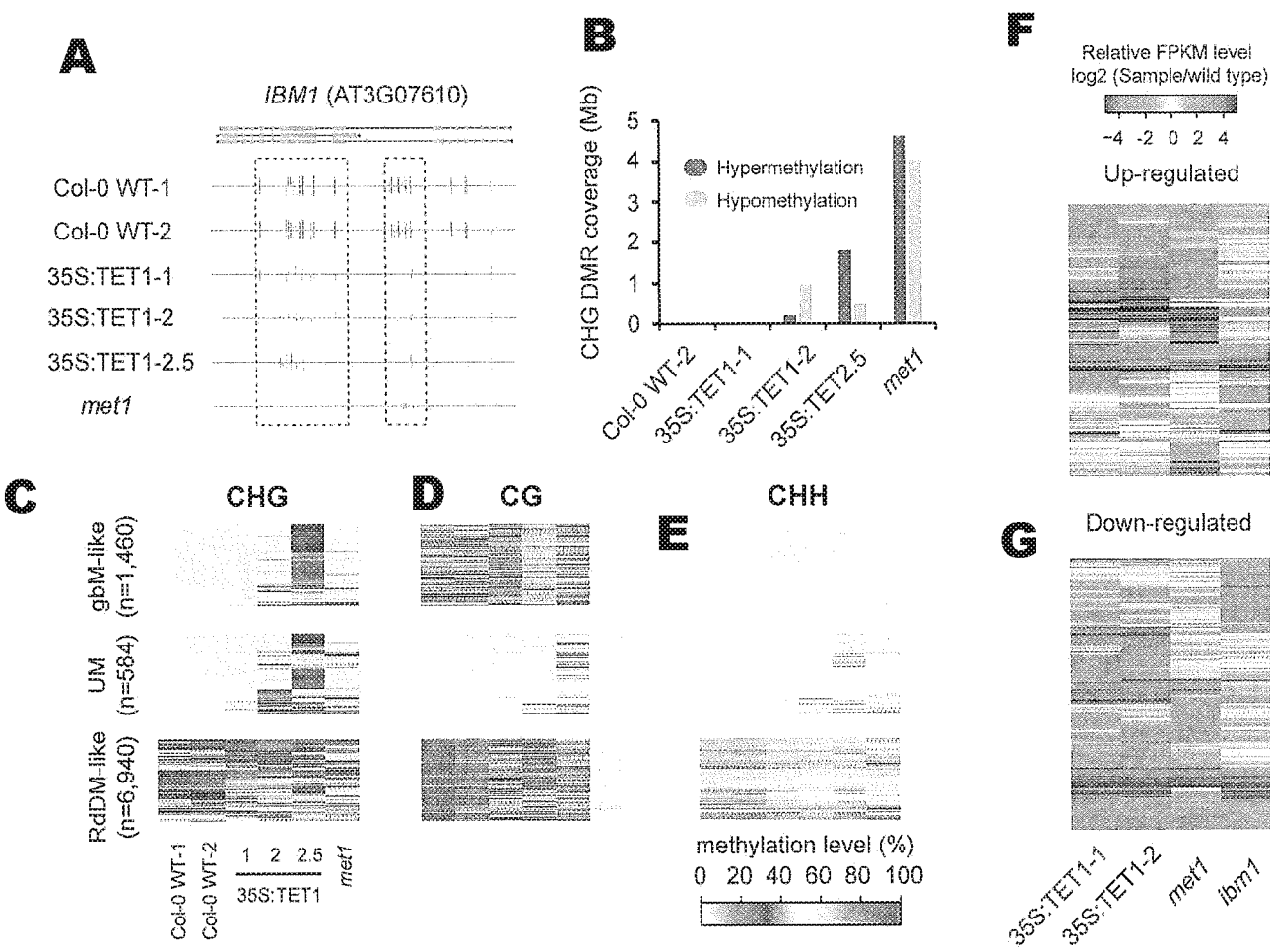


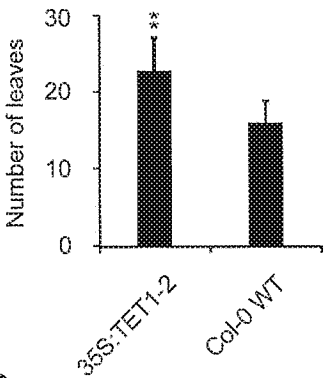
FIG. 3

A

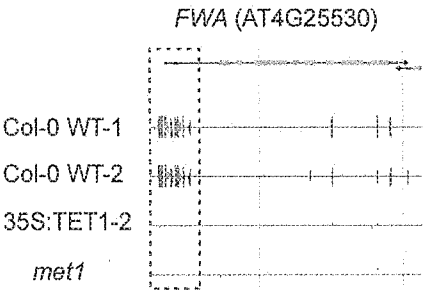


35S:TET1-2 Col-0 WT

B



C



D

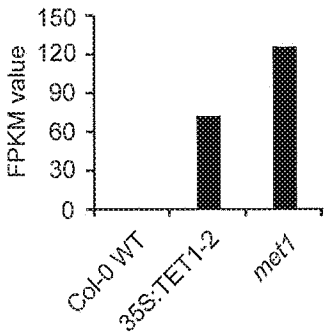


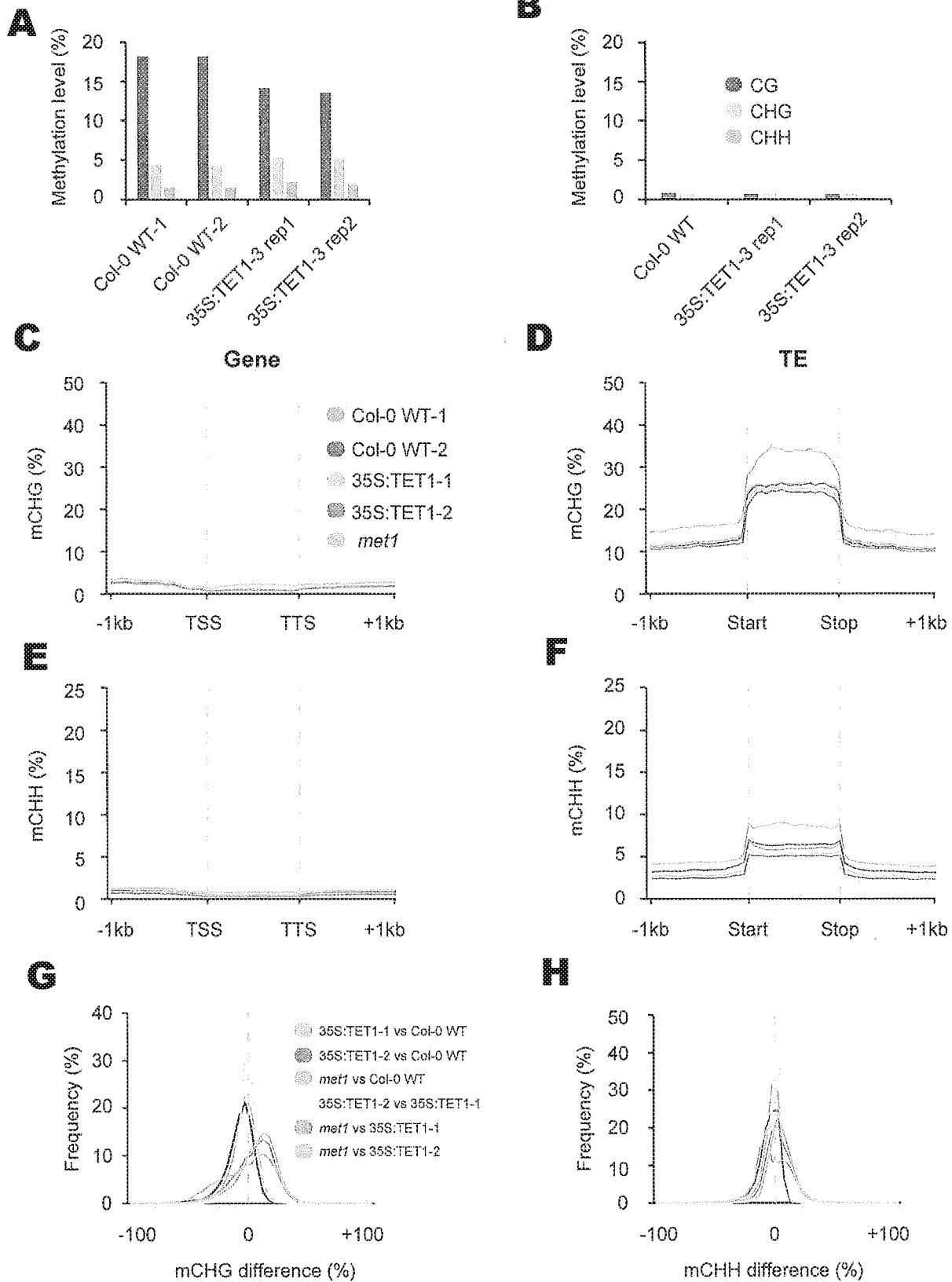
FIG. 4

FIG. 5-01

TET1, SEQ ID NO:1

```

1  MSRSRHARPS  RLVRKEDVKN  KKKNSQLRKT  TKGANKNVAS  VKTLPSPGKLG  QLIQERDVKK
61  KTEPKPPVPV  RSLLTRAGAA  RMNLDRTEVL  FQNPESLTCN  GFTMALRSTS  LSRRLSQPPL
121 VVAKSKKVPL  SKGLEKQHDC  DYKILPALGV  KHSENDVPM  QDTQVLPDIE  TLIGVQNPPL
181 LKGKSQETTQ  FWSQRVEDSK  INIPTHSGPA  AEILPGPLEG  TRCGEGLFSE  ETINDTSGSP
241 KMFAQDTVCA  PFPQRATPKV  TSQGNPSIQL  EELGSRVESL  KLSDSYLDPI  KSEHDCYPTS
301 SLNKVIPDLN  LRNCLALGGS  TSPTSVIKFL  LAGSKQATLG  AKPDHQEAFE  ATANQQEVSD
361 TTSFLGQAFG  AIPHQWELPG  ADPVHGEALG  ETPDLPEIPG  AIPVQGEVFG  TILDQQETLG
421 MSGSVVPDLP  VFLPVPPNPI  ATFNAPSKWP  EPQSTVSYGL  AVQGAIQILP  LSGHTEPQSS
481 SNSEKNSLPP  VMAISNVENE  KQVHISFLPA  NTQGFPLAPE  RGLFHASLGI  AQLSQAGPSK
541 SDRGSSQVSV  TSTVHVNTT  VVTMPVPMVS  TSSSSYTLL  PTLEKKKRKR  CGVCEPCQOK
601 TNCGECTYCK  NRKNSHQICK  KRKCEELKKK  PSVVVPLEVI  KENKRPQREK  KPKVLKADFD
661 NKPVNGPKSE  SMDYSRCGHG  EEQKLELNPH  TVENVTKNED  SMTGIEVEKW  TQNKKSQOLD
721 HVKGDFSANV  PEAEKSKNSE  VDKKRTKSPK  LFVQTVRNGI  KHVHCLPAET  NVSFKKFNIE
781 EFGKTLENN  YKFLKDTANH  KNAMSSVATD  MSCDHLKGRS  NVLVFQQPGF  NCSSIPHSSH
841 SIINHHASIH  NEGDAQKTPE  NIPSKEPKDG  SPVQPSLLSL  MKDRRLTLEQ  VVAIEALTQL
901 SEAPSENSSP  SKSEKDEESE  QRTASLLNSC  KAILYTVRKD  LQDPNLQGE  PKLNHCPSLE
961 KQSSCNTVVF  NGQTTLSNS  HINSATNQAS  TKSHEYSKVT  NSLSLFIPKS  NSSKIDTNKS
1021 IAQGIITLDN  CSNDLHQLPP  RNNVEYCNQ  LLDSSKKLDS  DDLSCQDATH  TQIEEDVATQ
1081 LTQLASIIKI  NYIKPEDKKV  ESTPTSLVTC  NVQOKYNQEK  GTIQQKPPSS  VHNHNGSSLT
1141 KQKNPTQKKT  KSTPSRDRRK  KKPTVVSQYE  NDRQKWEKLS  YMYGTICDIW  IASKFQNFQO
1201 FCPHDFPTVF  GKISSSTKIW  KPLAQTRSIM  QPKTVFPPLT  QIKLQRYPE  AEEKVKVEPL
1261 DSLSLFHLKT  ESNGKAFTDK  AYNSQVQLTV  NANQKAHPLT  QPSSPPNQCA  NVMAGDDQIR
1321 FQQVVKEQLM  HQRLPTLPGI  SHETPLPESA  LTLRNVNVVC  SGGITVVSTK  SEEEVCSSSF
1381 GTSEFSTVDS  AQKNFNDYAM  NFFTNPTKNL  VSITKDSELP  TCSCLDRIQ  KDKGPYYTHL
1441 GAGPSVAAVR  EIMENRYGQK  GNARIEIVV  YTGKEGKSSH  GCPIAKWVLR  RSSDEEKVLC
1501 LVRQRTGHH  PTAVMVVLIM  VWDGILPMA  DRLYTELTEN  LKSYNGHPTD  RRCTLNENRT
1561 CTCQGIDPET  CGASFSGGCS  WSMYFNGCKF  GRSPSPRRFR  IDPSSPLHEK  NLEDNLQSLA
1621 TRLAPIYKQY  APVAYQNQVE  YENVARECRL  GSKEGRFFSG  VTACLDLFC  PHRDIHNMNN
1681 GSTVVCITLR  EDNRSGLVIP  QDEQLHVLPL  YKLSDTDEFG  SKEGMEAKIK  SGAEIVLAPR
1741 RKKRTCFTQP  VPRSGKKRAA  MMTEVLAH  KIRAVEKKPIPR  IKRKNNSTTT  NNSKPSSLPT
1801 LGSNTETVQP  EVKSETEPHF  ILKSSDNTKT  YSLMPSAPHP  VKEASPGFSW  SPKTASATPA
1861 PLKN DATASC  GFERSSTPH  CTMPSGRLSG  ANAAAADGPG  ISQLGEVAPL  PTLSAPVMEP
1921 LINSEPSTGV  TEPLTPHQPN  HQPSFLTSPQ  DLASSPMEED  EQHSEADEPP  SDEPLSDDPL
1981 SPAEEKLPHI  DEYWSDESEH  FLDANIGGVA  IAPAHGSVLI  ECARRELHAT  TPVEHPNRNH
2041 PTRLSLVFYQ  HKNLNKPQHG  FELNKKIFE  KEAKNKKMKA  SEQKDQAANE  GPEQSSEVNE
2101 LNQIPSHKAL  TLTHDNVTV  SPYALTVHAG  PYNHWV

```

FIG. 5-02

TET2, SEQ ID NO:2

```

1  MEQDRTNHVE  GNRLSPFLIP  SPPICQTEPL  ATKLQNGSPL  PERAHPEVNG  DTKWHSFKSY
61  YGIPCMKGSQ  NSRVSPDFTQ  ESRGYSKCLQ  NGGIKRTVSE  PSLSGLLQIK  KLKQDQKANG
121 ERRNFGVSQE  RNPGESSQPN  VSDLSDKKES  VSSVAQENAV  KDFTSFSTHN  CSGPENPELQ
181 ILNEQEGKSA  NYHDKNIVLL  KNKAVLMPNG  ATVSASSVEH  THGELLEKTL  SQYYPDCVSI
241 AVQKTTSHIN  AINSQATNEL  SCEITHPSHT  SGQINSAQTS  NSELPPKPAA  VVSEACDADD
301 ADNASKLAAM  LNTCSFQKPE  QLQQQKSVEE  ICPSPAENNI  QGTTKLASGE  EFCSGSSSNL
361 QAPGGSSERY  LKQNEMNAY  FKQSSVFETD  SFSATTTPPP  PSQLLLSPPP  PLPQVPQLPS
421 EGKSTLNGGV  LEEHHHYPNQ  SNTTLLREVK  IEGKPEAPPS  QSPNPSTHVC  SPSMPLSERP
481 QNNVCVNRNDI  QTAGTMTVPL  CSEKTRPMSE  HLKHNPPIFG  SSGELQDNCQ  QLMRNKEQEI
541 LKGRDKEQTR  DLVPPTQHYL  KPGWIELKAP  RFHQEAESHLK  RNEASLPSIL  QYQPNLSNQM
601 TSKQYTGNSN  MPGGLPRQAY  TQKTTQLEHK  SQMYQVEMNQ  GQSQGTVDQH  LQFQKPSHQV
661 HFSPKDHLPK  AHVQSLCQTR  FHFQQRADSQ  TEKLMSPVLK  QHLNQQASET  EPFSNSHLLQ
721 HKPHKQAAQT  QPSQSSHLPO  NQQQQKLLQI  KNKEELLQTF  PHPQSNNDQQ  REGSFFGQTK
781 VEECFHGENQ  YSKSSEFETH  NVQMGLLEEVQ  NINRRNSPYS  QTMKSSACKI  QVSCSNNTHL
841 VSENKEQTH  PELFAGNKTO  NLHHMQYFPN  NVIPKQDLLH  RCFQEQEQKS  QQASVLQGYK
901 NRNDQMSGQQ  AAQLAQQRYL  IHNHANVFVP  PDQGGSHQT  PPQKDTQKHA  ALRWHLLQKQ
961 EQQQTQQPQT  ESCHSQMHRP  IKVEPGCKPH  ACMHTAPPEN  KTWKKVTKQE  NPPASCDNVQ
1021 QKSIIETMEQ  HLKQFHAKSL  FDHKALTLKS  QKQVKVEMSG  PVTVLTRQTT  AAELDSHTPA
1081 LEQQTTSSEK  TPTKRTAASV  LNNFIESPSK  LLDTPIKNLL  DTPVKTYQDF  PSCRCVEQII
1141 EKDEGPFFYTH  LGAGPNVAAI  REIMEERFQ  KGKAIRIERV  IYTGKEGKSS  QGCPKAKWV
1201 RRSSEEKLL  CLVRERAGHT  CEAAVIVILI  LVWEGIPLSL  ADKLYSELTE  TLRKYGTLTN
1261 RRCALNEERT  CACQGLDPET  CGASFSGGCS  WSMYYNGCKF  ARSKIPRKFK  LLGDDPKKEE
1321 KLESHLQNL  TLMAPTYKKL  APDAYNNQIE  YEHPAPECRL  GLKEGRPFSG  VTACLDCAH
1381 AHRDLHNMQN  GSTLVCTLTR  EDNREFGGKP  EDEQLHVLPL  YKVSVDVEFG  SVEAQEEKKR
1441 SGAIQVLSSF  RRVKRLAEP  VKTCRQRKLE  AKKAAAEKLS  SLENSSNKNE  KEKSAPSRTK
1501 QTENASQAKQ  LAELLRLSGP  VMQSQSQPQP  LQKQPPQPPQ  QORPQQQQPH  HPQTESVNSY
1561 SASGSTNPYM  RRPNPVSPYP  NSSHTSDIYG  STSPMNFYST  SSQAAGSYLN  SSNPMNYPG
1621 LLNQNTQYPS  YQCNGNLSVD  NCSPYLGSYS  PQSQPMDLYR  YPSQDPLSKL  SLPPIHLYQ
1681 PRFGNSQSFT  SKYLGYNQON  MQGDGFSSCT  IRPNVHHVGK  LPPYPHEMD  GHFMGATSRL
1741 PPNLSNPMD  YKNGEHHSPS  HIIHNSAAP  GMFNSSLHAL  HLQNKENDML  SHTANGLSKM
1801 LPALNHDRTA  CVQGGHLKLS  DANGQEQPL  ALVQGVASGA  EDNDEVWSDS  EQSFDPDIG
1861 GVAVAPTHGS  ILIECAKREL  HATTPKPNP  RNHPTRISLV  FYQHKSMEP  KHGLALWEAK
1921 MAEKAREKEE  ECEKYGPDYV  PQKSHGKKVK  REPAEPHETS  EPTYLRFIKS  LAERTMSVTT
1981 DSTVTTSPYA  FTRVTGPYNR  YI

```

7/19

FIG. 5-03

TET3, SEQ ID NO:3

```

1 MSQFQVPLAV QPDLFGLYDF PQRQVMVGSF PGSGLSMAGS ESQLRGGGDG RKKRKRCGTC
61 EPCRRLENCG ACTSCTNRRT HQICKLRKCE VLKKKVGLLK EVEIKAGEGA GPWGQGAAVK
121 TGSELSPLVDG PVPQMDSGP VYHGDSRQLS ASGVPVNGAR EPAGPSLLGT GGPWRVDQKP
181 DWEAAPGPAH TARLEDAHDL VAFSAVAEAV SSYGALSTRL YETFNREMSR EAGNNSRGPR
241 PGPEGCSAGS EDLDTLQTAL ALARHGMKPP NCNCDGPECP DYLEWLEGKI KSVVMEGGEE
301 RPRLPGLPFP GEAGLPAPST RPLLSSEVPQ ISPQEGPLPS QSALSIAKEK NISLQTAIAI
361 EALTQLSSAL PQPSHSTPQA SCPLPEALSP PAPFRSPQSY LRAPSWPVVP PEEHSSFAPD
421 SSAFPPATPR TEFPEAWGTD TPPATPRSSW PMPRPSPDPM AELEQLLGSA SDYIQSVFKR
481 PEALPTKPKV KVEAPSSSPA PAPSPVLQRE APTPSSEPDT HQKAQTALQQ HLHHKRSLEL
541 EQVHDTSFPA PSEPSAPGWW PPPSSPVPRL PDRPPKEKKK KLPTPAGGPV GTEKAAPGIK
601 PSVRKPIQIK KSRPREAQPL FPPVRQIVLE GLRSPASQEV QAHPPAPLPA SQGSAPVPLPP
661 PSRLALFAPS PSRDSLLPPT QEMRSPSPMT ALQPGSTGPL PPADDKLEEL IRQFEAEFGD
721 SFGLPGPPSV PIQDPENQQT CLPAPESPPA TRSPKQIKIE SSGAVTVLST TCFHSEEGGQ
781 EATPTKAENP LTPTLSGFLE SPLKYLDTPPT KSLLDTPAKR AQAEFPTCDC VEQIVEKDEG
841 PYYTHLGSGP TVASIRELME ERYGEKGKAI RIEKVIYTCK EGKSSRGCPI AKWVIRRHTL
901 EEKLLCLVRH RAGHHCQNAV IVILILAWEG IPRSLGDTLY QELTDTLRKY GNPTSRRRCL
961 NDDRTCACQG KDPNTCGASF SFGCSWSMYF NGCKYARSKT PRKFRLAGDN PKEEEVLRSK
1021 FQDLATEVAP LYKRLAPQAY QNQVTNEEIA IDCRLGLKEG RPFAGVTACM DFCAHAKDQ
1081 HNLYNGCTVV CTLTKEDNRC VGKIPEDQL HVLPLYKMAN TDEFGSEENQ NAKVGSGAIQ
1141 VLTAFPREVR RLPEPAKSCR QRQLEARKAA AEKKKIQKEK LSTPEKIQE ALELAGITSD
1201 PGLSLKGGLS QQGLKPSLKV EPQNHFSSEK YSGNAVVESE SVLGNCRPSD PYMNSVYSY
1261 HSYAQPSLT SVNGFHASKYA LPSFSYYGFP SSNPVFPSQF LGPGAWGHS SSGSFEKKPD
1321 LHALHNSLSP AYGGAEFAEL PSQAVPTDAH HPTPHHQQPA YPGPKYLLP KAPLLHSVSR
1381 DSPPFAQSSN CYNRSIKQEP VDPLTQAEV PRDAGKMCKT PLSEVSQNGG PSHLWQYSG
1441 GPSMSPKRTN GVGGSWGVSF SGESPAIVPD KLSSFGASCL APSHFTDGQW GLFPGEQQA
1501 ASHSGGRLRG KPWSCKEFGN STSALAGPSL TEKPWALGAG DFNSALKGSP GFQDKLWNP
1561 KGEEGRIPAA GASQLDRAWQ SFGLPLGSSE KLFALKSEE KLWDPPSLEE GPAEEPPSKG
1621 AVKEEKGGGG AEEEEELWS DSEHNFLDEN IGGVAVAPAH GSILIECARR ELHATTPKK
1681 PNRCHPTRIS LVFYQHKNLN QPNHGLALWE AKMKQLAERA RARQEEAARL GLGQQAELY
1741 GKRRKWGGTV VAEPQKKEK GVVPTQALA VPTDSAVTVS SYAYTKVTGP YSRWI

```

8/19

FIG. 6-01

```

SEQ ID NO:1      MSRSRHARPSRLVRKEDVNKKKKNSQLRKTTKGANKNVASVKTLSPGKLGKLIQERDVKK
SEQ ID NO:2      -----
SEQ ID NO:3      -----

SEQ ID NO:1      KTEPKPPVPVRSLLTRAGAARMNLDRTFVLFQNPESLTCNGFTMALRSTLSRRLSQPPL
SEQ ID NO:2      -----
SEQ ID NO:3      -----

SEQ ID NO:1      VVAKSKKVPLSKGLEKQHDCDYKILPALGVKHSENDSVPMQDTQVLPDIETLIGVQNPSL
SEQ ID NO:2      -----
SEQ ID NO:3      -----

SEQ ID NO:1      LKGKSQETTQFWSQRVEDSKINIPTHSGPAAEILPGPLEGTRCGEGLFSEETLNDTSGSP
SEQ ID NO:2      -----
SEQ ID NO:3      -----

SEQ ID NO:1      KMFQADTVCAPFPQRATPKVTSQGNPSIQLEELGSRVESLKLSDSYLDPFKSEHDCYPTS
SEQ ID NO:2      -----
SEQ ID NO:3      -----

SEQ ID NO:1      SLNKVIPDLNLRNCLALGGSTSPTSVIKFLLAGSKQATLGAKFDHQEAFEATANQQEVSD
SEQ ID NO:2      -----MEQDR
SEQ ID NO:3      -----

SEQ ID NO:1      TTSFLGQAFGAIPHQWELPGADPVHGEALGETPDLPEIPGAI PVQGEVFGTILDQQETLG
SEQ ID NO:2      TNHVEG---NRLSPFLIPSPPICQ-----TEPLA-----
SEQ ID NO:3      -----MSQFQV-----
                      : :

SEQ ID NO:1      MSGSVVPDLPVFLPVPPNPIATFNAPSKWPEPQSTVSYGLAV-----
SEQ ID NO:2      -----TKLQNGSPLPERAHPEVNGDTKWHSEFKSY--YGIPCMKGSQNSRVSPDFTQESR
SEQ ID NO:3      -----PLAV-----
                      *

SEQ ID NO:1      -----QGAIQILPLG-----SGHTPQSSS----NSEKNSLPPVMA
SEQ ID NO:2      GYSKCLQNGGIKRTVSEPSLSGLLQIKKLQDQKANGERFRNFGVSQERNPGESSQPNVSD
SEQ ID NO:3      -----QPDLPGLYDFPQ---RQVMVGSFPGSGLS-----
                      * : : * .

SEQ ID NO:1      IS-----NVENE--RQVHISFLPANTQGFPLAPERGLFHASLGI-----
SEQ ID NO:2      LSDKKESVSSVAQENAVKDFTSFSTHNCSE--PENPELQILNEQEGKSANYHDKNIVLLKN
SEQ ID NO:3      -----MA-----

SEQ ID NO:1      -AQLSQAGPSKSDRGSSQVSVTSTVHVNTTVVTMPVPMVSTS--SSSYTTLPTLEKKK
SEQ ID NO:2      KAVLMPNGATVS---ASSVEHTH--GELLEKTLQYYPDCVSIQVQKTTSHINAINSQATN
SEQ ID NO:3      -----GSESQLRGGGDGRKK
                      : :

SEQ ID NO:1      RRRCGVCEPCQ-----QKTNCGECTYCKNR--KNSHQICKRRK
SEQ ID NO:2      ELSCEITHPSHTSGQINSQAQTSNSELPPKPAAVVSEACDADDADNASKLAAMLNTCSFQK
SEQ ID NO:3      RRRCGTCEPCR-----RLENCGACTSC---TNRRTHQICKLRK
                      . * .*, : * : *, : *

SEQ ID NO:1      CEELKKKPSVVVPLEVIKENKRPQREKKF----KVLKADFNDKPVNGPKSESMDYSRC--
SEQ ID NO:2      PEQLQQQKSVFEICPSPAENNIQGTTKLASGEEFCGSSSNLQAPGSSSERVLKQNMNG
SEQ ID NO:3      CEVLKKKVGLLKEVEIKA---GEGAGPWGQGAAVKTGS---ELSPVDGPVPGQM---DSGP
                      * *:: : : : : . * :

```

9/19

FIG. 6-02

```

SEQ ID NO:1  --GHGEEQKLELNPHTVENVTKNE-DSMTG-----IEVEKWTQNKKSQQLTDHVKGDFSA
SEQ ID NO:2  AYFKQSSVFTKDSFSA--TTTPPPSOLLSPPLPQVPQLPSEGKSTLNGGVLEEHH--
SEQ ID NO:3  VYHGDSRQLSASGVVNGAREPAGPSLLGTGGPWRVDQKPDWEAAPGP-AHTARLEDAHD
      . : . : . : * : . :

SEQ ID NO:1  NVPEAEKSKNSEVDKRTKSPKLFVQTVR----NGI-----KRVHCLPAETN--
SEQ ID NO:2  -----HYPNQS---NTTLLREVKIEGKPEAPPSQSPNPSTHVCSFSPMLSE
SEQ ID NO:3  LVAFSA--VAEAVSSYGALSTRLYETFNREMSRE---AGNNSRGPRPGPEGCSAGSE---
      * * ; *

SEQ ID NO:1  ----VSFKKFNIEEFKGTLENNSYKFLKDTANHKNAMSSVATDMSCDHLKGRSNVLVFQO
SEQ ID NO:2  RQNNNCVNRNDIQT-----GTMTVPLCSEKTRPMSEHLKHNP
SEQ ID NO:3  -----DLDTL-----QT-ALALAR-----HGMKP
      ::: : : :

SEQ ID NO:1  PGFNCSSIPHSSH-----SIINHASIHNEDQPKTPE
SEQ ID NO:2  FIFGSSGELQDNCQQLM---RNKEQ-EILKGR-DKEQTRDLVP-----
SEQ ID NO:3  PNCNCDGP---ECPDYLEWLEGIKSVVMEGGEERPLPGPLPPEAGLPAPSTRPLLS
      * . . . . :

SEQ ID NO:1  NIPSKEPKDGSPVQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSENSSPSKSEKDE--E
SEQ ID NO:2  -----
SEQ ID NO:3  EVPGISPQEGPLPLSQSALSIAKEKNISLQTAIAIEALTQLSSALFPQSHSTPQASCPLE

SEQ ID NO:1  SEQR-----TASLLNSCKAILYT-----VRKDLQDPNLQGEPPKLNHCPSLEK
SEQ ID NO:2  -----PTQHYLKPGWIELKAPRFHQAESHLKRNEASLPSILQYQPN-----LSN
SEQ ID NO:3  ALSFPAPFRSPQSYLRAPSWPVVPEEHS---SFA-----PDSSAFPPA-----TPR
      : * * .

SEQ ID NO:1  QSSCNTVVFNGQTTTLENSHINSATNQASTKSHEYSKVTNSLSLFIKPSNSSKIDTNKSI
SEQ ID NO:2  QMT--SKQYTGNSNMPG-----G----LPROAYTQKTTQL---EHKSQMYQVEMNQOQ
SEQ ID NO:3  TEF--PEAW--GTDTPF-----A-----TPRSSWPMP-----
      : : . :

SEQ ID NO:1  AQGIIT----LDNCSNDLHQLPP----RNNEVEYCNQLDSSKKLSDDDLSCQDATHTQI
SEQ ID NO:2  SQGTVDQHLQFQKPSHQVHFSKTDHLPKARVQSLCGTRFHQQRADSQTEKLM---SPVL
SEQ ID NO:3  -----RPSDPMAELEQLLGSASDYIQSVFKRPEALP--TKPKV
      . . . . :

SEQ ID NO:1  EEDVATQLTQLASII--KINYIKPEDKKVESTPT--SLVTCNVQO--KYN-QEKGTIQOK
SEQ ID NO:2  KOHLNQASSETEPFSNSHLLQHKPHKQAAQTQPSQSSHLQONQQQQLQIKNKEEILQT
SEQ ID NO:3  K--VEAPSSSPAPAPS--PVLQRE-----APTSPSEPETHQKAQTALQ-----QH
      : : : : * : * : : *

SEQ ID NO:1  PPSSVHN----NHGSSLTKQKNPTQKKTKSTPSRDRRKKK----PTVVSQENDRQKWEK
SEQ ID NO:2  FPHPQSNNDQQREGSFFGQTKVEECFHGENQYSKSSEFETHNVQMGLLEVQNINRRNSPY
SEQ ID NO:3  LHH-----KRSLFLEQVHDTSFAPSEP-----
      . * : :

SEQ ID NO:1  LSYMYGTICDIWIA-----SKFQ-----NFG
SEQ ID NO:2  SQTMKSSACKIQVSCSNNTLHVSSENKEQTTHPELFAGNKTQNLHMQYFPNNVIPKQDLL
SEQ ID NO:3  -----SAFGWWPPSSPVERLP
      :

SEQ ID NO:1  QFCPHD----FPTVFGKISS-----TRIWKPLAQTR-SIMQPKTVFPPLTQIKLQR
SEQ ID NO:2  HRCFQEQEQK-----SQQASVLQG
SEQ ID NO:3  DRPPKEKKKKLPTPAGGPVGTEKAAPGIKPSVRKPIQIKKSRPREAQPLFPFPVRQIVLEG
      . :: * :

SEQ ID NO:1  YPESAEK-----VKVEPLDSLFLHL-----
SEQ ID NO:2  YKNRNQDMSGQAAQLAQQRYLHNHANVFPVFDQGGSHQTTPPQKDTQKHAA--LRWHL
SEQ ID NO:3  LRSFAS-----QEVQ---AHPPAPLPASQGSAPVLPPEPSLALFAPSPSRDSL
      . * . .

SEQ ID NO:1  KTES---NGKAFTDKAYNSQVQ--L-----TVNANQKAHPLTQPSSPPNQO
SEQ ID NO:2  LQKQEQQQTQEQTESCHSQMHRPIKVEPGCKPHACMHTAPPENKTWKKVTRQENPPASC
SEQ ID NO:3  LPPTQEMRSPSP-----M---TALQFGST-----GPLPPADDK-----
      . : :

```

[illegible]

11/19

FIG. 6-04

```

SEQ ID NO:1      EPLINS-----EPSTGV---TEPLTPHQ-----PNHQPSFLTSP-----
SEQ ID NO:2      F---MGATSRLPPN---L-SNPNM--DY--KNGEHHSPSHIHNYSAAFGMFNSSLHAL
SEQ ID NO:3      WGLFPGEGQQAASHSGGRLRGKFWSPCKFGNSTSALAGPSL/TEKPWALGAGDFNSALKGS
                  .          : *          .          **

SEQ ID NO:1      -----QDLASSPMEEDE-----
SEQ ID NO:2      HLQNKENDMLSHNTANGLSKMLPALNH---DR-----TACVQGGLH---KLSD---
SEQ ID NO:3      PGF---QDKLWNPMKGEEGRIPAAGASQLDRAWQSFGLPLGSSEKLFALKSEEKLWDPF
                  : *          : .

SEQ ID NO:1      --QHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDEHIFLDANIGCVATAPAHGSVL
SEQ ID NO:2      --ANGQ-----EKQPLALVQ--GVASGAEDNDEVWSDSEQSFLLDPDIGGVAVAPTHGSIL
SEQ ID NO:3      SLEEGPAEPPSPKGAVKEEK--GGGGAEEEEELWSDSEHNFLDENIGGVAVAPAHGSIL
                  ..          ..          :          . : * ***** : *** ;*****:***:***:*

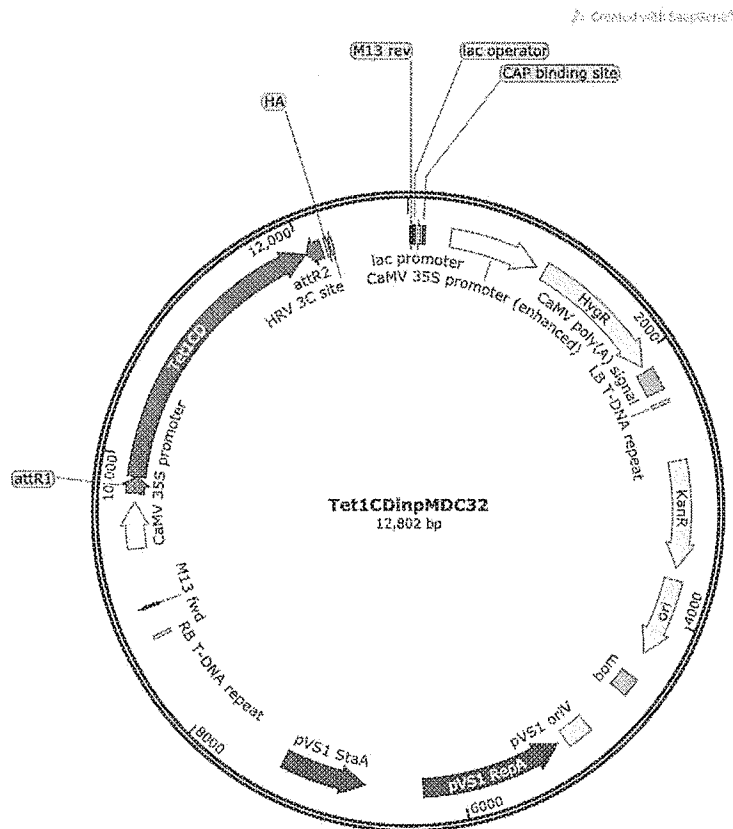
SEQ ID NO:1      IECARRELHATTPEHPNRRNHPTRLISLVFYQHKNLNKPQHGFELNKKIKFEAKEAKNKKMK
SEQ ID NO:2      IECAKRELHATTPLKNPNRRNHPTRISLVFYQHKS MNPKHGLALWEAKMAEKA---REKE
SEQ ID NO:3      IECARRELHATTPLKKPNRCHPTRISLVFYQHKNLNQPNHGLALWEAKMKQLAERARARQ
                  *****:*** *****:*****.:*:***: * : *;          : :

SEQ ID NO:1      AS-----EQK-----DQAANEGFEQSSEV--NELNQIPSHKALTLTHTDNVTV
SEQ ID NO:2      EECEKYGPDYVPQKSHGKKVKREP---AEPHETSEPTYLRFIKSLAERTMSVTTDSTVTT
SEQ ID NO:3      EEAARLGLGQQAELYGKKRWGCTVVAEPQQKEKKG-----VVPTRQALAVPTDSAVTV
                  .          *          *.:...:          :.....: *..**

SEQ ID NO:1      SPYALTHVAGPYNHV
SEQ ID NO:2      SPYAFTRVTGPYNRYI
SEQ ID NO:3      SSYAYTKVTGPYSRWI
                  * * * *:***:..:

```

12/19

FIG. 7A**FIG. 7B-01**

>Tet1CDinpMDC32 (12,802 bp)

CGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCACACAAACATACGAGCCGGAAGCATAAAGTGT
AAAGCCTGGGGTGCCATATGAGTGAGCTAACTCACATTAATGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCTGCATTAATGAATCGGCCAAGCGCGGGAGAGGCGGTTTGGCTATTGGCTAGAGCAGCTTGCCAAACATGGTGGAGCACGA
CACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAGGAGGCTATTGAGACTTTTCAACAAAGGGTAATATCGG
GAAACCTCCTCGGATTCCATTGCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCAT
CATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGT
GGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCGTCTACTCCA
AGAATATCAAAGATACAGTCTCAGAAGACCAGGAGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATT
CATTGCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAA
GGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTT
CAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGAC
CTTCTCTATATAAGGAAGTTCATTTTCAATTTGGAGAGGACACGCTGAAATCACCAGTCTCTCTACAAATCTATCTCTCTCGAGC
TTTCGCGAGATCCCGGGGGCAATGAGATATGAAAAGCCTGAACTCACCGGACGCTCTGTCGAGAAGTTTCTGATCGAAAAGTTTCG
ACAGCGTCTCCGACCTGATGACGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTG
CGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTTCATCGGCCGCGCTCCCGATTCCGGAAGT
GCTTGACATTGGGGAGTTTAGCGAGAGCCTGACCTATTGCATCTCCCGCGCTGCACAGGCTGTCAGTTGCAAGACCTGCCTGAAA
CCGAAGTCCCGCTGTTCTACAACCGGTGCGGAGGCTATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGC
CCATTCCGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCA
AACTGTGATGGACGACACCGTCACTGCGTCCGTGCGCGAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCGAAGTCC
GGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCTTGACGGACAATGGCCGCATAACAGCGGTCTTGAAGTGGAGCGAGGCG
ATGTTCCGGGATTCCCAATACGAGGTGCGCAACATCTTCTCTGAGGCGGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTT
CGAGCGGAGGATCCGGAGCTTGCAGGATCGCCACGACTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCT
TGGTTGACGGCAATTTTCATGATGACAGCTTGGGCGCAGGCTCGATGCGACGCAATCGTCCGATCCGGAGCGGGAGTGTCCGGCGT
ACACAAATCGCCCGCAGAAGCGCGGCGTCTGGACCGATGGCTGTGTAGAGTACTCGCGGATAGTGGAAACCGACGCCCGACAC
TCGTCCGAGGGCAAAGAAATAGAGTAGATGCCGACCGATCTGTGATCGACAAGCTCGAGTTTCTCCATAAATGTGTGAGTAG

FIG. 7B-02

TTCCAGATAAGGGAATTAGGGTTCTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAACCCCTTAGTATGTATTTGTATTTGT
AAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAAATCCAGTACTAAAATCCAGATCCCCGAATTAATTCGGCGTTAA
TTCAGTACATTAAAAACGTCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGCCACGAGCC
AGCCAACAGCTCCCCGACCGCAGCTCGGCACAAAATCACCCTCGATACAGGCAGCCCATCAGTCCGGGACGGCGTCAGCGGGAG
AGCCGTTGTAAAGCGGCAGACTTTGCTCATGTTACCGATGCTATTGCGGAAGAACGGCAACTAAGCTGCCGGGTTTGAACACGGAT
GATCTCGCGGAGGGTAGCATGTTGATTGTAACGATGACAGAGCGTTGCTGCCTGTGATCACCGCGGTTTCAAATCGGCTCCGTCG
ATACTATGTTATACGCCAATCTTGAACCAACTTTGAAAAGCTGTTTTCTGGTATTTAAGGTTTGTAGATGCAAGGAACAGTGAA
TTGGAGTTCGTCCTTGTATTAATTAGCTTCTTGGGGTATCTTTAAATACCTGTAGAAAAGAGGAAGGAAATAATAAATGGCTAAAATG
AGAATATCACCGGAATTGAAAAAATGATCGAAAAATACCGCTGCGTAAAAGATACGGAAGGAATGTCTCTGCTAAGGTATATAA
GCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGACGGACAGCCGCTATAAAGGGACCACCTATGATGTGGAACGGGAAAAGG
ACATGATGCTATGGCTGGAAGGAAAGCTGCTGTTCCAAAGGTCTGCTGACTTTGAACGGCATGATGGCTGGAGCAATCTGCTCATG
AGTGAGGCGCATGGCGTCTTTGCTCGGAAGAGTATGAAGATGAACAAAGCCCTGAAAAGATTATCGAGCTGTATCGGGAGTGCAT
CAGGCTCTTTCACTCCATCGACATATCGGATTGTCCCTATACGAATAGCTTAGACAGCCGCTTAGCCGAATTGGATTACTTACTGA
ATAACGATCTGGCCGATGTGGATTGCGAAAACCTGGGAAGAGACACTCCATTTAAAGATCCGCGCGAGCTGTATGATTTTTTAAAG
ACGGAAAAGCCGAAGAGGAACCTTGTCTTTTCCACGGCGACCTGGGAGACAGCAACATCTTTGTGAAAGATGGCAAAGTAAGTGG
CTTTATTGATCTTGGGAGAAGCGGCAGGGCGGACAAGTGGTATGACATTGCCTTCTGCGTCCGGTCGATCAGGGAGGATATCGGGG
AAGAACAGTATGTGAGCTATTTTTTACTTACTGGGGATCAAGCCTGATTGGGAGAAAATAAAATATATATTTTACTGGATGAA
TTGTTTTAGTACCTAGAATGCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGA
CCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCTTTTTTTCTGCGCTAATCTGCTGCTTGCAAACAAAAAACCCGCTAC
CAGCGGTGGTTTGTGTTGCCGATCAAGAGCTACCAACTCTTTTTCCGAAGTAAGTGGCTTCAGCAGAGCGCAGATACCAATACT
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACC
AGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGGCT
GAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAAGCACTACACCGAAGTACGATACCTACAGCGTGTGATGAGAAAGC
GCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACAGGGAGGATTCACAG
GGGAAACGCTTGGTATCTTTATAGTCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
GGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTTACGGTTCTTGGCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCG
TTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCGAGCCGACGACCGAGCGCAGCGA
GTCAGTGAGCGAGGAAGCGGAAGAGCGCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATGCTGCA
CTCTCAGTACAATCTGCTCTGATGCCCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCTATGGCTGCGCCCC
GACACCGCCCAACACCGCTGACGCGCCTGACGCGCTTGTCTGCTCCGCGATCCGCTTACAGACAAGCTGTGAGCTTCTCGGGG
AGCTGCATGTGTAGAGGTTTTCACCGTCTACCCGAAACGCGGAGGCGAGGGTGCCTTGATGTGGGCGCGCGCGGTGAGTGGCG
ACGGCGCGGCTTGTCCGCGCCTGGTAGATTGCCTGGCGTAGGCCAGCCATTTTTGAGCGGCCAGCGCGCGGATAGGCCGACGC
GAAGCGGCGGGCGTAGGGAGCGCAGCGACCGAAGGGTAGGCGCTTTTTGAGCTCTTTCGCTGTGCGCTGGCCAGACAGTTATGC
ACAGGCCAGCGGGTTTTAAGAGTTTTAATAAGTTTTAAGAGTTTTAGCGGAAAAATCGCCTTTTTCTCTTTTATATCAGTCA
CTTACATGTGTGACCGGTTCCCAATGTACGGCTTTGGGTTCCCAATGTACGGGTTCCGCTTCCCAATGTACGGCTTTGGGTTCCCA
ATGTACGTGCTATCCACAGGAAAGAGACCTTTTCAGCTTTTTTCCCTGCTAGGGCAATTTGCCCTAGCATCTGCTCCGTACATTA
GGAACCGCGGATGCTTCGCCCTCGATCAGGTTGCGGTAGCGCATGAGTACGAGTACGGGCGAGCCTGCCCTCTCTCTCAAAT
CGTACTCCGGCAGGTCATTTGACCCGATCAGCTTGCGCACGGTGAAACAGAACTTCTTGAACCTCTCCGGCGCTGCCACTGCGTTG
TAGATCGTCTTGAACAACCATCTGGCTTCTGCCCTTGCCCTGCGCGCGCGGCTGCCAGGCGGTAGAGAAAACGGCCGATGCCGGGATC
GATCAAAAAGTAATCGGGGTGAACCGTCAGCACGTCCGGGTTCTTGCTTCTGTGATCTCGCGGTACATCCAATCAGCTAGCTCGA
TCTCGATGTACTCCGGCGCGCCGGTTTCGCTCTTTACGATCTTGTAGCGGCTAATCAAGGCTTACCCCTCGGATACCGTCACCAGG
CGGCCGTTCTTGGCCTTCTTGTAGCTGCTATGGCAACGTGCGTGGTGTAAACCGAATGCAGGTTTCTACAGGTCGTCTTTCTG
CTTTCCGCTATCGGCTCGCGCGCAGAACTTGAGTACGTCGCGCAACGTGTGGACGGAACACGCGCGCGGCTTGTCTCCTTCCCT
CCCGGTCGTGTTTATGATTGCTGTTAGTGGTAAGCAACCGCATCAGTACAGGTCGTAATCCACACACTGGCCATGCGCGCGGCG
CCTGCGGAAACCTCTACGTGCCGCTGGAAGCTGCTAGCGGATCACCTCGCCAGCTCGTGGTTCACGCTTCGACAGACGGAAC
GGCCACGTCCATGATGCTGCGACTATCGCGGCTGCCACGTATAGAGCATCGGAACGAAAAATCTGGTTGCTCGTCCGCTTGG
GCGGCTTCTAATCGACGCGCACCGGCTGCCGCGGTTGCCGGATTCTTTCGCGATTGATCAGCGCGCGCTTGCACGATTCA
CCGGGGCGTCTTCTGCCTCGATGCGTTGCCGCTGGGCGGCTGCGCGGCTTCAACTTCTCCACAGGTCATCACCCAGCGCGC
GCCGATTTGTACCGGGCGGATGGTTTGGACCGCTCACGCGGATCTCTCGGCTTGGGGGTTCCAGTGCCATTGCAGGGCGCGCAG
ACAACCCAGCGCTTACCGCTGGCCAAACCGCCGTTCCCTCCACACATGGGGCATTCCACGGCGTGGTGCTGGTTGTTCTTGATT
TTCCATGCCGCTCTTTAGCCGCTAAAATTCATCTACTCATTTATTTCTTGTCTCATTTACTCTGGTAGCTGCGCGATGTATTCA
GATAGCAGCTCGGTAATGGTCTTGGCTTGGCGTACCGCTACATCTTCACTTGGTGTGATCCTCCGCGGCAACTGAAAGTTGAC
CGCTTTCATGGCTGGCGTGTCTGCCAGGCTGGCCAACTGTCAGCCTTGTCTGCTGCGTGGCTCGGACGCGCGGCACTTAGCGTGT
TTGTGCTTTTGTCTCAATTTCTCTTTACCTCATTAACCTCAAATGAGTTTTGATTTAAATTCAGCGCGCAGCGCTGGACCTCGCGGG
CAGCGTCGCGCTCGGGTCTGATTCAAGAACGGTGTGCCGCGCGCGGCGAGTGCCTGGGTAGCTCACGCGCTGCGTGATACGGGAC
TCAAGAATGGGAGCTCGTACCGCGCCAGCGCTCGGCAACCTCACCGCGATCGCGCTGCGCTTGTATCGCGCGGACACGACAAA
GGCGCTTGTAGCTTCCATCCGCTGACCTCAATGCGCTGCTTAAACAGCTCCACAGGTCGCGCGGCGGCGGCGGATGTCTGTAAGGCG
TTGGCTGCACCGGAATCAGCACGAAGTGGCTGCTTGTATCGCGGACACAGCCAGTCCGCGCGCTGGGCGCTCGTGCATCACT
ACGAAGTGCGCCGCGCGGATGGCTTCACTGCGGCTCAATCGTCCGGCGGTGATGCCGACAACGGTTAGCGGTTGATCTTCCG
CACGGCCGCGCAATCGCGGGCACTGCCCTGGGGATCGGAATCGACTAACAGAACATCGGCCCCGCGGAGTTGCAGGGCGCGGGCTA
GATGGGTTGCGATGGTCTTGTGCTGACCCGCTTTCTGGTTAAGTACAGCGATAACCTTCATGCGTTCCCTTGCCTATTTGTT
TATTTACTCATCGCATCATATACGACGCGACCGCATGACGCAAGCTGTTTTACTCAAATACACATCACCTTTTTAGACGGCGCGCG
TCGGTTTCTTACGCGCCAAAGCTGGCCGCGCAGGCCGCGGCTTGGCATCAGACAAACCGGCCAGGATTCATGACGCGCACGCT

TGAGACGTGCGCGGGCGGCTCGAACACGTACCCGGCCGCGATCATCTCCGCCTCGAFTCTTTCGGTAATGAAAAACGGTTTCGTCTT
GGCCGTCTCTGGTGCGGTTTCATGCTTGTCTCTTGGCGCTTCATTCTCGGGCGGCCGACAGGGCGTCGGCCTCGGTCAATGCGTCTCT
CACGGAAGGCACCGCGCCGCTTGGCCTCGGTGGGCGTCACTTCTCGCTGCGCTCAAGTGCGCGGTACAGGGTCGAGCGATGCACG
CCAAGCAGTGCAGCGCCCTCTTTCACGGTGGCGGCTCTCTGGTGCATCAGCTCGCGGGCGTCGCGCATCTGTGCCGGGGTGAGGTT
AGGGCGGGGGCAAACCTTACCGCTCGGCCCTTGGCGGCTCGCGCCGCTCGGGTGCGGTGCATGATTAGGGAACGCTCGAACC
CGGCAATGCCGGGCAACAGGTCAACACCATGCGCGCGGCCGCGCGTGGTGGTGTGCGGCCACGGCTCTGCCAGGTCACGAGCCC
GCGCCGCTCTCTGGATGCGCTCGGCAATGTCCAGTAGGTGCGGGTGCTGCGGGCCAGGCGGTCTAGCCTGGTCACTGTCACAAC
GTCGCCAGGGCGTAGGTGGTCAAGCATCTTGCCAGCTCGGGCGGTGCGCCTGGTGCCGGTGATCTTCTCGGAAAACAGCTTGG
TGCAGCCGCGCGCTGCACTTCCGCCCCTTGGTTGGTCAAGTCTCTGGTGTGCGGTGCTGACGCGGGCATAGCCAGCAGGCCAGCG
CGGCGCTCTTGTTCATGGCGTAATGTCTCCGGTTCTAGTCGCAAGTATTCTACTTTTACGCACTAAAAACCGGCACAAGAAACCC
CCAGGAAAAGGGCAGGGCGCAGCTGTGCGCTAATCTAGGACTTGTGCGACATCTGTTTTCAGAAACAGCGCTGCATGAACGT
AGAAGCCGATGCACTATAGCAGCGGAGGGTTGGATCAAAGTACTTTGATCCCCAGGGGAACCCCTGTGGTTGGCATGCAATACA
AATGGACGAACGGATAAACCTTTTACGCCCTTTTAAATATCCGTTATTCTAATAAACGCTCTTTTCTCTTAGGTTTACCGCCAA
TATATCTGTCAAACACTGATAGTTTAACTGAAGCGGGAAACGACAATCTGATCCAAGCTCAAGCTGCTCTAGCATTCGCCATT
CAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCTCTTCGTATTACGCCAGCTGGCGAAAGGGGGATGTCTGCAAGGC
GATTAAAGTTGGGTAACGCCAGGCTTTTCCAGTCACGACGTTGTAACAGCAGCGCCAGTGCCAAAGCTTGGCGTGGCTGCAGGTCAA
CATGGTGGAGCAGCACACTTGTCTACTCCAAAATATCAAAGTATCAGTCTCAGAAGACCAAGGGCAATTGAGACTTTTCAAC
AAAGGTAATATCCGGAACCTCTCGGAATCCATTGCCAGTATCTGTCACTTTATTGTGAAGTATGTGAAAAGGAAGGTGGC
TCCTACAAATGCCATCATTTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACC
CACGAGGAGCATCGTGGAAAAGAAGACGTTCCAACCCAGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCAGCACA
CACTTGTCTACTCCAAAATATCAAAGTACAGTCTCAGAAGACCAAGGGCAATTGAGACTTTTCACAAAGGTTAATATCCGGA
AACCTCTCTCGGAATCCATTGCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCTTACAAATGCCATG
TTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCGAGGAGCATCGTGG
AAAAAGAAGACGTTCCAACCCAGTCTTCCAAGCAAGTGGATGATGATATCTCACTGACGTAAAGGATGACGCAATCCAC
TATCCTTCGCAAGACCCTTCTCTATATAAGGAAGTTCATTTTATTGGAGAGGACCTCGACTCTAGAGGATCCCCGGGTACCGGG
CCCCCTCTCGAGGCGCGCAAGCTATCAAACAAGTTTGTACAAAAAGCTGAACGAGAAAACGTAATATGATATAAATATCAATATA
TTAAATTAGATTTTGCATAAAAAACAGACTACATAAATACTGTAAAACACAACATATCCAGTCACTATGCACCATGGAAGTGGCCAC
CTGCAGCTGTCTTGATCGAGTTATACAAAAAGACAAAGGCCCATATATACACACCTTGGGGCAGGACCAAGTGTCTGTGTCA
GGGAATCATGGGAATAGGTATGGTCAAAAAAGGAAACGCAATAAGGATAGAAATAGTAGTGATACACCGGTAAAGAAGGGAAAAGC
TCTCATGGGTGTTCAATTGTCTAAGTGGTTTGAAGAAGACAGCAGTGAAGAAAAGAGTCTTTGTTGGTCCGCGAGCGTACAGG
CCACCAGTGTCCAAGTGTGTGATGGTGGTGCTCATCATGGTGTGGGATGGCATCCCTCTTCCAATGGCCGACCGGCTATACACAG
AGCTCACAGAGAATCTAAAGTCATACAATGGGCACCTTACCGACAGAAGATGCACCTCAATGAAAATCGTACCTGTACATGTCAA
GGAATTGATCCAGAGACTTGTGGAGCTTCATTCTCTTTGGCTGTTCATGGAGTATGACTTTAATGGCTGTAAGTTTGGTAGAAG
CCCAAGCCCCAGAAGATTTAGAATTGATCCAAGCTCTCCCTTACATGAAAAAACCTTGAAGATAACTTACAGAGTTTGGCTACAC
GATTAGCTCCAATTTATAAGCAGTATGCTCCAGTAGCTTACCAAAATCAGGTGGAATATGAAAATGTTGCCCGAGAATGTCCGCTT
GGCAGCAAGGAAGGTGCTCCCTTCTCGGGCTCAGTCTTCCCTGGACTTCTGTCTCATCCCCACAGGACCATCACAAATGAA
TAATGGAAGCATGTGGTTTGTACCTTAACTCGAGAAGATAACCGCTCTTTGGGTGTATTCTCCTCAGATGAGCAGTCCATGTGC
TACCTCTTTATAAGCTTTTCAGACACAGATGAGTTTGGCTCCAAGGAAGGAATGGAAGCCAAGATCAAATCTGGGGCCATCGAGGTC
CTGGCACCCCGCGGCAAAAAAGAACGTGTTTCACTCAGCCTGTTCCCGCTTCTGGAAAGAAGAGGGCTGCGATGATGACAGAGGT
TCTTGCACATAAGATAAGGGCAGTGGAAAAGAAACCTATTCGCCAATCAAGCGGAAGAATAACTCAACAACAACAAACAGTA
AGCCTTCGTCACTGCCAACCTTAGGGAGTAACAGTATGAGACCGTGCAACCTGAAGTAAAAAGTGAACCGAACCCCATTTTATCTTA
AAAAGTTTCAGACAACATAAAACCTTATTCGTATGCGCATCCGCTCTCACCAGCTGAAAGAGGCATCTCCAGGCTCTCTCGTGTG
CCGAAGAGCTGCTTCAGCCACACCGATCTCCACTGAAGAATGACGCAACAGCCCTCATCGGGTTTCAGAAAGAAGCAGCATCCG
ACTGTACGATGCCTTCGGGAAGACTCAGTGGTGCCATGCAGCTGCTGCTGATGGCCCTGGCATTTACAGCTTTGGCGAAGTGGCT
CCTCTCCCCACCCTGTCTGCTCTGTGATGGAGCCCTCATTAATTCTGAGCCTTCCACTGGTGTGACTGAGCCGCTAACGCCTCA
TCAGCCAAACACCAGCCCTCTTCTCTCACTCTCCTCAAGACCTTGCTCTTCTCCAATGGAAGAAGATGAGCAGCATTCTGAAG
CAGATGAGCCTCCATCAGACGAACCCCTATCTGATGACCCCTGTCACTGCTGAGGAGAAATTGCCCCACATTGATGAGTATTGG
TCAGACAGTGAACACATCTTTGGATGCAAAATTTGGTGGGTGGCCATCGCACTGCTCAGGCTCGGTTTGGATTGAGTGTGC
CCGGCGAGAGCTGCACGCTTACCACCTCTTGGAGCAGCCCAACCGTAATCATCCAAACCCGCTCTCCCTGTCTTTACAGCACA
AAAACCTAAATTAAGCCCCAACATGGTTTGAAGTAAACAAGATTAACTTTGAGGCTAAAGAAGCTAAGAATAAGAAATGAAGGCC
TCAGAGCAAAAAGACCAGGCAGCTAATGAAGGTCCAGAACAGTCTCTGAAGTAAATGAATTGAACCAAATTCCTTCTCATAAAGC
ATTAACATTAACCCATGACAATGTTGTACCGGTGTCCCTTATGCTCTCACACACGTTGCGGGGCCCTATAACCATTTGGGTCTGAC
ATAGTGACTGGATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCA
TTTTACGTTTCTCGTTCACTTCTTTGTACAAAGTGGTTTCAGAGCTAGCGGCGGCCGACAGGTTCTTACCCTTATGACGTTCTCTGA
TTATGATCCTCGGTGGCTCGAAGTTTGTTCAGGGTCCACTGGGTTCCGAGATCCAACAGCTCCGAAACATCATCACCCCTC
CAGGAACCATTTGTGGCCCCATGGCTGGTTGGTAGTCAAGTCCCTGTGCAAAAACGAAGCCAAAGTAGATCAAGGTCAACCTATA
TTAGTCTTAGAGGCAATGAAGATGGAGCAGTGTGTGAAGCACCATCCTCTGGAAGCATACAGGACCTCAAAGTTAAAGCAGGCCA
ACAGGTTTCAGATGGCAGTGTCTTATTCAGGATCAAAGGGTAATTAATTAAGTACTGTTCTAGAGCGGCCGCCCCACCGCGGTGGAGCT
CGAATTTCCCGGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGGCATGATTATCATATAA
TCTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCC
GCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAACTAGGATAAATATCGCGCGCGGTGTCTATGTTAC
TGAATT

FIG. 7B-04

GAAGTGGCCACCTGCAGCTGTCTTGATCGAGTT
ATACAAAAAGACAAAAGGCCCATATTATACACACCTTGGGGCAGGA
CCAAGTGTGTGCTGCTGTCTCAGGGAAATCATGGAGAATAGGTATGGT
CAAAAAGGAAACCGCAATAAGGATAGAAATAGTAGTGTACACCGGT
AAAGAAGGGGAAAAGCTCTCATGGGTGTCCAATTGCTAAGTGGGT
TTAAGAAGAAGCAGTGATGAAGAAAAGTTCTTTGTTTGGTCCGG
CAGCGTACAGGCCAACCCTGTCCAAGTGTGTGATGGTGGTGTCTC
ATCATGGTGTGGGATGGCATCCCTCTTCCAATGGCCGACCGGCTA
TACACAGAGCTCACAGAGAATCTAAAGTCATACAATGGGCACCCCT
ACCGACAGAAGATGCACCCCTCAATGAAAATCGTACCTGTACATGT
CAAGGAATTGATCCAGAGACTTGTGGAGCTTCATTCTCTTTGGC
TGTTTCATGGAGTATGTACTTTAATGGCTGTAAGTTTGGTAGAAGC
CCAAGCCCCAGAAAGATTTAGAATTGATCCAAGCTCTCCCTTACAT
GAAAAAACCCTTGAAGATAACTTACAGAGTTTGGCTACACGATT
GCTCCAATTTATAAGCAGTATGCTCCAGTAGCTTACCAAATCAG
GTGGAATATGAAAATGTTGCCCGAGAATGTGGCTTGGCAGCAAG
GAAGGTTCGTCCCTTCTCTGGGGTCACTGCTTGCTGGACTTCTGT
GCTCATCCCCACAGGGACATTCACAACATGAATAATGGAAGCACT
GTGGTTTGTACCTTAACTCGAGAAGATAACCGCTCTTTGGGTGTT
ATTCTCAAGATGAGCAGCTCCATGTGCTACCTCTTTATAAGCTT
TCAGACACAGATGAGTTTGGCTCCAAGGAAGGAATGGAAGCCAAG
ATCAAATCTGGGGCCATCGAGGTCTGGCACCCCGCCGCAAAAAA
AGAACGTGTTTCACTCAGCCTGTTCCCGTTCTGGAAAGAAGAGG
GCTGCGATGATGACAGAGGTCTTGACATAAGATAAGGGCAGTG
GAAAAGAAACCTATTCCCCGAATCAAGCGGAAGAATAACTCAACA
ACAACAAACACAGTAAGCCTTCGTCACTGCCAACCTTAGGGAGT
AACACTGAGACCGTGCAACCTGAAGTAAAAAGTGAAACCGAACCC
CATTTTATCTTAAAAAGTTCAGACAACACTAAAACTTATTCGCTG
ATGCCATCCGCTCCTCACCCAGTGAAAGAGGCATCTCCAGGCTTC
TCCTGGTCCCGAAGACTGCTTCAGCCACACCAGCTCCACTGAAG
AATGACGCAACAGCCTCATGCGGGTTTTCAGAAAGAAGCAGCACT
CCCCACTGTACGATGCCTTCGGGAAGACTCAGTGGTGCCAATGCA
GCTGCTGCTGATGGCCCTGGCATTTCACAGCTTGGCGAAGTGGCT
CCTCTCCCCACCCTGTCTGCTCCTGTGATGGAGCCCCCTATTAA
TCTGAGCCTTCCACTGGTGTGACTGAGCCGCTAACGCCCTCATCAG
CCAAACCACCAGCCCTCCTTCCTCACCTCTCCTCAAGACCTTGCC
TCTTCTCCAATGGAAGAAGATGAGCAGCATTCTGAAGCAGATGAG
CCTCCATCAGACGAACCCCTATCTGATGACCCCTGTACCTGCT
GAGGAGAAATTGCCCCACATTGATGAGTATTGGTCAGACAGTGAG
CACATCTTTTGGATGCAATATTGGTGGGTGGCCATCGCACCT
GCTCACGGCTCGGTTTTGATTGAGTGTGCCCGGCGAGAGCTGCAC
GCTACCACTCCTGTTGAGCACCCCAACCGTAATCATCCAACCCGC
CTCTCCCTTGCTCTTTTACCAGCACAAAACCTAAATAAGCCCCAA
CATGGTTTTGAAGTAAACAAGATTAAAGTTTGAGGCTAAAGAAGCT
AAGAATAAGAAAATGAAGGCCTCAGAGCAAAAAGACCAGGCAGCT
AATGAAGTCCAGAACAGTCTCTGAAGTAAATGAATTGAACCAA
ATTCTTCTCATAAAGCATTAACATTAAACCATGACAATGTTGTC
ACCGTGTCCCTTATGCTCTCACACACGTTGCGGGGCCCTATAAC
CATTGGGTCTGA

gggggttttaggtctttccattactttttaaatgttttttctgttactgtctccgcgatctgatttttacgacaatatagagtttcgggttt
tgtcccattccagtttgaaaataaaggctccgtcttttaagtttgcgtgatcgataaacctgtgaagattgagtcagtcgatttat
tggatgatccattcttcacgtgttttttcttgcttcgaagttctgtataaccagatttgtctgtgtgcgatttgtcattacctagcc
gtgtatcgagaactagggtttttcgagtcgaattttgcccccttttggttatactcgtgttcgataacgattcatctggattagggtttt
aagtggtgacgttttagtattccaattttcttcaaaatttagttatggataatgaaaaatcccaattgactgttcaattttcttgtaa
atgcgcagatggctcgtaccaagcaaacgcgtcgtaaagtcacccggagggtaaagctccaaggaagcaacttgctactaagggtttt
ttccttcttgtctcttttttcaaaataacttgtgttgtgaagttgaatgttaatctccttctttattaaacctcaggtcgtcgt
aatctgcaccaactactggtggagtcgaagaaaccacatcgttaccgtcctggaactgttgcctcctccggtttgtcccttcttcgatt
tgtatgtgattctttgagattatgtaacattgtgtgttaacattcctcttatcttttggtgttcagtgaaatccgtaagtaccaga
agagtactgagttgcttatcaggaaactgccatttcagaggctagtcctgtgagattgtcgaagatttcaagactgatttgcgtttc
cagagccatgctgtcttagctctccaggaagctgcagaagcatatcttgttggctctcttgaagacactaaccttttgtgccattca
tgccaagcgtgtgaccataatgcccaaagacattcagctcgcctcgtatcagaggtgaacgcgcttaagccaaccaagaatccg
agatttggttcaagtagtttttgttttttatgaaagcaagatcttaattgctgtgttttttaaattctcgtgggtatgtagtagtaag
agtagtaagaagctcgaactaagaatgttgtgttctttaagcttattattatgtgtgttgcgaactctttttaactctttgtt
aaattcttatgttttcttaactcgtttgttctgcaccattgctttctcttcaaaataatggctgatttttttctataattttggatt
tgttgaatgctttcttttaaaagaatcagatttttgacttttgaccaaagaaaaataataatatcagacgataaaaatagacggctc
cgataaaaactaacctaaaaataaggaaataagtttctctttgaaccaaatttttctttctttgaccaatagatctttttgtcaacc
tcttaaaataattcttagtcaatcttctaataaacccattggccattaccaaaaattcctcggaacgcgtgaataaaaaacattct
atcat

FIG. 8-02

ctcggcaaacgccataaactacagcgctogagtttagcgggtgttggtgggtctggtgaagtgaaccggatttgcctctcagatagc
tatttttagacgggaagcttcattctgttgagaatgttcgagaaagctcaatgcagccttataaagtcattctgggagtggaagaaag
aagatgtagaaaagaaaaaagactgatctttgactagagtagtatctaccagagaatatatatattgcaagcaatgtgatgacgtgt
tcattctataaaaactcaaagctcggtttagttcttttgcctatgagcttttgcagattaaccataaaaatogaacctttatgtttacaag
atccttttgcctccaagtctactggctccggttgaaagtttaaagattagatcagtagattaccttttggccctacgttaaagtttg
tttattttaaatacaagtccttggtgatttttatcactttctcactttctcaacgacactgattcgttttcaaagcatttgcgttctt
gatctttttcacgagggcgatttctgaagaacacgagaatttgaaattctgggaaaagctttctcgaactttatctgagtaaattga
cagagagaatcgaaaaaagaatg

FIG. 8-03

SEQ ID NO:9

GmuCoreIN

GTGCGACTATAAAACGCCCCCTACACCCTCCTCCTTCCTTCGCGAGTTCAATTCCAATATATTCCATTCTCTCTGTGTATTTCCCTACC
TCTCCCTTCAAGGTTAGTCGATTTCTTCTGTTTTCTTCTTCGTTTCTTCCATGAATTGTGTATGTTCTTTGATCAATACGATGTT
GATTTGATTGTGTTTTGTTTGGTTTCATCGATCTTCAATTTTCATAATCAGATTCCAGCTTTTATTATCTTTACAACAACGTCCTTA
ATTTGATGATTCTTTAATCGTAGATTTGCTCTAATTAGAGCTTTTTCATGTCAGATCCCTTTACAACAAGCCTTAATTGTTGATTTC
ATTAATCGTAGATTAGGGCTTTTTTCATTGATTACTTCAGATCCGTTAAACGTAACCATAGATCAGGGCTTTTTTCATGAATTACTT
CAGATCCGTTAAACAACAGCCTTATTTTTTATACTTCTGTGGTTTTTCAAGAAATTGTTTCAGATCCGTTGACAAAAAGCCTTATTC
GTTGATTCTATATCGTTTTTCGAGAGATATTGCTCAGATCTGTTAGCAACTGCCTTGTTTGTGATTCTATTGCCGTGGATTAGGG
TTTTTTTTTACGAGATTGCTTCAGATCCGTACTTAAGATTACGTAATGGATTTTGATTCTGATTTATCTGTGATTGTTGACTCGAC
AGGATCCATG

SEQ ID NO:10

Gmupri

GGGCCCCAATATAACAACGACGTCGTAAACAGATAAAGCGAAGCTTGAAGGTGCATGTGACTCCGTCAGATTACGAAACCGCCAACT
ACCACGCAAAATGCAATTCTCAATTTCTAGAAAGGACTCTCCGAAAATGCATCCAATACCAAATATTACCCGTGTCATAGGCACCA
AGTGACACCATACATGAACACGCGTCACAATATGACTGGAGAAGGGTTCCACACCTTATGCTATAAAACGCCCCACACCCCTCCTC
CTTCCTTCGCGAGTTCAATCCAATATATTCCATTCTCTCTGTGTATTTCCCTACCGCACCCCTTCAAGGATCCATG

SEQ ID NO:11

GmuCore

GTGCGACTATAAAACGCCCCCTACACCCCTCCTCCTTCCTTCGCGAGTTCAATTCCAATATATTCCATTCTATCTGTGTATTTCCCTAC
CGCACCCCTTCAAGGATCCATG

SEQ ID NO:12

ZmRbcs1 Promoter

AGCTCGctccctttaatctggcgctagatctgcatccgcggttgcaaagataaatggcacatntagtggttattttgcaatacc
tttcatagtagatatccctaaatgcagtttttaggcagtggtgggaattaaataaacatttttaggaggagtttagattacctttc
tttctgtgatgactgatacagacgtggggaattcaaatgcaactctagcgaaagttcatatatttttcataaatagctgaggctggg
gtaattattttttgtagaaaaatagaataggtggaatggtttggggaaggcgtaggcgctcgtggacgacgcccataaaagacaa
gaggcggaattgccatgaattcgaggtagctaagtaaggcgcatgcataatatgcaaaaaattctactgtcactttccaatttca
atgcgctgcaaaacaagccatccctggaaactgacttgaattcagcccaattctgtagatccaaacagggccggcgctcagtgcctca
ggtgagagagcagcagacgatgcaaagagccaaaagtggaaagcagacgcagccgaagccgaagcccaagcccaaaactgtttgtc
tttggccagaaccgcgacgagcctaaactgccgcttctcctatctacaagtcctggcacatcacgcatagtccaaccatggcgc
gcaggcgataaaggcgcgccacggggacgcgacatgtggtggcggaagcgatcaggatagggccaggctggccggggcgggccacgg
gagaacgggtggccactcgtcccacatccgcttctgtcctgtactgcgtcctgcccccaacgagagccggagccggccatccc
gtcgcacactctccccctctatatatgcgctcggtgtgggggagcctactacaggacgacccaagcaagcaagcaagcagcgagta
catacatactaggcgagccaggcaGAATT

19/19

FIG. 8-04

SEQ ID NO:13

OsAnt1 Promoter:

AGCTTAGGAAGTGATTTTTAGCGTAGCTGTGTTTGTAGCGTAATTGCGTAAAGTCCTTTCAATTTTGCTATATCTCACTCGAAAGA
TTTTTTCTTATCTCTCACTCGATTTTCTCACTCAAATTTACAGTGTATTTTCTTGTAAGTTACAGTGTAAATTTATGAACTTACAC
TGTAACTTTTGTAAGTTACACTGTAATTTTGAATCTTCACATGTAAATTTTAAATTTTGTATTGGATTGGTCTTTTTCTTGAGG
ATATGGTAATTTAATGTTTCATTATGGTGTTCCTTAATTGCTTTTTGCTTTTATTATATCTATCGGATTTTAATACAAAGATTAAA
AATCTGTGTGATACGATTATAAAAAATCTTTCGAAAGATGTATAGGTACTCCCAAGCCCTTTTAAGAAAGTTTTTCAAGACAAAAGT
TTTTGGATGAAAGGTAGTTATAGGGAAAAAGGAATGTGCGTTTATGTTTATTTGCATTGCTTATTGGCAACCAAAAACTAATCTAT
AAGTAAATCTTTTATATACGTGCGCTTAATAATTCAAAGCAAATTCATGTAAATAAAATGCGATGAAGAACTTTAAAAAGTTA
TCAAATTTAGATTTTATTAAATTTTAGTTTACAAGAGCGCTACGATGAAGGCTTTAAAAAGATGGGAAAAATAAACCTTTGACCTT
TCTGGACTTCACCAAACAGCTCACGCTTTCGGCTTCGTGCCGTCTCGTCCCGTGCTACTGCTACCCCTCCTGACCCACCCGCCA
CTCCACGCTCCCTTCTCCTCCCTTCCCGTGACACACAGTCCCCACTCCACCGCTCCGTATAAGTATCCCTTCTTACCGCCGGC
CAGCCACAGCCACCGCTCCCCACCCACCCCGATCCCTCCCCGCGTACGGGCGCAGAAGGAACCCGTCTTCTAGAAGGAGGA
GGAGGGCTACCTCTCTCTCTCTCTCTTCTGCCGAAT

SEQ ID NO:14

Nos promoter:

AATTCAAGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGACCCCCCGCGATGACGCGGGACAAGCCGTTTTACGTTTGGAAC
GACAGAACCGCAACGTTGAAGGAGCCACTCAGCCGCGGGTTTCTGGAGTTTAATGAGCTAAGCACATACGTCAGAAACCATTTATG
CGCGTTCAAAAGTCGCCTAAGGTCACTATCAGCTAGCAAATATTTCTTGTCAAAAATGCTCCACTGACGTTCCATAAATTCCTTC
GGTATCCAATTAGAGTCTCATATTCACCTCTCAATCCAGATCTCGACTCTAGGGGGATCTACC

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/029354

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A01H 1/02; A01H 4/00; C12N 15/113; C12N 15/82 (2017.01)

CPC - A01H 1/02; C12N 15/113; C12N 15/8241; C12N 15/8261 (2017.05)

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)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/412; 435/6.12; 435/7.1; 800/260 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013/0323220 A1 (THE GENERAL HOSPITAL CORPORATION) 05 December 2013	1-3, 14-20, 34
---	(05.12.2013) entire document	---
Y		4-13, 21-33
Y	WO 2015/054106 A1 (MONSANTO TECHNOLOGY LLC) 16 April 2015 (16.04.2015) entire document	4-13, 21-33
P,X	US 2017/0016017 A1 (FROMM) 19 January 2017 (19.01.2017) entire document	1-34
A	US 2013/0177912 A1 (CHILDREN'S MEDICAL CENTER CORPORATION et al) 11 July 2013 (11.07.2013) entire document	1-34
A	WO 2014/093622 A2 (THE BROAD INSTITUTE, INC. et al) 19 June 2014 (19.06.2014) entire document	1-34

☐ 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

"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

29 July 2017

Date of mailing of the international search report

16 AUG 2017

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774