



(51) International Patent Classification:
C07K 14/435 (2006.01)

(21) International Application Number:
PCT/GB2011/001305

(22) International Filing Date:
6 September 2011 (06.09.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1014818.7 7 September 2010 (07.09.2010) GB

(71) Applicant (for all designated States except US): **MEDICAL RESEARCH COUNCIL** [GB/GB]; 2nd Floor David Phillips Building, Polaris House, North Star Avenue, Swindon SN2 1FL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HACKETT, Jamie, A.** [GB/GB]; MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU (GB). **MEEHAN, Richard, R.** [IE/GB]; MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU (GB).

(74) Agents: **NAISMITH, Robert, Stewart et al.**; Marks & Clerk LLP, 120 Bothwell Street, Glasgow G2 7JS (GB).

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

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

(54) Title: GENE REGULATION

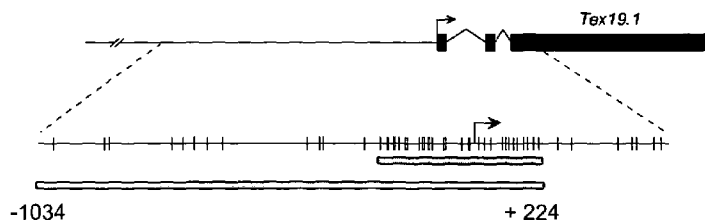


FIGURE 18

(57) Abstract: The present invention provides methods of identifying genes primarily and causally regulated by epigenetic events such as, for example, DNA methylation. Additionally, the invention provides nucleic acid constructs encoding genes and/or promoter sequences identified as primarily and causally regulated by epigenetic events, operably linked to reporter elements. Such constructs find particular application in the studies designed test the effects of agents on epigenetic events.

GENE REGULATION

FIELD OF THE INVENTION

The present invention relates to methods of identifying genes which are primarily and causally regulated by epigenetic events such as, for example, DNA methylation. The invention also provides uses of such genes and transgenic animals.

BACKGROUND OF THE INVENTION

Over 30 years ago it was proposed that DNA methylation may have an important role in controlling gene activity (Holliday and Pugh, 1975; Riggs, 1975). Subsequent work has borne out this hypothesis and demonstrated that DNA methylation is associated with long-term transcriptional silencing through its involvement in biological phenomena such as X chromosome inactivation, transposon silencing and genomic imprinting (Reik, 2007). Inactivation of the maintenance cytosine methyltransferase, *Dnmt1*, can lead to global hypomethylation and activation of genes in developing embryos (Jackson-Grusby et al., 2001; Li et al., 1993). Although the importance of CpG methylation as a stable contributor to transcriptional silencing has been firmly established, the precise role and specific global targets of strictly methylation-dependent gene repression are yet to be fully determined (Sasai and Defossez, 2009).

DNA methylation can affect gene expression by inhibiting transcription factor (TF) binding or alternatively, by acting as a ligand for methyl-CpG binding proteins (MeCPs) which recruit histone modifying complexes to establish a silent chromatin state (Bird, 2002). Whereas inactivation of *Dnmt1* results in an embryonic lethal phenotype, mice simultaneously lacking *Mecp2*, *Mbd2* and *Kaiso* MeCPs undergo normal mouse embryogenesis implying they are not essential in early development and do not significantly contribute to DNA methylation-dependent gene silencing (Martin, I et al., 2009). A possible explanation for the relative viability of mice lacking MeCPs is that there are few (if any) genes that are solely dependent on DNA methylation for their silencing. Instead a model can be envisaged whereby MeCPs reinforce transcriptional repression at promoters that is established initially through chromatin based silencing mechanisms (e.g. Polycomb) and subsequently by CpG methylation as a final repressive lock on transcription (Fouse et al., 2008; Mohn et al., 2008). Genome wide profiling supports this possibility as there is a convergence between the presence of DNA methylation and repressive histone marks at regulatory regions (Mohn et al., 2008). At the same time it is clear that there are sets of gene

Confirmation copy

promoters where histone marks (H3K4me3, H3K9me3 and H3K27me3) are absent but DNA methylation is present (Mikkelsen et al., 2007; Meissner et al., 2008). These are potentially genes that are regulated exclusively by DNA methylation, and perhaps MeCPs, during development and in differentiated tissues.

5 DNA methylation is bimodally distributed at gene promoters, with low-CpG density promoters (LCP) generally being methylated and CpG island (CGI) promoters, which have a high-CpG density, remaining largely unmethylated throughout development (Meissner et al., 2008). The presence of DNA methylation at LCP genes is not correlated with transcriptional repression implying that a high local
10 concentration of methyl-CpGs is required to effectively silence expression (Bird, 2002). In agreement with this, the rare characterised examples of methylated CGI promoters are strongly associated with transcriptional repression (Meissner et al., 2008; Mohn et al., 2008). However, importantly it is unclear if methylated CGI promoters represent a cause or a consequence of gene silencing.

15 SUMMARY OF THE INVENTION

Epigenetic events control gene expression without altering the underlying DNA sequence of an organism. One of the principle mechanisms of epigenetic gene regulation is DNA methylation in which methyl groups are added to CpG sites within DNA sequences. Levels of DNA methylation are correlated with gene expression and
20 low levels of methylation are typically associated with increased gene expression and high levels with reduced expression or gene silencing.

While the precise role of *in vivo* DNA methylation events in gene regulation and control is unknown, the present invention provides a method which allows users to identify genes for which DNA methylation is the primary and causal regulator of
25 expression. In particular, the invention allows users to identify genes for which *in vivo* DNA methylation events are the primary and causal regulator of gene expression.

Accordingly and in a first aspect, the present invention provides a method of identifying genes regulated by DNA methylation, said method comprising the steps of:

30 (a) identifying de-repressed genes in a cell exhibiting global hypomethylation;
(b) identifying de-repressed genes in cells treated with 5-aza deoxycytosine (5-aza dC); and

(c) allowing the cells treated in step (b) to recover in the absence of 5-aza dC and thereafter identifying genes which remain de-repressed;

wherein genes identified as being de-repressed in each of steps (a), (b) and (c) are regulated by DNA methylation.

One of skill will appreciate that step (a) need not be conducted first and the individual steps of the method provided by the first aspect of this invention may be executed in any order. For example, one may first identify de-repressed genes in cells treated with 5-aza dC. Subsequent steps may involve allowing the cells treated with 5-aza dC to recover in the absence of 5-aza-dC and identifying genes which remain de-repressed followed by a further step in which de-repressed genes are identified in globally hypomethylated cells. One of skill will appreciate that while the 5-aza dC treatment step and the subsequent recovery step may be combined (i.e. a cell is first treated with 5-aza dC and de-repressed genes identified and then the cells is allowed to recover in the absence of 5-aza dC and de-repressed genes identified after the recovery period), the 5-aza treatment and recover steps may be executed separately. Where the treatment step is not immediately followed by the 5-aza recovery procedure (as outlined as step (c) of the method above), the recovery step may begin with treatment of cells with 5-aza dC – said cells then being allowed to recover in the absence of 5-aza dC for a period of time.

In one embodiment, the method of identifying genes regulated by DNA methylation, comprises the steps of:

(a) identifying de-repressed genes in a cell exhibiting global hypomethylation;
(b) determining whether or not the genes identified in step (a) are also de-repressed in cells treated with 5-aza deoxycytosine (5-aza dC); and

(c) allowing the cells treated in step (b) to recover in the absence of 5-aza dC and thereafter determining whether or not the genes identified in step (b) remain de-repressed;

wherein genes identified as being de-repressed in step (c) and/or depressed in each of steps (a), (b) and (c), are regulated by DNA methylation.

In a yet further embodiment, the method of identifying genes regulated by DNA methylation, comprises the steps of:

(a) identifying genes de-repressed in cells treated with 5-aza deoxycytosine (5-aza dC);

(b) allowing the cells treated in step (a) to recover in the absence of 5-aza dC and thereafter determining whether or not the genes identified in step (a) remain de-repressed; and

(c) identifying de-repressed genes in a cell exhibiting global hypomethylation; wherein genes identified as being de-repressed in each of steps (a), (b) and (c) are regulated by DNA methylation.

The inventors have designated this method a form of 'epigenetic recovery' analysis, the results of which provide cohorts of genes relying exclusively and causally on DNA methylation events – particularly in vivo DNA methylation events, to regulate their expression. The method according to the first aspect of this invention may be used as a means to determine whether or not a gene or genes is/are regulated by epigenetic events such as DNA methylation. Such a method is particularly useful as, contrary to common thinking, the inventors have discovered that of the genes previously thought to be exclusively regulated by epigenetic events, only a sub-set are actually regulated in this way. As such, the method allows the user to identify genes primarily and causally regulated by, for example, DNA methylation, for potential use in assays for screening compounds for epigenetic dysregulation activity.

The cells referred to in the first aspect of this invention may be eukaryotic cells. Eukaryotic cells may include, for example, mammalian, insect, fungal and/or plant cells. Insofar as the invention concerns mammalian systems and gene methylation therein, the cells used in the methods provided by this invention, may be mammalian cells derived from, for example, primates (including human), rodents (for example mouse, rat, rabbit and guinea pig cells) and ungulates (for example porcine or ruminant (ovine, bovine) cells).

While it is possible to utilise cells obtained from or provided by, a mammalian subject, the method provided by this invention may utilise cells obtained or derived from existing mammalian cell lines. In addition, it should be understood that the cells may be adult type cells derived from a variety of different tissues. Cells of foetal and/or embryonic derivation may also be used.

Foetal and/or embryonic cells, including foetal/embryonic stem cells, may be totipotent or pluripotent and obtained from either established stem cell lines and/or developing embryonic material (perhaps experimentally created blastocysts or surplus foetal material created as part of a fertility program). In other embodiments, stem cells may be obtained through application of cell reprogramming techniques which generate induced pluripotent cells (iPs) from differentiated adult (somatic) cells.

In relation to the methods described herein and in particular the global hypomethylation step (i.e. the step in which de-repressed genes are identified in

globally hypomethylated cells), global (cell) hypomethylation may be induced by engineering the cell to lack one or more genes associated with the regulation of DNA methylation. Additionally or alternatively, the cell may be modified so as to comprise one or more mutated genes, wherein said mutated genes are associated with the regulation of methylation. By way of example, a cell for use in the methods described above and exhibiting global hypomethylation, may comprise a mutated version of a gene encoding an enzyme involved in DNA methylation such as, for example, the gene encoding DNA (cytosine-5)-methyltransferase 1 (Dnmt1); in other embodiments, the cell may be manipulated such that it lacks a copy of a gene encoding an enzyme associated with DNA methylation.

Cells comprising a mutated Dnmt1 gene or lacking a copy of that gene exhibit global hypomethylation and may be referred to as Dnmt1 null cells (Dnmt1⁻) cells.

One of skill will appreciate that global hypomethylation in cells may be induced through a variety of other means including, for example the modulation of LSH (PASG, HELLS); a SNF2 family-related putative chromatin-remodelling ATPase that is required to maintain global DNA methylation and for selective methylation at multiple loci (including pluripotency genes but not germ cell specific genes) in somatic cells and satellite DNA (Xi et al., 2009). Additionally, or alternatively, chemicals capable of removing methyl groups from DNA, such as, for example, 5-azacitidine may also be used.

One of skill will appreciate that the mutation and/or ablation of genes involved in DNA methylation may induce less desirable events, including apoptosis. To avoid such problems, cells for use in the methods provided by this invention may be further modified so as to suppress or reduce the occurrence of processes such as apoptosis. For example, the cell may be further modified so as to reduce or ablate the function of genes involved in or associated with, apoptosis. More specifically, the cell may be modified to exhibit reduced or ablated p53 expression. In one embodiment, the cells used in step (a) comprise a modified (for example mutated) *Trp53* gene or may lack the *Trp53* gene. Such cells may be known as *Trp53* null (i.e. *Trp53*⁻) cells.

In a particular embodiment, the globally hypomethylated cells for are mouse embryonic fibroblast (MEF) Dnmt1⁻ *Trp53*⁻ double null cells. Cells of this type are hypomethylated and do not undergo p53 mediated apoptosis.

In order to determine whether or not genes are de-repressed upon global hypomethylation, the expression of genes in any one of the cell types described herein

may be compared with the expression of the same genes in a normal/control cell (i.e. a cell of the same type but which has not been modified to exhibit global hypomethylation). Genes which appear upregulated relative to expression observed in the normal/control cell may be considered de-repressed upon induction of global
5 hypomethylation.

One of skill will appreciate that any degree of gene upregulation may represent de-repression however genes exhibiting approximately 2-fold greater expression in a globally hypomethylated cell as compared with expression observed in a normal/control cell, may be regarded as genes which are de-repressed upon global
10 hypomethylation. In other embodiments, upregulation greater than about 3-fold, 4-fold, 5-fold, 6-fold, 7-fold or 8-fold, as compared with expression observed in a normal/control cell, may be regarded as indicative of a de-repressed gene.

Levels of gene expression can be determined using, Northern blot or cDNA based procedures. In other cases, techniques comprising polymerase chain reaction
15 (PCR) procedures may be particularly useful. Such techniques may include quantitative (or real-time) PCR and reverse transcription polymerase chain reaction (RT-PCR); a variant of polymerase chain reaction (PCR). In RT-PCR an RNA strand is first reverse transcribed into its DNA complement (complementary DNA, or cDNA) using the enzyme reverse transcriptase, and the resulting cDNA is amplified
20 using traditional or real-time PCR. One of skill will be familiar with these techniques and further information may be found in Sambrook, MacCallum & Russell, *Molecular Cloning: A Laboratory Manual* (3rd ed.), Cold Spring Harbor Laboratory Press, 2001 (said reference being incorporated herein by reference).

As stated, in order to identify genes regulated by DNA methylation, de-repressed genes identified in globally hypomethylated cells must also be shown to be
25 de-repressed following treatment with 5-aza dC – this part of the procedure may be known as the 5-aza dC treatment step.

Cells to be treated with 5-aza dC may be eukaryotic as described above. Advantageously, the cells treated with 5-aza dC are the same type of cell used in the
30 global hypomethylation step. For example, where Dmmt1⁻¹ MEF cells are used to identify de-repressed genes in globally hypomethylated cells, the step of treating cells with 5-aza dC may use primary MEF cells. Furthermore, where ablation of a gene associated with apoptosis (for example *Trp53*) is used as a means of overcoming (p53) induced apoptosis, the cells treated with 5-aza dC may also be modified so as to

exhibit reduced or ablated p53 expression. In some embodiments, the cells may be Dmmt1/Trp53 null cells (i.e. Dmmt1^{-/-}Trp53^{-/-}).

Treatment of cells with 5-aza dC may comprise contacting cells with 5-aza dC for a predetermined length of time and then assessing levels of gene expression. As stated, genes which exhibit a level of expression greater (for example, at least about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold or 8-fold higher) than levels of expression identified in normal or control cells (for example the control cells used to compare results obtained from globally hypomethylated cells) which have not been treated with 5-aza dC, may be regarded as genes which are de-repressed. In one embodiment, 5-aza dC may be added for at least about 1, 2, 4, 8, 12, 24, 26, 48 or 72 hours before levels of gene expression are assessed.

As above, levels of gene expression can be determined using, Northern blot, Southern blot or polymerase chain reaction (PCR) based procedures and further information on each of these techniques may be found in Sambrook, MacCallum & Russell, *Molecular Cloning: A Laboratory Manual* (3rd ed.), Cold Spring Harbor Laboratory Press, 2001 (said reference being incorporated herein by reference).

In order to identify genes which are regulated by DNA methylation, genes identified as de-repressed in cells treated with 5-aza dC (and which may previously have been shown to be de-repressed in globally hypomethylated cells), must also be shown to remain de-repressed following a period of recovery in the absence of 5-aza dC – otherwise known as the recovery step.

In the recovery step, cells subjected to treatment with 5-aza dC are allowed to recover from said treatment (perhaps by withdrawing or removing 5-aza dC from the cells) and are incubated for a period of time, advantageously a predetermined period of time, in the absence of 5-aza dC. After said period of time, genes that remain de-repressed are identified.

Advantageously, the period of recovery in the absence of 5-aza dC may extend from about 1 to about 30 days, from about 5 to about 25 days, from about 10 to about 20 days and preferably at least about 14 to 15 days.

Genes which are shown to be de-repressed following execution of each of the global hypomethylation, 5-aza dC treatment and recovery steps, may be considered as regulated by DNA methylation.

In one embodiment, the method of identifying genes regulated by DNA methylation, may comprise the steps of:

- (a) identifying de-repressed genes in a *Dnmt1*⁻¹ *Trp53*⁻¹ MEF cell;
- (b) identifying de-repressed genes in a primary MEF cell and/or a *Trp53*⁻¹ MEF cell, treated with 5-aza deoxycytosine (5-aza dC); and
- (c) allowing the cells treated in step (b) to recover in the absence of 5-aza dC
- 5 for at least about 14 days and determining whether or not the genes identified in step (b) remain de-repressed;

wherein genes which are identified as de-repressed in each of steps (a), (b) and step (c), may be regulated by DNA methylation.

In a further embodiment, the methods described herein may comprise an

10 additional step in which genes identified as being primarily and causally regulated by epigenetic events such as DNA methylation, are further subjected to analysis in cells which lack de novo methylation activity. For example, useful cells may be modified to lack the DNA (cytosine-5-)-methyltransferase 3 beta (*Dnmt3b*) enzyme. Cells modified to lack this enzyme may be any of the cells described herein including, for

15 example mouse embryonic fibroblasts. Cells which lack *Dnmt3b* expression may be known as *Dnmt3b* ^{-/-} cells. Genes identified by the methods described herein and shown to be activated, induced and/or expressed in *Dnmt3b* ^{-/-} cells, may be concluded as regulated by DNA methylation.

Following execution of the methods described herein, the inventors have

20 determined that the genes identified in Figure 4 are genes for which DNA (promoter) methylation is the primary and causal regulator of expression. Specifically, the inventors have identified the following genes for which DNA (promoter) methylation is the primary and causal regulator of expression:

- (i) *Piwi12* (ENSMUSG00000033644)
- 25 (ii) *Tex19.1* (ENSMUSG00000039329)
- (iii) *Mov10l1*: ENSMUSG00000015365
- (iv) *Dazl* (ENSMUSG00000010592)
- (v) *Tex19.2* (ENSMUSG00000039337)
- (vi) *Tex13* (ENSMUSG00000042386)

(vii) Taf7l (ENSMUSG00000009596)

The inventors have designated the seven genes listed above, "Rec-UP" genes.

Intriguingly both *Tex19.1* and *PiwiL2* are critically dependent on *de novo* methylation catalysed by Dnmt3b but are not associated with repressive histone
 5 modifications. Remarkably, only *Tex19.1* from this methylation-dependent cohort relies on a methyl-CpG binding protein (Kaiso) to maintain complete repression.

The *Tex19.1* promoter sequence is provided below as SEQ ID NO: 1. The First exon and transcriptional start site is underlined.

SEQ ID NO: 1: Tex19 Promoter region (-2000 - +385bp)

```

10  TTTGGTTTTGTGTTGTTTTTTTTGCTGTTTGTCTGGTTTTTGTGTTTGCATGGGTGCTTGGG
    ACCGATGCCTGAAGAGCAAATGCTTTGCCGACTGAGCTCTCTCCCCAGCTTTTTCAAAGC
    CTCTTGAAAGTGATCCTTATTTTAAAATTCAACTTTCAGCAGTGTGGGCCAAGCTGGTAA
    CAGAAAGAAGGCAGTGAAGAAGCTGTGGGAAGCCGGCAGTGGTAGATTAACTTATTGTTG
    CTGCTGTGAGAAAATATGACCAAAGTCACTTGCAGAAGAAAAATTCTCTTTTGGCTAATG
15  GTTCCAGATGGATAGACCTTAATGGCAGAAGAGGCAACAGGGGGTCAGGACAGGAAGGTG
    AGAGATGACTTCTGTAACCACAAACATGGAGATGAGAGAAGATCGGAGGTAGGTGAGACT
    ATGAACTCTCAAAAGCTAATCACAGGGGCACATCTGTCAGCCGACCCACCTAAAGGTCC
    CCAAAGCAGCCCCACTGACAGAGAAGCAAGTGTTAAAATATCCAGTCCTATGGGGATGGG
    GGACACATCTCAACACCACGAATGTTGACAATTGGACACTAAGGTCAAAGGCAAGAGTC
20  AGGAGAGACATCTGGCTAGTTACAGCTGAGGAGGAGACAGCTGAGTAGGTCAAGAGCACA
    CAGAAACAAGGCTGAAGGTGAGGGTGGAGGGTAATGTGGGATAAAGGATAGAAGACTACA
    CAGGGCCAGTGAAGACTTCCGGGTAAGTGAAGATCGCAACTCCACATAGCAGCTTTAC
    TTATAAAAATCCCAGAGAGCAATGAGATACTCTTTATAAGGTGAGTGGTTAAAGGGATAT
    ATTCTTTTAAATGGAGTATTACTGAACTCTGAAGTGAAACTTGCTTCCTGCTGTCCCAGAG
25  CACCCAGGAAACCAAGCTTCATGATCCTAAGGAAGGAAACCATGGTAGTTTCTGGATGCT
    TGAGGCCAGACGTGACCTTCTGGACAGGGCAAAATGGTTGCAGAAGTCGAGTCCAGGGTC
    TGGGGAGAGCAGGGCCCTGAGGGCAGCCCCGAGGGCCTGTAGAAAGAGTGAAACTGGCCC
    TGTGATGACAGAGGAACGTTGCTACACCAGTATATCCAACCTGGTGGGATGTGCAGAGTG
    GACTGCGATGCTGCTCACTGATTTTAACAAAATATATCTGTATGCTCCAAGGCTGAATTGG
30  AGGGTGGAATACCTTCTATTTTCTGCTTGGTTTTTCTCTGAGTCCCAAACCTCCTTTAAGG
    ATAAACTCTTCAAAGAATTATTTTGGGTTTTGTTTTTGTGTTGTTGTTTTGTAAGA
    TTTGCCTCTGCCTCCCAATTGCTGGGATTAAAGGCATGTGCCACTGCCTGGCTAAGATTT
    ATTTATTTATTTTAAGAGTACTCTGTTGCTCTCTTCAGAGACACCAGAATAGGGCATCGG
    ATCCCATTACAGATGGTTGTGAGCTATCAGGTGATTGCTGCGAATGAACTCGACCTTTGG
35  AAGTGCAGTCAGTGCTCTTAACCCCTGAACCATCTCTCCAGCCCCTACAAAGAATTTCTT
    TAAAGAGTCTGTTTAAAGAACAAAAGACCCATAGACAGGCAACTGTGCTTAGGCTGCAGT
    TGAGAACTGAGCGAGCAGTTTCCCAAGTGTTAAAGAACTTTTACTAAAGAGACAGGGAA
    GAAGGAATCGACTTTTCCAAACAAGATGTGCCGCGGGCCCAACTCTGAGCCCGGGCCTTC
    GAGTCCCTCGAAGGCACAGGGATGGGGACATCGGGCACCGCTAGAGCACACAGGAAGCAC
  
```

ACAGGGCACTCAGGTGGCGCAAGGGAGCAGCGCATGCGCGCTGTAGGGGGCGCCAGGCCG
 CATGTGCGCCTTCAGAGCCAGAGTGACTCAGGGCCAGAGTGACTCAGCACCTTCGCTTCC
 TGCTGCTTTGGGACCTCATGGGAGATATGTAAATGAGCTGGAGCATCGCGGGACTTGGCC
 CTTTATAAGGCCGCGTTCGT CCTCCCTGGTGTTCAGTGCTTAGCGAGGAGAGCCTGGCT
 5 CTGGAGTCCATTAGCTGGCACCTGGGCGCTCAGTCAGGGAGGTAAGGATGGAAAGGGACT
 GGGCCGGGGCTGAGCCGGGCTGGGTGCGGTGGGCCTGGGTGCGGTGGGAGCACGGTGGG
 CAGAGAGGAGATCCTCCTGGCGTAGGCTTCCCGGCCTCTTTGGTTGGTAGCCGGGCGCAA
 GTGTGAGGTGCGGGCCCTTGAGCCCTGAGGACACCTCTGATGTCACCCAGCCTGTCCAGA
 GTGAACAGGGAACCCCGGGGGCCAGAGCCCCAACACCCTCTCCTCAGGGGAGCCCTGCAG
 10 CGGGCCCAGCAGGGTGAGCCAAACCTCATTTTCTCTGTCTTTCTAG

The PiwiL2 promoter sequence is provided below as SEQ ID NO: 2. The First exon and transcriptional start site is underlined.

SEQ ID NO: 2: PiwiL2 Promoter region (-2000 - +340bp)

AATTACTGGGATGATTGATGCTAGTTAGTTGGAGCTAAGAAATTAGTAGTGGTTAAGAAG
 15 ATACCAGCATCACTATGGTGAATCTTCTGAGAAGTGTTTCCTGGGAGCACAGAGAAGCTG
 TGTCTAGAGACACCCATGGTTGTACCTTGAGTTGGAAACCAGACTTGGTAAAGTGTAAG
 AGTCACCCAGGTGGTATTGGTTTTGAACATGAAGATGCTTGGCACTGTGAGAGGCCAGG
 AGAGGCCATTGGTGAAGGTGCAGCCTCAGTGGCAGTTGAAGGCCCCGGTCTGAAGGGTTC
 ATGTAGAGAAGTTGAGGCTTGGCACCATGAAAAGAGTCTATAAGAGGTTATTTGTGAAAG
 20 TGCAGTCCAGTTGTAGCAGAAGACAGCATTGTTTTGGTGATGCCAGTACCATGGGATGGT
 CACCAAGAACAGCAGCAACAGTGGAGTACAGACAGCTGGAGCCTAGAAGACAACGTGTGT
 GCTACAAAGGGCAGGGGTGGAGAAGTGACCCAAGTCCTTGGAGGAGCCTAGAAGATCACG
 AGTCAACCCAGACCTTGGATGGTCAGAGTTTGATTTTGCTTTTGATTGTGACTGTGCTC
 TGATATTTTTCCCTCTTGAAGTAAGAAAGTATTTTAGTGGAGCCACAGTTAAGAGACTT
 25 TGAATTATAAACTTGAATCTCAAAAGAGATTAGATATTTTAAAGGGACTGAAATTTAAC
 ATGTAAGAACTTGTAAAGATTGTGGGACTTTTAAGGTTATTTAGATCTTGGCGATGAATA
 AAGTAAGGGTTGAGGCTTAAAGTGATGTGTTTGTGTGTCAAGTTGACCGGGAGTCAATTG
 TACTGGCTGGTTTTGTGTGTCAACTTGACACAAGCTAGAGTTATCTCAGAGAAAGGTGCC
 TCCCTTGAGGAAATGCCTCCATGAGATCCAGCTGTAAGGAATTTTCTCAATTAGTGATCA
 30 AGCAGGGAGGGCCAGCCCATCTTGGGTGGTGCCATCCCTGGAATGGTAGTCTTGGGTTC
 TATAAGAAAGCAAGCTGAGTAACCAGCTATAAGAAAGTAAACCAGTAAGTAATATCTCTC
 CATGGTCTCTGCATCAGCTCCTGCTTCCTGACCTGCTTGAGTTCTGTCTTGACTTCCTTT
 GGTGATGAAAAGCAATGTGGAAAGTGTAAGCTAAATAAATCCTTTCTCCCAACTTCCT
 TCTTGGTCATGATGTTATTTACCAATTGCACCTTAGGGCCTTTAACCTGGGAGGGCAGAG
 35 CAGCTCACTATCAGCATGAACATATTAGCTAAACTGTTTACAGGAGCAGCCATTGAAAG
 GGATCCAGGGCAACCCCTACGCCTCGAAACTCACTTTAAAATGACTCACATAACTTCTTT
 TTTTGTGTTGTTGGTGGGGTTTTTTTGTGTTTTGTGTTTTGGGCGGGGTGGGGGCAGTTTC
 GAGACAGGGTTTCTCTGTGTAGCCCTGGCTGCCCTGGAACCTCACTTTGTAGACCAGGCTG
 GTCTCTAACTCAGAAATCTGCCTGCCTCTGCCTCCCGGGATTAAAGGCGTGCCACCTGTT
 40 TGAGAGCAACCTCCACATAGCTAGACTGCGTGAAAGGGTGAGCCAATGAGGTGAGGGACT
 CTTTGTAGTGGTAGGTTGTTTGGCCAATGAGCTGAGCCCTTGGGCTAGGGATCCGGAAAT

CAGGACCTCGGGCACGAGGCTCATTTCCAAGAGGCAGATGAGGCTGTGTTCTGCCTAGACA
 GGGGGCTTCAGGACCCTGACGACCCTAGGCCAGGTGAGAGGGCGAGACTCAGATGTGGAC
 GGCTAGGAGGGTGAAAGGAGCTGTGGTGCAGGCTTGCCGGGCTGTTTTTAAAAAGTTTCG
 GGCTCCTGGTGTGAGCGGGGCGCTCTCAGCCAGGGCCCGCCACCCTTTTGGGTGGCA
 5 TCTGCCCCACGCCCCCTCCCCACGGGCACGAGGTGGCCTGGGCGGGGTCAGGGTGGACTG
 GAATTCGTTATTCCAGGCCG AGTGTGTGGGAGGAACGCAGGGGCTGGAATAGGAGGGAA
AGGAGGTGGCTCCAGGAGAGAGCGAGAGAGGGAGCGCTCGCATCGGGGCTCAGTGGCACC
AGACCTAAAAAGAAATCTAGGCAAGGCTCCGGCACAGTCCACGTGGTGGAAAGGTACTTT
 GCTGTGCGCCGACCTTGGGTGGGATTTGCTTTGACGGACTTGGTCTCAGGTTAGGCCTGAG
 10 GCCTAGTGCCTCTGCCCAGAGGGCACACGTGGAGACCTGGAGAAGCTAGGCCTGGCCTGG
 GACCCTTCACTGCTCGGTGGGTGAGATCTAACCAGGAGAAAACCGCTTGAAGGAAAGA
 C

A number of diseases and/or conditions, including neoplastic or proliferative
 disorders such as cancer and the like, are caused or contributed to, by aberrant
 15 epigenetic events, including aberrant DNA methylation, affecting the expression of
 one or more genes. In some cases, genes or regulatory elements associated therewith,
 may become hypo- or hypermethylated leading to aberrant expression and disease.

Typically, hypomethylation may be associated with increased gene expression
 whereas hypermethylation is associated with gene silencing.

20 Compounds able to modulate epigenetic events may find application in the
 treatment of diseases having an epigenetic aetiology. Such diseases may include
 cancer and therapeutic compounds may be known as epigenetic dysregulators - i.e.
 compounds which modulate epigenetic events.

One of skill will appreciate that genes for which methylation is the primary
 25 and causal regulator of expression may be exploited as a means of studying the role of
 epigenetic events in cell processes and/or in assays to monitor or determine the effect
 of a test agent (for example a candidate drug compound) upon epigenetic events such
 as, for example, DNA methylation. While very few genes are dependent upon DNA
 methylation for silencing, using the methods described herein, the inventors have
 30 identified a subset of germline-specific genes and promoters which rely on promoter
 methylation to initiate and maintain their developmental expression pattern.

Accordingly, a second aspect of this invention provides the use of a gene for
 which methylation is the primary and causal regulator of expression and/or a promoter
 associated therewith, for studying, evaluating and/or monitoring the role of epigenetic
 35 events, in particular DNA methylation, in cell processes.

In a third aspect, the invention provides the use of a gene for which methylation is the primary and causal regulator of expression and/or a promoter associated therewith, for determining or evaluating the effect of a test agent on epigenetic events.

5 In one embodiment, the genes and/or promoters used in the second and/or third aspects of this invention are one or more of the genes and/or promoters described herein (see for example the genes listed in Figure 4). By way of example, the second and/or third aspects of this invention may utilise the Tex19.1 and/or PiwiL2 sequences or promoter sequences described herein.

10 One of skill will readily appreciate that the uses provided by the second and third aspects of this invention may be *in vitro* and/or *in vivo* uses. In one embodiment, an *in vitro* use, may comprise a synthetic system further comprising components of a particular cell process. *In vivo* systems may comprise cell and/or non-human animal based systems. Further discussion regarding suitable *in vitro*
15 and/or *in vivo* systems is given below.

One of skill will appreciate that in order to evaluate, monitor and/or study methylation events, the activity and/or expression of a gene for which methylation is the primary and causal regulator of expression and/or a promoter associated therewith, may be determined or monitored. By way of example, the activity and/or expression
20 of a gene for which methylation is the primary and causal regulator of expression and/or a promoter associated therewith may be determined and/or monitored while an *in vitro* and/or *in vivo* system is subjected to certain conditions and/or contacted with one or more factors. In this way, it may be possible to determine whether or not those conditions and/or factors influence and/or modulate epigenetic events such as DNA
25 methylation. For example if, under certain conditions, the activity of a gene for which methylation is the primary and causal regulator of expression and/or a promoter associated therewith, is found to be modulated (i.e. exhibits and increase and/or decrease in activity and/or expression), one might conclude that those conditions have an effect upon epigenetic events such as DNA methylation.

30 Similarly, if when contacted with a particular factor or factors, the activity of a gene for which methylation is the primary and causal regulator of expression and/or a promoter associated therewith, is found to be modulated (i.e. exhibits and increase and/or decrease in activity and/or expression), one might conclude that that, or those, factor(s) has/have an effect upon epigenetic events such as DNA methylation.

It should be understood that gene and/or promoter activity may be determined via the use of reporter elements (genes) under the control of the promoter and/or polymerase chain reaction (PCR) based techniques such as quantitative (q)-PCR and/or reverse transcriptase (RT)-PCR. Further information on each of these techniques may be obtained from CSH Protocols edited By David Crotty, Cold Spring Harbor Laboratory Press (monthly journal) and The Condensed Protocols From Molecular Cloning: A Laboratory Manual by Sambrook, MacCallum & Russell (2006: CSHLP).

In one embodiment, activity and/or expression of a gene for which methylation is the primary and causal regulator of expression and/or a promoter associated therewith, is determined *in vivo*, in response to one or more imposed or applied condition(s) and/or one or more administered factor(s).

In a further embodiment, the present invention provides use of the Tex19.1 gene, for studying, monitoring and/or evaluating the effect of one or more imposed and/or applied conditions and/or one or more factors, on epigenetic events such as DNA methylation.

In a fourth aspect, the invention provides an assay for studying, evaluating and/or monitoring the role of epigenetic events in cell processes and/or determining, monitoring and/or evaluating the effect of a test agent on epigenetic events, said assay comprising the step of contacting a test agent with a system comprising a nucleic acid encoding an element primarily and causally regulated by an epigenetic event and determining the expression of said element, wherein modulation of expression indicates that the test agent modulates epigenetic events.

As stated an epigenetic event may include DNA methylation and the assay provided by the fourth aspect of this invention provides a means for determining or evaluating the effect of a test agent on DNA methylation.

The assay provided by the fourth aspect of this invention may find application as a screen for compounds potentially useful in the treatment of diseases and/or conditions having an epigenetic aetiology – especially those diseases or conditions caused or contributed to by aberrant DNA methylation.

However, one of skill will appreciate that when treating diseases and/or conditions which do not have an epigenetic aetiology, it may be necessary to ensure that a therapeutic compound or compounds has no effect on epigenetic gene regulation. Accordingly, the assay provided by the fourth aspect of this invention

may find further application as a screen for compounds which have no effect upon epigenetic events such as, for example, DNA methylation.

The system for use in the assay of this invention may comprise an *in vitro* or *in vivo* system. An *in vitro* system may comprise a cell based assay whereas an *in vivo* system may utilise a non-human animal.

A cell based system may comprise a mammalian cell modified to include a nucleic acid encoding an element primarily and causally regulated by an epigenetic event. The cell may be a mammalian cell derived from, for example, a human or rodent. Mammalian cells may be derived from a number of different adult tissues including, for example, the skin, extracellular matrix, muscle, tissues of the cardiac/pulmonary system, tissues of the gut, nerve tissues, genitourinary tissue (including ovarian and testicular tissues), tissue of the respiratory system and/or renal tissue. In other embodiments, the cell may be an adult stem cell or an embryonic stem cell. Adult stem cells may be derived from a variety of tissues and/or structures including, for example, the skin, gut, tooth and/or hair follicles and embryonic stem cells may be derived from early stage and/or developing embryos (or embryos created as part of a fertility protocol or experiment). However, in order to avoid the use and destruction of embryonic material, embryonic stem cells may be obtained from existing cell lines and/or through re-programming procedures yielding so-called induced pluripotent stem cells (iPs).

In a further embodiment, the system may comprise a non-human animal modified or engineered to comprise exogenous nucleic acid encoding an element primarily and causally regulated by an epigenetic event – such non-human animals may otherwise be referred to as transgenic non-human animals, the exogenous nucleic acid being a transgene. A non-human animal for use in the assay provided by the fourth aspect of this invention may also be referred to as a non-human animal model.

Exemplary animals may include mammals, particularly mammals belonging to the order Rodentia, especially mice, rats, guinea pigs and rabbits. However, one of skill in this field will appreciate that other non-human transgenic animals for use in the assays provided by this invention may include ungulates, such as ovine, bovine and/or porcine species.

In one embodiment, the element encoded by the nucleic acid comprises a genetic promoter activated/silenced by methylation. Typically such promoters are

CpG-island promoters associated with genes which are themselves primarily and causally regulated by methylation. Exemplary genetic promoters for use in this invention include promoters associated with any of the Rec-UP genes described herein.

- 5 In one embodiment, the genetic promoter is the HCP *Tex19.1* promoter which regulates the *Tex19.1* gene. Additionally and/or alternatively, the genetic promoter is the *PiwiL2* promoter which regulates the *PiwiL2* gene encoding a member of the Argonaute family of proteins.

- 10 Epigenetic events such as DNA methylation may contribute to the activity of, for example, genetic promoters. In particular, the level of promoter activity may be directly linked to the level of methylation and low levels of methylation may result in increased promoter activity while high levels of methylation may lead to reduced activity and/or promoter silencing. As such, where the assay of this invention exploits a nucleic acid encoding a genetic promoter, a test agent which modulates epigenetic
15 activity may be detected by a corresponding modulation of promoter activity.

- While (relative) promoter activity may be directly determined (see PCR based methods described above), it is perhaps most common to determine promoter activity indirectly by determining the expression of a reporter gene operably linked to, or under the control of, the promoter. As such, the element encoded by the nucleic acid,
20 may further comprise a reporter element operably linked to, or under the control of, the genetic promoter described above. In this way, expression of the reporter gene can be used to determine or evaluate the extent to which a test agent modulates epigenetic events occurring at or within the genetic promoter.

- A reporter gene may be regarded as any gene which yields a detectable
25 product – for example mRNA and/or protein. Suitable reporter genes for use in this invention may include, for example, genes which encode optically detectable (for example fluorescent or luminescent) chromophores and/or moieties detectable by mass. Useful reporter genes may include *GFP* (encoding green fluorescent protein), luciferase genes, *dsred* (encoding red fluorescent protein). Further information
30 regarding suitable reporter genes may be obtained from Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques, (2nd edition), by Carey, Peterson, & Smale, CSHL Press, 2008; The Condensed Protocols From Molecular Cloning: A Laboratory Manual Sambrook & MacCallum, CSHL Press, 2006; and Molecular

Cloning: A Laboratory Manual (Third Edition) by Sambrook & MacCallum, CSHL Press, 2001; each of which is incorporated herein by reference.

5 In view of the above, this invention also provides nucleic acid constructs comprising a nucleic acid sequence encoding a gene and/or promoter primarily and causally regulated by methylation (identified using, for example, the methods provided by the first aspect of this invention) and a reporter element operably linked thereto. Additionally, the invention provides a cell (for example an isolated or *in vitro* cell) comprising the nucleic acid construct of this invention.

10 One embodiment of this invention provides a nucleic acid construct (and a system, for example a cell based system) comprising a Tex19.1 promoter operably linked to a green fluorescent protein (GFP for example enhanced (e) GFP) gene; wherein the GFP or eGFP element is a reporter element). Advantageously, the Tex19.1 promoter may be methylated and one of skill will appreciate that a construct of this type and a cell comprising the same, may find application in the testing of
15 agents for effects on methylated (silenced) genes. Specifically, a construct comprising a methylated (silenced) Tex19.1 promoter operably linked to GFP or eGFP, may be used to identify agents which de-methylate genes – a de-methylation event manifesting as expression of the GFP or eGFP element from the construct, through activation (expression) of the tex19.1 promoter.

20 An exemplary Tex19.1 promoter – GFP (eGFP) construct is shown in Figure 21. This reporter construct (and others described herein) directs ubiquitous expression when non-methylated but is silenced when methylated. Reporter constructs provided by this invention may be referred to as reporters of perturbation of promoter specific DNA methylation.

25 Additionally, the invention provides animals, for example transgenic animals, harbouring a nucleic acid construct of this invention. Suitable animals are defined in detail above and may be used to test or screen agents for an effect upon DNA methylation.

30 A nucleic acid construct comprising a promoter and a reporter gene operably linked thereto, may be created recombinantly using PCR and cloning techniques. For example, nucleic acid fragments comprising encoding a promoter or a reporter gene to be fused or conjugated there to, may be cloned into a cloning vector (for example a plasmid) and replicated in a suitable system (for example a mammalian or bacterial

cell). The replicated vector may then be harvested and subjected to a restriction protocol to excise the nucleic acid encoding the promoter and/or reporter gene.

Without wishing to be bound by theory, one of skill will readily understand that test agents which modulate epigenetic events such as, for example, DNA
5 methylation, may increase the level of DNA methylation at the CpG sites of a genetic promoter. This in turn may lead to a readily detectable modulation (an increase or decrease) in reporter gene expression.

Of course, in the absence of an agent capable of modulating epigenetic events such as DNA methylation, the genetic promoter of the element encoded by the nucleic
10 acid of the system, may exhibit some degree of activity – referred to hereinafter as a baseline level of activity. This will in turn lead to a baseline level of reporter gene expression.

When the assay provided by the fourth aspect of this invention is used to screen a test agent which is also an epigenetic dysregulator, the activity of the
15 promoter may be modulated by the test agent. Where a test agent is able to increase levels of DNA methylation, this may cause a reduced level of promoter activity (or promoter silence) as compared to the baseline level of promoter activity. Conversely, where a test agent decreases levels of DNA methylation, this may result in increased promoter activity as compared to the baseline level of promoter activity. The
20 modulation in promoter activity by the test agent may in turn result in a modulation of reporter gene expression as compared to the baseline level of reporter gene expression.

In one embodiment, the present invention comprises an assay for determining or evaluating the effect of a test agent on epigenetic events, said assay comprising the
25 step of contacting a test agent with a mammalian cell comprising a nucleic acid encoding a reporter gene under the control of a genetic promoter primarily and causally regulated by an epigenetic event and determining the expression of said reporter gene, wherein modulation of reporter gene expression indicates that the test agent modulates epigenetic events.

30 In a further embodiment, the assay for determining or evaluating the effect of a test agent on epigenetic events, comprises the step of contacting a test agent with a system comprising a nucleic acid encoding a reporter gene under the control of the *Tex19* promoter and determining the expression of said reporter gene, wherein

modulation of reporter gene expression indicates that the test agent modulates epigenetic events.

One of skill will appreciate that the results of any of the assays described herein may be compared to the results of a control method in which the system has not been contacted with a test agent or has been contacted with an agent known to have no effect upon epigenetic events such as, for example, DNA methylation. Such a system may be referred to as a “control system”.

Again, without being bound by theory, it is suggested that a test agent which reduces reporter gene expression to a level below the baseline level of expression, may exert its effect by increasing DNA methylation at CpG elements within the promoter controlling expression of the reporter gene. In such cases, the test agent may be regarded as a modulator of epigenetic events. Where a test agent induces reporter gene expression to a level above the baseline level of expression, the test agent may decrease DNA methylation within the promoter controlling reporter gene expression – test agents which induce increased reporter gene expression (as compared to baseline expression) may also be regarded as modulators of epigenetic events.

Test agents which do not induce modulation of reporter gene expression may be regarded as having no effect on epigenetic events.

Where the system comprises a cell-based system, the test agent may be contacted to the system by any suitable means including, for example, by supplementation of the cell culture medium with an appropriate concentration of the test agent, transfection and/or direct injection of the test agent into a cell or cells.

Transfection protocols are most suitable for test agents which comprise nucleic acids such as for example, antisense oligonucleotides and/or iRNA or siRNA molecules. Transfection of cells with exogenous nucleic acid may comprise chemical transfection (perhaps using calcium phosphate), electroporation and/or heat-shock.

Where the system is an *in vivo*, non-human animal based system, a test agent may be administered orally or parenterally to said non-human animal. Parenteral administration may include intravenous, intramuscular, subcutaneous, intracapsular, intraperitoneal, intrarectal, intranasal and/or transdermal (through the skin using, for example, a skin patch or transdermal iontophoresis) administration. In other embodiments, the test agent may be formulated for administration to a non-human transgenic animal via a nasal spray or inhalant. However, one of skill will appreciate

that the route of administration of a particular test agent may depend, in part, on the chemical nature of the agent, and can be determined by those skilled in the art. One or more test agents may be administered together or separately via the same or different routes of administration. Furthermore, test agent can be combined with a pharmaceutically acceptable excipient, carrier or diluent and/or formulated as a tablet, a solution or in suspension form, as appropriate.

A number of techniques may be used to modify a cell or non-human animal such that it expresses a nucleic acid encoding, for example, a promoter regulated by DNA methylation and a reporter gene operably linked thereto. By way of example, the nucleic acid encoding the promoter and/or reporter gene may be cloned into an expression vector. The expression vector (comprising, for example, a nucleic acid encoding a promoter and an operably linked reporter gene) may then be transfected (using any of the protocols described herein) into a host cell. Where the assay is a cell based system, the host cell may be any of the cell types described herein. Where the assay utilises a non-human animal model, the cell may be an appropriate germ-line cell – one of skill will appreciate that modification of a germ-line cell with a nucleic acid encoding, for example a promoter and a reporter gene operably linked thereto, may lead to an adult non-human animal which expresses in one or more adult tissues or cell types, the promoter/reporter gene construct.

In a fifth aspect, the present invention provides a transgenic animal comprising a nucleic acid encoding an element primarily and causally regulated by an epigenetic event. In one embodiment, the transgenic animal encodes a promoter regulated by methylation and a reporter gene under control of (or operably linked to) said promoter.

Advantageously, the transgenic animal is a non-human animal, typically of the Rodentia order as described above. In a further embodiment, the transgenic animal comprises a nucleic acid encoding the *Tex19* promoter and a reporter gene operably linked to (or under control of) said promoter.

In a further (sixth) aspect, the present invention provides an assay comprising one or more immobilised nucleic acid fragments (or probes) representative of one or more of the genes (and/or promoters associated therewith) described herein (or identified by the methods provided by this invention). Additionally or alternatively, the assay may comprise nucleic acid fragments (or probes) comprising sequences complementary to sequences encoding the genes provided by this invention or

identified by methods described herein. In one embodiment, the assay may take the form of a gene “chip” in which the one or more nucleic acid fragments are immobilised at discrete points or addresses on a substrate. Advantageously, the assay may take the form of an array or microarray. One of skill will appreciate that an assay of this type may be useful as a means of determining whether or not a gene for which methylation is the primary and causal regulator of expression, is affected by one or more conditions imposed on, or applied to, an in vitro/in vivo system and/or by one or more factor(s) contacted therewith.

For example, following contact with one or more factor(s) (including test agents and/or candidate drugs) or application of one or more conditions to an in vitro/in vivo system, it may be possible to extract nucleic acid (for example mRNA), subject that nucleic acid to further processing techniques such as RT-PCR (described above) and analyse the resulting nucleic acid profile to determine whether or not the activity and/or expression of any gene or genes have been modulated by the factor(s) and/or conditions. By contacting the nucleic acid resulting from RT-PCR procedures with a gene-chip according to the sixth aspect of this invention – said gene chip comprising fragments of nucleic acid complementary to sequences of any of the genes described herein (or identified by methods of this invention), it may be possible to determine which, if any of the genes primarily and causally regulated by DNA methylation, exhibit modulated activity and/or expression. One of skill will appreciate that during RT-PCR procedures, amplified nucleic acid fragments may be labelled with fluorescent probes/moieties to facilitate detection when hybridised to probes immobilised on the “gene-chip”.

DETAILED DESCRIPTION

The present invention will now be described in detail and with reference to the following Figures which show

Figure 1. Transient exposure to 5aza-dC selectively activates genes regulated by DNA methylation. (A) Schematic of 5aza-dC-recovery assay. (B) Venn-diagram showing overlap between the Aza-Up (blue) and Rec-Up (grey) gene sets. Most genes are re-silenced during recovery. (C) Percentage of genes with somatically methylated CpG island (CGI) promoters upregulated by 5-azadC (Aza-Up) or after 14 days recovery (Rec-Up). (D) Gene ontology analysis of the Aza-Up (blue) and Rec-Up (grey) gene sets at hierarchical level BP2. Only germline associated ontologies remain significantly upregulated in Rec-Up. (E) RT-PCR of germline and somatic

genes in NIH/3T3 fibroblasts (con) exposed to 5aza-dC (aza) and after 3, 8 and 14 days cellular recovery. *Gapdh* is a loading control. (F) RT-PCR showing germline-specific genes are mis-expressed in primary MEFs (pMEFs) and *p53*^{-/-} MEFs (P) treated with only 5aza-dC (48 and 96 hour exposure) or 5aza-dC and TSA (HDAC inhibitor) and in *Dnmt1*^{-/-} MEFs (DP). (G) Venn diagram cross-referencing all three complementary experiments (Rec-Up, D1-Up, P-aza) to identify a minimal gene set that is regulated directly by DNA methylation. Not to scale. Overlapping genes (17) are strongly enriched in germ-line associated genes ($p = 0.0005$) at BP2. *** = $p < 0.001$.

Figure 2. Increasing the fold-change threshold stringency after 5aza-dC exposure enriches for methylated CGI promoters. (A) Scatter plot showing that increasing the fold-change threshold at which genes are considered significantly upregulated after 5aza-dC treatment (NIH/3T3) is strongly correlated with an enrichment of methylated CGI promoter genes. This indicates that increasing the threshold stringency (i.e. taking the most upregulated genes) selects for normally somatically methylated CGI genes, which are promising methylation dependent candidates. The black point indicates CGI promoter methylation genome-wide (i.e. 1-fold). (B) Volcano plot demonstrating the relationship between expression fold-change among CGI genes and promoter methylation after transient exposure to 5aza-dC in NIH/3T3 cells (left) and 14 days cellular recovery (right). Note upregulated genes exhibit no significant enrichment of highly methylated promoters immediately after 5aza-dC exposure (left), indicating the drug primarily affects gene expression through indirect mechanisms. In contrast the recovery set demonstrates there is preferential recovery of indirect (nonmethylated) targets but continual expression of methylation-dependent genes (methylated CGI genes – top right of each graph).

Figure 3. Only germline-specific genes are continually activated by 5aza-dC exposure. (A) Bisulphite methylation analysis of *Tex19.1* in NIH/3T3 cells before (control) and after (aza) 5aza-dC exposure and following a 14 day recovery period under normal culture conditions (aza-rec). Cells that retain global DNA methylation after 5aza-dC exposure may have a selective growth advantage leading to a gradual titration of demethylated cells over generations. (B) RT-PCR of germline-specific genes in *p53*^{-/-} MEFs after 5aza-dC treatment and at 3, 8, 14 days after drug withdrawal (recovery). *Krt8* and *Wnt4* are somatic controls that re-impose transcriptional silencing during the recovery period. (C) RT-PCR showing germline-

specific genes (*Tex19.1* and *PiwiL2*) are activated by 5aza-dC but not TSA in primary MEFs (pMEF) and p53^{-/-} MEFs (P). Hypomethylated *Dnmt1*^{-/-}; p53^{-/-} (DP) MEFs also exhibit *Tex19.1* and *PiwiL2* activation. (D) Semi-quantitative RT-PCR of pluripotent-associated genes. The black triangle indicates increasing concentrations of input cDNA. Primary MEFs and ES cells are negative and positive controls respectively. Number of cycles is shown on right hand panel. -RT; no reverse transcriptase control.

Figure 4. Candidate methylation-dependent genes derived from cross-referencing complementary experimental approaches. Table of genes that were upregulated >6-fold in all three experimental approaches (aza-recovery, *Dnmt1*^{-/-}; p53^{-/-} Aza MEFs), which represent the final methylation-dependent candidate genes. Shown is the gene name, expression fold-change in each assay and primary tissue of expression. N/D: not determined.

Figure 5. Germline CGI promoters are methylated in somatic cells and drive transcription independently of germline-restricted transcription factors. (A) Bisulphite sequencing demonstrates that the *Tex19.1*, *Tex13* and *PiwiL2* CGI promoters are hypermethylated in primary MEFs. Filled circles - methyl-CpGs and empty circles - non-methylated CpGs. Each horizontal clone represents a unique genomic allele (B) *Tex19.1* CGI promoter has tissue specific methylation patterns; it is hypermethylated in non-expressing brain and spleen but hypomethylated in expressing sperm or purified E13.5 germ cells. (C) The non-methylated *Tex19.1*, *Tex13* and *PiwiL2* promoters drive strong luciferase reporter expression after transfection into somatic cells (Neura2a and 293T cells). These promoters are not regulated by tissue-specific transcription factors. *Rnh2*, *Tex11* and *Oct4* are negative control reporter constructs. (D) The methylated *Tex19.1* and *PiwiL2* promoter-reporter constructs are strongly repressed relative to unmethylated controls in 293T cells. Shown are normalised luciferase values from three independent experiments +/- S.E.M.

Figure 6. *Tex19.1* is hypomethylated in DP cells and is repressed by promoter specific CpG methylation. (A) Bisulphite sequencing of *Tex19.1* in control P MEFs and hypomethylated DP MEFs. (B) Promoter specific methylation of different portions of the *Tex19.1* promoter. The indicated regions were digested out of the vector, *in vitro* CpG methylated, re-ligated back into the luciferase vector (pGL3-basic), linearised and size separated to identify single-copy insertions. ~5ng was used per transfection which imposes a severe limit on the potential fold repression.

Observed relative repressions are not quantitative but are proof of principle. The difference in fold-repression between the -224 to +58 construct and the larger *Tex19.1* reporters can be attributed to the presence of a functional repressor motif within the +58 to +198 sequence, which represses transcription >7-fold below vector only in isolation, when methylated or unmethylated.

Figure 7. Promoter CpG methylation causally regulates *Tex19.1* and *PiwiL2* expression. (A) *Tex19.1* and *PiwiL2* expression is silenced in embryoid body (EB) differentiated (2, 4, 7 and 10 days) ES cells. *Oct4* expression monitors differentiation and *Gapdh* is a loading control. -RT, no reverse transcriptase control. (B) The *Tex19.1* and *PiwiL2* promoters are progressively methylated coincident with silencing in differentiating ES cells. (C) *de novo* methylation of the *Tex19.1* and *PiwiL2* promoters (y-axis right, blue line) occurs in parallel with gene silencing (y-axis left, grey line) in differentiating EBs (X-axis, days). (D) RT-PCR showing 5aza-dC can reactivate germline associated genes but not pluripotent genes (*Oct4* and *Nanog*), in EBs. (E) ES cells lacking *de novo* methyltransferases (*Dnmt[3a, 3b]*^{-/-}) fail to fully silence *Tex19.1* and *PiwiL2* expression during EB formation. (F) *Dnmt3b* but not *Dnmt3a* is required to silence *Tex19.1* and *PiwiL2* expression in EBs.

Figure 8. *Dnmt3b* specifically targets *de novo* methylation and gene silencing to *Tex19.1*. (A) Quantitative RT-PCR of *Tex19.1* expression from WT J1 and mutant ES cells during embryoid body (EB) differentiation. WT and *Dnmt3a*^{-/-} ES cells can impose strong silencing on *Tex19.1* but *Dnmt3b*^{-/-} ES cells are unable to silence expression. (B) Bisulphite methylation analysis of *Tex19.1* before (Day 0) and after (day 10) EB differentiation of the indicated genotype ES cells. In the absence of *Dnmt3b*, *Tex19.1* fails to acquire *de novo* methylation and gene silencing. (C) RT-PCR demonstrating that *Tex19.1* silencing fails in *Dnmt3b*^{-/-} and *Dnmt[3a, 3b]*^{-/-} ES cells differentiated with retinoic acid (RA) relative to mock (M) treated.

Figure 9. Silences *Tex19.1* is not marked by histone modifications and is methylated in the absence of polycomb complexes. (A) Schematic of *Tex19.1* CpG structure and primer locations. Blue rectangles indicate *Tex19.1* exonic structure. Below is the CpG distribution with each vertical line representing a CpG dinucleotide. The *Tex19.1* promoter CpG island is shown in red. Shown as black bars are the regions (Up & Pr) that are amplified for ChIP analysis. The bisulphite amplified region (grey bar) is shown above the CpG map. (B) Native-ChiP for the indicated histone mark/variant at the promoter proximal (*Tex19-Pr*) and upstream promoter

region (*Tex19-Up*) of the *Tex19.1* promoter is nonexpressing P cells and expressing DP cells. Inactive *Oct4* and active *B-actin* are controls. Importantly, *Tex19.1* has no significant enrichment of any marks when it is repressed (in P MEFs) indicating histone marks do not significantly contribute to *Tex19.1* silencing. Compare enrichment of the repressive mark H3K27me3 at *Tex19.1* and *Oct4* in P MEFs. The enrichment of H3K4me2 at *Tex19.1* in DP cells is likely a downstream consequence of demethylation, which may directly occlude H3K4me3 thereby maintaining gene silencing. (C) The *Tex19.1* promoter can still acquire *de novo* methylation (and silencing) in the absence of the Eed, an essential component of the PRC2 complex.

Figure 10. Germline-specific genes are depleted in histone modifications and rely on Dnmt3b mediated *de novo* methylation for silencing. (A) *Eed*^{-/-} ES cells silence *Tex19.1* similarly to wild-type cells upon retinoic acid (RA) differentiation. *Oct4* expression is a differentiation and *Gapdh* a loading control. (B) Testis-specific genes (see text for definition) are highly enriched for CGI promoters lacking histone modifications in MEFs. No modification in blue, H3K4me3 in light grey, H3K27me3 in dark grey and bivalent (H3K4me3 plus H3K27me3) in black. (C) RT-PCR shows that *Tex19.1*, *PiwiL2* and *Dazl* are robustly expressed in *Dnmt3b*^{-/-} MEFs (E12.5 and E13.5) compared to heterozygous cells. (D) Expression array analysis indicates that 27 genes are significantly up regulated in *Dnmt3b*^{-/-} MEFs (D3b-Up). Blue bars indicate the gene is expressed specifically in the germline or placenta. (E) D3b-Up is enriched for methylated CGI gene promoters in MEFs. (F) Bisulphite sequencing indicates that *Tex19.1*, *Dazl* and *PiwiL2* are hypomethylated in *Dnmt3b*^{-/-} MEFs, whereas *Tex13*, which remains silent, has a wild-type pattern of methylation. *** = *p* < 0.001.

Figure 11. Kaiso enhances methylation-dependent silencing of *Tex19.1*. (A) RT-PCR of candidate and germline-specific genes in MeCP mutant fibroblasts reveals only *Tex19.1* is significantly derepressed. DP cells are included as a positive expression control and *Gapdh* as a loading control. (B) The *Tex19.1* promoter is hypermethylated in wild type (WT), *Mbd2*^{-/-}, *Kaiso*^{-/-} and triple-null (*Mbd2*^{-/-}, *Kaiso*^{-/-}, *Mecp2*^{-/-}) (MKO) tail fibroblasts. (C) qRT-PCR analysis of *Tex19.1* expression in wild-type and *Kaiso*^{-/-} fibroblasts transiently transfected and FACS sorted for xKaiso-GFP (GFP+) or absence of xKaiso-GFP (GFP-). (D) qRT-PCR of *Tex19.1* (upper panel) and *Oct4* (lower panel) expression in wild-type and *Kaiso*^{-/-} ES cells during embryoid body (EB) differentiation. Relative expression to undifferentiated ES cells

(day 0, set as 1) is shown and normalised to *Gapdh*. (E) DNA methylation profile of the *Tex19.1* promoter in wild-type and *Kaiso*^{-/-} ES cells during EB formation. (F) ChIP of endogenous Kaiso in wild-type and *Kaiso*-null fibroblasts. Enrichments were determined by qRT-PCR of the *Tex19.1* promoter region from two independent experiments. Error bars are S.E.M. (G) The Kaiso zinc-finger domain (ZF1-3) only binds a methylated *Tex19.1* reporter. ZF1-3 fused with the VP16 activation domain (xZF-VP16) or VP16 alone was transiently co-transfected into 293T cells with either an unmethylated, *HhaI* methylated or *SssI* methylated *Tex19.1* luciferase reporter. xZF-VP16 specifically activates *SssI* methylated *Tex19*, shown as a ratio to VP16 alone. Mock methylated or *HhaI* methylated *Tex19.1* were not activated. Shown are normalised luciferase values from three independent experiments +/- S.E.M. * = $p < 0.05$; *** = $p < 0.001$.

Figure 12. *Tex19.1* is partially depressed in *Kaiso*^{-/-} fibroblasts and is transcribed from the canonical annotated promoter. (A) qRT-PCR of *Tex19.1* expression in mutant MeCP cell lines. DP is shown as a positive control and demonstrates there is only partial derepression of *Tex19.1* in mutant MeCP cell lines, implying that alternative mechanisms apart from Kaiso must contribute to methylation-mediated silencing of *Tex19.1*. (B) Mapping of the *Tex19.1* transcriptional start site (TSS) (expected size = 314bp). Sequencing of products indicates that *Tex19.1* expressing DP, *Mbd2*^{-/-} and *Kaiso*^{-/-} cells all transcribe *Tex19.1* from the annotated TSS. In contrast overcycling the 5'RACE PCR to map very low background level *Tex19.1* transcripts from silenced and methylated WT and P MEFs, indicates that these initiate from a non-canonical TSS ~51bp upstream of the annotated TSS.

Figure 13. PGC purity and *Tex19.2* methylation status in PFCs. (A) Staining for endogenous *Oct4* expression in FACS sorted E13.5 germ cells expressing an Oct4-GFP transgene. This analysis suggested E10.5 PGCs were >95% pure (not shown) and E13.5 were >98% pure. (B) Left panel: Methylation analysis of the *Tex19.2* promoter in PGCs and somatic cells at E10.5 and in male and females at E13.5. Right-panel: qRT-PCR analysis of *Tex19.2* expression for the indicated samples. Note that *Tex19.2* shows similar dynamics to *Dazl* whereby the promoter is methylated and silenced in the germline prior to entry in to the gonad (E10.5) but demethylated and transcriptionally activated by E13.5.

Figure 14. DNA methylation-dependent germline-specific genes exhibit two distinct profiles of gene expression during germ cell development. (A) DNA methylation and expression profile of *Dazl* during primordial germ cell (PGC) development (E10.5, E13.5 female and E13.5 male) (top) and matched somatic cells from the testes (bottom). In the expression panel (right) testis derived somatic cells and PGCs were assayed by qRT-PCR relative to 18S RNA. (B) As in A, DNA methylation and expression profile of *Tex19.1* and *PiwiL2* during PGC development. (C) Germline genes fall into one of two distinct categories depending on their expression change in PGCs between E10.5 and E13.5. The first set (left peak) is expressed prior to entry into the gonad at ~E10.5 and therefore exhibit limited or no change in expression. These loci, like *Tex19.1*, would be predicted to be hypomethylated at E10.5. The second set (right peak) is strongly upregulated by E13.5 and like *Dazl*, would be predicted to be hypermethylated at E10.5 and demethylated coincident with expression by E13.5. (D) DNA methylation and expression profile of *Tex19.1* in the embryonic portion of E6.5, E7.5 and E8.5 embryos.

Figure 15. *Tex19.1* protein expression in the post-implantation embryo but silenced by E7.5. (A) *Tex19.1* is hypomethylated (left panels) and expressed (right panel) at E9.5 in PGCs. This confirms that the hypomethylated state of *Tex19* at E10.5 (Fig 4.2) is not due to premature genomic reprogramming. (B) Anti-*Tex19.1* staining (brown precipitate) in E6.5 embryo can be seen in the ectoplacental cone (ec), extraembryonic cell lineages (ex.e) and the epiblast (epi), (C) is IgG control. (D) In E7.5 embryos *Tex19.1* becomes downregulated in epiblast-derived tissues in the embryo with some faint residual staining present in embryonic mesoderm (mes) and extraembryonic mesoderm (ex.m). *Tex19.1* staining is present in extraembryonic tissues (ex.e) at this stage). (E) is IgG control. (F) *Tex19.1* staining is not detectable in embryonic tissues at E8.5. *Tex19.1* staining is absent in the hindgut endoderm where primordial germ cells are located at this stage. Panel shows posterior region of E8.5 embryo. hg, hind-gut; al, allantois. (G) At E9.5 *Tex19.1* staining is only detected in the placenta (pl). No staining was detected in embryonic tissues. (H) 9.5 dpc IgG control. (I) Higher magnification picture of the placenta at E9.5, *Tex19.1* staining is present in the trophoblast and spongiotrophoblast layers. (J) IgG control. Scale bars are 200microns.

Figure 16. Methylated CGI genes are depleted of promoter histone modifications. (A) Scatter graph plotting the upregulated CGI genes from each expression microarray we performed according to whether they are methylated (X-axis) and their histone modification state (Y-axis). We find a strong correlation between the presence of a methylated CGI promoter and the absence of histone modifications among upregulated genes implying that promoter CGI methylation may function exclusively of other epigenetic systems to direct gene silencing.

Figure 17. Dnmt3b target genes are not enriched in E2f6 binding sites. (A) The genes upregulated in E13.5 *Dnmt3b*^{-/-} (null-mutant) MEFs (p1) exhibit no enrichment of the core E2f6 binding site (shown below – exact matches) in their promoter (-1000 - +500bp). In fact, we observed a depletion of core E2f6 binding sites of genes derepressed in *Dnmt3b*^{-/-} MEFs relative to promoters genome-wide. (B) Hidden markov model analysis ($p < 0.0001$) suggests the complete E2f6 binding motif (shown below) is also not significantly enriched among *Dnmt3b*^{-/-} upregulated genes relative to the genome. Importantly both analyses here used the actual occurrence of the core and complete E2f6 binding motifs at genome-wide promoters as the baseline and not the predicted occurrence based on nucleotide probability – which greatly underestimates the true prevalence of E2f6 sites. Our analysis is consistent with a model whereby E2f6 is not a general mechanism for targeting Dnmt3b to target genes but does not rule out some specific loci attracting Dnmt3b through E2f6.

Figure 18. Schematic of Tex19.1 promoter. Rectangles indicate Tex19.1 exonic structure. Below is the CpG distribution with each vertical line representing a CpG dinucleotide. The Tex19.1 promoter CpG island is shown in grey. The region in below corresponds to the region that was used to drive luciferase and eGFP expression in transient assays and in cell lines.

Figure 19. Schematic of Tex19.1 CpG deletion constructs. The region in black corresponds to the region that was used to drive luciferase and eGFP expression in transient assays and in cell lines. Each promoter construct was transiently transfected into 293T cells and reporter expression was normalised to co-transfected renilla luciferase. Shown is expression relative to empty vector (pGL3-b), which is set at 1. Scale is calibrated to log 10. Promoter regions are relative to transcriptional start site (TSS).

Figure 20. Promoter methylation of the *Tex19.1* reporter represses transcription. *Tex19.1* reporters were *in vitro* methylated or mock methylated and

transiently transfected into 293T cells. Shown is relative repression as determined by the ratio of methylated reporter expression (M) to unmethylated reporter expression (U).

Figure 21. *Tex19.1* promoter-*GFP* construct. The *Tex19.1* promoter was cloned into a EGFP-n1 plasmid, in place of the cytomegalovirus (CMV) promoter. In the resulting construct, *Tex19.1* promoter drives the expression of *GFP*.

Figure 22. Flow-diagram illustrating the design of the system aiming to monitor 'demethylating properties' of novel molecules (compound 'x'). The training compound is 5-azadeoxycytidine which inhibits cytosine methyltransferases (DNMTs).

Figure 23. Treatment of *Tex19.1*-GFP-Me NIH3T3 stable cell line with 5-azadeoxycytidine (5-aza dC, 0.5uM). The drug was dissolved in DMSO, which was used as a negative control (A). Three days treatment with 5-aza-dC were enough to induce a considerable *GFP* expression in the cell line analysed (B). Notice the bright fluorescence in B that is absent in A. Endogenous *Tex19.1* expression was also induced after exposure to 5-aza-dC.

Materials and Methods

Cell culture and tissue samples

DNA methylation deficient E9.5 *Dnmt1*^{n/n} *p53*^{-/-} (DP), control *p53*^{-/-} (*Dnmt1*^{n/+}) (P) and primary MEFs were maintained as described (Lande-Diner et al., 2007) as were ES cells (E14/J1) (Ollinger et al., 2008). Embryoid body (EB) differentiation was induced through LIF withdrawal and cell aggregation in hanging drops for 2 days, followed by suspension culture for 5 days and then as adherent colonies. Where indicated retinoic acid (1μM) was used for 72hrs to induce differentiation.

Aza-recovery assay

Low density cells were treated with 1μM 5aza-dC for 3 days, fresh 5aza-dC media was added daily. Cells were subsequently washed and allowed to recover with normal media *in situ* for 5 days and then passaged every 3 days.

Germ cell collection

PGCs were purified from outbred MF1 mice carrying the *Oct4-GFP* transgene. The dissected hind-gut (E9.5) or urogenital ridges (E10.5 & E13.5) from embryos were trypsinized and FACS sorted for GFP using a FACSAriaII SORP (Becton Dickson). Purity of PGCs was determined by Oct4 staining.

DNA constructs

Indicated promoter regions were PCR amplified and directionally cloned into pGL3-basic (Promega) using the *Bgl*III and *Mlu*I sites and used in luciferase reporter assays as described (Dunican et al., 2008).

5 Quantitative RT-PCR

Total RNA from cell lines or tissues was isolated with trizol (Invitrogen). After DNase (Turbo (Ambion)) treatment, cDNA was transcribed with superscript III reverse transcriptase (Invitrogen). RT-PCR was performed with platinum *taq* (Invitrogen) for an optimised number of cycles. Quantitative real-time PCR was
10 carried out with Brilliant II SYBR green qPCR mix (Stratagene) using an iCycler (Biorad). Primer sequences available upon request.

Microarray and bioinformatics analysis

Total RNA was biotinylated and amplified with a TotalPrep RNA amplification kit (Illumina). Pre- and post- labelled RNA was bioanalysed (Agilent)
15 to confirm RNA integrity and hybridized to Illumina mouse Ref-8 v2.0 BeadChips. Background subtraction, normalisations and statistical analyses were carried out using BeadStudio data analysis software modules. Microarray expression changes were calculated as a ratio of normalised expression between samples, with a $p < 0.01$ detection threshold and t-test used to determine significance. Raw DNA methylation
20 sequencing data from pMEFs (Meissner et al., 2008) was mapped to CGI promoters with R statistical analysis software and used to determine CpG methylation levels at target genes. Promoter histone modification states were determined using the same raw data set.

Bisulphite sequencing

25 DNA was extracted from trizol preps or phenol-chloroform extracted from cells after proteinase K digestion. Bisulphite conversion, PCR amplification and sequence analysis was carried out as described (primer sequences obtainable upon request) (Dunican et al., 2008).

ChIP

30 Crosslinked chromatin was prepared and sonicated as described (Dunican et al., 2008). Precleared supernatant was supplemented with 5ug of anti-Kaiso (ab12723) or control IgG antibodies and immunoprecipitated overnight at 4°C with rotation. Protein G slurry was added and washes were carried out. Immunoprecipitated material was eluted twice with elution buffer (0.1M NaHCO₃, 1%

SDS (w/v)) at 30°C and crosslink's reversed at 67°C overnight. Samples were purified and quantitated by qPCR.

Results

5 A cohort of germline-specific genes are derepressed in hypomethylated somatic cells.

We developed an epigenetic disruption and recovery assay to identify candidate genes that are strictly dependent on promoter CpG methylation for silencing in somatic cells (see scheme in Figure 1A). Exposure to the DNA methyltransferase inhibitor, 5 aza-deoxycytidine (5aza-dC), can induce gene activation through global
10 hypomethylation but also has significant indirect effects on expression of nonmethylated or secondary target genes (Link et al., 2008; Palii et al., 2008). We predicted that genes regulated primarily by DNA methylation would be activated by transient exposure to 5aza-dC and remain continually expressed during a recovery period, on the basis that somatic cells lack significant *de novo* methyltransferase
15 activity. In contrast, indirectly activated genes would re-impose silencing following withdrawal of the drug. We profiled the global expression changes in NIH/3T3 fibroblasts after 72hrs exposure to 1μM 5aza-dC and following 14 days cellular recovery after drug removal. Mapping the expression changes to a global DNA methylation dataset (Meissner et al., 2008) indicated that imposing increasingly
20 stringent fold-change thresholds was linearly correlated ($r^2=0.98$) with enrichment for upregulated genes that have somatically methylated CGI promoters (Figure 2). This indicates that methylated CGI genes generally exhibit stronger transcriptional activation in response to 5aza-dC compared to nonmethylated or indirectly activated genes. Consequently we employed a highly stringent 6-fold relative expression
25 threshold with a <0.01 detection *p*-value to enrich for genuine methylation-dependent candidate genes. This identified 344 (1.9%) significantly upregulated genes after three day 5aza-dC (Aza-Up) exposure and 49 (0.3%) genes significantly upregulated following 14 day recovery (Rec-Up) under normal conditions (Figure 1B). The Rec-Up gene set potentially represents genes that have escaped an epigenetic repression
30 mechanism that cannot be reimposed after disruption by 5aza-dC exposure.

To confirm our recovery assay selectively enriches for methylation-dependent candidate genes we mapped the Aza-Up and Rec-Up gene sets to an enhanced global CGI promoter methylation dataset (Meissner et al., 2008). Aza-Up contained a 2.7x enrichment of somatically methylated genes (>0.75) relative to the genome (5.4%

versus 2.0% expected; *Fishers exact* $p=0.44$) (Figure 1C), implying that >94% of CGI genes induced by 5aza-dC are not methylated and are likely to be indirectly activated by the drug (Palii et al., 2008). In contrast, Rec-Up contained a 13.4x enrichment of somatically methylated CGI genes (26.7% versus 2.0%; $p=5.6^{-07}$) (Figure 1C),
 5 suggesting the recovery assay selectively enriches for genes that are unable to re-impose DNA methylation-mediated transcriptional silencing following promoter demethylation. Consistent with this, we noted that the *Tex19.1* CGI promoter is demethylated after 1 μ M 5aza-dC treatment in a significant subpopulation of NIH/3T3 cells (~50%) and remains hypomethylated and expressed following 14 days recovery
 10 (Figure 3).

Gene ontology (GO) analysis using the DAVID annotation tool (biological process level 2) indicated the Aza-Up gene set was significantly enriched in multiple categories broadly corresponding to stress response and germline-specific classes (*response to stress* $p=0.005$; *immune response* $p=2.9e^{-12}$; *gamete generation*
 15 $p=0.038$) (Figure 1D, left). In contrast, the Rec-Up gene set resolved to only germline-associated categories (*gamete generation* $p=0.038$; *sexual reproduction* $p=0.049$), implying these are truly epigenetically regulated (Figure 1D right panel). RT-PCR confirmed that germline-specific genes are induced by 5aza-dC in NIH/3T3 fibroblasts and remain derepressed at 3, 8 and 14 days following drug withdrawal
 20 (Figure 1E). In contrast, control somatic genes rapidly re-imposed gene silencing during cellular recovery (Figure 1E). We had comparable results employing our recovery assay in primary MEFs and $p53^{-/-}$ MEFs (Figure 3).

To refine our list of methylation-dependent candidate genes we also analysed mouse embryonic fibroblasts (MEFs) homozygous for a hypomorphic mutant of
 25 *Dnmt1* (*dnmt1^{n/n}*, $p53^{-/-}$) (DP-cells) relative to their sibling $p53^{-/-}$ cells (P-cells) (Lande-Diner et al., 2007). Inactivation of *p53* prevents induction of apoptosis in *Dnmt1^{-/-}* somatic cells (Jackson-Grusby et al., 2001). Utilizing a stringent 6-fold relative expression threshold we identified 190 genes (1.0%) upregulated in DP cells (D1-Up). Consistent with the Rec-Up gene set there was a significant enrichment in
 30 germline-specific genes within the D1-Up set (*reproductive process* $p=4.0e^{-03}$; *gamete generation* $p=0.018$). We also arrayed sibling P-cells after 3 days exposure to 1 μ M 5aza-dC which also revealed strong enrichment for germline-specific categories (*gamete generation* $p=0.023$; *reproductive process* $p=0.025$). We validated our

results by RT-PCR of primary MEFs and P-cells treated with 5aza-dC and in DP cells (Figure 1F). All germline-specific genes we examined were strongly derepressed in all hypomethylated cells irrespective of the mechanism of global demethylation. Exposure to 5aza-dC along with the histone deacetylase (HDAC) inhibitor TSA did not significantly enhance derepression, with the exception of *Pramell1* and *Mov10l1* in pMEFs (Figure 1F). Germline-specific genes were not activated by TSA alone (data not shown). We did not observe mis-expression of pluripotent specific genes (*Rex1*, *Sall1*, *Nanog* and *Stella*) in DP cells or in 5aza-dC treated MEFs, implying that additional silencing mechanisms act at these genes (Figure 3). Our data suggests that germline associated genes are preferentially derepressed in hypomethylated somatic cells.

To generate a minimal list of methylation-dependent candidate genes we identified upregulated genes in common between the Rec-Up gene set, the D1-Up gene set and 5aza-dC treated P-MEFs. This contained 16 single-copy genes and IAP repeat elements as strong candidates for direct regulation by promoter DNA methylation (Figure 1G and Figure 4). As suggested by the GO analysis, a significant number of these genes (10/16) (*Tex19.1*, *PiwiL2*, *Rps4y2*, *Rhox5*, *Dazl*, *MageA11*, *Tex13*, *Mov10l1*, *Rhox4d* and *Tex19.2*) are predominantly expressed in the germline (testis) or placenta. Our top hits are *Tex19.1* and *PiwiL2* both of which have been implicated in the inhibition of transposable genetic elements in the germline (Aravin et al., 2008; Ollinger et al., 2008).

DNA methylation represses candidate genes promoters

Nine of the ten testis-specific genes in the minimal list have CGI promoters. Bisulphite sequencing of *Tex19.1*, *Tex19.2*, *Tex13* and *PiwiL2* confirmed that these promoters are heavily methylated in non-expressing primary MEFs (Figure 5A), brain and spleen but are hypomethylated in E13.5 germ cells, sperm and *Dnmt1*^{-/-} (DP) MEFs in which they are expressed (Figure 5B and 6). Promoter methylation at these loci therefore strongly correlates with their expression status *in vivo*.

To directly test the dependence of the *Tex19.1*, *Tex13* and *PiwiL2* promoters on DNA methylation for transcriptional repression we generated promoter luciferase reporter constructs for each gene. We also tested 2 control germline-specific genes (*Rnh2* and *Tex11*) that were not included in the Rec-Up gene set along with an *Oct4* reporter in human 293T cells and mouse Neuro2a cells. Transient transfection demonstrated that all three *bona fide* candidate gene promoters (*Tex19.1*, *Tex13* and

PiwiL2) could drive strong expression in somatic cells compared to the luciferase vector only (Figure 5C). In contrast *Oct4*, *Tex11*, and *Rnh2* are not active in this assay. This suggests that the *Tex19.1*, *Tex13* and *PiwiL2* promoters do not rely on germline-specific activators to drive transcription, which is consistent with an epigenetic system directing their tissue-specific expression.

To test if DNA methylation could repress transcription from these promoters we assayed expression from *in vitro* methylated *Tex19.1* and *PiwiL2* reporters. This resulted in 115-fold and 150-fold repression respectively relative to mock methylated controls (Figure 5D). To confirm that promoter methylation, as opposed to methylation of the entire vector backbone, can direct transcriptional repression we also generated *Tex19.1* reporter constructs with only the promoter region specifically DNA methylated. Transfection of these into 293T cells resulted in strong repression of luciferase activity (up to 12-fold) relative to mock-methylated constructs (Figure 6). We conclude promoter DNA methylation of *Tex19.1* and *PiwiL2* reporters is sufficient to impose a strong transcriptional silencing state.

***De novo* DNA methylation causally regulates *Tex19.1* and *PiwiL2* in differentiating ES cells.**

We used ES cells to investigate whether promoter DNA methylation at our candidate genes is the causal mechanism of transcriptional regulation or a downstream consequence of the gene expression state. Both *Tex19.1* and *PiwiL2* are expressed in E14 ES cells (day 0) and are silenced by day 4 of differentiation into embryoid bodies (EB) (Figure 7A). In day 0 ES cells the promoters of *Tex19.1* and *PiwiL2* are hypomethylated but acquire *de novo* methylation as EB differentiation proceeds (Figure 7B). Importantly, DNA methylation temporally accumulates in parallel with silencing, consistent with a causal role (Figure 7C; compare 7A & 7B) and unlike *Oct4*, where DNA methylation occurs subsequent to transcriptional repression and is therefore a secondary repressive system at the *Oct4* locus (Feldman et al., 2006). Treatment of day 7 EBs for 3 days with high (10 μ M) or low (1 μ M) concentrations of 5aza-dC was sufficient to reactivate both *Tex19.1* and *PiwiL2* (Figure 7D). Other germline associated genes that are expressed in undifferentiated ES cells (*Mov10l1*, *Tex19.2*, *Mael*, *Dazl* and *Ant4*) and silenced in EBs are also re-activated by 5aza-dC. This suggests that germline-specific genes are not being silenced through changes in transcription factor availability during differentiation, as they can be readily

reactivated by demethylation. As with somatic cells, 5aza-dC treatment was not sufficient to reactivate either *Oct4* or *Nanog* after EB associated silencing (Figure 7D).

To investigate the precise role of promoter DNA methylation at *Tex19.1* and *PiwiL2*, we used *Dnmt[3a, 3b]^{-/-}* deficient ES cells, which lack *de novo* methylation activity (Okano et al., 1999). Early passage *Dnmt[3a, 3b]^{-/-}* ES cells retain global methylation due to Dnmt1 activity and can differentiate *in vitro* (Gilbert et al., 2007). As observed with E14 ES cells, *Tex19.1* and *PiwiL2* were strongly silenced during EB differentiation of wild-type J1 ES cells but importantly, remained expressed during differentiation of *Dnmt[3a, 3b]^{-/-}* ES cells (Figure 7E). In contrast, *Oct4* was silenced in both wild-type and double mutant EBs. Even after 15 days differentiation, *Tex19.1* and *PiwiL2* were intrinsically unable to fully silence in *Dnmt[3a, 3b]^{-/-}* EBs, suggesting that *de novo* promoter methylation is the critical mark mediating silencing at these loci (data not shown). The partial repression of *PiwiL2* (7-fold in *Dnmt[3a, 3b]^{-/-}* EBs compared to >50-fold in wild-type EBs) may reflect a change in the availability of specific transcription factors or indicate that other epigenetic systems have a secondary influence on *PiwiL2* regulation. In contrast, *Tex19.1* expression remained completely unaffected during differentiation of *Dnmt[3a, 3b]^{-/-}* EBs, suggesting that promoter methylation is the exclusive mechanism for silencing this gene.

To determine the relative contribution of each *de novo* methyltransferase in targeting methylation to the *Tex19.1* and *PiwiL2* promoters, we induced EB formation in ES cells lacking either *Dnmt3a* or *Dnmt3b*. This indicated that *Tex19.1* and *PiwiL2* specifically require *Dnmt3b* to fully silence transcription in EBs (Figure 7F). qRT-PCR confirmed that both wild-type and *Dnmt3a*-null ES cells imposed strong repression of *Tex19.1* but *Dnmt3b^{-/-}* ES cells could not impose silencing (Figure 8). Bisulphite analysis revealed that the *Tex19.1* and *PiwiL2* promoters failed to acquire *de novo* methylation in *Dnmt3b^{-/-}* EBs, but are fully methylated in *Dnmt3a^{-/-}* EBs (Figure 8), indicating that Dnmt3b has a specific role in targeting DNA methylation to these loci. These data provide strong evidence that Dnmt3b dependent promoter CpG methylation is necessary and sufficient for silencing *Tex19.1* expression.

Methylation-dependent genes are depleted of histone modifications.

It is possible that histone modifications contribute either to target DNA methylation to the *Tex19.1* locus or to provide a secondary repressive mechanism.

However, genome-wide histone modification profiling by Mikkelsen et al (2007) suggests the *Tex19.1* promoter is not enriched for H3K27me3, H3K36Me3 and H3K9Me3 in all assayed cell types. We experimentally validated the relative depletion of histone modifications at the *Tex19.1* promoter by performing native chromatin-immunoprecipitation from P and DP cells. The repressive polycomb complex 2 (PRC2) associated H3K27me3 mark is absent from, or only weakly deposited, at silenced *Tex19.1* in P-MEFs. (Figure 9). Moreover mutant ES cells lacking the PRC2 component *Eed* could silence and *de novo* methylate *Tex19.1* comparably to wild-type cells upon retinoic acid (RA) differentiation, indicating polycomb mediated epigenetic repression is functionally dispensable for *Tex19.1* silencing (Figure 10A and Figure 9). As expected, *Dnmt3b*^{-/-} ES cells failed to silence *Tex19.1* following RA differentiation (Figure 8). Another possibility is that *G9a*, which catalyses H3K9 dimethylation, may have a role in repressing *Tex19.1* as it has been shown to target DNA methylation to several gene promoters (Dong et al., 2008). However, *Tex19.1* is not mis-expressed in *G9a*^{-/-} MEFs or when *G9a* activity is inhibited in somatic cells. GO analysis also demonstrates that germline associated genes are not preferentially mis-expressed in *G9a*^{-/-} MEFs (Sampath et al., 2007). A combination of Bix-01294 (a *G9a* inhibitor) and TSA does not activate *Tex19.1* expression in P-MEFs (data not shown) (Kubicek et al., 2007). Taken together these data reinforce the view that DNA methylation is the exclusive epigenetic mark that mediates *Tex19.1* silencing in somatic cells.

To establish if depletion of histone modifications is a general attribute of germline specific genes, we generated and mapped a data set of CGI containing testis-specific genes (defined as expressed >10-fold in testis than all somatic tissues in the BioGPS database) to their histone modification state in MEFs (Mikkelsen et al., 2007). Testis-specific genes are highly enriched in promoters which lack all assayed histone modifications (*Fishers exact* $p=0.00019$) (Figure 10B). This is consistent with a role for DNA methylation independently of other epigenetic mechanisms at germline-specific genes. Both the D1-Up ($p=3.8e^{-8}$) and Rec-Up ($p=0.00006$) sets are also enriched in CGI promoters associated with an absence of histone modifications. Downregulated genes in all arrays exhibited no significant enrichment for promoters depleted of histone marks.

***Dnmt3b* directs *de novo* methylation of germline-specific CGI promoters *in vivo*.**

As Dnmt3b is responsible for targeting methylation to *Tex19.1* and *PiwiL2* in differentiating ES cells, we investigated the possibility that Dnmt3b has a specific role in *de novo* methylating germline associated CpG island promoters during development. We derived *Dnmt3b*^{-/-} primary MEFS from E12.5 and E13.5 day embryos and demonstrated that *Tex19.1* and *PiwiL2* are robustly induced compared to WT and heterozygous controls (Figure 10C). Global expression analysis indicates that only 27 (0.15%) genes are significantly upregulated in these cells (D3b-Up) (Figure 10D). Strikingly, D3b-Up genes are highly enriched in only germline specific GO categories (gamete generation $p=0.004$; sexual reproduction $p=0.0005$). Mapping the D3b-Up gene set to their methylation profile (Meissner et al., 2008), indicated that it contained a 20x enrichment of methylated (>0.75) CGI promoters relative to the genome; 40.0% versus 2.0% expected ($p=4.6e^{-12}$) (Figure 10E). Consistent with this, the CGI promoters of *Tex19.1*, *PiwiL2* and *Dazl* are hypomethylated in *Dnmt3b*^{-/-} MEFs relative to WT cells (Figure 10F). In contrast *Tex13*, a germline-specific gene that is not derepressed in *Dnmt3b*^{-/-} MEFS, remains fully methylated in the mutant cells. This implies a functional role for Dnmt3b in silencing a specific subset of germline-specific genes in somatic cells via *de novo* methylation of CGI promoters after implantation (Li, 2002). The D3b-Up set is also highly enriched in CGI genes that lack associated histone modifications ($p = 2.1e^{-21}$).

***Tex19.1* is a specific target for silencing by Kaiso.**

Since we have identified a novel subset of genes regulated exclusively by DNA methylation, we investigated whether the methyl-CpG binding proteins Kaiso, MeCP2 and Mbd2, mediate transcriptional repression at these loci (Sasai and Defossez, 2009). *Tex19.1* was expressed in triple-null (*Mbd2*, *Mecp2*, *Kaiso*)^{-/-} (MKO) and *Kaiso*-null fibroblasts (Figure 11A & 12) and had limited expression in *Mbd2*-null cells. Bisulphite analysis revealed that *Tex19.1* was fully methylated relative to wild-type fibroblasts in all the mutant cell lines (Figure 11B). *Tex19.1* therefore represents a novel example of a heavily methylated CpG-dense promoter driving expression above background. It requires Kaiso for complete DNA methylation-dependent transcriptional silencing in fibroblasts. However, the level of *Tex19.1* expression in *Kaiso*-null and MKO cells is significantly lower to that observed in DP cells or wild-type testis by qRT-PCR, implying that DNA methylation imposes partial silencing through alternative (possibly direct) mechanisms at this

locus (Figure 12). Transcript mapping indicates that DP, *Mbd2*^{-/-} and *Kaiso*^{-/-} cells all transcribe *Tex19.1* from the same canonical TSS (Figure 12).

To confirm that Kaiso can directly contribute to *Tex19.1* silencing, we over-expressed GFP-tagged *Xenopus* Kaiso (xKaiso) in *Kaiso*^{-/-} fibroblasts. qRT-PCR
 5 analysis of GFP positive (xKaiso rescued) and GFP negative (control) cells demonstrated that ectopic expression of xKaiso can further silence *Tex19.1* by ~5-fold (Figure 11C). Additionally, *Kaiso*^{-/-} ES cells cannot impose full *Tex19.1* silencing during EB formation; *Tex19.1* expression was only reduced ~4-fold in differentiated *Kaiso*^{-/-} EBs compared to full silencing (>100-fold) in wild-type EBs (Figure 11D).
 10 This is despite the fact that *de novo* methylation of *Tex19.1* proceeds normally in both cell types and is consistent with the fibroblast analysis (Figure 11E). In contrast both ES cell lines can silence *Oct4* expression during EB formation. These data strongly support a critical role for Kaiso in the maintenance of gene silencing at *Tex19.1*. Indeed, the *Tex19.1* proximal promoter is enriched with the symmetrically methylated
 15 CGCG tetranucleotides that Kaiso preferentially binds to in a methylation-dependent manner (Prokhortchouk et al., 2006). ChIP of endogenous Kaiso protein demonstrated that it was significantly enriched at the methylated *Tex19.1* promoter in wild-type but not *Kaiso*^{-/-} fibroblasts (*students T-test* $p=0.02$) (Figure 11F). In addition a fusion protein between the zinc-finger domain of *Xenopus* Kaiso, which is
 20 responsible for methyl-CpG binding, and the viral VP16 transcriptional activation domain specifically activated a methylated *Tex19.1* reporter relative to VP16 but not its unmethylated counterpart ($p=0.00009$) (Figure 11G) indicating that Kaiso preferentially localizes to the methylated *Tex19.1* promoter.

Apart from *Tex19.1*, no significant derepression of other candidate or
 25 germline-specific genes was observed in *Kaiso*^{-/-} fibroblasts (Figure 11A). *PiwiL2* was weakly expressed in *Mbd2*^{-/-} and MKO fibroblasts suggesting that Mbd2 may contribute to full silencing at this locus. Expression profiling of *Kaiso*^{-/-} fibroblasts indicated that 31 transcripts are upregulated relative to WT cells, with the top target being *Tex19.1*. In agreement with the RT-PCR analysis, *Tex19.1* and *PiwiL2* are the
 30 only germ cell specific genes that are misregulated upon expression arraying MKO cells (data not shown). This implies that the tested MeCPs are not required to maintain silencing of methylated germline-specific genes in somatic cells. Moreover gene targets upregulated in *Kaiso*^{-/-} cells were not significantly enriched in methylated CGI genes ($p=0.29$). *Tex19.1* was the only gene commonly upregulated between

Dnmt1^{-/-} and *Kaiso*^{-/-} fibroblasts. The lack of general gene activation is consistent with genetic studies that suggest MeCPs do not act as global regulators of gene expression. However, the data is also consistent with Kaiso being a specific mediator of methylation dependent silencing at *Tex19.1*, which therefore represents an exception as a gene that is causally regulated by DNA methylation through recruitment of a methyl-CpG binding protein.

DNA methylation regulates lineage specific expression of *Tex19.1* in PGCs.

Previous reports have suggested that DNA demethylation correlates with the temporal activation of 3 germline-specific genes during post-migratory germ cell development (Maatouk et al., 2006). Primordial germ cell (PGC) specification occurs in the proximal epiblast of gastrulating mouse embryos at E6.25 resulting in a pool of 40-50 PGCs located in extra-embryonic mesoderm at the base of the allantois by E7.25. These nascent PGCs migrate into the hindgut endoderm around E8.5, colonising the emerging gonads between E10.5-E11.5, then enter a phase of epigenetic reprogramming that includes global loss of DNA methylation and histone modifications (Hajkova et al., 2008; Hajkova et al., 2002). This epigenetic reprogramming event has been proposed to be responsible for erasure of DNA methylation at the *Dazl* promoter, and activation of *Dazl* expression in post-migratory germ cells. We investigated if *Tex19.1* and other germline methylation-dependent genes identified here are also temporally regulated by promoter DNA demethylation during PGC development. We isolated E10.5 (>95% pure) and E13.5 germ cells (>98% pure) from embryonic gonads based on expression of a transgenic *Oct4*-GFP marker (Figure 13). As demonstrated previously, *Dazl* was methylated in E10.5 PGCs but hypomethylated by E13.5 and this correlated with strongly increased transcript levels at E13.5 (Figure 14A) (Maatouk et al., 2006). *Dazl* remained methylated in the somatic cells of the gonad throughout this period.

Unexpectedly, we found the *Tex19.1* promoter was hypomethylated in PGCs at E10.5 prior to epigenetic reprogramming and also in male and female germ cells at E13.5 (Figure 14B). *Tex19.1* was also hypomethylated and expressed in migrating PGCs isolated from the hind-gut region of E9.5 embryos (Figure 15). As expected, *Tex19.1* was heavily methylated in the somatic cells of the gonad and hind-gut. qRT-PCR indicated there was no significant change in *Tex19.1* expression between E10.5 and E13.5 PGCs (Figure 14B). Therefore unlike *Dazl*, *Tex19.1* is unmethylated and expressed in PGCs, prior to entry into the gonad and epigenetic reprogramming at

E11.5. *PiwiL2* followed a similar pattern to *Tex19.1* and was demethylated and expressed at E10.5 (Figure 14B). In contrast, *Tex19.2* was methylated in E10.5 PGCs and underwent demethylation and transcriptional activation by E13.5 in male and female PGCs comparable to *Dazl* (Figure 13). Expression analysis of an additional
 5 eight germline-specific genes, which are derepressed by 5aza-dC and in *Dnmt1*^{-/-} MEFs, demonstrated that there is a bimodal pattern at this stage such that genes are either strongly activated in PGCs between E10.5 and E13.5 or exhibit no significant expression change (Figure 14C). Based on methylation analysis of *Tex19.1*, *PiwiL2*, *Dazl* and *Tex19.2*, we predict that somatically methylated CGI genes expressed in
 10 PGCs prior to entry into the gonad would already be hypomethylated whereas the genes activated by E13.5 undergo promoter demethylation and activation coincident with epigenetic reprogramming. These latter genes represent novel targets crucially reliant on DNA methylation for lineage-specific expression and temporal activation in post-migratory PGCs similarly to *Dazl*. What determines the different temporal
 15 activation profiles is not clear but may be related to specific gene functions.

As *Tex19.1* is expressed in E9.5 PGCs, we investigated the developmental regulation and promoter methylation dynamics of *Tex19.1* during post-implantation development. *Tex19.1* protein and RNA is expressed in the pluripotent epiblast cells in E6.5 embryos, is downregulated during gastrulation, and can only be weakly
 20 detected in late E7.5 embryos. *Tex19.1* protein expression is also detected in extra-embryonic tissues during post-implantation development (Figure 15). Bisulphite sequencing of the embryonic portions of E6.5-E8.5 embryos showed that the *Tex19.1* promoter is methylated at E7.5 and E8.5 but interestingly, is in an intermediate hypomethylated state at E6.5 (Figure 14D). This suggests that embryonic *Tex19.1* is
 25 undergoing a process of *de novo* methylation at E6.5, which is complete by E7.5. Importantly, *Tex19.1* acquires DNA methylation prior to repression during post-implantation development again suggesting that promoter methylation is the causal mechanism of silencing in somatic tissues. Moreover, this data is consistent with expression of *Tex19.1* being regulated by Dnmt3b mediated *de novo* methylation, as
 30 Dnmt3b is maximally expressed at ~E6.5 (Watanabe et al., 2002).

It is not clear if *Tex19.1* in PGCs escapes embryonic *de novo* methylation at ~E6.5 or is specifically demethylated in PGCs prior to E9.5. However, we did not detect any cells expressing *Tex19.1* protein in the base of the allantois or in the hindgut endoderm in E8.5 embryo's (Figure 15), suggesting that *Tex19.1* is not

expressed in nascent PGCs and the *Tex19.1* promoter may be methylated then subsequently demethylated in PGCs before E9.5. PGCs undergo an epigenetic reprogramming event at ~E8.5 (Hajkova et al., 2008) that has been suggested to involve limited loss of DNA methylation at as yet unidentified genomic loci (Seki et al., 2007). The *Tex19.1* promoter is therefore a good candidate locus to be regulated by the proposed lineage-specific DNA demethylation event in ~E8.5 PGCs. Indeed it is possible that the cohort of methylation-dependent genes we have identified as expressed in E10.5 PGCs (Figure 14C) represent novel loci activated by promoter demethylation during epigenetic reprogramming of ~E8.5 PGCs.

10 **Discussion**

Role of promoter CpG methylation

We set out to unequivocally identify genes that are regulated exclusively by DNA methylation with the expectation that this gene set would also be potential targets for methyl-CpG binding repressor proteins (MeCPs). Our initial screen employed a novel ‘epigenetic disruption and recovery’ assay in which somatic cells are exposed to 5aza-dC and then allowed to recover for 14 days after drug exposure. We hypothesized that because the mechanism for targeting *de novo* methylation is largely inactive in normal differentiated cells (Okano et al., 1999), *bona fide* methylation-dependent genes would be continually expressed once promoter CpG methylation has been removed. Our results indicate that this assay selectively enriches for upregulated genes that are normally silenced and associated with somatically methylated CGI promoters. It is possible that this ‘hit and run’ strategy will be useful in identifying genes that are epigenetically silenced by DNA methylation in other contexts e.g. tissue specifically and cancer; and distinguish them from genes (e.g. *Oct4*) that rely on multiple independent epigenetic silencing mechanisms (Feldman et al., 2006; Illingworth et al., 2008). Indeed it is notable that >94% of CGI genes initially induced by 5aza-dC are activated indirectly from non-methylated promoters, largely as a stress response to 5aza-dC mediated cytotoxicity.

We generated a minimal cohort of candidate genes that rely directly on promoter CpG methylation by identifying genes upregulated in common between our three complimentary experimental approaches (Rec-Up, D1-Up, Paza-Up). The small number of candidate genes that responded to demethylation in multiple assays indicated there is a limited (but none the less important) role for DNA methylation ‘only’, in maintaining silencing of mostly germ cell specific genes in somatic cells. In

other contexts DNA methylation can combine with other repressive mechanisms to provide a secondary lock on silencing (Fouse et al., 2008; Tachibana et al., 2008). The top hit from our screen, *Tex19.1*, relies exclusively on promoter DNA methylation to initiate and maintain its silencing. This view is supported by an absence of repressive

5 histone modifications at the *Tex19.1* promoter. We find a strong correlation between the presence of DNA methylation at promoter proximal regions and the absence H3K27me3, H3K4me3 and H3K9me3 among upregulated genes in each array (Figure 16). This suggests that the presence of DNA methylation and histone modifications at CGI promoters are incompatible, when the former is the primary regulatory

10 mechanism. Indeed, analysis of the histone modification state of all testis-specific genes indicates that there is a highly significant general depletion of histone marks at germline-specific CGI promoters. Considered with the gene ontology analysis of the Rec-Up and D1-Up gene sets, which revealed strong enrichments for germline-specific transcripts, our data strongly supports a direct and *causal* regulatory role for

15 promoter DNA methylation at a significant cohort of CGI-associated germline-specific genes exemplified by *Tex19.1*. Promoter CpG methylation therefore directs the lineage-specific cellular memory of at least a subset of germline-specific genes.

Our data indicates Dnmt3b is responsible for *de novo* methylation of *Tex19.1*, *Dazl* and *PiwiL2*, during development. Gene's upregulated in *Dnmt3b*^{-/-} MEFs are

20 highly germline-specific and are strongly enriched in CGI promoters that are somatically methylated and lacking repressive histone marks. It is therefore likely that Dnmt3b has a functional role in directing methylation to specific, largely germline-associated, CpG islands during embryogenesis. How Dnmt3b targets specific CGI promoters is an important question for future studies. Previous work has

25 indicated that the transcriptional repressor E2f6, which interacts with Dnmt3b, is required for silencing and subsequent hypermethylation of *Tex12* and *Tuba3* in somatic cells (Hervouet et al., 2009; Pohlers et al., 2005). However, we did not observe any significant enrichment of consensus E2f6 binding sites among the D3b-Up genes relative to genome-wide promoters (Figure 17). Moreover, array analysis of

30 *E2f6*^{-/-} MEFs identified only one gene (*Stag3*), that is also present in our D3b-Up list (Storre et al., 2005) implying that E2f6 binding is not the primary mechanism for targeting Dnmt3b mediated *de novo* methylation to the gene set we have identified.

Regulation of germline-specific genes by promoter CpG methylation.

Why do only germline-specific genes preferentially utilize DNA methylation for silencing? Hypomethylated germline-specific promoters are not exposed to the high rate of mutation associated with methyl-CpGs and therefore have the unique advantage of preserving their CpG density during successive generations (Gehring et al., 2009). In contrast somatic cell-specific genes regulated by CpG methylation would, by definition, be silenced and hypermethylated in the germline, with a high risk of progressively losing CpGs by deamination. This would lead to reduced promoter CpG density over generations and, once below a certain threshold, loss of a strict methylation-dependent silencing mechanism. For a gene to be entirely regulated by DNA methylation, it must therefore satisfy several conditions. It should; (i) have a CpG-dense promoter (ii) be mainly expressed in the germline (for evolutionary stability), (iii) utilize ubiquitous transcription factors to drive expression and, (iv) not have gained additional upstream epigenetic regulation. Additionally it must attract tissue specific *de novo* methylation during development, which may be dependent on Dnmt3b. The above criteria imposes a severe limit on the number of genes that can theoretically be regulated primarily by promoter CpG methylation and accounts for the small number of *bona fide* methylation-dependent targets that have been characterized up to now. *Tex19.1* satisfies all the above requirements and represents a model gene that is strictly regulated by DNA methylation.

A proposed role of DNA methylation is maintenance of transposon inactivation - the 'genome defence' hypothesis (Yoder et al., 1997). A drawback of this mechanism is that PGCs undergo phases of epigenetic reprogramming resulting in partial demethylation of retroelements (Hajkova et al., 2002), which could conceptually lead to transient transposon expression and mutagenic genomic insertion. *Tex19.1* may participate in a retrotransposon silencing mechanism, as the class II LTR-retrotransposon MMERVK10C is significantly post-transcriptionally upregulated in *Tex19.1*^{-/-} sperm (Ollinger et al., 2008). Intriguingly, this function may be linked to the distinct methylation-dependent regulation of *Tex19.1* expression; potential activation of methylation-dependent retroelements may be coupled by concomitant activation of *Tex19.1*. Similarly, *PiwiL2* has an important role in suppressing retroelements post-transcriptionally or by re-targeting CpG methylation to them via piRNAs (Aravin et al., 2008). It is striking that *PiwiL2* and *Tex19.1* are responsive to promoter demethylation and could therefore utilise global demethylation as an expression cue to maintain suppression of aberrant retroelement activity.

Role of Methyl-binding proteins

The methyl-CpG binding protein, Kaiso, contributes to DNA methylation mediated silencing of *Tex19.1* in mouse cells. Although Kaiso is a critical component that ensures full silencing of *Tex19.1*, other mechanisms such as steric hindrance of DNA binding factors, inhibition by alternative MeCPs or direct repression by methyltransferases may also contribute to silencing, as *Tex19.1* is only partially derepressed in the absence of Kaiso (Dunican et al., 2008; Sasai and Defossez, 2009). We were unable to find any significant derepression of other methylation-dependent genes in (*Mbd2*, *Mecp2*, *Kaiso*)-null cells. The absence of general activation of methylation-dependent genes, considered with the relatively mild phenotypes of MeCP-mutant mice, suggests that MeCP2, Mbd2 and Kaiso do not have a primary role in mediating silencing at methylated CGIs *in vivo*. We favour a direct model for methylation-dependent silencing rather than recruitment of alternative MeCPs not tested here, based on previous analyses that MeCPs largely bind to non-overlapping genomic loci (Ballestar et al., 2003; Klose et al., 2005). It is possible that direct interference by DNA methylation is sufficient to prevent activation of methylated CGIs by blocking methyl-CpG sensitive transcription activators. Indeed, it has recently been demonstrated that the CXXC protein Cfp1 targets the activating modification H3K4me3 to unmethylated CGIs but is occluded from methylated CGIs (Thomson et al., 2010). Interestingly, *Tex19.1* also gains H3K4me3 following promoter demethylation and this may be sufficient to direct its strong expression (Figure 9). Direct transcription factor inhibition by DNA methylation can account for the different transcriptomic outcomes between hypomethylated *Dnmt1* mutants and MeCP-deficient cell lines, where DNA methylation patterns are maintained (Skene et al., 2010). It is possible that MeCPs are recruited to methylation-dependent genes to impose an additional layer of repression but are not strictly necessary to retain silencing due to the direct effect of DNA methylation *per se*. We speculate that *Tex19.1* uniquely requires Kaiso to impose full transcriptional silencing because of the inherent strength of the *Tex19.1* promoter (Fig 5C), which drives reporter expression comparably to the highly active CMV promoter. The strength of the *Tex19.1* promoter in somatic contexts also accounts for it being the top target in our global expression analyses of hypomethylated cells.

In summary we have identified a subset of germline-specific genes that causally rely on a single epigenetic mechanism, promoter DNA methylation, to

initiate and maintain their developmental expression pattern. Moreover, it is likely that MeCPs are not the primary mechanism for mediating the effects of promoter CpG methylation in this developmental context. Determining the precise mechanism of DNA methylation-dependent silencing *in vivo* is an interesting question for future research.

Reference List

- Aravin,A.A., R.Sachidanandam, D.Bourc'his, C.Schaefer, D.Pezic, K.F.Toth, T.Bestor, and G.J.Hannon. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31**: 785-799.
- Ballestar,E., M.F.Paz, L.Valle, S.Wei, M.F.Fraga, J.Espada, J.C.Cigudosa, T.H.Huang, and M.Esteller. 2003. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J.* **22**: 6335-6345.
- Bird,A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**: 6-21.
- Dong,K.B., I.A.Maksakova, F.Mohn, D.Leung, R.Appanah, S.Lee, H.W.Yang, L.L.Lam, D.L.Mager, D.Schubeler, M.Tachibana, Y.Shinkai, and M.C.Lorincz. 2008. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* **27**: 2691-2701.
- Dunican,D.S., A.Ruzov, J.A.Hackett, and R.R.Meehan. 2008. xDnmt1 regulates transcriptional silencing in pre-MBT *Xenopus* embryos independently of its catalytic function. *Development* **135**: 1295-1302.
- Feldman,N., A.Gerson, J.Fang, E.Li, Y.Zhang, Y.Shinkai, H.Cedar, and Y.Bergman. 2006. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* **8**: 188-194.
- Fouse,S.D., Y.Shen, M.Pellegrini, S.Cole, A.Meissner, N.L.Van, R.Jaenisch, and G.Fan. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**: 160-169.
- Gehring,M., W.Reik, and S.Henikoff. 2009. DNA demethylation by DNA repair. *Trends Genet.* **25**: 82-90.
- Gilbert,N., I.Thomson, S.Boyle, J.Allan, B.Ramsahoye, and W.A.Bickmore. 2007. DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J. Cell Biol.* **177**: 401-411.
- Hajkova,P., K.Ancelin, T.Waldmann, N.Lacoste, U.C.Lange, F.Cesari, C.Lee, G.Almouzni, R.Schneider, and M.A.Surani. 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**: 877-881.

- Hajkova,P., S.Erhardt, N.Lane, T.Haaf, O.El-Maarri, W.Reik, J.Walter, and M.A.Surani. 2002. Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**: 15-23.
- 5 Hervouet,E., F.M.Vallette, and P.F.Cartron. 2009. Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics*. **4**: 487-499.
- Holliday,R. and J.E.Pugh. 1975. DNA modification mechanisms and gene activity during development. *Science* **187**: 226-232.
- 10 Illingworth,R., A.Kerr, D.Desousa, H.Jorgensen, P.Ellis, J.Stalker, D.Jackson, C.Clee, R.Plumb, J.Rogers, et al. 2008. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS. Biol.* **6**: e22.
- Jackson-Grusby,L., C.Beard, R.Possemato, M.Tudor, D.Fambrough, G.Csankovszki, J.Dausman, P.Lee, C.Wilson, et al. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* **27**: 31-39.
- 15 Klose,R.J., S.A.Sarraf, L.Schmiedeberg, S.M.McDermott, I.Stancheva, and A.P.Bird. 2005. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19**: 667-678.
- Kubicek,S., R.J.O'Sullivan, E.M.August, E.R.Hickey, Q.Zhang, M.L.Teodoro, S.Rea, K.Mechtler, J.A.Kowalski, C.A.Homon, T.A.Kelly, and T.Jenuwein. 2007. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Mol. Cell* **25**: 473-481.
- 20 Lande-Diner,L., J.Zhang, I.Ben-Porath, N.Amariglio, I.Keshet, M.Hecht, V.Azuara, A.G.Fisher, G.Rechavi, and H.Cedar. 2007. Role of DNA methylation in stable gene repression. *J. Biol. Chem.* **282**: 12194-12200.
- 25 Li,E. 2002. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**: 662-673.
- Li,E., C.Beard, and R.Jaenisch. 1993. Role for DNA methylation in genomic imprinting. *Nature* **366**: 362-365.
- 30 Link,P.A., M.R.Baer, S.R.James, D.A.Jones, and A.R.Karpf. 2008. p53-inducible ribonucleotide reductase (p53R2/RRM2B) is a DNA hypomethylation-independent decitabine gene target that correlates with clinical response in myelodysplastic syndrome/acute myelogenous leukemia. *Cancer Res.* **68**: 9358-9366.
- 35 Maatouk,D.M., L.D.Kellam, M.R.Mann, H.Lei, E.Li, M.S.Bartolomei, and J.L.Resnick. 2006. DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**: 3411-3418.
- Martin,C., I, J.Hansen, D.Leaford, S.Pollard, and B.D.Hendrich. 2009. The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PLoS. One.* **4**: e4315.

- Meissner,A., T.S.Mikkelsen, H.Gu, M.Wernig, J.Hanna, A.Sivachenko, X.Zhang, B.E.Bernstein, C.Nusbaum, D.B.Jaffe, et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766-770.
- 5 Mikkelsen,T.S., M.Ku, D.B.Jaffe, B.Issac, E.Lieberman, G.Giannoukos, P.Alvarez, W.Brockman, T.K.Kim, R.P.Koche, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553-560.
- Mohn,F., M.Weber, M.Rebhan, T.C.Roloff, J.Richter, M.B.Stadler, M.Bibel, and D.Schubeler. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**: 755-766.
- 10 Okano,M., D.W.Bell, D.A.Haber, and E.Li. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.
- 15 Ollinger,R., A.J.Childs, H.M.Burgess, R.M.Speed, P.R.Lundegaard, N.Reynolds, N.K.Gray, H.J.Cooke, and I.R.Adams. 2008. Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS. Genet.* **4**: e1000199.
- 20 Palii,S.S., B.O.Van Emburgh, U.T.Sankpal, K.D.Brown, and K.D.Robertson. 2008. DNA methylation inhibitor 5-Aza-2'-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B. *Mol. Cell Biol.* **28**: 752-771.
- Pohlers,M., M.Truss, U.Frede, A.Scholz, M.Strehle, R.J.Kuban, B.Hoffmann, M.Morkel, C.Birchmeier, and C.Hagemeier. 2005. A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr. Biol.* **15**: 1051-1057.
- 25 Prokhortchouk,A., O.Sansom, J.Selfridge, I.M.Caballero, S.Salozhin, D.Aithozhina, L.Cerchietti, F.G.Meng, L.H.Augenlicht, J.M.Mariadason, et al. 2006. Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell Biol.* **26**: 199-208.
- Reik,W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**: 425-432.
- 30 Riggs,A.D. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**: 9-25.
- Sampath,S.C., I.Marazzi, K.L.Yap, S.C.Sampath, A.N.Krutchinsky, I.Mecklenbrauker, A.Viale, E.Rudensky, M.M.Zhou, B.T.Chait, and A.Tarakhovsky. 2007. Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol. Cell* **27**: 596-608.
- 35 Sasai,N. and P.A.Defossez. 2009. Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int. J. Dev. Biol.* **53**: 323-334.
- Skene,P.J., R.S.Illingworth, S.Webb, A.R.Kerr, K.D.James, D.J.Turner, R.Andrews, and A.P.Bird. 2010. Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Mol. Cell* **37**: 457-468.

Storre,J., A.Schafer, N.Reichert, J.L.Barbero, S.Hauser, M.Eilers, and S.Gaubatz. 2005. Silencing of the meiotic genes SMC1beta and STAG3 in somatic cells by E2F6. *J. Biol. Chem.* **280**: 41380-41386.

5 Tachibana,M., Y.Matsumura, M.Fukuda, H.Kimura, and Y.Shinkai. 2008. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J.* **27**: 2681-2690.

Thomson,J.P., P.J.Skene, J.Selfridge, T.Clouaire, J.Guy, S.Webb, A.R.Kerr, A.Deaton, R.Andrews, K.D.James, D.J.Turner, R.Illingworth, and A.Bird. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**:
10 1082-1086.

Watanabe,D., I.Suetake, T.Tada, and S.Tajima. 2002. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech. Dev.* **118**: 187-190.

Yoder,J.A., C.P.Walsh, and T.H.Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**: 335-340.
15

Claims

1. A nucleic acid construct encoding a gene and/or promoter primarily and causally regulated by methylation and a reporter element operably linked thereto.
- 5 2. A cell comprising a nucleic acid construct encoding a gene and/or promoter primarily and causally regulated by methylation and a reporter element operably linked thereto.
- 10 3. Use of the nucleic acid construct of claim 1 or cell of claim 2, for studying, evaluating and/or monitoring the role of epigenetic events in cell processes and/or for determining or evaluating the effect of a test agent on epigenetic events.
- 15 4. The nucleic acid construct of claim 1, cell of claim 2 or use of claim 3, wherein the gene and/or promoter, primarily and causally regulated by methylation, is obtainable by a method comprising the steps of:
 - (a) identifying de-repressed genes in a cell exhibiting global hypomethylation;
 - (b) identifying de-repressed genes in cells treated with 5-aza deoxycytosine (5-aza dC); and
 - (c) allowing the cells treated in step (b) to recover in the absence of 5-aza dC
- 20 and thereafter identifying genes which remain de-repressed;
wherein steps (a), (b) and (c) may be executed in any order and genes, or promoters associated therewith, identified as being de-repressed in each of steps (a), (b) and (c) are primarily and causally regulated by methylation.
- 25 5. The nucleic acid construct of claim 1, cell of claim 2 or use of claim 3, wherein the gene and/or promoter primarily and causally regulated by methylation is selected from the group consisting of:
 - (i) Tex19.1 (ENSMUSG00000039329)
 - 30 (ii) Piwil2 (ENSMUSG00000033644)
 - (iii) Tex19.2 (ENSMUSG00000039337)

(iv) *Mov10l1*: ENSMUSG00000015365

(v) *Dazl* (ENSMUSG00000010592)

(vi) *Tex13* (ENSMUSG00000042386)

(vii) *Taf7l* (ENSMUSG00000009596)

5

6. The nucleic acid construct of claim 1, cell of claim 2 or use of claim 3, wherein the nucleic acid construct encodes the *Tex19.1* gene and/or *Tex19.1* promoter operably linked to a reporter element.

10

7. The nucleic acid construct of claim 1, cell of claim 2 or use of claim 3, wherein the nucleic acid construct encodes the *PiwiL2* promoter operably linked to a reporter element.

15

8. The nucleic acid construct, cell or use of claims 6 or 7, wherein the reporter element is Green Fluorescent Protein (GFP).

9. A transgenic animal comprising the nucleic acid construct of claim 1, 4, 5, 6, 7 or 8.

20

10. A method of studying, evaluating and/or monitoring the role of epigenetic events in cell processes and/or determining, monitoring and/or evaluating the effect of a test agent on epigenetic events, said method comprising the step of:

25

(a) contacting a test agent with a system comprising a nucleic acid encoding an element primarily and causally regulated by an epigenetic event;

(b) determining the expression of said element; and

(c) comparing the level of expression with that of a control system not contacted with the test agent;

wherein any modulation of expression following contact with the test agent indicates that the test agent modulates epigenetic events.

30

11. The method of claim 10, wherein the system comprises the nucleic acid construct of claim 1, 4, 5, 6, 7 or 8

12. The method of claims 10 or 11, wherein the system is an *in vitro* cell based system.

13. The method of claims 10, 11 or 12, wherein the system comprises the cell of claim 2, 4, 5, 6, 7 or 8.

14. A method of identifying genes regulated by DNA methylation, said method comprising the steps of:

- (a) identifying de-repressed genes in a cell exhibiting global hypomethylation;
- (b) identifying de-repressed genes in cells treated with 5-aza deoxycytosine (5-aza dC); and
- (c) allowing the cells treated in step (b) to recover in the absence of 5-aza dC and thereafter identifying genes which remain de-repressed;

wherein steps (a), (b) and (c) may be executed in any order and genes identified as being de-repressed in each of steps (a), (b) and (c) are regulated by DNA methylation.

15. The method of claim 14, wherein the cell exhibiting global hypomethylation is selected from the group consisting of:

- (a) a cell contacted with a chemical capable of removing methyl groups from DNA;
- (b) a cell lacking one or more genes associated with the regulation of DNA methylation; and
- (c) a modified cell wherein one or more genes associated with the regulation of methylation is/are mutated to effect global hypomethylation.

16. The method of claims 14 or 15, wherein the cell exhibiting global hypomethylation comprises a mutated version of a gene encoding DNA (cytosine-5)-methyltransferase 1 (Dnmt1).

17. The method of claims 14, 15 or 16, wherein the cells exhibiting global hypomethylation further comprise a modified *Trp53* gene.

18. The method of claims 14, 15, 16 or 17, wherein the cells treated with
5 5-aza deoxycytosine (5-aza dC) are primary mouse embryonic fibroblasts, optionally harbouring a modified *Trp53* gene.

19. The method of claims 14, 15, 16, 17 or 18 wherein the cells treated in
step (b) are allowed to recover in the absence of 5-aza dC for at least about 14 hours.
10

20. The method of claims 14, 15, 16, 17, 18 or 19, wherein genes
identified as regulated by DNA methylation, are further subjected to analysis in cells
which lack de novo methylation activity, wherein those genes shown to be activated,
induced and/or expressed in cells which lack de novo methylation activity, are
15 regulated by DNA methylation.

21. The method of claim 20, wherein the cells which lack de novo
methylation activity are cells which lack a functional DNA (cytosine-5-)-
methyltransferase 3 beta (*Dnmt3b*) enzyme.
20

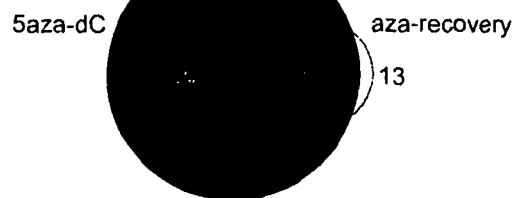
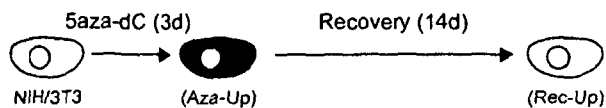
22. An assay comprising one or more immobilised nucleic acids
representative of one or more genes and/or associated promoters, primarily and
causally regulated by an epigenetic event and/or nucleic acids comprising sequences
complementary thereto, wherein the genes and/or associated promoters, primarily and
25 causally regulated by an epigenetic event are identifiable by a method comprising the
steps of:

- (a) identifying de-repressed genes in a cell exhibiting global hypomethylation;
- (b) identifying de-repressed genes in cells treated with 5-aza deoxycytosine (5-
aza dC); and
- 30 (c) allowing the cells treated in step (b) to recover in the absence of 5-aza dC
and thereafter identifying genes which remain de-repressed;

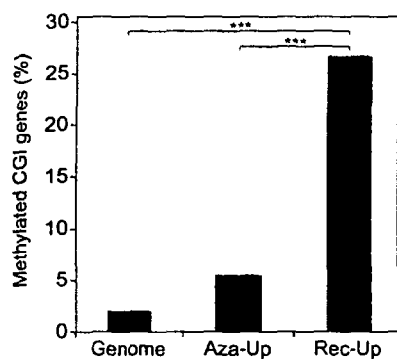
wherein steps (a), (b) and (c) may be executed in any order and genes
identified as being de-repressed in each of steps (a), (b) and (c) are regulated by DNA
methylation.

b.

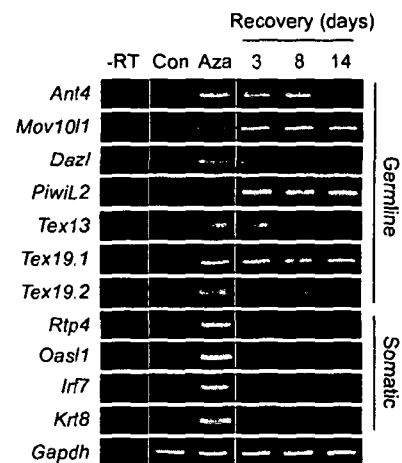
PCT/GB2011/001305



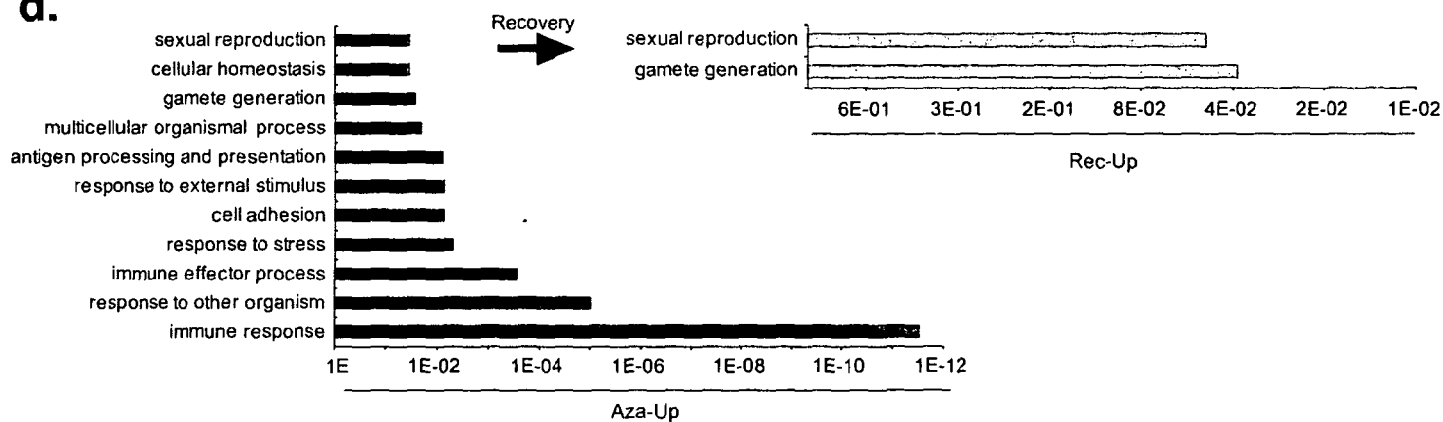
c.



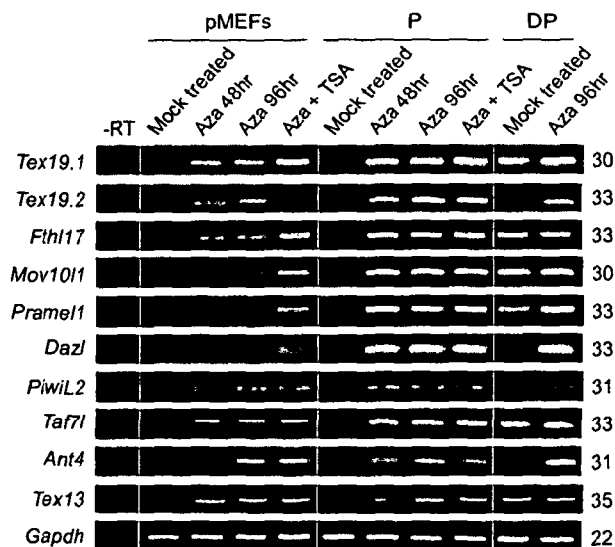
e.



d.



f.



g.

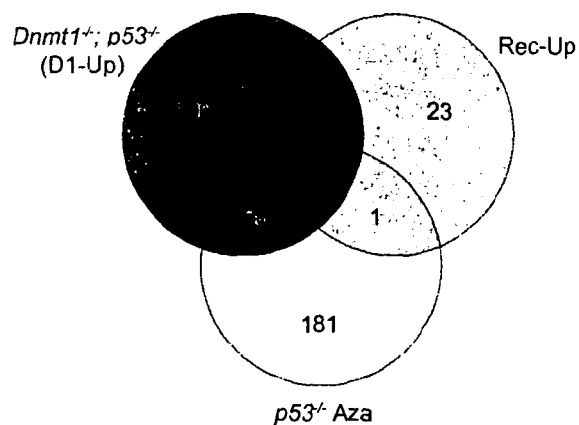
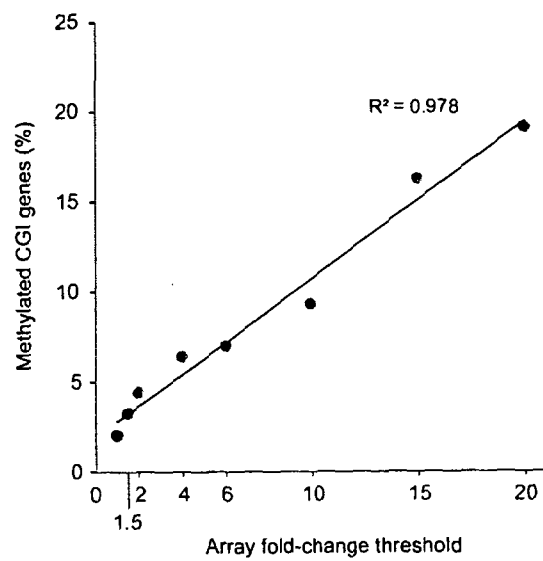


FIGURE 1

a.



b.

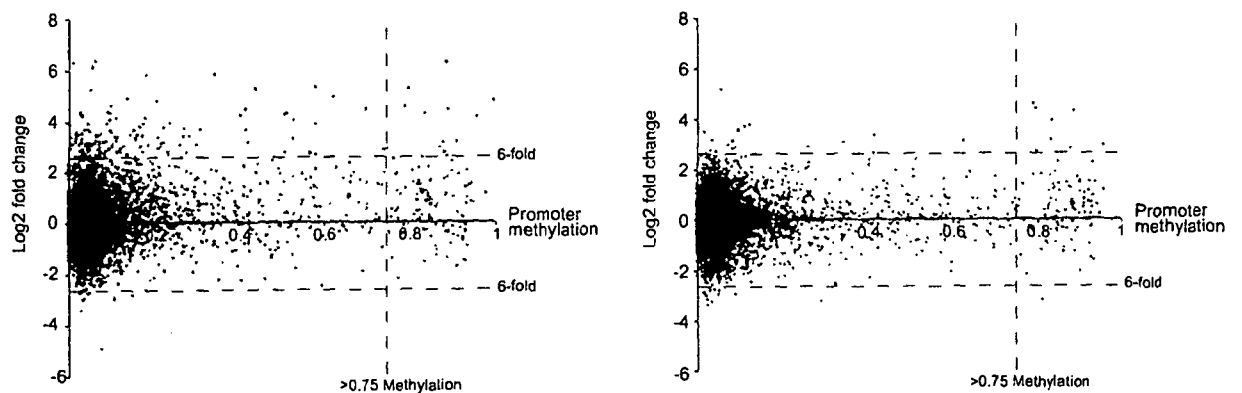


FIGURE 2

FIGURE 3

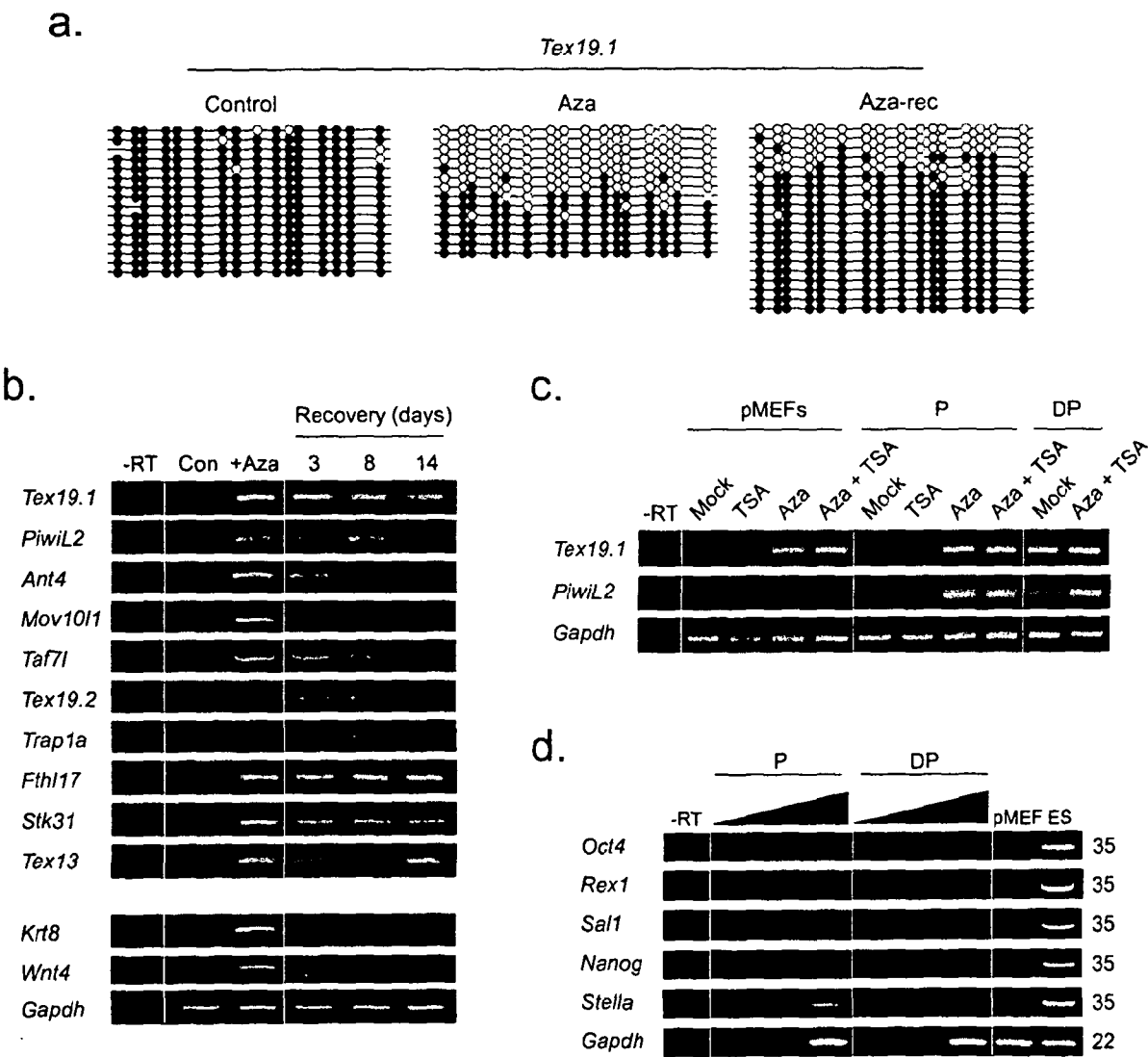


FIGURE 4

Gene	Fold Change			Tissue-specificity
	Aza-Rec	Dnmt1 ^{-/-}	p53 ^{-/-} Aza	
<i>Tex19.1</i>	36.1	430	549	Testis
<i>Rhox5</i>	49.0	306	54.1	Testis/Placenta
<i>Mabp</i>	15.3	93.1	9.2	Testis
<i>Tex13</i>	14.9	48.1	32.8	Testis
<i>Mob101</i>	18.2	31.2	22.0	Testis
<i>Rps4y2</i>	7.9	30.0	83.3	Testis
<i>Ptfr</i>	6.1	19.7	63.1	Testis
<i>Rhox4d</i>	6.5	16.9	17.3	Testis/Placenta
<i>Tex19.2</i>	6.2	6.1	34.7	Testis
<i>Tex19.2</i>	15.2	10.8	56.9	Testis
<i>Id4</i>	15.6	24.4	13.5	Testis
<i>Slc15a3</i>	6.1	9.5	12.9	Semi-Ub
<i>Gsta2</i>	8.7	27.2	37.5	Liver
<i>Akr1c12</i>	8.9	14.9	29.9	Liver
<i>Slc12a1</i>	28.5	207	205	Kidney
<i>Gpr97</i>	10.8	45.0	52.2	Blood/Placenta
<i>Xim2</i>	35.1	19.7	24.7	Blood/Placenta

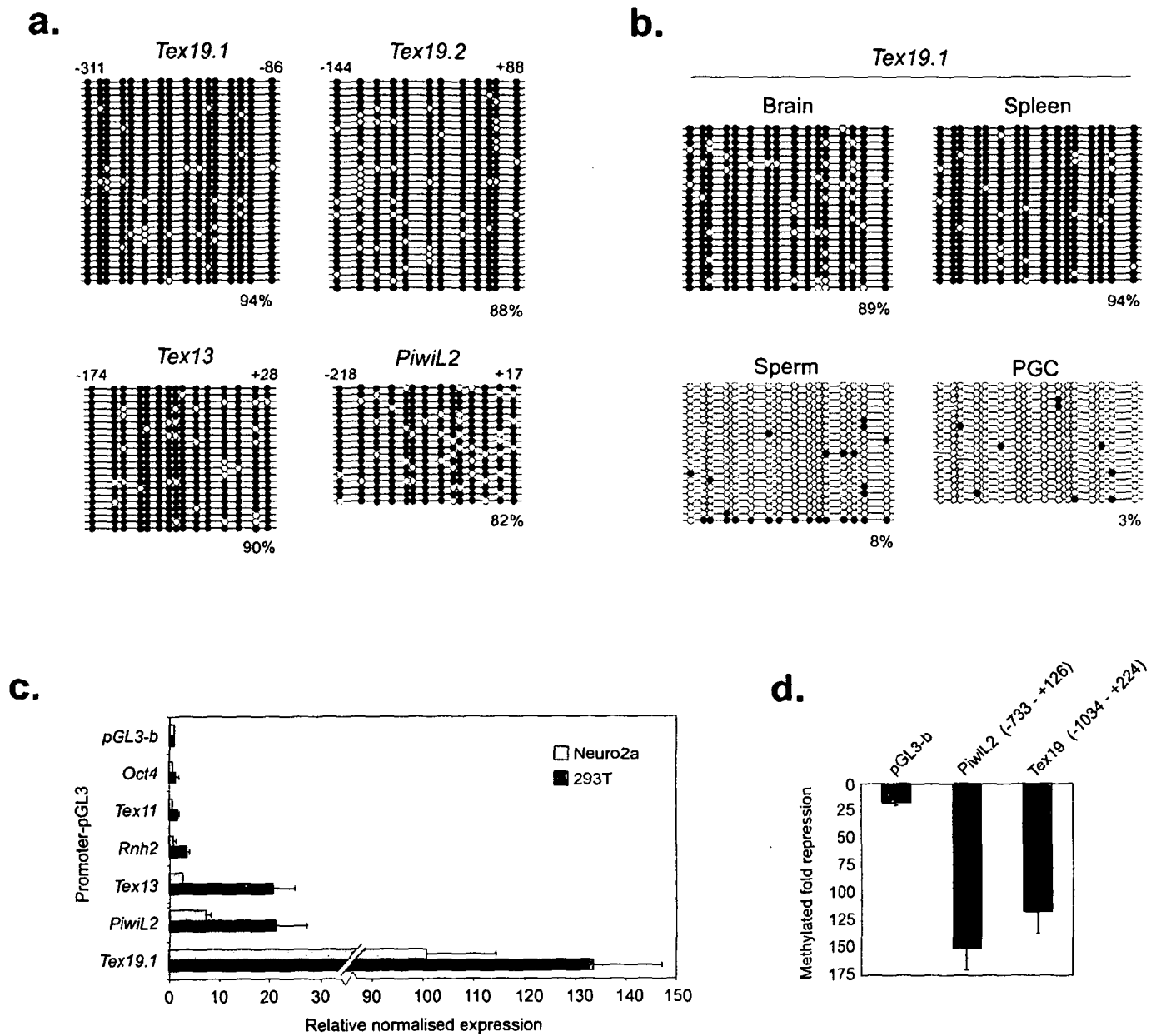
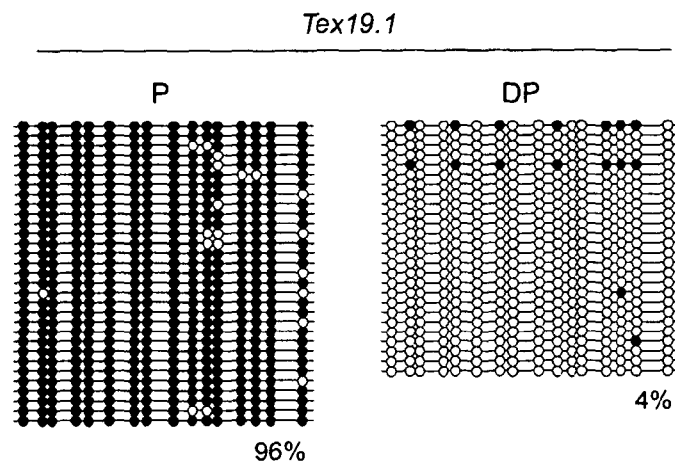


FIGURE 5

a.



b.

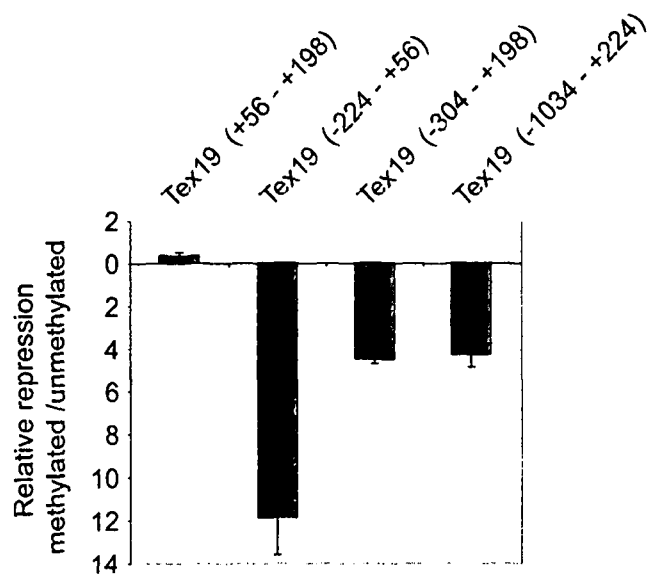
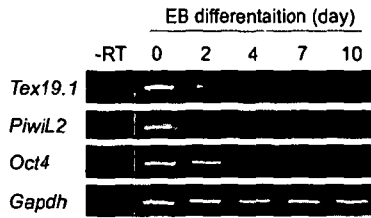
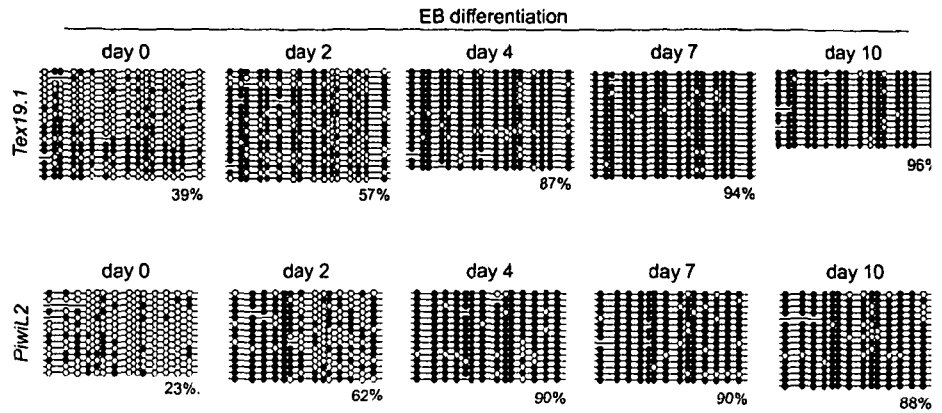


FIGURE 6

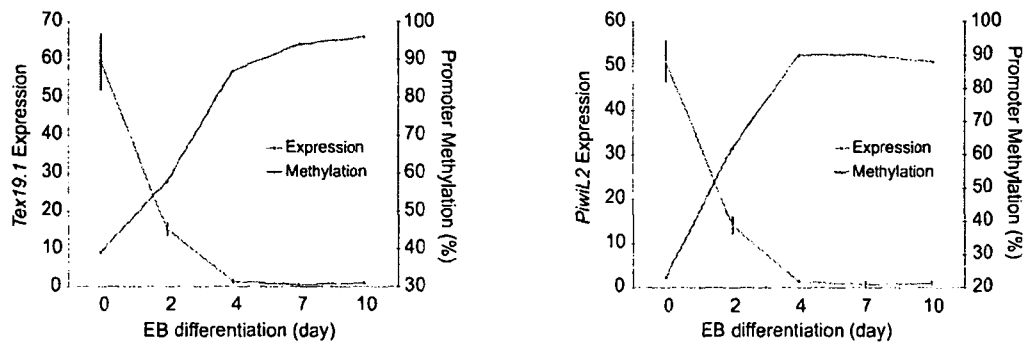
a.



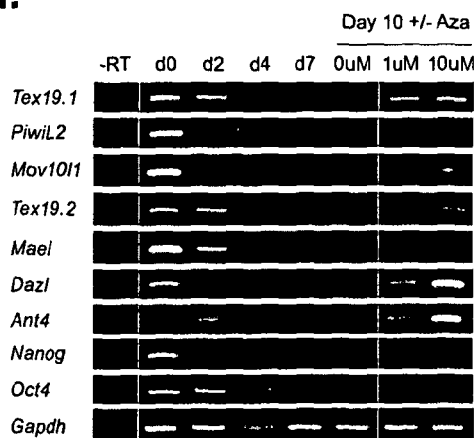
b.



c.



d.



e.



f.

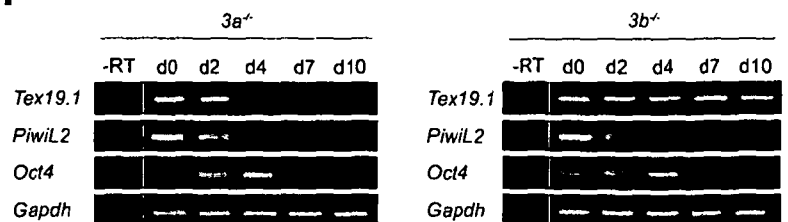
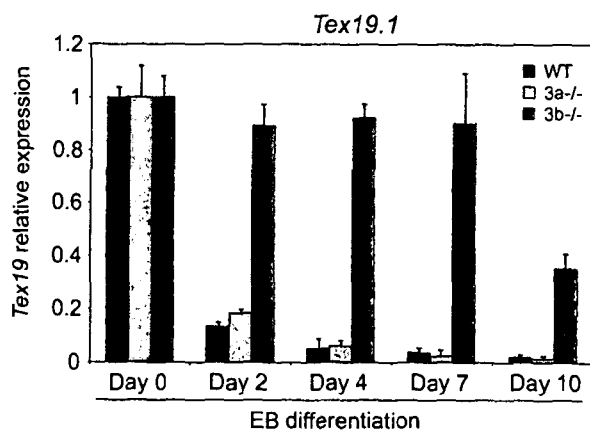
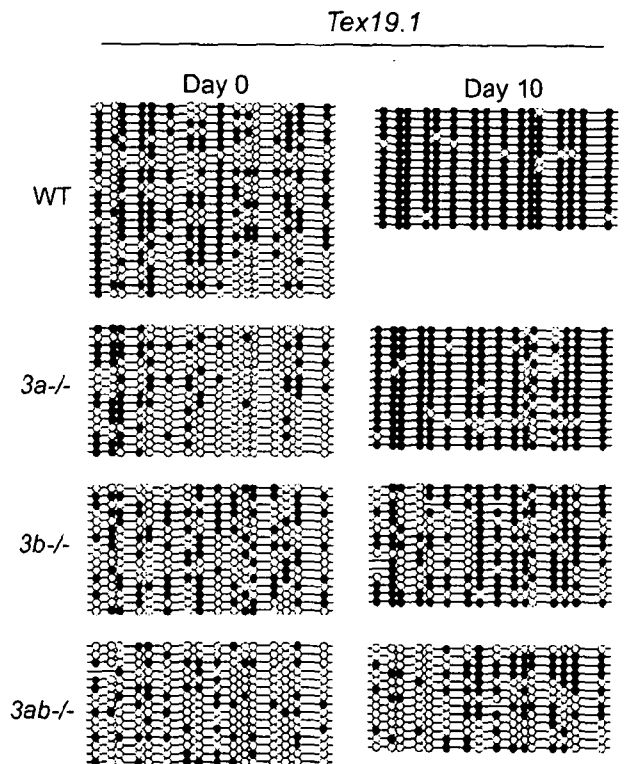


FIGURE 7

a.



b.



c.

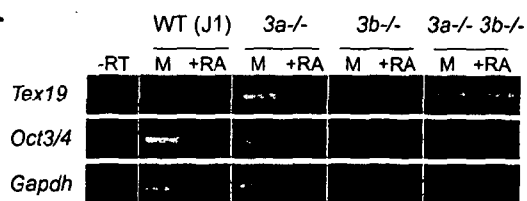


FIGURE 8

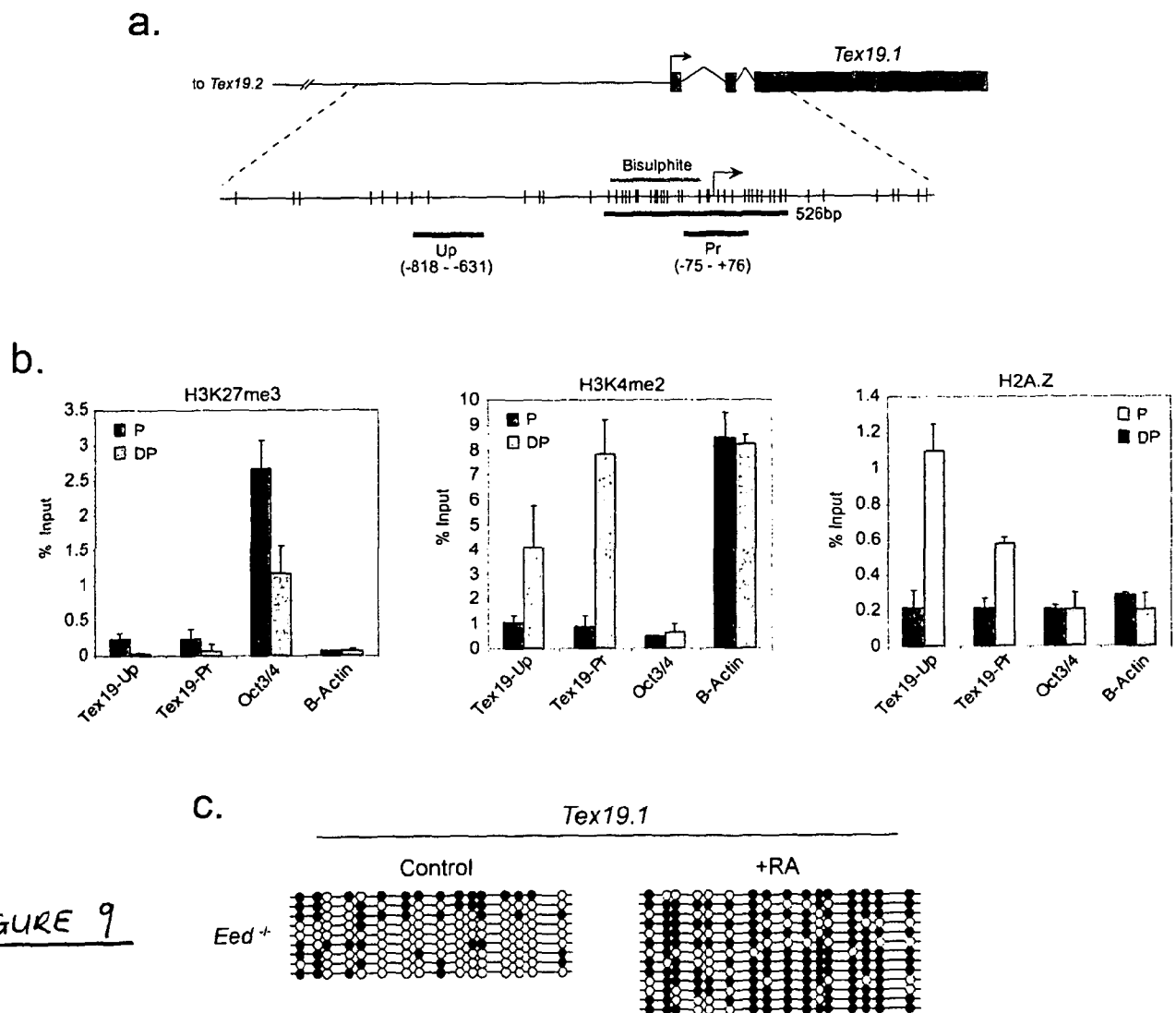


FIGURE 9

10/21

WO 2012/032287

PCT/GB2011/001305

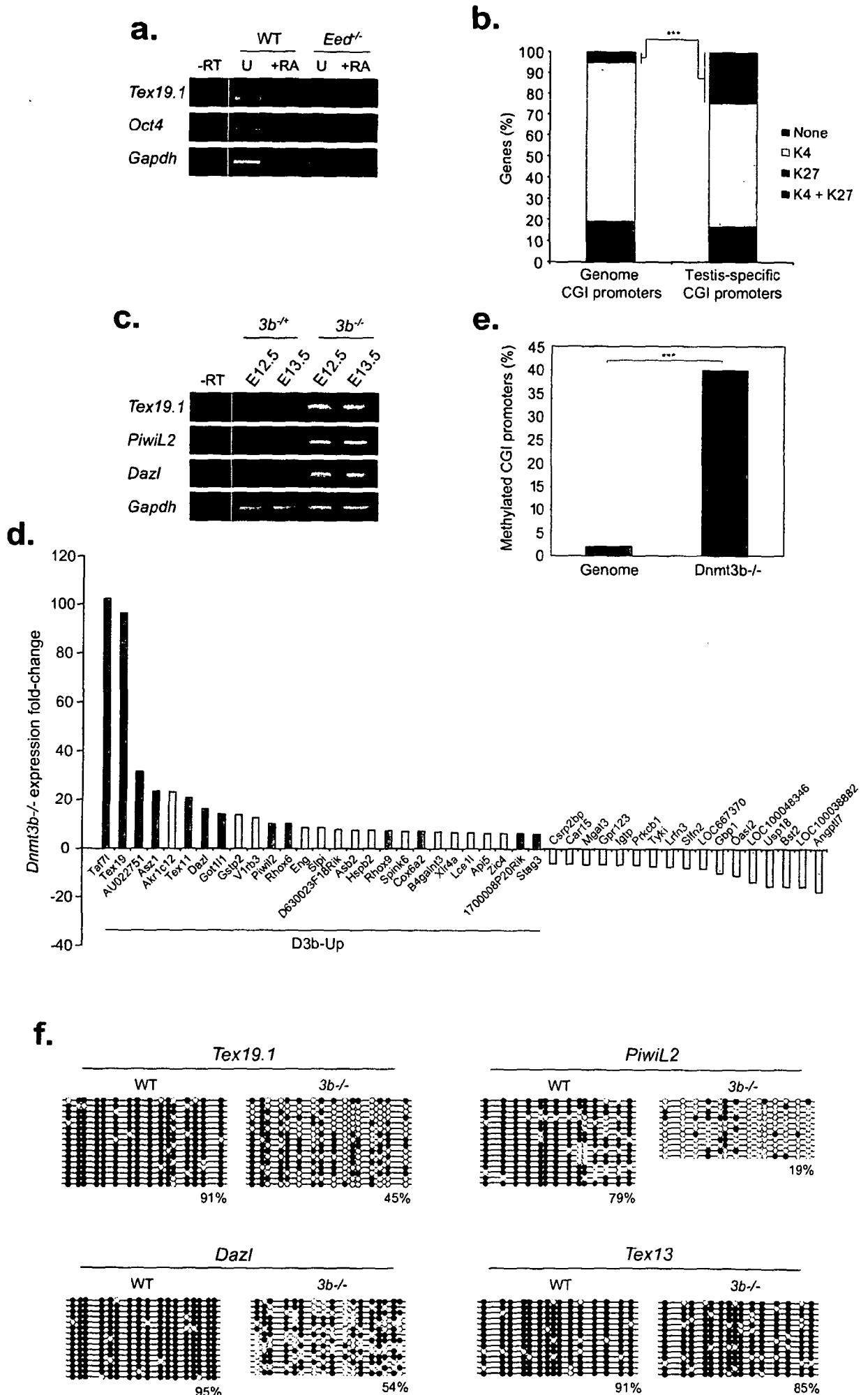


Fig. 10

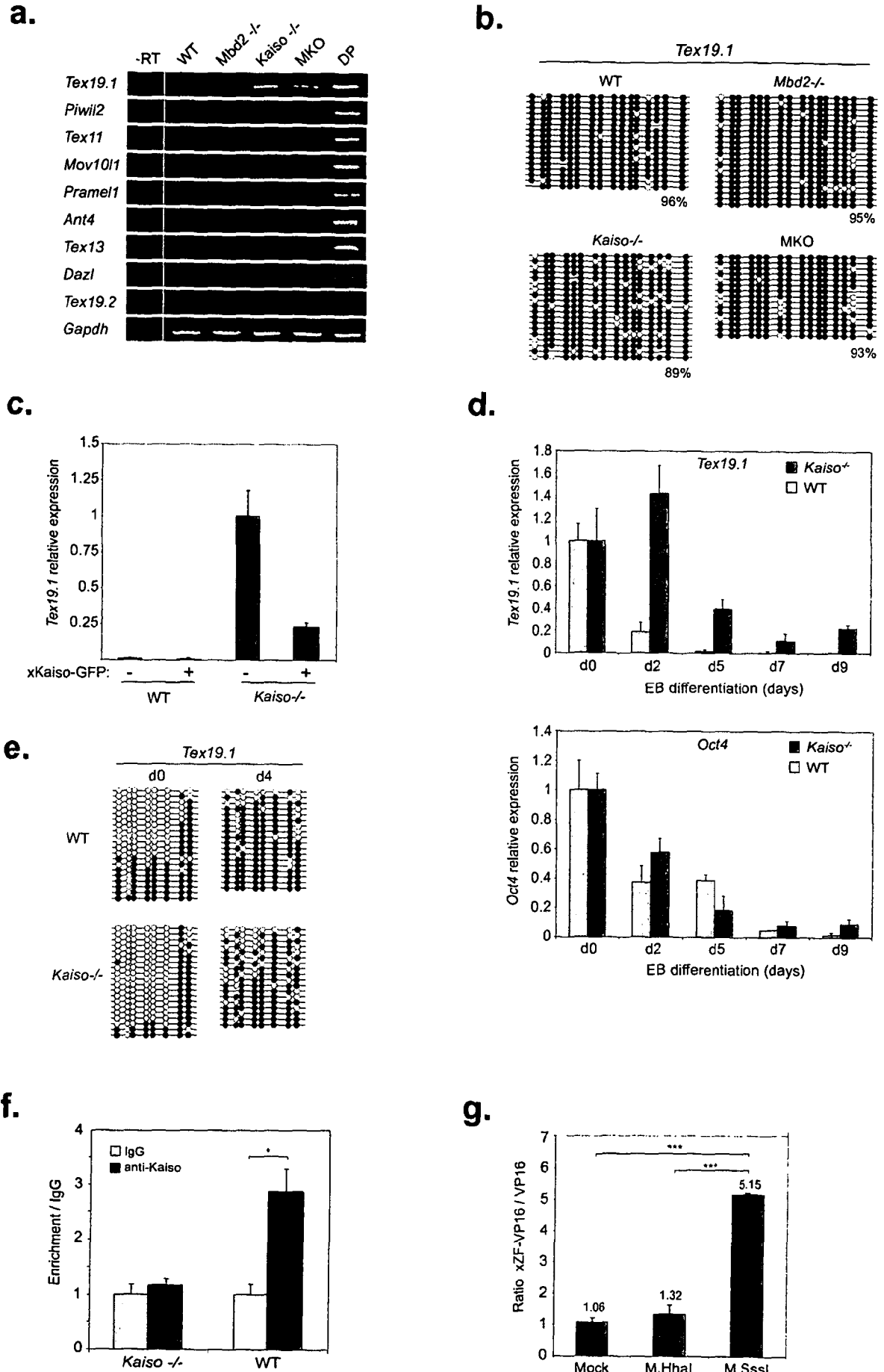
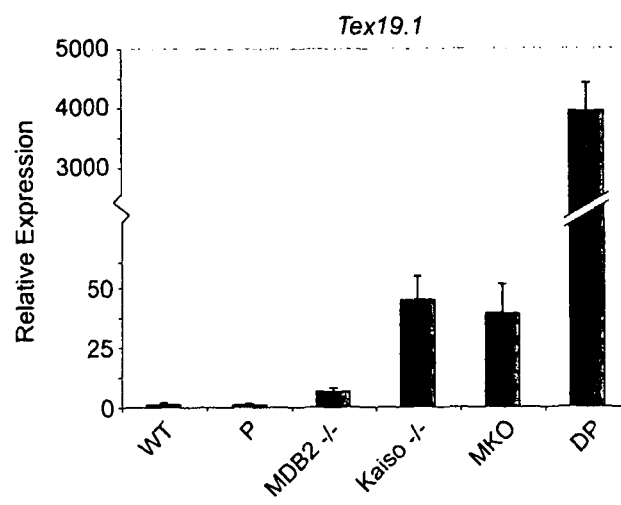


FIGURE 11

a.



b.

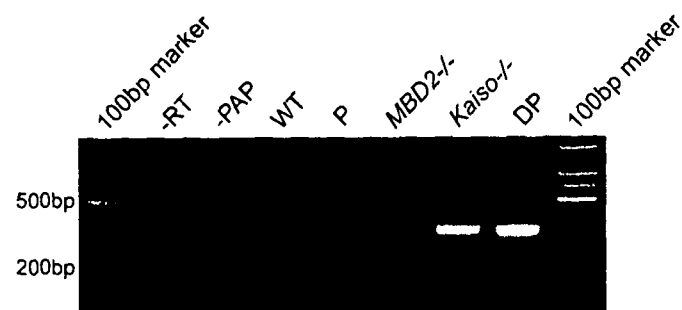
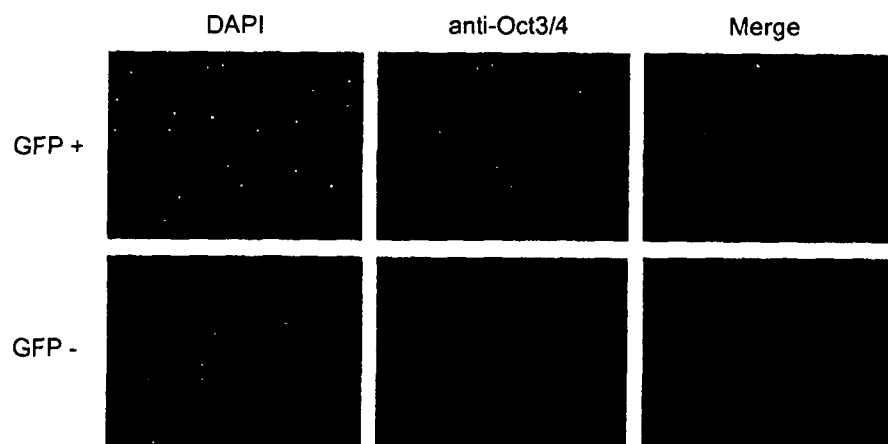


FIGURE 12

a.



b.

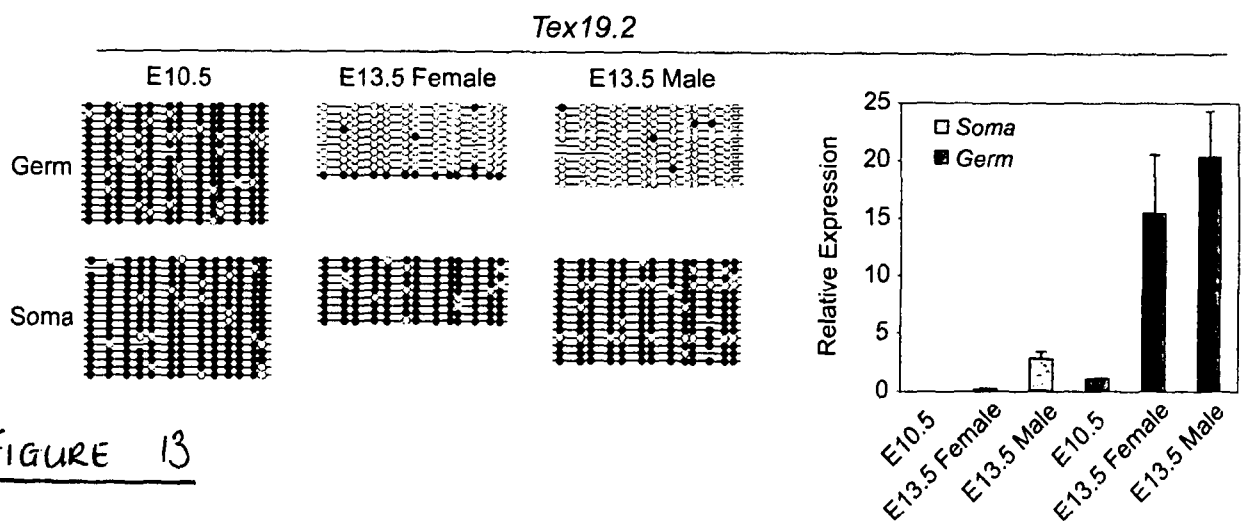


FIGURE 13

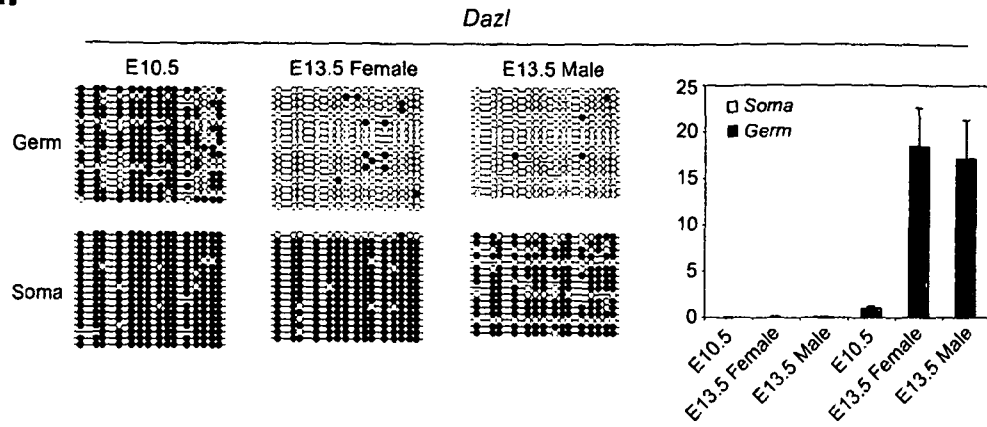
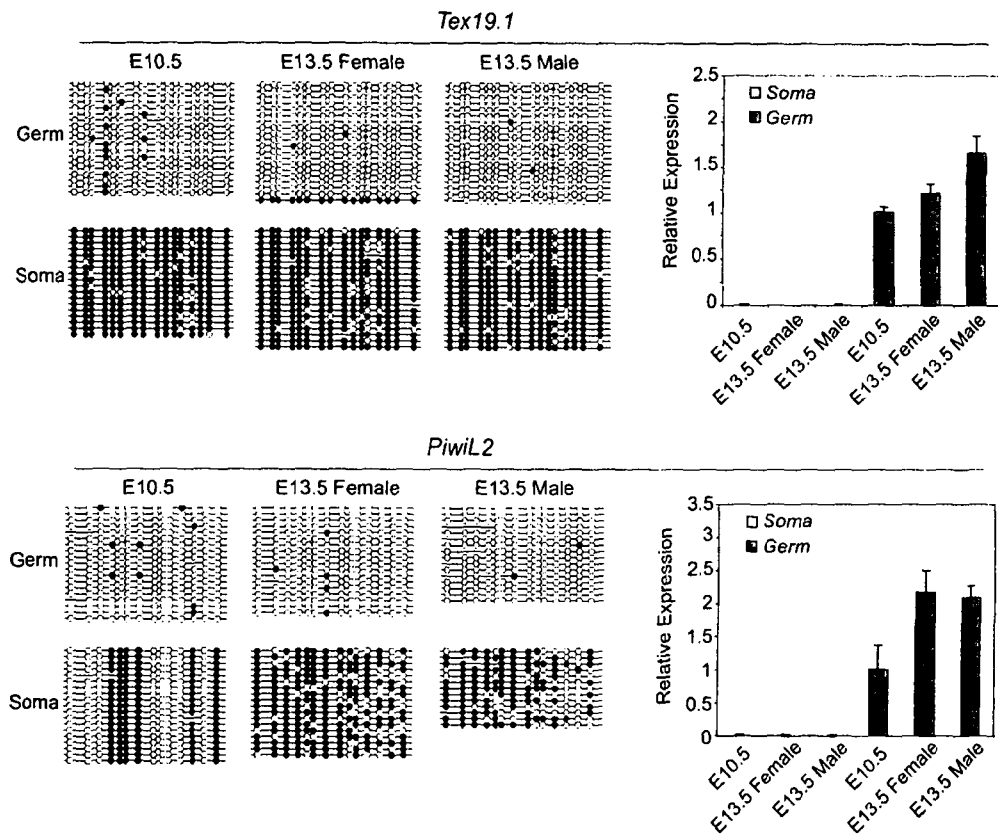
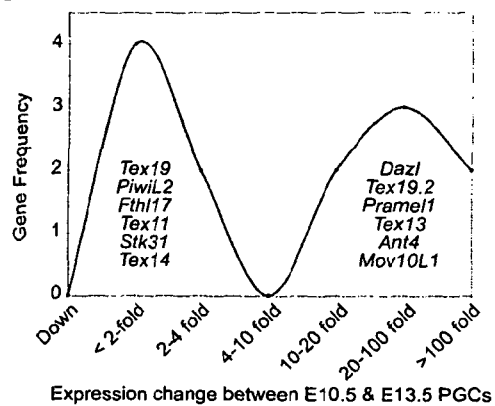
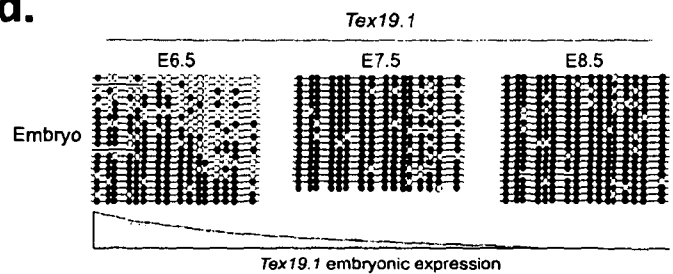
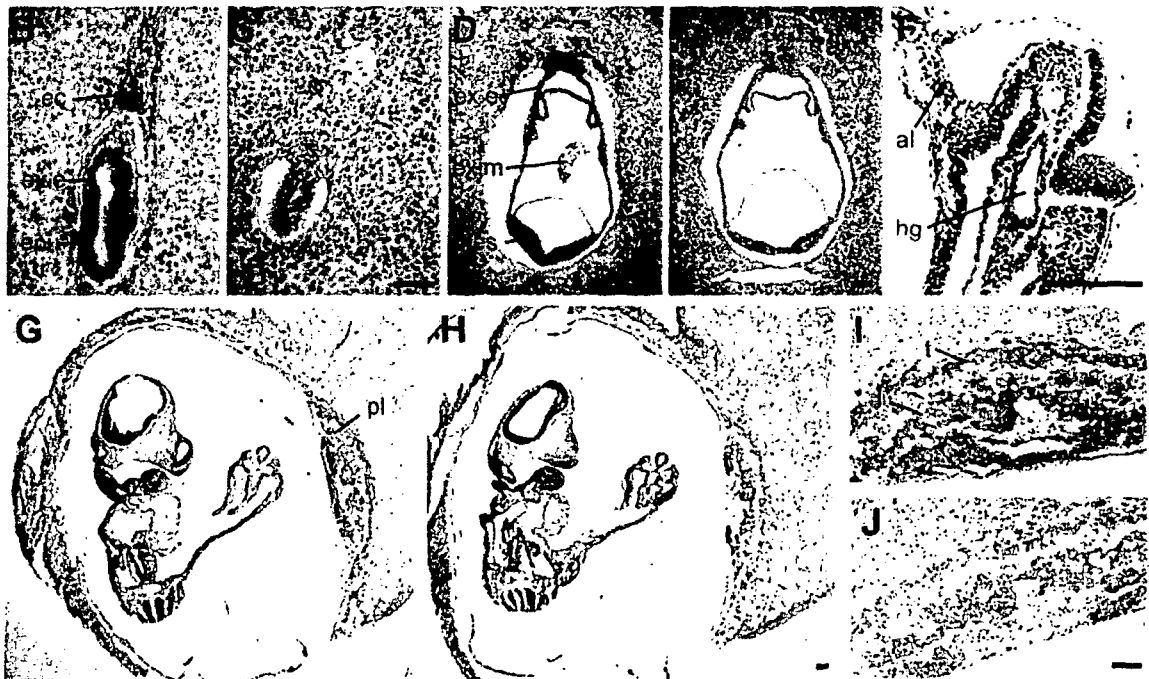
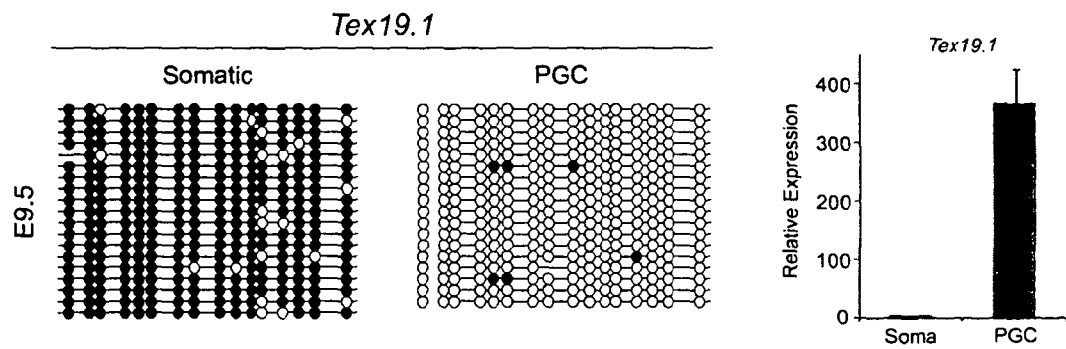
a.**b.****c.****d.**

FIGURE 14

FIGURE 15

a.



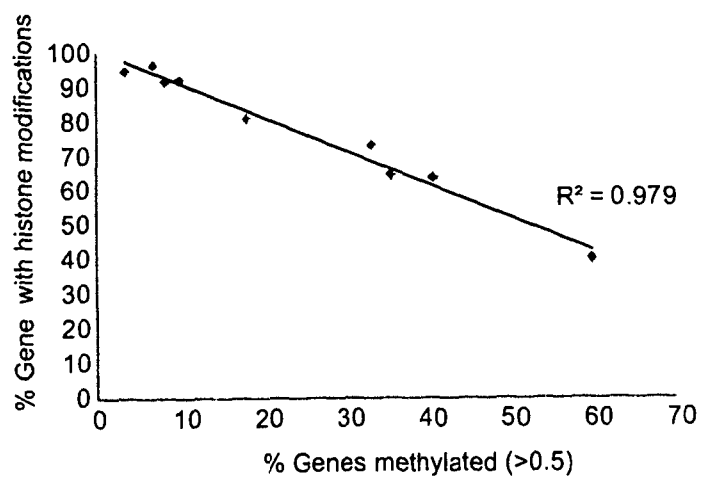


FIGURE .16

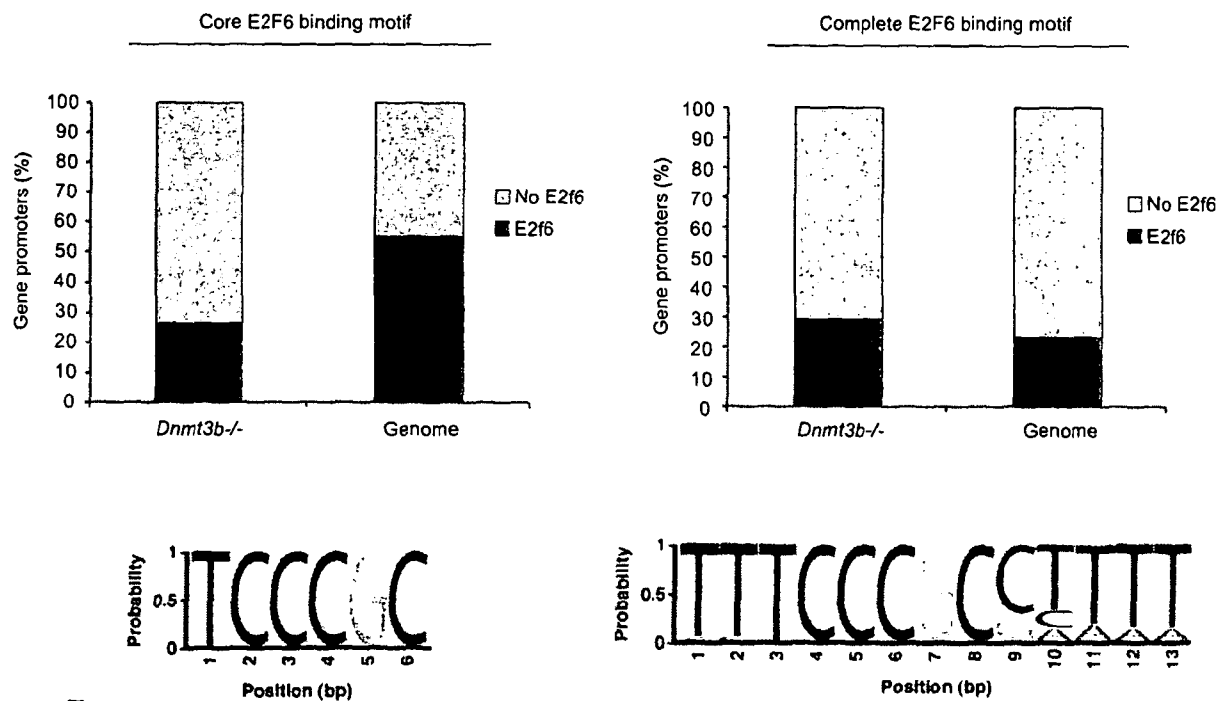
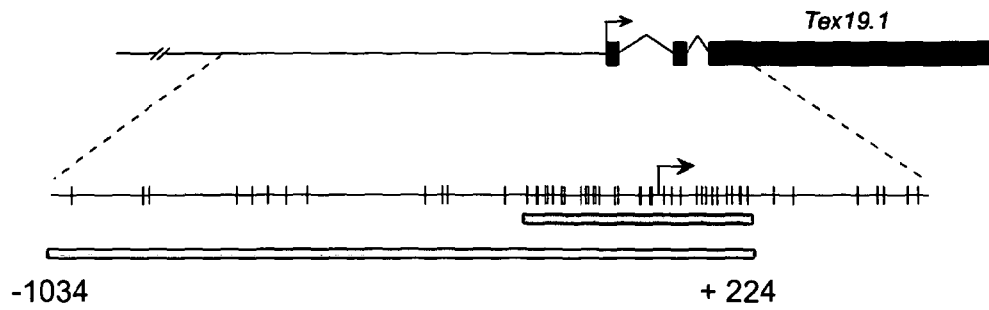
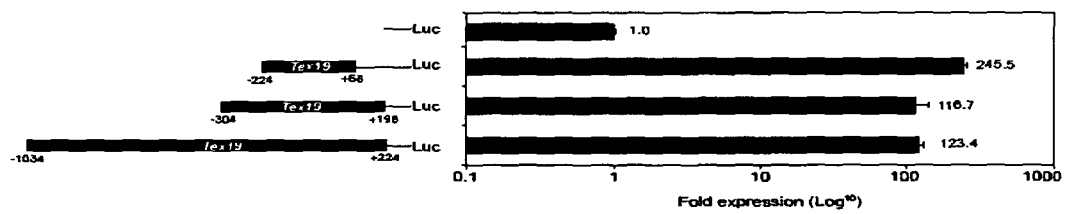


FIGURE 17

18/21

**FIGURE 18****FIGURE 19**

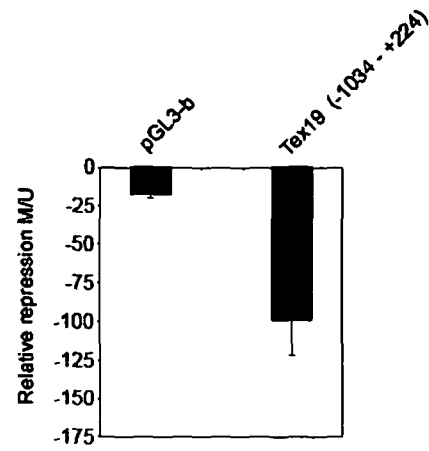


FIGURE 20

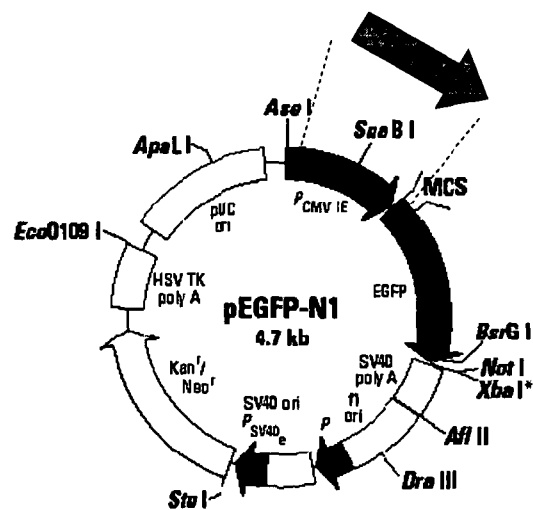
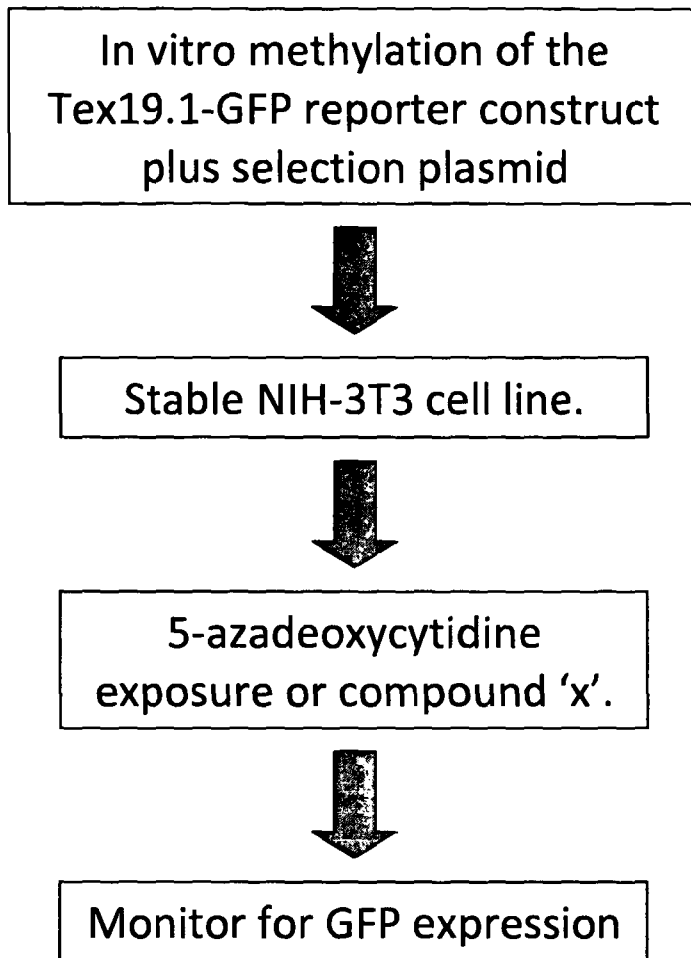


FIGURE 21

20/21

**FIGURE 22**

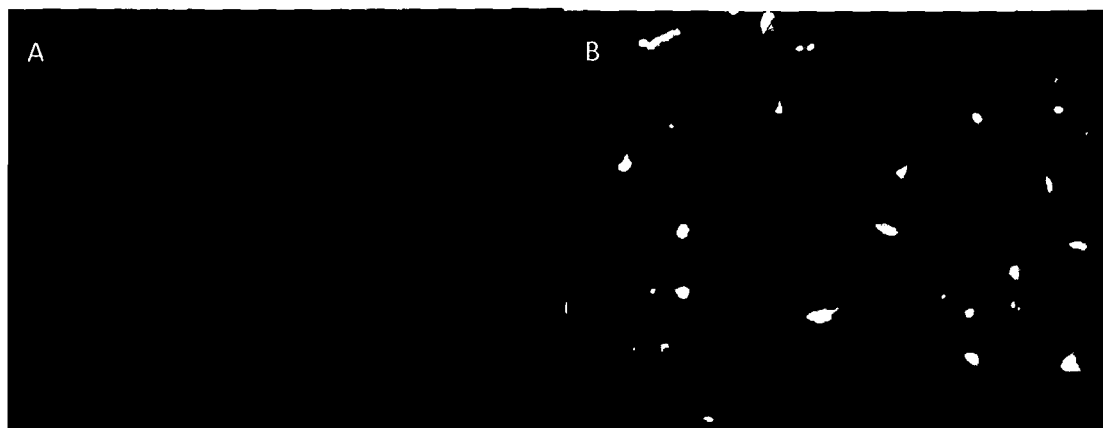


FIGURE 23