LECTURE 4

DNA METHYLATION

DNA methyl transferases methylate DNA

Maintenance of DNA methylation patterns by DNMTs

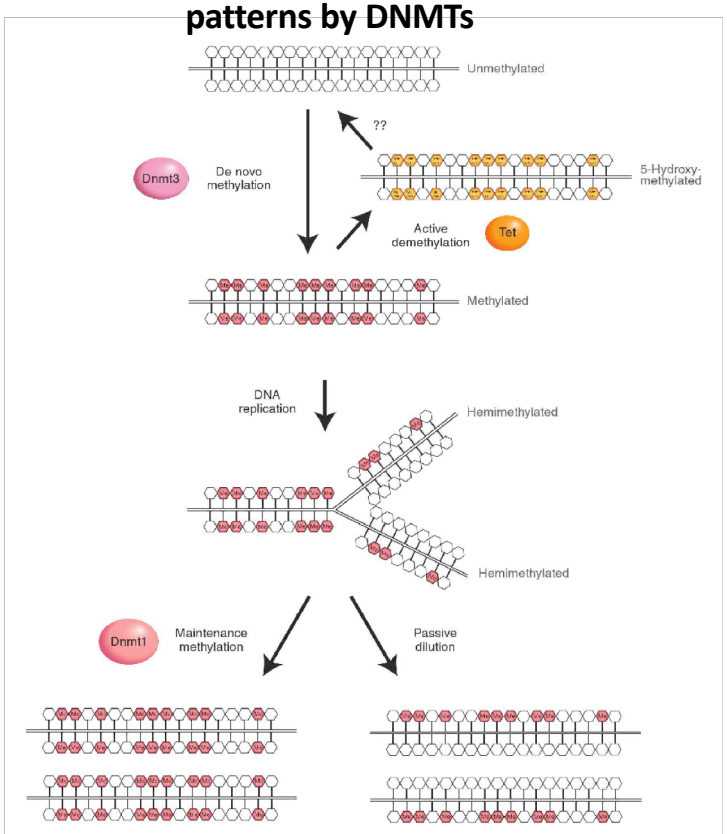


Figure 2. De novo methylation and maintenance methylation of DNA. A stretch of genomic DNA is shown as a line with self-complementary CpG pairs marked as vertical strokes. Unmethylated DNA (*top*) becomes methylated "de novo" by Dnmt3a and Dnmt3b to give symmetrical methylation at certain CpG pairs. On semiconservative DNA replication, a progeny DNA strand is base-paired with one of the methylated parental strands (the other replication product is not shown). Symmetry is restored by the maintenance DNA methyltransferase, Dnmt1, which completes half-methylated sites, but does not methylate unmodified CpGs.

Discovery of function and DNMT family members:

DNMT1: discovered first

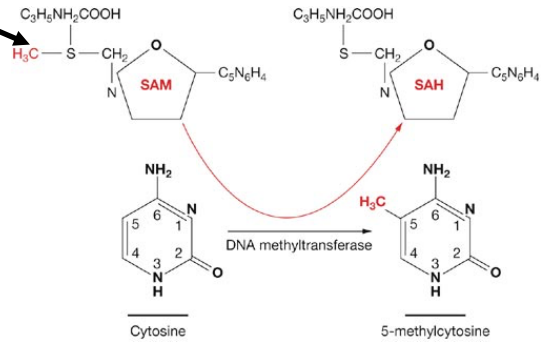
Cell extract + DNA containing CpG repeats + ^{14}C labelled -CH3 in AdoMet (SAM) → radioactive -CH3 transferred to DNA

Next step: Purification of enzymatic activity from cell extract → 200kDa complex containing a protein with specific DNA methyl transferase activity: **DNMT1**

Biochemical characterization of substrate specificity:

This enzyme is 7– to 100-fold more active on hemimethylated DNA as compared with un-methylated substrate *in vitro*

^{14}C



Discovery of de novo DNMTs:

Sequence of DNMT1 was used to look For genes with similar sequence (sequence homology)

→ Discovery of de-novo DNMTs that work efficiently work on un-methylated DNA (DNMT3a, 3b)

→ De-novo DNMTs cannot efficiently methylate hemi-methylated DNA

LOSS OF DNA METHYLTRANSFERASES IS LETHAL DURING EMBRYONIC MOUSE DEVELOPMENT

Table 1. Function of mammalian DNA methyltransferases

DNA methyltransferase	Species	Major activity	Major phenotypes of loss of function
Dnmt1	Mouse	Maintenance methylation of CpG	Genome-wide loss of DNA methylation, embryonic lethality at embryonic day 9.5 (E9.5), abnormal expression of imprinted genes, ectopic X-chromosome inactivation, activation of silent retrotransposon. In cancer cell lines, it leads to cell cycle arrest and mitotic defects.
Dnmt3a	Mouse	De novo methylation of CpG	Postnatal lethality at 4–8 wk, male sterility, and failure to establish methylation imprints in both male and female germ cells
Dnmt3b	Mouse	De novo methylation of CpG	Demethylation of minor satellite DNA, embryonic lethality around E14.5 days with vascular and liver defects. (Embryos lacking both Dnmt3a and Dnmt3b fail to initiate de novo methylation after implantation and die at E9.5.)
DNMT3B	Human	De novo methylation of CpG	ICF syndrome: immunodeficiency, centromeric instability, and facial anomalies. Loss of methylation in repetitive elements and pericentromeric heterochromatin.

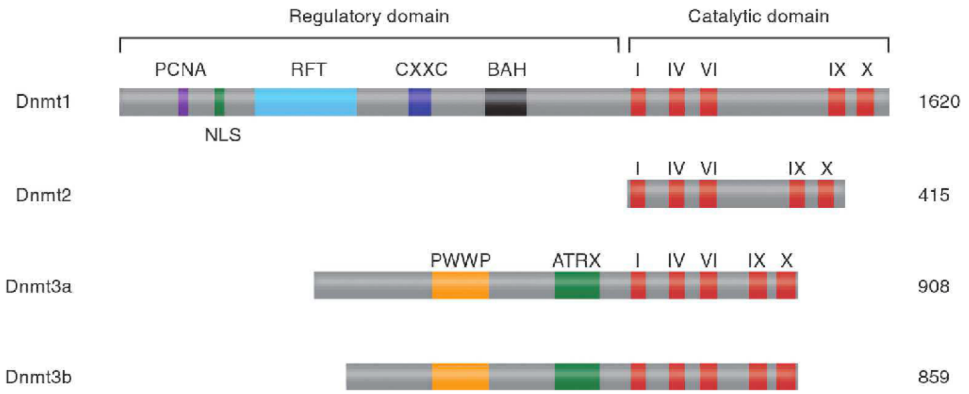
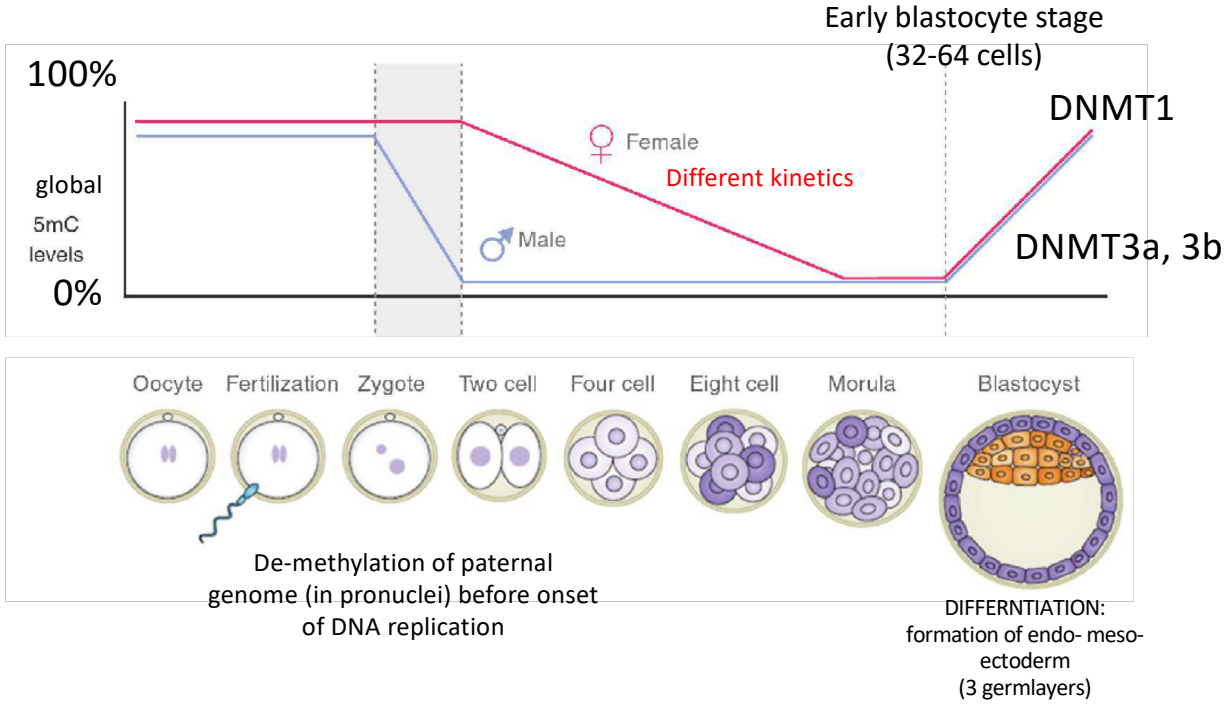


Figure 4. Mammalian DNA methyltransferases. The catalytic domains of Dnmt1, Dnmt2, and the Dnmt3 family members are conserved (the signature motifs, I, IV, VI, IX, and X, are most conserved in all cytosine methyltransferases), but there is little similarity among their amino-terminal regulatory domains. Domain abbreviations: PCNA, PCNA-interacting domain; NLS, nuclear localization signal; RFT, replication foci-targeting domain; CXXC, a cysteine-rich domain implicated in binding DNA sequences containing CpG dinucleotides; BAH, bromo-adjacent homology domain implicated in protein–protein interactions; PWWP, a domain containing a highly conserved "proline-tryptophan-tryptophan-proline" motif involved in heterochromatin association; ATRX, an ATRX-related cysteine-rich region containing a C2-C2 zinc finger and an atypical PHD domain implicated in protein–protein interactions.

DNA METHYLATION IS ABUNANT IN THE GENOME AND IS SUBJECTED TO DRAMATIC ALTERATIONS DURING EMBRYOGENESIS



70%- 80% of CpG dinucleotides are methylated in the genome

DNA methylation levels are high in fertilized Oocytes that contain the paternal and maternal genome (carries characteristic methylation patterns)

Paternal and maternal methylation patterns are rapidly erased (exception: imprinted genes maintain paternal and maternal methylation information). → the paternal and maternal methylation epigenome is cancelled

DNA methylation levels remain low during the first cell division events until the blastocyst stage

In the blastocyst stage cell differentiation programs are activated and genes need to be regulated on the epigenetic level → DNA methylation is increasing (loss of DNMT1, DNMT3a or DNMT3b is lethal → establishment and maintenance of DNA methylation is impaired)

70%- 80% of CpG di-nucleotides are methylated in the human genome!
Remember only 2% of the genome encode for mRNAs
98% is noncoding DNA that contains a large proportion of transposable elements, repeat sequences, etc…

ON THE SINGLE GENE LEVEL:
CpG islands (CGIs) are short sequences stretches with variable DNA methylation that regulate promoter activity

NOTE: single CpGs are generally hyper-methylated (60-90%)
CpG islands are differentially methylated, but are generally demethylated

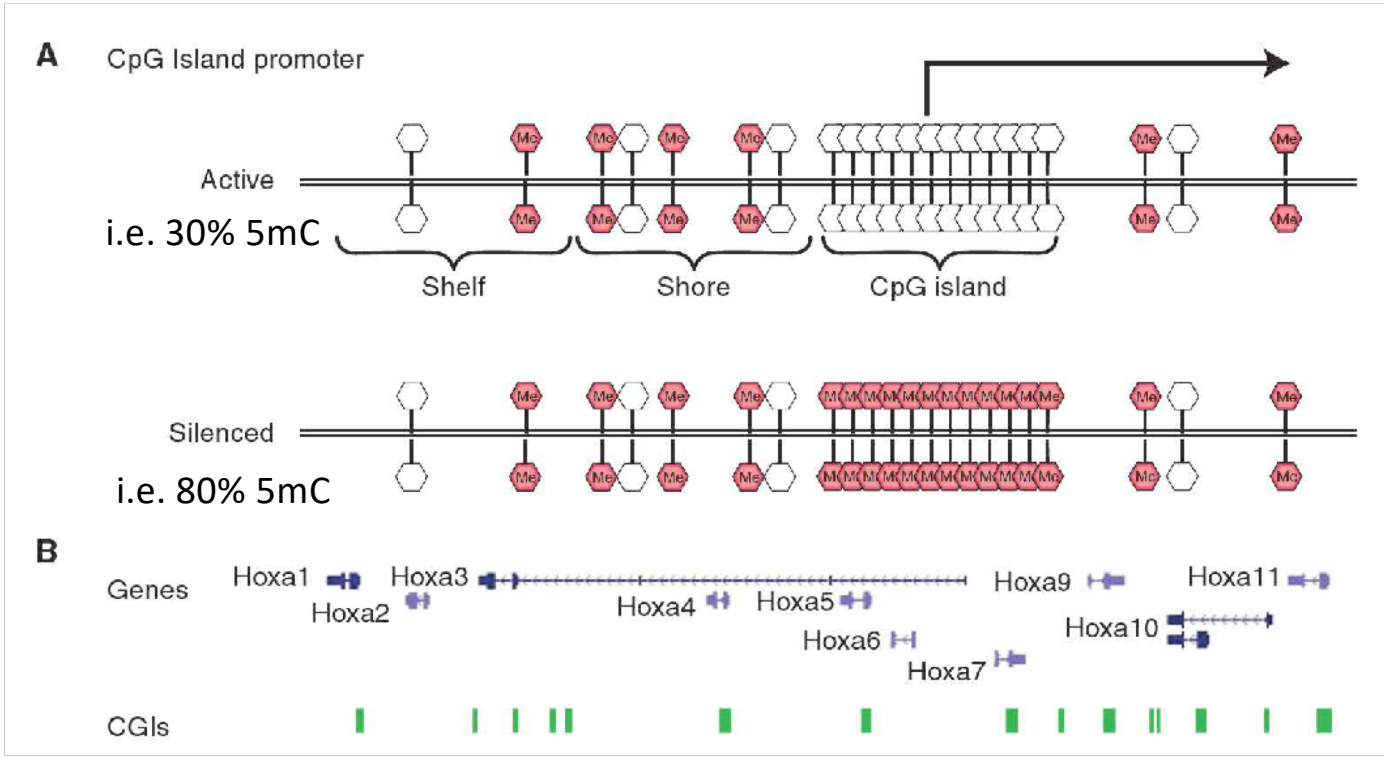
CpG islands (CGIs) have a length of ca. 1kb

60% of human genes are controlled by CGIs containing promoters that allow tissue/cell specific gene expression

CpG islands can overlap with the first exon (methylation level in 1st exon is good predictor of gene expression

CpG islands located <2kb from promoter: shores

CpG islands located <2-4kb from promoter: shores



TRANSCRIPTIONAL REGULATION BY METHYL-DNA BINDING PROTEINS

Interference with transcription factor binding

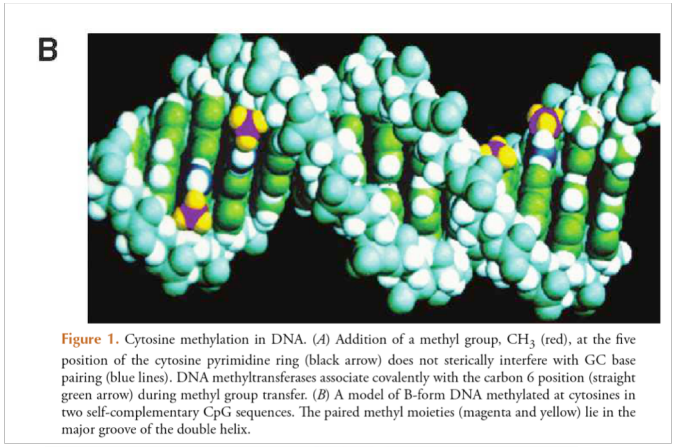
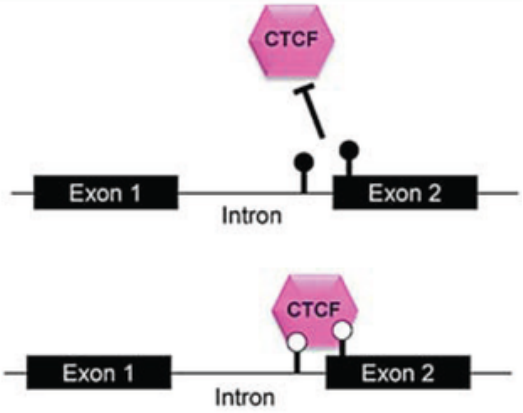


Figure 1. Cytosine methylation in DNA. (*A*) Addition of a methyl group, CH_3 (red), at the five position of the cytosine pyrimidine ring (black arrow) does not sterically interfere with GC base pairing (blue lines). DNA methyltransferases associate covalently with the carbon 6 position (straight green arrow) during methyl group transfer. (*B*) A model of B-form DNA methylated at cytosines in two self-complementary CpG sequences. The paired methyl moieties (magenta and yellow) lie in the major groove of the double helix.

Methylated DNA obtains different structure:
Transcription factors cannot bind anymore
→ DNA methylation sensitive transcription factors



Example: CTCF
Unmethylated DNA CTCF binds → activation of expression
Methylated DNA: CTCF does not bind → no activation

Note: CTCF is a major epigenetic regulator that is involved in controlling genomic imprinting, enhance activation,…

EPIGENTIC READERS OF DNA METHYLATION

Transcriptional regulation by methyl-DNA binding proteins

Table 2. Functions of methyl-CpG binding proteins

MBP	Major activity	Species	Major phenotypes of loss-of-function mutations
MeCP2	Binds mCpG with adjacent run AT-rich run Transcriptional repressor	Mouse	Delayed onset neurological defects including inertia, hind-limb clasping, nonrhythmic breathing, and abnormal gait. Postnatal survival ~10 wk.
MECP2	Binds mCpG with adjacent AT run Transcriptional repressor	Human	Heterozygotes suffer from Rett syndrome, a profound neurological disorder characterized by apraxia, loss of purposeful hand use, breathing irregularities, and microcephaly
Mbd1	Binds mCpG via MBD; a major splice form is also able to bind CpG via a CxxC domain	Mouse	No overt phenotype, but subtle defects in neurogenesis detected
Mbd2	Binds mCpG Transcriptional repressor	Mouse	Viable and fertile, but show reduced maternal nurturing behavior. Defective gene regulation in T-helper cell differentiation leading to altered response to infection. Highly resistant to intestinal tumorigenesis.
Mbd3	Core component of NuRD corepressor complex Does not show strong binding to mCpG	Mouse	Early embryonic lethal
Mbd4	DNA repair protein that binds mCpG and T:G mismatches at mCpG sitesThymine DNA glycosylase that excises T from T:G mismatches	Mouse	Viable and fertile. three- to fourfold increase in mutations at CpG sites. Increased susceptibility to intestinal cancer correlates with C to T transitions within the *Apc* gene. Mbd4 functions to minimize the mutability of 5-methylcytosine.
Kaiso	Binds mCGmCG and CTGCNA Transcriptional repressor	Mouse	No overt phenotype. Small but significant delay in tumorigenesis on Min background.

Several proteins were identified to have affinity to methylated CpG
but do no have affinity to un-methylated CpG → mediate transcriptional silencing

→ CpG METHYL BINDING DOMAIN PROTEIN (MBD) FAMILY : MeCP1, MeCP2, Mbd1, Mbd2, Mbd2, Mbd4

→ Kaiso (unrelated protein)

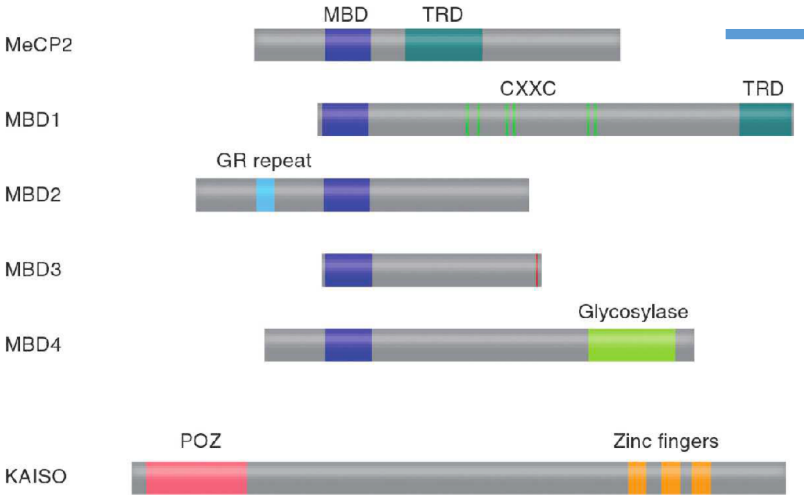
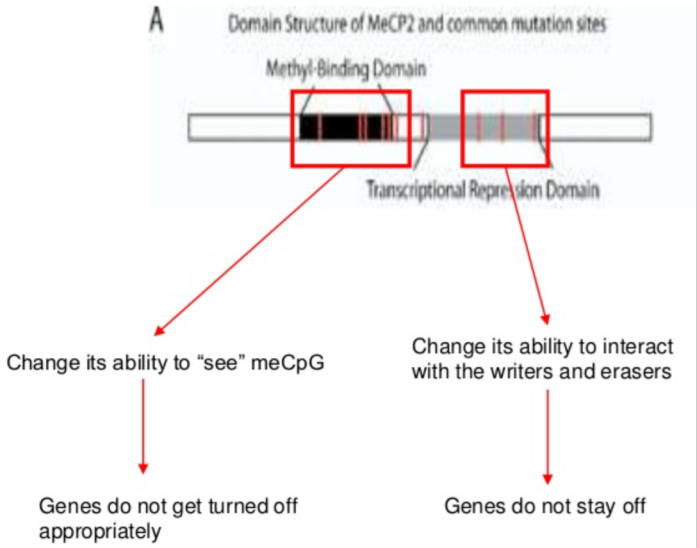


Figure 8. Proteins that bind methyl-CpG. Five members of the MBD protein family are aligned at their MBD domains (purple). Other domains are labeled and include TRD; CXXC domains, which are zinc fingers, some of which are implicated in binding to nonmethylated CpG; GR repeats that may bind; a T:G mismatch glycosylase domain that is involved in repair of 5mC deamination. Kaiso lacks the MBD domain, but binds methylated DNA via zinc fingers (orange) and possesses a POB/BTB domain that is shared with other transcriptional repressors. Domain abbreviations: MBD, methyl-CpG binding domain; TRD, transcriptional repression domain; POZ, poxvirus and zinc finger, a protein–protein interacting domain.

→ **RETT SYNDROME**

How does MeCP2 effect the brain function?

- Through it's job as a reader of epigenetic bookmarks
- The wide array of functions that MeCP2 performs ALL contribute to Rett syndrome.
- The different mutations have different effects on the presentation of the disease.
- In addition since each person is different based on their personal epigenetics, the disease will be individual as well.



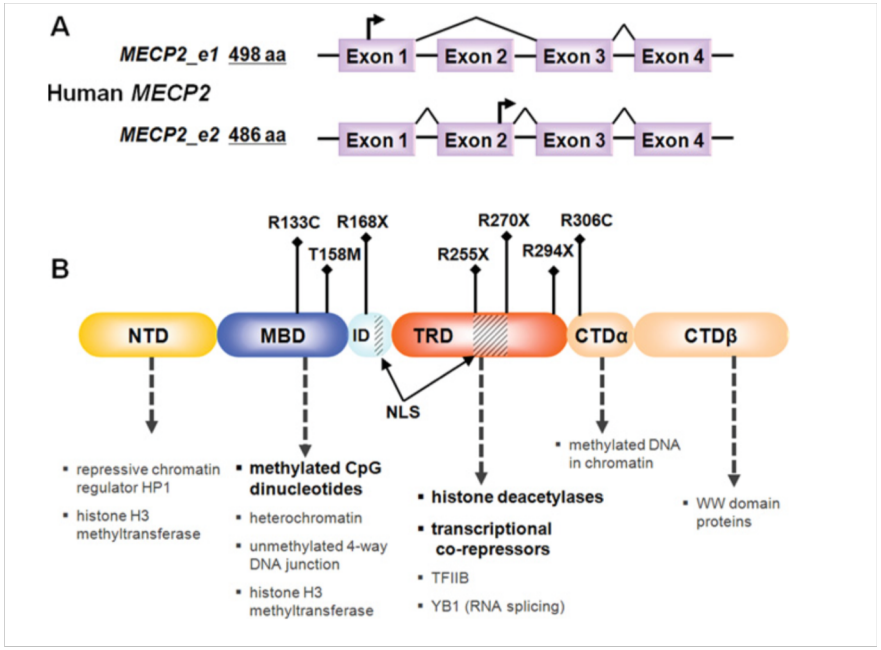


Figure 1 Composition of MeCP2: gene structure, splicing patterns and putative functional domains

(**A**) Splicing patterns generating the two mRNA isoforms of *MECP2*, _e1 and _e2. The two isoforms generate two protein isoforms of MeCP2 with differing N-termini due to the use of alternative translation start sites (bent arrows) and the absence or presence of exon 2 in the transcript. (**B**) Apart from the N-terminus, both MeCP2 isoforms are identical and contain several functionally distinct domains: NTD, N-terminal domain; MBD, methylated DNA-binding domain; ID, interdomain; TRD, transcription repression domain; CTD, C-terminal domain; NLS; nuclear localization signals. Locations of seven of the most common point mutations in RTT are indicated (◆). Below each domain are indicated major (bold) and other (grey) interactors and functions. HP1, heterochromatin protein 1; TFIIB, transcription factor IIB; YB1, Y-box-binding protein 1.

EPIGENTIC READERS OF DNA METHYLATION

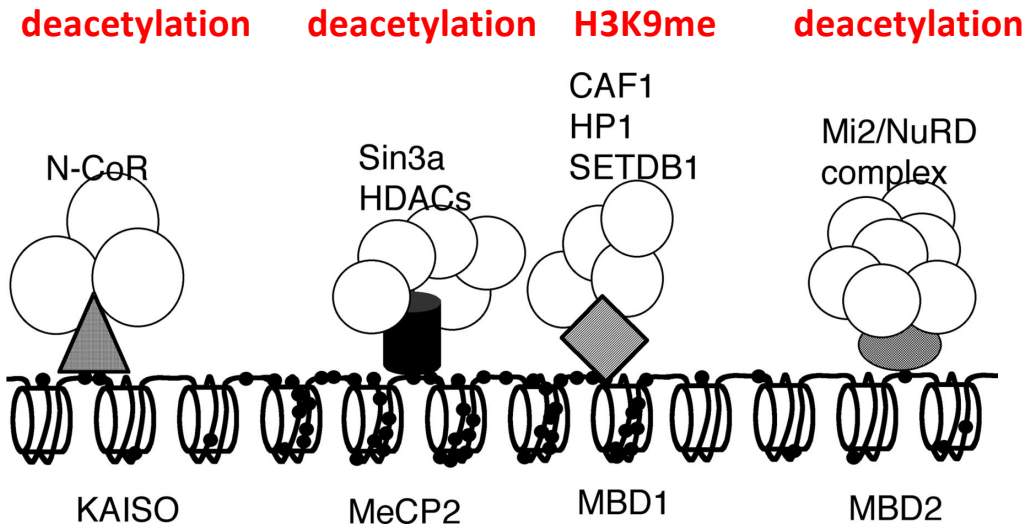
Transcriptional regulation by methyl-dna binding proteins

Table 2. Functions of methyl-CpG binding proteins

MBP	Major activity	Species	Major phenotypes of loss-of-function mutations
MeCP2	Binds mCpG with adjacent run AT-rich run Transcriptional repressor	Mouse	Delayed onset neurological defects including inertia, hind-limb clasping, nonrhythmic breathing, and abnormal gait. Postnatal survival ~10 wk.
MECP2	Binds mCpG with adjacent AT run Transcriptional repressor	Human	Heterozygotes suffer from Rett syndrome, a profound neurological disorder characterized by apraxia, loss of purposeful hand use, breathing irregularities, and microcephaly
Mbd1	Binds mCpG via MBD; a major splice form is also able to bind CpG via a CxxC domain	Mouse	No overt phenotype, but subtle defects in neurogenesis detected
Mbd2	Binds mCpG Transcriptional repressor	Mouse	Viable and fertile, but show reduced maternal nurturing behavior. Defective gene regulation in T-helper cell differentiation leading to altered response to infection. Highly resistant to intestinal tumorigenesis.
Mbd3	Core component of NuRD corepressor complex Does not show strong binding to mCpG	Mouse	Early embryonic lethal
Mbd4	DNA repair protein that binds mCpG and T:G mismatches at mCpG sitesThymine DNA glycosylase that excises T from T:G mismatches	Mouse	Viable and fertile. three- to fourfold increase in mutations at CpG sites. Increased susceptibility to intestinal cancer correlates with C to T transitions within the *Apc* gene. Mbd4 functions to minimize the mutability of 5-methylcytosine.
Kaiso	Binds mCGmCG and CTGCNA Transcriptional repressor	Mouse	No overt phenotype. Small but significant delay in tumorigenesis on Min background.

Methyl-CpG binding proteins are present in transcriptional co-repressor complexes

MeCP2: component of the Sin3A HDAC complex
Mbd3: component of the NuRD HDAC complex
Mbd1: interacts with HDAC3. Mbd1 and HDAC3 are recruited by the PML-RARalpha hybrid protein to silence gene expression in Acute promyelocytic leukemia
MBD1: interacts with the H3K9 HMTase SETDB1



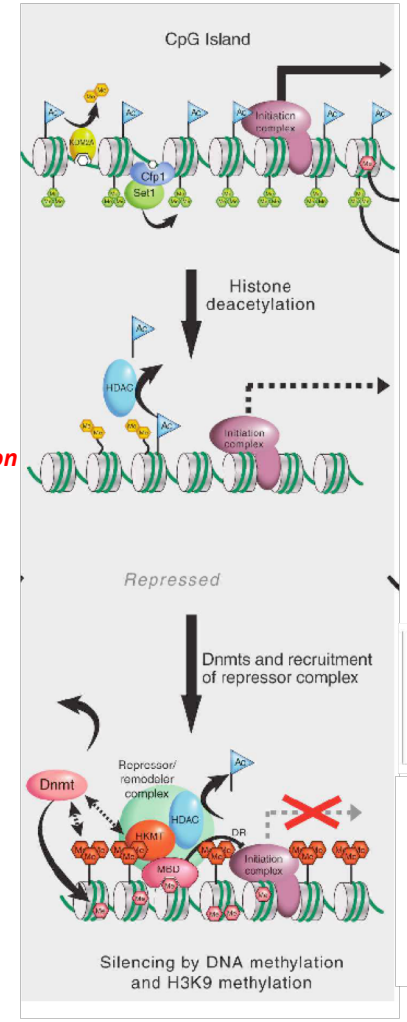
Collaboration to repress genes

TRANSCRIPTIONAL REGULATION BY METHYL-DNA BINDING PROTEINS
RECRUITMENT OF Methyl-CpG binding proteins and co-repressor complexes

Active

Transition

Inactive



MeCP2: components of the Sin3A HDAC complex
Mbd3: component of the NuRD HDAC complex
Mbd1: interacts with HDAC3.

Example: Mbd1 and HDAC3 are recruited by the PML-RARalpha hybrid protein (specialized transcription factor) to silence gene expression in cancer
MBD1: interacts with the H3K9me3 HMTase SETDB1

DNA methylation collaborates with other chromatin modifying complexes to repress gene expression

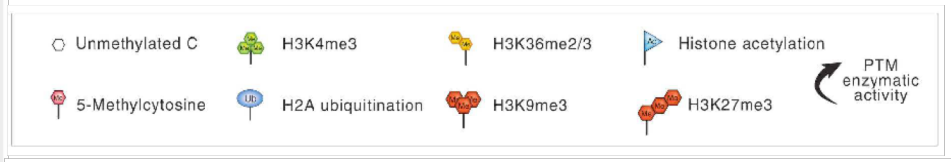


Figure 9. Recruitment of corepressors by methyl-CpG binding proteins. A hypothetical transition between an active, nonmethylated gene promoter and a repressed promoter whose silence is attributable to DNA methylation, as mediated by complexes containing an MBD protein such as MeCP2 (gray shading). The transition phase represents an intermediate step during which transcription is silenced and DNA methylation occurs. MeCP2 is envisaged to recruit the NCoR histone deacetylase (HDAC) complex and histone lysine methyltransferase (HKMT) activity to the methylated sites.

ON THE SINGLE GENE LEVEL:
CpG islands (CGIs) are short sequences stretches with variable DNA methylation that regulate promoter activity

NOTE: single CpGs are generally hyper methylated (60-90%)
CpG islands are differentially methylated, but are generally demethylate

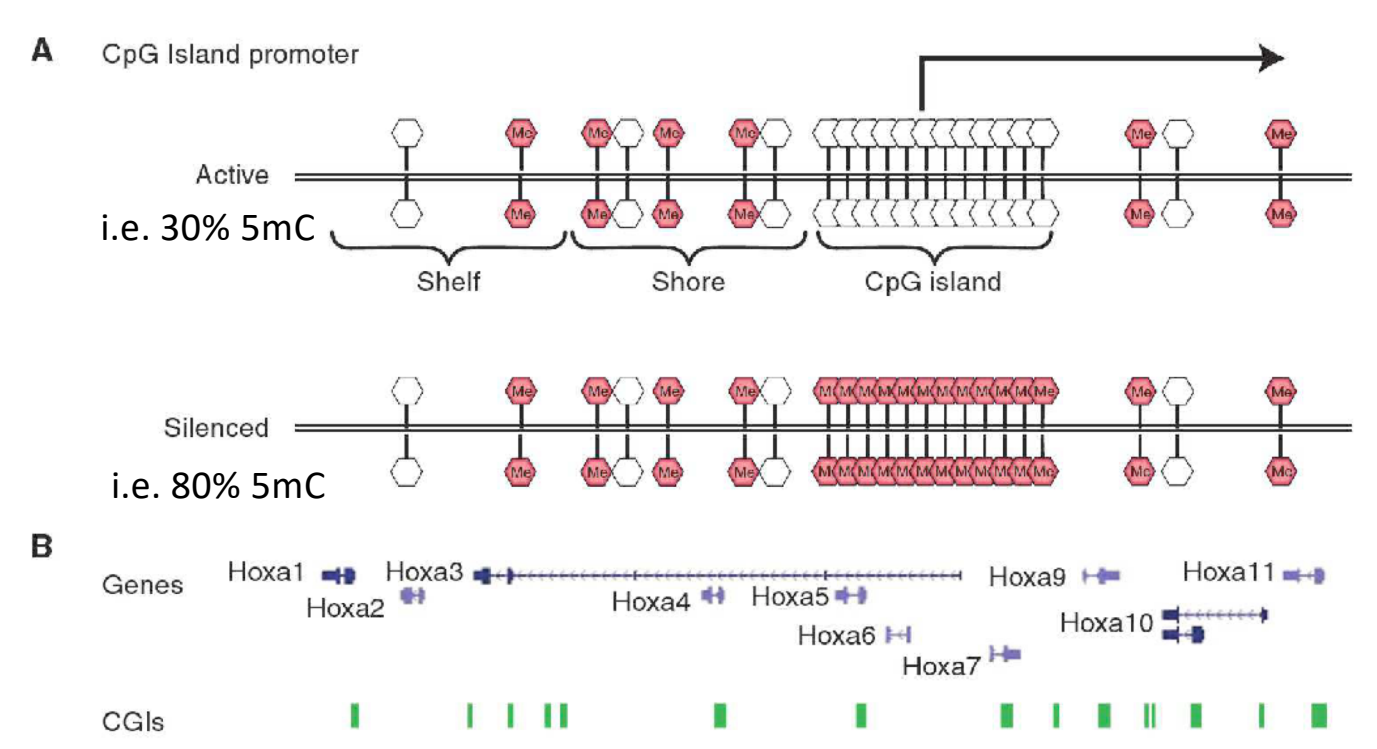
CpG islands (CGIs) have a length of ca. 1kb

60% of human genes are controlled by CGIs containing promoters that allow tissue/cell specific gene expression

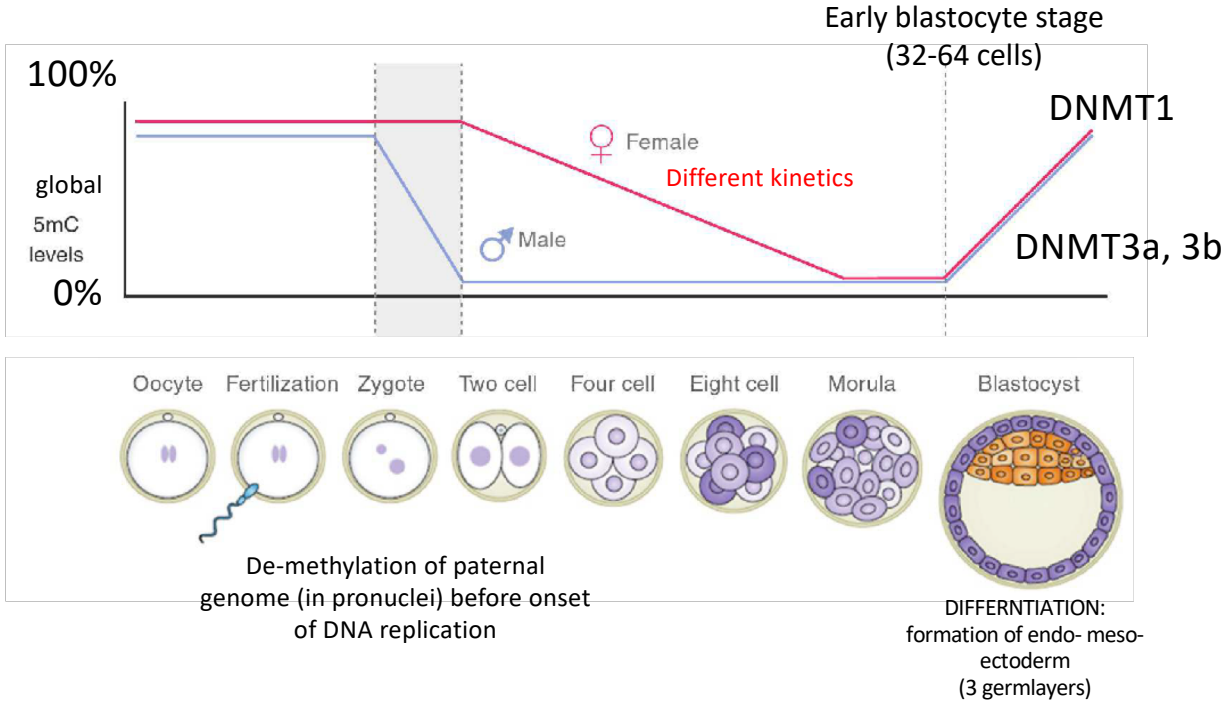
CpG islands can overlap with the first exon (methylation level in 1st exon is good predictor of gene expression

CpG islands located <2kb from promoter: shores

CpG islands located <2-4kb from promoter: shores



DNA METHYLATION IS ABUNANT IN THE GENOME AND ISSUBJECTED TO DRAMATIC ALTERATIONS DURING EMBRYOGENESIS



70%- 80% of CpG dinucleotides are methylated in the genome

DNA methylation levels are high in fertilized Oocytes that contain the paternal and maternal genome (carries characteristic methylation patterns)

Paternal and maternal methylation patterns are rapidly erased (exception: imprinted genes maintain paternal and maternal methylation information). → the paternal and maternal methylation epigenome is cancelled

DNA methylation levels remain low during the first cell division events until the blastocyst stage

In the blastocyst stage cell differentiation programs are activated and genes need to be regulated on the epigenetic level → DNA methylation is increasing (loss of DNMT1, DNMT3a or DNMT3b is lethal → establishment and maintenance of DNA methylation is impaired)

60%- 90% of CpG di-nucleotides are methylated in the human genome!
Remember only 2% of the genome encode for mRNAs
CpG islands are differentially methylated

DNA METHYLATION IS REVERSIBLE: DNA DEMETHYLATION BY Tet-family proteins

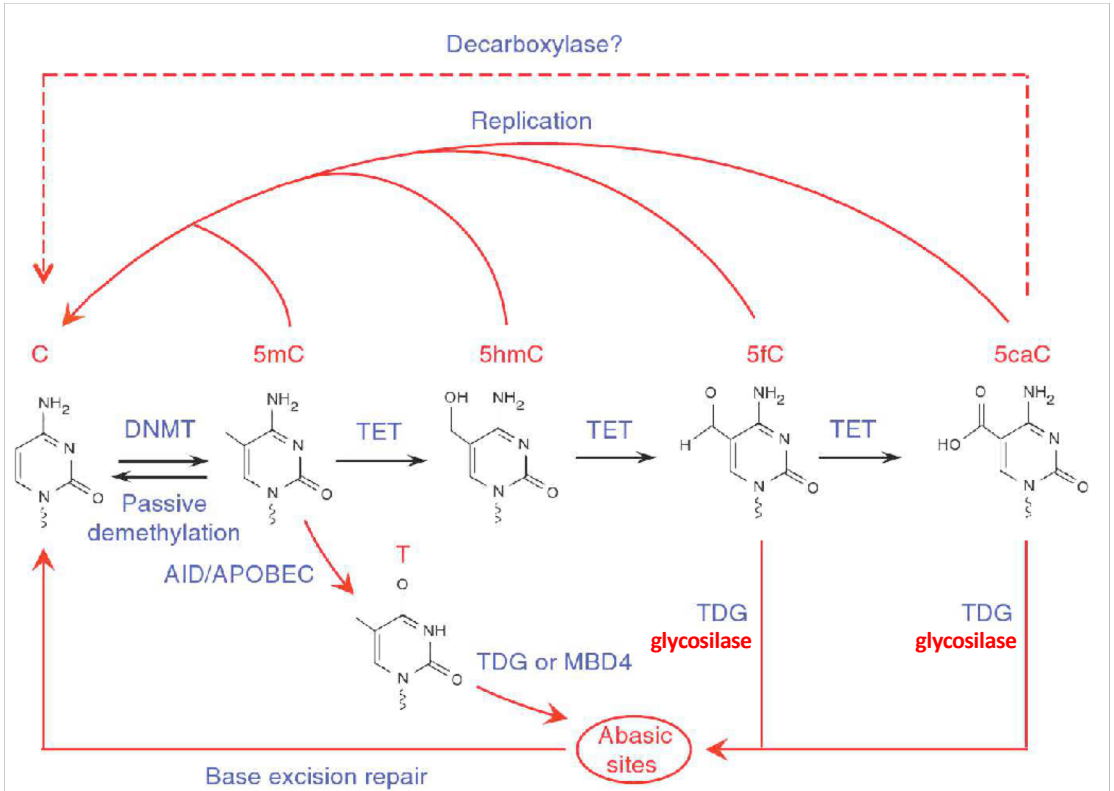


Figure 6. Model of Tet-initiated DNA demethylation pathways. DNA methylation (5mC) is established and maintained by DNMT. 5mC can by oxidized by Tet family of dioxygenases to generate 5hmC, 5fC, and 5caC. Because the oxidized 5mC derivatives cannot serve as substrates for DNMT1, they can be lost by replication-dependent passive demethylation. 5hmC can be deaminated by AID/APOBEC to become 5hmU, which together with 5fC and 5caC can be excised by glycosylases such as TDG, followed by DNA repair to generate C. Alternatively, a putative decarboxylase may convert 5caC to C.

Tet-family proteins mediate DNA demethylation

template DNA strand		neo-synthesized DNA
5mC: 5-methylcytosine		DNMT1: GC-m3 m3-CG
	Tet family proteins ↓ oxidation	
5hmC: 5-hydroxymethylcytosine		GC 5hmCG
	Tet family proteins ↓ oxidation	
5fC: 5-formylcytisine		GC 5fcCG
	Tet family proteins ↓ oxidation	
5caC: 5-carboxylcytosine		GC 5caCG

5mC, 5hmC and 5fC are abundant in the cell
5caC is present only at very low abundance

DNMT1 has exclusive specificity for 5mC

DNA METHYLATION IS REVERSIBLE: DNA DEMETHYLATION BY Tet-family proteins

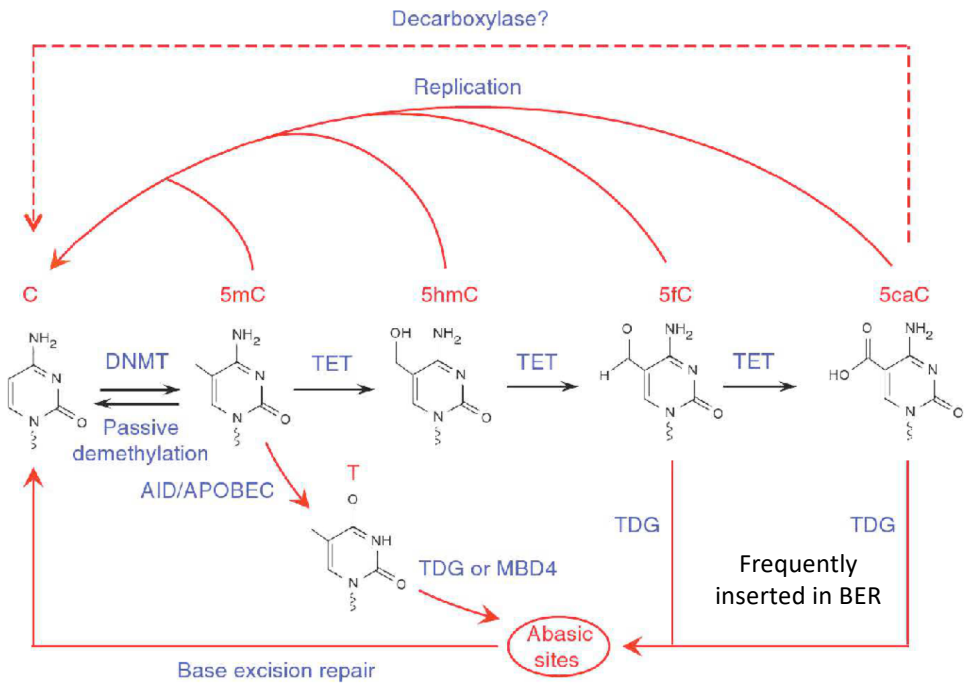
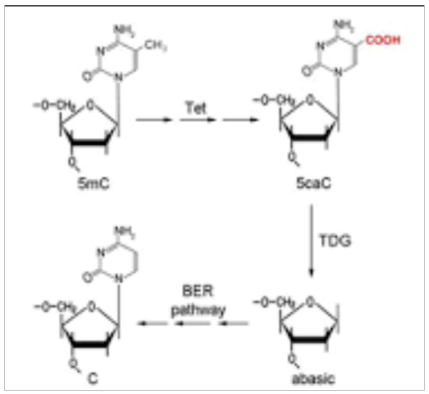
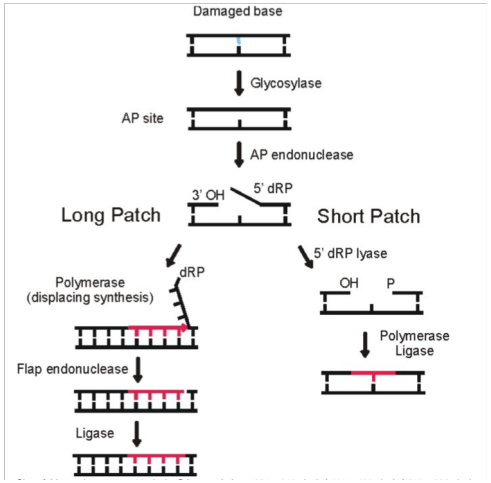


Figure 6. Model of Tet-initiated DNA demethylation pathways. DNA methylation (5mC) is established and maintained by DNMT. 5mC can by oxidized by Tet family of dioxygenases to generate 5hmC, 5fC, and 5caC. Because the oxidized 5mC derivatives cannot serve as substrates for DNMT1, they can be lost by replication-dependent passive demethylation. 5hmC can be deaminated by AID/APOBEC to become 5hmU, which together with 5fC and 5caC can be excised by glycosylases such as TDG, followed by DNA repair to generate C. Alternatively, a putative decarboxylase may convert 5caC to C.



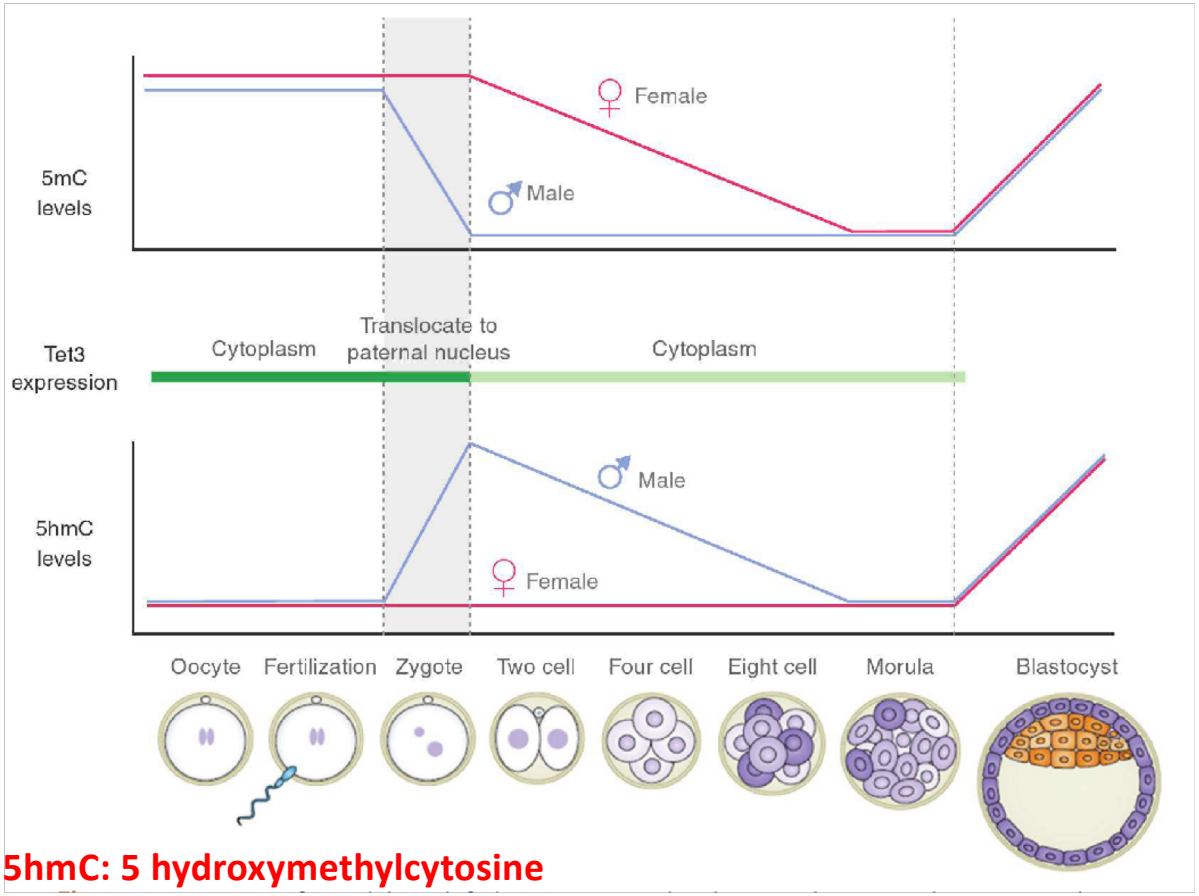


The protein encoded by this gene belongs to the TDG/mug DNA glycosylase family. Thymine-DNA glycosylase (TDG) removes thymine moieties from G/T mismatches by hydrolyzing the carbon-nitrogen bond between the sugar-phosphate backbone of DNA and the mispaired thymine. With lower activity, this enzyme also removes thymine from C/T and T/T mispairings. TDG can also remove uracil and 5-bromouracil from mispairings with guanine. Interestingly, TDG knockout mouse models showed no increase in mispairing frequency suggesting that other enzymes, like the functional homologue MBD4, may provide functional redundancy. This gene may have a pseudogene in the p arm of chromosome 12.Additionally, in 2011, the human thymine DNA glycosylase (hTDG) was reported to efficiently excises 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), the key oxidation products of 5-methylcytosine in genomic DNA. Later on, the crystal structure of the hTDG catalytic domain in complex with duplex DNA containing 5caC was published, which supports the role of TDG in mammalian 5-methylcytosine demethylation.

Check textbooks: glycosilases cleave off bases from sugar → apyrimidic/apurinic site → BER pthway

DNA METHYLATION IS REVERSIBLE:
ACTIVE AND PASSIVE DNA DEMETHYLATION

DNA de-methylation of the paternal and maternal genome has different kinetics



5hmC: 5 hydroxymethylcytosine

PASSIVE DNA DEMETHYLATION

Successive rounds of DNA methylation reduce the amount of 5mC. In this situation DNMT1 is excluded from the Nucleus! (only transient presence of oocyte specific version of DNMT1 at the 8 cell stage)

MATERNAL GENOME: slow de-methylation of DNA

ACTIVE DNA DEMETHYLATION

Enzymatic activity rapidly de-methylates 5mC

PATERNAL GENOME: fast de-methylation of DNA

→ ***In zygotes Tet3 is localized to the PATERNAL nucleus***

→ *Paternal DNA is demethylated*

→ *High levels of* 5hmC: 5-hydroxymethylcytosine, 5fc: 5-formylcytisine and 5caC: 5-carboxylcytosine *were detected at high levels in the paternal nucleus*

→ *BER machinery concentrated in pronucleus*