LECTURE 4

DNA METHYLATION

DNA METHYLATION CONTROLS GENE EXPRESSION

FACTS:

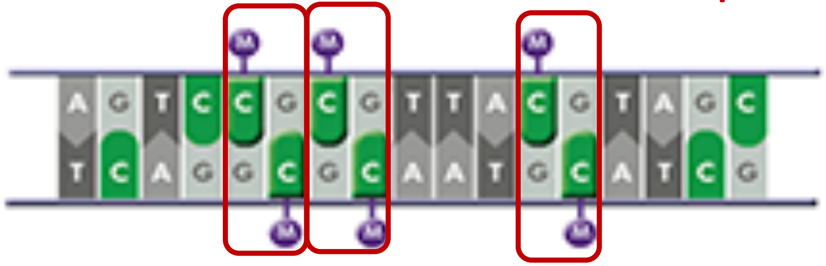
1. **DNA methylation is created at CpG di-nucleotide motifs**

2. **An accumulation of CpG is called “CpG island” (CGI)**

3. **CpG islands are enriched at promotes and sequence elements that are important for gene expression control. In some cases, CpG islands can be also located in distant locations.**

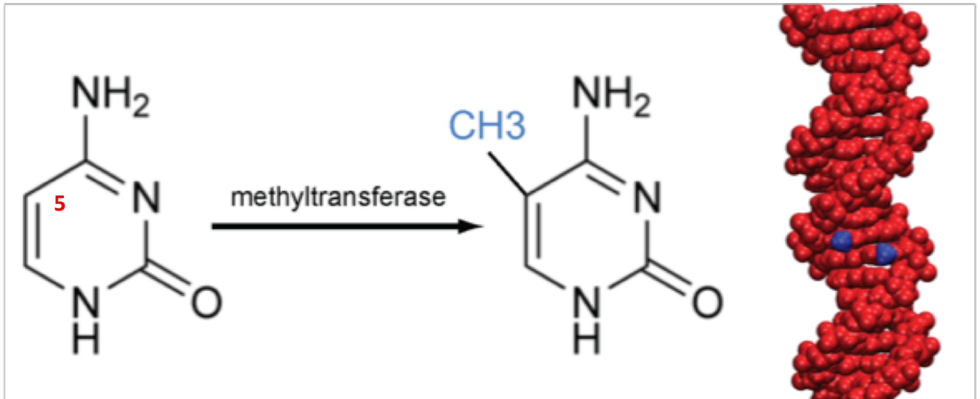
4. ***CpG methylation (=“DNA methylation”) is directly linked with stable, inheritable gene silencing***

ADVANTAGE OF DNA METHYLATION AT CpG



CpGs are self-complementary
Di-nucleotide in paired stand also contains methylation
Methylation patterns can be maintained during DNA replication

Epigenetic modification that is imposed on genomic DNA

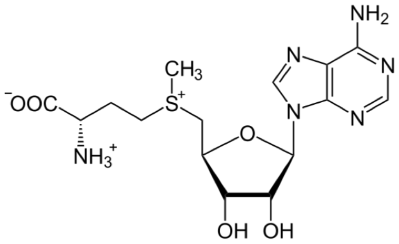


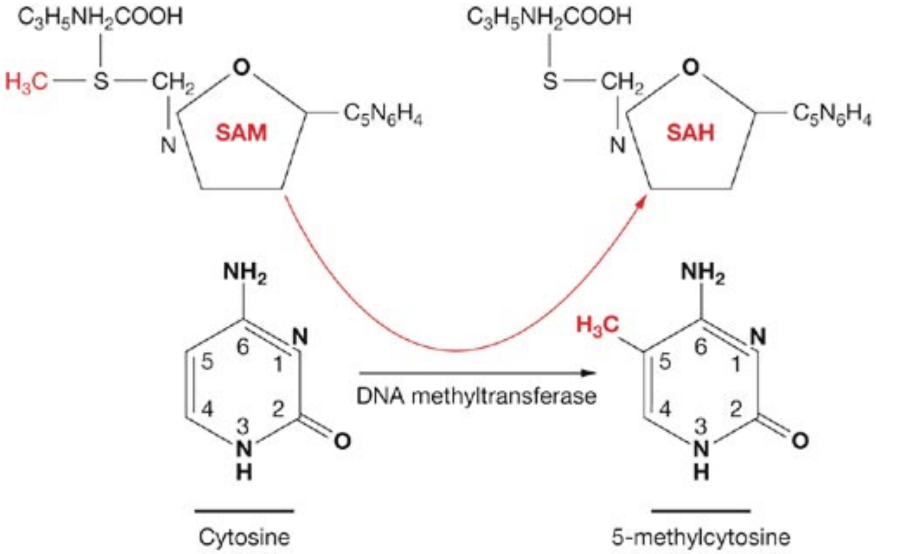
CYTOSINE

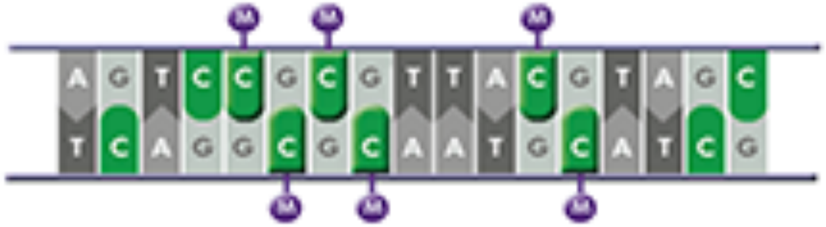
METHYL-CYTOSINE

DNA METHYLTRANSFERASES CATALYZE DNA METHYLATION

DNA methyltransferases (DNMTs) transfer a methyl-group from AdoMet (SAM) to Cytosine located in a CpG dinucleotide







S-adenosyl-l-methionine (AdoMet) or (SAM),

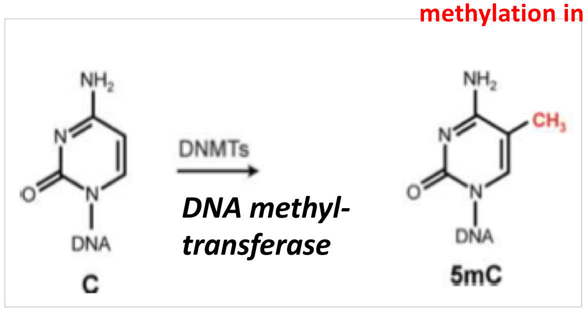
The source of the methyl group is S-adenosyl-l-methionine (AdoMet) or (SAM), which is converted to S-adenosyl-l-homocysteine (AdoHcy) in the reaction.

S-Adenosyl methionine is a common cosubstrate involved in methyl group transfers, transsulfuration, and aminopropylation.
SAM = enzymatic cofactor
<u>SAM is after ATP the most commonly used cofactor used by the cell</u>
Although these anabolic reactions occur throughout the body, most **SAM-e is produced and consumed in the liver**. More than 40 methyl transfers from SAM-e are known, to various substrates such as nucleic acids, proteins, lipids and secondary metabolites. It is made from adenosine triphosphate (ATP) and methionine by methionine adenosyltransferase. SAM was first discovered in Italy by Giulio Cantoni in 1952.

Mapping DNA methylation at CpG islands
BISULFITE SEQUENCING

Methylation of cytosine at CpG dinucleotides is an important epigenetic regulatory modification in many eukaryotic genomes.

DNA methylation was found to be located genome-wide with a pattern of low methylation in proximity to promoters and high gene body-methylation in highly-expressed genes → methylation pattern can identify transcribed DNA (gene)

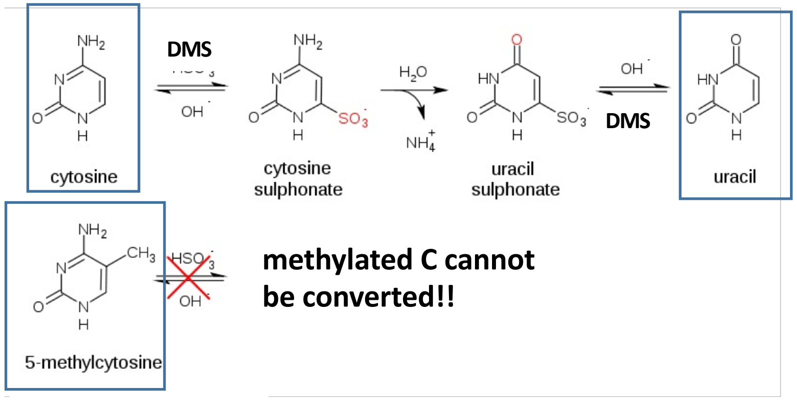


DNA methyl-transferase

active gene

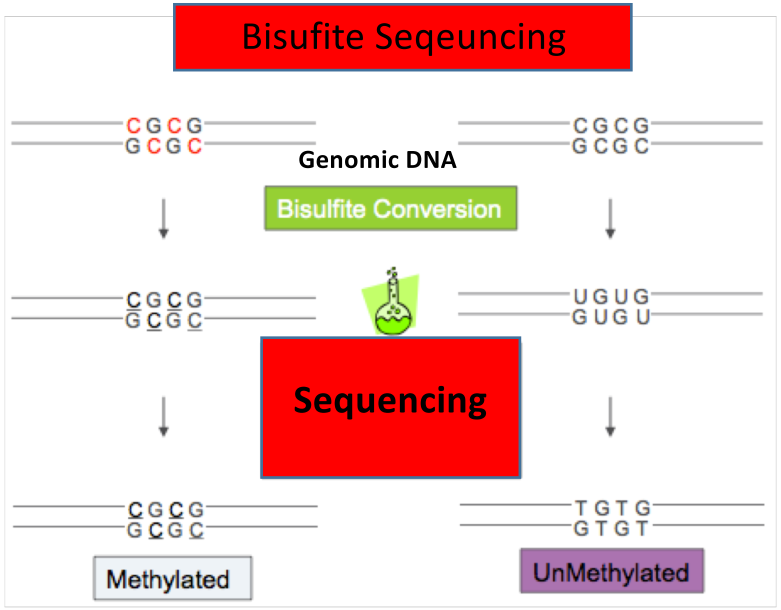
silenced gene

Bisulfite conversion: C→U conversion using dimethyl sulfate



methylated C cannot be converted!!

DMS = Dimethyl sulfate



Bisufite Seqeuncing

Genomic DNA

Bisulfite Conversion

Sequencing

Methylated

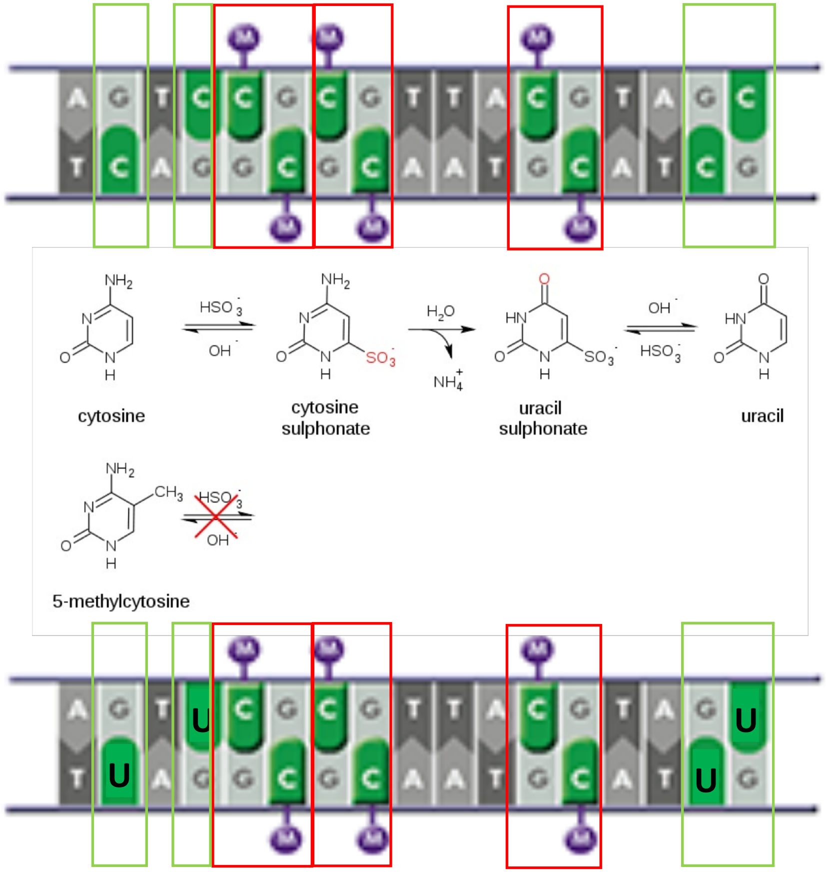
UnMethylated

Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected. Thus, bisulfite treatment introduces specific changes in the DNA sequence that depend on the methylation status of individual cytosine residues, yielding single-nucleotide resolution information about the methylation status of a segment of DNA.

Mapping DNA methylation at CpG islands
BISULFITE SEQUENCING

No DNMT targets *DNMT targets* *DNMT target* *DNMT target*

Genomic DNA



DNA for Sequencing

Bisulfite conversion

Seqeuncing of both strands reveals C → U (T) transition

Compare with genomic sequence
C→U sequence change = DNA methylation
C→C no sequence change = no DNA methylation

Mapping DNA methylation at CpG islands
BISULFITE SEQUENCING

An example:
hTERT encodes the telomerase gene
hTERT elongates telomeres thereby protecting cancer cells from replicative senescence

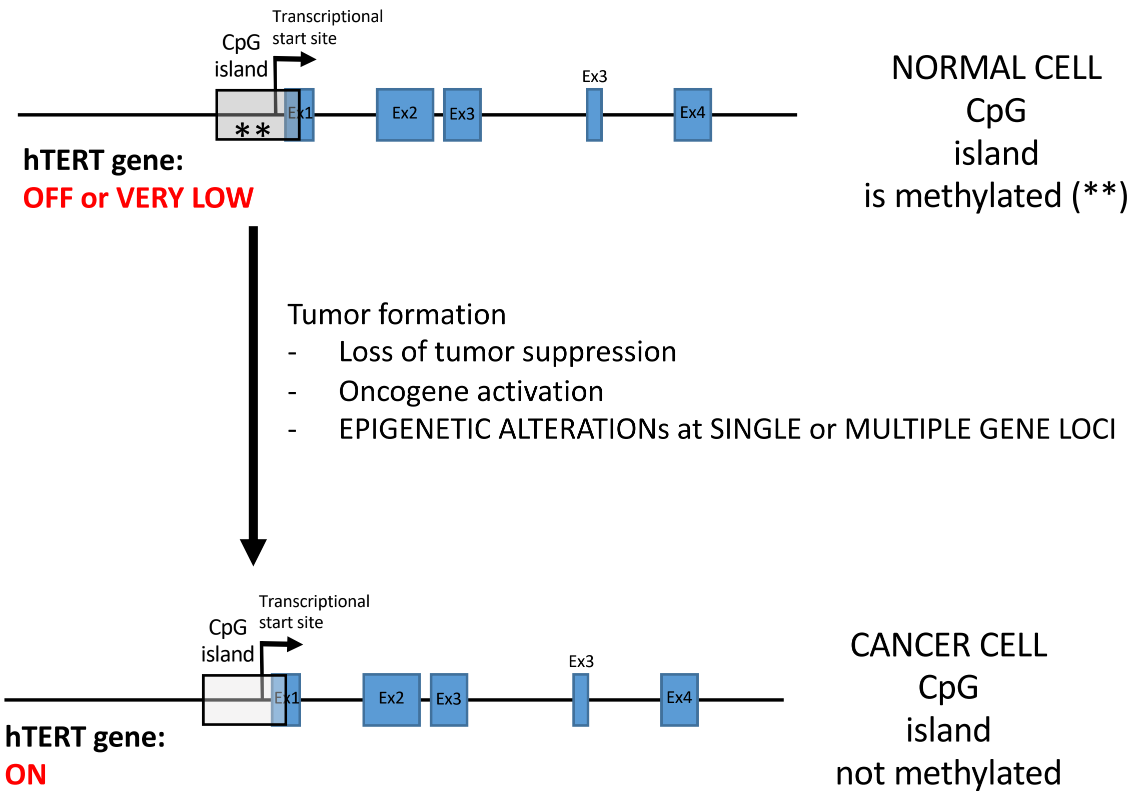
→ Cancer cells do not age and can proliferate forever!!

→ Therefore 90% of human cancers express telomerase. NOTE: telomerase gene is silenced by DNA methylation at defined positions in differentiated cells.

→ Bisulfite sequencing showes that cancer cells have a de-methyalted CpG island located at the hTERT promoter.

Note: CpG islands can alos overlap with the 1st intron of the gene and impact on promoter activity!!!!

primary or differentiated cell



hTERT gene:
OFF or VERY LOW

NORMAL CELL
CpG
island
is methylated (**)

Tumor formation
- Loss of tumor suppression
- Oncogene activation
- EPIGENETIC ALTERATIONs at SINGLE or MULTIPLE GENE LOCI

hTERT gene:
ON

CANCER CELL
CpG
island
not methylated

Cancer cell

Mapping DNA methylation at CpG islands of individual genes
BISULFITE SEQUENCING

Prepare DNA from normal cell and tumor cells

Purify DNA and perform bi-sulfite conversion (DMS)
(Unmethylated C → U; Methylated C → C)

Amplify your region of interest using gene specific primers = CpG island in the TERT promoter

Purify DNA fragment obtained by PCR

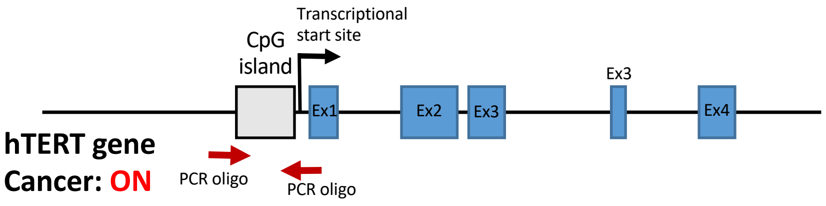
Clone fragment into Plasmid

Transform bacteria with plasmid (one bacteria receives only one plasmid!!!)

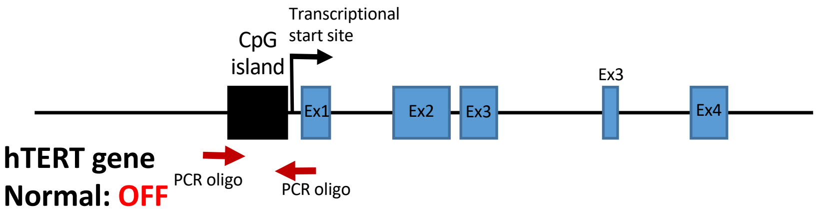
Pick single colonies and grow individual (10-15) small cultures for mini prep

Purify amplified plasmid DNA from 10-15 individual bacterial cultures

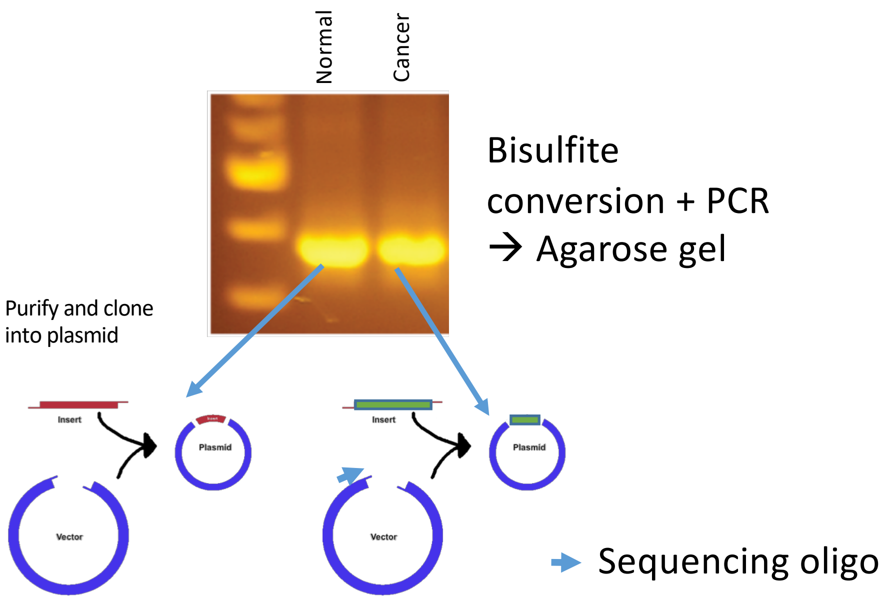
Sequence inserts using a primer that anneals to the vector DNA, adjacent to the insertion site of the PCR product
REMEMBER: only a single type of plasmid is sequenced
The plasmid-insert refers to a single copy of hTERT CpG island prepared from the normal and cancer cell cultures !!!!
To have a good representation of DNA methylation pattern, you need to sequence at least 10-15 clones



hTERT gene
Cancer: ON



hTERT gene
Normal: OFF



Bisulfite conversion + PCR
→ Agarose gel

Purify and clone into plasmid

Sequencing oligo

Mapping DNA methylation at CpG islands
BISULFITE SEQUENCING

→ Sequence inserts using a primer that anneals to the vector DNA, adjacent to the insertion site of the PCR product

REMEMBER: only a single type of plasmid is sequenced
The plasmid-insert refers to a single copy of hTERT CpG island prepared from the normal and cancer cell cultures !!!!

To have a good representation of DNA methylation pattern, you need to sequence at least 10-15 clones

RESULTS:

1. Align obtained DNA sequences obtained from normal and cancer cell to reference human DNA sequence
2. Compare C→U conversion in normal versus tumor cell (see image on right hand side)

INTERPRETATION:

Parameter 1: Quality control of your bisulfite conversion:

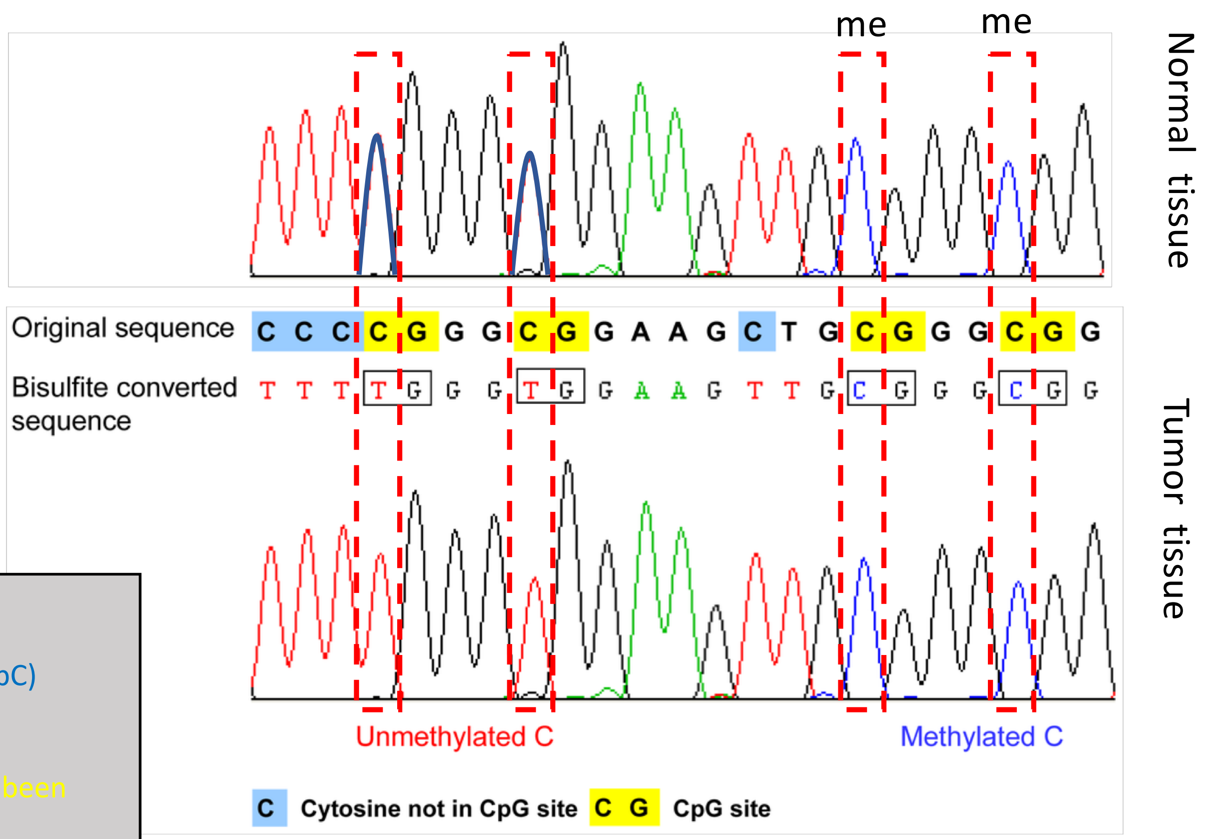
Have all C that aren't located 5' of to a neighboring G (CpT, CpA, CpC) converted to U???

Parameter 2:

What is the frequency of C located in CpG dinucleotides that have been converted to U? = non methylated il cell

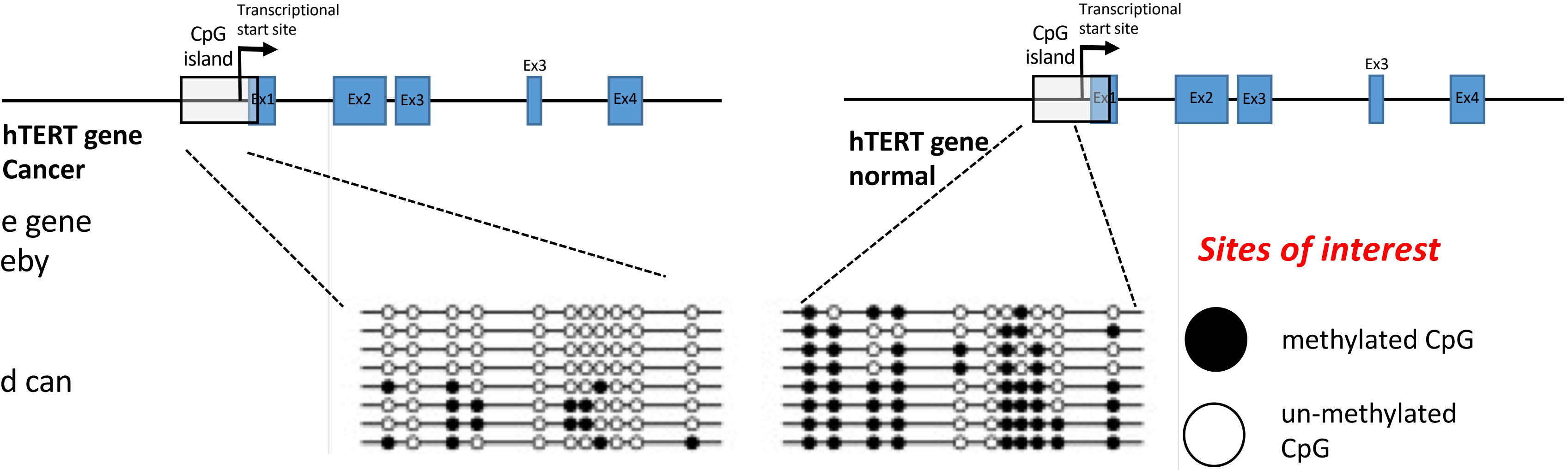
What is the frequency of C located in CpG dinucleotides that have NOT been converted to U? = methylated in cell

Result from Sanger sequencing using Dye terminators



Compare with genomic sequence
Bisulfite conversion: C→U sequence change = DNA methylation
C→C no seqeunce change = no DNA methylation

Mapping DNA methylation at CpG islands
BISULFITE SEQUENCING



cancer cell

→ Many conversions

normal cell

→ Few conversions

Sites of interest

● methylated CpG

○ un-methylated CpG

An example:
hTERT encodes the telomerase gene
Tert elongates telomeres thereby protecting cancer cells from replicative senescence

→ Cancer cells do not age and can proliferate forever!!

→ Therefore 90% of human cancers express telomeres; but telomerase is silenced by DNA methylation in differentiated cells.

→ Bisulfite sequencing showed that cancer cells have a de-methyalted CpG island located at the hTERT promoter

Mapping DNA methylation at CpG islands
BISULFITE SEQUENCING

An example:

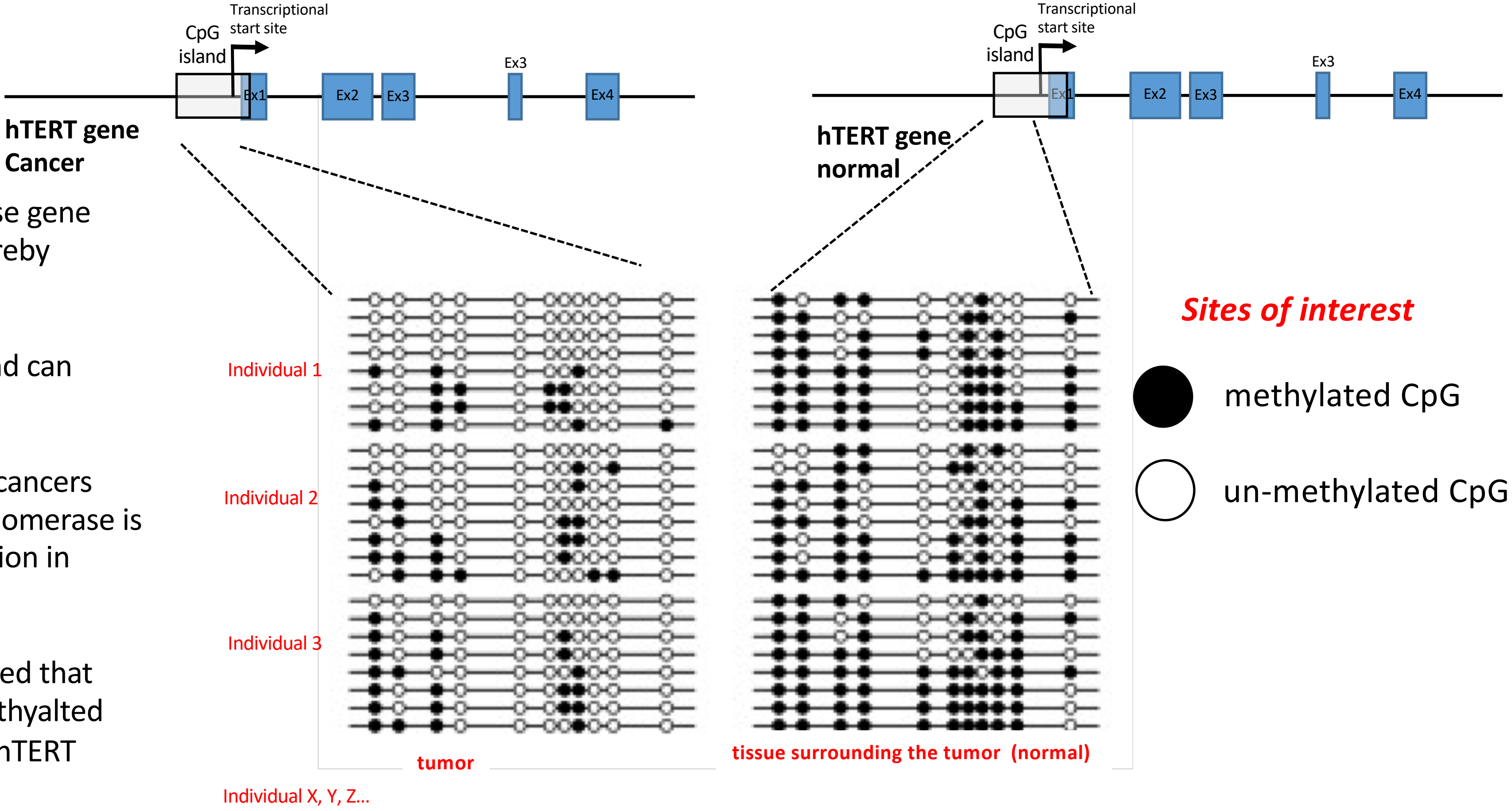
hTERT encodes the telomerase gene

Tert elongates telomeres thereby protecting cancer cells from replicative senescence

→ Cancer cells do not age and can proliferate forever!!

→ Therefore 90% of human cancers express telomeres; but telomerase is silenced by DNA methylation in differentiated cells.

→ Bisulfite sequencing showed that cancer cells have a de-methyalted CpG island located at the hTERT promoter



ANALYSIS OF PATIENT TUMORS + ADJACENT TISSUE

High throughput mapping DNA methylation at CpG islands
METHYLATED DNA IMMUNOPRECIPITATION: METHYL-DIP

Methyl-DIP works similar to ChIP:

***ESSENTIAL TO GET BIOLOGICAL INFORMATION:* 2 experimental samples:**

- Control (normal cells)
- Experimental sample (cancer cells)

Prepare genomic DNA (carries DNA methylation marks of CpG)

Sonicate DNA (see also ChIP)

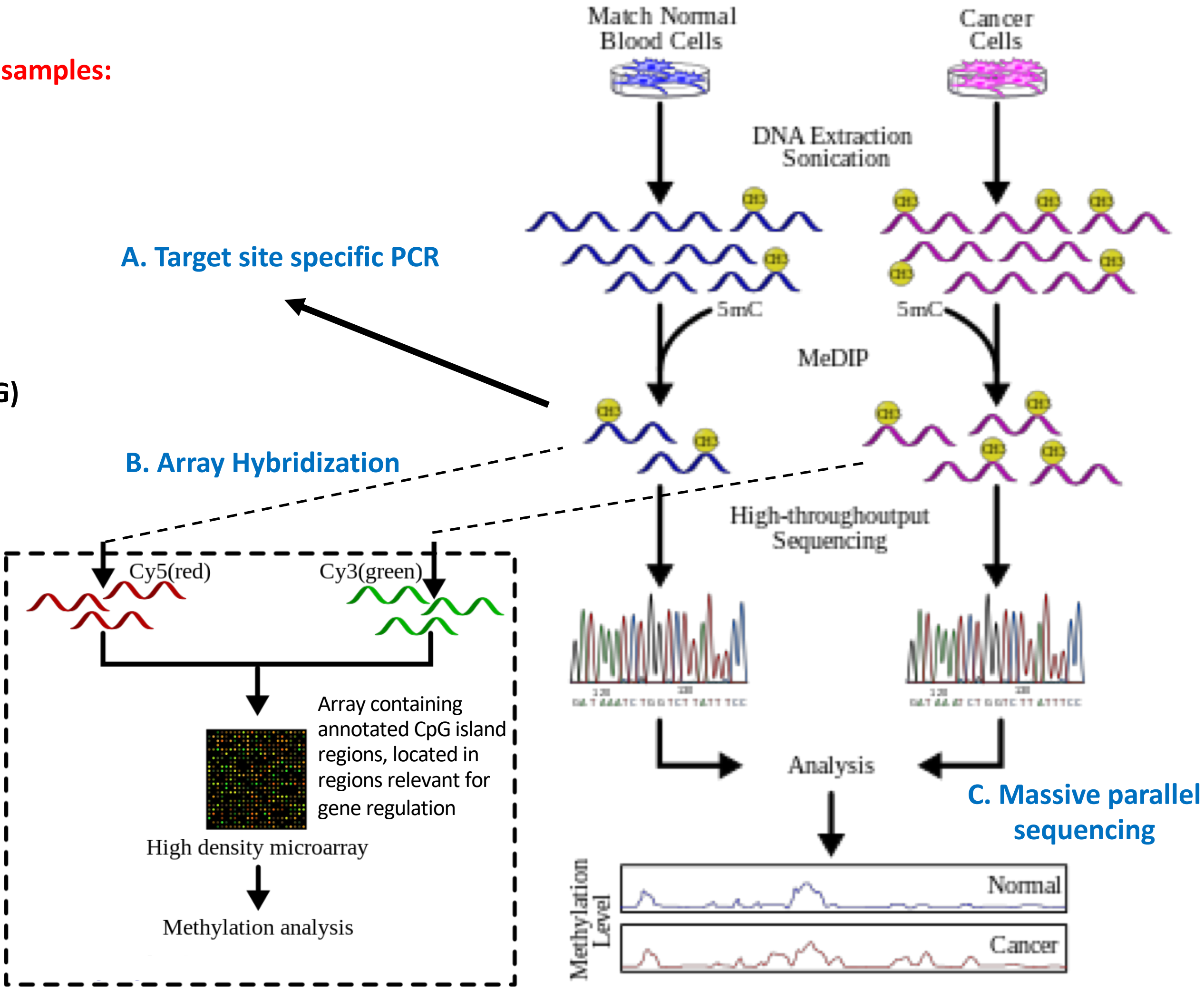
Immunoprecipitation using an anti-methyl-CpG specific antibody (monoclonal, discriminates between CpG and met-CpG)

Washing of precipitate

DATA ACQUISITION

Main methods to analyse DNA methylation

1. **PCR on specific CpG islands of interest**
2. **Differential labelling (Cy3-control; Cy5-cancer methyl-DNA) followed by hybridization to genome array**
3. **Massive parallel sequencing (different approaches)**



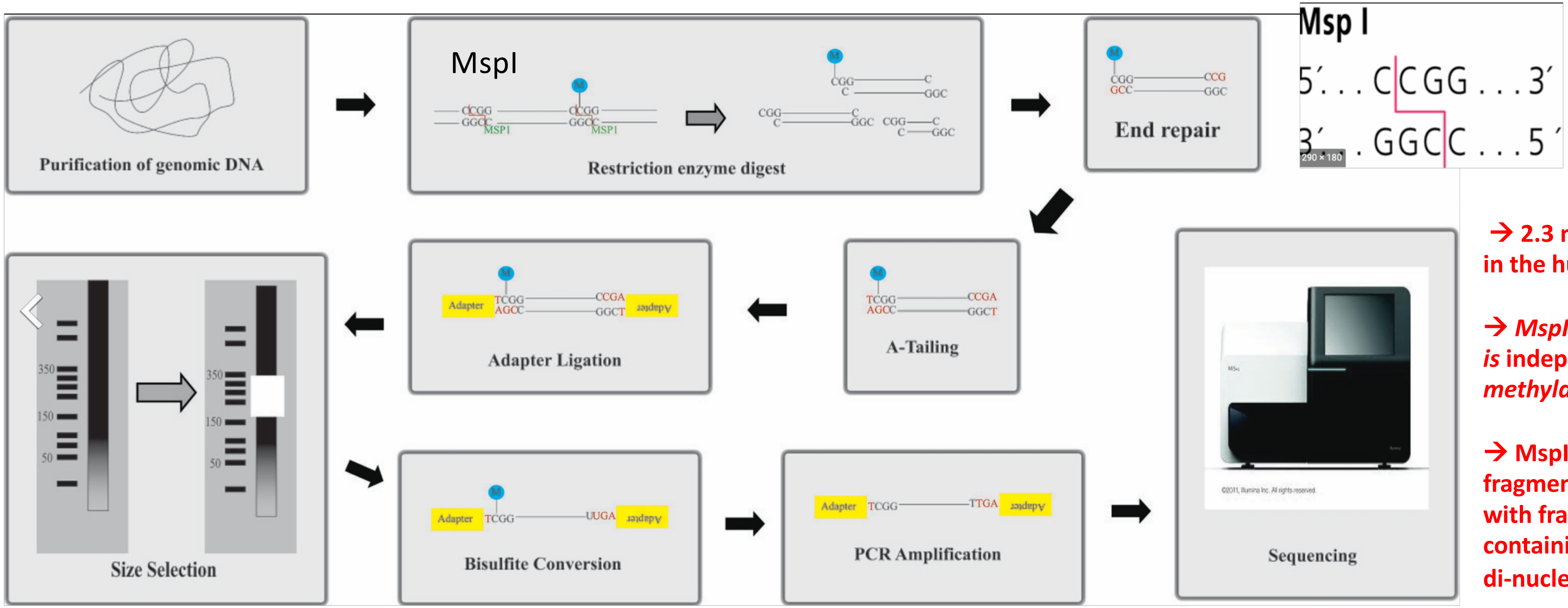
A. Target site specific PCR

B. Array Hybridization

Array containing annotated CpG island regions, located in regions relevant for gene regulation

C. Massive parallel sequencing

High throughput mapping DNA methylation genome wide:
REDUCED REPRESENTATION BISULFITE SEQEUNCING (RRBS)



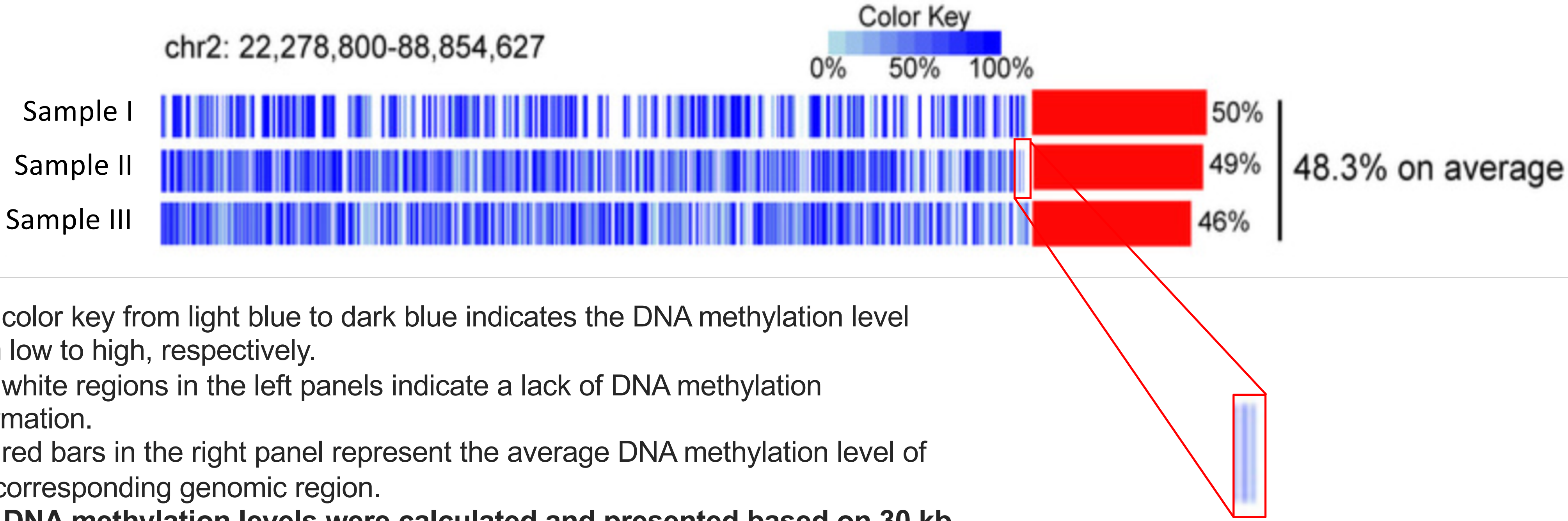
→ 2.3 million CCGG sites in the human genome

→ MspI cutting after CpG is independent of methylation

→ MspI digest ensures fragmentation of genoms with fragments containing terminal CpG di-nucleotides

Reduced representation bisulfite sequencing (RRBS) is an efficient and high-throughput technique used to analyze the genome-wide methylation profiles on a single nucleotide level.
This technique combines restriction enzymes and bisulfite sequencing in order to enrich for the areas of the genome that have a high CpG content.
Due to the high cost and depth of sequencing needed to analyze methylation status in the entire genome.
The fragments that comprise the reduced genome but still includes the majority of promoters, as well as regions such as repeated sequences that are difficult to profile using conventional bisulfite sequencing approaches.

Mapping DNA methylation genome wide:
REDUCED REPRESENTATION BISULFITE SEQEUNCING (RRBS)



The color key from light blue to dark blue indicates the DNA methylation level from low to high, respectively.

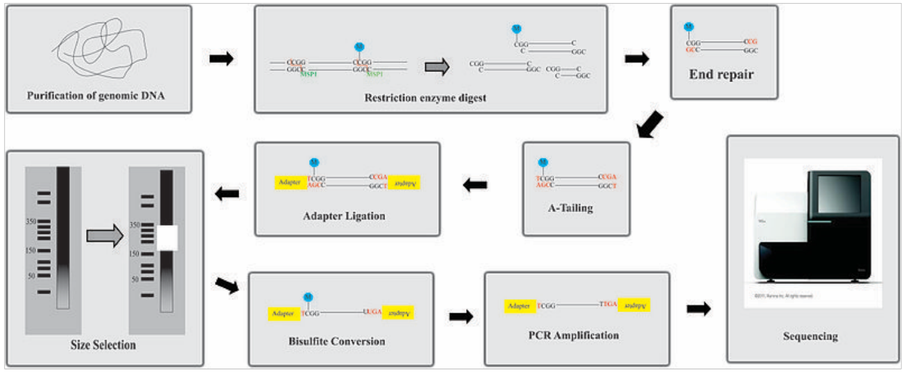
The white regions in the left panels indicate a lack of DNA methylation information.

The red bars in the right panel represent the average DNA methylation level of the corresponding genomic region.

The DNA methylation levels were calculated and presented based on 30 kb windows, only if these windows have more than 5 CpG sites covered.

For each bar in the diagram the 30 kb sequnece Obtained by DNA seqeuncing and methyaltion status are known!
→ Information on CpG di-methylation on single nucleotide level

Mapping DNA methylation genome wide:
REDUCED REPRESENTATION BISULFITE SEQEUNCING (RRBS)



Enzyme digestion: First, genomic DNA is digested using a methylation-insensitive restriction enzyme. It is integral for the enzymes to not be influenced by the methylation status of the CpGs (sites within the genome where a cytosine is next to a guanine) as this allows for the digestion of both methylated and unmethylated areas. MspI is commonly used. This enzyme targets 5’CCGG3’ sequences and cleaves the phosphodiester bonds upstream of CpG dinucleotide. When using this particular enzyme, each fragment will have a CpG at each end. This digestion results in DNA fragments of various sizes.End repair and A-tailing: Due to the nature of how MspI cleaves double stranded DNA, this reaction results in strands with sticky ends. End repair is necessary to fill in the 3’ terminal of the ends of the strands. The next step is adding an extra adenosine to both the plus and minus strands. This is referred to as A-Tailing and is necessary for adapter ligation in the subsequent step. End repair and A-Tailing is done within the same reactions, with dCTP, dGTP and dATP deoxyribonucleotides. In order to increase the efficiency of A tailing, the dATPs are added in excess in this reaction.Sequence adapters: Methylated sequence adapters are ligated to the DNA fragments. The methylated adapter oligonucleotides have all cytosines replaced with 5’methyl-cytosines, in order to prevent the deamination of these cytosines in the bisulfite conversion reaction. For reactions to be sequenced using Illumina sequencers, the sequence adapters are used to hybridize to the adapters on the flow cell.Fragment purification: The desired size of fragments is then selected to be purified. The different sizes of the fragments are separated using gel electrophoresis and are purified using gel excising. According to Gu et al., DNA fragments of 40-220 base pair are representative of the majority of promoter sequences and CpG islands[2]Bisulfite conversion: The DNA fragments are then bisulfite converted, which is a process that deaminates unmethylated cytosine into a uracil. The methylated cytosines remain unchanged, due to the methyl group protecting them from the reaction.PCR amplification: The bisulfite converted DNA is then amplified using PCR with primers that are complementary to the sequence adapters.PCR purification: Before sequencing, the PCR product must be free of unused reaction reagents such as unincorporated dNTPs or salts. Thus, a step for PCR purification is required. This can be done by running another electrophoresis gel or by using kits designed specifically for PCR purification.Sequencing: The fragments are then sequenced. When RRBS was first developed, Sanger sequencing was initially used. Now, next generation sequencing approaches are used. For Illumina sequencing, 36-base single-end sequencing reads are most commonly performed.Sequence alignment and analysis: Due to the unique properties of RRBS, special software is needed for alignment and analysis. Using MspI to digest genomic DNA results in fragments that always start with a C (if the cytosine is methylated) or a T (if a cytosine was not methylated and was converted to a uracil in the bisulfite conversion reaction). This results in a non-random base pair composition. Additionally, the base composition is skewed due to the biased frequencies of C and T within the samples. Various software for alignment and analysis is available, such as Maq, BS Seeker, Bismark or BSMAP. Alignment to a reference genome allows the programs to identify base pairs within the genome that are methylated.

DNA methyltransferases methylate DNA

FACTS:

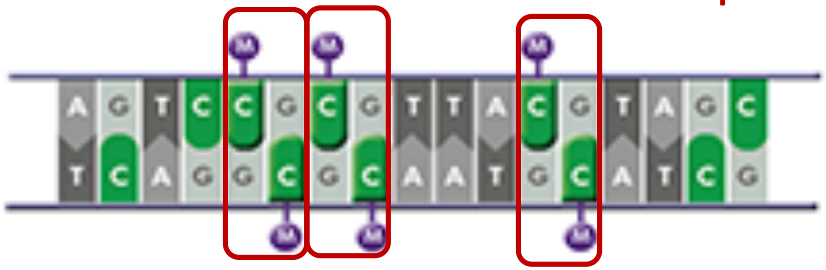
1. DNA methylation is created at CpG di-nucleotide motifs

2. An accumulation of CpG genes is called CpG island

3. CpG islands are enriched at promoters and other, more distant sequence elements that are important for gene expression control

4. CpG methylation (="DNA methylation") is linked with Stable gene silencing

ADVANTAGE OF DNA METHYLATION AT CpG



CpG Di-nucleotide are self-complementary

PROPOSED MODEL:

Methylation patterns can be maintained during DNA replication

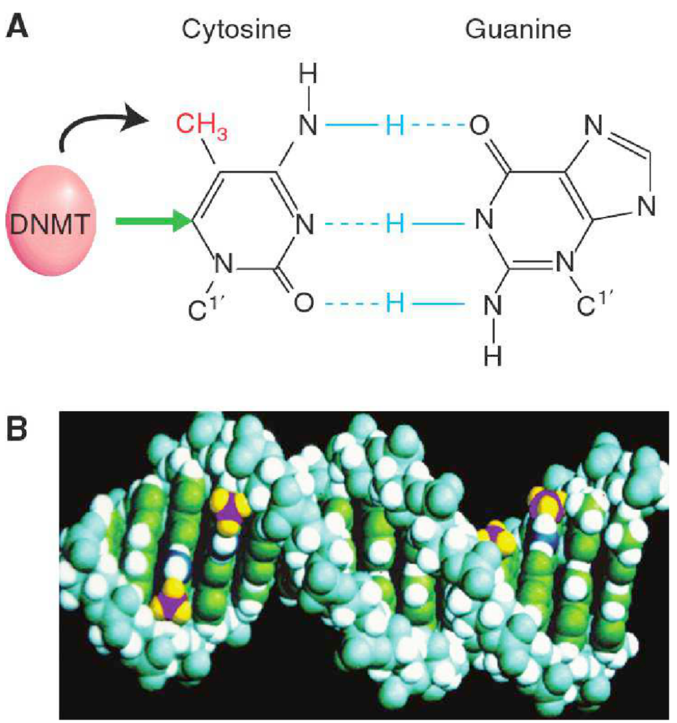


Figure 1. Cytosine methylation in DNA. (*A*) Addition of a methyl group, CH_3 (red), at the five position of the cytosine pyrimidine ring (black arrow) does not sterically interfere with GC base pairing (blue lines). DNA methyltransferases associate covalently with the carbon 6 position (straight green 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.

De-novo DNMTs → place new DNA methylation

Maintenance DNMTs → propagate methylation after replication

DNA methyl transferases methylate DNA

Maintenance of DNA methylation patterns by DNMTs

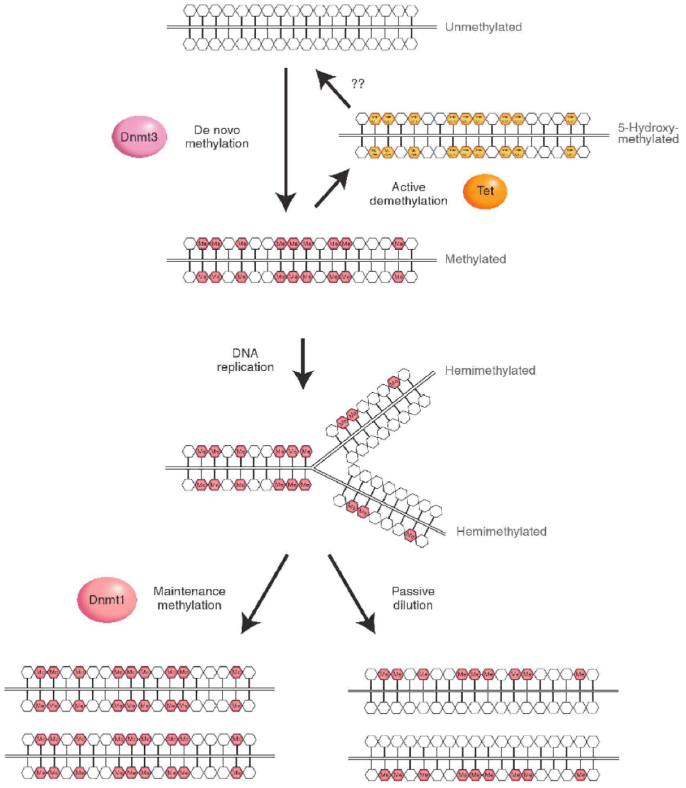


Figure 2. De novo methylation and maintenance methylation of DNA. A stretch of genomic DNA is shown as a line with self-complementary CpG pairs marked as vertical strokes. Unmethylated DNA (*top*) becomes methylated "de novo" by Dnmt3a and Dnmt3b to give symmetrical methylation at certain CpG pairs. 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:

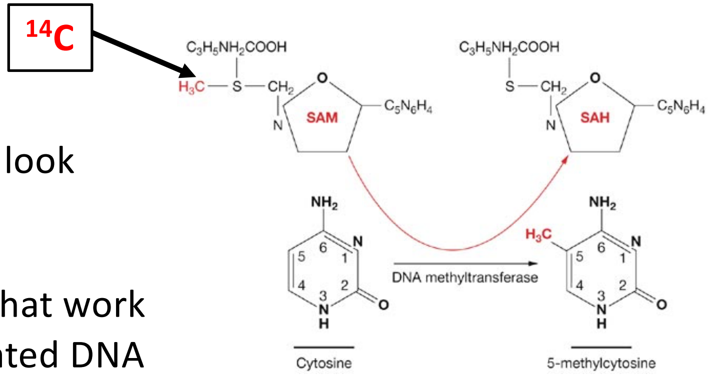
DNMT1: discovered first

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*



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

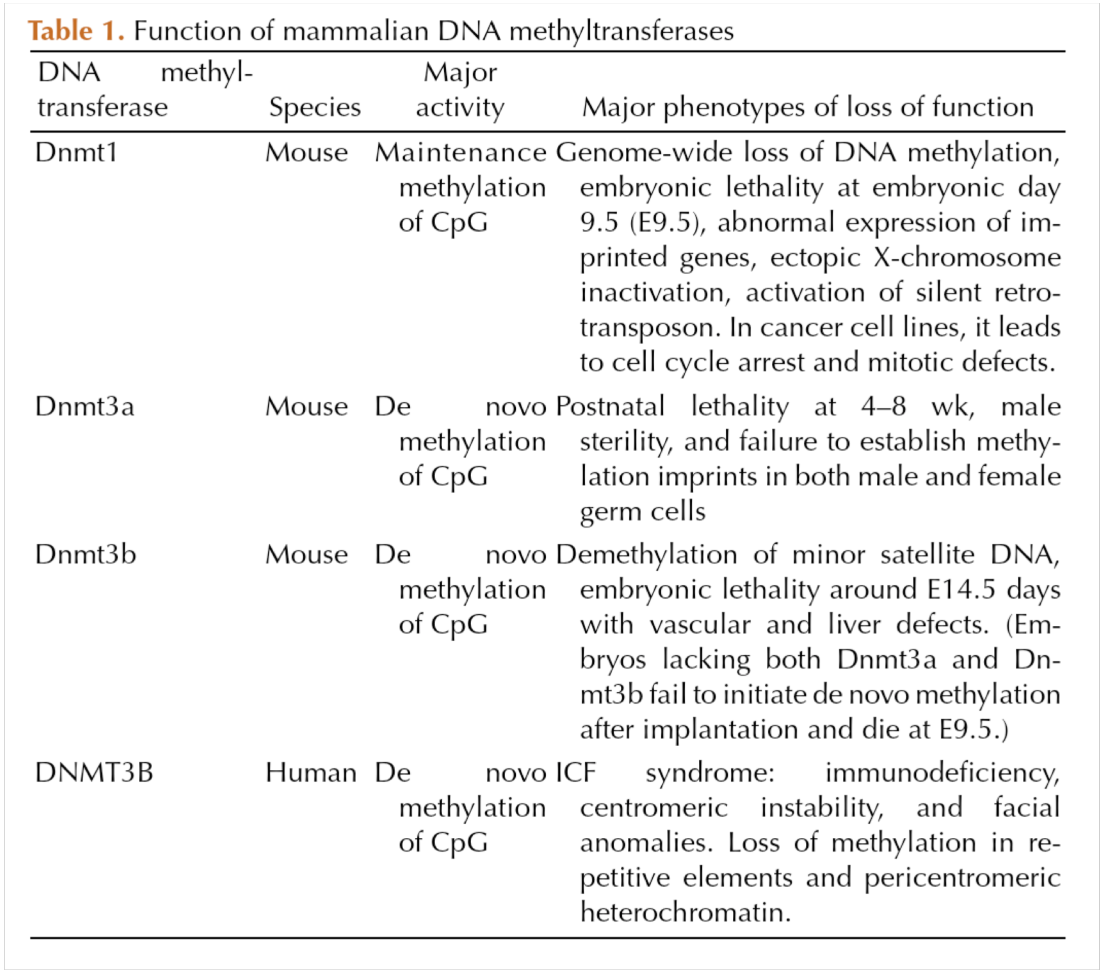


Table 1. Function of mammalian DNA methyltransferases

DNA methyl-transferase	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.

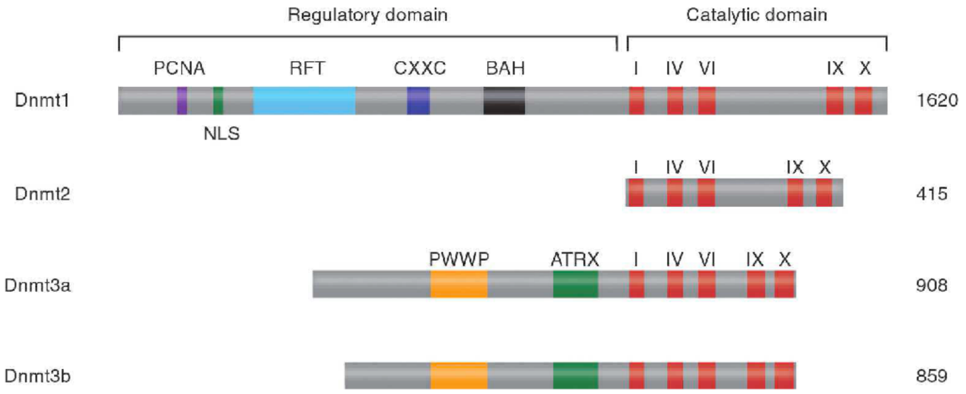
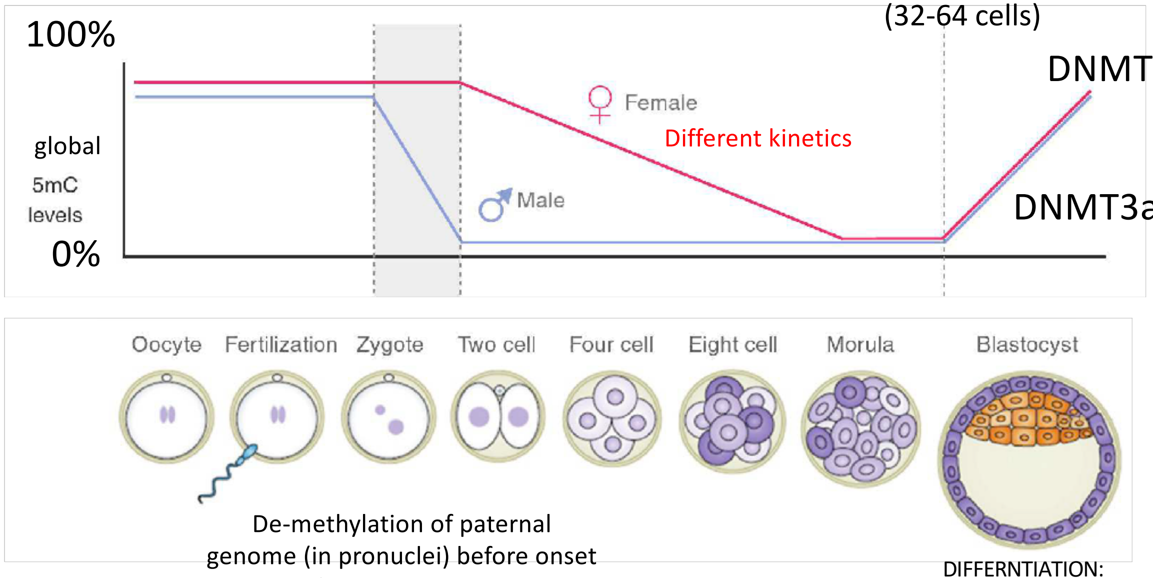


Figure 4. Mammalian DNA methyltransferases. The catalytic domains of Dnmt1, Dnmt2, and the Dnmt3 family members are conserved (the signature motifs, I, IV, VI, IX, and X, are most conserved in all cytosine methyltransferases), but there is little similarity among their amino-terminal regulatory domains. 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



Early blastocyte stage
(32-64 cells)

DNMT1

DNMT3a, 3b

De-methylation of paternal
genome (in pronuclei) before onset
of DNA replication

DIFFERNTIATION:
formation of endo- meso-
ectoderm
(3 germlayers)

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

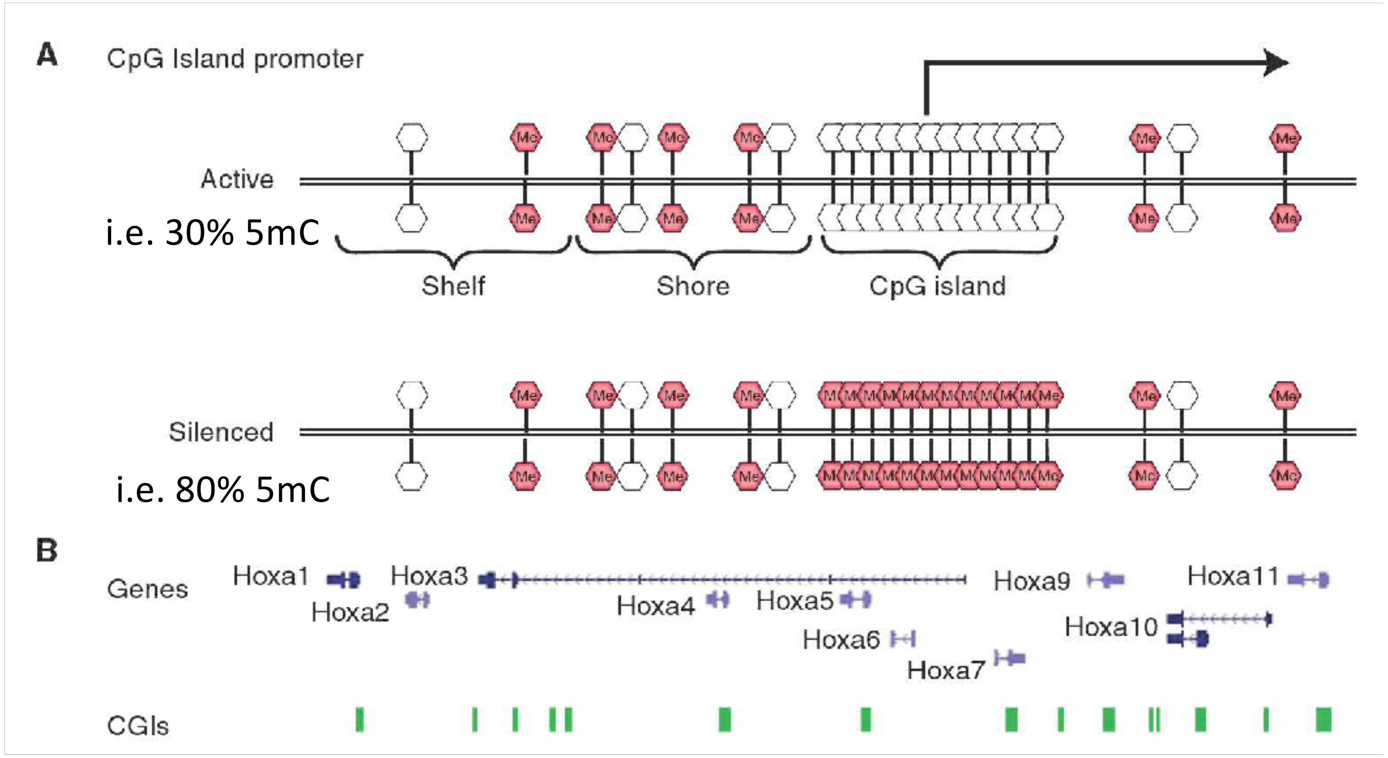
CpG islands (CGIs) have a length of ca. 1kb

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

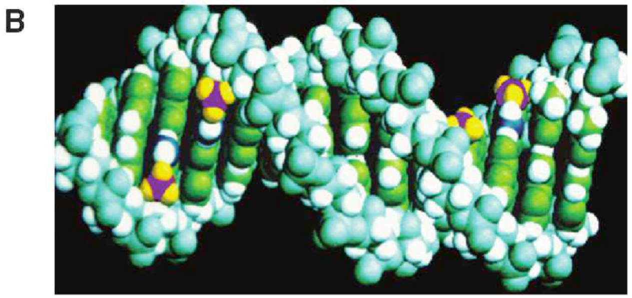
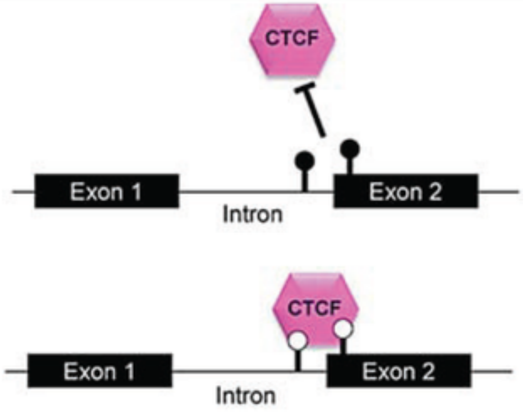


Figure 1. Cytosine methylation in DNA. (*A*) Addition of a methyl group, CH_3 (red), at the five position of the cytosine pyrimidine ring (black arrow) does not sterically interfere with GC base pairing (blue lines). DNA methyltransferases associate covalently with the carbon 6 position (straight green 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

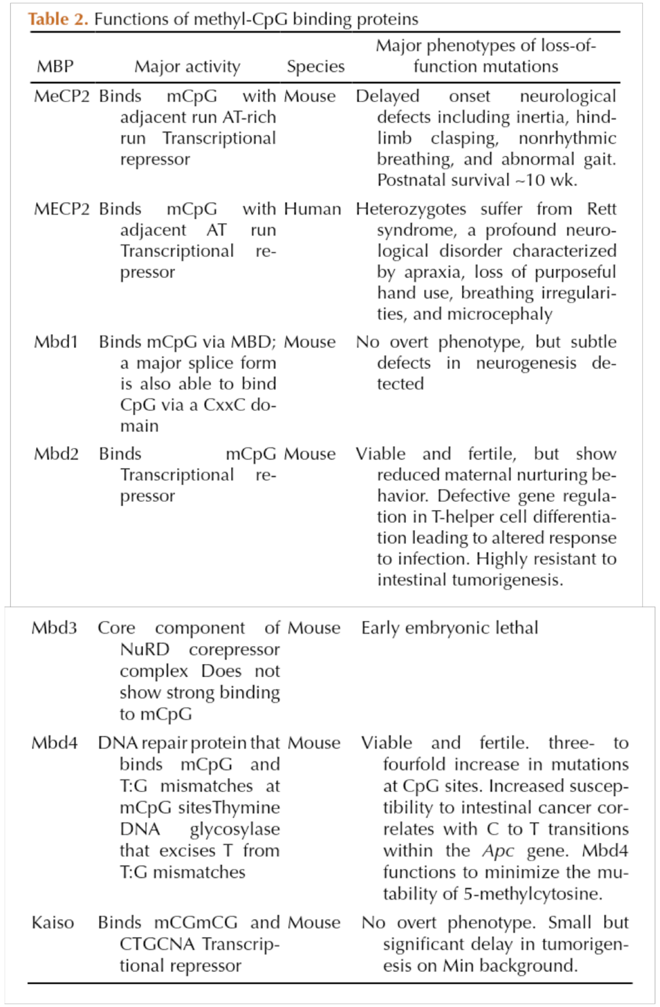
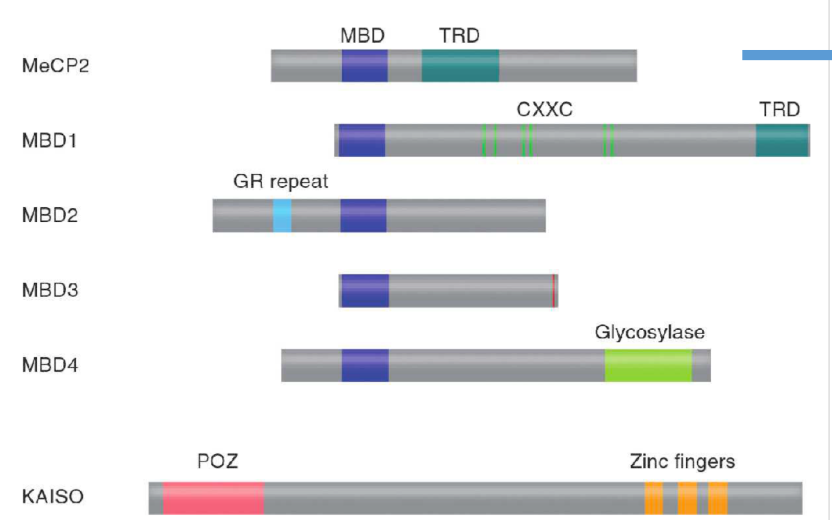


Table 2. Functions of methyl-CpG binding proteins

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)

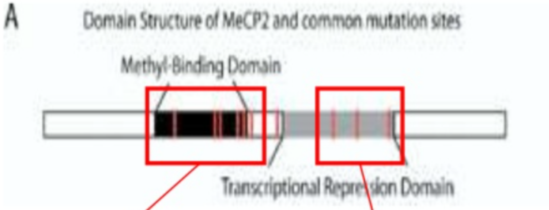


→ **RETT SYNDROME**

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.

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.



Change its ability to "see" meCpG

Genes do not get turned off appropriately

Change its ability to interact with the writers and erasers

Genes do not stay off

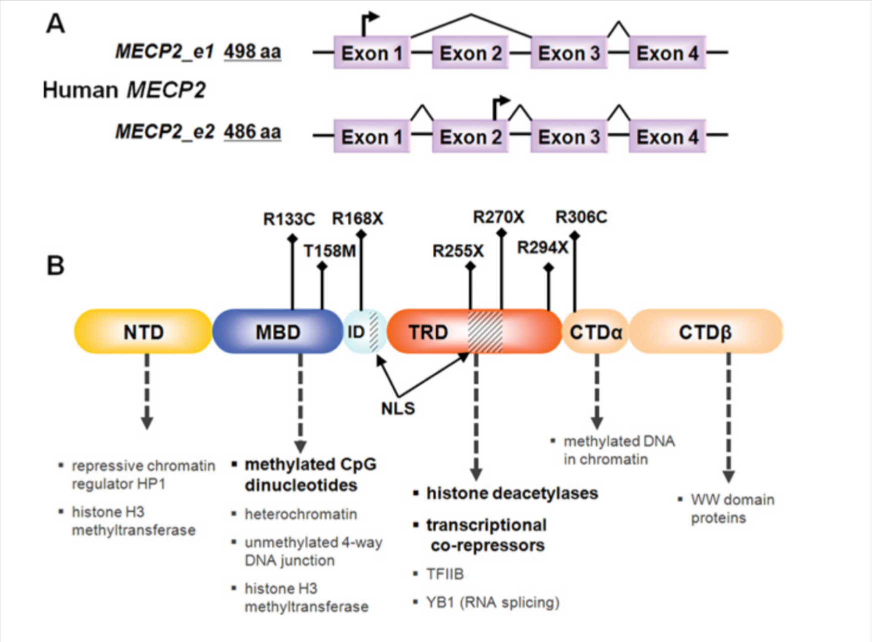


Figure 1 Composition of MeCP2: gene structure, splicing patterns and putative functional domains

(**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

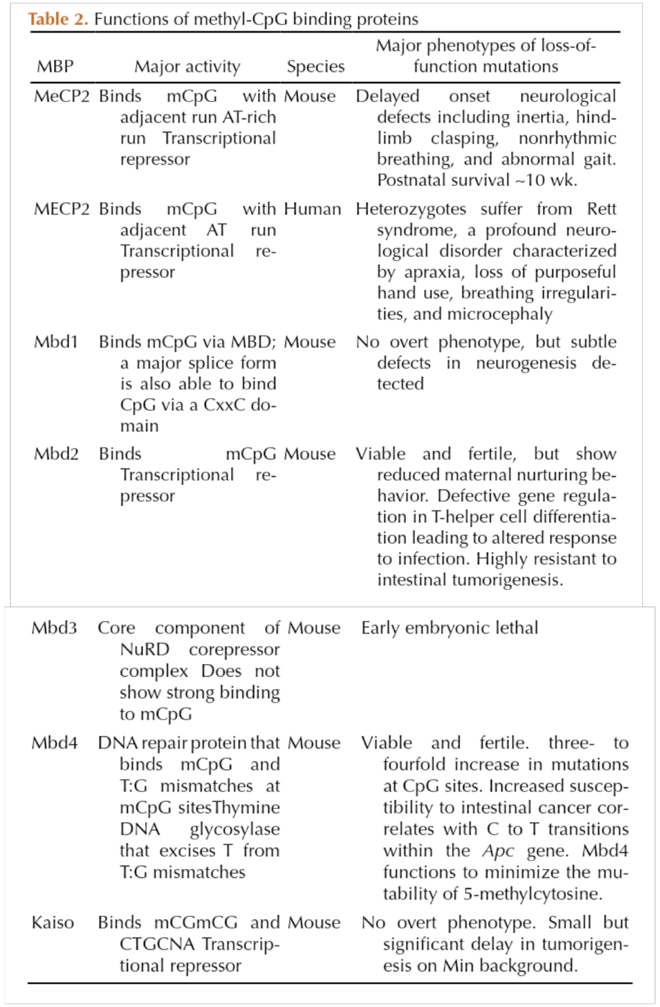


Table 2. Functions of methyl-CpG binding proteins

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.

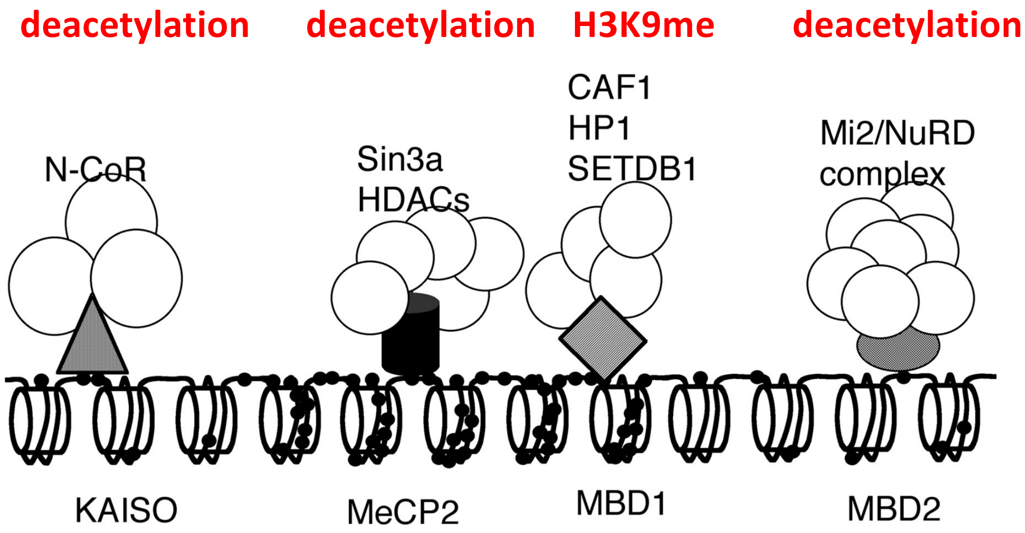
Methyl-CpG binding proteins are present in transcriptional co-repressor complexes

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



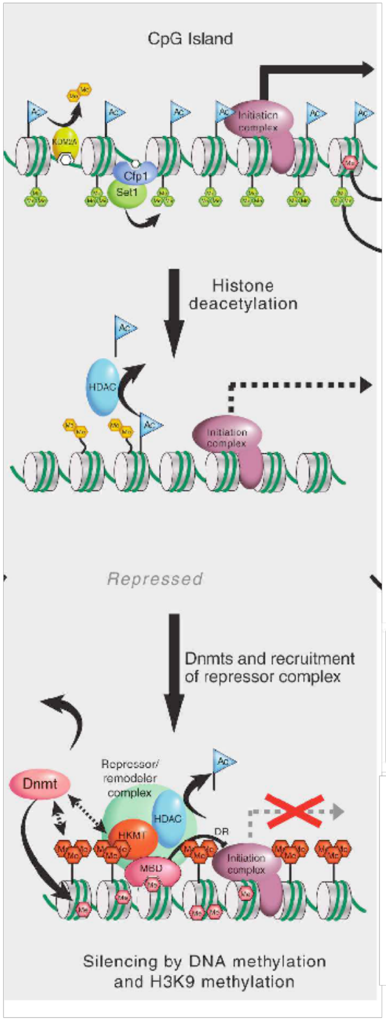
Collaboration to repress genes

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

Active

Transition

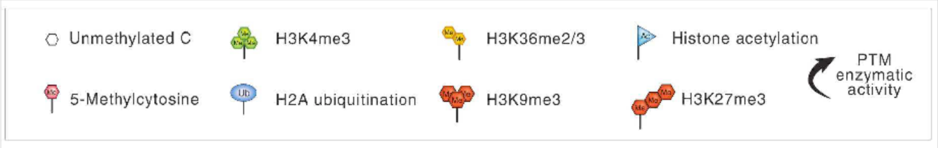
Inactive



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



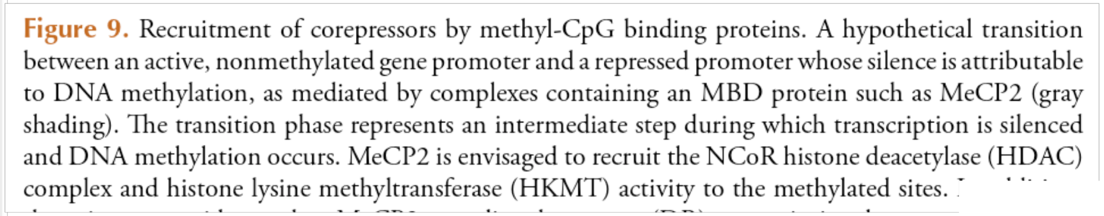


Figure 9. Recruitment of corepressors by methyl-CpG binding proteins. A hypothetical transition between an active, nonmethylated gene promoter and a repressed promoter whose silence is attributable to DNA methylation, as mediated by complexes containing an MBD protein such as MeCP2 (gray shading). The 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

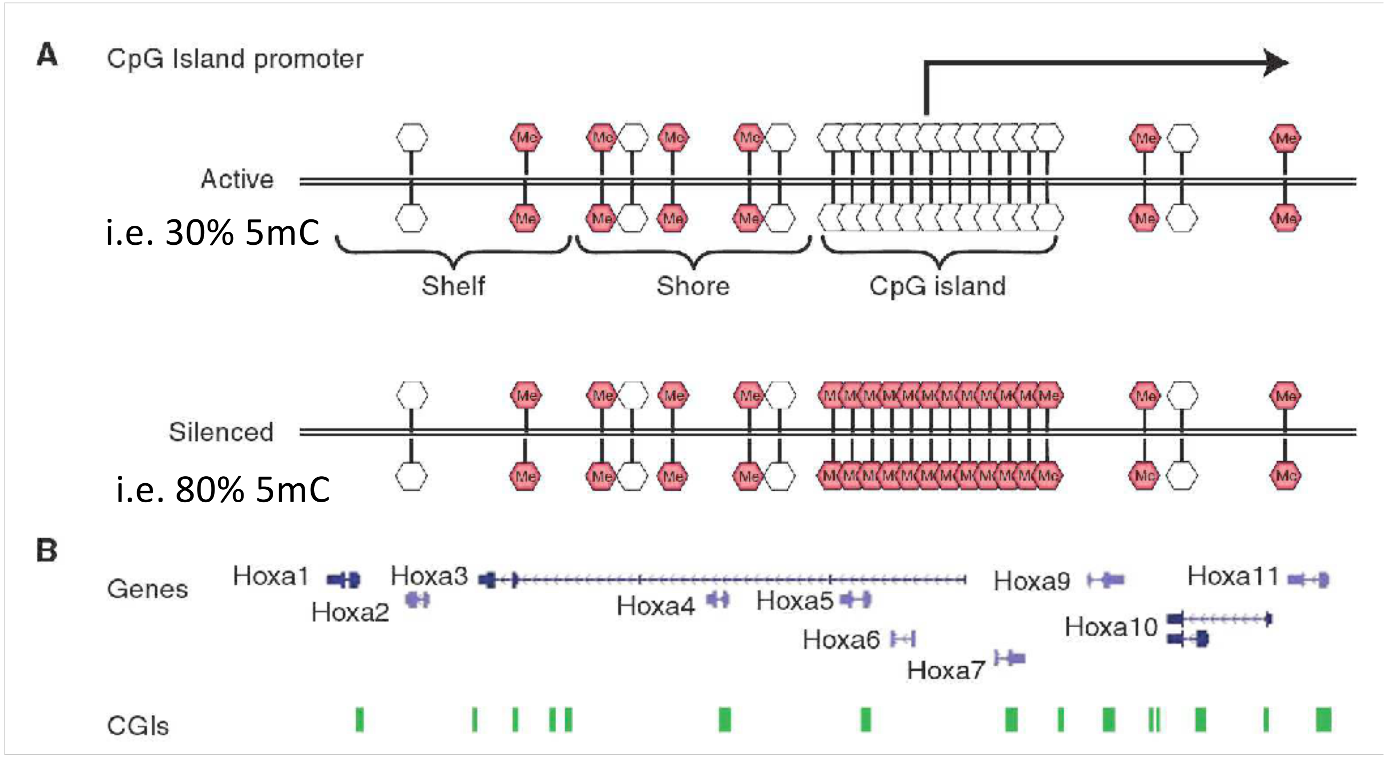
CpG islands (CGIs) have a length of ca. 1kb

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

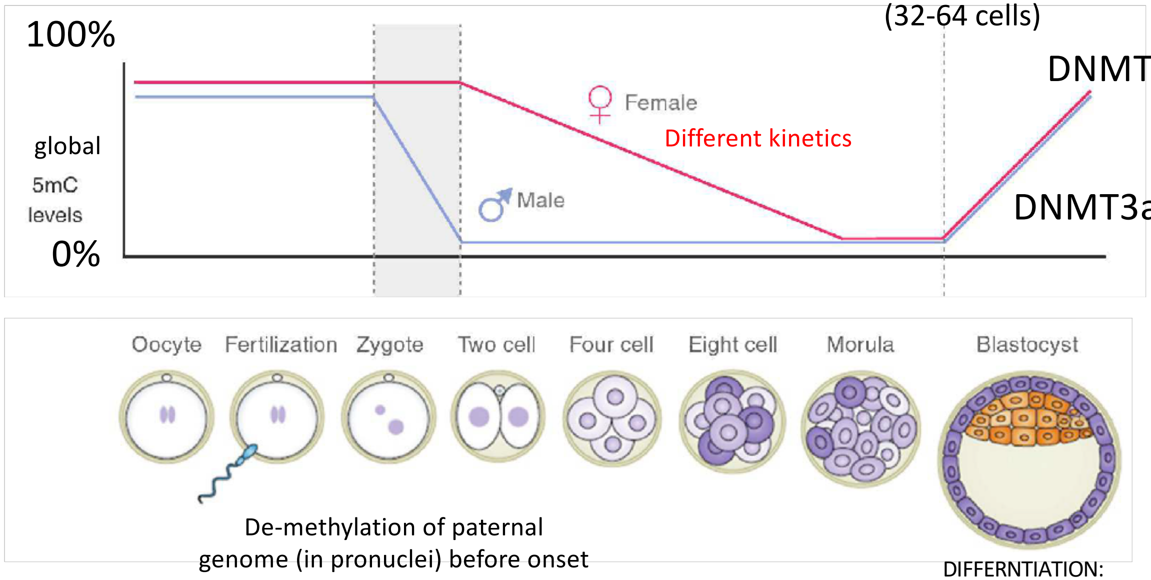
CpG islands located <2-4kb from promoter: shores



i.e. 30% 5mC

i.e. 80% 5mC

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



Early blastocyte stage (32-64 cells)

DNMT1

DNMT3a, 3b

Different kinetics

De-methylation of paternal genome (in pronuclei) before onset of DNA replication

DIFFERNTIATION:
formation of endo- meso- ectoderm
(3 germlayers)

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

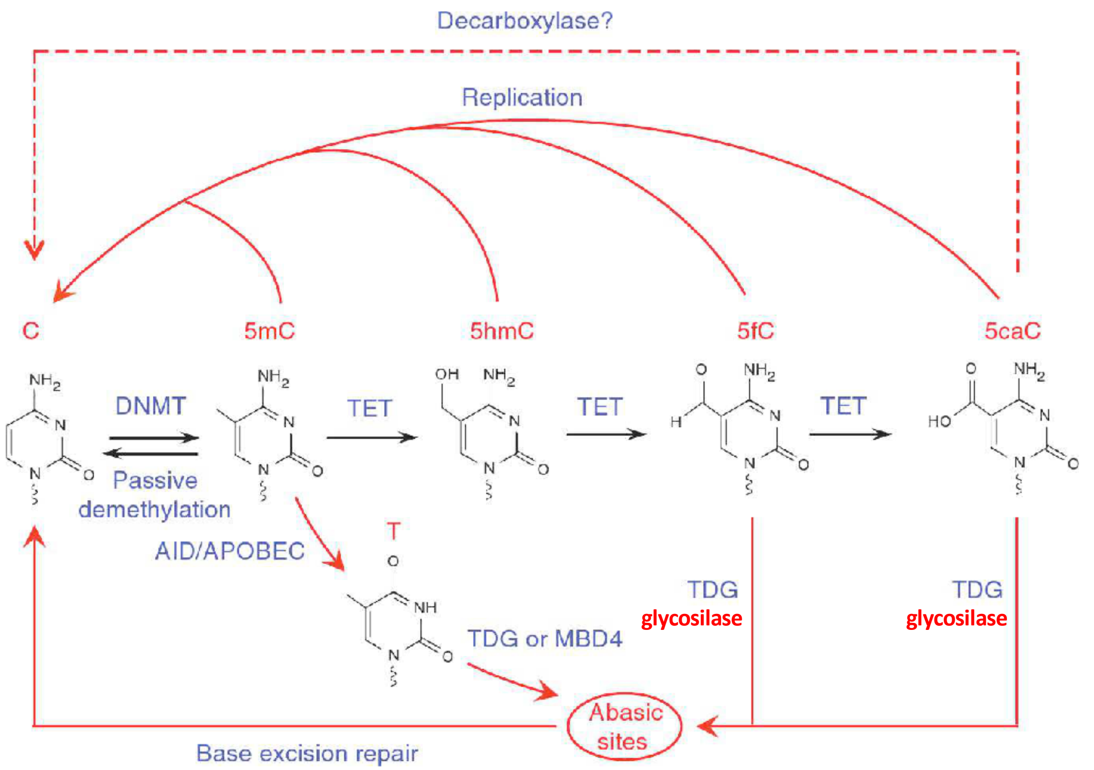


Figure 6. Model of Tet-initiated DNA demethylation pathways. DNA methylation (5mC) is established and maintained by DNMT. 5mC can by oxidized by Tet family of dioxygenases to generate 5hmC, 5fC, and 5caC. Because the oxidized 5mC derivatives cannot serve as substrates for DNMT1, they can be lost by replication-dependent 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

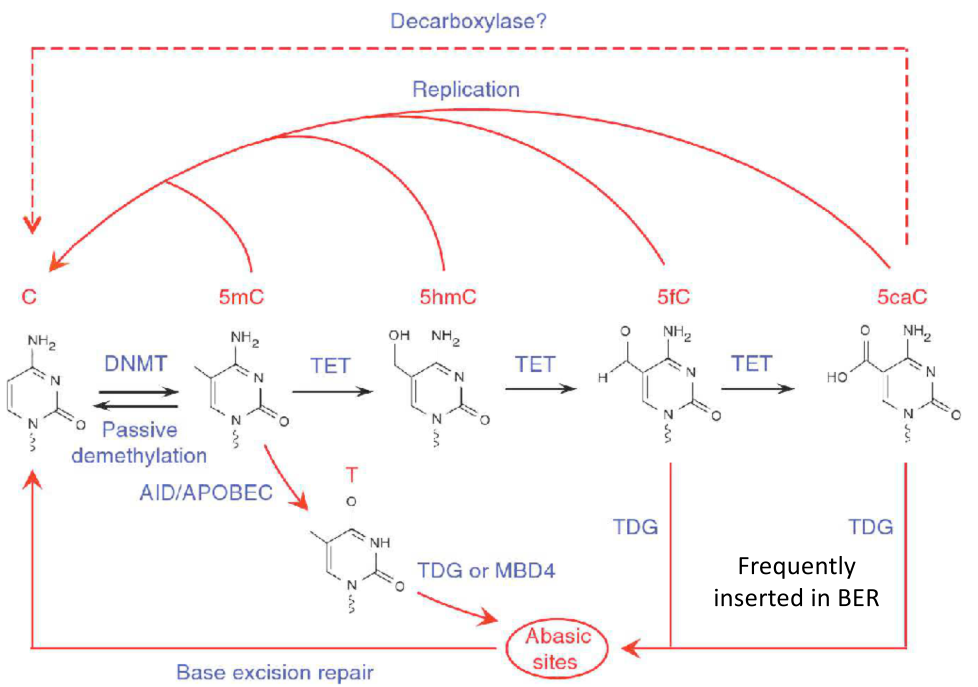
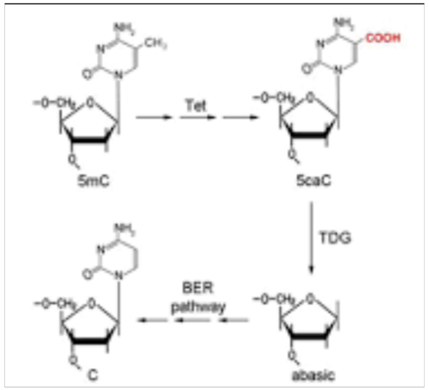
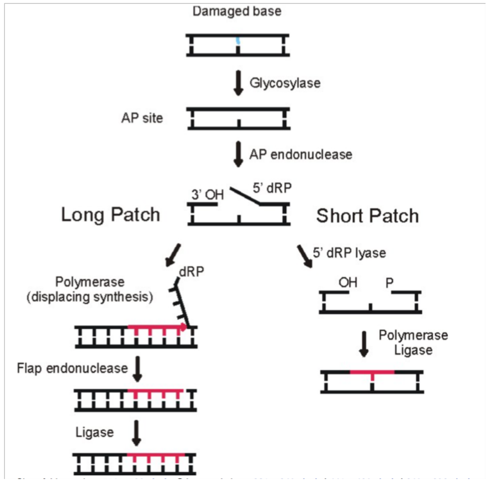


Figure 6. Model of Tet-initiated DNA demethylation pathways. DNA methylation (5mC) is established and maintained by DNMT. 5mC can by oxidized by Tet family of dioxygenases to generate 5hmC, 5fC, and 5caC. Because the oxidized 5mC derivatives cannot serve as substrates for DNMT1, they can be lost by replication-dependent 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.



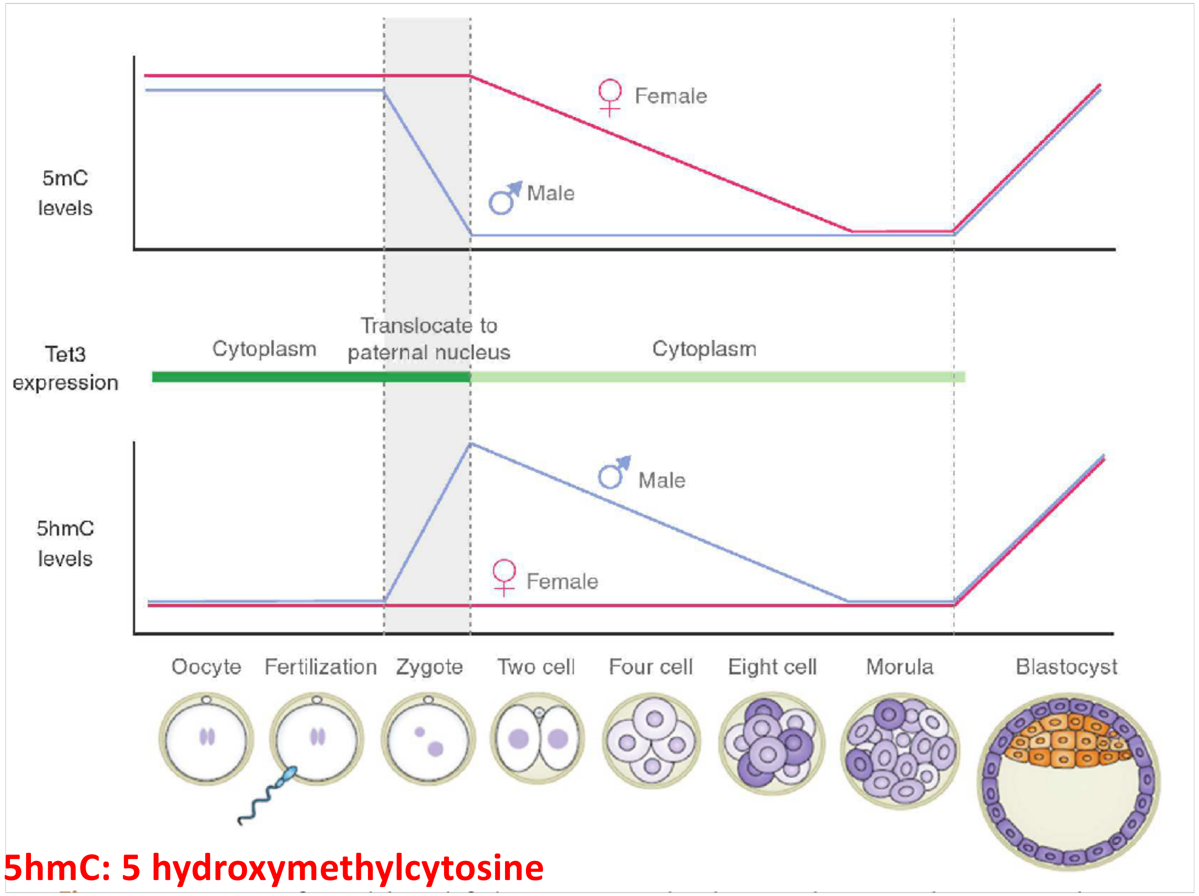


The protein encoded by this gene belongs to the TDG/mug DNA glycosylase family. Thymine-DNA glycosylase (TDG) removes thymine moieties from G/T mismatches by hydrolyzing the carbon-nitrogen bond between the sugar-phosphate backbone of DNA and the mispaired thymine. With lower activity, this enzyme also 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*