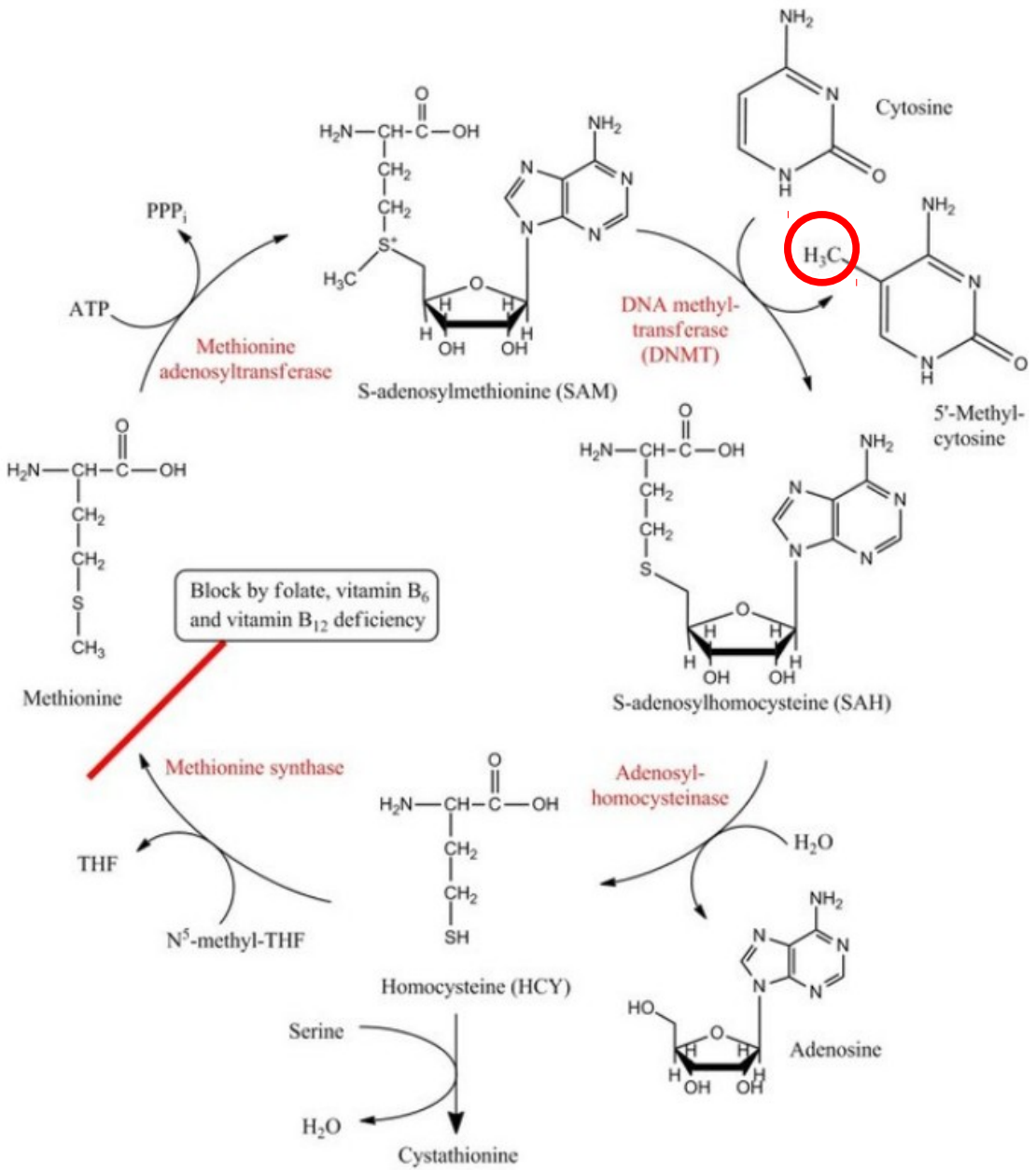
Part 2

DNA methylation

Roles of DNA methylation

- Long term silencing of genes
- Silencing of repetitive elements (e.g. transposons)
- X-chromosome inactivation
- Establishment and maintenance of imprinted genes.
- Suppression of viral genes / other exogenous elements
- Control of genome stability
- Control of apoptosis
- Deregulation leads to carcinogenesis.

DNA methylation occurs on cytosine



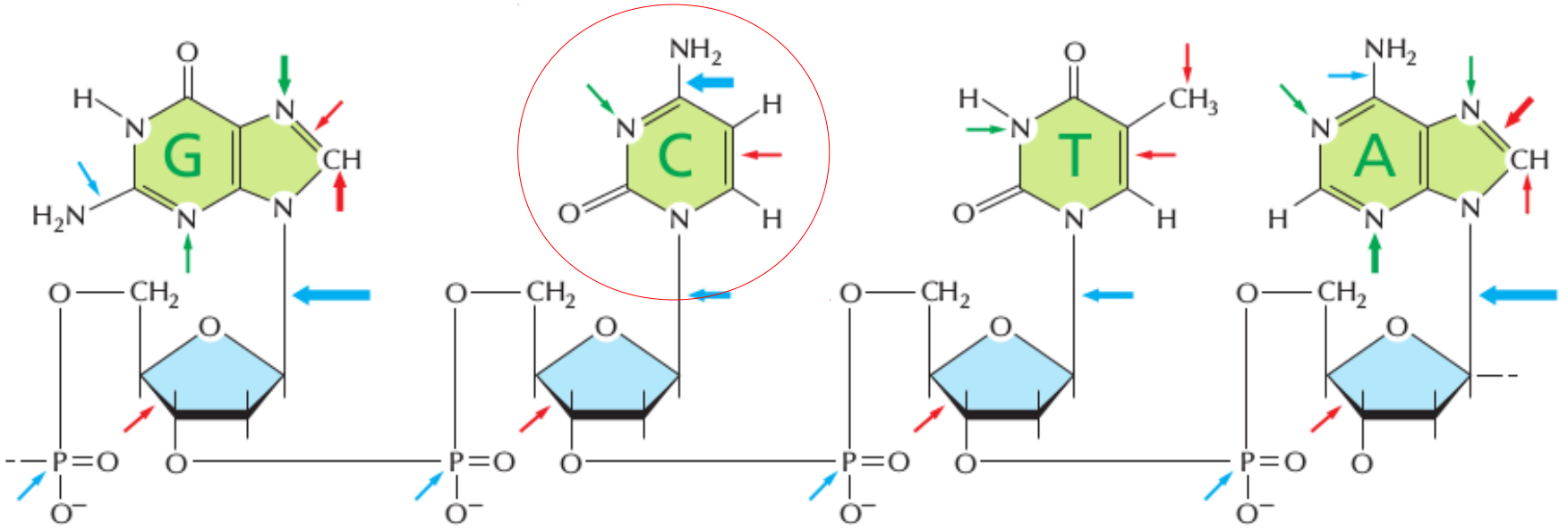
Lesions to the DNA

TABLE 5–3 Endogenous DNA Lesions Arising and Repaired in a Diploid Mammalian Cell in 24 Hours

DNA lesion	Number repaired in 24 h
Hydrolysis	
Depurination	18,000
Depyrimidination	600
Cytosine deamination	100
5-Methylcytosine deamination	10
Oxidation	
8-oxo G	1500
Ring-saturated pyrimidines (thymine glycol, cytosine hydrates)	2000
Lipid peroxidation products (M1G, etheno-A, etheno-C)	1000
Nonenzymatic methylation by *S*-adenosylmethionine	
7-Methylguanine	6000
3-Methyladenine	1200
Nonenzymatic methylation by nitrosated polyamines and peptides	
O^6-Methylguanine	20–100

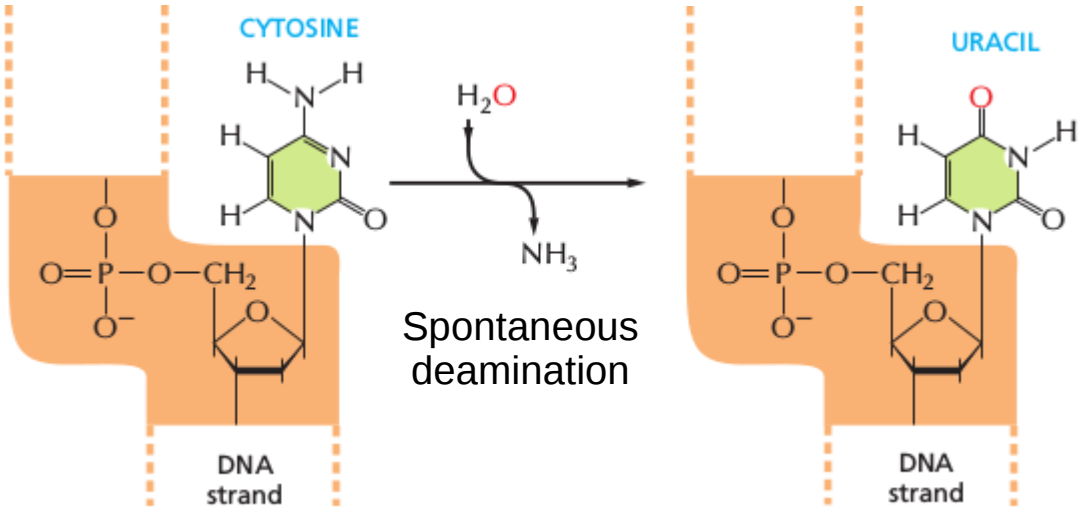
The DNA lesions listed in the table are the result of the normal chemical reactions that take place in cells. Cells that are exposed to external chemicals and radiation suffer greater and more diverse forms of DNA damage. (From T. Lindahl and D.E. Barnes, *Cold Spring Harb. Symp. Quant. Biol.* 65:127–133, 2000.)

Spontaneous modifications of nucleotides

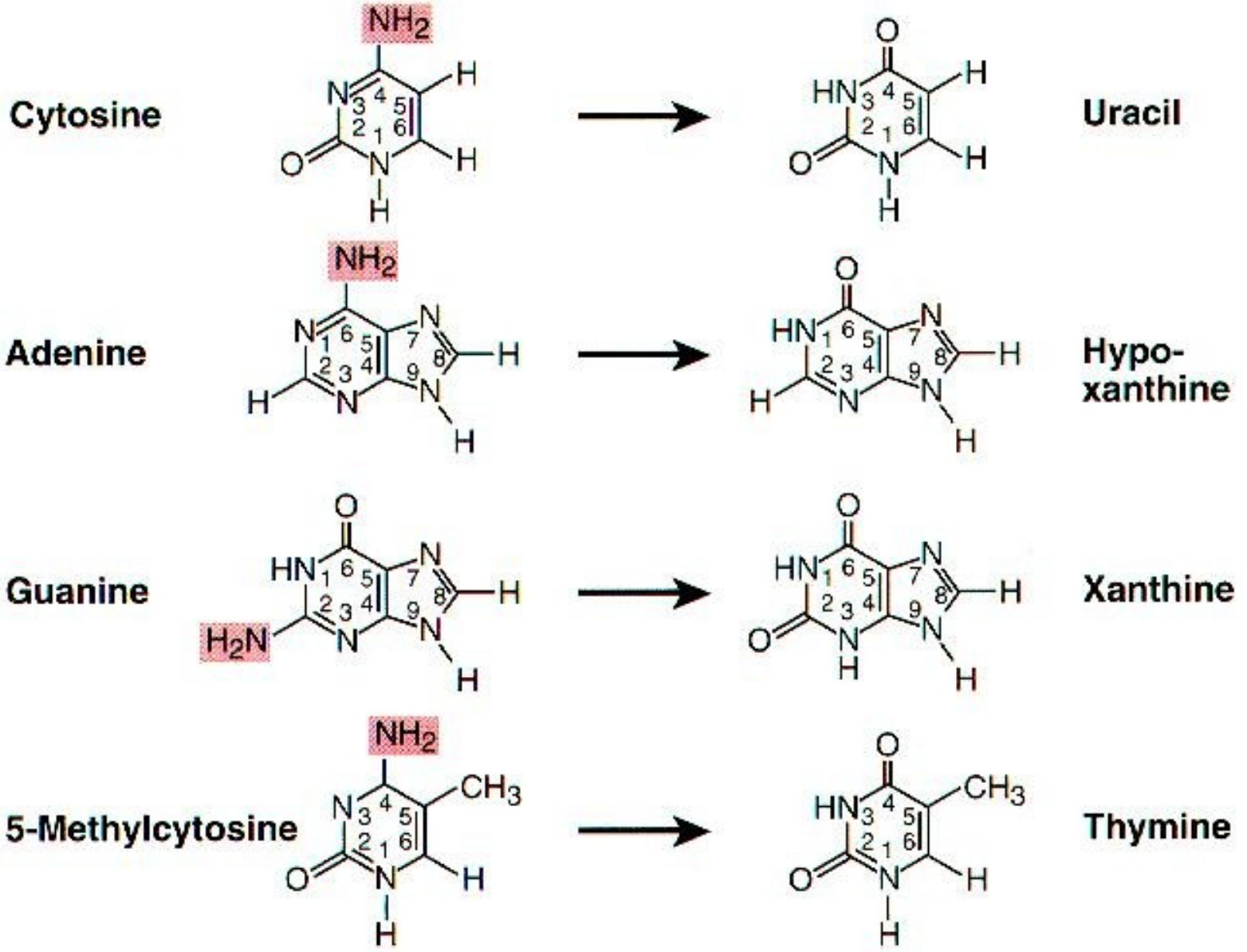


oxidative damage
hydrolytic attack
Uncontrolled methylation

Width of arrow = frequency



Effects of spontaneous deaminations



The peculiarity of the CG dinucleotide in vertebrates

Human G+C: 42%.

Expected CG dinucleotide frequency = 4.41%.

Observed CG dinucleotide frequency = 1%

} 4 times lower than expected !

Why this bias exists in vertebrates?

1. 5mC tends to deaminate to T.
2. the repair system (BER), specific for CG dinucleotides, is inefficient

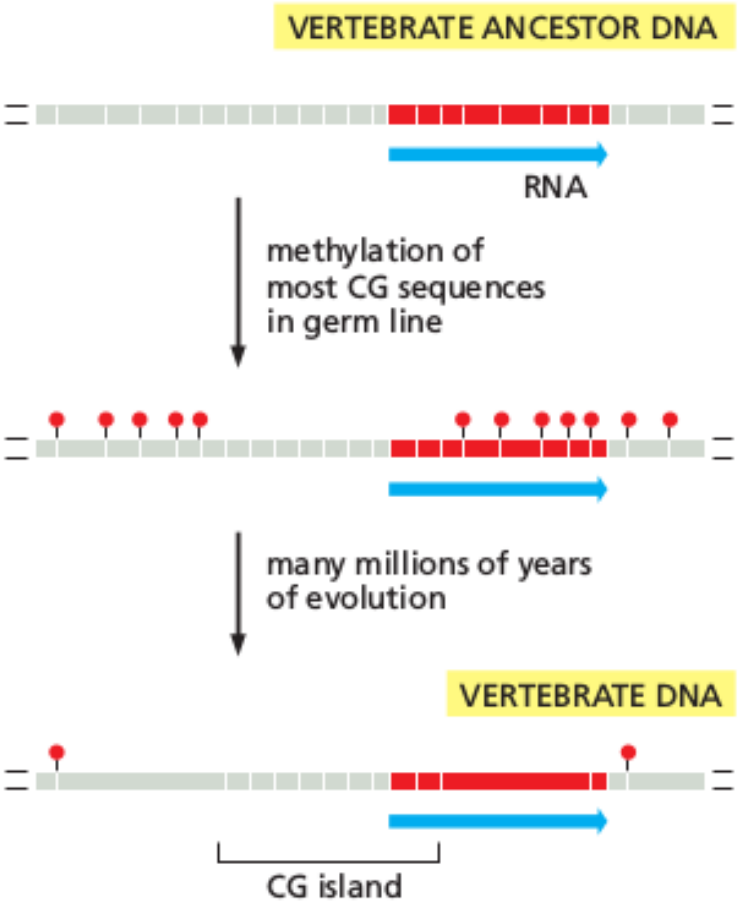
=> GCs tend to be cleared

During the course of evolution, more than 3/4 CGs have been lost in this way, leaving vertebrates with a remarkable deficiency of this dinucleotide.

...BUT...

CG nucleotides are overrepresented in specific regions of the vertebrate's genome:

the CpG islands.



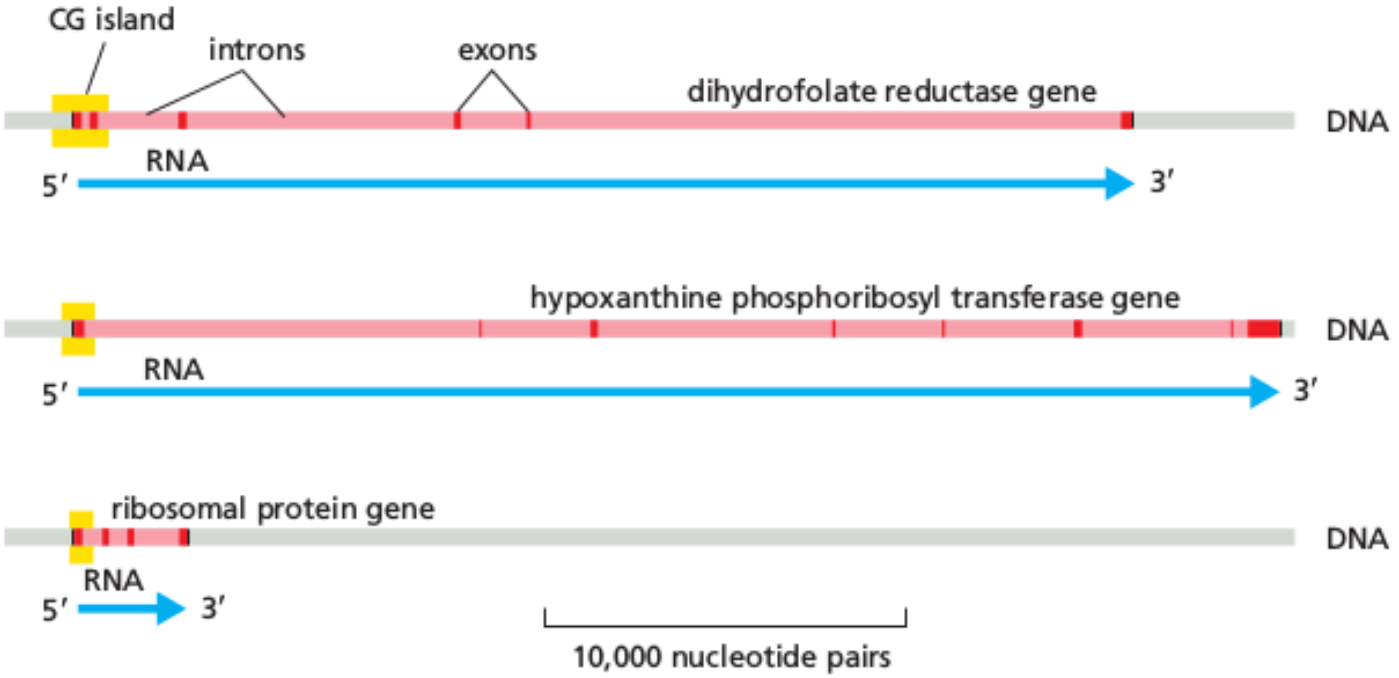
CG dimer distribution in genomes: the CpG islands

The CG sequences that remain are very unevenly distributed in the genome; they are present at 10 times their average density in selected regions, called CpG islands, which average 1000 nucleotide pairs in length.

The human genome contains roughly 20,000 CG islands and they usually include promoters of genes, among which:

- 60% human protein-coding genes

- virtually all housekeeping genes (genes essential for cell viability)



CG dimer distribution in genomes: the CpG islands

Definition: a region >200 bp, and a G+C percentage that is greater than 50%, and with an observed-to-expected CpG ratio that is greater than 60 %

Random genome region

AATTTCAAATATGTGCAAAACTAAATGATATATTGTTTATGGACACAGACATAGGACTAT
AAAGAAAGCAAGGAATGGTGGACAAAACATTTGAGATAATGATCACTTCTAAAGAAAAGG
AAAGAAAAGGGCTTAGGATCGGGCCCACAGGGTGCTTGTAATCTATTCATAATATTCTAG
CTCTTAAACTGGTTTCATGGGTGGATTCTATGTGTGTTCATATTATAGTAATCATTTTAC
AGTCTTATAAATATTATCTTGTAATGGTTCCACATTTGCTAAAACAAGAAACAATGTAAT
TAAAATGTATCTAGATATTGAAAACAATTTAAGGCAAAGGAAATGTTACATTTTATAATT
ATTTATTGGGCCACTCATCTTCCCTGGAAACAATAATTTTAACTATATATCAACTAACCT
CTGTATACTGTTTCCTTTTTTCACTGTATTTTTTAAATTTTATTAGTTATCTCTCTTACG
GCATTCTAGTACTAAAGAGGAGAGACCACGGATACTTAGTATATGGCCTTCTGCCAAAAT
TCTGACGCAGGTTTCAAAGTTAGGACCTCATTTACATCTGAGTGGAAACTGCTACTTGAA
GGTAAAGGCTAATTCTATTAAACCTAATATAAAGCTTAATTTAAATCTATTTTTAAGTCA
GTGTTGGTCTGTGTATCATAAAGAGGGCACAGATAGGTAAGGGACAGAAAGCTGGAAGAG
ATTAATTAAAATAGGGAAGAAACACTTAAGATCTACCCTCTTGGCAAATTTCAAGTATAC
AATACCGTATCGTTAACTTTAGTCATTAACTTTAGTCGTTAAAGTTATCGTTAACTTTGA
TGTTGTACATTAGGGTCTCTAGAAAGTATTCGTCTTGCATAACTGAAACCTTGTACCTCT
TGACCAACATCTGCCCACTTCCCATTGCCCCTAGCCCCTGGCAACCACTATTCTACTCTC
TCCTATGAGTTTGACTATTTTAGAGTTCATATATAATTGAGATTATACAGTTTTTAATTT
ACAATGTATGCTTATATCAAGTTATACATCTTTAATATATGCAATTTTAAATTGTCAATT
ACACTTCAACAGAGTTGGAAAAAATAAAAATAAGAAAGACATTGTTATTATTAGTAAAAA
GTAATACATTTGTCATATTCCAAAAAATAAAAAACAGGGAGGTCAAGTCATGAATGGCTA
CGTCAGATGGGCTGGGCCTTGACTGCGTTCGCCACAGAATTGGAAATTTAAGATAAAGGT
GCACTAAAACTTAAGGACAAATTTAAAGGGCCATTCTAATTCCGTATTCTGATAACATGT
CAGAATAGGCAAGATTCTGATAACATGTCAGAATAGGAAAGCAAACTTTAAAGGCTCAGA
GCCAGAGATCAAGCCAAAATTCAGACAGCAAGAGTCAGGTATCCTGACAAGGGAGTAGGG
CTGTATACTGTTTCCTTTTTTCACTGTATTTTTTAAATTTTATTAGTTATCTCTCTTACG
GCATTCTAGTACTAAAGAGGAGAGACCACGGATACTTAGTATATGGCCTTCTGCCAAAAT
TCTGACGCAGGTTTCAAAGTTAGGACCTCATTTACATCTGAGTGGAAACTGCTACTTGAA
GGTAAAGGCTAATTCTATTAAACCTAATATAAAGCTTAATTTAAATCTATTTTTAAGTCA
GTGTTGGTCTGTGTATCATAAAGAGGGCACAGATAGGTAAGGGACAGAAAGCTGGAAGAG
ATTAATTAAAATAGGGAAGAAACACTTAAGATCTACCCTCTTGGCAAATTTCAAGTATAC
AATACCGTATCGTTAACTTTAGTCATTAACTTTAGTCGTTAAAGTTATCGTTAACTTTGA
TGTTGTACATTAGGGTCTCTAGAAAGTATTCGTCTTGCATAACTGAAACCTTGTACCTCT
TGACCAACATCTGCCCACTTCCCATTGCCCCTAGCCCCTGGCAACCACTATTCTACTCTC

(5') flank of a gene

CGTTTCTATAATGGACATGTTGGTTCCTTGTATTTTAAAACGCAGCCACAAACTGAGGCG
ATAGCGATCACGTTATGCAGTCCCCAATCGTCCAAGCCGAATAGGCTGCATGAAGTATCA
CTCAAAAACATTCATTACCGTCGGTGTGTTGAGTGGTAACGCCAGTGTGCGAATTCATTA
ATGGAAATCACGGGACCCCGAGCGGTTGGAGATCTGCCACACCCCTCTGCCTGCGTTAAG
ACAGCCTGACCCGACCTCTGGTAAATTAATGCGCAAAATGGCGGTTGCACGTCTGTTTCC
CCACCCCGGCGCACAATCCTGAGGAAGGACGTCACGTCACCAGGGCAACGACGCTCTTGC
GTAAAGGCCCGGCCCAAGGGAACGTTCAGGGCGTCTCGGCTTTCCCCGCTGCTGCTTCTG
CTAGGCCCAGTGCGAGACCAGAGCACGAGCGACTCCCGTCGTCCCCGGCCAGGCAGATGT
TGGCCTAGTCCTGGCGCGAACGAAGCGCGCTATTTCCCTGCTTCCTCTAGGCCAAGCCTG
CTTTACGGCAGGGCCCGCCTCGGGAGCGAGCACAGACCGGGGCAGCGAGGCCAGCCAGGC
GCCGACGAGGTCCCCGAACGCGCACGCGCTCCGTTCAGCTCCGGGTGGCGGCCGCCGGAG
TAGACGTTAGCCATGGAAACCGAGAGCTGGCCCGGGCGGGGCCGCGGTGAGCTCGTTATT
CGGCCGCCGCAGCTTTTCTGCCTCCGCATTCGGGCACTAACCAACCTCCCGGCGGGAGCG
CCCAGCCCGACTGAGACAACACCAGAATCCCTTGAGTCCTGCGTAAGATGTCAGAGTCCT
CAGGAGCACCGCCGACTCCAGTGCTCTCTTAACAGTTTGGCACTCCCTGGAGAGGTTTCC
TAGCCTGAATGGCAAGTCCATGCCATCTTCCGACCAGTACAGAAGACAAAGGGGCTTTTT
AGCCCTCGCCTCTTCCTCCCCAACCCGCAGCCCCTGCGTAAGGGGCAGCCGGGCCTTCGG
GAAGCTCCACGCAACCCACAAAGGACCCGGGATCACTGCAGAACCCCCGGAGACCCGGAC
ACGCCTCGGCAGCTCCGAGCCCGGCCTCCCACTCGGGAGAGGCTTTGGCCGGTCCTCACC
TGGGCGTCGTCCCTGCCACTCGGCCCATAGTAGCGTCAGGTCGCCGTCAGCCCTCAGGAA
GCGCAGCAGCTGCACCAAGAGCAGGAGCAGCGCGCACAGCACCAGCAGCCACAGCAGCAG
CTCCCAGTTCATTGCGGCCGCGCACGCCCAGCTCGGGGGGAAGAAGACGGCCCGCACCAG
AGTCGCGTCGCTGCCCTGCGGGATCGCAGCGCCACCCCTTCGGCCAGCCCAGAGCCGCAC
TGCCCCGGCTCCGCCCAGGGAGCAGGCCGGGAGGAGGAGCGGCTCGGGCGCGCCGGGCTC
GTAAAGGCCCGGCCCAAGGGAACGTTCAGGGCGTCTCGGCTTTCCCCGCTGCTGCTTCTG
CTAGGCCCAGTGCGAGACCAGAGCACGAGCGACTCCCGTCGTCCCCGGCCAGGCAGATGT
TGGCCTAGTCCTGGCGCGAACGAAGCGCGCTATTTCCCTGCTTCCTCTAGGCCAAGCCTG
CTTTACGGCAGGGCCCGCCTCGGGAGCGAGCACAGACCGGGGCAGCGAGGCCAGCCAGGC
GCCGACGAGGTCCCCGAACGCGCACGCGCTCCGTTCAGCTCCGGGTGGCGGCCGCCGGAG
TAGACGTTAGCCATGGAAACCGAGAGCTGGCCCGGGCGGGGCCGCGGTGAGCTCGTTATT
CGGCCGCCGCAGCTTTTCTGCCTCCGCATTCGGGCACTAACCAACCTCCCGGCGGGAGCG
AGCCCTCGCCTCTTCCTCCCCAACCCGCAGCCCCTGCGTAAGGGGCAGCCGGGCCTTCGG
GAAGCTCCACGCAACCCACAAAGGACCCGGGATCACTGCAGAACCCCCGGAGACCCGGAC

CpG islands and methylation status

Given an organism, the methylation status **within** CpG islands is peculiar:

- some positions are constantly methylated

- some positions are constantly unmethylated

- some position are methylated or not ***depending on the cell type***.

The unmethylated state is maintained by sequence-specific **DNA-binding proteins** that

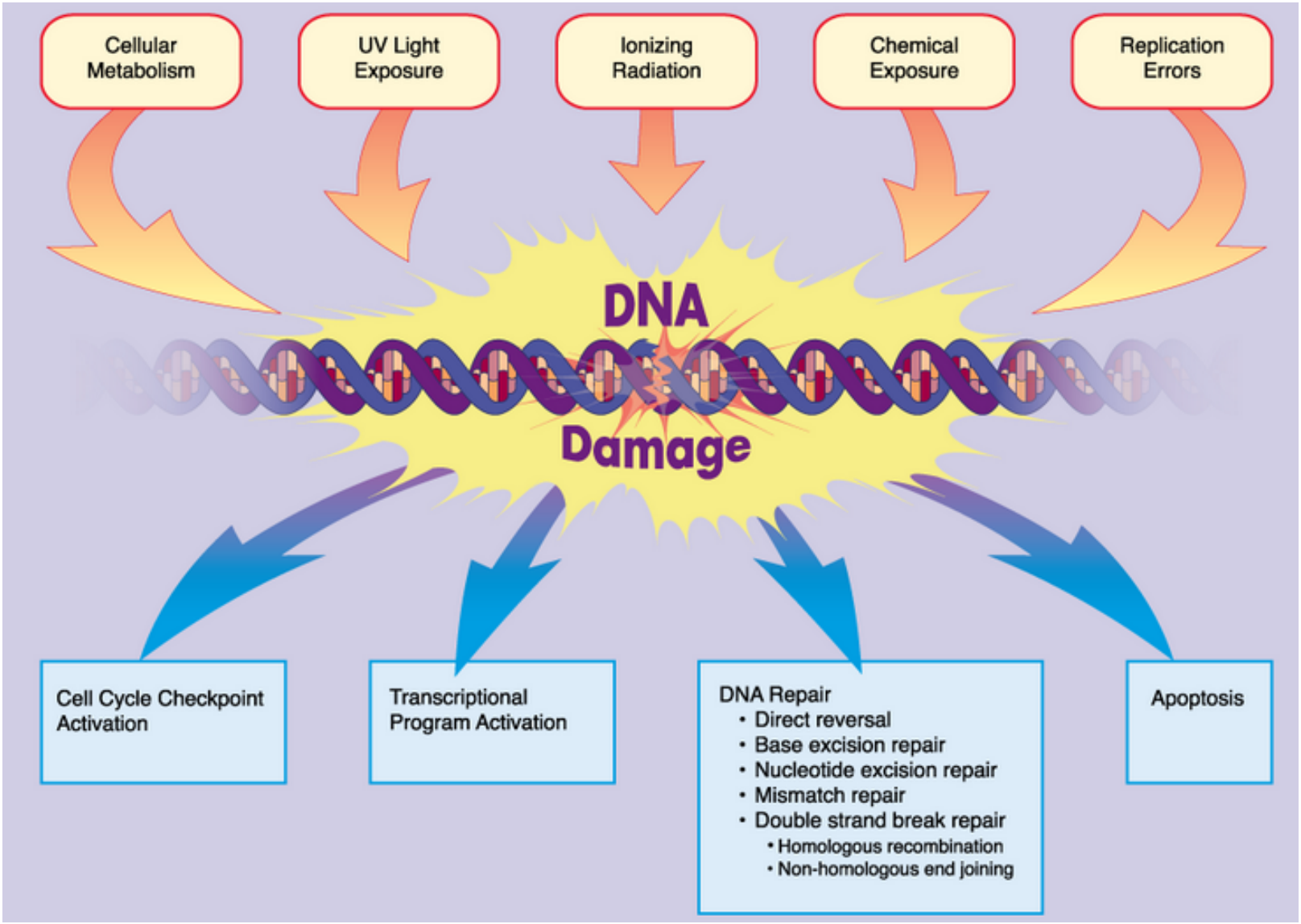
- protect the DNA from methyl transferases.

- recruit DNA demethylases

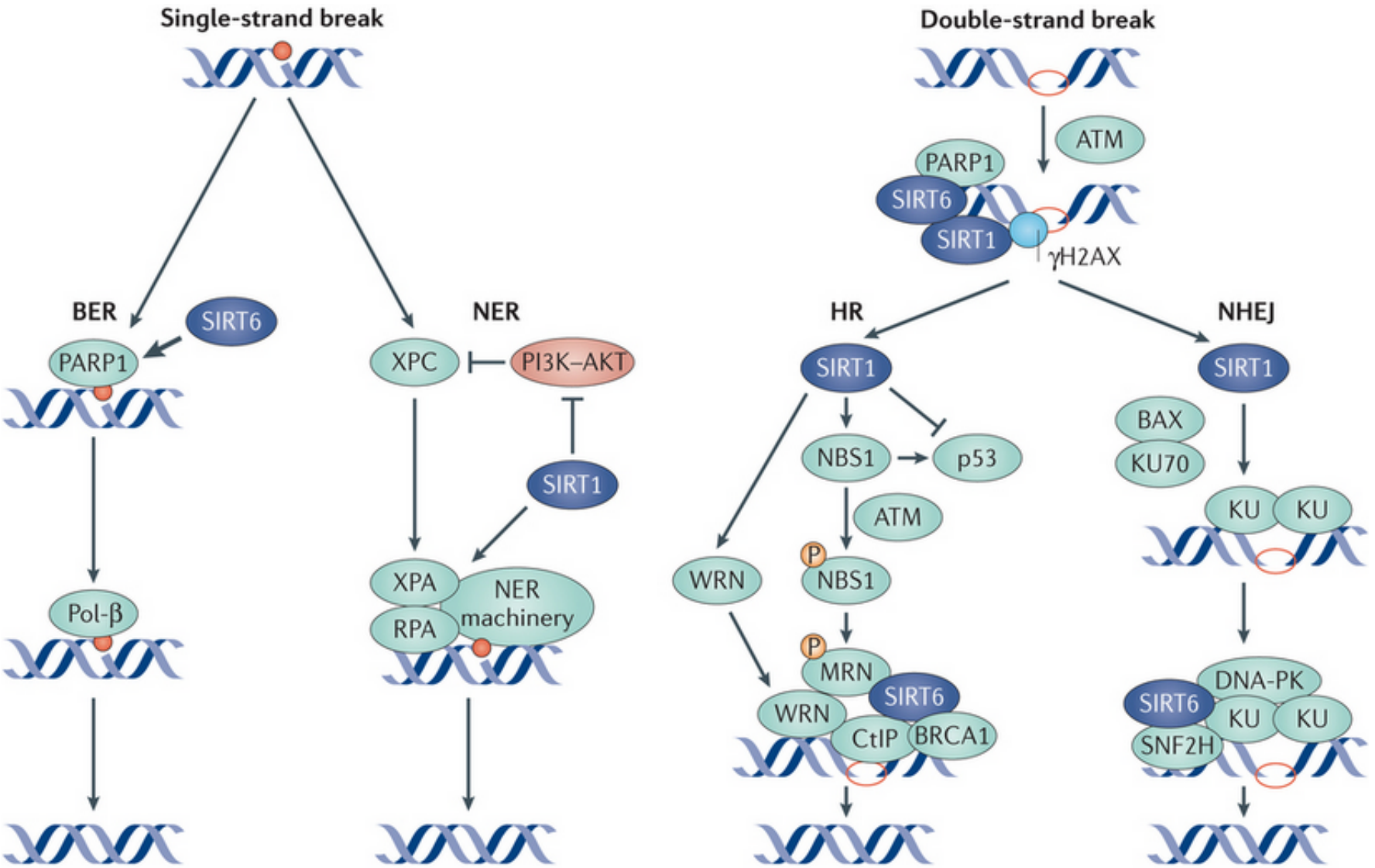
In practice, CG dinucleotides in these islands were saved by evolutionary destructions because the C remains unmethylated in the germ line.

As with DNaseI sensitivity, under-methylation is a consequence of commitment to a particular pattern of gene expression, and is associated with the change in chromatin structure observed in active or potentially active genes.

Genome surveillance



Genome surveillance: BER, NER, HR, NHEJ



The DNA Mismatch Repair (MMR) pathway

Replicative polymerases are exceptionally fast and precise, but some incorporation error may remain. The MMR machinery operates **during replication** and involves MutS and MutL proteins.

The repair process consists of the following steps:

A. MutS (hetero-dimer of MSH genes) slides through the replicated double helix just after the polymerase and identifies mismatches (DNA distortions)

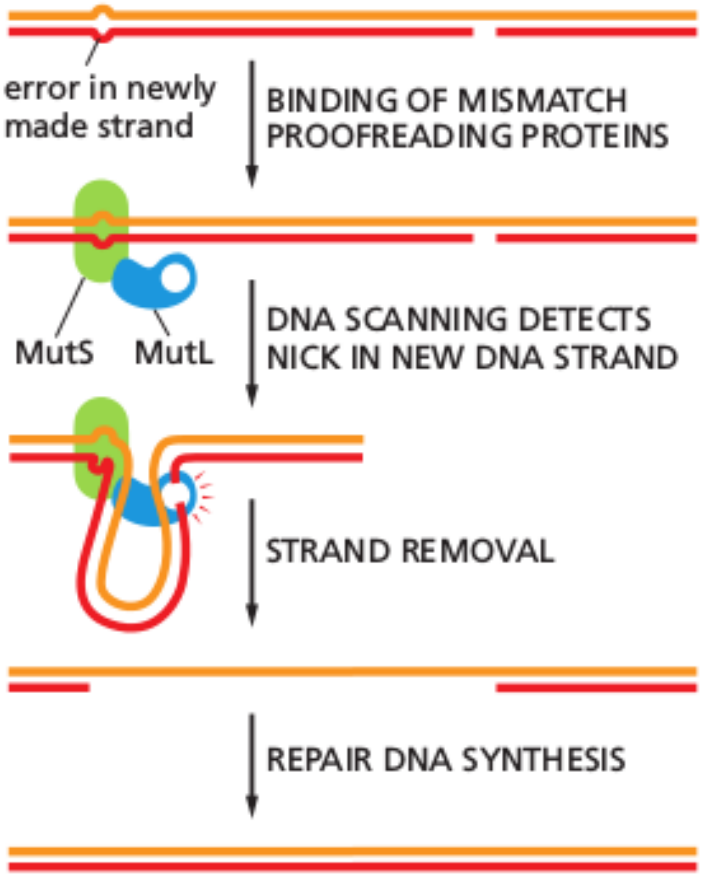
B. MutLα (MLH1-PMS2 hetero-dimer with ATPase activity) is recruited to lesion point

C. PCNA (Proliferating Cell Nuclear Antigen, a modulator of polymerase δ) guides MutLα to the daughter strand

D. MutLα nicks the daughter strand, leading to strand removal

E. Repair DNA polymerase fills the gap

F. DNA ligase joins extremities and closes the nick.



The base excision repair (BER) pathway

A. A point mutation is present in DNA

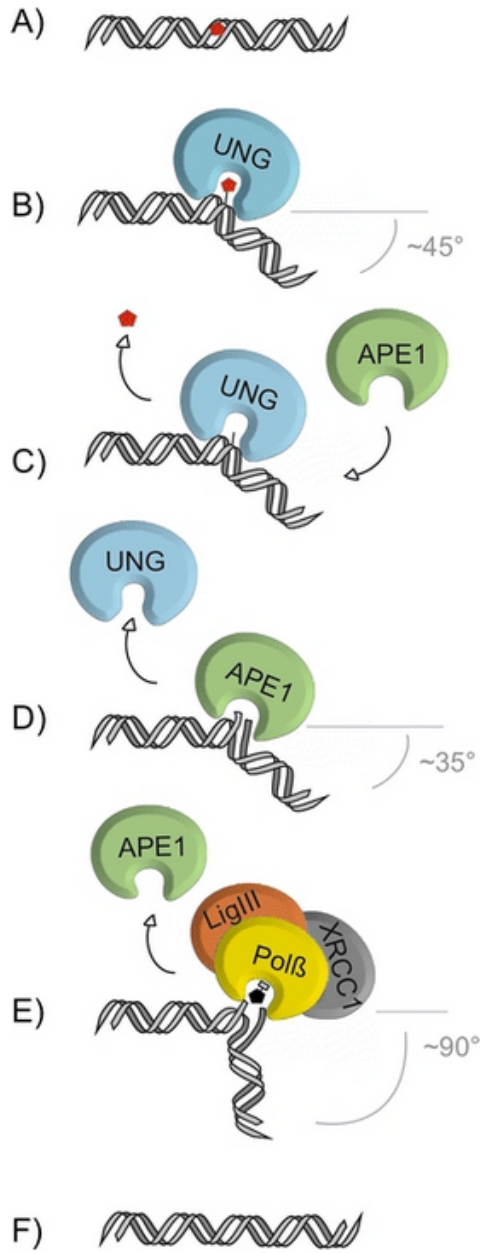
B. The mutation specifically recruit a **DNA glycosylase**, that flips the base out of the base-stack into its catalytic site pocket where specific contacts examine the substrate base and position it for nucleophilic attack to the N-glycosidic bond

C. DNA glycosylase detaches the excised base, leaving an empty space called the "**AP site**" (apurinic/apirimidinic)

D. The AP-endonuclease, **APE1**, cleaves the phosphate backbone 5' to the AP site, leaving a 3'OH and a 5'dRP

E. Polymerase β (**Polβ**) hydrolyzes the 5'dRP and fills in the single nucleotide gap. DNA ligase III (LigIII), supported by the scaffold protein XRCC1 reform the covalent bold.

F. The original sequence is restored.

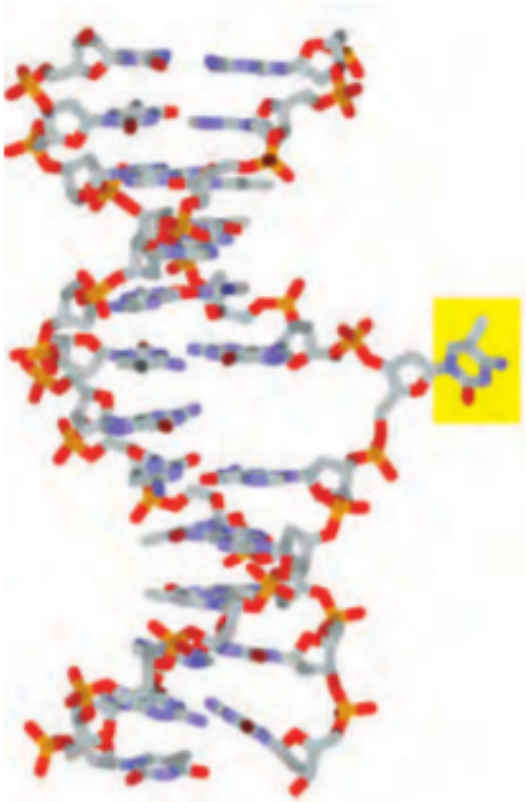


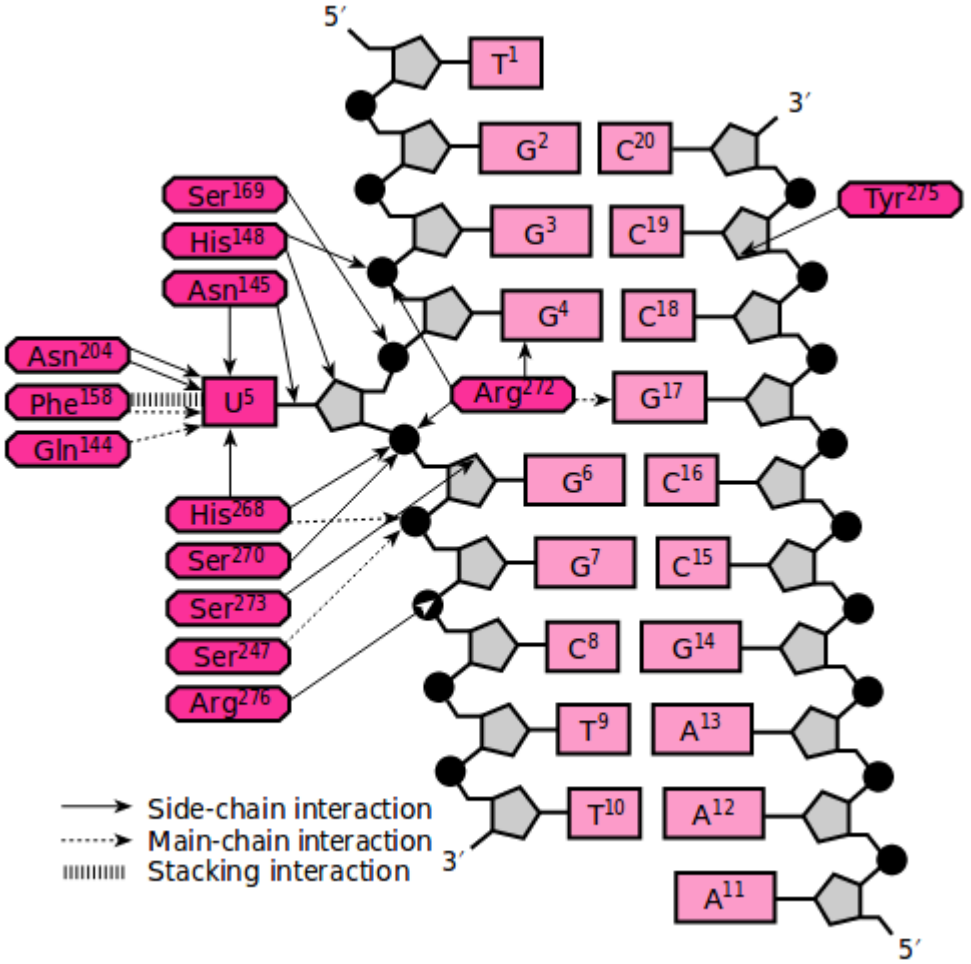
Note 1. The DNA glycosylase does not leave the AP-site until APE1 is present, and favors its attachment.

Note 2. DNA is bent during the catalytic action of DNA glycosylase. This favor the exposure of nearby DNA sequence, supporting the arrival of repair proteins.

DNA glycosylases flip out bases from DNA

DNA glycosylases are able to interrogate the surface of DNA bases by flipping them out of the helix into a selective active site pocket.





Kockran et al. 1997 Biochem. J. 325, 1-16

DNA glycosylases: the DNA repair swiss-knife

At least 11 DNA glycosylases have been identified in mammals, that can be classified into four structurally distinct super-families:

- uracil DNA glycosylases (UDGs)

- helix-hairpin-helix (HhH) glycosylases

- 3-methyl-purine glycosylase (MPG)

- endonuclease VIII-like (NEIL) glycosylases

Type of base lesion		Name	Physiological substrates
Uracil in ssDNA dsDNA	UNG	Uracil-N glycosylase	U, 5-FU, ss and dsDNA
	SMUG1	Single-strand-specific monofunctional uracil DNA glycosylase 1	U, 5--hmU, 5-FU, ss and dsDNA
Pyrimidine derivates in mismatches	MBD4	Methyl-binding domain glycosylase 4	T, U, 5-FU, εC, opposite G, dsDNA
	TDG	Thymine DNA glycosylase	T, U, 5-FU, εC, 5-hmU, 5-fC 5-caC; opposite G, dsDNA
Oxidative base damage	OGG1	8-OxoG DNA glycosylase 1	8--oxoG, FaPy, opposite C, dsDNA
	MYH	MutY homolog DNA glycosylase	A opposite 8--oxoG, C or G, 2--hA opposite G, dsDNA
Alkylated purines	MPG	Methylpurine glycosylase	3--meA, 7-meG, 3-meG, hypoxanthine, εA, ss and dsDNA
Oxidized, ring-fragmented or --saturated pyrimidines	NTHL1	Endonuclease III-like 1	Tg, FaPyG, 5-hC, 5-hU, dsDNA
	NEIL1	Endonuclease VIII-like glycosylase 1	Tg, FaPyG, FaPyA, 8-oxoG, 5--hU, 5--hC, ss and dsDNA
	NEIL2	Endonuclease VIII-like glycosylase 2	As NTHL1 and NEIL1
	NEIL3	Endonuclease VIII-like glycosylase 3	FaPyG, FaPyA, prefers ssDNA

Jacobs and Schär 2012 Chromosoma 121(1):1-20

The nucleotide excision repair (NER) pathway

A key characteristic is that it can recognize an extraordinarily wide range of covalent mutations (e.g. T^T) that destabilize the DNA duplex. It is inherently coupled to transcription, repairing immediately needed genes, but it can also act globally.

A. XPC-RAD23B binds the site of the lesion (globally).

B. TFIIH (the general transcription factor recruiting RNA PolII) is recruited and, with its multiple subunits, unwinds DNA.

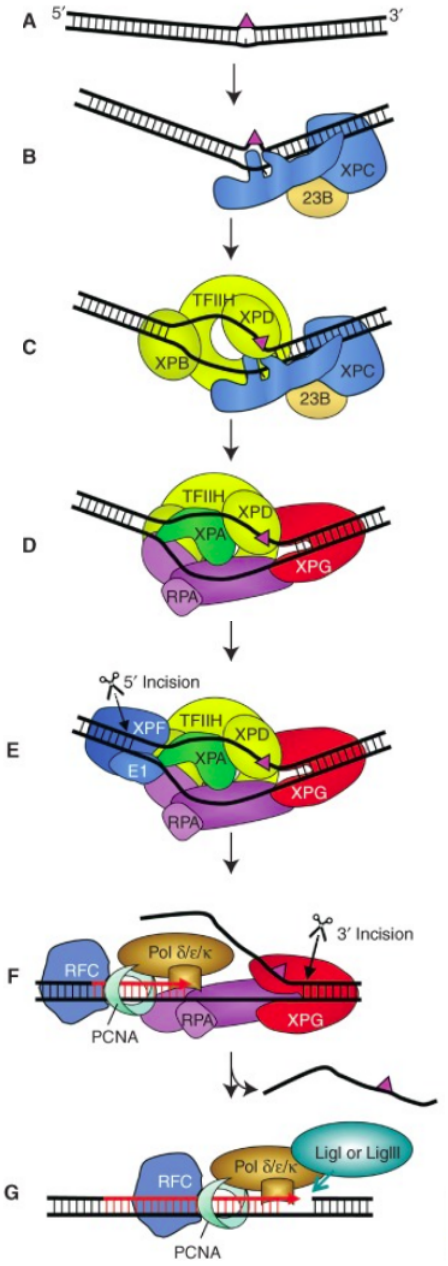
C. If lesion is confirmed, XPA and RPA are recruited that act as “precision pointers”. XPG endonuclease is also bound (inactive)

D. On the opposite site, the ERCC1-XPF endonuclease is recruited and, upon interaction with XPA, cuts the DNA.

E. The PCNA clamp brings Pol δ/κ/ε that displace TFIIH and starts polymerizing from XPF cut site, until XPG site is reached.

F. XPG is activated and it produces a second nick on the DNA, on the same strand as XPF.

G. The newly synthesized DNA is eventually sealed by ligase.



Schärer 2013 Cold Spring Harb Perspect Biol. 5(10):a012609

Homology directed repair (HDR) pathway

This mechanism is active in cases of double strand breaks (DSB) and is frequent after DNA replication (phase S or G2), when the two daughter DNA molecules lie close together and one can serve as a template for repair of the other.

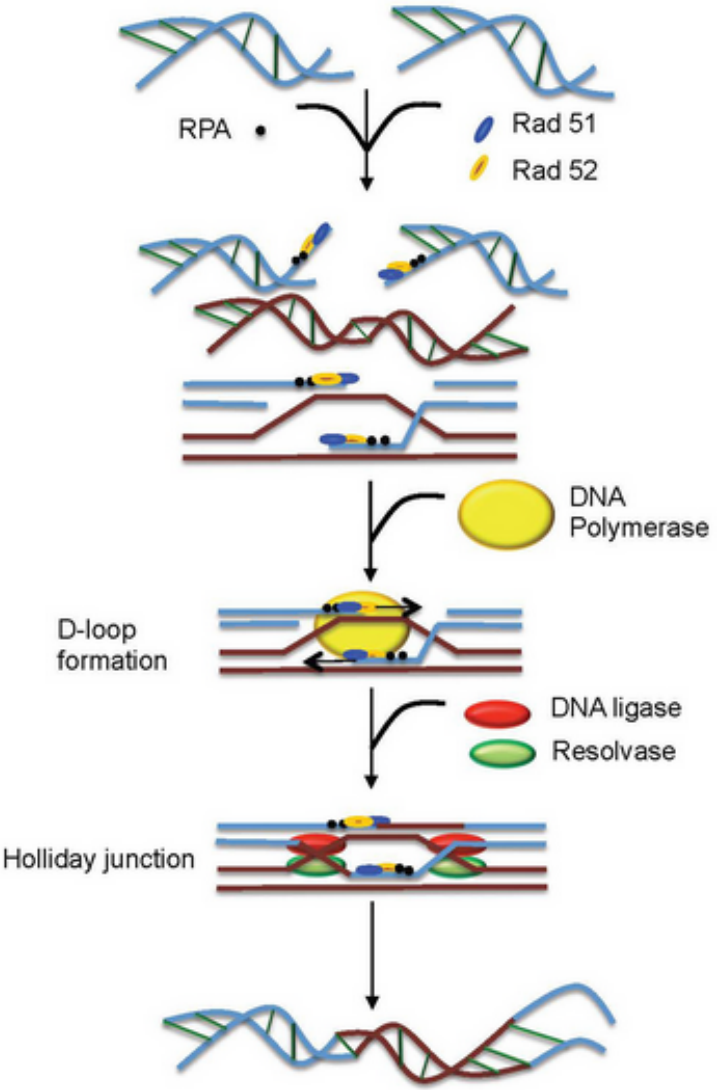
A. The ends of the broken DNA are covered with RPA (Replication Protein A) to protect from chromosome rupture.

B. The double helix is resected by specific nucleases to have an overhang extremity (SS)

C. Rad51 bind SS and drives a strand invasion by exposing 3 nucleotides: an ATP-dependent 3-by-3 nt scanning on the the sister chromosome.

D. Once stable base-pairing is established, an accurate DNA polymerase extends the invading strand by using the information provided by the undamaged template molecule

E. Strands are displaced and further repair synthesis and ligation restore the two original DNA double helices



Non-homologous end joining (NHEJ) pathway

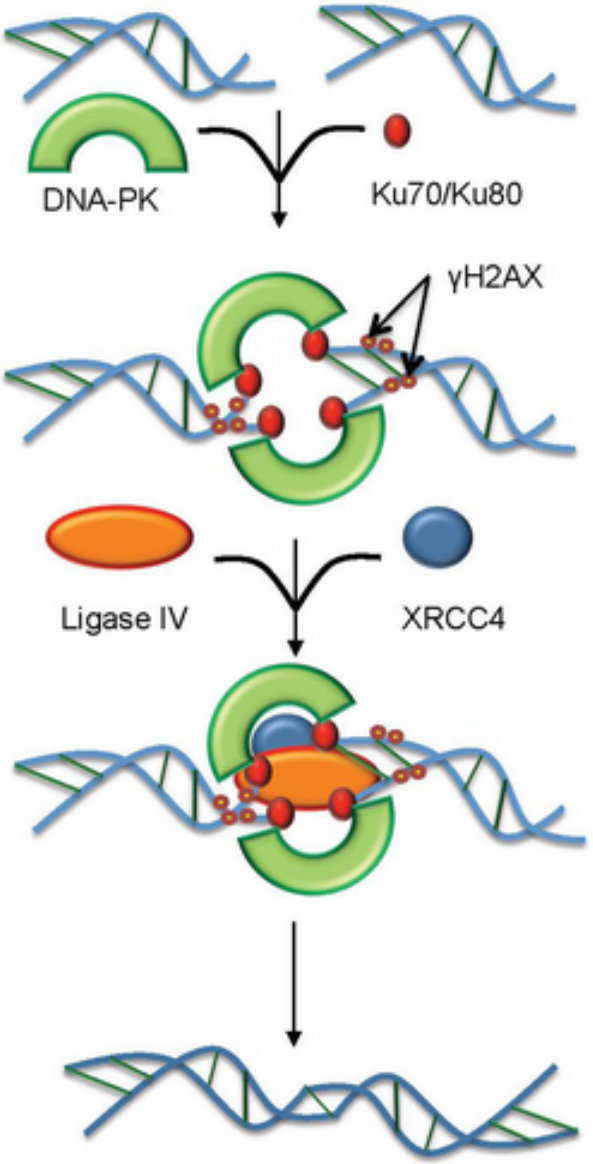
It happens before duplication and is a risky repair because ends are forced to associate regardless their compatibility.

A. the Ku hetero-dimer quickly localize to DSBs due to its extraordinary affinity (2 nM) for DNA free ends. Ku forms a ring that slide through DNA, contacting the sugar backbone, not the bases.

B. Ku scaffolds the big NHEJ complex composed by
- DNA-Pkcs (PI3K family as ATM),
- XRCC4, a tetrameric scaffold protein
- XLF, that form a filament to bridge DNA ends

C. depending on the lesion, several other enzymes are recruited and activated:
- PNKP, polynucleotide kinase-phosphatase
- Polμ (polymerase) and RecQ (recombinase)
- APLF, an exo/endo nuclease.
- DNA Ligase IV (the last player)

This complex can mediate end-end repair because it contains everything needed to resect ends, fill gaps, remove blocking end groups and make ends ligatable.



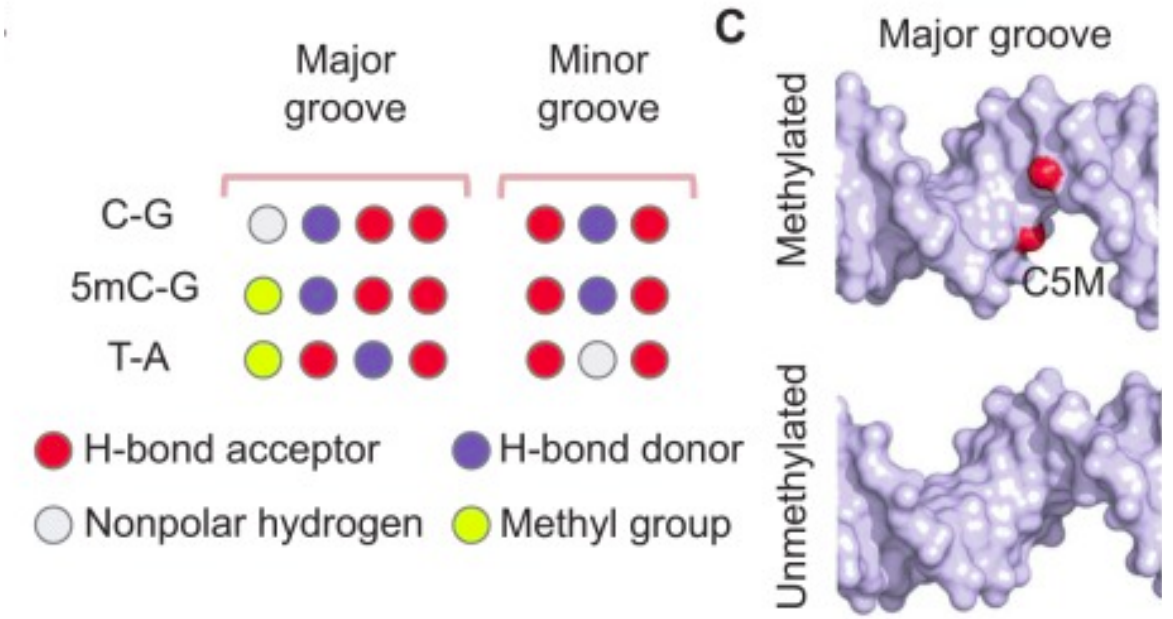
Davis and Chen 2013 Transl Cancer Res. 2(3):130-143

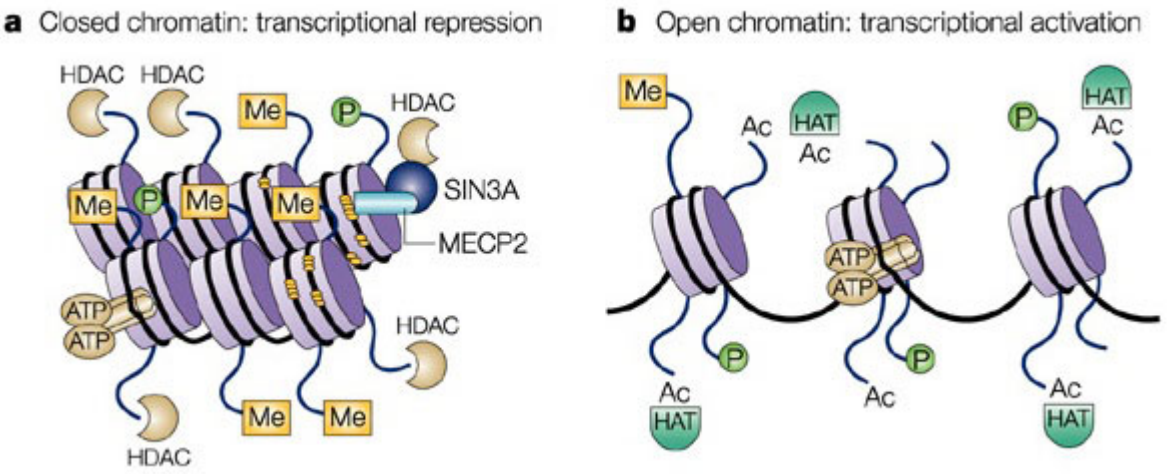
Effect of methylation on DNA accessibility

The methyl groups on methylated cytosines lie in the major groove of DNA and interfere directly with the binding of proteins (transcription regulators as well as the general transcription factors) required for transcription initiation.

In addition, cells contain a repertoire of proteins that bind specifically to methylated DNA, impairing DNA access to transcription factors.

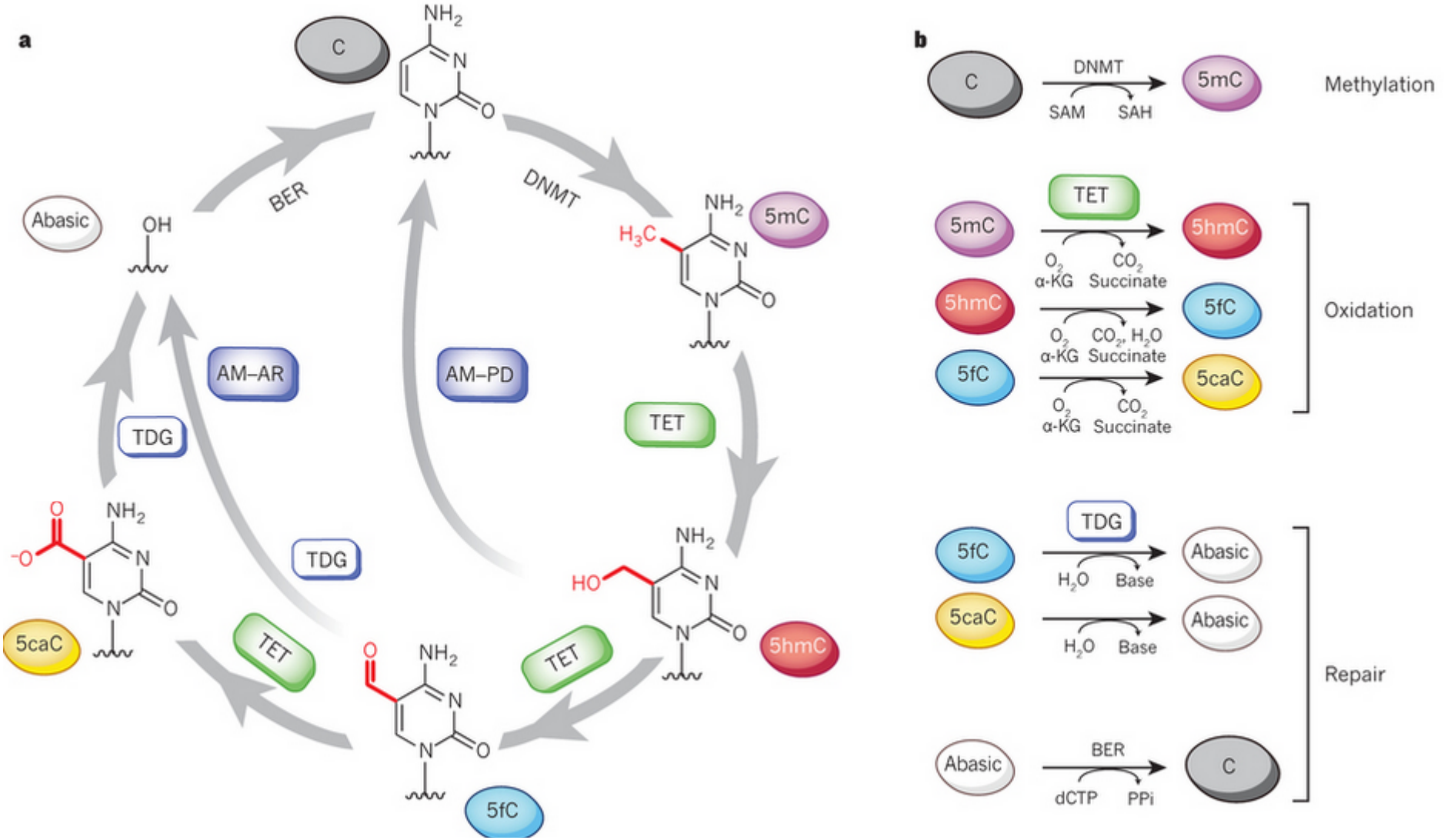
The best characterized of these proteins associate with histone modifying enzymes, leading to a repressive chromatin state where chromatin structure and DNA methylation act synergistically.





Dantas Machado et al. 2015 Brief Funct Genomics 14(1):61-73

Control of DNA methylation status



Kohly and Zhang 2013 Nature 502:472–479

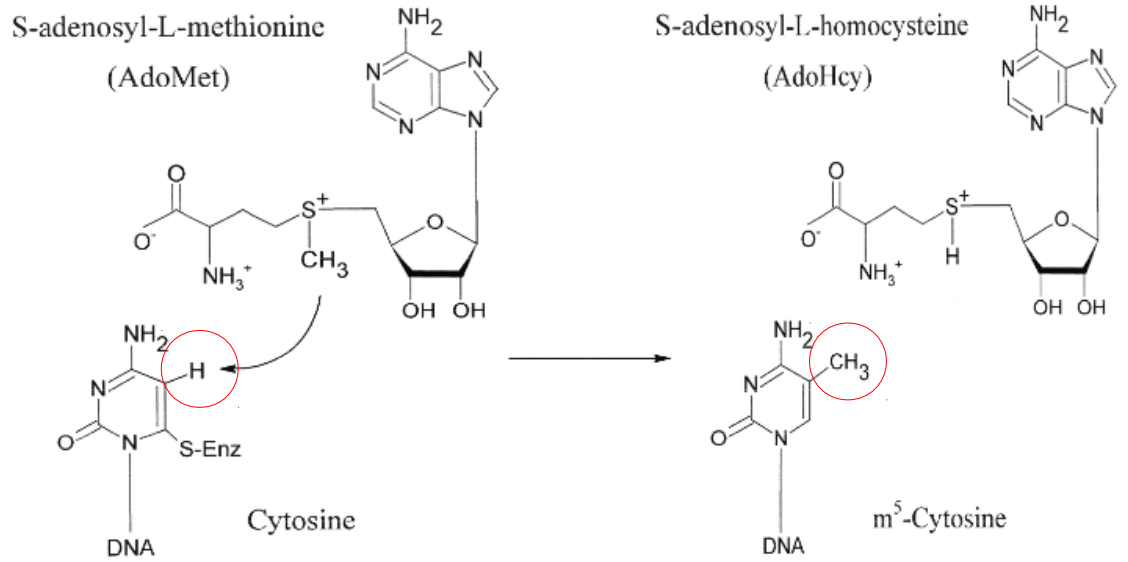
DNA methylation

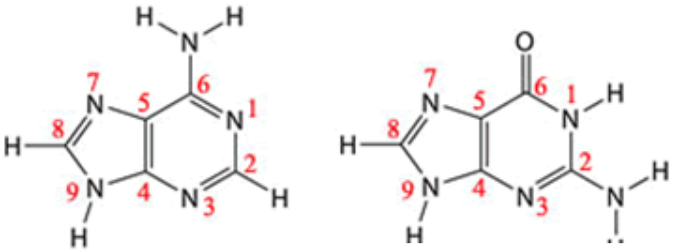
The methyl group can be attached at the N6 position of adenine, or at the N4 or C5 positions of the cytosine residues.

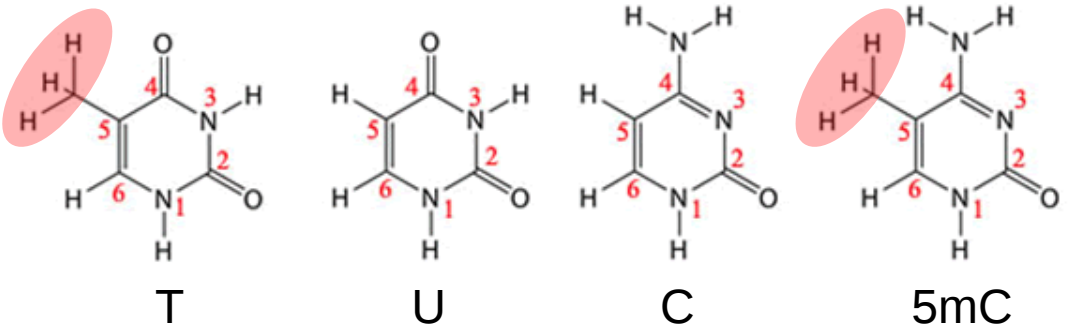
In mammals we only observe methylation at position 5 of the cytosine (C) residue within the CG dinucleotides (**CpG**). The modified base (m5C) is designated as the fifth base of DNA.

The methylation of C leads to the formation of a structure that is similar to T, that in turn is formed by methylation of U.

This observation has important evolutionary consequences: **does DNA evolved from RNA?**







DNA methyl transferases (DNMT)

DNMT enzymes are the writers of CH3-based DNA epigenetic mechanisms: they are active in association with the DNA replication machinery

1. **DNMT1**
2. **DNMT2**

maintenance

3. **DNMT3A**
4. **DNMT3B**

de novo (embryo)

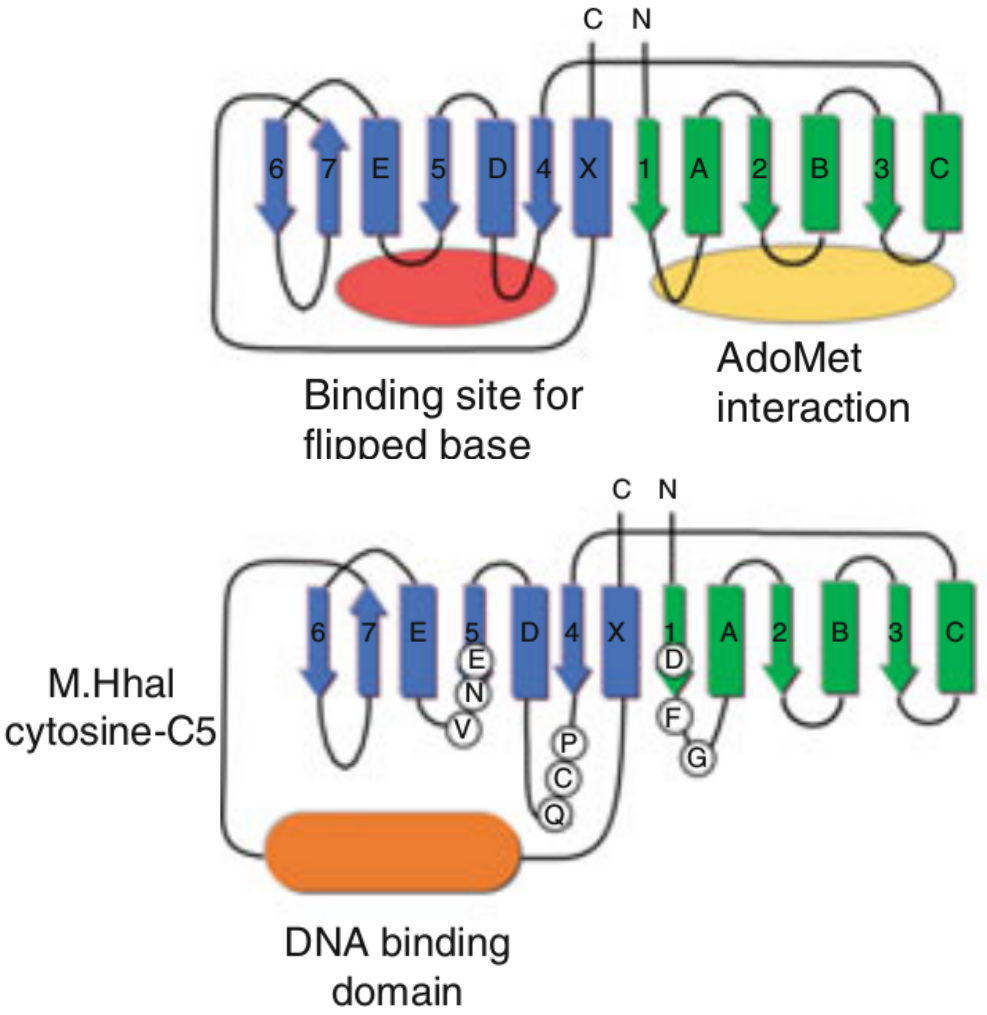
5. **DNMT3L**

helps maintenance

The catalytic mechanism of DNMTs requires that the C base is flipped out from the DNA structure, so a conformation that is similar to that of AdoMet.

The evolution of DNMTs seems related to a duplication of the Rossman fold.

Paired Rossman fold

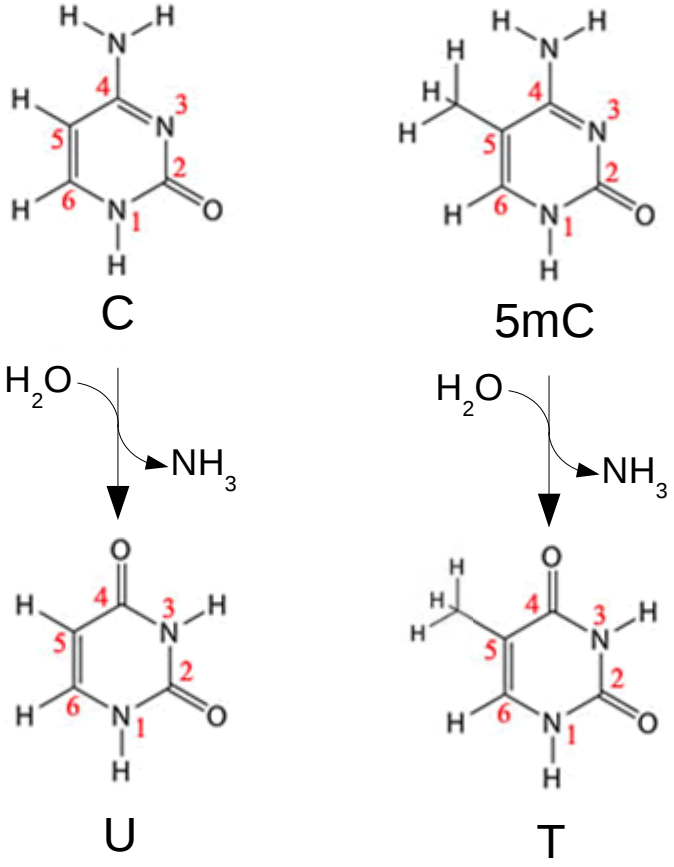


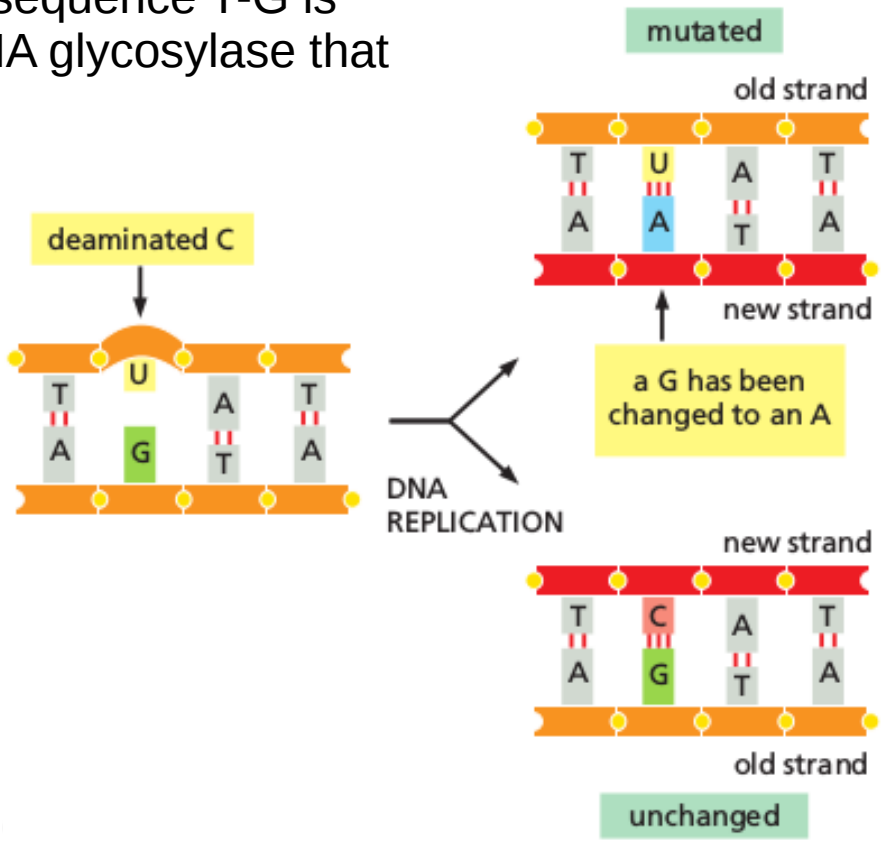
Consequences of C or 5mC spontaneous deamination

Deamination of cytosine or 5-methyl cytosine give rise to uracyl or thymine: in both cases the pairing with G is compromised, causing CG to TA transition mutations if not repaired.

- Mismatched uracyl is corrected by uracil DNA glycosylases.

- Mismatched base pairs involving T in the sequence T-G is corrected by a special but **inefficient** DNA glycosylase that removes the T.





Note: the DNA molecule is stable by design. The presence of T rather than U in DNA allows the repair system to distinguish a deaminated C from a naturally occurring U.

Inheritance of DNA methylation patterns to daughter cells

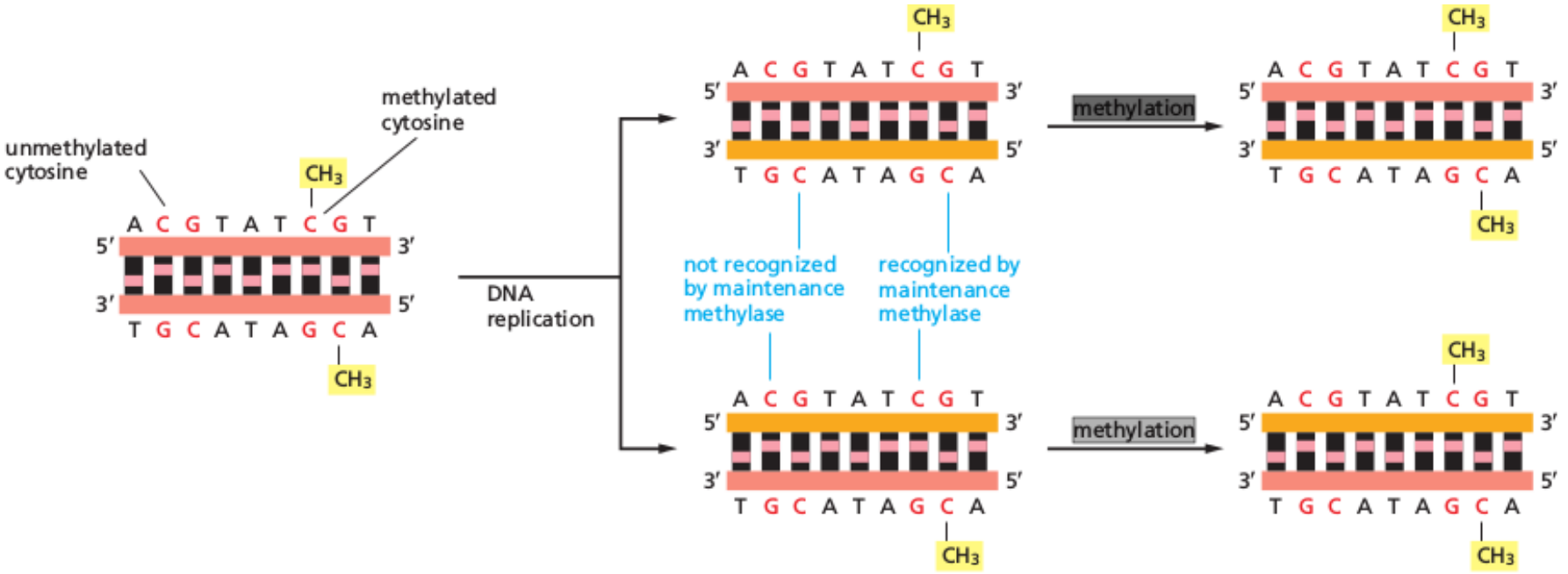
Methylation can be inherited in somatic cells due to the semi-conservative nature of DNA replication.

The replication of DNA creates a methylated filament (from the original cell) and an unmethylated filament (newly synthesized).

The so-called *maintenance methyl transferase* (DNMT1) selectively

- binds methylated C in CG dinucleotides
- methylates the C of the GC in the reverse complement strand.

This enzyme allows patterns of methylation to be inherited by daughter cells.



Methylated patterns are maintained by a paired system

Methylation **inheritance** is possible because de novo DNA methylases are recruited (an in turn recruit), histone reader and writer proteins.

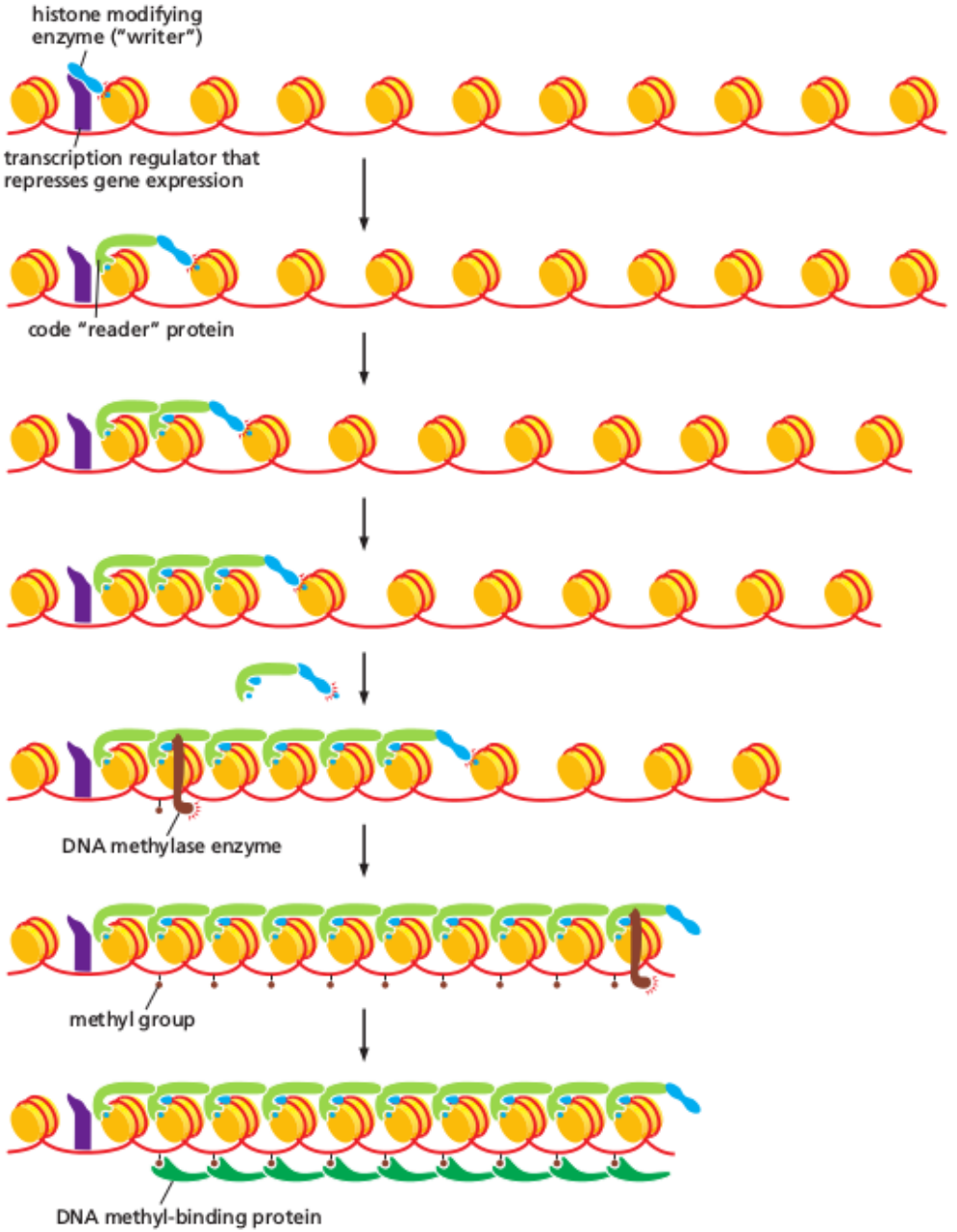
- A de novo DNA methylase is attracted by the histone reader

- The DNMT methylates nearby cytosines in DNA

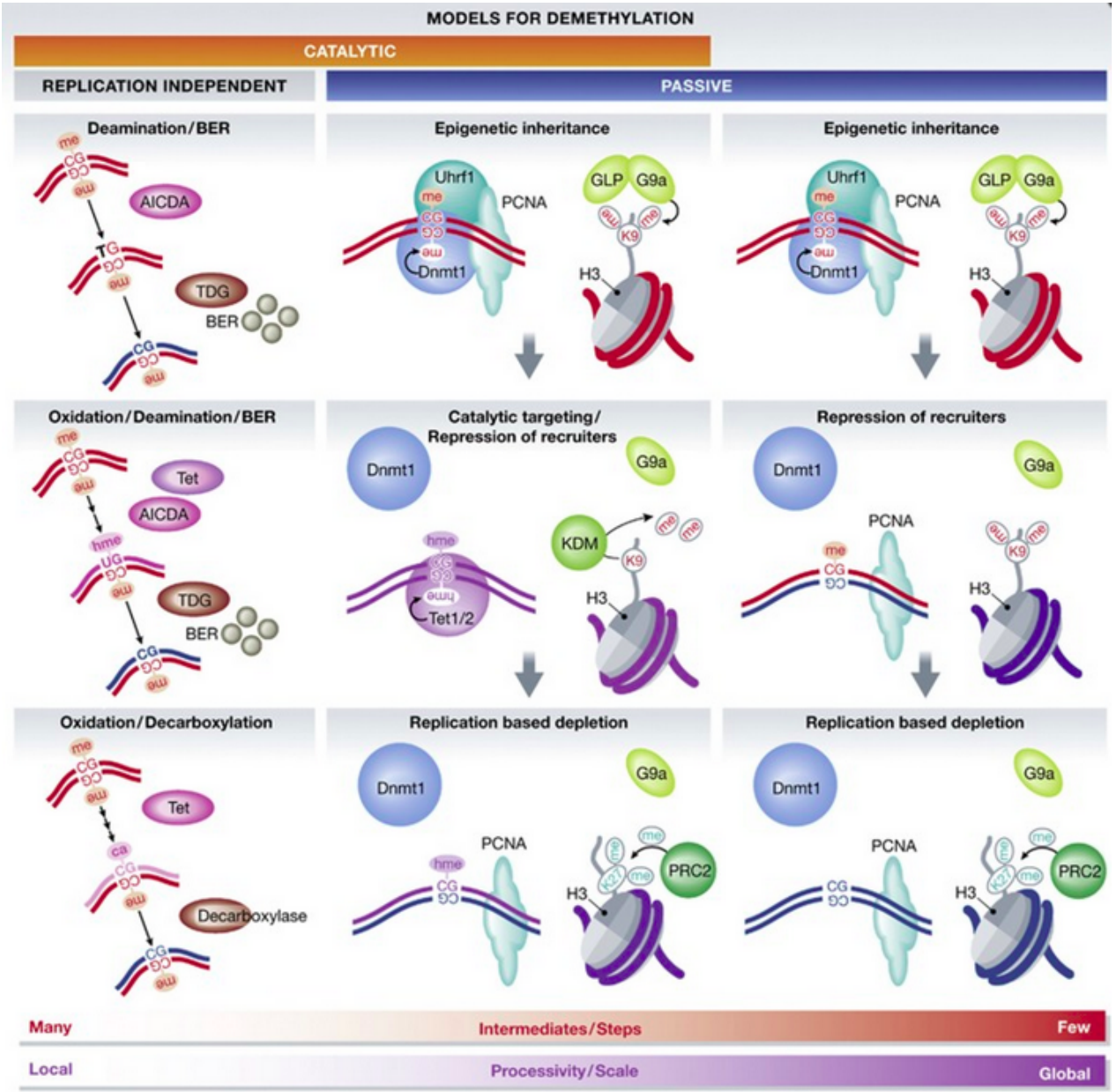
- C^{me} is bound by DNA methyl-binding proteins.

- During DNA replication, some of the modified (blue dot) histones will be inherited by one daughter chromosome

- The modification induces reconstruction of the same pattern of chromatin modifications as in the mother cell.



DNA demethylation: restoring citosine

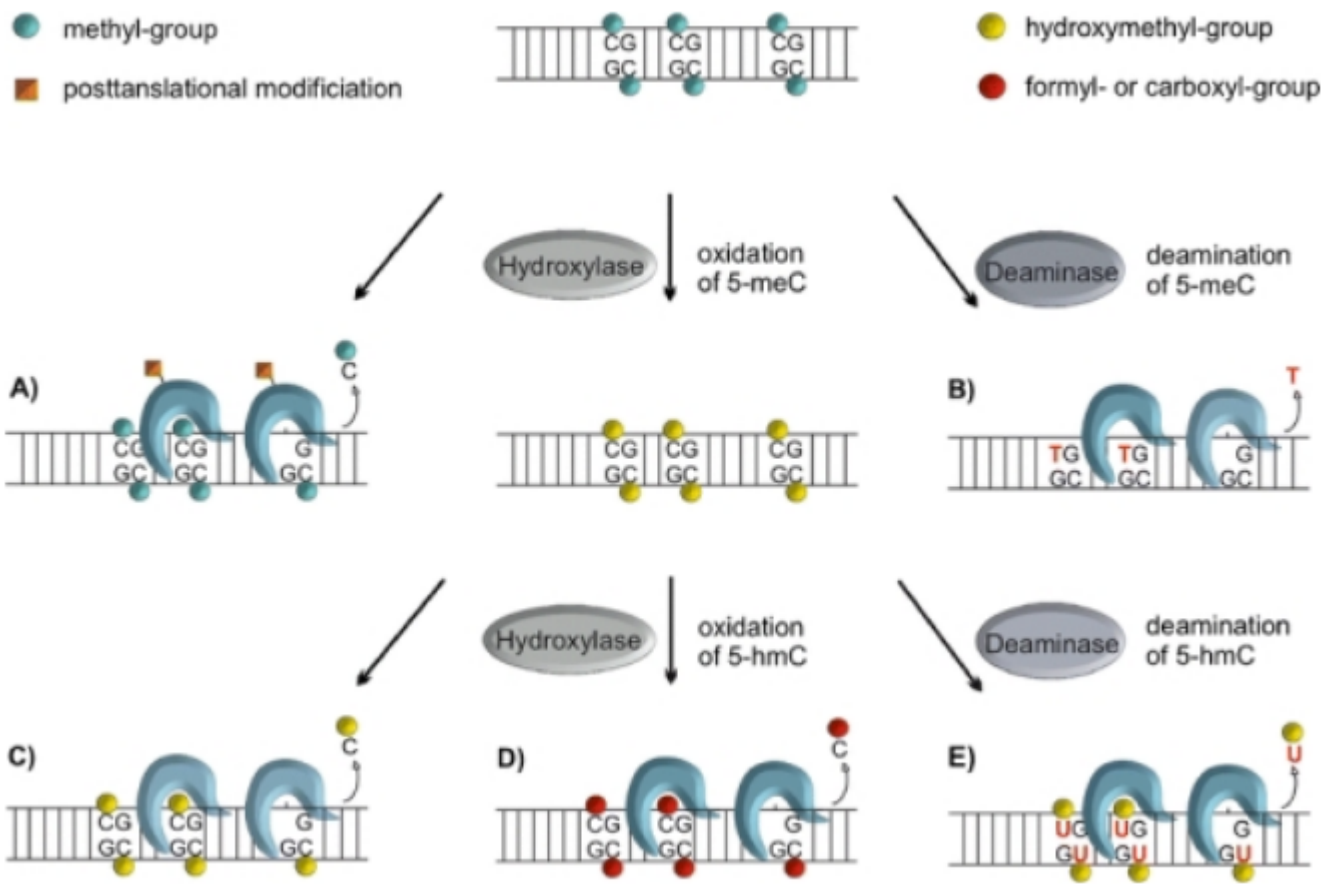


Smith and Meissner 2013 EMBO J. 32(3): 318-321

DNA demethylation: restoring citosine

The conversion of 5-meC to C in DNA can occur in different ways:

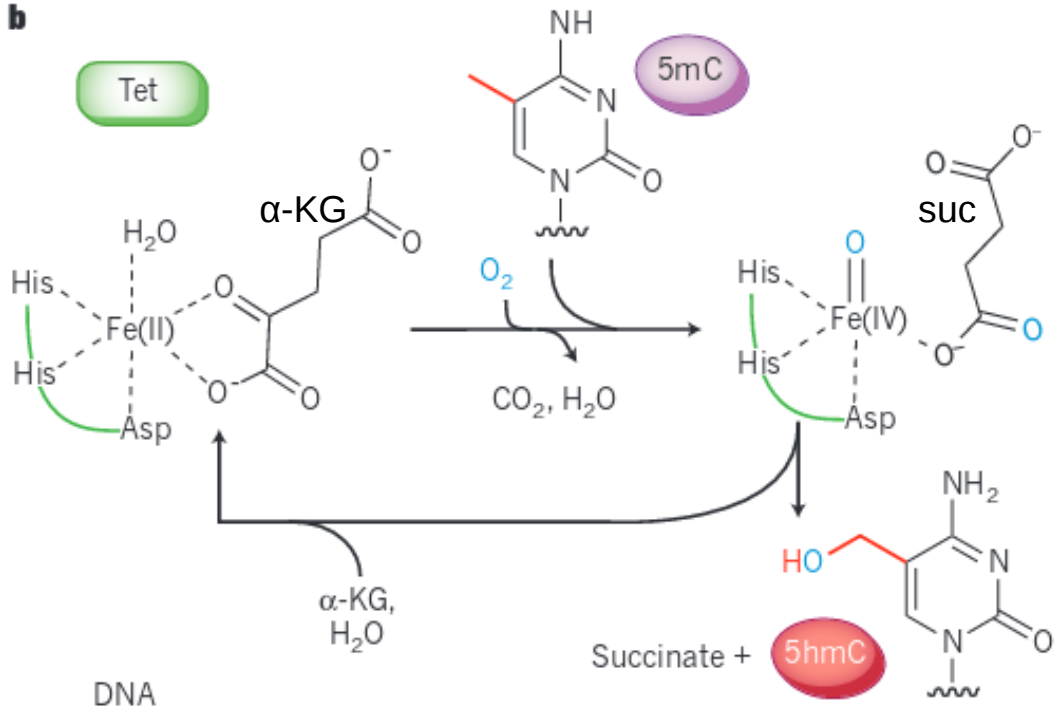
1. direct removal of the methyl group (?)
2. replacement of 5-meC with C by nucleotide excision repair (NER)
3. deamination/oxidation of 5-meC followed by base excision repair (BER).

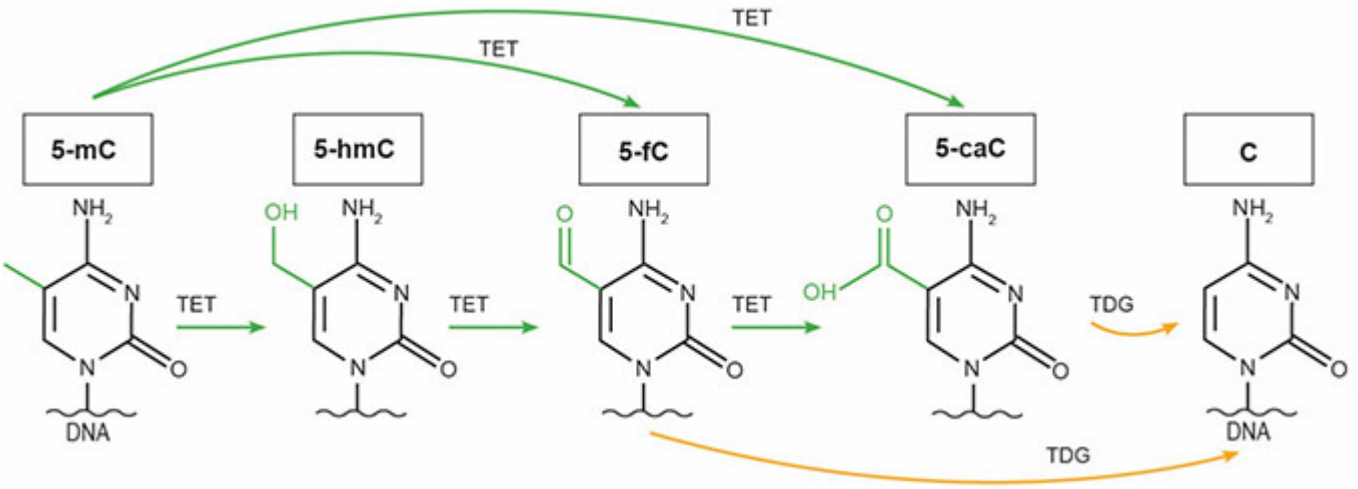


DNA demethylases (DMT): TET Methylcytosine Dioxygenase

TET (1,2,3) has a double-stranded β-helix fold that contains the crucial metal-binding residues found in the family of **Fe(II)** / **α-KG**-dependent oxygenases.

All three TET enzymes have a CD (Cys-rich and DSBH regions) domain that uses O_2 to decarboxylate α-ketoglutaric acid generating a high-valet iron oxide that converts 5mC to 5hmC (as well as 5hmC and 5caC). The reaction release succinate.

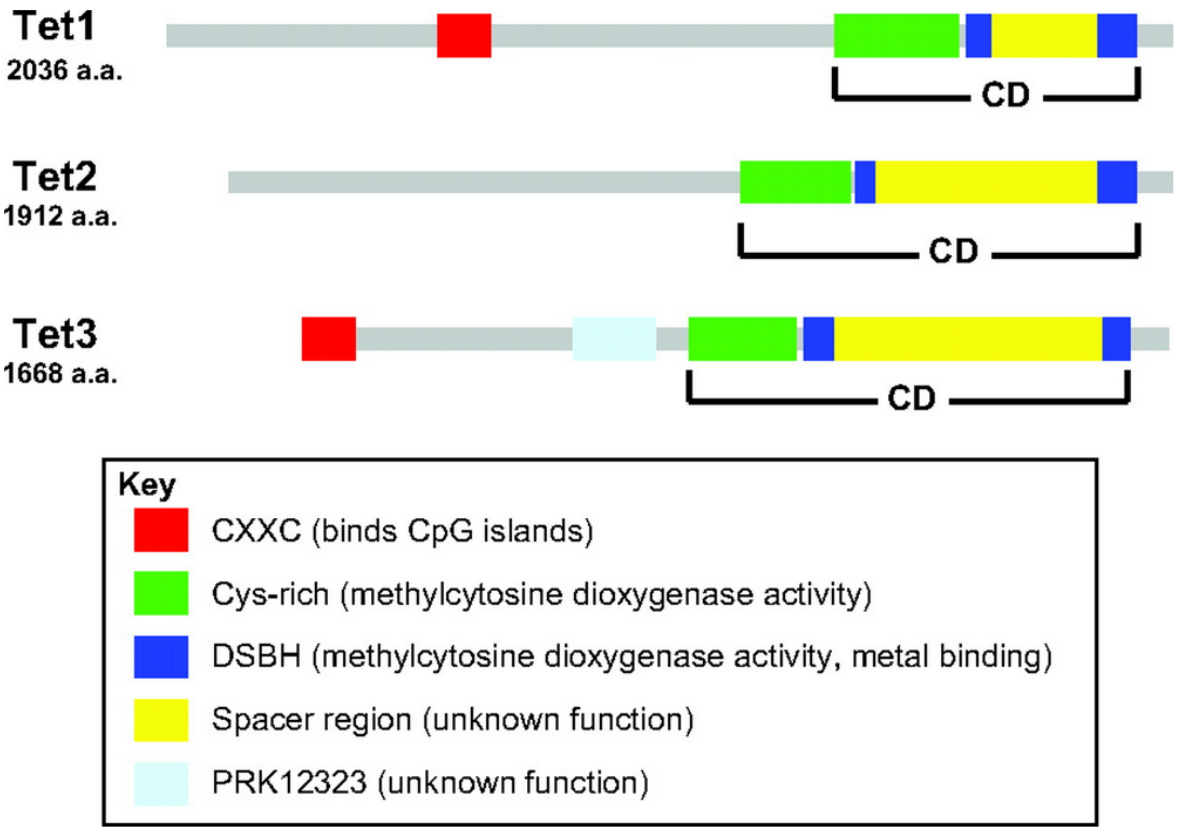




DNA demethylases (DMT): the TET family

Apart from the catalytic domain (CD), TET1 and TET3, but not TET2, have a CXXC domain. CXXC domains typically bind unmethylated CpG dinucleotides, but in TET enzymes they bind 5mC and 5hmC. The exact role of the CXXC domain is not fully understood, but it is likely involved in targeting the enzyme to specific regions of the genome.

TET1 and TET2 are found in embryonic stem cells, while TET3 is found in germ line cells.



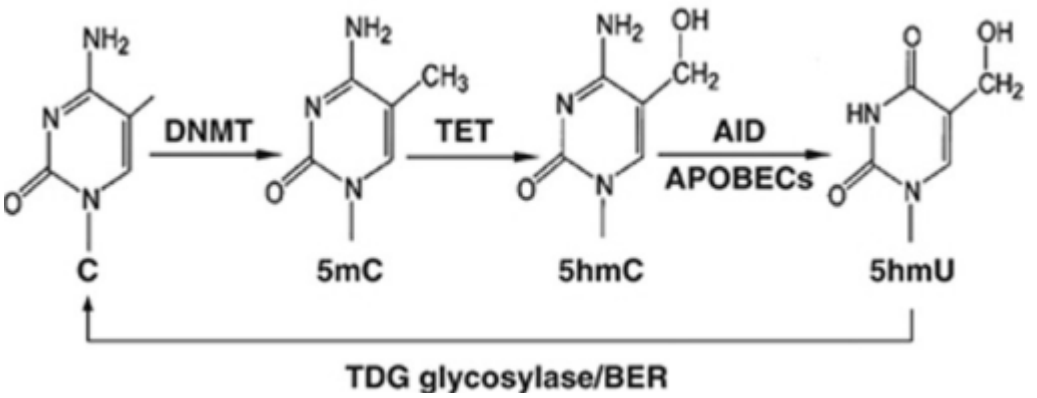
DNA deaminase: AID/APOBEC

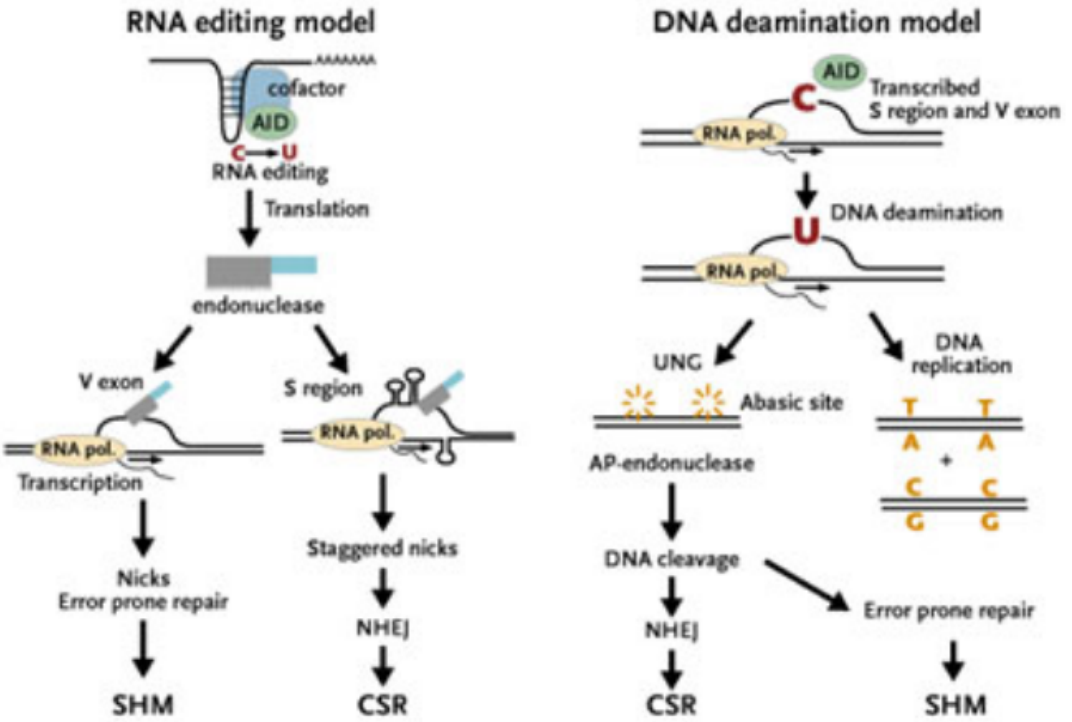
Among the AID/APOBEC family, AAI and APOBEC1 are expressed in mammalian **oocytes** and **embryos** at stages when global DNA demethylation occurs.

They can deaminate 5meC to thymine in vitro which, followed by G-T mismatch repair, could lead to DNA demethylation.

AID/APOBEC are mostly studied for their role in inducing C→U mutations and affecting somatic rearrangement of immunoglobulin genes during the immune system diversity explosion, a early stage of immune-competence development.

In this context, it is unclear if the mutation is induced in DNA or RNA.



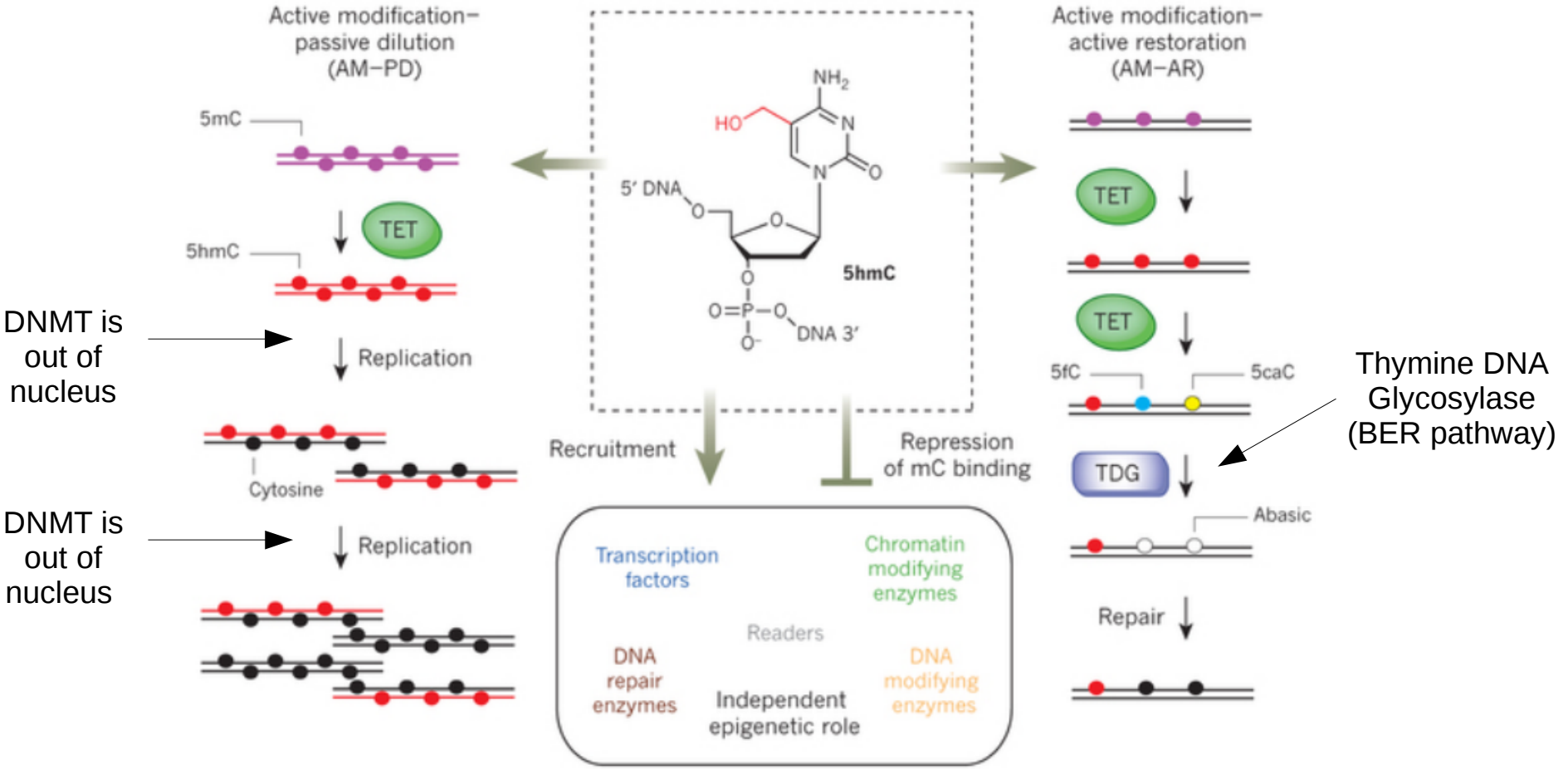


DNA demethylation: active or passive modes

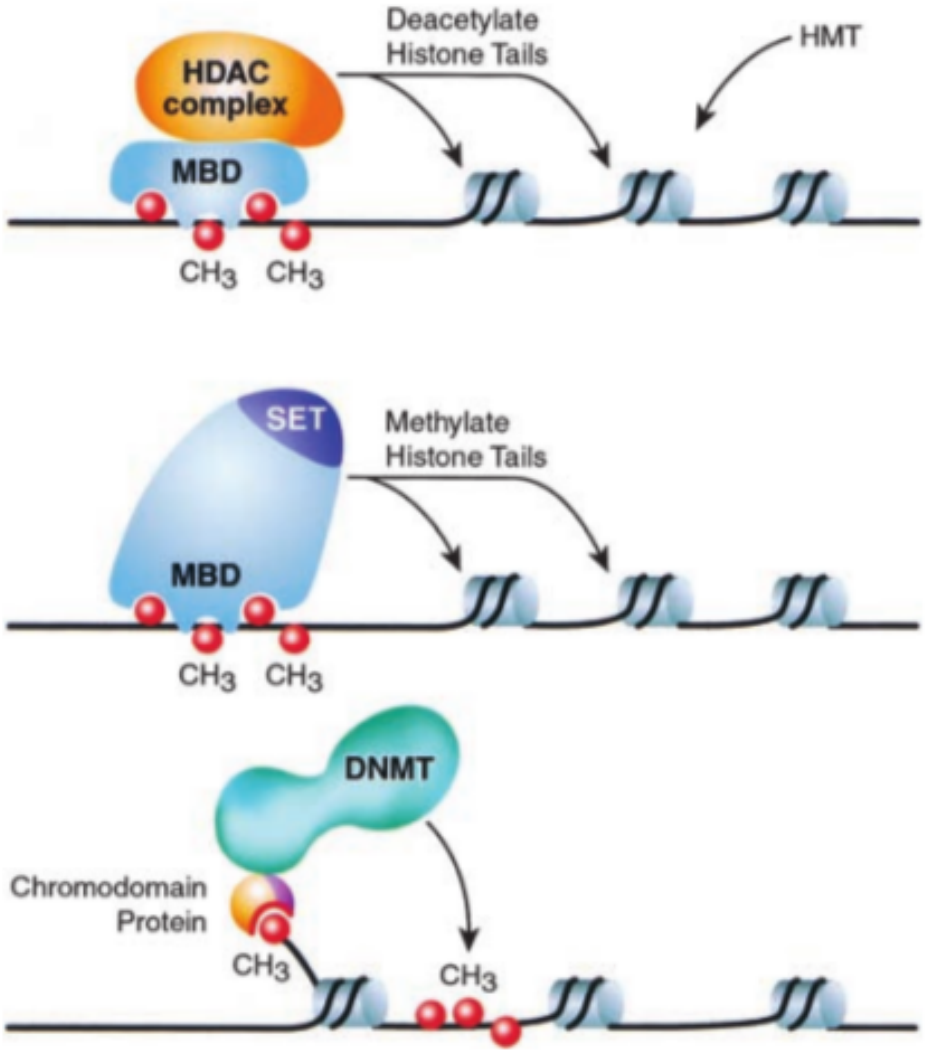
Removal of methyl groups from 5mC can be achieved in two different ways, all starting from oxidation of 5mC by enzymes of the TET family:

1. AM-PD: active modification (AM) followed by passive dilution (PD, duplications)

2. AM-AR: active modification (AM) followed by active restoration (AR, by BER)



Relationship between DNA methylation and histone modifications



Methyl-CpG-binding proteins (MBD domain) recruit HDAC complex to deacetylate histone so that the histone tails will be suitable for subsequent methylation by HMTs.

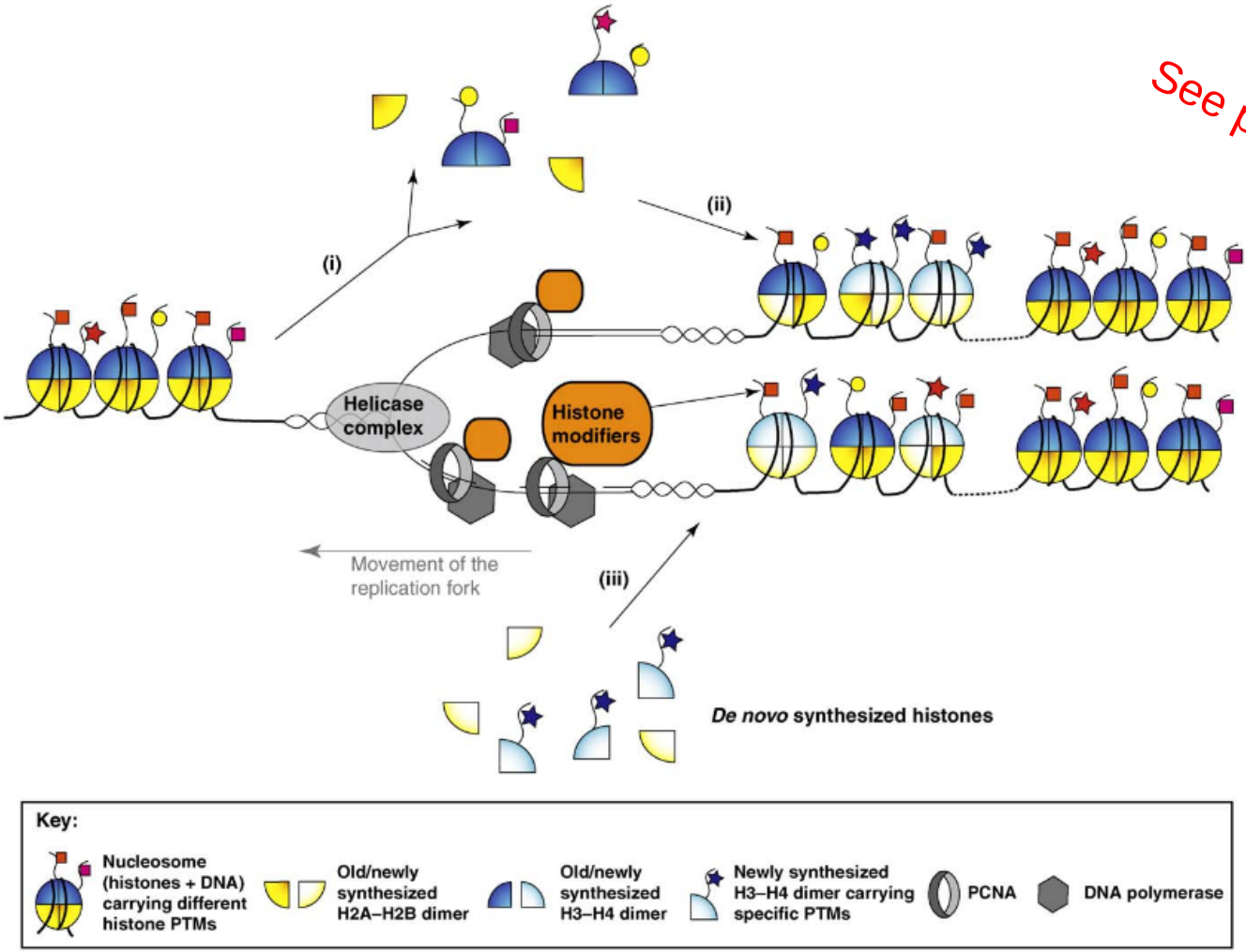
In chromatin domains where histones are hypoacetylated, the MBD domain-containing HMTs may bind directly and methylate the histones.

Methylated histone tails may recruit DNMTs to methylate DNA for long-term gene silencing.

Methylation memory after DNA replication

During DNA replication (modified) histones disassemble to favor accessibility, and during the S phase a mass of new histones is produced. The methyation status of the DNA has influenced histone PTMs. New and old histones mix and assemble just after the replication fork, and this proximity ensure a track record of the previous DNA Methylation status.

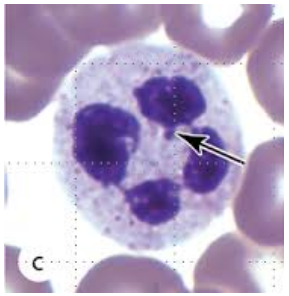
See part 1 slide 50



Barth Imhof 2010 Trends Bioch Sci. 35(11): 618-626

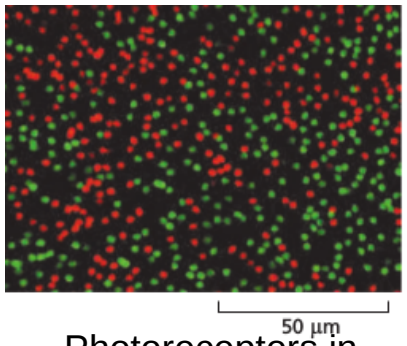
Inactivation of X chromosome in females

The most striking example of inherited (mitotic, cell-to-cell) chromatin structure is the X-inactivation in mammals, where one (random) X chromosome of a female embryo gets inactivated by methylation to avoid dosage problems of XX with respect to XY. The progeny of that embryonic cell will maintain the X-inactivation.

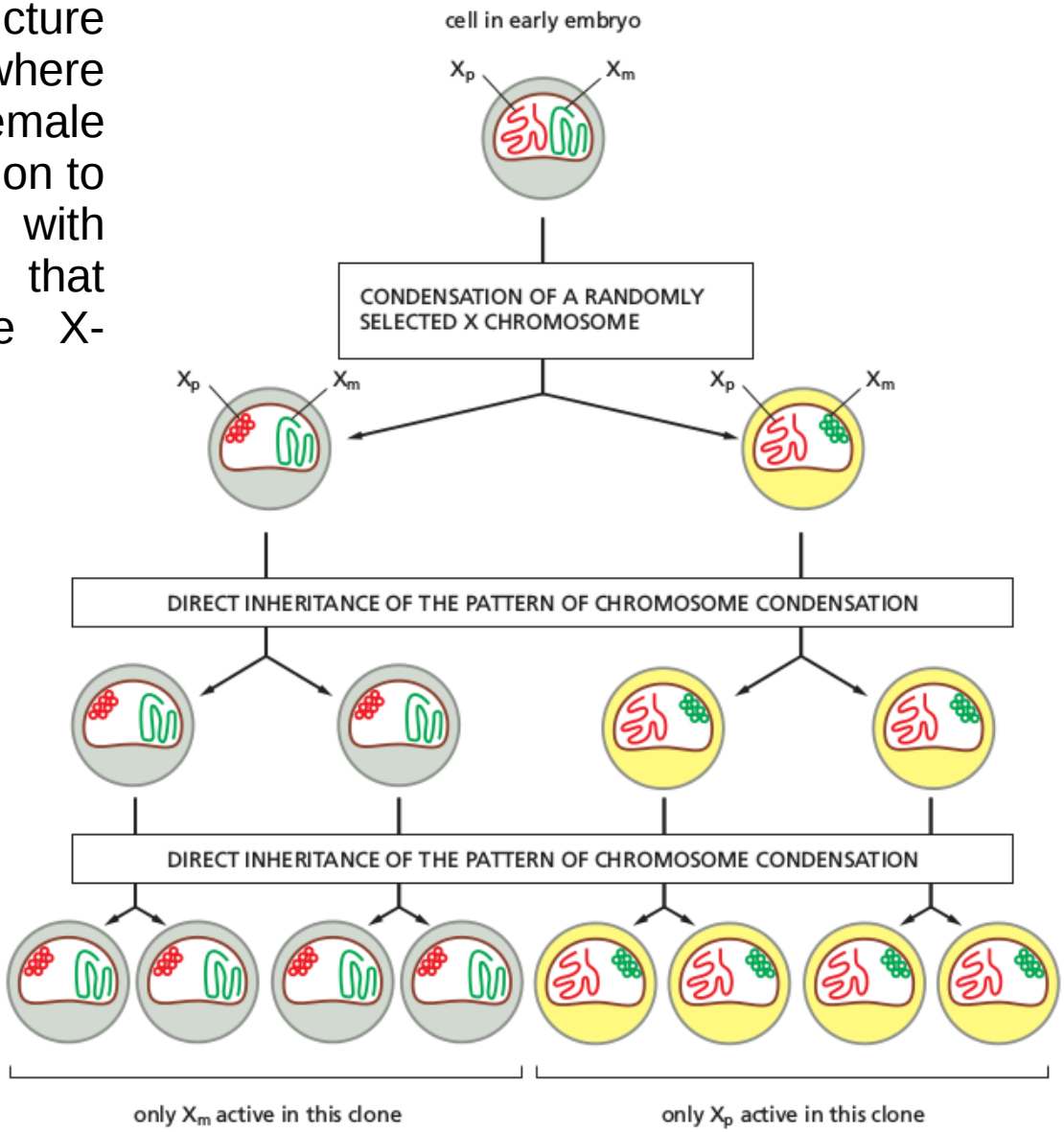


Drum Stick (Barr body) in female neutrophils

Female Calico Cat

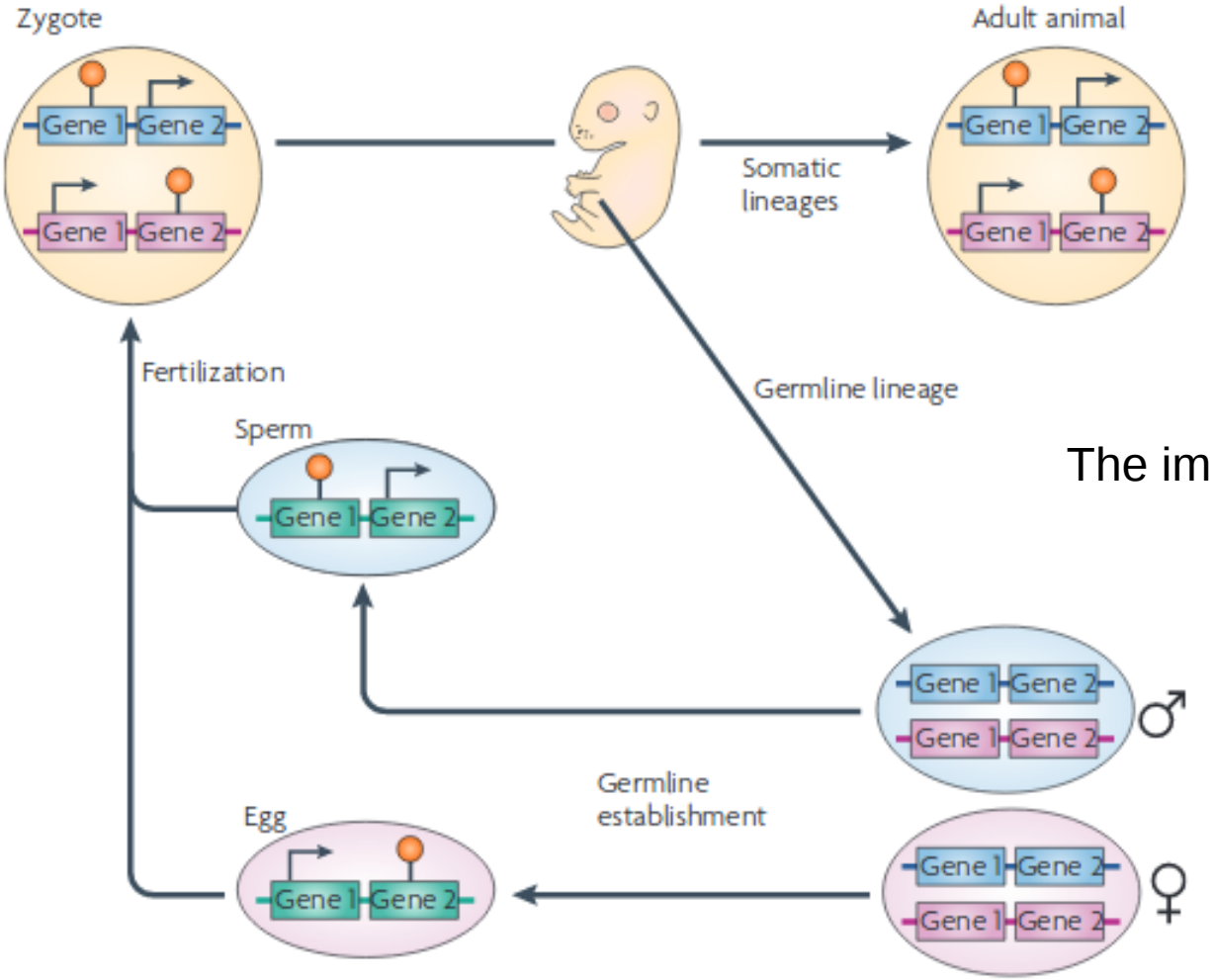


Photoreceptors in female mice



DNA methylation and imprinting

In a small fraction of mammalian genes (less than 1%) one of the two alleles that is inherited by the offspring is partially or completely switched off. The decision as to which one is silenced depends on which allele was inherited from the mother and which from the father. The existence of imprinted genes is an evolutionary enigma, as they effectively nullify the advantages of diploidy.

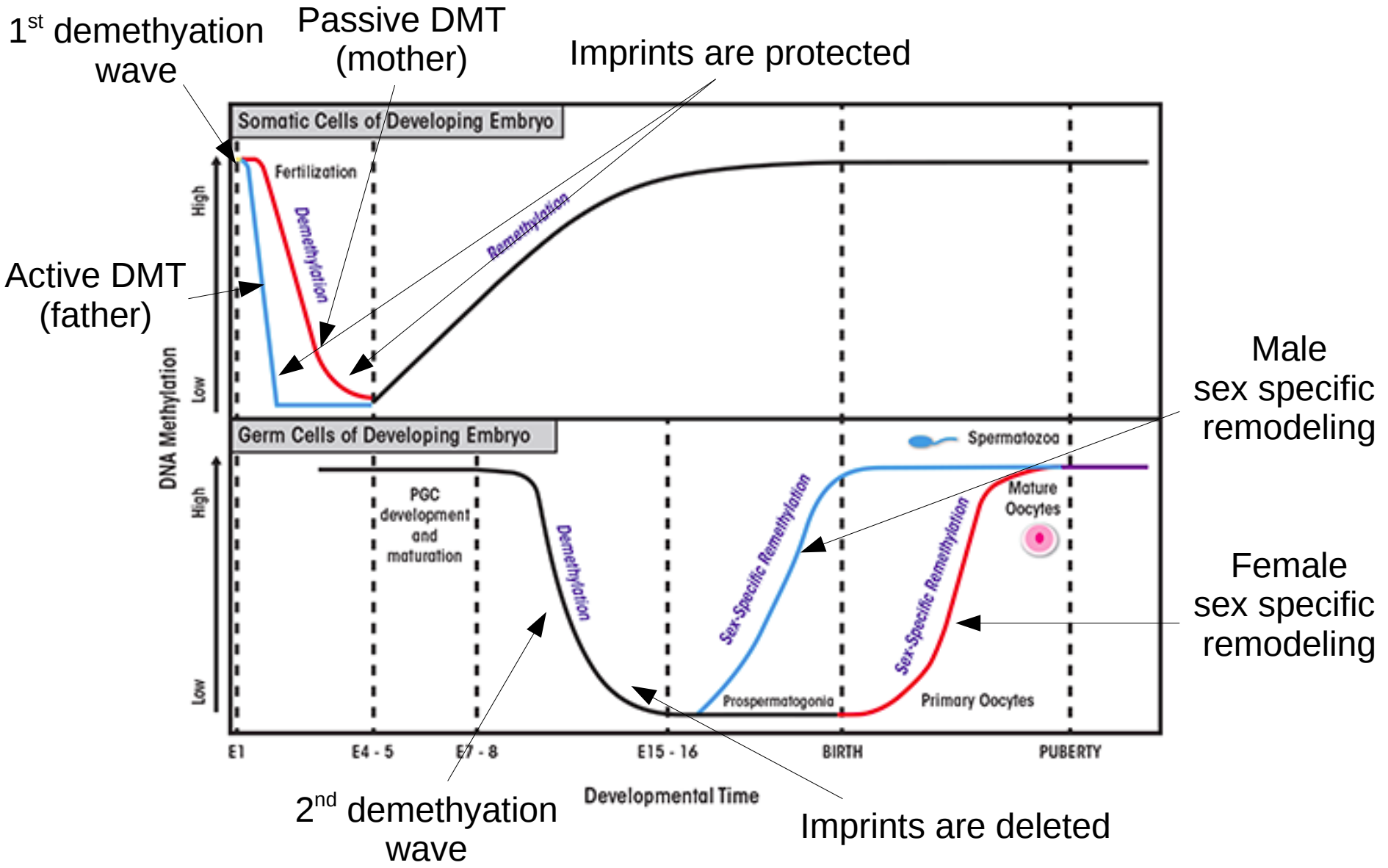


The imprinting cycle

Bartolomei and Ferguson-Smith 2011 Cold Spring Harb Perspect Biol. 3(7):a002592

Restore imprinting after demethylating wave

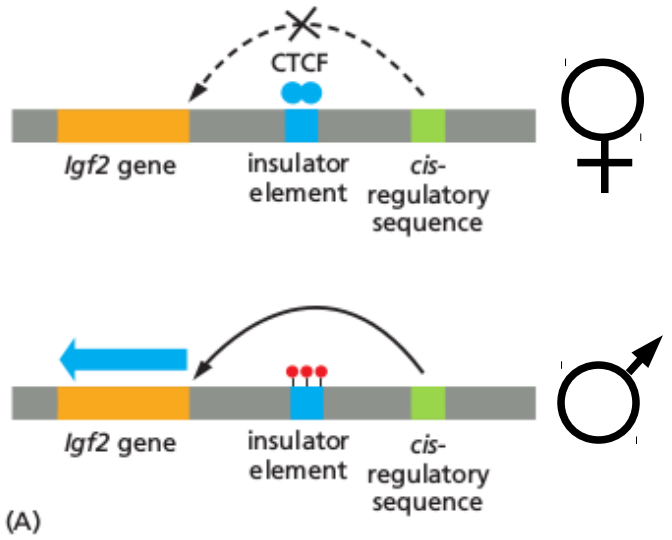
After the demethylating wave (during fertilization), *de novo*, sex specific imprints (mediated by DNMT3A/L and guided by histone acetylation) are placed on germ cells.



Examples of imprinting mechanisms

Mother: a protein called CTCF binds to an insulator sequence blocking communication between cis-regulatory sequences and the Igf2 gene, that is repressed.

Father: the insulator is methylated and CTCF cannot bind, so Igf2 is expressed.

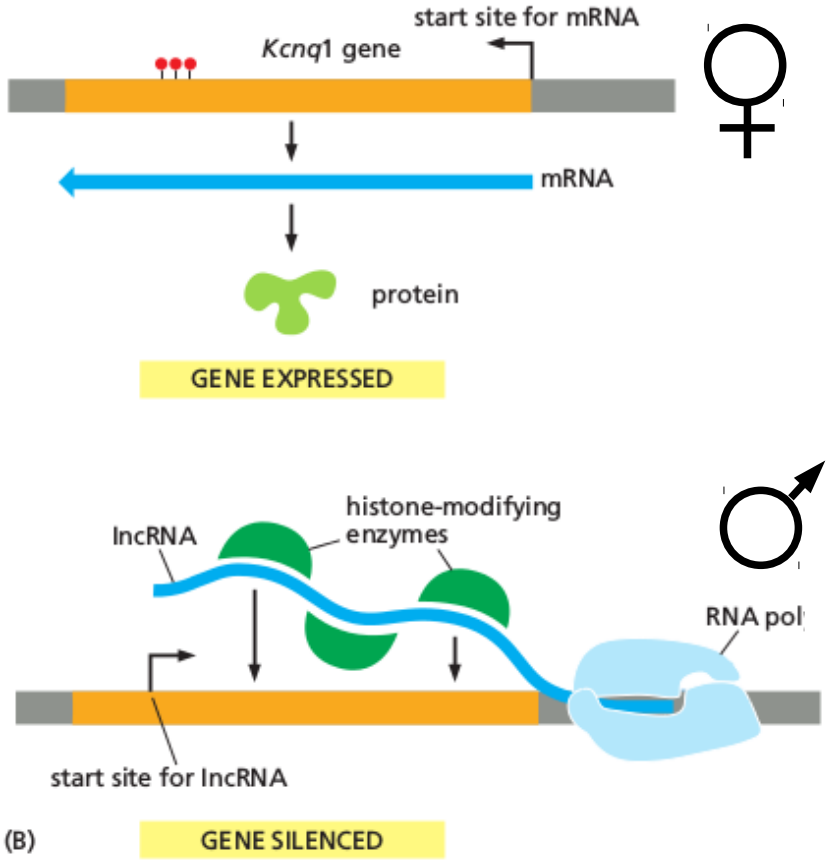


In simpler cases, the imprinting simply inhibits by methylation one of the two alleles.

Frequently maternal imprinting act on gene promoter regions, while paternal imprinting on the transcribed gene region.

Mother: the synthesis of a lncRNA is blocked by methylation. lncRNA cannot therefore interfere with the expression of Kcnq1 gene, that in turn is expressed.

Father: lncRNA is synthesized, remains in place and directs chromatin remodeling and repression of Kcnq1.

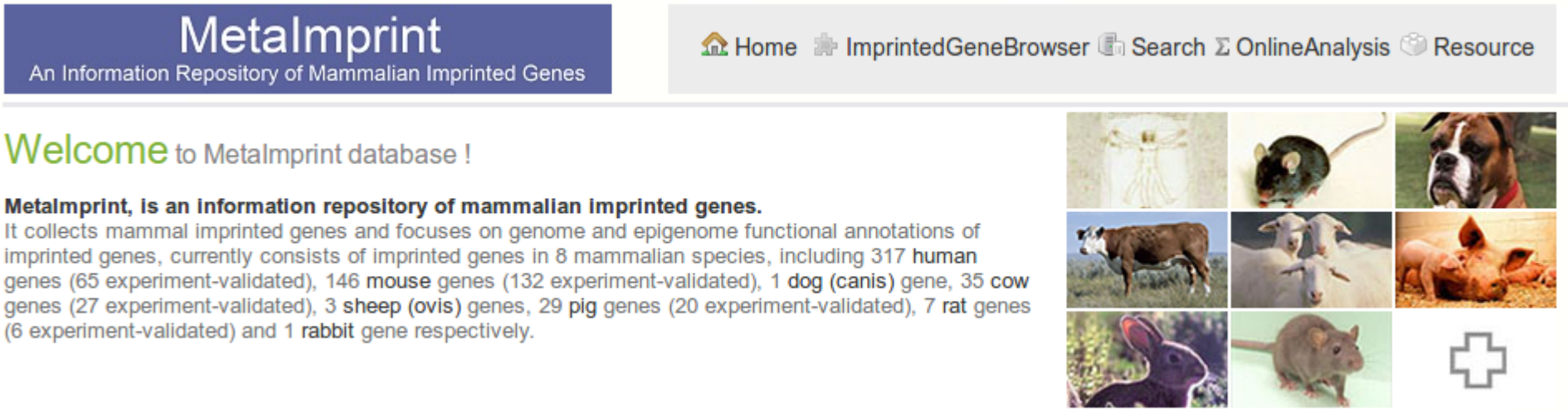


Imprinting control regions

Many imprinted genes reside in 1 Mb clusters throughout the genome. These clusters typically contain

- **at least one non-coding RNA**
- **both maternally and paternally expressed imprinted genes.**

Clusters are controlled by a discrete DNA element termed an **Imprinting Control Region** (ICR) that reside in a region that can be far from the controlled clusters.



bioinfo.hrbmu.edu.cn/MetaImprint/

Bartolomei and Ferguson-Smith 2011 Cold Spring Harb Perspect Biol. 3(7):a002592

Features of imprinted genes

Gametic imprints are placed on paternal derived genes during sperm production and on maternal derived genes during egg formation.

- Many of the imprinted genes identified to date are involved fetal and placental growth.

- Several imprinted genes instead have their role in specific tissues.

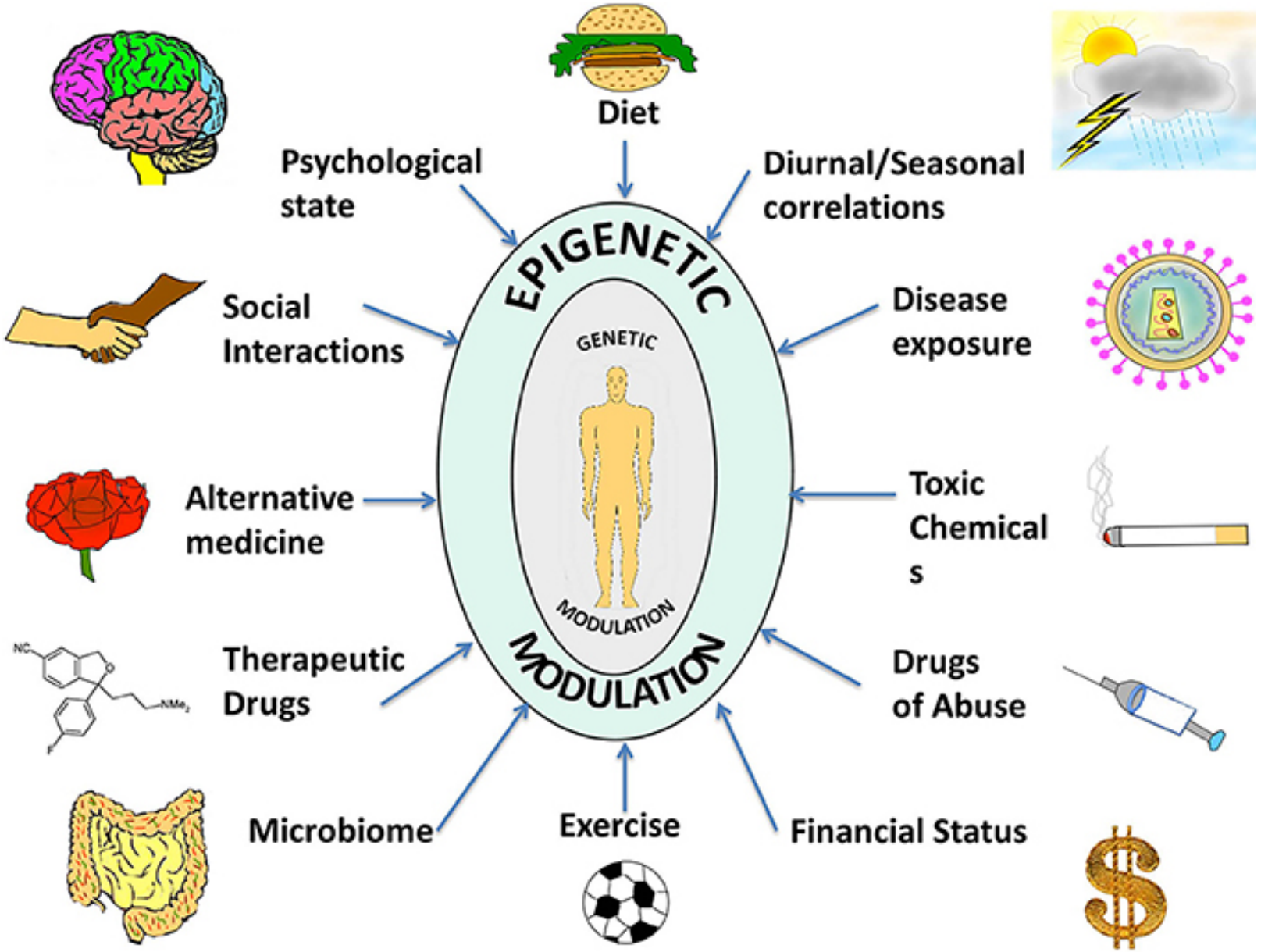
In general, imprinted genes play roles in key biological processes, such as cell proliferation, metabolic adaptation, as well as neurological processes and behavior.

Consequently, genomic imprinting must be correctly regulated during the entire life-span

SPECIES	NUMBER OF GENES
Human	305
Mouse	179
Pig	81
Cattle	28
Sheep	20
Birds	18
Horse	3
Dog	1
Cat	1
Rabbit	1
Primates	2

Bartolomei and Ferguson-Smith 2011 Cold Spring Harb Perspect Biol. 3(7):a002592

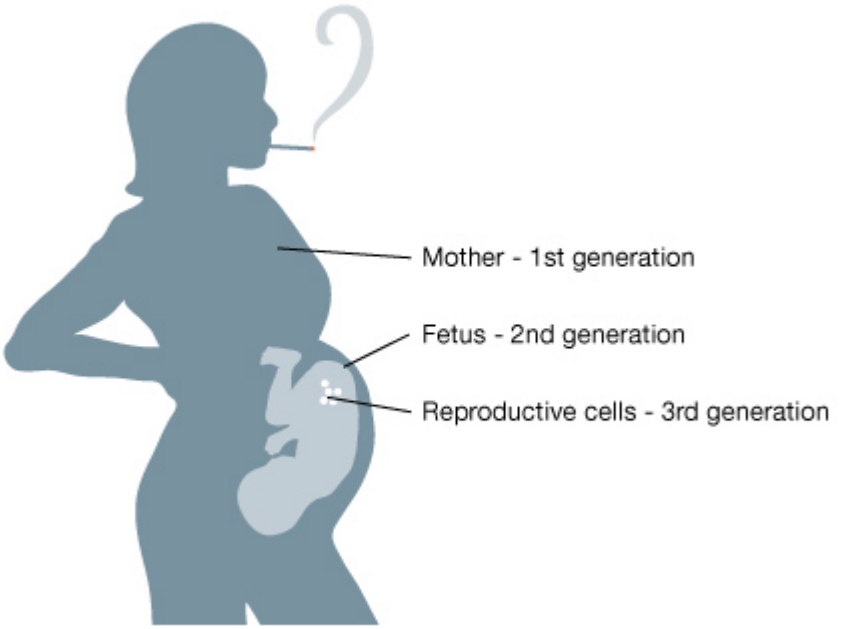
Epigenetics and environment



Epigenetics and inheritance

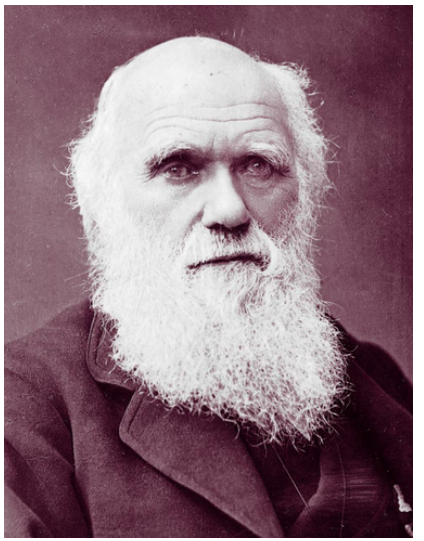
In a pregnant mother, three generations are directly exposed to the same environmental conditions at the same time.

Only an epigenetic trait that continues into the 4th generation could be said as inherited and not due to direct exposure.



It is very difficult to establish epigenetic inheritance

Transgenerational epigenetic inheritance has also the potential to be adaptive, and in some cases, might even respond to environmental challenges, with major implications for heredity, breeding and evolution.



Heard and Martienssen 2014 Cell. 157(1): 95-109
Penny 2015 Genome Biol Evol. 7(6):1758-1760

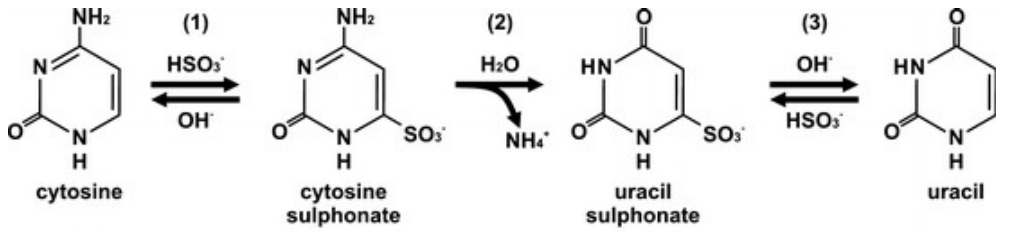
-omics for methylation: the bisulfite-seq

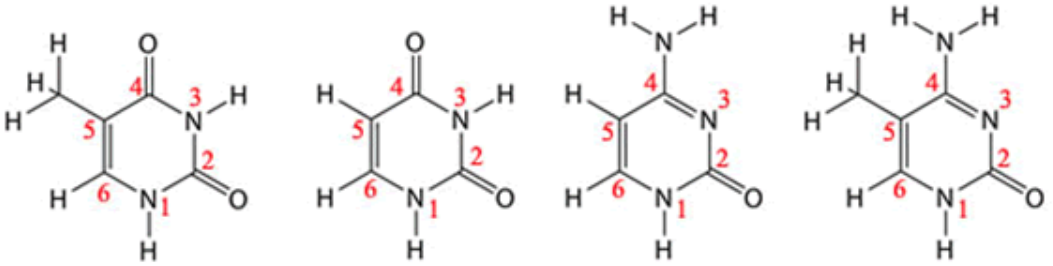
The basic idea of DNA bisulfite treatment is to force cytosine to deaminate (to uracyl).

5-methyl citosine is resistant to treatment.

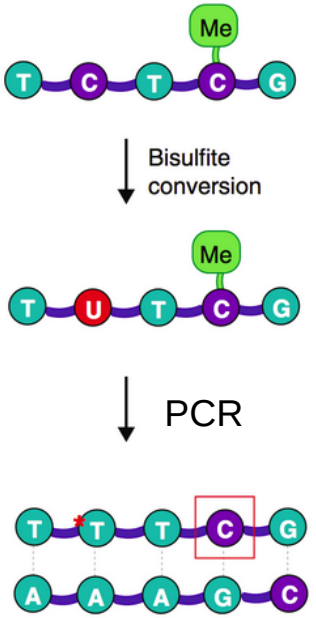
This time we don't have error repair correction, because DNA is simply amplified by Taq polymerase (type A pol) that cannot distinguish U and T (both valid), and they incorporate A in the new strand, generating an A-T pair.

In a typical set-up, treated and untreated DNA samples, are sequenced by NGS: unmethylated C will be read as T, and by comparison with untreated samples the C → T mutation pattern will indicate the original methylation status





T U C 5mC



Important

U/T equivalence is typical of type A pol (Taq). Type B pol (Pfu) would not be able to amplify bisulfite treated DNA.

Genome-wide output of bisulfite-seq