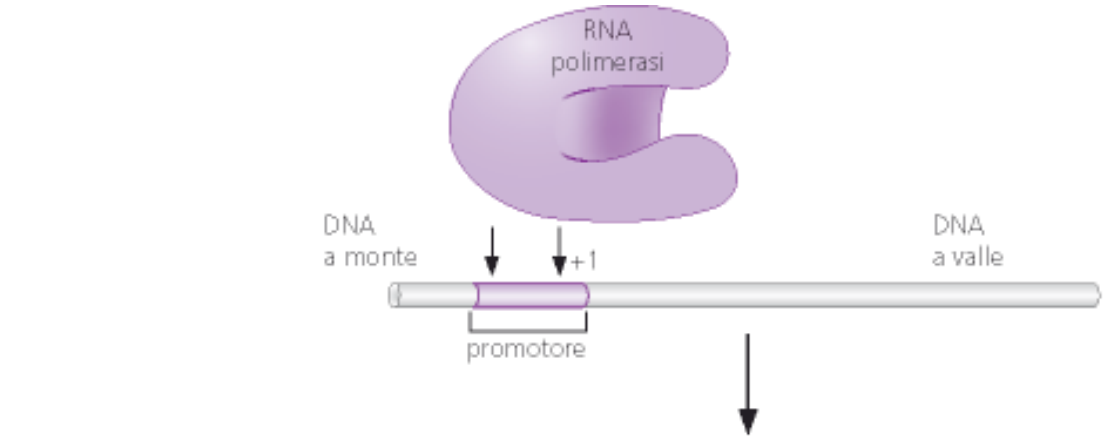
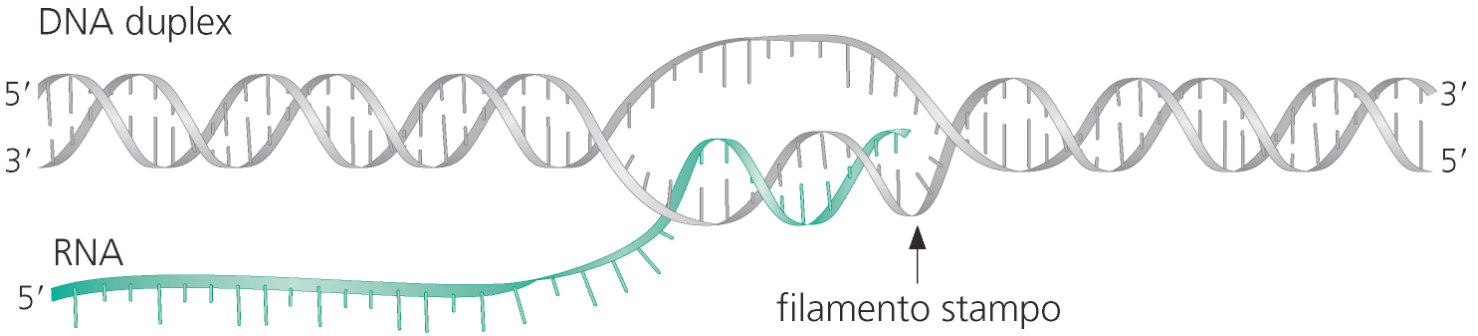
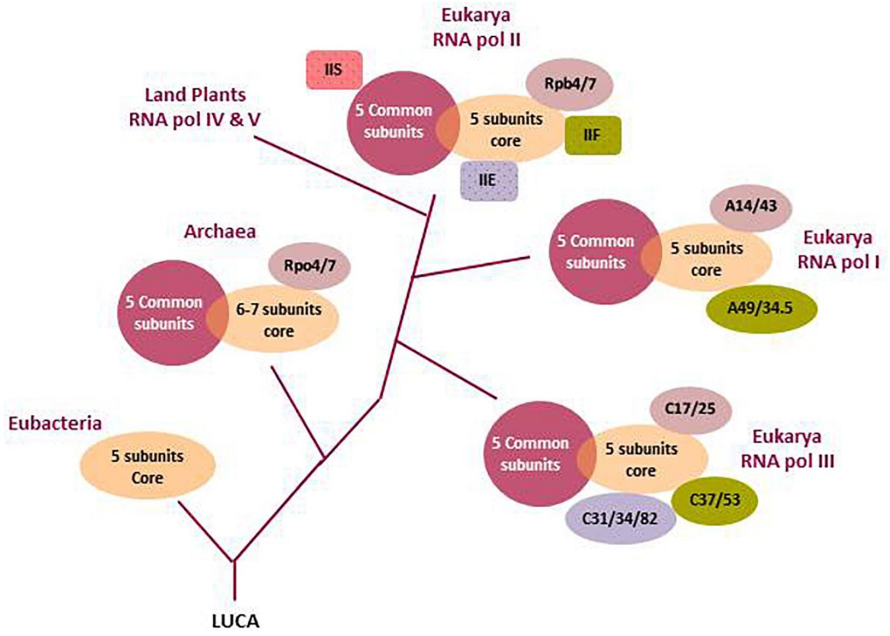
RNA transcription





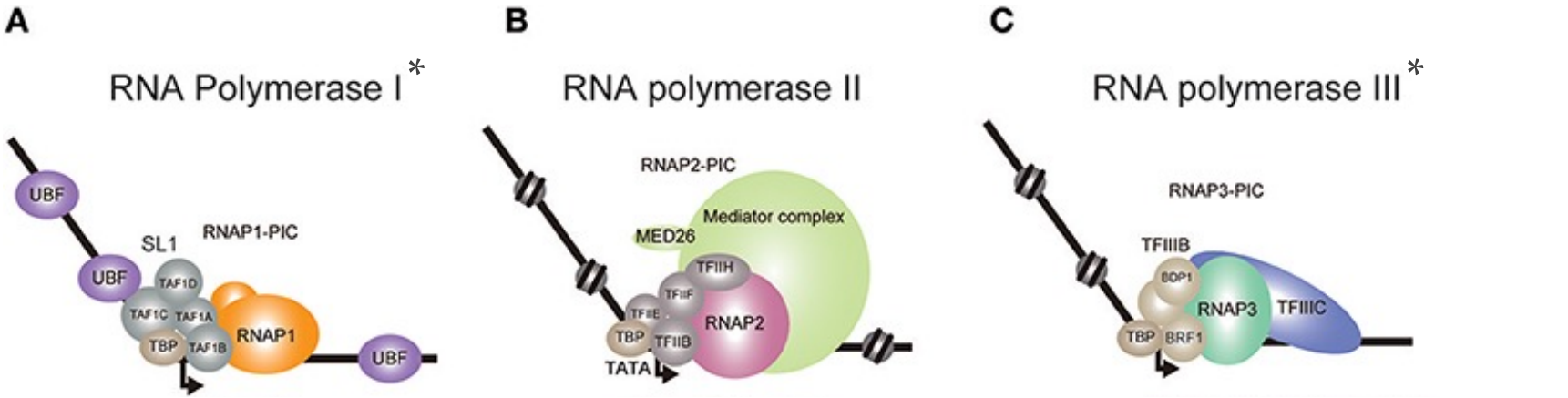
RNA polymerases

All living beings have RNA pols with a core of five to seven subunits. After *Eubacteria* separation, the common ancestor of *Archaea* and *Eukarya* added additional peripheral subunits. Finally, after eukaryote emergence, the *Archaea*-derived nucleus started to develop specialized RNA polymerases. Specialized RNA pols I and III integrated some transcription factors as permanent subunits which, in RNA pol II, remain independent (TFIIS, TFIIF, TFIIE).



Eubacteria	Archaea	Eukarya RNA pol I	Eukarya RNA pol II	Eukarya RNA pol III
β'	Rpo1	A190	Rpb1	C160
β	Rpo2	A135	Rpb2	C128
α	Rpo3	AC40	Rpb3	AC40
α	Rpo11	AC19	Rpb11	AC19
	TFS	A12	Rpb9+ TFIIS	C11
	Rpo5	Rpb5	Rpb5	Rpb5
ω	Rpo6	Rpb6	Rpb6	Rpb6
	Rpo8	Rpb8	Rpb8	Rpb8
	Rpo10	Rpb10	Rpb10	Rpb10
	Rpo12	Rpb12	Rpb12	Rpb12
	Rpo4	A14	Rpb4	C17
	Rpo7	A43	Rpb7	C25
		A49	TFIIFα	C37
		A34.5	TFIIFβ	C53
	Rpo13		TFIIE(α/β)/TFIIF	C31/34/81

RNA polymerases



47S pre-rRNA

mRNAs, lnRNAs, snRNAs

tRNAs, 5S rRNA, U6 snRNA

Types of RNAs transcribed from active regulatory elements

Type	Acronyms	Estimated number	Half-life	Processing features
pre-messenger RNA	pre-mRNA	19,954	minutes to hours	capped spliced polyA tail
long non-coding RNA	lncRNA lincRNA	30,000	minutes to hours	capped spliced polyA tail
enhancer RNA	eRNA	40,000-65,000	minutes	capped
promoter -associated RNA	uaRNA paRNA PROMPT	Gene and lncRNA promoters	minutes	capped

* RNA Pols I and III make up over 80% of the transcriptional activity in a growing cell.

Eukaryotic Transcriptional Regulation

1. Level of Chromatin (DNA accessibility)

- DNA methylation
- Histone modifications
- Histone remodeling
- Nucleosome composition

2. Level of DNA (Interaction with basal transcription machinery)

- Regulatory sequences (enhancers, silencers)
- Transcription factors (activators, repressors)

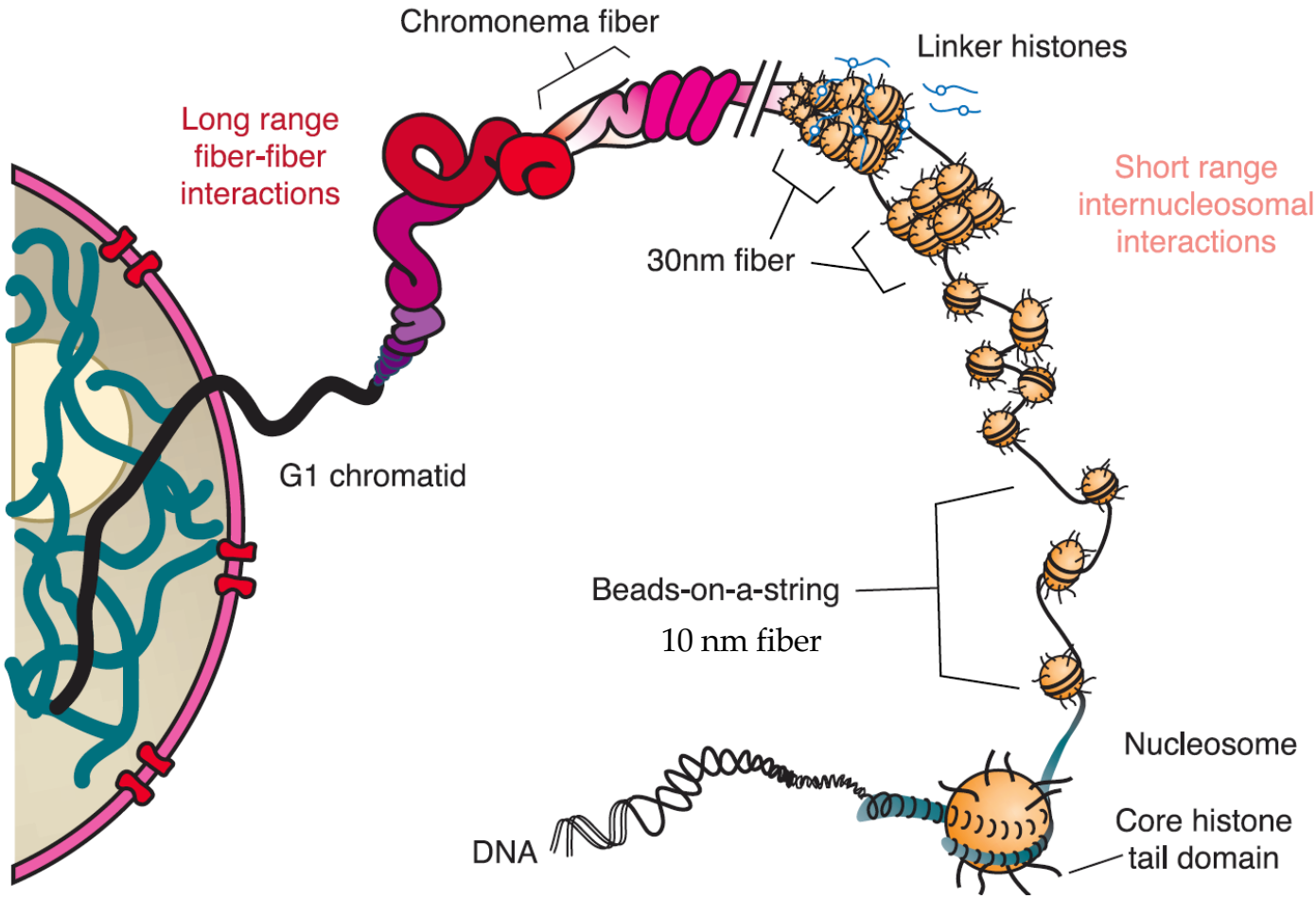
3. Level of Regulatory RNA (Interaction with DNA, RNA or protein)

- Small and long non-coding RNAs

Eukaryotic Transcriptional Regulation

-Level of Chromatin-

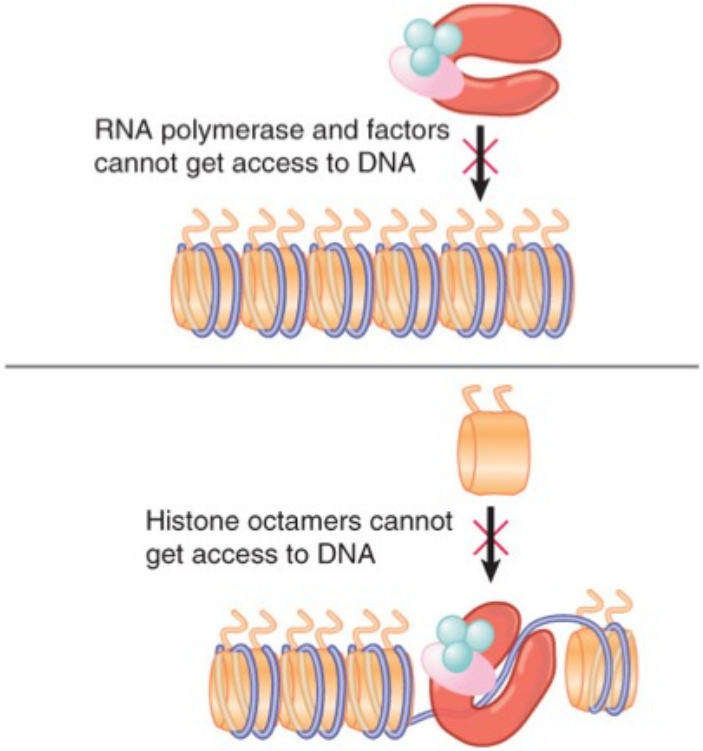
Chromatin-structure - reminder



- **Several levels of condensation**
  - 10 nm extended fiber with nucleosomes like beads-on-a-string (7x compaction)
  - 30 nm condensed fiber (40-50x compaction)
  - And higher

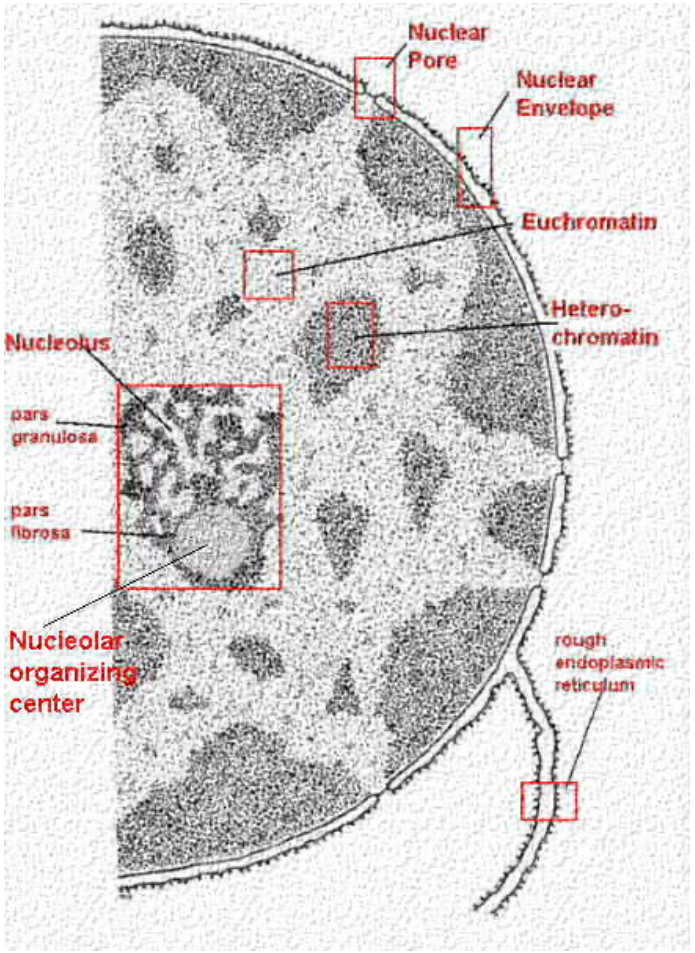
Chromatin compaction influences activity of DNA in transcription

Chromatin imposes a major limitation to transcription by three eukaryotic RNA pols preventing their direct targeting to gene promoters, which probably explains why all nuclear RNA pols are first engaged in pre-initiation complexes before starting transcription.



Chromatin compaction influences activity of DNA in transcription

- **Heterochromatin** -transcriptionally silent
- **Euchromatin** - transcriptionally active



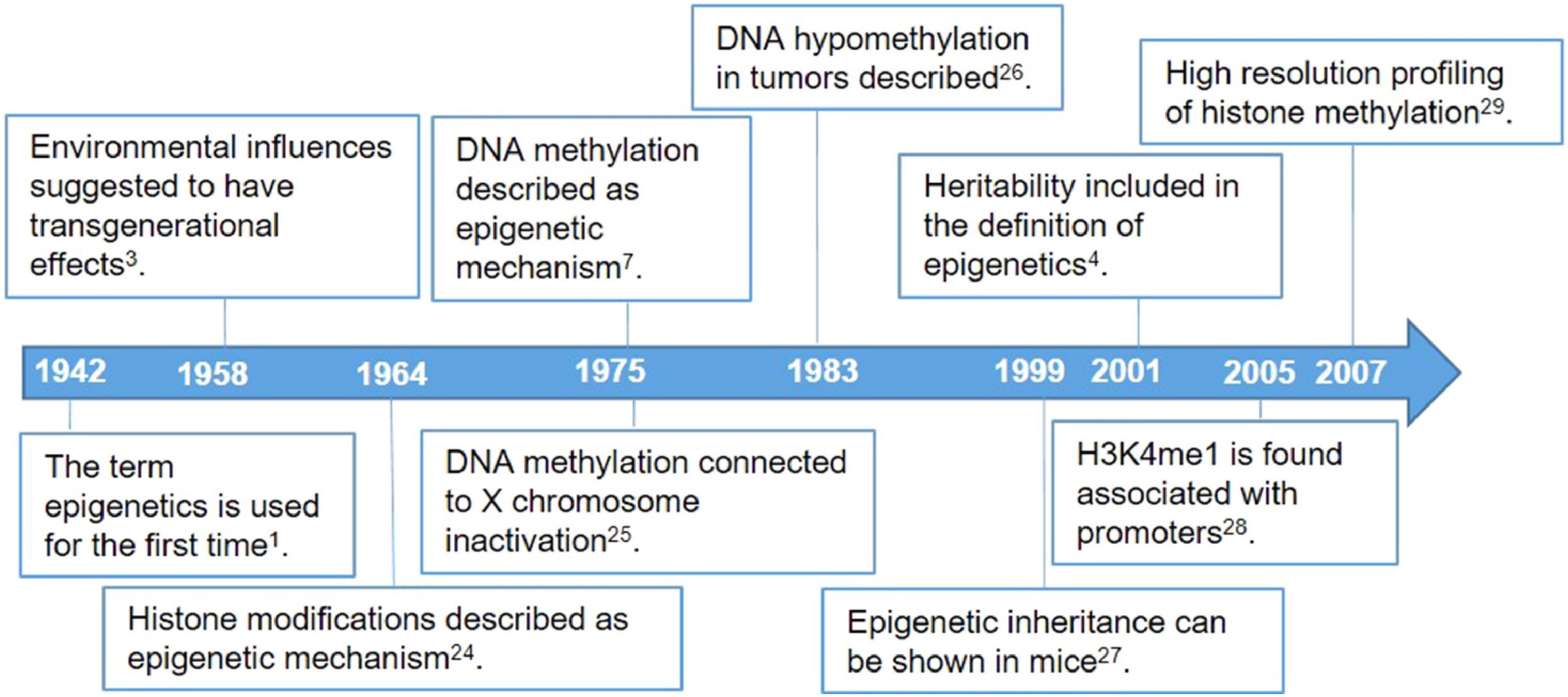
Miescher, Flemming, Kossel and Heitz defined nucleic acids, chromatin and histone proteins, which led to the cytological distinction between euchromatin and heterochromatin (from 1869 to 1928).

Epigenetic control

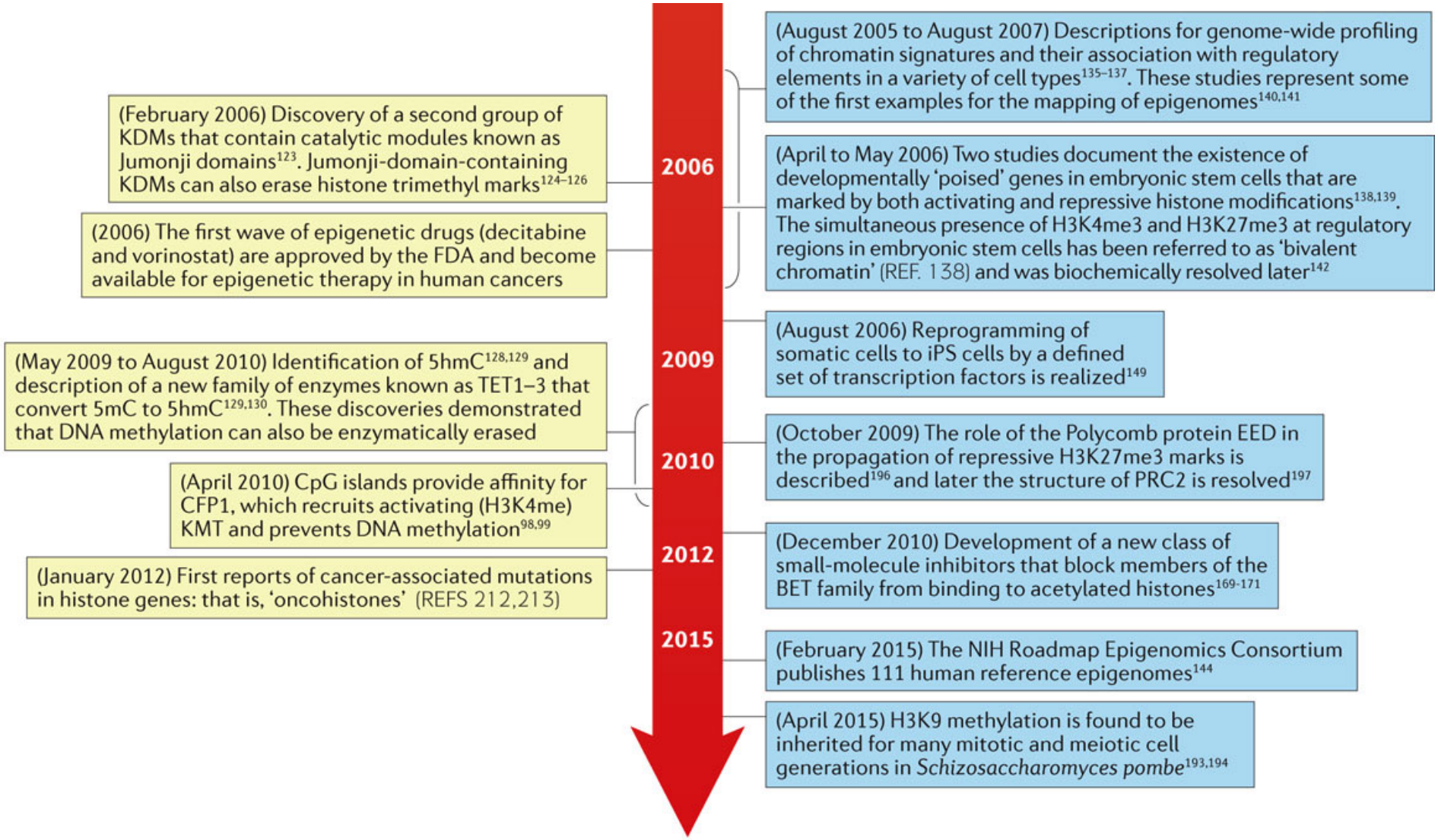
1942, Waddington coined the term 'epigenetics', which he defined as changes in phenotype without changes in genotype, to explain aspects of development for which there was little mechanistic understanding.

2016, epigenetic mechanisms transduce the inheritance of gene expression patterns without altering the underlying DNA sequence but by adapting chromatin, which is the physiological form of our genetic information.

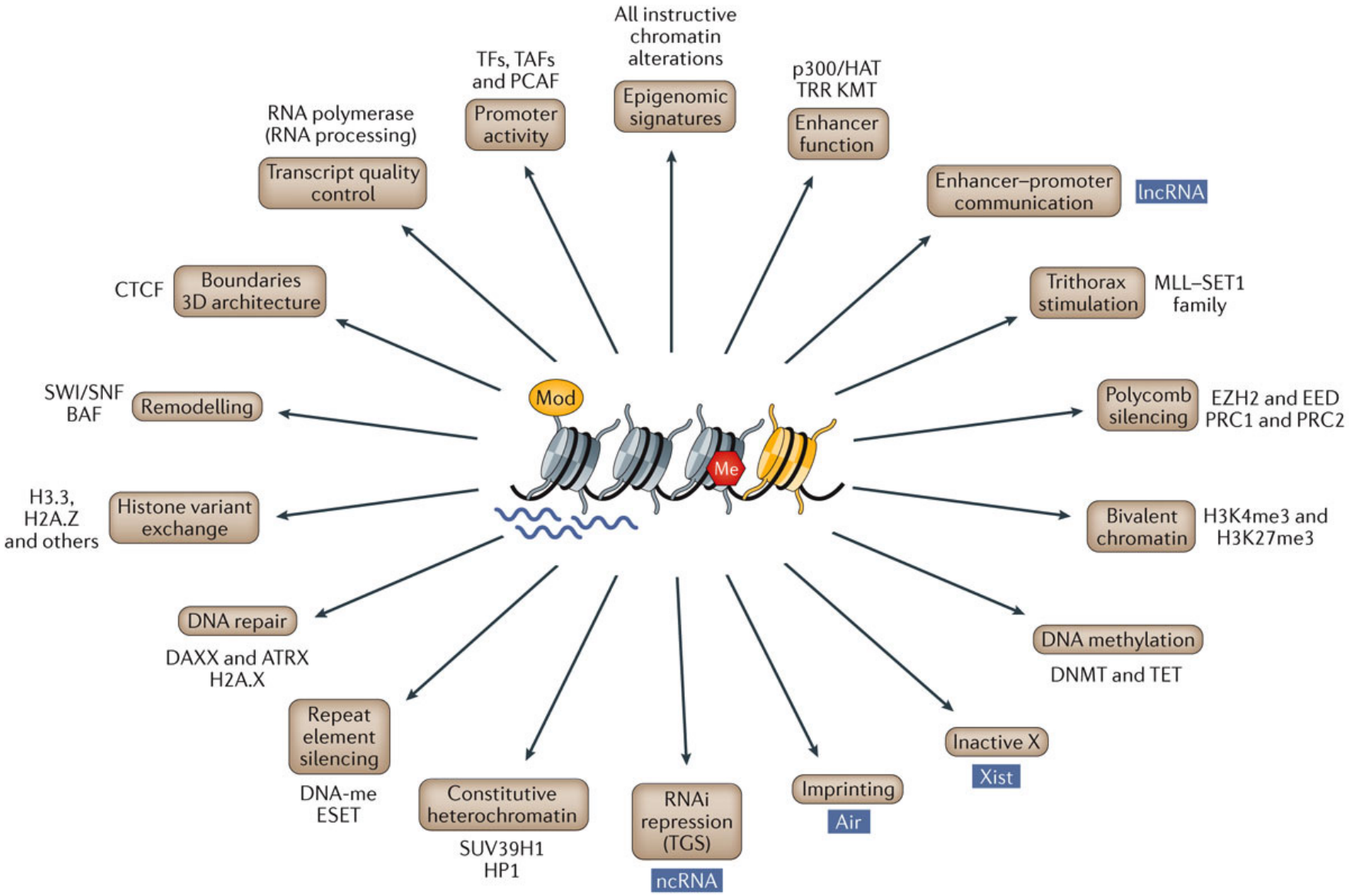
Timeline of major discoveries and advances in epigenetic research



Timeline of major discoveries and advances in epigenetic research



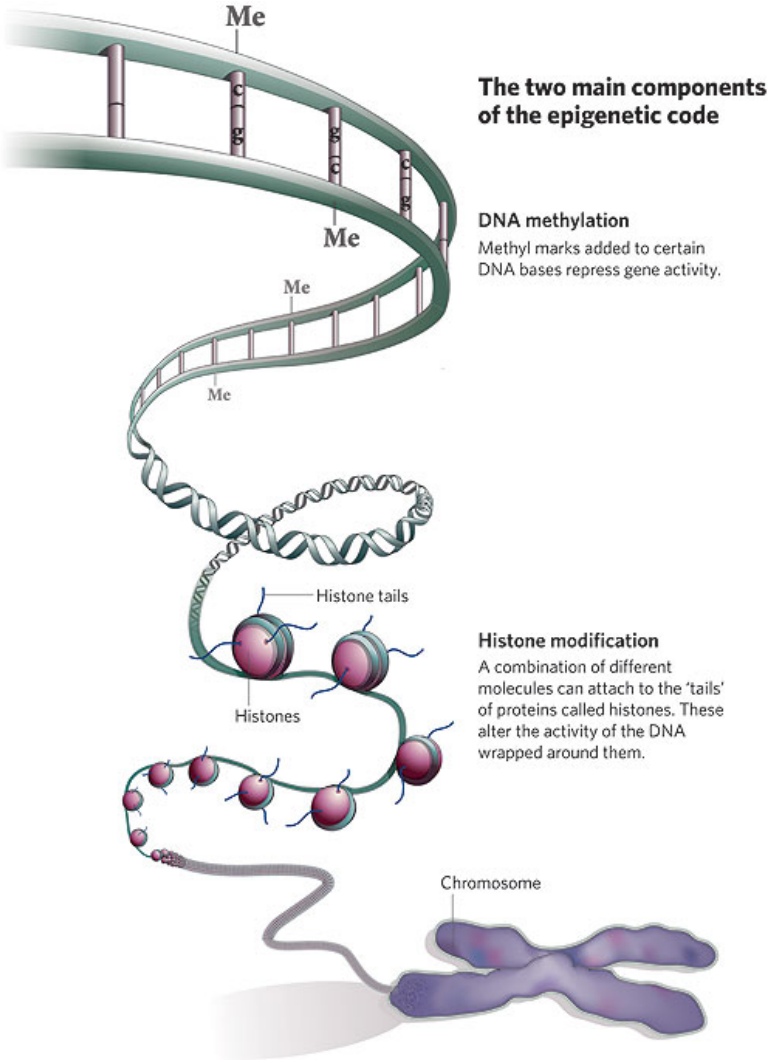
Key examples of chromatin contribution to epigenome function



Nature Reviews | Genetics

Molecular Mechanisms that regulate transcription

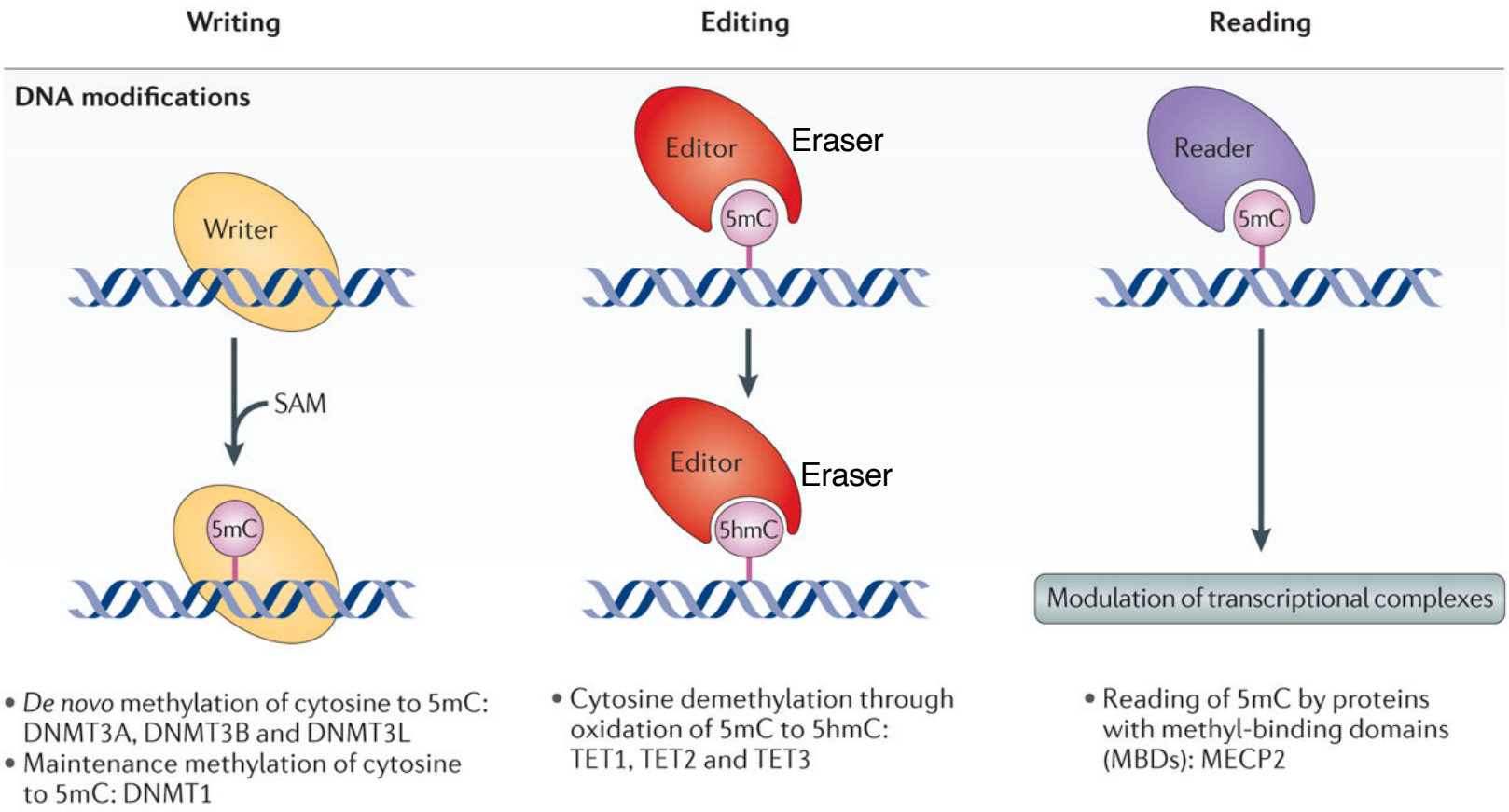
- DNA methylation
- Histone modifications



DNA methylation

DNA modifycation

DNA methylation recruited specific proteins that modulates transcription. Moreover, specific proteins (Erasers or Editor) can recognize this modification

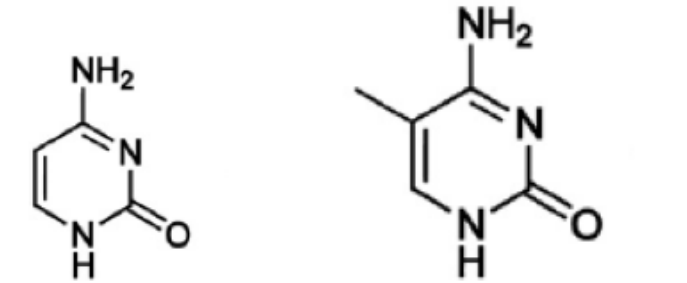


It is a frequent modification in plants (~14%) and mammals (~8%) , very rare in flies (~0.03%), not present in yeast and nematodes.

DNA methylation

- The predominant epigenetic modification of DNA in mammalian genomes is methylation of **cytosine nucleotides (5-MeC)**.

- The primary target sequence for DNA methylation in mammals is **5'- CpG-3' dinucleotides**.

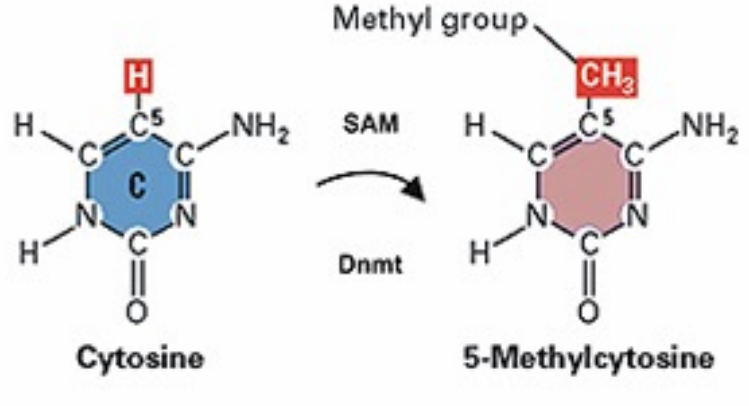


Cytidine
C

5-methyl cytidine
5mC

Roles of DNA methylation

Cytosine DNA methylation is a stable epigenetic mark that is crucial for diverse biological processes

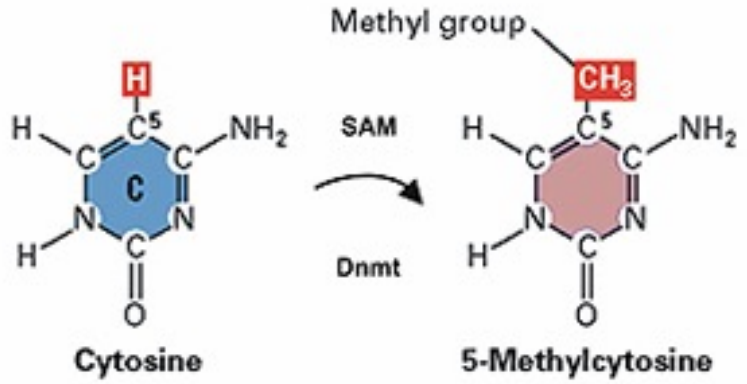


- Transcriptional silencing
- Protecting the genome from transposition
- Genomic imprinting
- X inactivation
- Tissue specific gene expression

Roles of DNA methylation

Sites of methylation:

- Inactive X
- Imprinted loci
- Transposon-derived sequences
- CpG islands

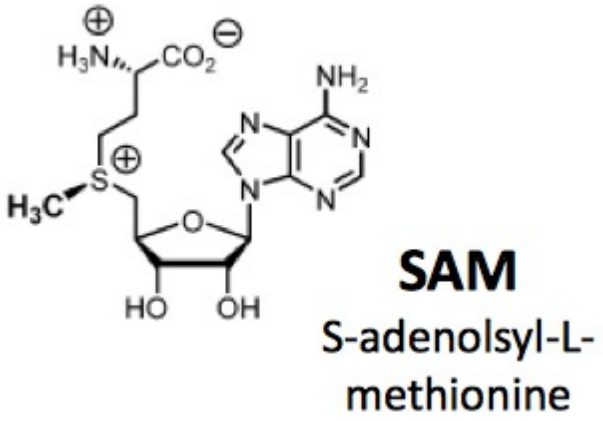


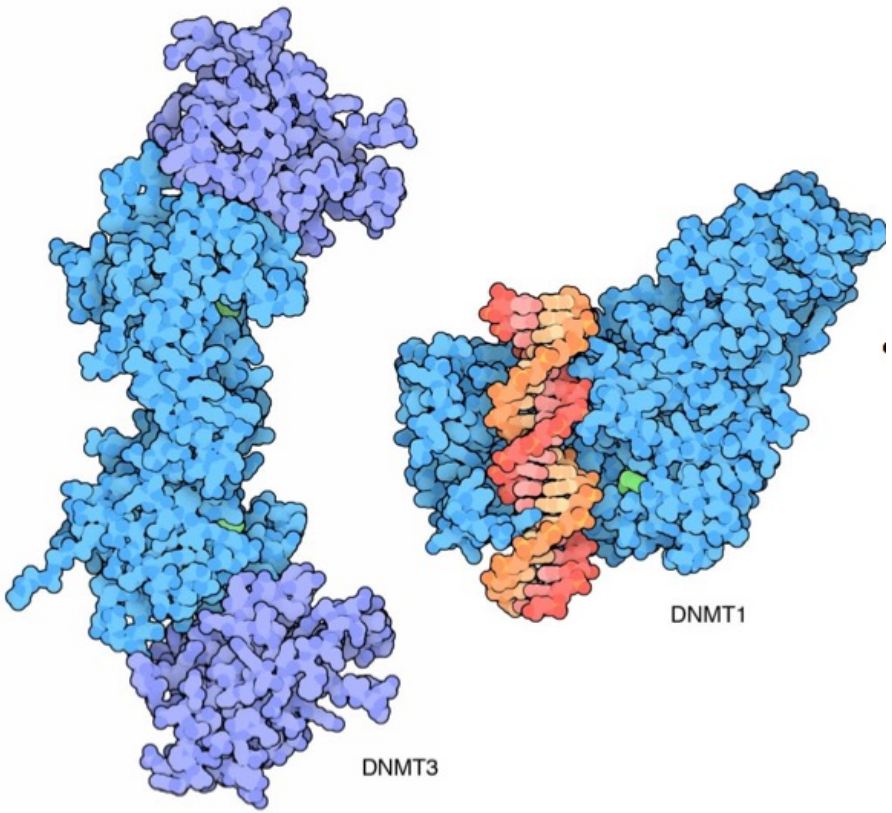
CpG Methyltransferases

DNMT1 -Maintenance methylation

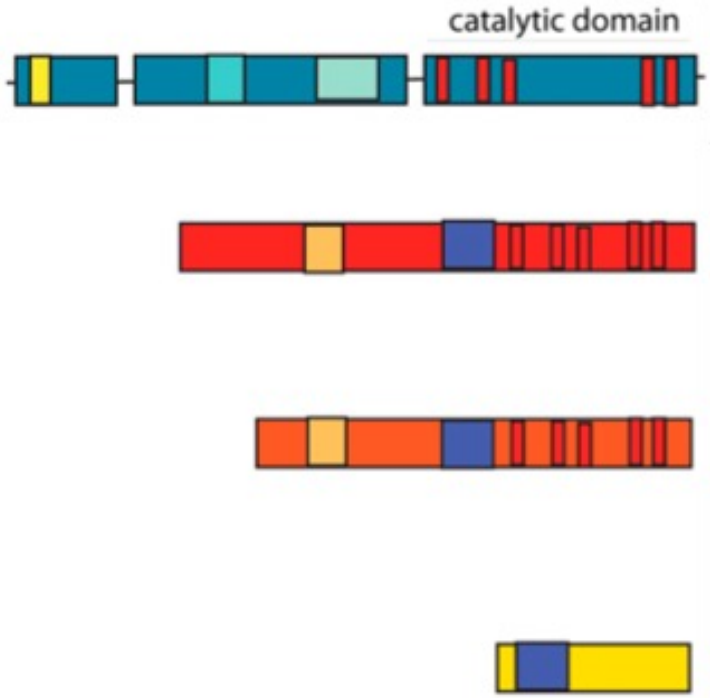
- Looks for hemimethylated CpG and maintains methylation pattern following replication

DNMT3 - De novo methylation





The DNMT family of enzymes



Family member	Function and effect on hematopoiesis with deletion
DNMT1	Maintenance DNA methyltransferase. Germline deletion embryonic lethal. Conditional deletion with *Mx*-Cre results in decreased HSC self-renewal and increased cycling of HSC's.
DNMT3A	*De novo* DNA methyltransferase. Germline deletion results in death by 4 weeks. Conditional deletion with Mx-Cre has minimal effects in primary mice or primarily transplanted mice. With serial transplantation however there is a huge increase in HSC's yet defect of differentiation .
DNMT3B	*De novo* DNA methyltransferase. Germline deletion embryonic lethal. Conditional deletion with *Mx*-Cre has minimal effects but Dnmt3a/Dnmt3b double knockout mice appear to have more pronounced effect on HSC's with some conflict in current reports.
DNMT3L	Catalytically inactive but stimulates enzymatic activity of DNTM3A. Germline knockout mice are viable but sterile. Hematopoietic phenotype of Dnmt3L knockout mice not reported.

Leucine zipper
PWWP domains
Zinc finger
PHD-like domains
Polybromo domain

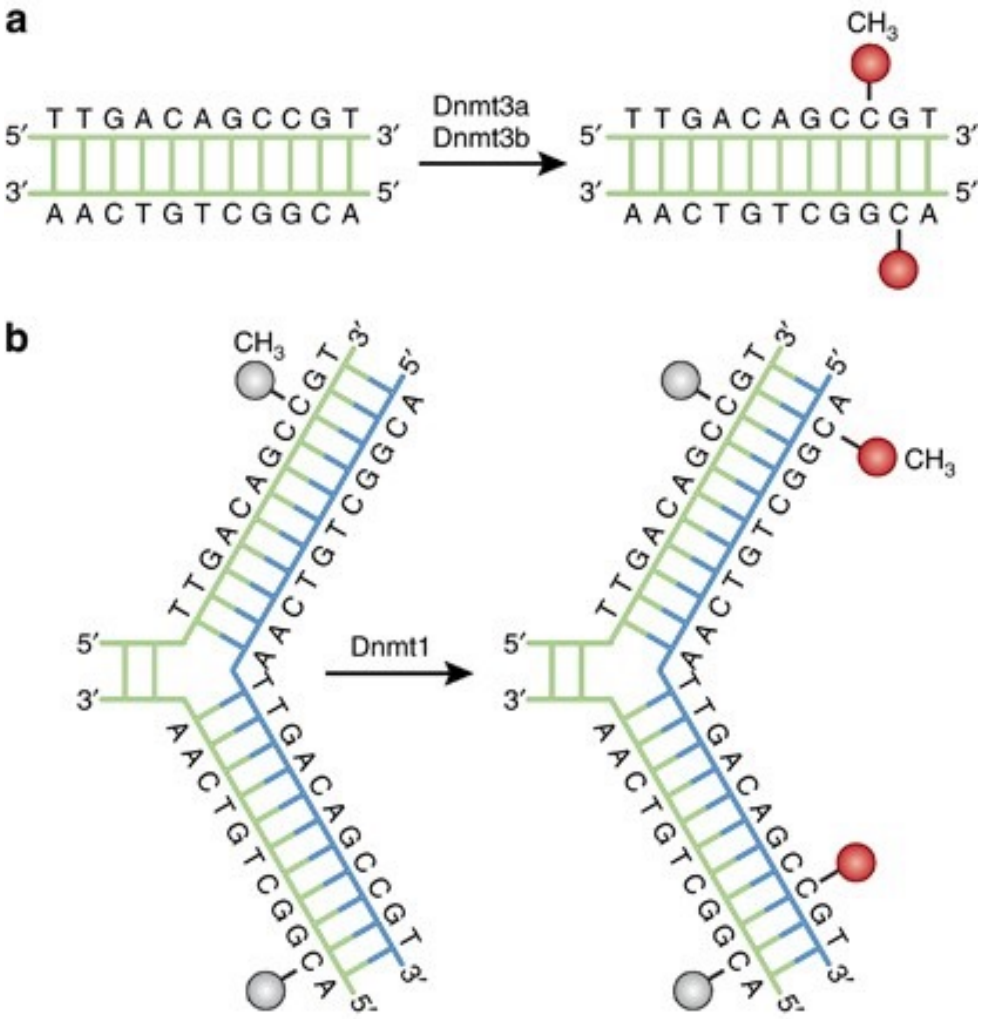
Although Dnmt3a is expressed relatively ubiquitously, Dnmt3b is poorly expressed by the majority of differentiated tissues with the exception of the thyroid, testes, and bone marrow.
Dnmt3L is mainly expressed in early development and is restricted to the germ cells and thymus in adulthood

Maintenance of DNA Methylation in Mammals

Dnmt1, which is highly expressed in mammalian tissues including the brain. Dnmt1 preferentially methylates hemimethylated DNA.

During DNA replication, Dnmt1 localizes to the replication fork where newly synthesized hemimethylated DNA is formed. Dnmt1 binds to the newly synthesized DNA and methylates it to precisely mimic the original methylation pattern present before DNA replication.
Additionally, Dnmt1 also has the ability to repair DNA methylation.

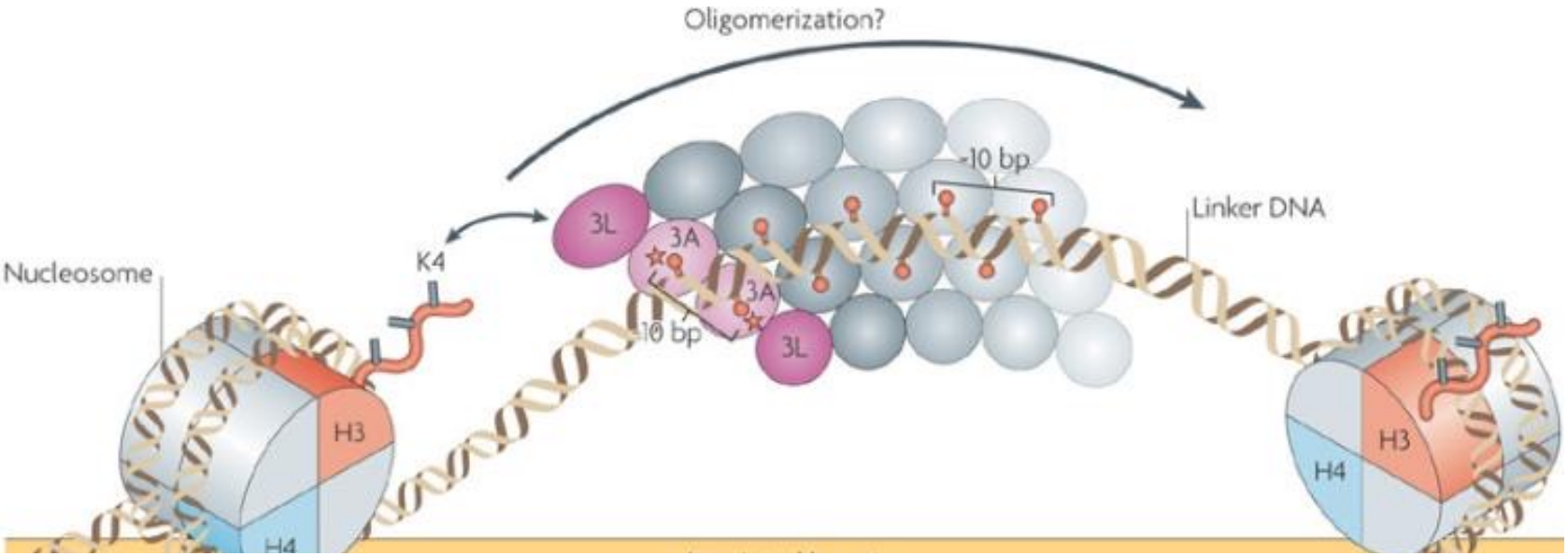
For this reason, **Dnmt1 is called the maintenance Dnmt** because it maintains the original pattern of DNA methylation in a cell lineage.



De Novo DNA Methylation in Mammals

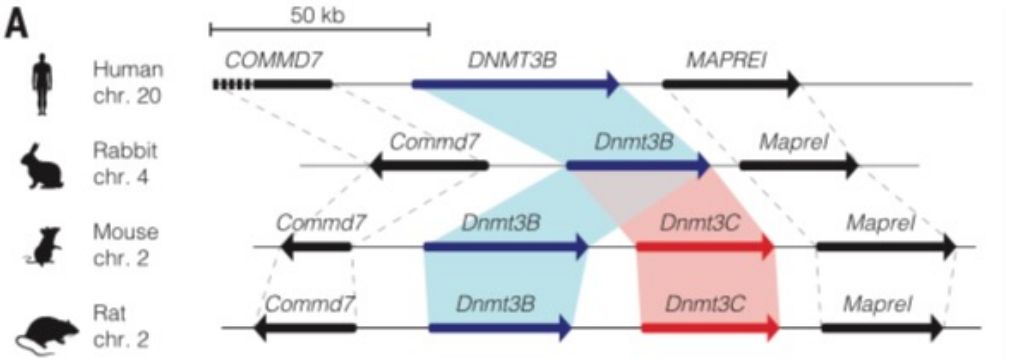
In mammals, DNA methylation patterns are established by the **DNA methyltransferase 3 (DNMT3)** family of *de novo* methyltransferases and maintained by the **maintenance methyltransferase DNMT1**. DNMT3L interacts with unmethylated H3K4 and recruits DNMT3A resulting in the formation of a tetrameric complex.

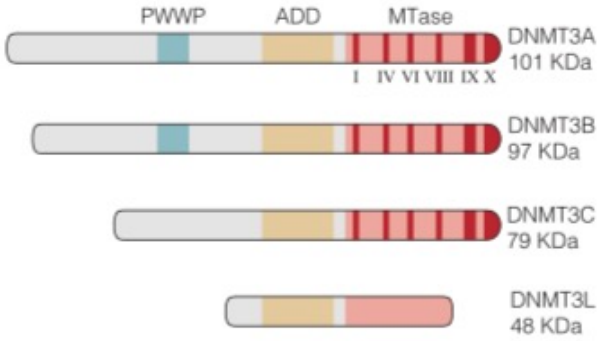
In addition to interactions with unmethylated H3K4 tails, other mechanisms for targeting DNA methylation to specific loci are interactions of DNMT3 with the **histone methyltransferases** G9a (H3K9me), Ezh2 (PRC2 component, H3K27me), suppressor of variegation 3-9 homologue 1 (SUV39H1) and SET domain bifurcated 1 (SETDB1) and **recruitment by lncRNAs**.



The DNMT3C methyltransferase protects male germ cells from transposon activity

Dnmt3C is a *de novo* DNA methyltransferase gene that evolved via a duplication of Dnmt3B in rodent genomes and was previously annotated as a pseudogene. DNMT3C is the enzyme responsible for methylating the promoters of evolutionarily young retrotransposons in the male germ line and this specialized activity is required for mouse fertility.





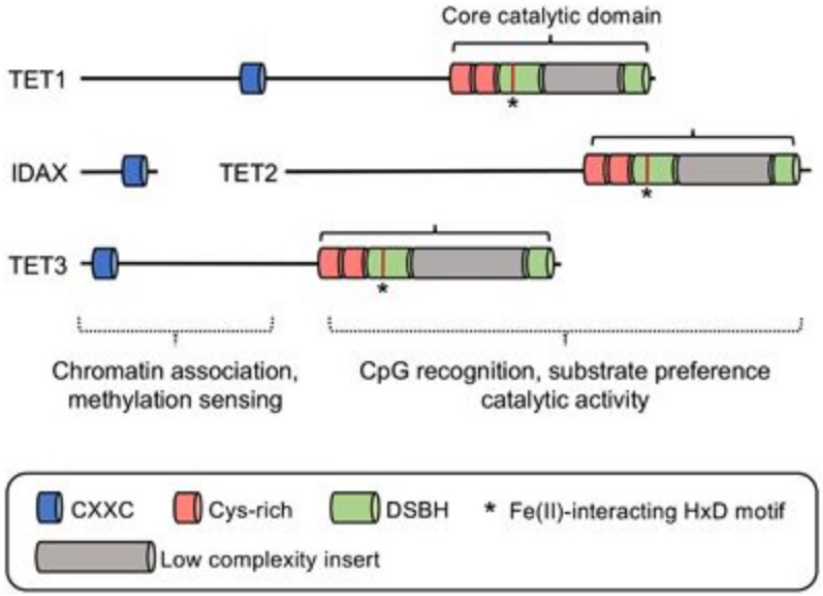
Barau et al., 2016 Nature

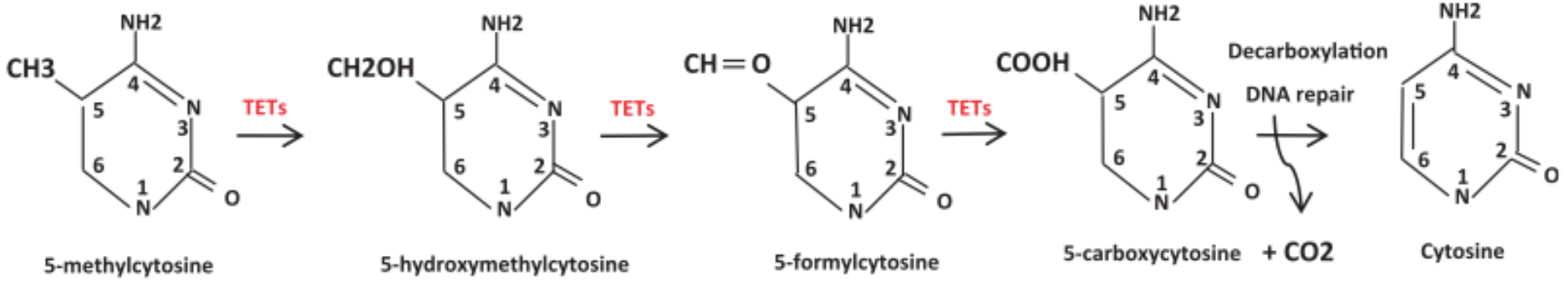
Removing the 5-methylcytosine (5mC)

There is no known mechanism in mammalian cells that can cleave the strong covalent carbon-to-carbon bond that connects cytosine to a methyl group

Active demethylation of 5mC to cytosine is mediated through a series of oxidative steps mediated by the **ten-eleven-translocation (TET)** enzymes, which require **oxygen, iron and α-ketoglutarate**.

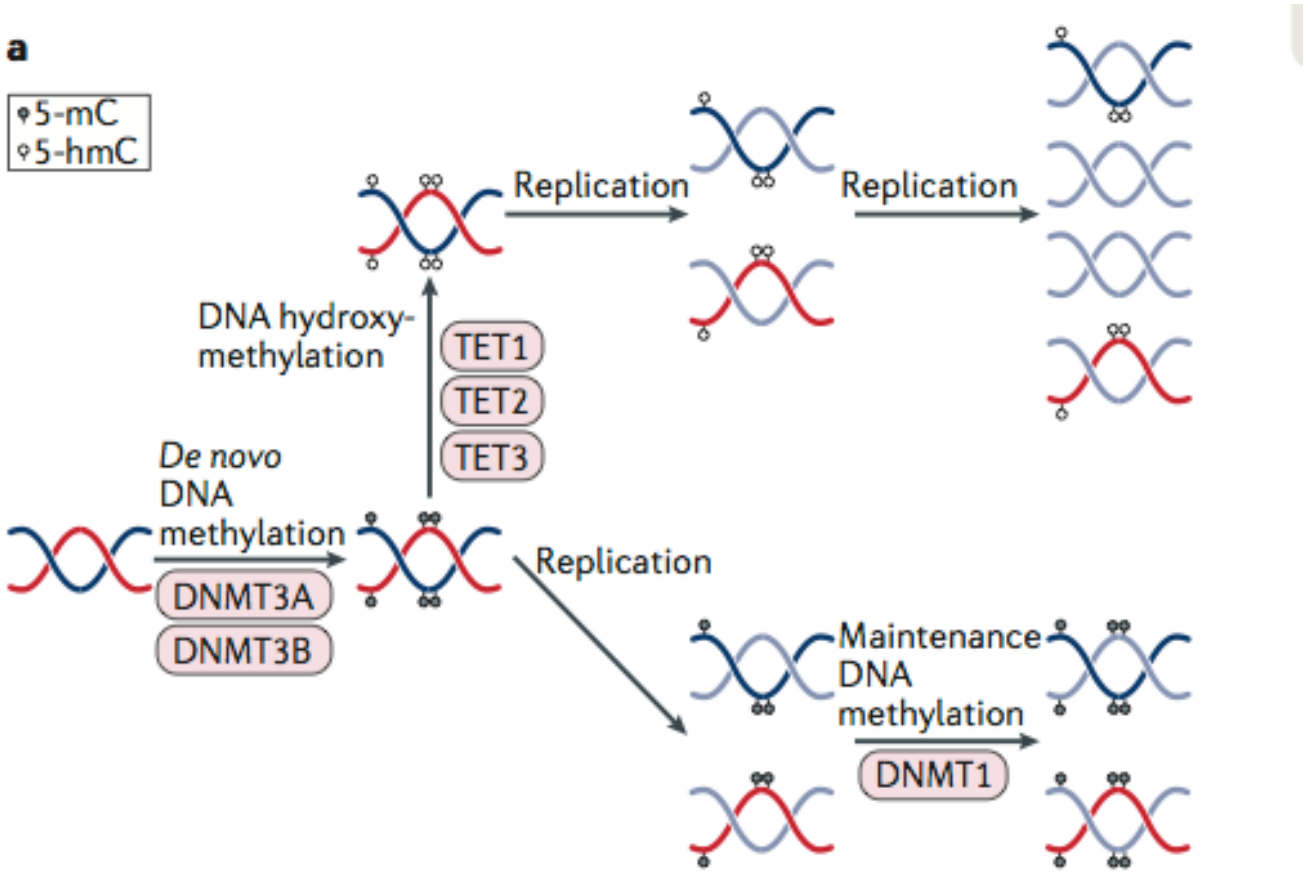
5mC is converted to **5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC)** and **5-carboxylcytosine (5-caC)**, which is then either decarboxylated to cytosine or recognized by thymine DNA glycosylase as part of the mismatch repair system (**BER**).





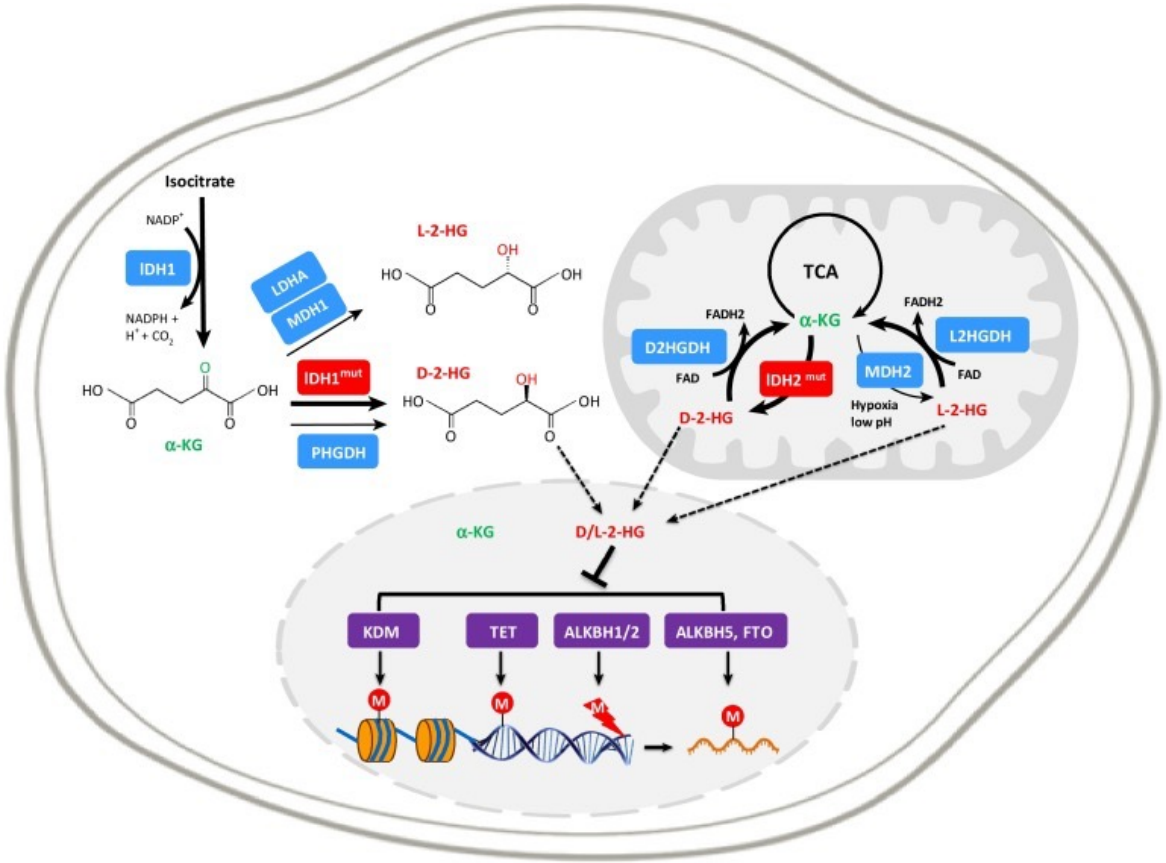
Methylation patterns are reproduced at each round of cell division

DNMT3A and DNMT3B are *de novo* methyltransferases that mark unmethylated DNA, whereas DNMT1 maintains DNA methylation after DNA replication, 5-methylcytosine (5-mC) is oxidized to 5-hydroxymethylcytosine (5-hmC) by enzymes of the TET family (TET1, TET2 and TET3); 5-hmC is not recognized by DNMT1, which results in loss of DNA methylation during replication.

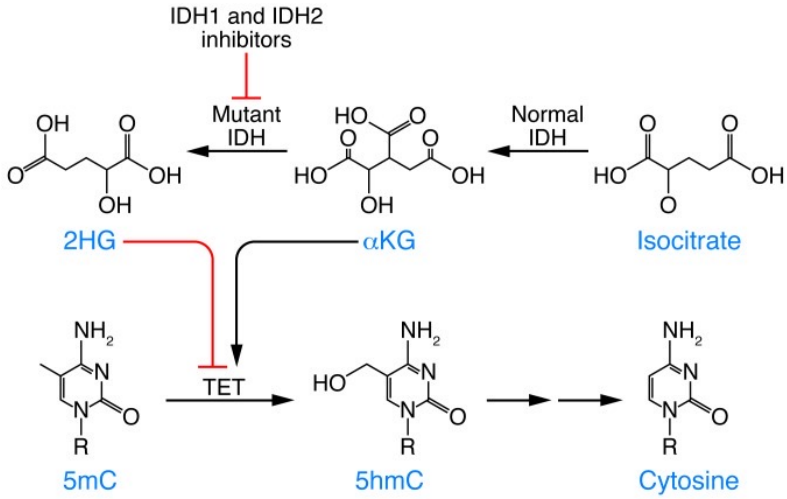


IDH1 and IDH2 mutations lead to epigenetic alteration in cancer

IDH1 and *IDH2* are the most frequently mutated metabolic genes in human cancer, resulting in neomorphic enzymes that convert α-ketoglutarate (α-KG) to 2-hydroxyglutarate (2-HG). 2-HG acts as an antagonist of α-KG to competitively inhibit the activity of α-KG-dependent dioxygenases, including those involved in histone and DNA demethylation.



Trends in Cancer



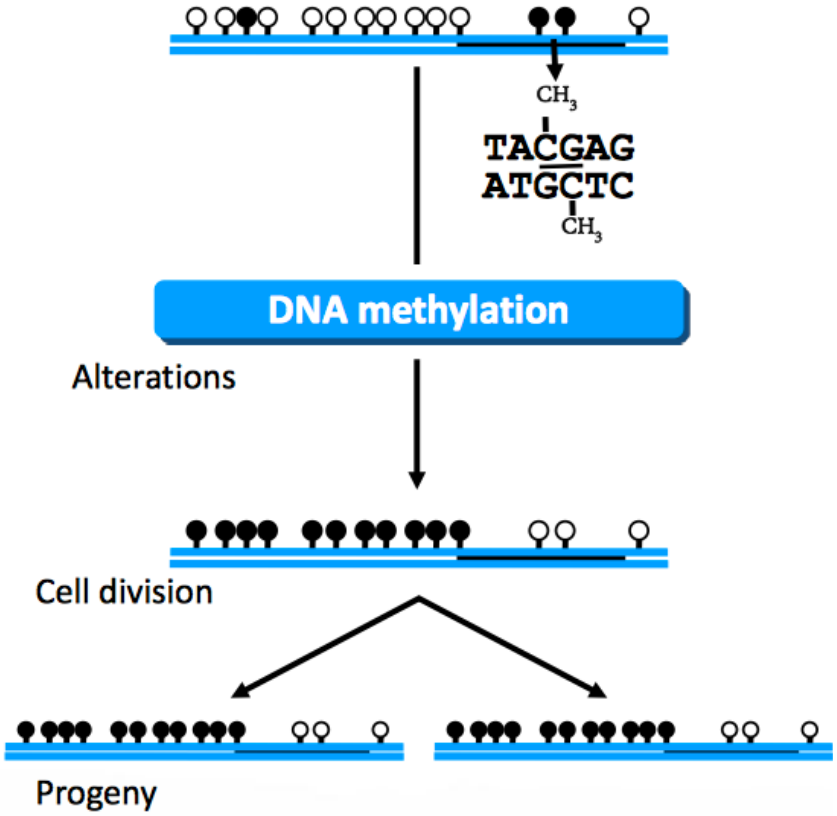
Acute myeloid leukemia

CpG Island

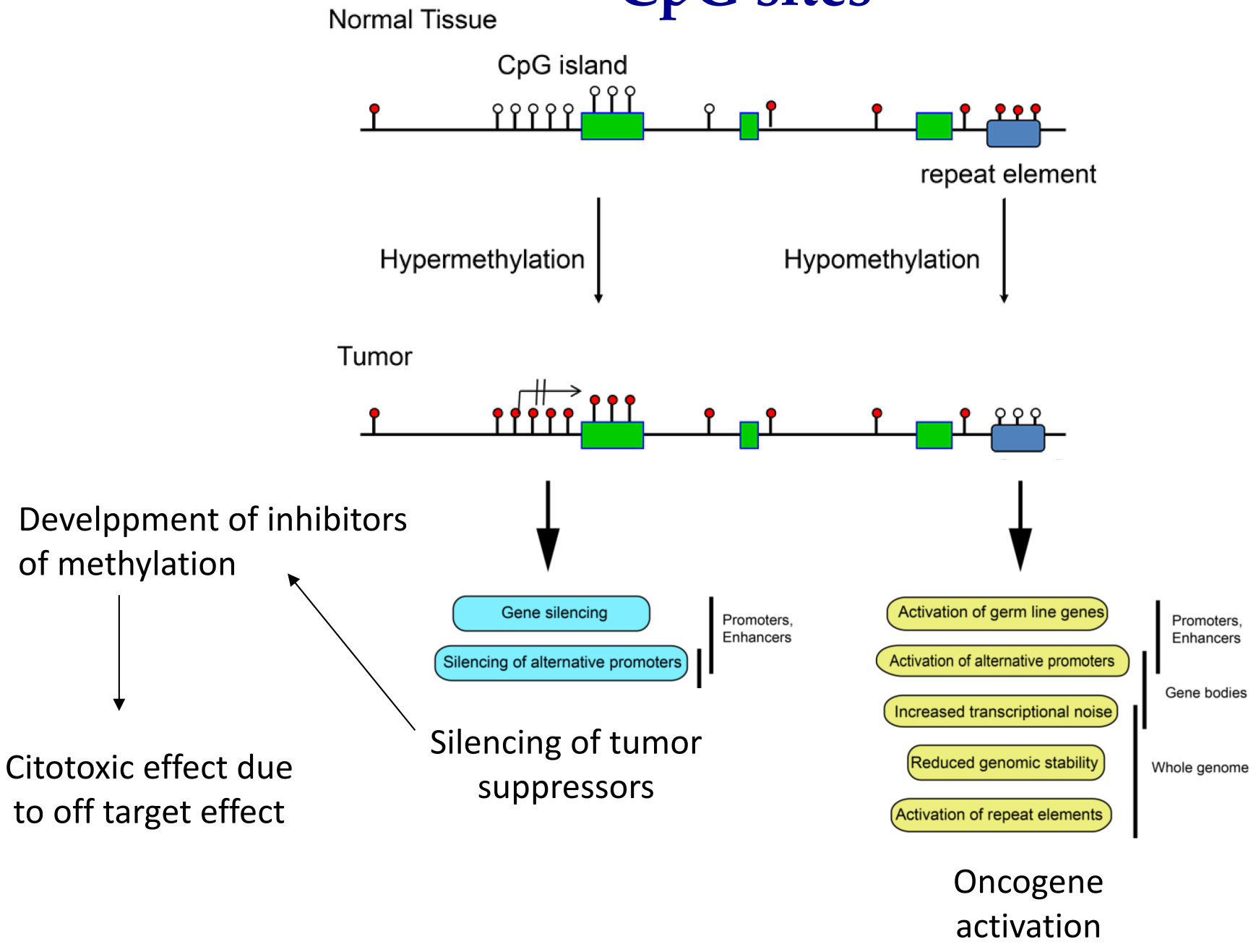
- **Definition**
  - ✓ Small stretches of about 300-3000bp
  - ✓ >50% GC content, regulatory regions

- **Genomic distribution**
  - ✓ ~70% of promoter regions contain CpG islands
  - ✓ Only 1% of remaining genome contains CpG islands
  - ✓ Frequently methylated in tumor cells

- **Methylation status**
  - ✓ Generally non-methylated in promoter of housekeeping and tumor suppessor genes
  - ✓ Typically methylated in transposable elements
  - ✓ Methylated in gene bodies influencing transcription and splicing

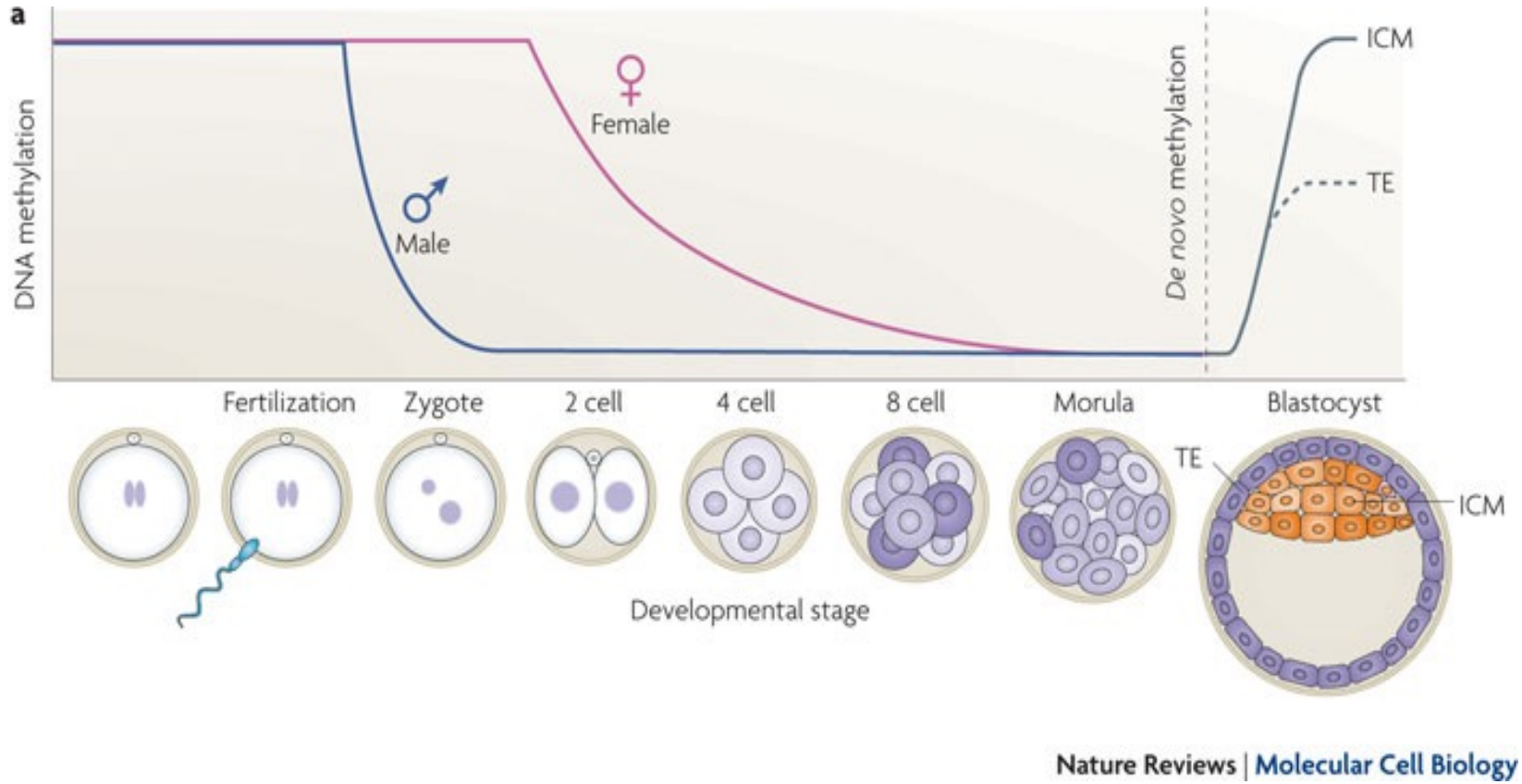


Carcinogenesis alters the methylation profile of CpG sites

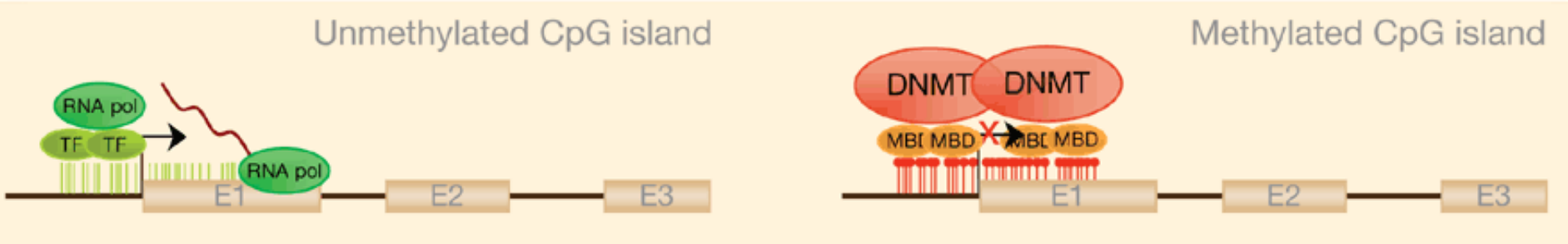


Dynamics of DNA methylation during development

Shortly after a sperm fertilizes an egg, the paternal genome rapidly undergoes genome-wide active DNA demethylation by the TET3 enzyme and remains demethylated following multiple rounds of cell division. During this time, the maternal genome experiences gradual, passive demethylation due to DNMT1 exclusion from the nucleus, hence the sharper demethylation slope for the paternal curve. *De novo* methylation patterns are established by the DNA methyltransferases DNMT3A and DNMT3B during the development of the blastocyst.



Methylated CpG Islands Inhibit Transcription



- Active promoters are usually unmethylated
- **Methylated DNA** recruits **methyl-CpG-binding domain proteins (MBD or MECP)** which may inhibit transcription by several mechanisms
- **Unmethylated CpG** islands recruit **Cfp1** which associates with a histone methyltransferase creating **H3K4me3 (transcriptional activation)**

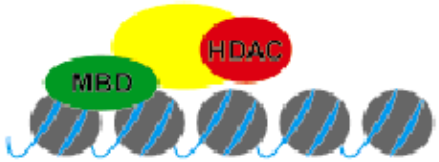
Readers of DNA Methylation

Whereas DNA methylation may itself reduce gene expression by impairing the binding of transcriptional activators, a second class of proteins with a high affinity for 5mC inhibits transcription factor binding.

DNA methylation is recognized by three separate families of proteins:
1- **the MBD proteins** that contain a conserved methyl-CpG-binding domain (MBD) confering them a higher affinity for **single methylated CpG sites**. This family includes MeCP2, the first identified methyl-binding protein, along with MBD1-4. **MBDs are more highly expressed in the brain than in any other tissue**, and many MBDs are important for normal neuronal development and function
2- **The UHRF** (ubiquitin-like, containing PHD and RING finger domain) proteins, including UHRF1 and UHRF2, are multidomain proteins that bind to methylated cytosines. The UHRF protein family first binds to Dnmt1 and then targets it to hemimethylated DNA in order to maintain DNA methylation, especially during DNA replication
3- **the zinc-finger proteins:** Kaiso (binds to consecutive methylated CpG sites), ZBTB4 and ZBTB38 (bind to single methylated CpG site). Similar to the MBD family, repress transcription in a DNA methylation-dependent manner.

Mechanisms for Repression Mediated by MBDs

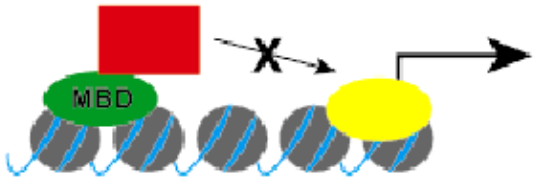
Model 1 - MBD proteins interact with HDAC to generate hypoacetylated, condensed chromatin



Model 2 - MBD proteins coat methylated loci occluding regulatory DNA

Model 3 - MBD proteins alter local DNA and or chromatin architecture

Model 4 - an MBD protein sequesters an essential transcription factor, preventing its function



nature genetics • volume 23 • october 1999

Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2

Ruthie E. Amir[1], Ignatia B. Van den Veyver[2,3], Mimi Wan[5], Charles Q. Tran[3], Uta Francke[5,6] & Huda Y. Zoghbi[1,2,4]

Rett syndrome1 (RTT) is a progressive neurodevelopmental disorder and one of the most common causes of mental retardation in females, with an incidence of 1 in 10,000–15,000. Patients with classic RTT appear to develop normally until 6–18 months of age, then gradually lose speech and purposeful hand use, and develop micro- cephaly, seizures, autism, ataxia, intermittent hyperventilation and stereotypic hand movements. After initial regression, the condition stabilizes and patients usually survive into adulthood. As RTT occurs almost exclusively in females, it has been proposed that RTT is caused by an X-linked dominant mutation with lethality in hemizygous males. Previous exclusion mapping studies using RTT families mapped the locus to Xq28. Using a systematic gene screening approach, we have identified mutations in the gene (*MECP2*) encoding X- linked methyl-CpG-binding protein 2 (MeCP2) as the cause of some cases of RTT. MeCP2 selectively binds CpG dinucleotides in the mammalian genome and mediates transcriptional repression through interaction with histone deacetylase and the corepressor SIN3A . In 5 of 21 sporadic patients, we found 3 *de novo* missense mutations in the region encoding the highly conserved methyl-binding domain (MBD) as well as a *de novo* frameshift and a *de novo* nonsense mutation, both of which disrupt the transcription repression domain (TRD). In two affected half-sisters of a RTT family, we found segregation of an additional missense mutation not detected in their obligate carrier mother. This suggests that the mother is a germline mosaic for this mutation. Our study reports the first disease-causing mutations in RTT and points to abnormal epigenetic regulation as the mechanism underlying the pathogenesis of RTT.

MeCP2 and Rett syndrome

Rett syndrome is a neurodevelopmental disorder linked to the X chromosome:

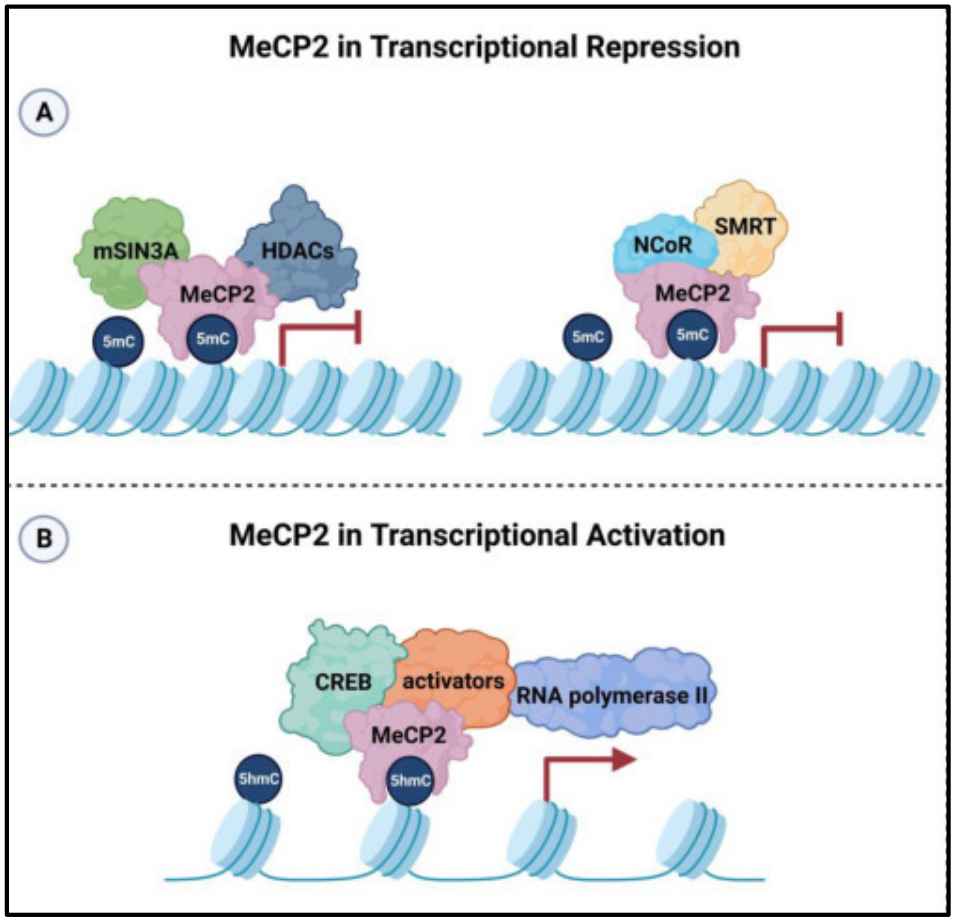
- Autistic traits
- Loss of acquired language
- Loss of motor skills

95% of Rett patients carries mutations on MECP2 gene

pre-genomic era:
MeCP2 represses specific genes in the brain

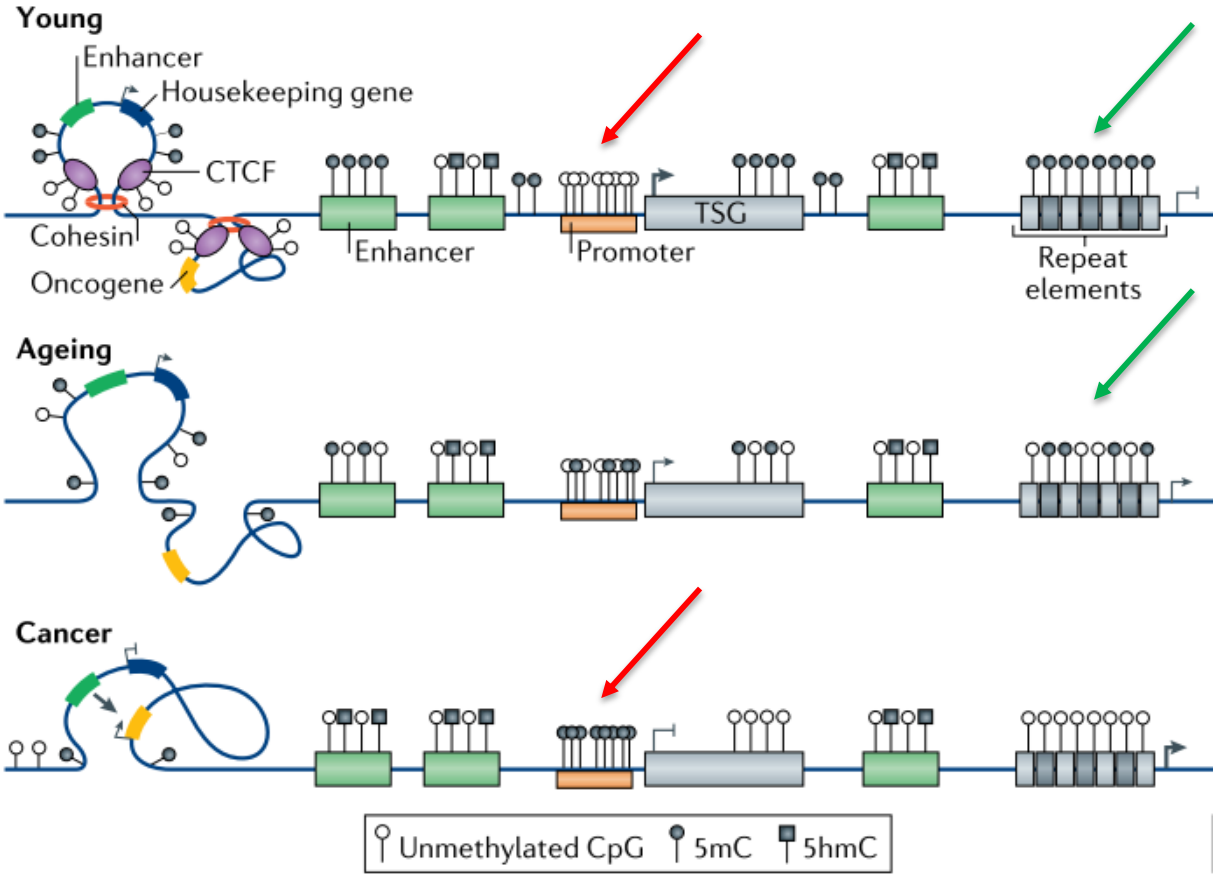
post-genomic era:
-60% MeCP2 binding sites fall outside the genes and only 6% in the CpG islands in the promoters.
-The modulation of gene expression mediated by MeCP2 is repressive for 377 genes while is activating for 2184.

Mode of action of MeCP2



Changes in DNA methylation

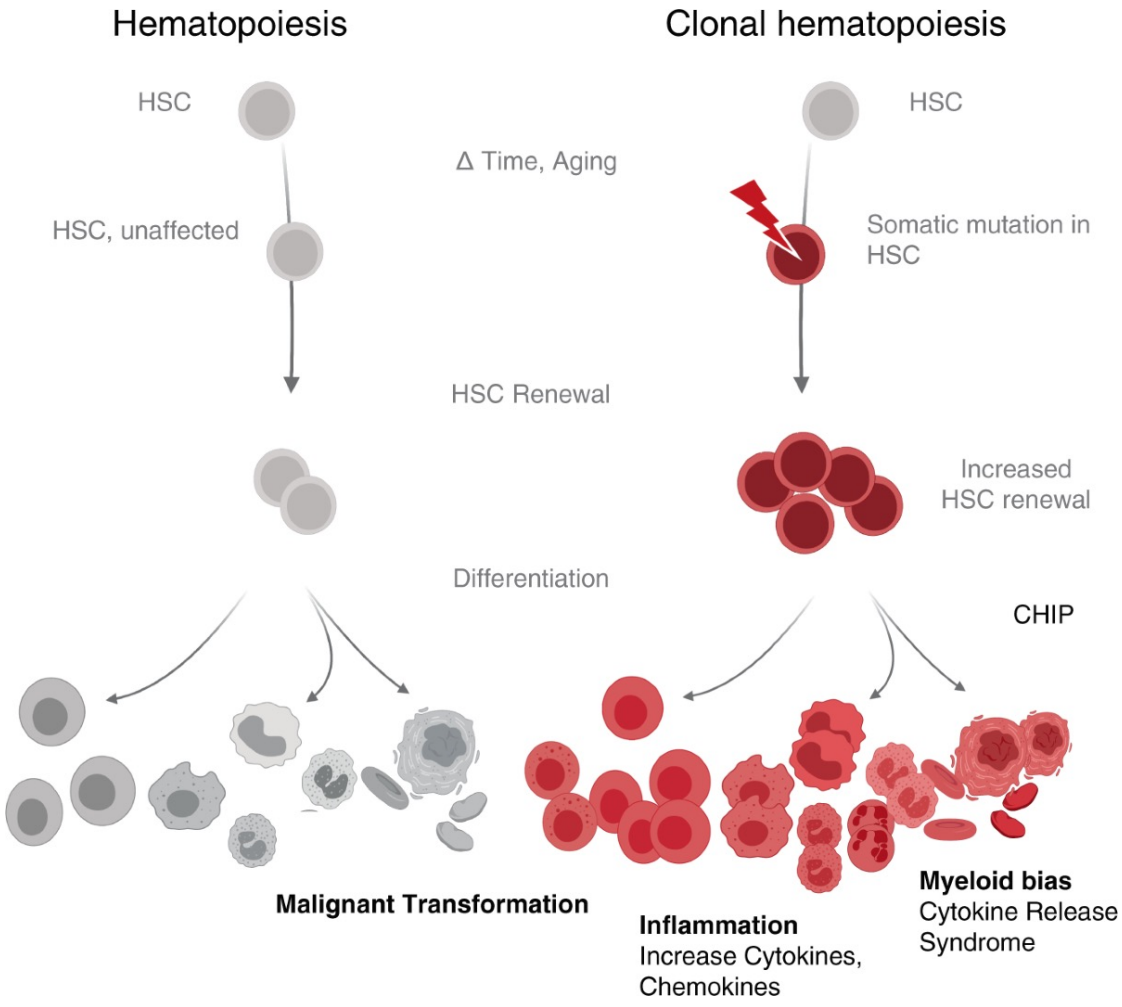
Loss of DNA methylation occurs primarily in constitutive-heterochromatin repeat regions in ageing, while hypermethylation primarily occurs at promoter CpG in cancer. Ageing is accompanied by selective loss and reorganization of heterochromatin and upregulation of transcripts from repeat elements, in particular retrotransposable elements, the genomic transposition of which is associated with the formation of DNA double-strand breaks and can negatively impact genome stability and cause disease.



Clonal hematopoiesis of indeterminate potential (CHIP)

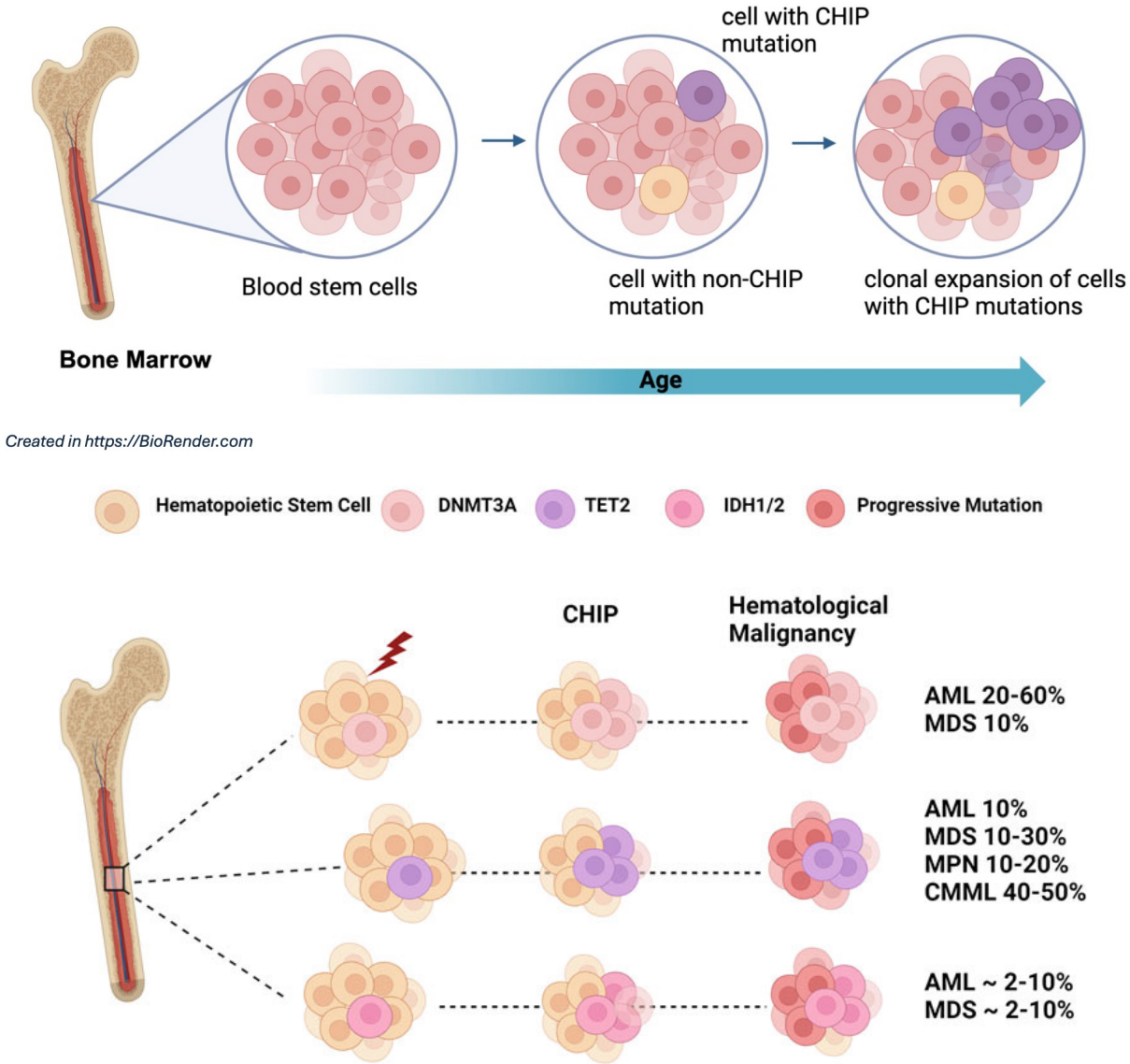
Clonal hematopoiesis of indeterminate potential (CHIP) is a common age-related phenomenon in which hematopoietic stem cells (HSCs) acquire leukemogenic mutations resulting in the selection and expansion of a genetically distinct subpopulation of blood cells. Mutations in **DNMT3A** and **TET2** are two of the most frequent genes found in CHIP.

During hematopoiesis, HSCs normally acquire DNAm patterns consistent with terminal cell lineage, but knockout of Dnmt3a in mice prevents HSCs from establishing new DNAm patterns, leading to a self-renewal pattern. Despite its opposing regulatory role in demethylation, knockdown of Tet2 led to a similar pattern of increased HSC self-renewal, and global loss of hydroxymethylation in HSCs



Clonal hematopoiesis of indeterminate potential (CHIP)

Selective pressures such as age, chemotherapy and environmental insults lead to the expansion of these mutant clones that may progress to myeloid malignancies. The subsequent acquisition of additional mutations in genes such as FLT3 and NPM1 can result in the transformation into overtacute myeloid leukemia (AML).



In addition, IDH mutations are also associated with CHIP in older populations.

Studying DNA Methylation

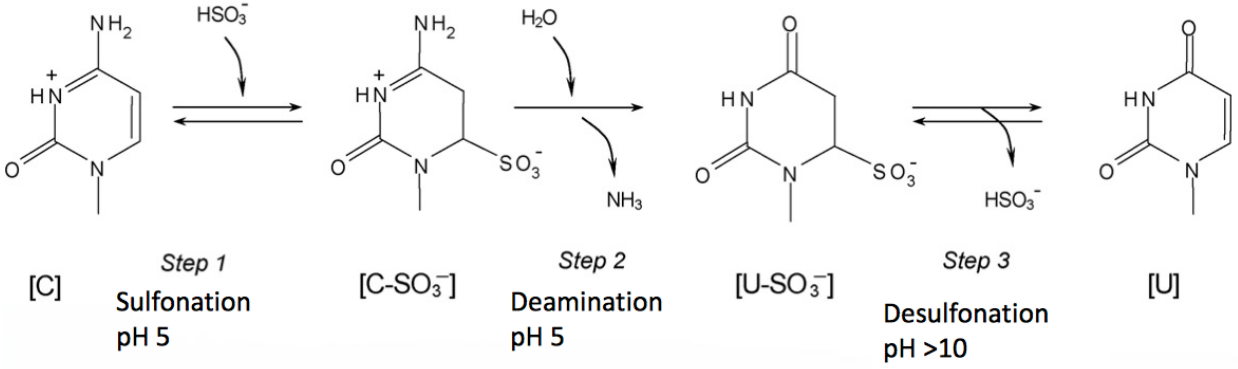
1. **Methylation-sensitive restriction enzymes combined with Southern blots or PCR**

   - Requires complete digestion of methylated DNA to avoid false positives and lacks sensitivity

2. **Methylation specific immunoprecipitation**

3. **<u>Bisulfite modification</u>** (<u>is the most widely used technique for studying DNA methylation</u>)

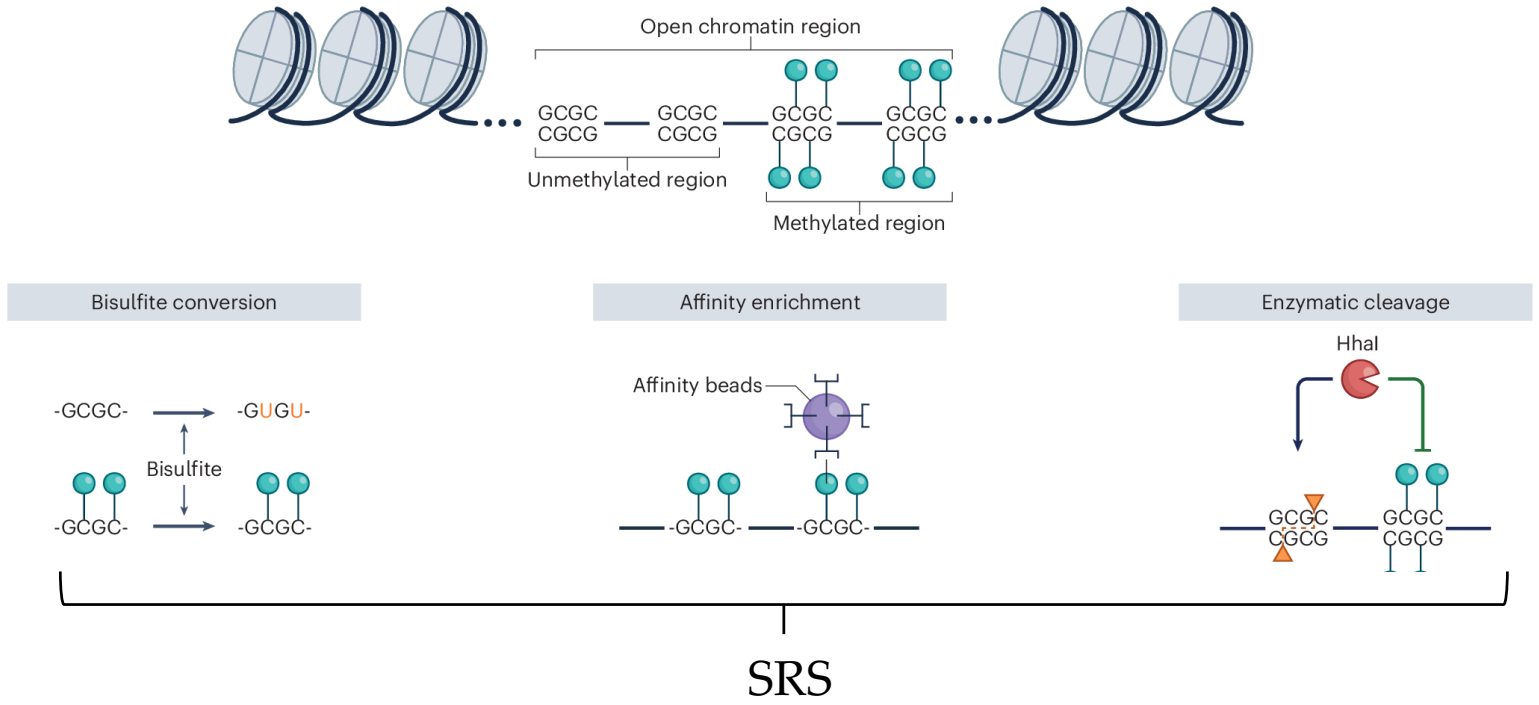
   - C's are converted to U's
   - <u>Methylated C's are NOT converted</u>

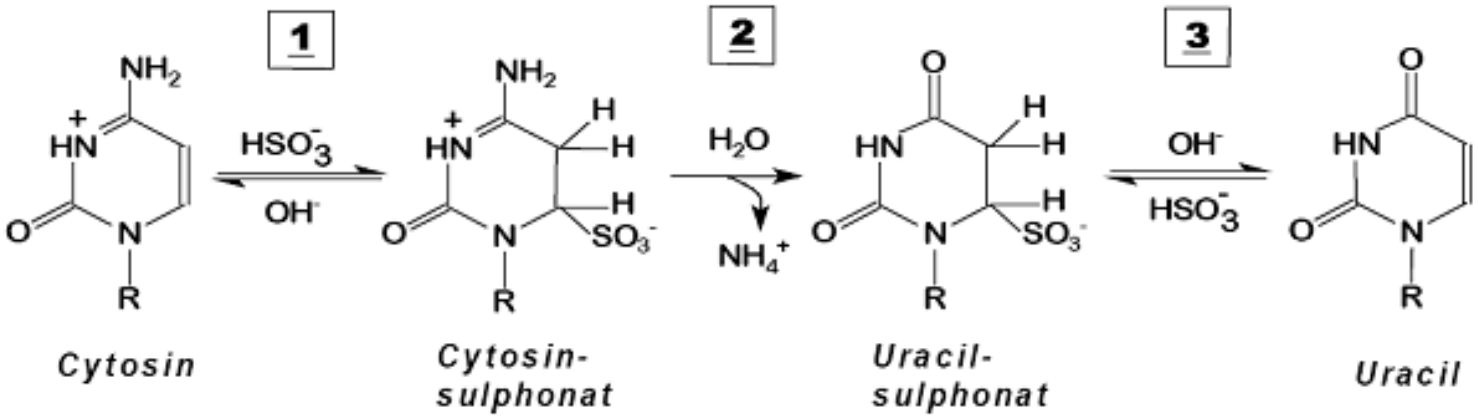


!!!!Standard molecular biology techniques such as PCR and cloning erase information about the position of methylated cytosines in DNA.

Studying DNA Methylation

Short Reads Sequencing (SRS) methods for detecting DNA methylation include bisulfite conversion, affinity enrichment and enzymatic cleavage. In bisulfite conversion, unmethylated cytosines are converted to uracil, while methylated cytosines remain unchanged, allowing their differentiation during sequencing. Affinity enrichment using methyl-binding proteins attached to beads then selectively isolates methylated DNA regions. Subsequent enzymatic cleavage with restriction enzymes (for example, HhaI) that cut at specific unmethylated sites leaves methylated regions intact.





Allele 1 (metilato)

```
---ACTCCACGG---TCCATCGCT---
---TGAGGTGCC---AGGTAGCGA---
```

Allele 2 (non metilato)

```
---ACTCCACGG---TCCATCGCT---
---TGAGGTGCC---AGGTAGCGA---
```

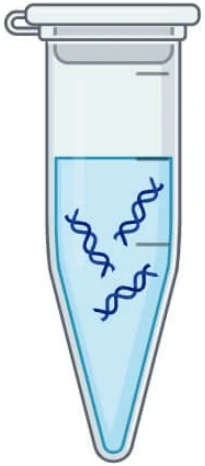
bisulfite treatment
Alkylation
denaturation

```
---AUTUUAUGG---TUUATCGUT---
---TGAGGTGUU---AGGTAGCGA---
```

```
---AUTUUAUGG---TUUATUGUT---
---TGAGGTGUU---AGGTAGCGA---
```

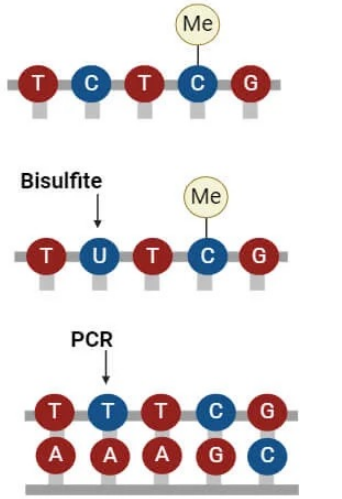
PCR and fragment sequencing to identify
polymorphisms generated by bisulfite treatment.

Methylation Sequencing Steps



DNA Extraction

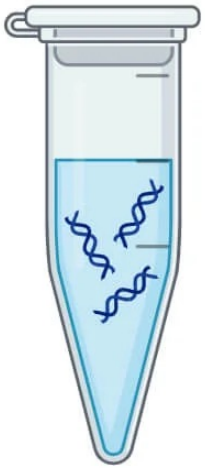
Library Preparation



Bisulfite Conversion

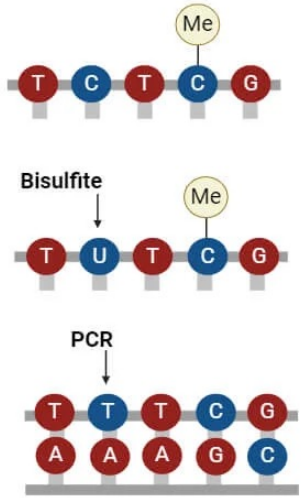
Methylated C remain C while the unmethylated C are converted in U (this will be read as T by the sequencing machine)

Methylation Sequencing Steps



DNA Extraction

Library Preparation



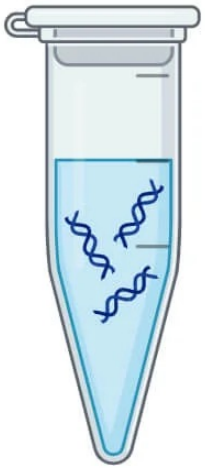
Bisulfite Conversion

PCR Amplification

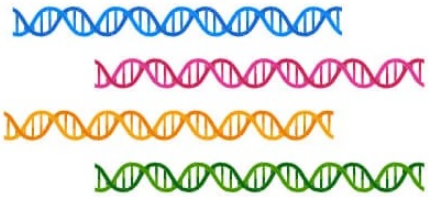
Sequencing

Data Analysis

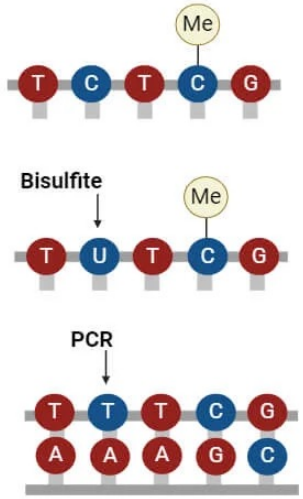
Methylation Sequencing Steps



DNA Extraction



Library Preparation



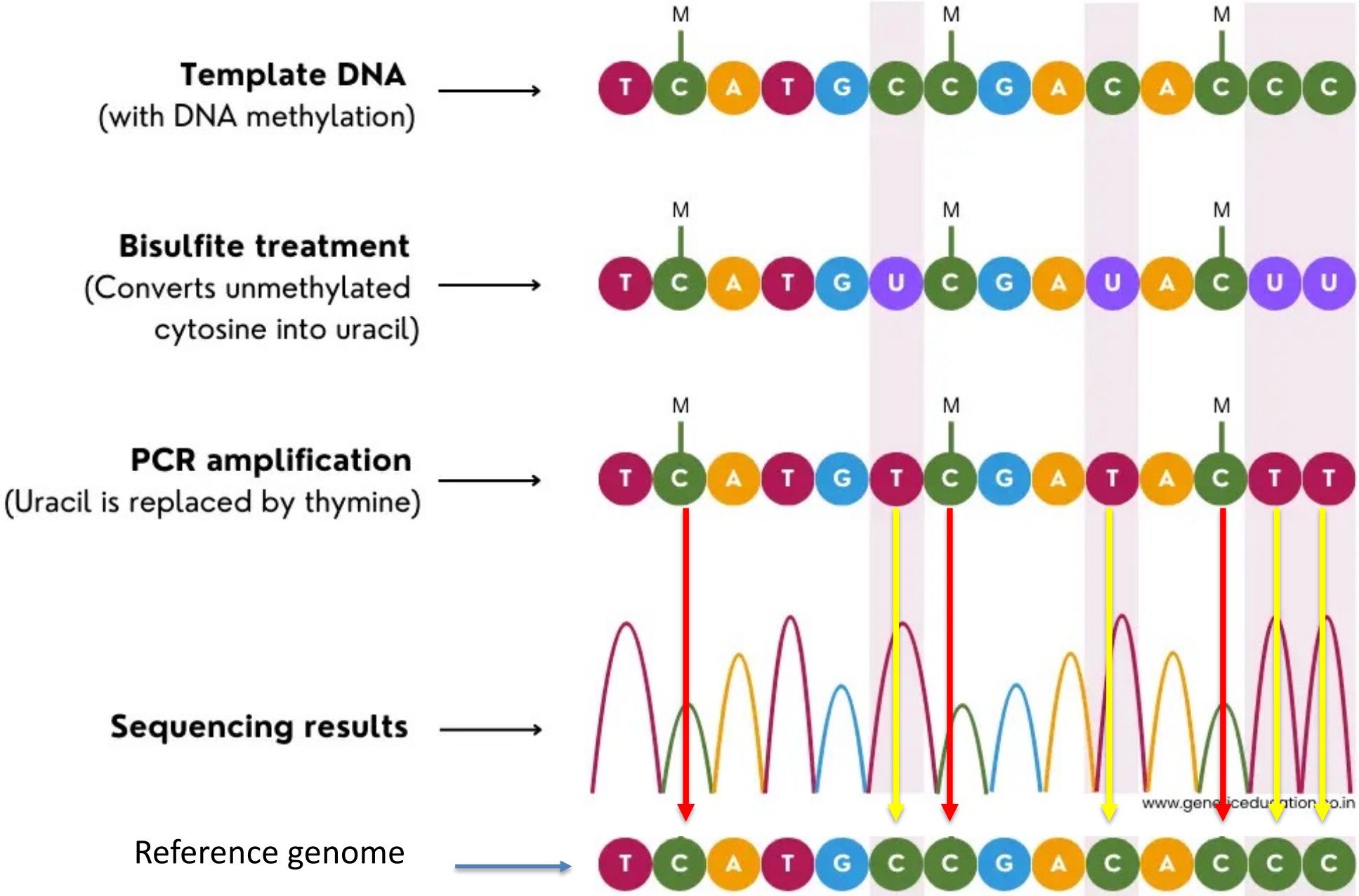
Bisulfite Conversion

Data Analysis

Sequences will be aligned on reference genome. C-T substitution means that the C was not methylated, conversely all the other kept Cs were originally methylated

Treated DNA ...GAUATGCGA...GATGTGTGACG...

Reference Genome ..GAUATGCGA...GACGTGCGACG...


Template DNA
(with DNA methylation)
Bisulfite treatment
(Converts unmethylated
cytosine into uracil)
PCR amplification
(Uracil is replaced by thymine)
Sequencing results
Reference genome
M
T C A T G C C G A C A C C C
T C A T G U C G A U A C U U
T C A T G T C G A T A C T T
www.genomiceducation.co.in