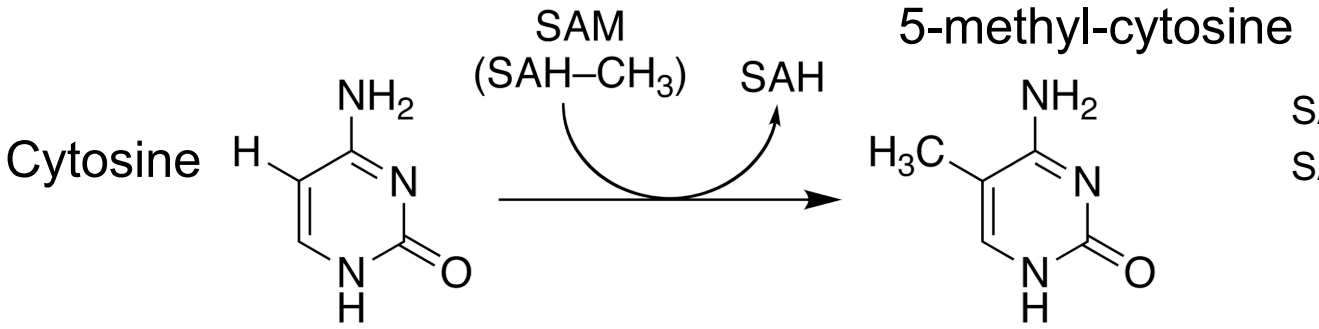
Cytosine methylation

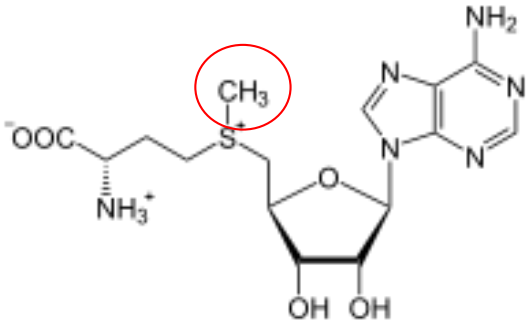
Observation: 3-6 % of all cytosines are methylated in human DNA.

This methylation occurs (almost) exclusively when cytosine is followed by a guanine base -> **CpG dinucleotide**.



SAM: S-adenosyl-methionine

SAH: S-adenosyl-homocysteine



As most CpGs serve as targets of DNA methyltransferases, about 70 - 80% of them are usually methylated.

BUT mammalian genomes contain much fewer (only 20-25 %) of the CpG dinucleotide than is expected by the G+C content (we expect 1/16 ≈ 6% for any random dinucleotide).

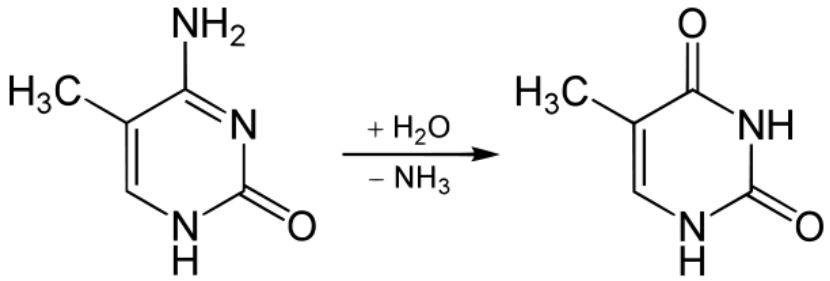
This is typically explained in the following way:

…. (see following page)

Esteller, Nat. Rev. Gen. 8, 286 (2007)
www.wikipedia.org

Cytosine methylation

5-Methylcytosine can easily **deaminate** to **thymine**.

5-methyl-cytosine 

thymine

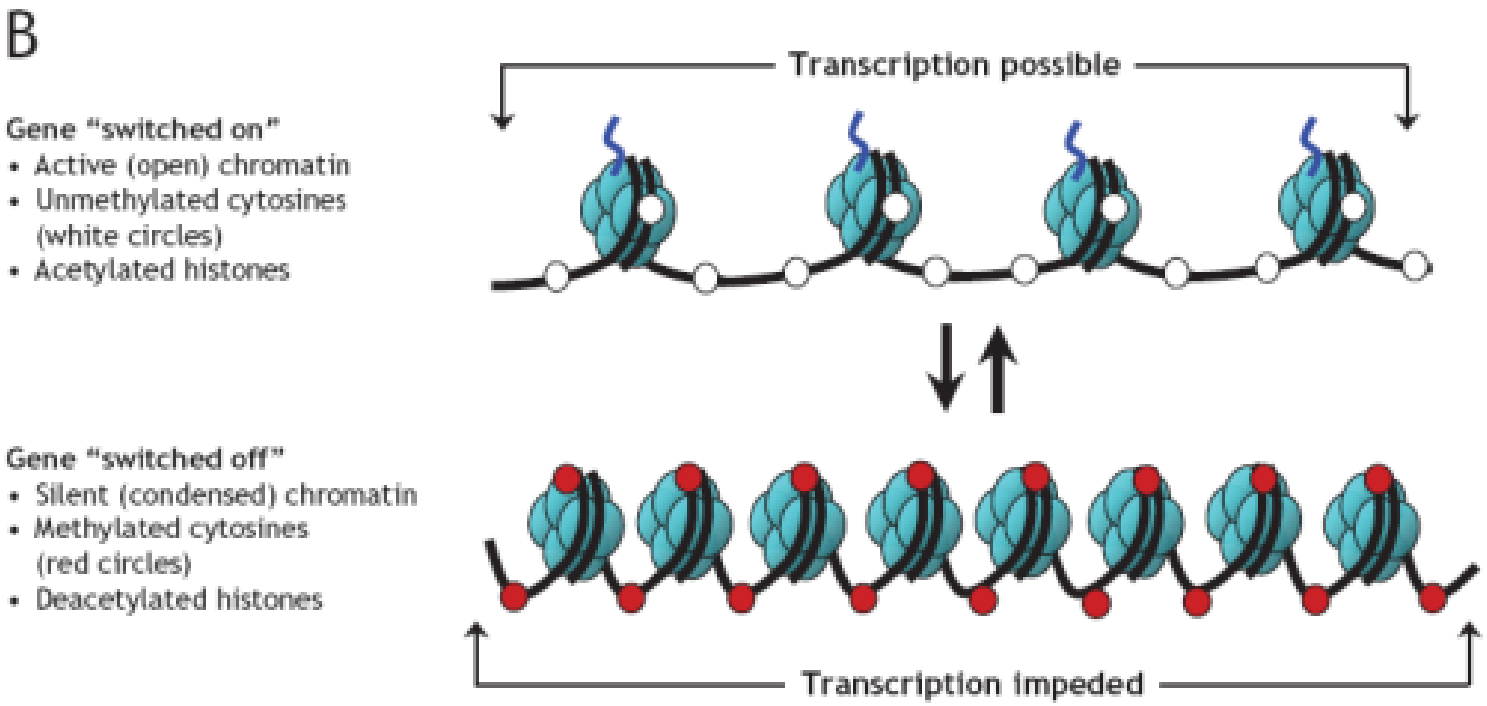
If this mutation is not repaired, the affected CpG is permanently converted to TpG (or CpA if the transition occurs on the reverse DNA strand).

Hence, methylCpGs represent **mutational hot spots** in the genome.
If such mutations occur in the germ line, they become heritable.

A constant loss of CpGs over thousands of generations
can explain the low frequency of this
special dinucleotide in the genomes of human and mouse.

Esteller, Nat. Rev. Gen. 8, 286 (2007)
www.wikipedia.org

chromatin organization affects gene expression



Schematic of the reversible changes in chromatin organization that influence gene expression:

genes are expressed (switched on) when the chromatin is **open** (active), and they are inactivated (switched off) when the chromatin is **condensed** (silent).

White circles = unmethylated cytosines;

red circles = methylated cytosines.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

DNA methylation

Typically, unmethylated clusters of CpG pairs are located in
tissue-specific genes and in essential **housekeeping genes.**

(House-keeping genes are involved in routine maintenance roles and are expressed in most tissues.)

These clusters, or **CpG islands**, are targets for proteins
that bind to unmethylated CpGs and initiate gene transcription.

In contrast, **methylated CpGs** are generally associated with silent DNA,
can block methylation-sensitive proteins and can be easily mutated.

The loss of normal DNA methylation patterns is the
best understood epigenetic cause of disease.

In animal experiments, the removal of genes that encode DNMTs is lethal;
in humans, overexpression of these enzymes has been linked
to a variety of cancers.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

CpG islands

CpG islands are characterized by a high density of CpG dinucleotides that can be targeted by DNA methylation.

CpG islands are regulatory elements and are often located in the promoter region of genes.

Criteria to define CpG islands:

Gardiner-Garden and Frommer: ≥ 200 bp length,
G + C ≥ 50%
$CpG_{obs}/CpG_{exp} \geq 0.6$

Takai and Jones: ≥ 500 bp length
G + C ≥ 55%
$CpG_{obs}/CpG_{exp} \geq 0.65$.

Hutter, Helms, Paulsen, Genomics 88, 323 (2006)

Beta-values measure fractional DNA methylation levels

After analysis of raw sequencing data + filtering of problematic regions etc

the degree of methylation is typically expressed as

fractional **beta value:** %mCG(i) / (%mCG(i) + %CG(i))

A beta value for CpG position *i* takes on values between

0 (position *i* not methylated) and 1 (position *i* fully methylated)

Methylation levels of neighboring sites are correlated

- Observation: methylation levels of neighboring CpG positions within 1000 bp are often correlated;
- distance between neighboring CpGs is ca. 100 bp (1% frequency)

- Idea: exploit this effect to „smoothen“ experimental data, e.g. when this is obtained at low coverage

Master thesis of Junfang Chen (February 2014):

Journal of Bioinformatics and Computational Biology
Vol. 12, No. 6 (2014) 1442005 (16 pages)
© Imperial College Press
DOI: 10.1142/S0219720014420050

Imperial College Press
www.icpress.co.uk

AKSmooth: Enhancing low-coverage bisulfite sequencing data via kernel-based smoothing

Junfang Chen*,†,‡, Pavlo Lutsik†, Ruslan Akulenko*,
Jörn Walter†,§ and Volkhard Helms*,§

*Center for Bioinformatics, Saarland University
Saarbrücken 66123, Germany

†Department of Genetics, Saarland University
Saarbrücken 66123, Germany
‡s9juchen@stud.uni-saarland.de