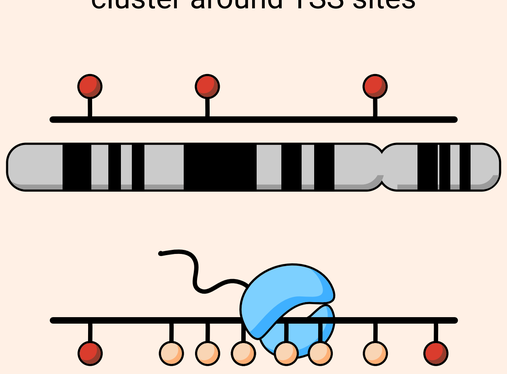
Role of DNA methylation in somatic mutagenesis of human tumors

David Mas-Ponte[1] and Fran Supek[1,2]

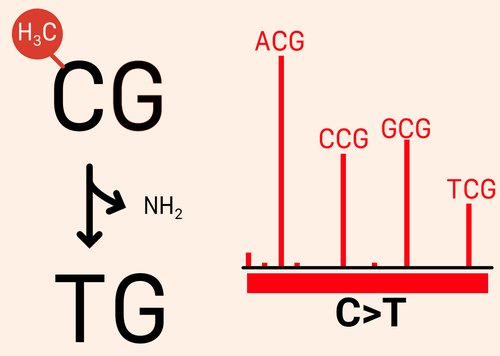
(1) Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology
(2) Catalan Institution for Research and Advanced Studies (ICREA)

@davidmasp

CpG DNA dinucleotides are sparse and typically methylated but cluster around TSS sites

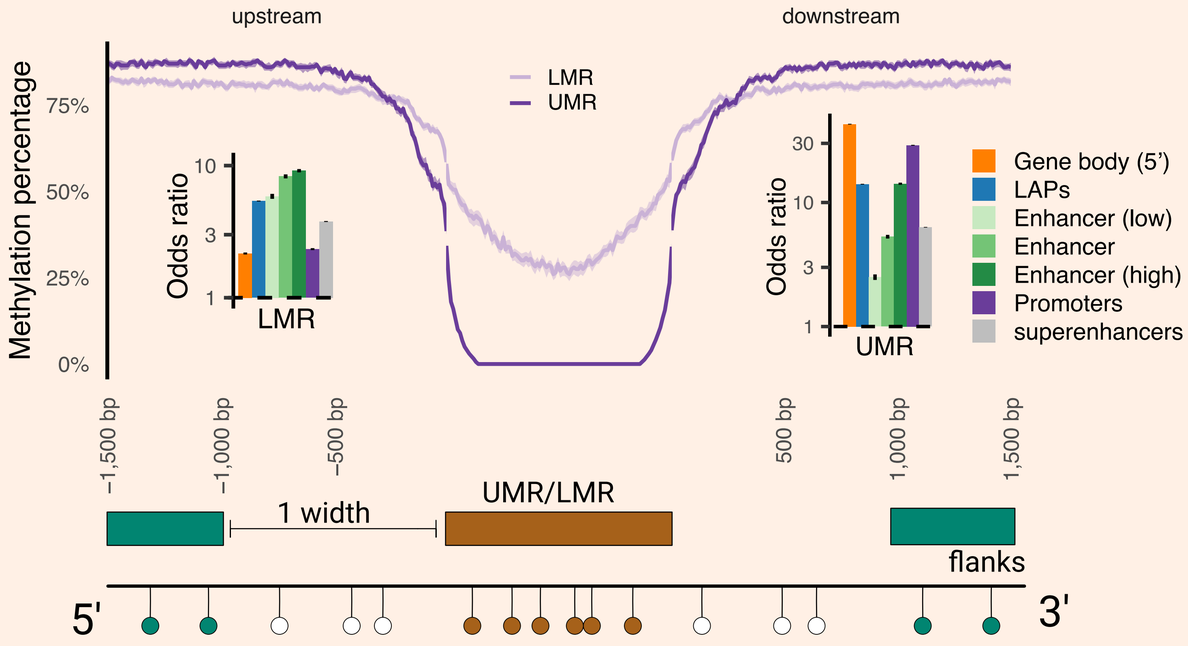


The methylated cytosine is less stable and often deaminates to thymine



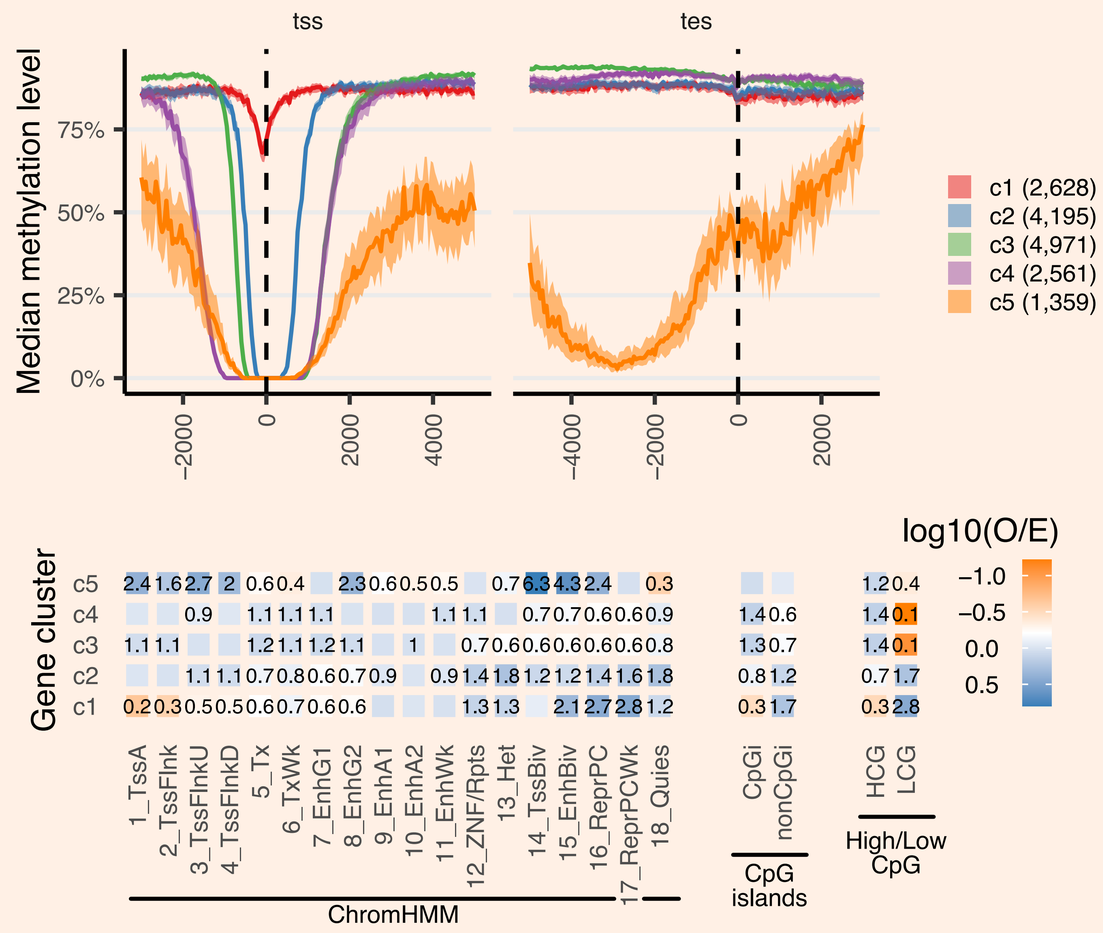
CpG methylation along the human genome

In humans, CpG dinucleotides in DNA are usually methylated at the cytosine nucleobase and have, globally, a low frequency in the genome. However, they are particularly enriched near transcription start sites (TSS) and other functional elements (A). These accumulations play an important role in the regulation of the adjacent gene where they are located[1]. When methylated, these accumulations block the binding of transcription factors switching off transcription. According to their methylated status, the genome can be segmented into unmethylated (UMR), low-methylated (LMR) and fully methylated regions, or the rest of the genome. While the UMRs are specifically associated to the promoter regions of genes, LMRs are more intergenic and enriched in enhancers[2,3] (B).



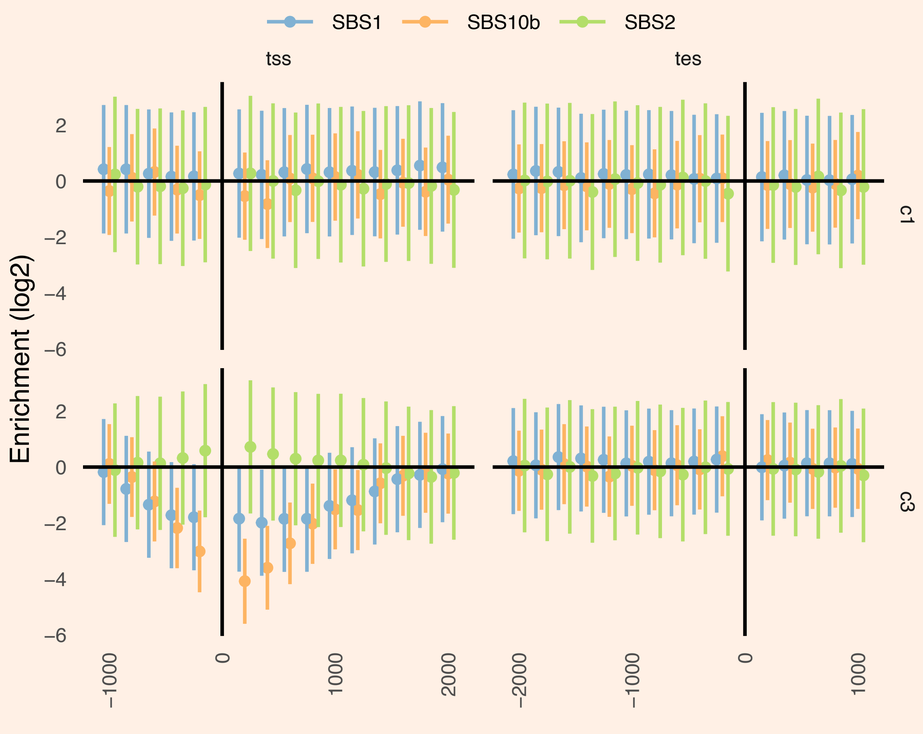
Stratification of genes according to their methylation profiles

Due to the known interaction between UMRs and transcription, we ought to explore if genes could potentially show distinct methylation accumulation profiles. We averaged methylation data from whole genome bisulfite sequencing (WGBS) in XXbp segments around 8 kilobases at both gene ends. We then run a principal component analysis (PCA) to extract the consistent variation and cluster individual genes in 5 clusters (D). Each group was associated with different transcriptional or chromatin states. C1 and C2 were composed by repressed genes with a fully methylated TSS or short hypomethylation window and with a depletion for active marks. C3 and C4 contained active genes with a wide hypomethylation. Finally, C5 is formed by short genes which contain the polycomb mark and present bivalent transcription (E).



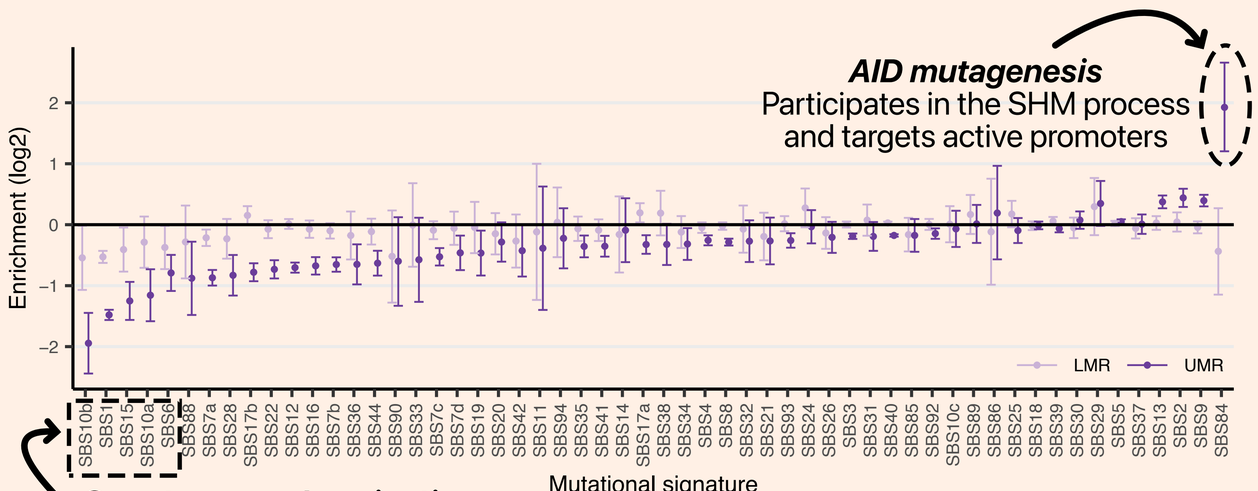
Mutation rate variability along the gene body

Consistent with our analysis, mutation rates along gene bodies also presented a substantial variation, particularly for gene clusters with wider hypomethylation, like C3, and for signatures associated with DNA methylation at UMRs, like SBS1, 10b and 2 (F). This change was not seen in gene clusters where the promoter remained methylated.



Mutation rate changes at UMRs and LMRs

First, we quantify the role of DNA methylation in modulating the mutation rate at both UMRs and LMRs for each extracted mutational signature in the PCAWG dataset. We see that mutational signatures attributed to DNA methylation4,5 (SBS1, SBS10b and SBS15) exhibit a depletion in UMRs and LMRs. Surprisingly, some signatures also show a relative increase in mutation rate with methylation: SBS84 and SBS9, associated with the somatic hypermutation (SHM) in lymphocytes and SBS2 and SBS13 associated with APOBEC mutagenesis.



Spontaneous deamination and replication errors

Mutation signatures associated with the spontaneous deamination of the 5mC and the introduction of uncorrected replication errors at methylated sites due to lack of either proofreading activity in the DNA polymerase or MMR.

Sequencing artifacts signatures and SBS10d SBS23 SBS87, excluded

This study highlights the role of DNA methylation in the modulation of mutation rates in tumor samples, particularly, around unmethylated regions, such as UMRs and LMRs. Due to their overlap with the 5' end of the TSS, this characteristic hypomethylation also represents an important determinant of the overall mutation burden of genes and other functional elements like enhancers and loop anchors. Due to this effect, the clustering of genes according to their differential methylation profiles or considering the methylation status of a region of interest may complement a better estimation of their baseline mutation rates.

1. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. (2012).
2. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res. (2013).
3. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature (2011).
4. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. Nat. Genet. (2015).
5. Poulos, R. C., Olivier, J. & Wong, J. W. H. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. Nucleic Acids Res. (2017).

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
European Research Council
Established by the European Commission
IRB BARCELONA
INSTITUTE FOR RESEARCH IN BIOMEDICINE

We acknowledge the ministry of science and innovation for funding the presenter (D.M.P., FPI Severo Ochoa grant BES-2017-079820) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (757700).