Effects of cytosine methylation on transcription factor binding sites

Yulia A. Medvedeva, Abdullah Khamis, Ivan V. Kulakovskiy, Wail Ba-Alawi, Md. Shariful I. Bhuyan, Hideya Kawaji, Timo Lassmann, Alister R.R. Forrest, Matthias Herbers, and the FANTOM consortium

King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center, Computer, Electrical and Mathematical Sciences and Engineering Division, Thuwal 23955-6900, Saudi Arabia, ju.medvedeva@gmail.com

Vladimir B. Bajic

King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center, Computer, Electrical and Mathematical Sciences and Engineering Division, Thuwal 23955-6900, Saudi Arabia, vladimir.bajic@kaust.edu.sa

Background: DNA methylation in promoters is strongly linked to downstream gene repression [1]. However, the question remains whether DNA methylation is a cause or consequence of gene repression. In the former case, DNA methylation may affect the affinity of transcription factors (TFs) towards their binding sites (TFBSs). In the latter case, gene repression caused by chromatin modification, is stabilized by DNA methylation. Until now, the above-mentioned scenarios have only been supported by non-systematic evidences and have not been tested for a wide spectrum of TFs.

Methods: To estimate DNA methylation we used ENCODE [2] data obtained by reduced representation bisulfite sequencing (RRBS) [3]; to evaluate genome-wide expression across various cell types, we used FANTOM5 (Forrest et al., *submitted*) data obtained by cap analysis of gene expression (CAGE) [4]. For TFBSs prediction we used the remote dependency model (RDM), which takes into account possible correlation of nucleotides at remote positions within TFBSs and was shown to effectively decrease false positive rate as compared to the widely used position weight matrix (PWM) model.

Results: In this study we found that for 16.6% of cytosines methylation profile and the expression profile of neighboring TSSs show significant negative correlation. We name CpG that correspond to such cytosines “traffic lights”. CpG “traffic lights” are mostly located within CpG island in gene promoters. We hypothesize that if CpG “traffic lights” are not induced by average methylation of a silent promoter, they may affect binding of TFs to their

binding sites and therefore regulate transcription. We observed a strong selection against CpG "traffic lights" within TFBSs, more pronounced for "core" position of the TFBS, supporting the damaging role of CpG "traffic light" for a TFBS. Surprisingly, we found selection to be stronger for repressors then for activators or multifunctional TFs.

Conclusions: In this work we suggest that single cytosine methylation may play a role in transcriptional regulation. In a way, this puts into a different perspective the current common perception of the link of methylation and gene expression. Our results allow us to suggest that blocking of TFBS by selective methylation is likely to be restricted to special cases and cannot be considered as a general regulatory mechanism of methylation-dependant transcription.

This work is part of the FANTOM5 project. Data downloads, genomic tools and co-published manuscripts are summarized here http://fantom.gsc.riken.jp/5/.

FANTOM5 was made possible by a Research Grant for RIKEN Omics Science Center from MEXT to Yoshihide Hayashizaki and a Grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT, Japan to YH. We would like to thank all members of the FANTOM5 consortium for contributing to generation of samples and analysis of the data-set and thank GeNAS for data production. YAM, AK, WBA, MdSIB and VBB were supported by the base research funds of VBB at KAUST. KIV was supported by Dynasty Foundation and Russian Foundation for Basic Research, 12-04-32082-mol_a.

1. Bird, A.P., *DNA methylation versus gene expression.* J Embryol Exp Morphol, 1984. **83 Suppl**: p. 31-40.
2. Dunham, I., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
3. Meissner, A., et al., *Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.* Nucleic Acids Res, 2005. **33**(18): p. 5868-77.
4. Salimullah, M., et al., *NanoCAGE: a high-resolution technique to discover and interrogate cell transcriptomes.* Cold Spring Harb Protoc, 2011. **2011**(1): p. pdb prot5559