An in-silico meta-analysis of cell-type-specific non-CpG methylation

Nowadays, the amount of raw data required for a genome-wide analysis is quite large and it is often difficult for even its original author(s) to spend enough time for its thorough analysis. Further, it would be practically even more difficult to perform meta-analysis, i.e., analysis of combined data from different sources, for such ‘big data’. Thus, there still seem to remain many hidden treasures in the forest of raw data repositories. Here we report such an example in our recent study.

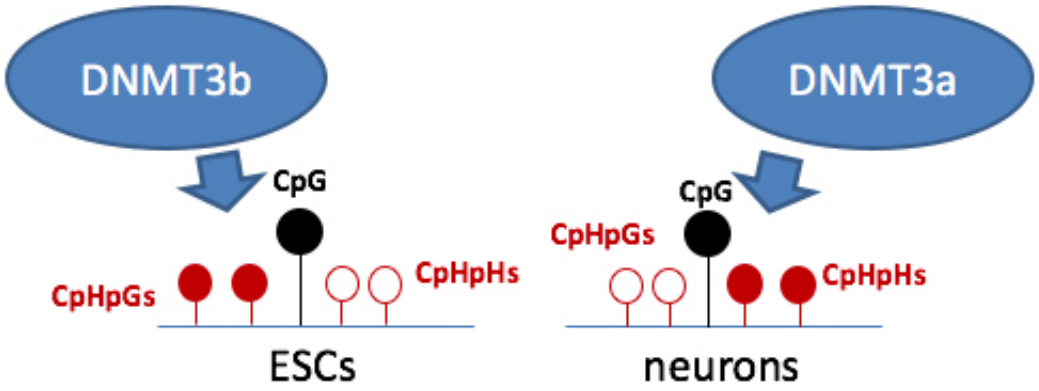

Fig. 1. Differential contribution of DNMTs on mCpHpH and mCpHpG results in differential distribution and function of mCpHs in ESC and neuron.

To understand how gene expression is regulated, it is important to understand how epigenetic information, i.e., the information that is not directly encoded in genome sequences, is involved in the entire mechanism. Several types of epigenetic information are known: DNA methylation, various modifications of histone tails, and various types of chromatin structures, such as the TAD (topologically associating domain) structure. Amongst them, DNA methylation occurs usually on the cytosine nucleotide in the CG dinucleotide, which is conventionally denoted as CpG. It is known that extensively methylated DNA regions are transcriptionally inactive and such regions can be genome-widely detected with several techniques, such as the whole genome bisulfite sequencing (WGBS). Consequently, many WGBS data in various cell types and/or developmental stages have been produced and stored in public databases.

Interestingly, it is also known that the methylation of DNA does not occur exclusively in the CpG context; such (exceptional) non-CpG-type methylations are called CpH methylation because H

means A, T, or C in the IUPAC code. Importantly, the CpH methylations are observed in limited cell-types, such as embryonic stem cells (ESCs) and neurons. Although it is known that the enzymes that cause CpH methylation also cause CpG methylation, several lines of evidences suggest that CpH methylations play some specific roles by themselves. However, their detailed mechanisms and roles are still not clear.

By combining several WGBS data of several cell types as well as their related data, such as the transcriptome data in these cells types, we performed a rather comprehensive meta-analysis of CpH methylation in (embryonic) stem cells and neurons. Then, we found that the nature of CpH methylation is rather different between ESCs and neurons: in ESCs, a methylation enzyme DNA methyltransferase 3b (DNMT3b) is preferentially transcribed and it mainly works in the CAG context whereas another enzyme DNMT3a is more transcribed in neurons and prefers the CAC context. Furthermore, in neurons, the methylated CACs are recognized by brain-disease-related proteins, such as MeCP2, while, in ESCs, DNMT3b interacts with a histone mark, H3K36me3, which is prevalent in actively transcribed genes. Our findings would be useful in more detailed understanding the cellular roles of CpH methylation and underlie the importance of further extensive meta-analyses of publically available data.

Jong-Hun Lee, Sung-Joon Park, Kenta Nakai
The Institute of Medical Science, the University of Tokyo, Japan

Publication

Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s.
Lee JH, Park SJ, Nakai K
Sci Rep. 2017 Sep 12