iMethyl: an integrative human DNA methylation variation database

Yuh SHIWA[1,2], Ryohei FURUKAWA[1], Tsuyoshi HACHIYA[1,2], Hideki OHMOMO[1], Mamoru SATOH[1,2], Jiro HITOMI[3], Kenji SOBUE[3], Makoto SASAKI[3] and *Atsushi SHIMIZU[1]. *Corresponding and Poster Presenter

[1] Division of Biomedical Information Analysis, [2] Division of Biobank and Data Management, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan. [3] Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Shiwa, Iwate 028-3694, Japan

Abstract

Locus-specific DNA-methylation (DNAm) signatures in blood cells are known to be associated with exposure to various environmental hazards, such as cigarette smoke and arsenic, and are considered risk factors for various disorders, such as type 2 diabetes, rheumatoid arthritis, cancer, and schizophrenia. Recent studies revealed that some DNAm signatures had a greater impact on disease risks when compared with genetic variants. Accordingly, DNAm signatures are promising biomarkers in the current era of precision medicine. However, few open databases include comprehensive DNAm profiles. Thus, we analyzed inter-individual DNAm variations in the blood cells of healthy subjects and presented the data as an open database.

We performed whole-genome bisulfite sequencing of purified monocytes and $CD4^+$ T cells collected from an apparently healthy Japanese population (n = 109; 54 men and 55 women; age, 35–75 years) and obtained comprehensive DNAm profiles covering ~90% of the CpG sites in the reference human genome. Based upon the DNAm profiles, we estimated the average and standard deviation (SD) of DNAm levels for ~24 million CpG sites, and then calculated histograms for the DNAm levels of each CpG site. In addition, we examined allele frequencies for ~4 million genetic variants and gene expression levels for ~16,000 genes.

We successfully implemented the obtained data pertaining to DNAm, genetic variants, and gene expressions, as an open database, named “iMethyl”, on a UNIX server with CentOS, Apache HTTP Server and JBrowse 1.12.1., which is available on our web site (http://imethyl.iwate-megabank.org). In the iMethyl browser, regions of interest (ROIs) can be specified using the gene symbols, dbSNP rsIDs, and genomic positions. On the ROIs, the average and SD of DNAm levels as well as SNVs and gene expression levels are shown regarding the 2 types of blood cells. By clicking on the bar in the CpG tracks, histograms of the DNAm levels for each CpG site appear in the pop-up window.

We provided an open database of reference DNAm level distributions for ~24 million CpG sites in two types of blood cells, which may help in the design of assay probes and identification of novel DNAm biomarkers.