A Visualization Tool for Computational Analysis of DNA Methylation Level Using Bisulfite Sequencing Data

Hongseok Tae*

Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA, The Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN 47405, USA

Abstract

Methylation of cytosine is a post-synthesis modification that does not affect the primary DNA sequence but greatly influences gene expression level and phenotypes of an organism. As high-throughput sequencing of bisulfite-treated DNA is the most efficient method to identify methylated sites, several tools to map sequencing reads on a reference are available. But tools to visualize and to interpret the methylation level of methylation sites are currently insufficient. Herein, we present a novel tool to visualize the methylation level of CpG sites.

Availability: MethylDraw is available at http://sourceforge.net/p/methyldraw.

Keywords: visualization, DNA methylation, bisulfite sequencing, epigenetics, MethylDraw

Introduction

Epigenetics is a growing research field which aims to investigate non-genetic factors that alter the expression level of an organism's genes and characterize phenotypes without changing DNA sequence (Bird, 2007). DNA methylation is one of the most studied epigenetic modifications and usually occurs in the CG-rich regions. Hypermethylation of cytosine residues at CpG dinucleotides in the promoter regions is often observed at DNA repair genes, metastasis inhibitor genes and tumor suppressor genes in cancer (Esteller *et al.*, 2001). By identifying aberrant methylation hot-spots in cancer cells, we can target these genes for therapeutic intervention and develop DNA methylation biomarkers for early detection, diagnosis, prognosis, and monitoring response to therapy.

High-throughput bisulfite sequencing is the current gold standard in analysis of DNA methylation patterns in a genome (Cokus *et al.*, 2008). Bisulfite treatment converts unmethylated forms of cytosine to uracil and resequencing identifies the methylated sites by comparing the bases of sequencing reads to those of a reference sequence after sequencing reads are mapped to the reference sequence by mapping programs including BSMAP (Xi and Li, 2009) and RMAP (Smith *et al.*, 2009). Several visualization programs including Methtool (Grunau *et al.*, 2000) and BiQAnalyzer (Bock *et al.*, 2005) are available to show the methylation level of CpG sites, but they do not provide a visualization interface to compare the methylation level of multiple genes and multiple conditions at the same time. Herein, we present MethylDraw which is a tool to draw images visualizing the methylation level of CpG sites for several samples treated in different conditions.

Features and Results

MethylDraw is composed of a set of PERL script programs and runs in a command line environment. It requires two different types of input files. The first type contains only CpG positions of genes in a reference genome. The second type contains the total number of sequencing reads and the number of methylated reads at each position for each condition, which is a generalized format to accommodate diverse outputs of mapping programs. It reads multiple input files for different conditions to support a genome association study which is widely used in polymorphism studies (Hong *et al.*, 2010), then creates HTML outputs which are easily accessible using a web browser.

The HTML files generated by MethylDraw consist of four different categories: average view, gene view, condition view and full view (Fig. 1). The average view shows an average methylation level per gene per condition in a single page. The gene view and the condition view show the methylation level of CpG sites for only a selected gene and a selected condition, respectively. And the full view shows the methylation level of CpG sites for all genes and conditions in a single page. Coloring of each cell on an image, which corresponds to a CpG site, is determined by the degree of two properties, coverage and methylation level, which are described by two bars on the left top parts of images in the HTML files. The color intensity represents the level of coverage and the height of the dark represents the de-
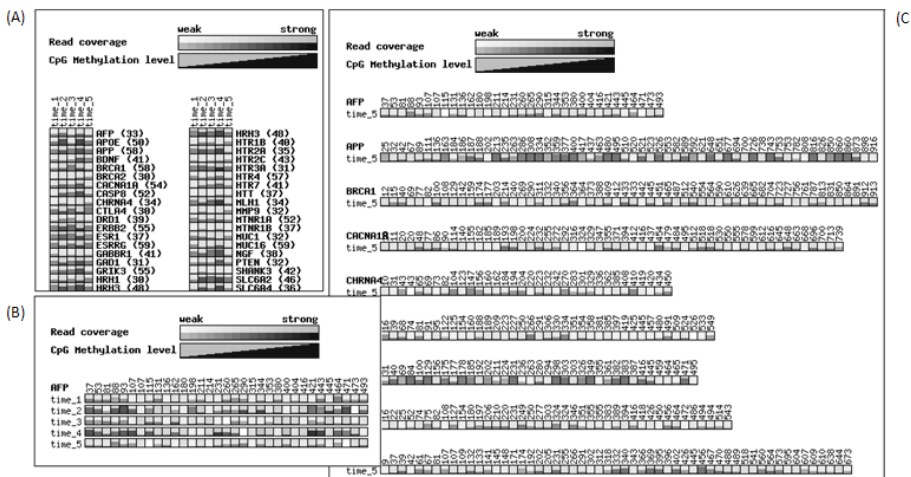
Fig. 1. Output images of Methyl-Draw. (A) The average view shows an average methylation level per gene per condition. (B) The gene view shows the methylation level of a selected gene for all conditions. (C) The condition view shows the methylation level of all genes for a given condition.

gree of methylation.

Discussion

We have developed a tool to visualize the methylation level estimated from high-throughput bisulfite sequencing results. MethylDraw supports a sequencing analysis pipeline using diverse tools by taking generalized input files, running on a command line environment and producing HTML files and image files. In the future, it will have a web interface to accept user inputs through a web browser.

Acknowledgements

References

Bird, A. (2007). Perceptions of epigenetics. *Nature* 447, 396-398.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* 21, 4067-4068.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215-219.

Esteller, M., Corn, P.G., Baylin, S.B., and Herman, J.G. (2001). A gene hypermethylation profile of human cancer. *Cancer Res.* 61, 3225-3229.

Grunau, C., Schattevoy, R., Mache, N., and Rosenthal, A. (2000). MethTools--a toolbox to visualize and analyze DNA methylation data. *Nucleic Acids Res.* 28, 1053-1058.

Hong, K.W., Kim, H.L., and Oh, B.S. (2010). Genome-Wide Association Studies of the Korea Association REsource (KARE). Consortium. *Genomics & Informatics* 8, 101-102.

Smith, A.D., Chung, W.Y., Hodges, E., Kendall, J., Hannon, G., Hicks, J., Xuan, Z., and Zhang, M.Q. (2009). Updates to the RMAP short-read mapping software. *Bioinformatics* 25, 2841-2842.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10, 232.