Bisulfighter: accurate detection of methylated cytosines and differentially methylated regions

Yutaka Saito 1,2, Junko Tsuji 3, Toutai Mituyama 1,2

1 Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Japan

2 Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Japan

3 Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, USA

Abstract

Analysis of bisulfite sequencing data usually requires two tasks: to call methylated cytosines (mCs) in a sample, and to detect differentially methylated regions (DMRs)between paired samples. Although numerous tools have been proposed for mC calling, methods for DMR detection have been largely limited. Furthermore, most previous studies have not provided experiments to benchmark mC calling and DMR detection in a systematic manner.

Here, we present Bisulfighter, a new software package for detecting mCs and DMRs from bisulfite sequencing data. Bisulfighter combines the LAST alignment tool for mC calling, and a novel framework for DMR detection based on hidden Markov models (HMMs). Unlike previous attempts that depend on empirical parameters, Bisulfighter can utilize the expectation-maximization algorithm for HMMs to adjust parameters for each dataset. We conduct extensive experiments in which accuracy of mC calling and DMR detection is evaluated on simulated data with various mC contexts, read qualities, sequencing depths, and DMR lengths, as well as on real data from a wide range of biological processes including pathogenesis and normal development. We demonstrate that Bisulfighter consistently achieves better accuracy than other published tools, providing greater sensitivity for mCs with fewer false positives, more precise estimates of mC levels, more exact locations of DMRs, and better agreement of DMRs with gene expression and DNase I hypersensitivity.

The source code is available at http://epigenome.cbrc.jp/bisulfighter.