

AKSmooth: Enhancing low-coverage bisulfite sequencing data via kernel-based smoothing

Junfang Chen 1,2, Pavlo Lutsik 2, Ruslan Akulenko 1, Jörn Walter 2 and Volkhard Helms 1

1 Center for Bioinformatics, Saarland University, Germany

2 Department of Genetics, Saarland University, Germany

Abstract

Whole-genome bisulfite sequencing (WGBS) is an approach of growing importance. It is the only approach that provides a comprehensive picture of the genome-wide DNA methylation profile. However, obtaining a sufficient amount of genome and read coverage typically requires high sequencing costs. Bioinformatics tools can reduce this cost burden by improving the quality of sequencing data. We have developed a statistical method AKSmooth that can accurately and efficiently reconstruct the single CpG methylation estimate across the entire methylome using low-coverage bisulfite sequencing (Bi-Seq) data. We demonstrate the AKSmooth performance on the low-coverage ($\sim 4\times$) DNA methylation profiles of three human colon cancer samples and matched controls. Under the best set of parameters, AKSmooth-curated data showed high concordance with the gold standard high-coverage sample (Pearson 0.90), outperforming the popular analogous method. In addition, AKSmooth showed computational efficiency with runtime benchmark over 4.5 times better than the reference tool. To summarize, AKSmooth is a simple and efficient tool that can provide an accurate human colon methylome estimation profile from low-coverage WGBS data. The proposed method is implemented in R and is available at <https://github.com/Junfang/AKSmooth>.