

EpiStatProfiler: a novel workflow for the qualitative analysis of DNA methylation

lunedì 28 febbraio 2022 19:15 (10)

Motivation: DNA methylation is an epigenetic modification, primarily occurring at CpG sites, that is involved in major biological mechanisms, such as the regulation of gene expression and the genome stability. Typically, association studies based on this modification are focused on the identification of genomic regions whose average DNA methylation differs among distinct conditions. However, studying the methylation status of cytosines at single-molecule level can provide additional insights about the cell-to-cell heterogeneity and the cell clonality within a sample. In this context, all the different combinations of CpGs methylation states that can be observed in a given locus are defined as epialleles. Several bioinformatic tools have been developed to extract epiallelic information from bisulfite sequencing data. Nevertheless, these tools have some limitations on the selection of the regions that can be profiled (e.g., number of CpG sites) and they do not provide support on the availability of dedicated statistical tests on the epiallele compositions derived from their output.

Methods: Here we present a novel workflow that can be used to retrieve epiallelic profiles from bisulfite sequencing data. In particular our workflow allows: data loading and filtering, regions design, and epialleles extraction. Dedicated statistical tests can then be used to identify regions that differ among groups based on their epiallelic composition.

Results: We developed EpiStatProfiler, a new R-package providing a library of functions that can be used to extract and summarise epialleles from any type of bisulfite sequencing data and to perform downstream statistical comparisons among different groups. The tool is intended to enable a customized selection of target regions, according to a set of user-defined parameters (minimum coverage, number of cytosines, minimum window size). Furthermore, it is also possible to analyse strand specific and non-CG methylation. Epialleles information is stored by EpiStatProfiler in two different outputs: a compressed 0-1 matrix containing the epialleles composition for each analysed region and an additional output containing basic features and multiple metrics derived from the profiled regions. EpiStatProfiler provides a set of functions to perform epiallele-based comparisons in longitudinal and cross-sectional studies. We believe that this package could represent a valuable tool to qualitatively analyse the methylation heterogeneity in a variety of systems, such as tumor evolution, cell differentiation and disease conditions.

Department

Department of Biology, Department of Molecular Medicine and Medical Biotechnology

Primary author(s): Mrs. SARNATARO, Antonella (Molecular Medicine and Medical Biotechnology, University Federico II); SCALA, Giovanni (Department of Biology, University Federico II)

Co-author(s): DE RISO, Giulia (University of Naples Federico II); COCOZZA, Sergio (University Federico II); PEZONE, Antonio (University of Naples Federico II); AMENTE, Stefano (University Federico II)

Presenter(s): Mrs. SARNATARO, Antonella (Molecular Medicine and Medical Biotechnology, University Federico II)