

Development a simple computational pipeline for processing Illumina methylation data

Lopatenko M.V.

*Institute of Biology and Biomedicine, Lobachevsky State University of Nizhni Novgorod,
Nizhni Novgorod, Russia
fasdan1999@gmail.com*

Key words: bioinformatics pipeline; epigenetics

In the context of continuous growth of DNA methylation data due to requests in gerontology and oncology, there is a need to improve the methods of their processing. At the same time, there is a demand for a simple and convenient solution that aggregates different software tools to increase the variability and accuracy of the analytical process [1]. Thus, we present a DNA methylation data processing pipeline that does not require a high threshold of entry, making it accessible to a wide range of scientific community. In this work, the developed pipeline was applied on the IDAT file (Illumina arrays raw data) from Illumina's Infinium arrays (27k, 450k, and EPIC), obtained by accessing the Zenodo database (<https://zenodo.org/>).

The pipeline itself included algorithms and libraries in R and Python for error correction, data normalization and identification of sites with differential methylation and had the following steps: 1) preparation of methylation IDAT files; 2) quality control (QC) and filtering analysis; 3) analysis of processed data and interpretation.

The obtained result can facilitate the construction of age calculation models for epigenetic clocks by integrating the developed pipeline with machine learning algorithms. Moreover, the pipeline is very flexible and can be modified to work with bisulfite sequencing (BS-seq) data, including their quality control, alignment, methylation calling, analysis and visualization of results. As a result, it will significantly increase the functionality and quality of the processed information.

References

1. Dafni A., Piferrer F. Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. *Front Marine Sci.* 2023;10:1096909