St. Jude Children's Research Hospital®

St. Jude BioHackathon

Title

Developing a methylation array analysis pipeline

Category

Processing Pipelines And Methods

Challenge

Currently, no good pipeline exists for analyzing methylation array data, particularly differentially methylated regions (DMRs). The ideal program would take raw idat files, perform quality control and normalization in minfi R with adjustable parameters, incorporate sex chromosome analysis, use window analysis to perform both outlier and case vs. control DMR analysis, find DMRs, plot DMRs, and annotate DMRs (by transcripts, ENCODE regulatory elements, repeat expansions, population frequency data, etc.) with additional functionality to perform clustering/dimensional reduction of methylation values. The ideal DMR plot will be interactive, allowing users to scan the genome (hover computer mouse over plot) and determine which samples have outlier methylation at each locus, query gene lists, plot specific samples, and batch export images.

Benefit

This pipeline will expedite discoveries in the realm of DNA methylation with a multitude of applications in neurological disease, cancer, etc. A straight-forward output that could be easily visualized concurrently with other ‘omics datasets would make integrative data exploration much easier.

Helpful Tools, Packages, or Software

Minfi R package, DMR window analysis (https://github.com/AndyMSSMLab/Epivariation-in-23K-samples), findDMR, plotly, ggplot, jbrowse.

Test Data

Any methylation array data (.idat file); PPMI cohort (https://www.ppmi-info.org/access-data-specimens/download-data)