Systems Imagination

About Jia

Jia Zeng, Ph.D., is a biomedical informatics scientist specializing in next-generation sequencing and cancer biomarker discovery. She previously worked as an Application Scientist in Oxford Nanopore Technologies in developing methods to enable the first wave applications of nanopore sequencing devices. Jia received her PhD degree in Biological Design and is currently pursuing a master's degree in Biomedical Informatics.

Systems Imagination, Inc.

The pipelines discussed here are currently implemented at Systems Imagination, Inc. Our multidisciplinary team is more than happy to discuss how our services can align with your needs.

Follow us:

🐦 @systemsimagin

in linkedin.com/company/systems-imagination-inc

References

1. Bock C. Analysing and interpreting DNA methylation data. Nature Reviews Genetics. 2012;13:705-719. http://www.nature.com/nrg/journal/v13/n10/full/nrg3273.html

Systems Imagination

Systems Imagination, Inc.
7730 E. Greenway Suite 205
Scottsdale, AZ 85260

480.422.2731
info@siiventures.com
www.systemsimagination.com

Bioinformatics Perspective

Jia Zeng, PhD

Advanced Genomic Analysis for NGS

Systems Imagination, Inc proudly offers our services to pharmaceutical companies, academic research groups and individual researchers at the intersection of translational genomics and computer science. Our multidisciplinary teams include experienced Knowledge Engineers, Scientists, Bioinformaticians and Mathematicians. Our analytical pipelines are carefully researched so that we only utilize peer-reviewed computational tools with proven performance and reliability.

Recommended Workflows for DNA Methylation Analysis

Quality Control and Alignment: Preprocessing DNA Methylation Data

The data for DNA methylation mapping come from three major sources: bisulfite sequencing, bisulfite microarrays, and enrichment-based sequencing. Different tools are used for processing data, quantifying the absolute and relative DNA methylation and quality control for raw data obtained from different DNA methylation assays[1].

• Bisulfite-sequencing uses bisulfite to induces mutation of unmethylated Cs to Us, which become Ts. Next generation sequencing maps the mutations.

• Bisulfide microarrays uses bisulfite treatment to convert Cs to Us. Genotyping are performed on a microarray that detects a selection of Cs.

• Enrichment-based methods, enrich methylated DNA samples then perform a comparison of the sequences based on experimental conditions.

Visualization and Statistical Analysis

Visualize data in methylation maps and identify differentially methylated regions.

Methylation information containing DNA methylation levels for each CpG in each sample can be extracted from the output sam file. The genome tracks can be visualized in a genome browser to inspect the selected regions or to compare global similarity and differences among samples[1].

To identify differentially methylated regions (DMRs), a suite of DMR detection method through statistical analysis was recommended. They include t-tests, Wilcoxon rank-sum tests, mixture models, Shannon entropy, logistic M values, feature selection and more[1].

Notably, if DNA methylation is tested at a large number of genomic loci, multiple hypothesis testing needs to be corrected 1. False discovery rate (FDR) can be obtained from uncorrected P values.

When the differences in DNA methylation are weak, two methods can be used to increase the statistical power[1]. The comparison can be expanded on larger genomic regions instead of single CpGs, or hierarchical models can be used for dealing with samples with stochastic noises.

Visualization and Interpretation

Verifying and validating differences in DNA methylation and interpreting the differences

Differentially methylated regions need to be confirmed both computationally and biologically. In order to verify the differences, quality control plots including volcano plots, Q-Q plots, and Manhattan plots[1]. Computational tools can be used to assist manual ranking and selecting promising differentially methylated regions for experimental verification.

Interpreting the genome-wide DNA-methylation landscape against annotated genome will help researchers discover biologically relevant trends. Several computational tools exist for assisting interpretation. The tools recommended by Bock's review feature several functions include searching for enrichment of gene functions and regulatory elements, assessing confounding factors in inferring causal relationships and other aspects[1].

Pipeline for DNA Methylation Analysis