Technical Application Series

DNA Methylation:
Microarrays and Bisulfite Sequencing (BS-Seq)

GeCo
Advanced Genomic Consulting Service

Background

The analysis of DNA methylation by microarray or bisulfite sequencing approaches makes it possible to obtain a pan-genomic vision of epigenetic modifications between different samples. GeCo's advanced bioinformatics analysis modules enable the ability to explore DNA methylation data in depth using reference methods. In addition to quickly extracting biological information and identifying groups with homogeneous profiles, our approach can be used to conduct differential methylation analyses, analyze pathways, determine patterns of transcription factors, and perform integrations with the transcriptome.

The available modules for the analysis of DNA methylation data are described below. Deliverables include high-quality figures and tables as well as a detailed material and methods to support rapid publication of the results.

Module 1 - Data Quality Control and Integration by Region of Interest and Global Distribution

This module restricts the data to reliable signals depending on the technology utilized by detecting intensity, identifying p-value for chips, and determining sequencing coverage. The module also explores integration of signals by region of interest (bisulfite-seq) which are, by default, defined in relation to genes (promoter/gene body) and CpG islands (island, shore, shelf) as shown in Figure 1. Additional types of regions (enhancers, chromatin domains, transcription factor binding sites) can also be customized for each project.

Results can be output in terms of quality control metrics (number of CpGs/regions of interest correctly detected in each sample) in the form of figures and tables and as figures which show the distribution of methylation levels for each sample by type of region (Fig. 2)

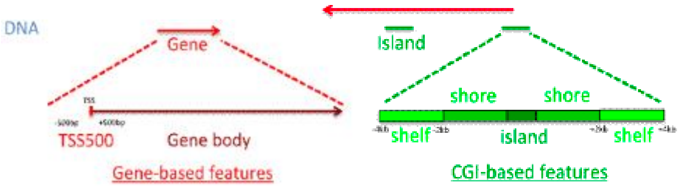

Figure 1. Regions of interest used for integration of bisulfite-seq signals.

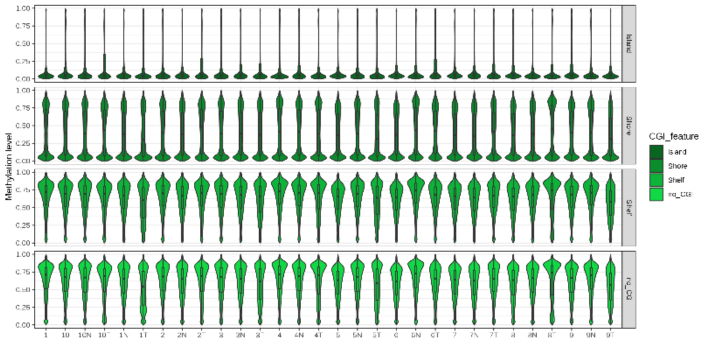

Figure 2. Distribution of methylation levels by type of region (CpG island, shore, shelf and off island) on a series of samples.

Module 2 - Unsupervised Classification of Methylation Profiles

This module enables the selection of the most variable CpGs/regions in the dataset based on standard deviation. The samples are classified on the basis of methyloma using different approaches. This includes a principal component analysis, t-SNE, hierarchical clustering, and consensus clustering (Fig. 3). The results are cross-referenced with annotations provided by the client (or from other omics) to identify significant associations.

Results are provided as figures illustrating the different classifications made (e.g., PCA, clusterings, heatmap), a proposed optimal classification, and tables showing the association of the groups with clinical and molecular annotations

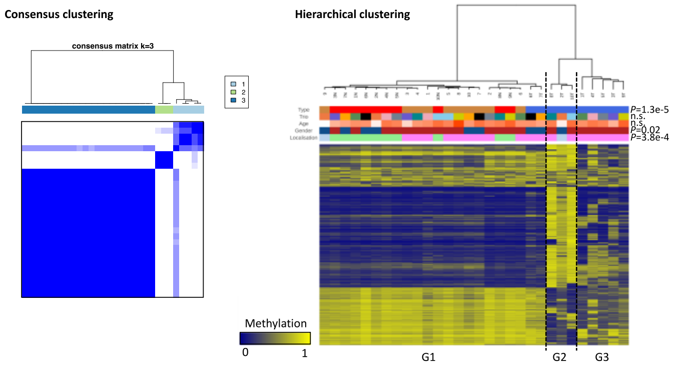

Figure 3. Unsupervised classification of methylation profiles. Consensus clustering (left) makes it possible to identify the most suitable partition (here in 3 groups). The hierarchical clustering and the heatmap (right) allow to visualize the associations to the annotations and the changes of methylation.

Module 3 - Differential Methylation Analysis

This module compares the methylation of each CpG/ region of interest between the groups resulting from the classification and/or groups of interest defined by the customer using methylKit.

The distribution of methylation changes by type of region (e.g., islands, promoters, gene bodies). A pathway enrichment analysis is also performed among hyper- and hypomethylated genes at the promoter or gene body level.

Outputs include tables and summary figures indicating the number of differentially methylated CpGs/regions of interest for each comparison, the CpGs/differentially methylated regions showing the average methylation in each group, the delta, p-value and q-values, and the significantly enriched pathways among differentially methylated genes, and graphical representations including volcano plots, scatterplots, and smooth scatterplots showing the overall changes in methylation and by type of region (Fig. 4)

a) Number of differentially mixed regions

b) Visualization of methylation changes

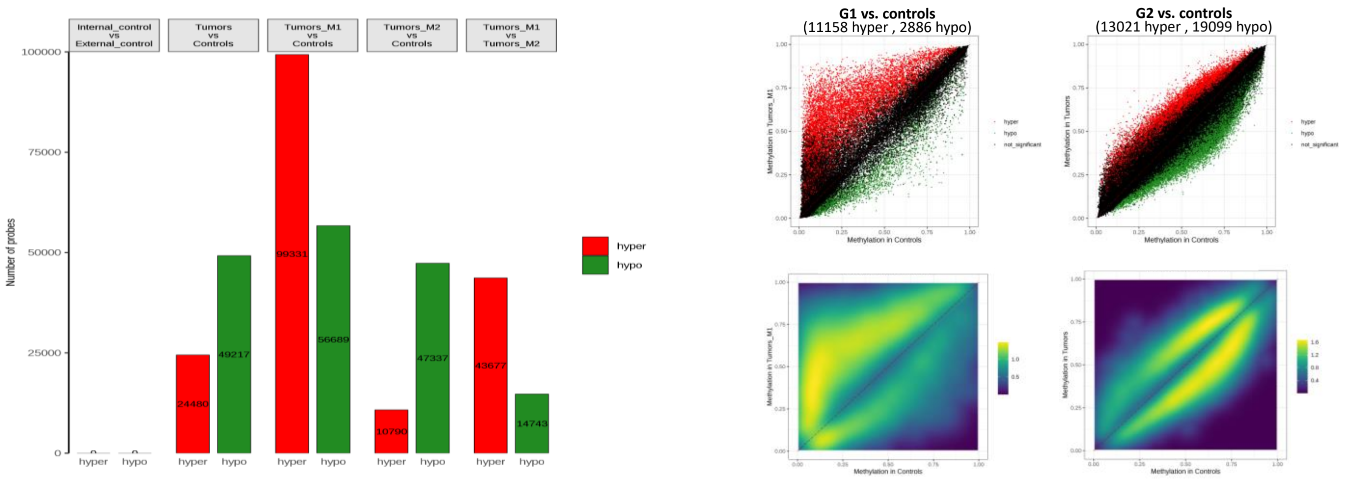

Figure 4. Differential methylation analysis. a) number of hyper- and hypomethylated regions in each comparison performed. b) sample visuals showing methylation changes in each group.

Module 4 - Analysis of Binding Patterns of Transcription Factors

Activation of a transcription factor is frequently accompanied by coordinated hypomethylation of its target sequences. To identify this type of event, this module looks for the binding motifs of enriched transcription factors among the CpGs/differentially methylated regions using the ELMER R/Bioconductor package.

Outputs include tables which show the patterns enriched among CpG and/or hyper methylated or hypomethylated regions in each group (Fig. 5) and - heatmaps representing the average methylation of targets for significant transcription factors.

a) Patterns enriched among hypermethylated regions in G2 group

b) Visualization of the average methylation for target sequences

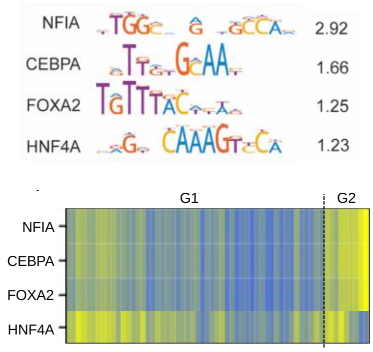

Figure 5. Patterns of enriched transcription factors among differentially methylated regions between the two groups

Module 5 - Methylation and Expression Data Crossing

If transcriptomic data (RNA-Seq) are available for the same series, the module is used to perform an integrated analysis of this data with the methylome. This includes a comparison of the lists of differentially expressed and differentially methylated genes, a pathway enrichment analysis among epigenetically deregulated genes, and the identification of genes whose expression is strongly correlated with the methylation of a neighboring CpG using the ELMER R/Bioconductor package.

This approach makes it possible to identify genes finely regulated by methylation (Fig. 6). Cross-referencing of the expression of transcription factors between groups and link with differentially methylated motifs is also performed, making it possible to robustly identify the transcriptional programs activated in each group.

Results are provided as tables and figure (Venn diagrams) indicating the overlap of differentially expressed and methylated genes, as figures and and tables representing the genes whose expression is coupled to the methylation of a neighboring CpG, and as figures representing the correlation between the expression of deregulated transcription factors and the methylation of their targets.

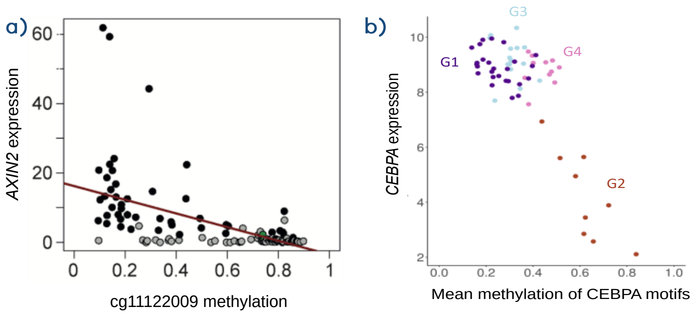

Figure 6. Cross-referencing of methylation and expression data. a) Identification of a CpG strongly linked to the expression of a neighboring gene. The expression of the AXIN2 gene is positively correlated with the methylation of the neighboring CpG cg11122009. b) Underexpression of a transcription factor and hypermethylation of its targets in a subgroup. The underexpression of the CEBPA transcription factor is accompanied by hypermethylation of its targets in the G2 group.

Examples of studies using GeCo modules for DNA methylation:

- Letouzé E, Martinelli C, Loriot C, et al. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*. 2013;23(6):739-752.
- Pilati C, Letouzé E, Nault JC, et al. Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations and the mechanisms of malignant transformation. *Cancer Cell*. 2014;25(4):428-441.
- Morin A, Goncalves J, Moog S, et al. TET-Mediated Hypermethylation Primes SDH-Deficient Cells for HIF2α-Driven Mesenchymal Transition. *Cell Rep*. 2020;30(13):4551-4566.e7.

Advanced Genomic
Consulting Service

A service brought to you by:

To learn more about GeCo or to contact us about your project,
email us at services@integragen.com

© 2020 IntegraGen SA. All rights reserved.