A framework for analysis of DNA methylation data

Flemming S[1,*], Grüning BA[1], Häupl T[2], Günther S[1]

[1] *Department of Pharmaceutical Bioinformatics, Institute for Pharmaceutical Sciences, University of Freiburg, Germany;*
[2] *Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin*
**e-mail: stephan.flemming@pharmazie.uni-freiburg.de*

Introduction

DNA Methylation data analysis may include raw data preprocessing, detection of significant differential methylation sites, network based approaches, or the analysis of correlation with gene expression data. The use of several tools is frequently associated with the problem of incompatible file formats and the need of expert knowledge for each application. Therefore, further advanced analyses are often omitted or cannot be sufficiently adapted to the user's requirements. Furthermore, sharing of workflows, checking of intermediate results, and the reproducibility of complex calculations may be difficult. To address these problems we have applied the workflow management system Galaxy [1]. The capability for data analysis is demonstrated with a publicly available data set.

Methods

We present a comprehensive framework that provides a broad range of functionality for the analysis of DNA methylation data. All tools have been integrated as modules and can be linked together in reusable and shareable workflows.
The framework can be used directly via web browser, without any programming skills. Application of tools and resulting data sets are automatically logged. Thus, the reproducibility is further increased and comparisons of different analysis approaches are facilitated.
Our framework supports the analysis of data exported from GenomeStudio2011 for Illumina Infinium Methylation Assays [2] as well as data stored in the FGEM (final gene expression matrix) file format [3]. We provide inhouse developed modules as well as widely used tools and methods for DNA methylation data evaluation, e.g. several libraries of the Bioconductor package. The following methods have been included.

Input
- File Upload (FGEM format)
- Repositories (NCBI-GEO)

Data Normalisation
- Background Correction
- Quantile Normalisation
- Distribution Fitting
- Regression

Visualisation
- Biplots (PCA)
- Heat Maps
- Scatter Plots
- Histograms

Significant Differential Methylation
- Fold Change
- Linear Model for Log-Ratios
- t-Test
- F-Test
- Bonferroni Correction

Further Analyis
- Correlation
- Clustering
- Principal Component Analysis

Network Analysis
- Comparative Analysis of Correlation Networks
- Modularity

Case Study

Publicly available data based on a high-throughput method using HumanMethylation27 BeadChip technology [4] were analysed. Genome-wide CD4+ T-cell DNA methylation levels of twelve biological replicates (12 lupus patients and 12 controls) were studied [5]. To identify differentially methylated CG sites between lupus patients and controls, we used the associative analysis as described by Dozmorov *et al.* [6]. After obtaining data normalization with a non-linear fitting approach followed by a regression method, methylation sites with a signal exceeding 6 SD above noise level and at least 1.2-fold deviation in methylation between patients and controls were filtered. We observed 250 hypomethylated loci compared to 121 hypermethylated loci in lupus patients, representing 248 and 120 unique genes. A comprehensive analysis of correlation networks [7] shows that the metabolism of glycine, serine and threonine and insulin receptor signalling are affected by significant changes in relationships of the corresponding genes.

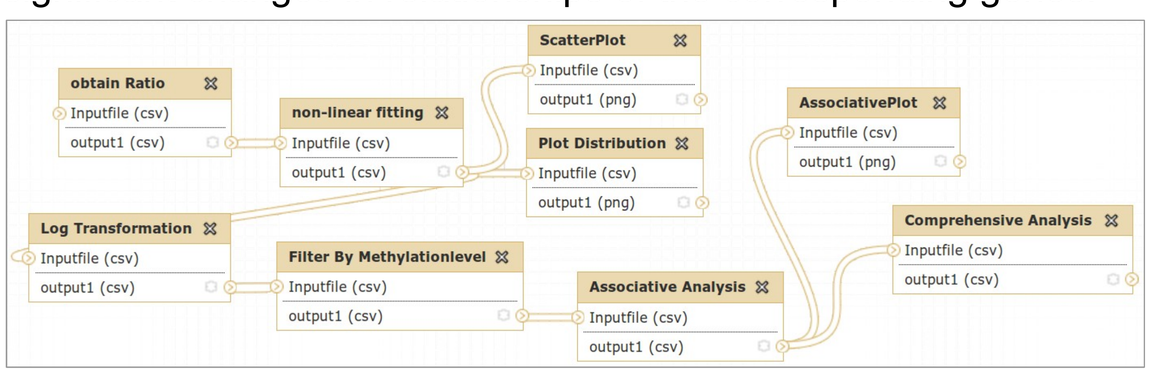

Fig. 1: Workflow of associated methylation analysis as described by Dozmorov *et al.* [6].

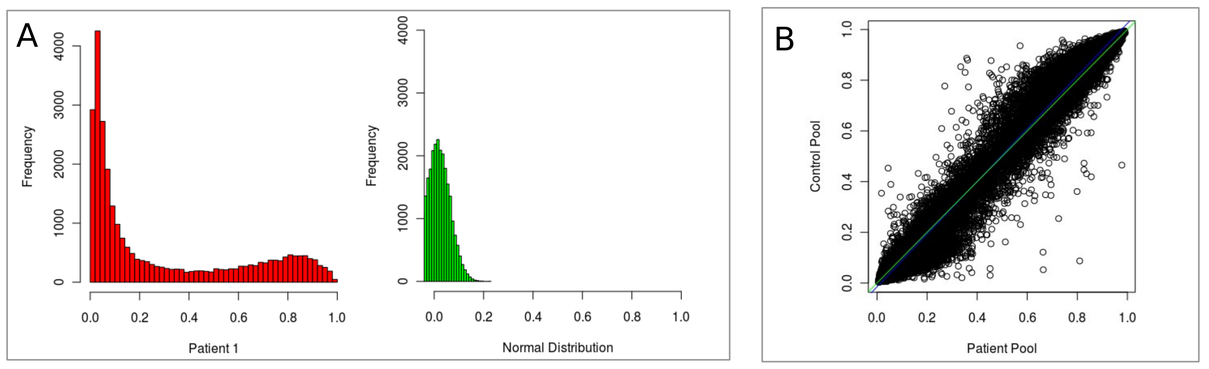

Fig. 2: Applied d*ata normalisation: (A) non-linear fitting, (B) linear regression.*

Future Prospects

For the integration of omics data, the implementation of additional methods is planned, e.g. for correlation analysis of methylation and gene expression data. Furthermore, we plan to increase the number of supported repositories for high-throughput data (e.g. ArrayExpress [8]). Thus, additional methods for raw data processing are required. With regard to our current field of research - dealing with methylation levels in cells of the immune system using Infinium HumanMethylation450 BeadChips [9] - the integration of related publicly available datasets will facilitate various experimental designs. The framework will be available via internet for third parties on request.

References

[1] Blankenberg D, Von Kuster G, Coraor N, *et al*. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010 Jan;Chapter 19:Unit 19.10.1-21.
[2] http://www.illumina.com/software/genomestudio_software.ilmn
[3] http://www.ebi.ac.uk/miamexpress/help/datafile_help.html
[4] http://www.illumina.com/technology/infinium_methylation_assay.ilmn
[5] Jeffries M, Dozmorov M, Tang Y, *et al*. Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus. Epigenetics. 2011 May;6(5):593-601.
[6] Dozmorov I, Lefkovits I. Internal standard-based analysis of microarray data. Part 1: analysis of differential gene expressions. Nucleic Acids Res. 2009 Oct;37(19):6323-39.
[7] Emmert-Streib. The chronic fatigue syndrome: a comparative pathway analysis. J Comput Biol. 2007 Sep;14(7):961-72.
[8] http://www.ebi.ac.uk/arrayexpress/
[9] http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn

Pharmaceutical Bioinformatics
University of Freiburg

The working group of Pharmaceutical Bioinformatics at the Institute for Pharmaceutical Sciences develops algorithms and software for pharmaceutical research. Our fields of research include the modeling of molecular interactions, prediction of biological effects of molecules, identification of potential new drug agents, and analysis of gene expression and methylation data. The working group is part of the University of Freiburg's Research Group Program of the Excellence Initiative of the federal and state governments.

http://www.pharmaceutical-bioinformatics.de/

Bundesministerium für Bildung und Forschung

UNI FREIBURG