Computational Methods for the DEEP Characterization of DNA Methylation BLUEPRINTs

Fabian Müller[1*], Thomas Lengauer[1], Christoph Bock[1,2,3]

[1] *Max Planck Institute for Informatics, Saarbrücken, Germany*
[2] *CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria*
[3] *Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria*
[*] *Contact: fmueller@mpi-inf.mpg.de*

Abstract
DNA methylation is associated with regulatory functions in health and disease. Genome-wide DNA methylation maps are now routinely generated in the context of large epigenome mapping projects such as IHEC.
While methods and computational pipelines for the initial processing of data originating from bisulfite-based sequencing and microarrays have been widely established, tools are scarce for the high-level analysis and interpretation of DNA methylation signatures.

To fill this gap, we developed the RnBeads software package (http://rnbeads.mpi-inf.mpg.de) which is now routinely employed by many laboratories. RnBeads facilitates comprehensive, interpretable analysis of large-scale DNA methylation datasets, implementing start-to-finish analysis by specifying only a few lines of code, while also providing functions and methods for in-depth exploration in the R programming language. The results pertaining to data preprocessing, adjusting for confounding factors, exploratory analyses, identifying and characterizing differential methylation and data export are presented in browsable analysis reports featuring method descriptions, data tables and publication-quality plots.

Furthermore, a framework for the global analysis of epigenetic patterns in repetitive DNA elements such as transposons is currently in development. This software tool allows for the quantification of epigenomic profiles based on a catalogue of reference sequences and provides data structures and methods for the analysis and visualization of epigenetic signatures of repetitive DNA subfamilies across multiple samples.

In order to dissect the DNA methylation dynamics during human blood cell differentiation, we apply the above methods to large collections of whole-genome bisulfite sequencing data generated in the BLUEPRINT and DEEP projects. We obtain a global view on the hematopoietic landscape by studying the methylomes of a large panel of differentiated blood cell types and delineate lineage-specific patterns. Furthermore, we find widespread changes in DNA methylation patterns during memory formation in $CD4^+$ T cells. Analyzing DNA methylation profiles originating from low-input sequencing, we provide a detailed account of lineage-specific methylation signatures in hematopoietic stem cells and early blood progenitors. Using statistical modeling we identify epigenetic signatures of cell identity that can be used to characterize within-cell-type heterogeneity.

Funding:
This work has been partially funded by the EU Grant HEALTH-F5-2011-282510 (BLUEPRINT) and by the BMBF Grant 01KU1216A (DEEP).