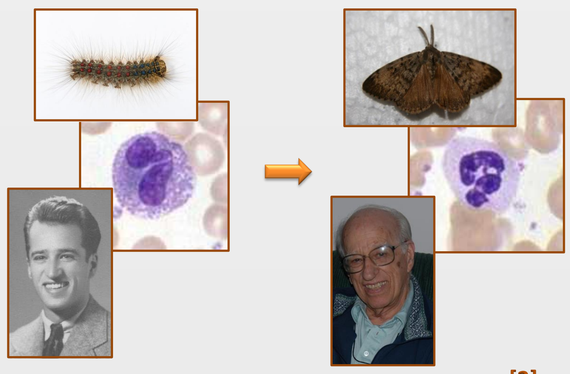
Network Of Silence

Flemming S.[1]; Bohleber S.[1]; Häupl, T.[2]; Günther, S.[1]

[1] *Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences, University of Freiburg*
[2] *Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin*

stephan.flemming@pharmazie.uni-freiburg.de

"Epigenetics: stably heritable phenotype without alterations in the DNA sequence."



[2]

Aim

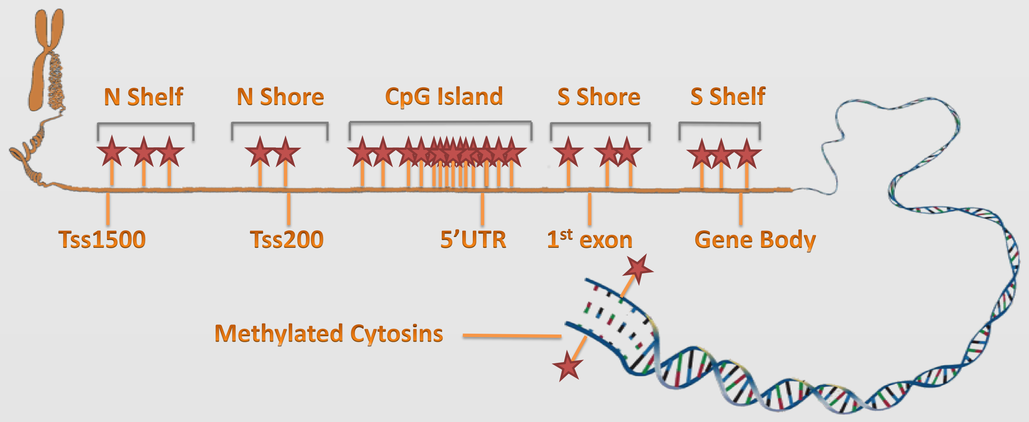
Changes in methylation levels of CpG dinucleotides correlate with **transcriptional repression** and **gene silencing**, although not all sites have the same impact on gene expression. The aim of this study is the generation of a network based on DNA methylation data

- to identify **cluster** of CpG sites with similar methylation levels among different tissues and conditions,
- to elucidate a relationship of **sequence motifs** and transcription factor binding sites with similar methylation patterns,
- to identify CpG sites which are more **predictive** for changes in **gene expression** than others,
- to assign similarities in methylation patterns to **functional proeperties**,
- to identify condition-specific CpG sites.

Epigenetics

Beside microRNA production and histone modification, **DNA methylation** is a well studied epigenetic modification and describes the covalent binding of a methylgroup to a cytosin within a cytosin-guanin dinucleotide (CpG site). CpG dinucleotides can be **methylated**, **unmethylated** and **hemimethylated**.

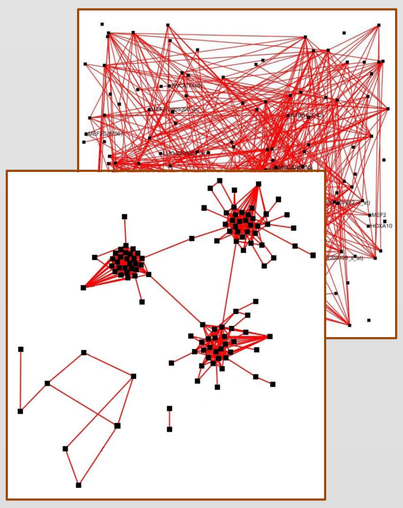
The approximately **28 million CpG sites** in the human genome are not equally distributed and occur mainly in clusters of high CpG density, called CpG islands (CGI), [1] which can be found in approximately 60% of all human gene promotors, but are also located in sections surrounding the transcription start site, gene body and sections that follow the translation terminal codon.



[3]

Network

A correlation network based on public available datasets derived with the **Illumina HumanMethylation450** platform [1] was generated. The platform provides a **genome-wide coverage** for methylation levels on single nucleotide resolution, which are represented with a **beta value** as a quantitative measure ranging from 0 for completely unmethylated to 1 for completely methylated.



$$\beta = \frac{Intensity\ M}{Intensity\ U + Intensity\ M + 100}$$

Clustering was obtained with the **k-means** algorithm. The network has the following properties:

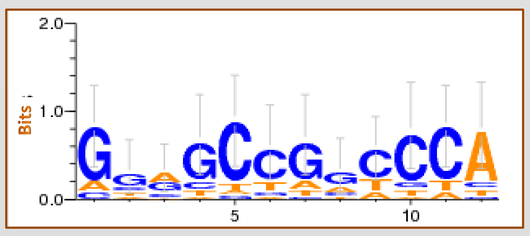
- approximately **482,000 nodes** representing CpGs for more than 23,000 genes
- more than **20 Mio edges** (threshold 0.95)
- ~5,300 samples (120 series) (Gene Expression Omnibus: GPL13534 [4])

Example

The following cluster consists of 37 CpGs representing 24 genes, including the glycoprotein CD4, which is located on the surface of immune cells such as T helper cells.

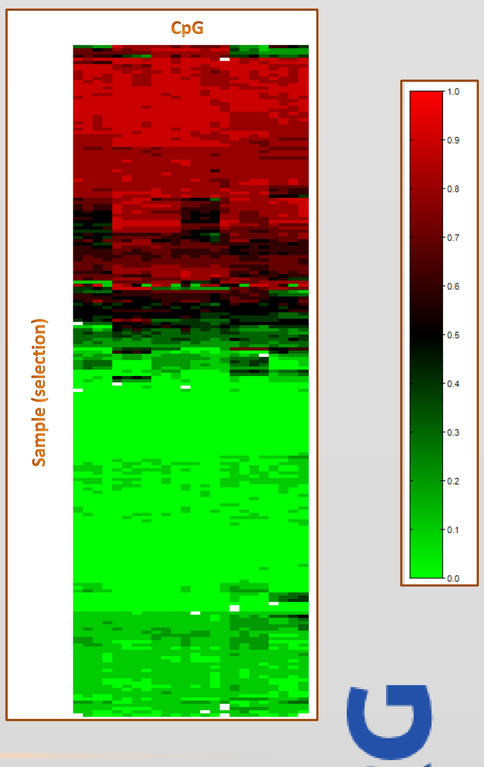
Sequence Logo

Graphical representation of the sequence conservation of nucleotides surrounding the CpGs within a cluster.



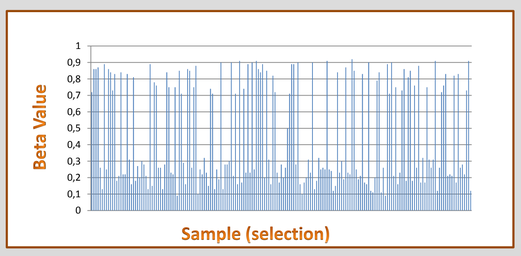
HeatMap

Representing the β-values of CpGs within one cluster among different conditions.



Condition Reference

Histogram as an overview over the beta values for a single CpG among different conditions.



Outlook

A **webservice** is currently under development and will provide a interface for queries on the existing network as well as an option to extend the network with own data. Cell specific patterns are difficult to identify with the current approach, if they are only represented within a small number of samples. A detailed analysis of the resulting clusters for unique methylation levels of a CpG in a specific sample series could solve this problem.

Pharmazeutische Bioinformatik

References

[1] Dedeurwaerder S., et al., Evaluation of the Innium Methylation 450K technology. Epigenomics, 3(6):771{84, December 2011
[2] http://www.flickr.com/photos/stevehart/3970336221/, http://commons.wikimedia.org/wiki/User:Lennert_B, http://commons.wikimedia.org/wiki/User:IvanTortuga, http://commons.wikimedia.org/wiki/User:Archaeodontosaurus
[3] Darryl Leja (NHGRI), Ian Dunham (EBI)
[4] Edgar R, et al., Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids , Res. 2002 ;30(1):207-10

UNI FREIBURG