Genome-wide mapping of DNA methylation variants affecting gene expression levels in gliomas with respect to their grade and *IDH* gene mutation status

Michal J Dabrowski[1], Agata Dziedzic[1], Rafal Guzik[2,3], Michal Draminski[1], Bartosz Wojtas[2], Karolina Stępniak[2], Bartłomiej Gielniewski[2], Jacek Koronacki[1], Bozena Kaminska[2]

[1] Institute of Computer Science, Computational Biology Lab, PAS, Warsaw, Poland; [2] Nencki Institute of Experimental Biology, Laboratory of Molecular Neurobiology, Warsaw, Poland;
[3]Andrzej Frycz Modrzewski Krakow University, Faculty of Medicine and Health Sciences, Department of Biochemistry, Poland.

1 Introduction

Epigenetic changes of DNA methylation status may affect gene expression and play a crucial role in glioma pathogenesis. Here we focused on differentially methylated promoters between various histopathological WHO glioma grades: pilocytic astrocytoma (PA, GI) and higher-grade glioma (diffuse astrocytoma and glioblastoma, hereafter HGG, GII-GIV). We used SeqCap Epi CpGiant Methylation panel and performed bisulphite conversion followed by Illumina NGS sequencing as well as gene expression (RNA-seq) analysis (Fig. 1). To analyze methylomes we developed CytoMeth tool which compiles a set of open source software enabling fast rough data processing.

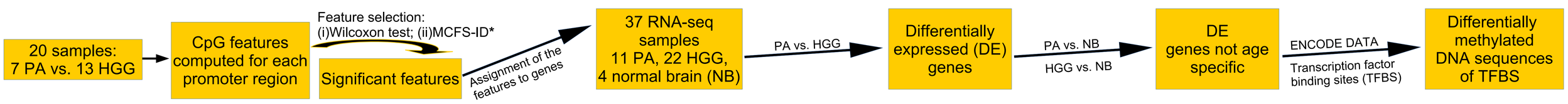

Fig. 1. Outline of the project. *Wilcoxon test with false discovery rate correction; MCFS-ID: Monte Carlo Feature Selection and Interdependency Discovery[1]

2A Results

At average, there were ~23 mln sites of well covered cytosines (coverage ≥ 10 reads) obtained for each sample. Majority of them were in the non-CpG context and ~3.3 mln were in CpG dinucleotides. The sum of unique promoter-specific CpG sites across all samples was above 2.5 mln, but the common part constituted only ~155,000 (Fig.2). Out of differentially expressed, not age specific genes that differed also in promoter methylation level, we chose the ones with the highest fold change in expression between PA and HGG. For these genes (Fig. 3), their description was obtained via text mining and based on that unsupervised clustering was performed.

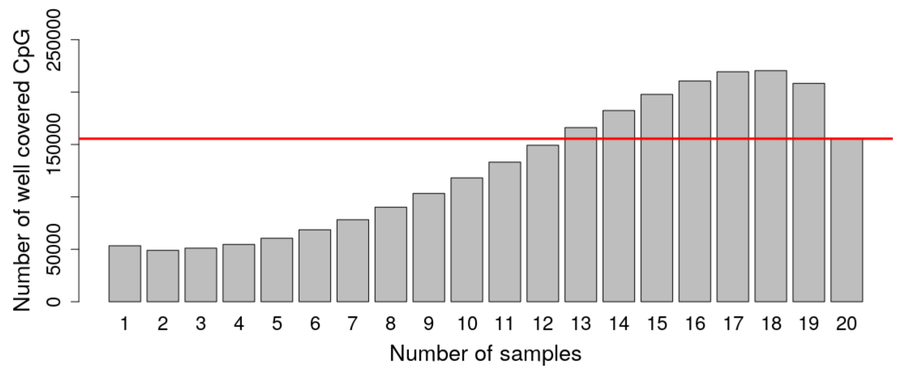

Fig. 2. The cumulative number of CpG sites within promoters vs. number of samples

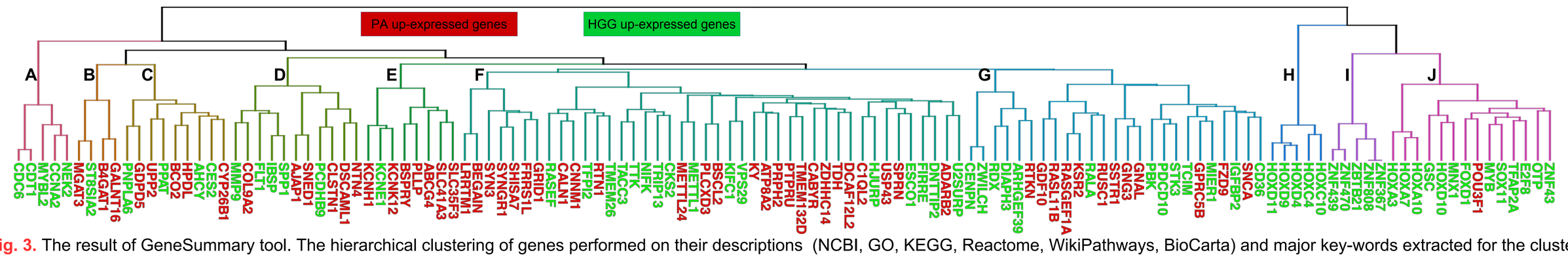

Fig. 3. The result of GeneSummary tool. The hierarchical clustering of genes performed on their descriptions (NCBI, GO, KEGG, Reactome, WikiPathways, BioCarta) and major key-words extracted for the clusters.

We obtained ten clusters of genes, named A-J (Fig. 3), characterized with the following specific key-words: (A) cell cycle, DNA replication, G transition; (B) glycosylation, n glycan, metabolism; (C) metabolic / catabolic process, retionic acid, metabolism; (D) receptor, cell adhesion, growth, k_akt, extracellular matrix; (E) transmembrane transport, potassium ion, ABC channel; (F) protein, gene, synaptic, spindle, splicing; (G) positive regulation, signal transduction, negative regulation, response; (H) homeobox genes, morphogenesis; (I) RNA transcription regulation, RNA polymerase; (J) development, differentiation, transcription, morphogenesis.

2B Results

Within differentially methylated promoters, we checked the distance of CpG sites to their corresponding transcription start sites (TSS) to make sure that detected differences concern the same genomic regions (Fig. 4). Further, using brain cell lines ENCODE ChIP-seq data, we annotated transcription factors motifs to promoters (Fig. 5) and localized transcription factor binding sites (TFBS). The differences in methylation level of TFBS sequences between PA and HGG samples were tested (Fig. 6).

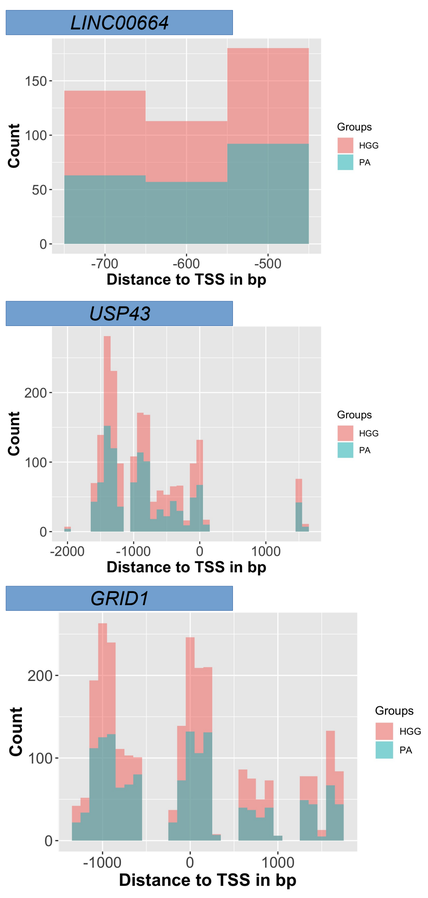

Fig. 4. The CpG sites position in relation to TSS

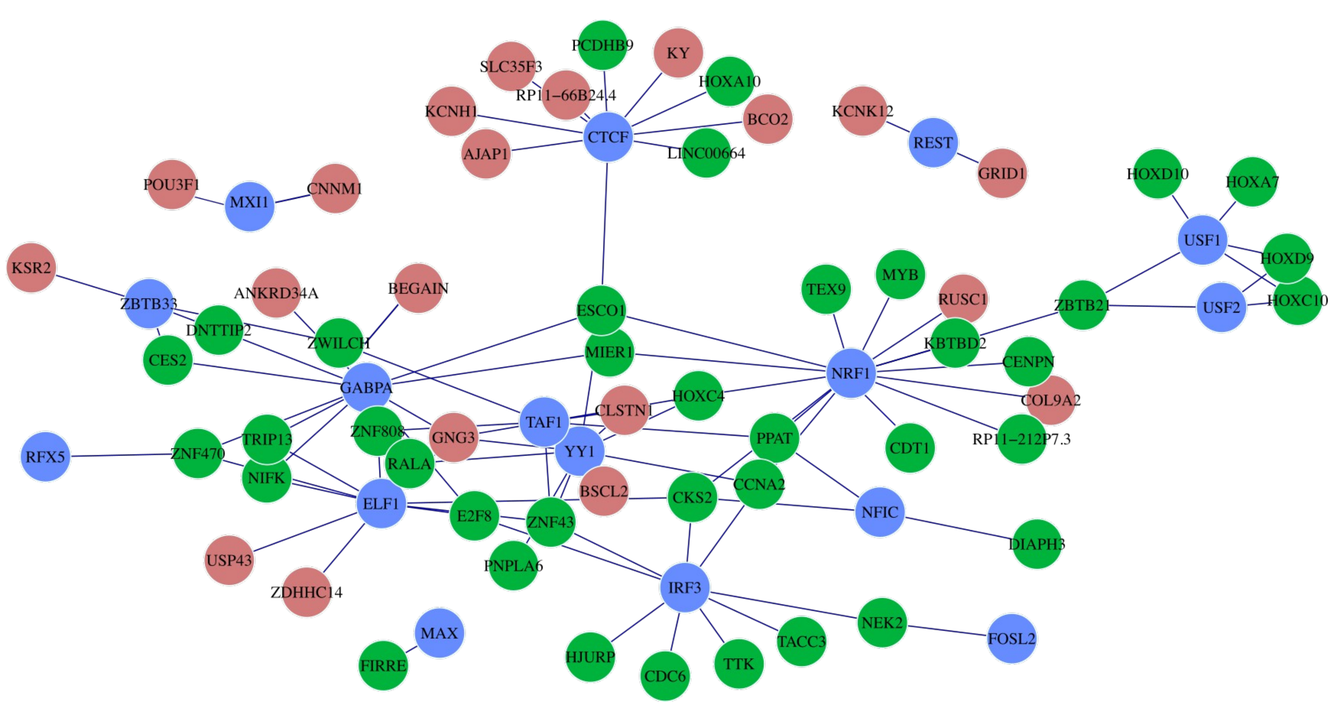

Fig. 5. Transcription factors (TF) whose motif sequence was detected in the gene-promoter region.
Blue – TF names; **Red** – gene names up-expressed in PA; **Green** – gene names up-expressed in HGG

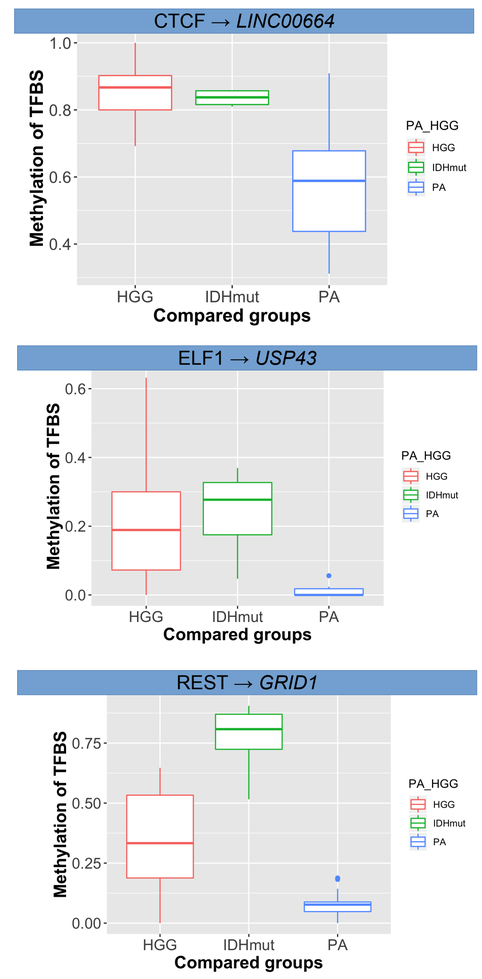

Fig. 6. The exemplary TFBS that differed in methylation level

3 Conclusions

- The clear functional division of specific genes, allowed by DNA methylation analysis, indicates their potential biological importance in pathways of gliomagenesis. The expression levels of some of those genes can be explained by specific TFBS methylation patterns. One interesting example is HOX group enclosing *HOXA7*, *HOXC10*, *HOXD10* and *HOXD9* genes, all up-expressed in HGG and regulated by USF1 and USF2.
- TFBS methylation pattern showed significant relation to tumor grade as well as *IDH1* mutation status. Higher methylation levels were observed in GII-GIV and *IDH1* mutants.

This project was supported by the NCN grant 2015/16/W/NZ2/00314

[1] Dramiński Michał and Koronacki Jacek. "rmcfs: An R Package for Monte Carlo Feature Selection and Interdependency Discovery." *Journal of Statistical Software* 85.1 (2018): 1-28.