DOI 10.18699/BGRSSB-2018-025

Motif analysis of regulatory SNPs reveals Krüppel-like transcription factors as putative tumor suppressors in colorectal carcinoma

O. Kel-Margoulis[1]*, A. Kel[1, 2], E. Wingender[1, 3]
[1] *geneXplain GmbH, Wolfenbuettel, Germany*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[3] *Institute of Bioinformatics, University Medical Center Göttingen, Georg August University, Göttingen, Germany*
** e-mail: olga.kel-margoulis@genexplain.com*

Key words: regulatory SNP, transcription factors, signal transduction, colorectal cancer, CRC

Motivation and Aim: Genome-wide associated studies (GWAS) reveal multiple SNPs correlated with predisposition to various cancers. Majority of such SNPs are found in non-coding regions of genome, which makes their interpretation a special task. It becomes evident that such SNPs mark regions of genomes that posses important regulatory functions. It is important to understand the role of regulatory SNPs in molecular mechanisms of cancer initiation and progression. In the current work we study SNPs that are revealed by eQTL while analyzing massive transcriptomics (RNA-seq) and genotyping data of a cohort of 103 patients with colorectal carcinoma (CRC) [1]. Many of them are also shown to be correlated with the change of expression of neighbor genes. We applied a machine learning techniques combined with graph search algorithms in order to reconstruct regulatory circuits involving genes located near regulatory SNPs.
Methods and Algorithms: In the current work we applied a novel approach, which we refer to as "walking" pathways. This approach is further development of "upstream analysis" [2]. First, we applied an improved version of Composite Module Analyst (CMA) [3] to reveal specific combinations of transcription factor binding sites (TFBS) in the genomic regions around regulatory SNPs. Next, we performed a graph search in signal transduction network for common regulatory nodes upstream of transcription factors that were revealed at the previous step. At this step we identified positive feedback loops, which direct the network search towards potential master-regulators of a self-inducing pathological state of the system.
Results: We found that binding sites for KLF transcription factors are significantly enriched around regulatory SNPs. We also found that KLF transcription factor binding sites are co-localized with binding sites for E2F1, TP53 and other factors forming composite elements in the regions around regulatory SNPs. Analysis of gene expression data in tumor samples revealed significant down-regulation of three KLF family members. Finally, the upstream analysis with positive feedback loops (walking pathways) identified several potential master-regulators. Among them we identified several cyclin/cdk complexes, which are significantly up-regulated in the tumor samples. The reconstructed signal transduction circuit involves cell cycle regulators, Cdk1, Cdc2, Plk1, p107 as well as transcription factors c-Myc, TGIF, PEA3 whose expression is highly unregulated in tumor samples.
Conclusion: These findings, together with the fact of significant down-regulation of three KLF family members, allow suggesting GKLF as a putative tumor suppressor. Down-regulation of GKLF in many CRC patients may lead to activation of oncogenic regulatory circuits. These results shed a light on the important role of regulatory SNPs in cancer genomes.
Acknowledgements: Supported by the Targeted Program "Research and development on priority directions of science and technology in Russia, 2014–2021", Contract No. 14.604.21.0101, unique identifier of the applied scientific project: RFMEFI60414X0101, and by the grant of the EU FP7 program: "SYSCOL".

References

1. Ongen H. et al. (2014) Putative cis-regulatory drivers in colorectal cancer. Nature. 512(7512):87-90.
2. Koschmann J. et al. (2015) "Upstream Analysis": An integrated promoter-pathway analysis approach to causal interpretation of microarray data. Microarrays. 4:270-286. DOI 10.3390/microarrays4020270.
3. Waleev T. et al. (2006) Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. Nucleic Acids Res. 34(Web Server iss.):W541-5.