Statistical methods for RNAseq coverage profiles and their applications in biological phenomena.

Alicja Szabelska[1,2], **Idzi Siatkowski**[1], **Michal Okoniewski**[2]

[1]*Poznan University of Life Sciences, Wojska Polskiego 28, 60-637 Poznan*
[2]*Functional Genomics Center Zurich, Winterthurerstrasse 190, 8104 Zurich*

Abstract

Nowadays next generation sequencing, mostly RNA-Seq is one of the most powerful tools that can be applied to gain more insight in genomic issues. RNA-Seq produces the enormous amounts of data. To be able to find interesting features in such data the most popular parametric approaches (see [1], [5]) summarize the data to "count data" on the gene level and apply negative binomial distribution to model it with usage of generalized linear model. Other popular non-parametric [3] or based on empirical Bayes [2] methods were also introduced to analyze RNA-Seq data. However, they are also based on count data, which means the initial information produced by sequencer is reduced and we do not take advantage of the power of next-generation sequencing tools. That is why we present the methods for seeking and statistical verification of the genomic features based on coverage profiles of interesting genome. According to [4] the first step of the analysis is to introduce a measures of dissimilarity of the profiles. The second step is the statistical validation of the outcomes with usage of permutation test.

During the talk the overall work for coverage based approach will be presented. In addition, the examples of usage of the method in biological experiment will be included.

Keywords: Differential expression, alternative splicing, RNA-Seq, coverage profiles, permutation tests

AMS subject classifications: 62P10, 92B15

Acknowledgements: Research supported by Poznan University of Life Sciences and Functional Genomics Center Zurich (Scientific Exchange Programme NMS^{ch} - no. 11.182)

Bibliography

[1] Anders, S., & Huber, W. (2010). Diferential expression analysis for sequence count data. *Genome Biology*, 11: R106.

[2] Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: empirical Bayesian methods for identifying diferential expression in sequence count data. *BMC bioinformatics*, 11(1), 422.

[3] Li, J., & Tibshirani, R. (2011). Finding consistent patterns: A nonparametric approach for identifying diferential expression in RNA-Seq data. Statistical Methods in Medical Research.

[4] Okoniewski, M. J., Leniewska, A., Szabelska, A., Zyprych-Walczak, J., Ryan, M., Wachtel, M., Morzy, T., Schaer, B., & Schlapbach, R. (2012). Preferred analysis methods for single genomic regions in RNA sequencing revealed by processing the shape of coverage. Nucleic Acids Research, 40(9), e63-e63.

[5] Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a bioconductor package for diferential expression analysis of digital gene expression data. Bioinformatics, 26: 139140.