Researchers obtain most accurate measures of gene expression

September 4 2014

RNA-sequencing data analysis method BitSeq developed by Academy Research Fellow Antti Honkela's research group and University of Manchester researchers has been found to be the most accurate gene transcript expression estimation method in a large international assessment. The method is based on probabilistic modelling which can capture the uncertainty related to the measurements.

RNA-sequencing allows measuring the gene expression of humans or other organisms. The method has recently become very popular in bioscience and medical research, and it is being adopted to clinical applications. Compared to previous methods, RNA-sequencing enables the study of alternative gene isoforms or transcripts, which are formed for example through the process of alternative splicing.

The analysis of the large amount of data produced by RNA-sequencing requires many advanced computational methods. Analysis of transcript level data is especially demanding and the differences between alternative methods can be large.

In the recent assessment the BitSeq method developed by the University of Helsinki and the University of Manchester researchers produced clearly the most reliable results in this task. In one subtask it could produce equally accurate results using only half the data needed by a very popular alternative method.

"The BitSeq method is based on probabilistic modelling that allows

comparing different possible origins for observed sequences that cannot be identified uniquely. This allows computing probability distributions over the expression levels of each transcript of every gene in a way that captures the uncertainty and possible sources of error in the measurements. Accounting for this uncertainty through probabilities is essential for the
accuracy of the method," Antti Honkela says.

More information: SEQC/MAQC-III Consortium. (2014). "A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality." Control Consortium. *Nature Biotechnology*. DOI: 10.1038/nbt.2957

Glaus, P., Honkela, A., & Rattray, M. (2012). "Identifying differentially expressed transcripts from RNA-seq data with biological variation." *Bioinformatics*:28(13), 1721?1728. DOI: 10.1093/bioinformatics/bts260

Provided by University of Helsinki