

УДК 616.24-006.6:577.21/.23:007:002.6:681.31

МОЛЕКУЛЯРНІ СИГНАТУРИ РАКУ ЛЕГЕНІВ ЗА ДАНИМИ СЕКВЕНУВАННЯ РНК І РАНЖУВАННЯ ГЕНІВ ЗА РІВНЕМ ЕКСПРЕСІЇ

В. Бондаренко, Ф. Еммерт-Стрейб¹

*Інститут молекулярної біології і генетики НАН України
Університет Технологій Тампере*

Гени, що належать до таких біологічних процесів як «клітинна міграція», «морфогенез епітелію», «розвиток кровоносних судин» та «відповідь на реактивні форми кисню», забезпечують найточнішу класифікацію пухлинних зразків та зразків здорової тканини легенів. Ми припускаємо, що ранжування генів за рівнем їх експресії в якості основи для досліджень транскриптому сприяє глибшому розумінню змін в транскриптомі клітини, що викликані клітинним переродженням.

MOLECULAR SIGNATURES OF LUNG CANCER BASED ON GENE EXPRESSION RANKING OF RNA-SEQ DATA

V. Bondarenko, Frank Emmert-Streib¹

*Institute of Molecular Biology and Genetics, NAS of Ukraine
¹Tampere University of Technology, Finland*

Introduction. Whole Transcriptome Shotgun Sequencing or RNA-Seq is an experimental methodology which allows quantification of all transcripts expressed in a population of cells. In biomedicine, it is gradually gaining popularity to become a prevalent method for transcriptome analysis. Sequencing of individual human transcriptomes has become a powerful tool in disease biomarkers discovery and setting up individual treatment therapy of cancer and other disorders.

Among all cancer-related diseases, lung cancer remains to be the leading cause of the majority of deaths worldwide. Because of diagnostic difficulties and high level of recurrence even after surgical treatment, many efforts have been made to identify genes with prognostic properties. In several studies, on the basis of microarray gene expression data, diagnostic and prognostic gene expression signatures have been identified for non-small cell lung cancer. Interestingly, there is almost no overlap between some of the published signatures i.e. the evidence indicates that none of the reported gene expression signatures are ready for clinical application.

Methods. We explored a large-scale RNA-seq dataset from a biomedical study of lung adenocarcinoma

(84 patients). First, we investigated influence of the technical variation in the sequencing depth on the gene expression ranks using repetitive data subsampling and statistics, such as rank correlation, median and maximum rank changes. Secondly, using gene expression ranks, we applied a Bayesian classifier as a diagnostic predictor for lung cancer. As a feature selection mechanism, we used prior biological information from the Gene Ontology (GO) project database.

Results and conclusions. On the basis of the gene expression rank changes between normal and cancer samples, we provide an accurate classification of cancer and non-cancer samples. We demonstrate, that genes belonging to processes of «cell migration», «epithelium morphogenesis», «blood vessels development» and «response to oxygen levels» result in best performance of the lung cancer and non-cancer transcriptomes classification. We suggest that our results are beneficial in terms of using gene expression ranking as a basis for whole-transcriptome studies and provide novel insights in our understanding of the transcriptome disruptions caused by cancer progression.