New insights from Affymetrix expression microarray data analysis

Andrey V. Marakhonov, N.S. Sadovskaya, E.A. Borzov,

Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moskvorechie Str., 1, Moscow, Russia, marakhonov@generesearch.ru

Ancha Baranova,

Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moskvorechie Str., 1, Moscow, Russia,

School of Systems Biology, David King Hall, MSN 3E1, George Mason University, Fairfax, VA, 22030, abaranov@gmu.edu

Mikhail Yu. Skoblov

Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moskvorechie Str., 1, Moscow, Russia, mskoblov@generesearch.ru

Recent advances in methodology have led to the extent studies of organisms at the level of nucleic acids. Microarray gene expression profiling offers an opportunity for genome-scale, quantitative evaluation of gene expression studies by simultaneously measuring expression levels for thousands of genes. Affymetrix expression arrays provide powerful tool for discovering global gene expression profile in different conditions. Now, expression microarrays used in a variety of applications, including understanding of pathologic changes in different conditions of organism [1], differentiation of pathologic conditions [2], investigation of molecular basis of response to different stimuli, toxins, environment exposure, stresses, etc. [3, 4], targets of miRNAs [5], transcription factors [6], decipher the signaling pathways, etc. Nevertheless the discrepancies in probe sets data remained to be wide-spread phenomenon which is usually underestimated or dissembled. So the correct annotation of probe sets remains the actual problem. The man goal of the study was to determine probe sets which could correctly represent the real gene expression profile in human transcriptome.

We downloaded GSE1133 data set from BioGPS service [7] which is generated using Affymetrix HG U113A platform. First of all we determined genes represented more than single probe set in U133A array. We have also updated array annotation by grouping probe

sets which were annotated to several genes into one group. We referred group as several probe sets to one gene. For further analysis we have chosen group contained two probe sets per gene as a model group. Total amount of two-probeset groups was 2759 in Affy U133A array including 68 cases when groups consist at least one probe set annotated to more than one gene.

The performed correlation analysis allowed us to distinguish the multiple probe sets containing genes into several groups. Highly correlated probe sets to the single genes could indicate the specific probe sets which are correctly annotated. Two other types of uncorrelated probe sets could be related to the different transcript variants and tissue-specific genes. The one of cause of uncorrelated data from probe sets to the single genes remained to be improper annotation of probe sets. We could also distinguish cases of interrelations of gene and its pseudogene.

Another intriguing insight from Affymetrix expression microarray data was analysis of *cis*-antisense partners cooperative expression.

This work was supported by the application no. 2012-1.3.1-12-000-1001-028 for grant of the Ministry of education and science of Russia (contract № 8589, dated September 24, 2012) and Personal Grant of President of Russian Federation for young PhD scientists for 2012-2013 (МК-5249.2012.4).

1. M.Torres-Martin et al. (2013) *Int J Oncol*, **42**:848–862.

2. A.E.Ivliev et al. (2010) *Cancer Res*, **70**:10060–10070.

3. M.Tokumoto et al. (2013) *J Toxicol Sci*, **38**:55–7.

4. A.B.Shirode et al. (2013) *Mol Carcinog*, DOI:10.1002/mc.21995.

5. S.Artmann et al. (2012) *PLoS One*, **7**:e38365.

6. E.G.Hagos et al. (2011) *Am J Cancer Res*, **1**:85–97.

7. C.Wu et al. (2009) *Genome Biol*, **10**:R130.