Student Sebastian
Biosystems Group, Institute of Automatic Control, Silesian University of Technology, 44-100 Gliwice, Poland
e-mail: sebastian.student@polsl.pl
Cichońska Anna
Student of the Faculty Of Automatic Control, Electronics And Computer Science, Silesian University of Technology, Gliwice, Poland
e-mail: anna.cichonska@gmail.com
Skonieczna Magdalena
Biosystems Group, Institute of Automatic Control, Silesian University of Technology, 44-100 Gliwice, Poland
e-mail: magdalena.skonieczna@polsl.pl
Joanna Rzeszowska-Wolny
Biosystems Group, Institute of Automatic Control, Silesian University of Technology, 44-100 Gliwice, Poland
e-mail: jrzeszowskawolny@yahoo.com

Microarray gene expression studies and real time RT-PCR validation for the DNA damage and repair pathway

Different low-level preprocessing methods for Affymetrix microarrays data were evaluated based on concordance with a real time RT-PCR method. The aim of low-level analysis is to measure gene expression levels, and to allow comparison of the results from more than one array. In this paper three of the most popular preprocessing methods: MAS5, RMA and GCRMA, were used. Expression of genes from the DNA damage and repair pathway were analyzed through the MAS5 - single array analysis algorithm, the GCRMA - probe-specific background correction and multiple array analysis, or RMA - mismatch probes ignored and multiple array analysis.
The data were derived from experiments conducted with the Affymetrix platform U133A. For biological testing the colorectal carcinoma HCT 116 cell line was chosen. The cells were irradiated with 4 Gy of ionizing radiation, and non-irradiated cells used as a control group. After microarray data analysis, real time RT-PCR was conducted. As an indicator for concordance between microarray experiments and real time RT-PCR, the percentage of genes with the same direction of changes in irradiated and non-irradiated cells was used. The computational analysis was finished with the PLS-based (partial least squares-based) gene selection method, which enables assignment of the biological meanings for the genes with the highest weights in the PLS model. The PLS method, in contrast to the PCA (principal component analysis) criterion based on maximization of the variance of a linear combination of genes, extracts components by maximizing the sample covariance between the class variable and linear combination of genes. The information for genes included in components described by PLS can be directly related to the biological meaning of this analysis.
The results show that data preprocessed with the RMA method for microarray data has the best concordance with real time RT-PCR assays. The biological validation for the best 10 genes with the highest weights in the PLS model proved its applicability in systems biology. Some of these genes (MSH2, RAD9A, XP) are sensors

for nucleic acid damage, and others (NTHL1, TDP1 DCLRE1A, ERCC2, POLI, MPG, TREX2) are engaged in mechanisms of DNA repair. Obviously, the best score was obtained for genes responsible for signaling cellular stress after ionizing radiation.

This work was supported by grants No. N N 518497639 from the Polish Ministry of Education and Science and BK 221/Rau1/20 from the Silesian University of Technology

References

[1] Affymetrix, *New Statistical Algorithms for Monitoring Gene Expression on GeneChip Probe Arrays***2001**.

[2] R. Irizarry et al, *Exploration, normalization, and summaries of high density oligonucleotide array probe level data* Biostatistics **4** 249-264.

[3] Z. Wu et al, *A Model Based Background Adjustment for Oligonucleotide Expression Arrays* Journal of the American Statistical Association **99** 909-917.

[4] A. Boulesteix, *PLS dimension reduction for classification with microarray data.* Statistical Applications in Genetics and Molecular Biology **3** Article33.