Roman Jaksik, Michał Marczyk, Joanna Polańska
Institute of Automatic Control, Silesian University of Technology
E-mail: `roman.jaksik@polsl.pl, michal.marczyk@polsl.pl,`
`joanna.polanska@polsl.pl`
Andrzej Polański
Institute of Informatics, Silesian University of Technology
E-mail: `andrzej.polanski@polsl.pl`

Microarray artifacts elimination algorithm based on image recognition techniques

Gene expression profiling studies conducted with the oligonucleotide microarrays are one of the most popular methods used to characterise global transcriptome changes induced by various chemical or physical factors. Despite their proven usefulness in the identification of genetic aberrations leading to various diseases, including cancer, they require very extensive data processing techniques in order to separate expression level changes resulting from various artifacts and measurement inaccuracy from those of biological origin. One of the crucial sources of measurement inconsistencies are artifacts originating from aberrations on the surface of the microarray which is scanned with a laser detector during the experiment. Most common sources of such aberrations are defects of the manufacturing process and impurities within target genomic material, which can significantly affect the final results. Data processing methods often fail to exclude outlying signal values resulting from such defects leading to artificially increased variation between replicate experiments, to decreasing statistical significance of between-sample studies, or to reduced accuracy of sample classification if the experiment aims to search for a factor induced genetic response signature. In this work we propose a novel detection and correction method based on signal intensities of other, unaffected, replicate probes. The method was implemented as a standalone windows application allowing for very easily processing of hundreds of microarray samples, within few minutes and visualizing the results of analyses on various processing steps. The usefulness of the presented method was evaluated by the analysis of a real microarray dataset with technical replicates showing large number of aberrations. Using common statistical methods based on replicate sample correlation the overall impact of artifacts processing was evaluated. The results show that image artifacts correction increases dataset integrity, proving that it is possible to separate image defects from inter sample variations of biological origin and specific features of the microarray chip, achieving higher quality of the analysed data.