Pawel Foszner
Silesian University of Technology
e-mail: Pawel.Foszner@polsl.pl
Roman Jaksik
Silesian University of Technology
e-mail: Roman.Jaksik@polsl.pl
Aleksandra Gruca
Silesian University of Technology
e-mail: Aleksandra.Gruca@polsl.pl
Joanna Polanska
Silesian University of Technology
e-mail: Joanna.Polanska@polsl.pl
Andrzej Polanski
Silesian University of Technology
e-mail: Andrzej.Polanski@polsl.pl

Efficient reannotation system for verifying genomic targets of DNA microarray probes

Systems for data cleaning for supporting analysis of results of DNA microarray experiments are becoming important elements of bioinformatics aspects of gene expression analysis [1]. It has been demonstrated that data cleaning at the level of microarray probes, based on most recent knowledge on genomic data, can substantially improve results of predictions of molecular classifiers. However, due to the difficulty of the whole genome browsing projects, available services and data for reannotation of microarray probes are still quite sparse. In our research we have created an efficient reannotation tool by combining the well known gene search tool BLAT [2] with appropriately designed database and tools for operations on it.

We show properties of our tool by using two Affymetrix chips HG U133A and HG 1.0 ST. In the Affymetrix microarrays, the gene intensity is calculated on the basis of gene probes consisting of 25-mer oligo-nucleotides. For many reasons, in many cases, the calculated value does not match the real expression. These reasons include single nucleotide polymorphism, adjusting the probe to another gene or intron. Our task was to check how many probes can truly determine gene expression. We have developed a database which contains information about how the probes are aligned to the latest human genome. Using those matches to the genome, for each probe we found mRNA and EST sequences. In our presentation we compare reannotation results for analyzed Affymetrix chips, based on two different built of Human Genome, HG18 and HG19. Improving the quality of data can be further verified by comparing the misclassification rates for classification of microarray data obtained using the official affymetrix CDF files and CDF file created by us. The information obtained from reannotations can help to update the CDF files, and can significantly improve the quality of classification.

Acknowledgements. This work was supported by the European Community from the European Social Fund.

References

[1] Ramil N. Nurtdinov, Mikhail O. Vasiliev,Anna S. Ershova, Ilia S. Lossev and Anna S. Karyagina, *PLANdbAffy: probe-level annotation database for Affymetrix expression microarrays* Nucleic Acids Research, 2010, **38** D726–D730.

[2] Kent,W.J., *BLAT–the BLAST-like alignment tool.* Genome Res. **12** 656–664.