**Aleksandra Pfeifer**
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch; Silesian University of Technology, Poland
e-mail: apfeifer@io.gliwice.pl
**Małgorzata Oczko-Wojciechowska**
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, Poland
**Michał Świerniak**
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch; Silesian University of Technology, Poland
**Michał Jarząb**
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, Poland
**Barbara Jarząb**
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, Poland

## Sources of variability in the gene expression profile of follicular thyroid tumours: SVD analysis of microarray data

Many attempts have been performed by microarray gene expression profiling of thyroid follicular tumours in order to find genes that distinguish adenomas and carcinomas. The two types of thyroid follicular tumours: adenomas (benign) and carcinomas (malignant) are indistinguishable before surgical procedure by classical pathology. A hypothesis that gene expression profiling by microarray test may aid in the diagnosis has not been fully verified. The aim of our study was to apply unsupervised methods of gene expression analysis to identify the main sources of variability in follicular tumors which may influence the feasibility of genetic testing in this disease. We performed microarray gene expression profiling in 45 follicular tumours by Affymetrix hgu133plus2 microarray. We performed Singular Value Decomposition (SVD) analysis of the whole dataset to identify the supergenes (modes) that characterise the main sources of variation and are more representative/stable than single transcripts. Next we analysed the biological meaning of the variability related to each supergene. We selected genes that contribute most to each of the supergenes and analysed them with different biological mining methods: gene ontology analysis, gene groups analysis and hierarchical clustering of samples. We revealed that the main sources of variance in the analysed dataset are related to the immune response (1st, 3rd and 6th supergenes), cell proliferation (2nd and 5th supergenes) and differentiation (2nd supergene). Among genes that contribute most to the 1st, 3rd and 4th supergene, many are related to the difference between thyroid carcinoma and normal thyroid tissue. As in the analysis we noted certain arbitrary steps, we also performed SVD analysis on the artificial microarray dataset to assess the influence of these parameters on the results. Comparison of SVD to other unsupervised methods will also be presented.