High throughput expression profiling of circulating tumor cells from breast cancer patients as potential therapy decision indicator

Katarina Kolostova 1,2, Marianna Romzova 2, Vladimir Bobek 1,3, Mikael Kubista 2,4, Sabine Kasimir-Bauer 5

1. Dept. of Tumor Biology, Third Faculty of Medicine, Charles University in Prague, Czech Republic. 2. Laboratory of Gene Expression, Institute of Biotechnology, Academy of Sciences, Prague. 3. Dept. of Surgery, Third Faculty of Medicine and Faculty Hospital Kralovske Vinohrady, Charles University Prague. 4. TATAA Biocenter, Gothenburg, Sweden. 5. Dept. of Gynecology and Obstetrics, University Hospital Essen, Germany.

Purpose of the Study
To assess the molecular profiles of circulating tumor cells (CTCs) from breast cancer (BC) patients (early and metastatic) and to evaluate their potential as prognostic and therapy indicators.

Gene Expression Analysis for Metastatic Breast Cancer
Cluster analysis of CTC-MBC expression data shows the most variable discriminating genes: EPCAM, MUC1, KRT19, AURKA, MCM, TOP2A, TWIST, and ALDH1 when comparing the positive and negative CTC-samples.

Key Findings
- CTCs were found in 42% of the metastatic and in 26% of the early BC patients.
- There is a difference between gene expression profiles of the CTCs within the metastatic process. The mRNA levels of the markers used in routine for CTC identification (GA7332, MUC1, KRT19) change with the progression of the disease.
- We conclude that CTCs enriched by the AdnaGen-technology are heterogeneous and do not definitively divide the CTC positive and CTC negative samples of the early BC patients. The biggest differences were found for EPCAM, KRT19, TOP2A.
- The predictive value of expression profiles in CTCs for the therapeutic interventions will be further prospectively evaluated.

Panel of markers: ADAM17, AKT2, ALDH1, AURKA, CD24A4, CD45, CTSF, CXCR1, ERBB2, ESR, GA7332, KI67, KRT19, KRAS, MCM4, MUC1, MTOR, MYC, PTEN, PARR, PGR, P3K, SATB1, SCGB2A, TOP2A, TP53, TWIST, VEGFA, VEGFR, WHSC1L1.

Analysis of the gene expression data was performed using the multivariate methods Hierarchical clustering and principal component analysis (PCA). The cut off value was set as 30 Cq. Missing data of duplicates were filled by interpolation. Data not obtained by amplification were filled by maximum Cq value of the tested gene plus 0.5.

We were not able to find genes with invariable expression that would be suitable for normalization. The data were normalized to the global expression of all genes and consequently, gene expression levels are relative to the other genes in the panel. For multivariate analysis the data were also associated to give the genes the same clustering weights.

Gene Expression Analysis for Early Breast Cancer
In contrast to expression profiles obtained in MBC, we can not definitively divide the CTC positive and CTC negative samples of the early BC patients. The biggest differences were found for EPCAM, KRT19, TOP2A.

Key Findings
- CTCs were found in 42% of the metastatic and in 26% of the early BC patients.
- There is a difference between gene expression profiles of the CTCs within the metastatic process. The mRNA levels of the markers used in routine for CTC identification (GA7332, MUC1, KRT19) change with the progression of the disease.
- We conclude that CTCs enriched by the AdnaGen-technology are heterogeneous and do not necessarily express EPCAM. Even CTCs with low EPCAM expression but with elevated KRT19 and AURKA levels may represent a subgroup of high risk metastatic BC patients.
- In early BC, after comparison of the CTC-positive and negative samples, we may conclude, that the biggest differences were found for EPCAM, KRT19, WHSC1L1, AURKA and TOP2A.
- The predictive value of expression profiles in CTCs for the therapeutic interventions will be further prospectively evaluated.

This work has been supported by Ministry of Health Grant – IGA – NS 9976, Grant Agency of Charles University no. 7708 and by Ministry of Education, Youth and Sports of the Czech Republic – Kontakt ME1045.

Panel of markers: AKT2, ALDH1, AURKA, CD24A4, CD45, CTSF, CXCR1, ERBB2, ESR, GA7332, KI67, KRT19, KRAS, MCM4, MUC1, MTOR, MYC, PTEN, PARR, PGR, P3K, SATB1, SCGB2A, TOP2A, TP53, TWIST, VEGFA, VEGFR, WHSC1L1.

Analysis of the gene expression data was performed using the multivariate methods Hierarchical clustering and principal component analysis (PCA). The cut off value was set as 30 Cq. Missing data of duplicates were filled by interpolation. Data not obtained by amplification were filled by maximum Cq value of the tested gene plus 0.5.

We were not able to find genes with invariable expression that would be suitable for normalization. The data were normalized to the global expression of all genes and consequently, gene expression levels are relative to the other genes in the panel. For multivariate analysis the data were also associated to give the genes the same clustering weights.

Gene Expression Data Analysis and Melt Curve Analysis software were used to obtain Cq- values and Tm. gPCR- results have been analyzed by GENEX v.s. 5.0 software (MultiD, SE).

This work has been supported by Ministry of Health Grant – IGA – NS 9976, Grant Agency of Charles University no. 7708 and by Ministry of Education, Youth and Sports of the Czech Republic – Kontakt ME1045.

We would like to thank our colleagues taking part in the project and helping us: Dr. Petra Statakov, Dr. Martina Kubecova, Dr. Milan Brychta, Dr. Josef Valchar, Dr. Zuzana Usiakova, Veronika Mlcekova, Dana Adorina, Eva Hroncova.