mRNA expression analysis and classification of colonic biopsy samples using oligonucleotide and cDNA microarray techniques

Orsolya Galamb

PhD thesis

2007, Budapest
Semmelweis University, 2nd Department of Internal Medicine

Semmelweis University Doctoral School
Clinical Medicine Doctoral School
Gastroenterology PhD program

Program leader: Zsolt Tulassay MD, DSc
Supervisor: Béla Molnár MD, PhD
INTRODUCTION

Colorectal cancer (CRC) is one of the most frequent cancers in the world with very high mortality. In Hungary CRC is also the second death-causing tumorous disease. In our country, approximately 8000 new colorectal cancer cases are registered, and almost 5000 CRC-related deaths appear in every year. Hence, the early diagnosis, the discrimination of genetically and expressionally different tumors and in the light of these, enhancement of the therapies become necessary. The five-year survival data also emphasize the importance of the early diagnosis of CRC. Five-year survival rate is 80-90% in early CRC, 60% in case of nodal involvement, while under 10% in metastatic CRC.

Using the earlier research methods one could analyse one or some selected disease-related genes, the analysis of whole molecular background was not possible. The recently available microarray techniques allow us simultaneous monitoring of the expression of thousands of genes.

The mRNA expression microarrays make possible studying expression and activation of large number of genes in a single hybridization, in different cells, tissues or distinct stages of the same cell/tissue type (such as physiological or pathological state, and with or without drug treatment). The essence of expression microarray technology is that complementary target oligonucleotides or cDNAs are fixed to solid surface, and single strand cDNA or cRNA probes which are complementary to the extracted sample RNA and labelled with modified nucleotides are hybridized to them. Intensity of the hybridization refers to the expression level of the tested genes. The microarray techniques not only give new information about the causes of diseases, but offer new diagnostic and differential diagnostic opportunities.

In my PhD thesis I have been searching for biomarkers of the development of colorectal carcinoma, and made gene expression analysis for colorectal disease classification using whole genomic oligonucleotide and cDNA microarray technology and colonic biopsy samples.
AIMS

The aims of my PhD work were

• to confirm the applicability of biopsy samples for microarray analysis,
• comparison of the oligonucleotide and cDNA microarray systems, analysis of their advantages, disadvantages, applications and limits,
• to develop and test validation assay system,
• to analyse the gene expression background of the colorectal adenoma-dysplasia-carcinoma sequence: to identify genes with altered expression indicating colorectal carcinogenesis,
• to search for altered biological pathways for explanation of the pathomechanism of these colonic diseases based on mRNA expression microarray results,
• to identify discriminatory genes between the main diagnostic groups of colorectal diseases, and to find gene expression-based relationship between the different colonic alterations.

MATERIALS AND METHODS

1. Samples

Totally, 456 colonic biopsy samples of 182 patients (33 healthy, 12 with ulcerative colitis (UC), 8 with Crohn’s disease (CD), 31 with early CRC (Dukes A and B stages), 37 with advanced CRC (Dukes C and D stages), 31 with non-dysplastic villous adenoma and 30 with dysplastic adenoma) were analysed. The samples taken for RNA isolation were placed into RNAlater Stabilization Reagent and stored at –80°C until the RNA extraction. Tissue sections made from diagnostic histological paraffin blocks were used for protein analysis. TMA immunohistochemistry was performed using 16 overlapping and 103 independent (not analysed on Affymetrix microarrays) set of patients making up a sum of 289 samples.

2. Microarray analyses

Total RNA was isolated from fresh frozen colonic biopsy samples. After the amplification and fluorescent labelling of the total RNA samples, Atlas cDNA microarray analysis was done from 22 sample pairs (6 IBD, 6 CRC, 10 adenoma and the corresponding healthy mucosa) and Affymetrix whole genomic oligonucleotide microarray analysis was performed
using 52 biopsy samples (8 healthy, 9 UC, 5 CD, 7 early CRC, 8 advanced CRC, 6 non-dysplastic adenoma, 9 dysplastic adenoma).

3. Statistical evaluation

a. Evaluation of cDNS microarray data. Scanned arrays were evaluated by the GenePix Pro 4.1 software. Automated spot detection using local background determination was done and feature extraction (ratio of medians, ratio of means, Cy3/Cy5) was performed. One-way (group) ANOVA and multivariate exploratory techniques (discriminant analysis, factor analysis and hierarchical cluster analysis) were performed by SAS 6.12 version statistical software. Hierarchical cluster analysis was done using Ward's method (Euclidean distances). Functional analysis and visualization of biological association network were done using Pathway assist 2.53 software.

b. Evaluation of Affymetrix oligonucleotide microarray data

After quality control analyses, two independent normalization methods (RMA and MAS5.0) were applied. PAM method (“Prediction Analysis for Microarrays”) was used for feature selection. Hierarchical cluster analysis using Genesis software was performed for visualization of discriminatory gene expression patterns. For the discriminant analysis of reduced variables SPSS 15.0 program was used.

In order to identify the adenoma-dysplasia-carcinoma sequence specific genes, for the pre-processing of the microarray data RMA background correction, quantile normalization and median polish summarization were applied. The disease-specific (healthy, non-dysplastic adenoma, dysplastic adenoma, early and advanced CRC) mRNA expression values in case of each transcripts were determined from the pre-processed microarray data. Kendall’s rank correlation analysis was performed for quantification of association between the expression level and the disease stages.

4. RT-PCR validation

The microarray results were confirmed using real-time RT-PCR. EGFR (epidermal growth factor receptor) and TOP1 (DNA topoisomerase 1) one-step real-time RT-PCR analysis (Roche LightCycler) were used for verification of the cDNA microarray results. Evaluation of relative ratios (diseased/normal/same patient) was prepared using RelQuant software. For the verification of the whole genomic microarray data, expression of 52 selected genes was measured using Taqman real-time PCR and Applied Biosystems Microfluid Card System. The SDS 2.2 software was used for data analysis. The extracted delta Ct values (which
represent the expression normalized to the ribosomal 18S expression) were grouped according to the histological groups, then the Student’s t-test was performed to compare the expression values between groups.

5. Protein level analysis
For protein level analysis of potential marker molecules, conventional EGFR immunohistochemistry, furthermore osteopontin and osteonectin tissue microarray immunohistochemistry were prepared using 289 tissue samples of 119 patients. Pearson khi-test and Fischer exact test were used for statistical evaluation of correlation between the protein marker expression and disease stage.

RESULTS

1. Identification of over- or downregulated genes in colorectal diseases

- CRC cases are characterized by upregulated genes in the DNA replication, cell cycle, extracellular matrix remodelling, transcription regulation, oncogenesis and growth factor related cell proliferation cell function groups; and downregulated genes in the DNA repair, tumor suppression and apoptosis cell function groups. Several cell adhesion- and transport-related genes were also found to be differentially expressed in CRC compared to the normal mucosa. In addition, altered expression of some immune regulatory and angiogenesis genes was also detected in CRC.

- Adenoma cases showed altered gene expression data in transport (like ABCA8, TRPM6), adhesion (such as CXCL12, CD44, ADAM-like decysin 1, claudin-1, integrin alpha 6), metabolism (like carbonic anhydrase I, phosphodiesterase-3A), cell proliferation (such as MET protooncogene, CXCL2, CXCL3 oncogenes, tumour-associated calcium transducer 2) and apoptosis functional groups.

- IBD cases are mainly featured by the gene expression changes of immune regulation, cell proliferation, transport processes and metabolism. Altered expression of genes involved in extracellular matrix remodeling and intracellular signal transduction were also detected.
2. Gene expression markers of the colorectal adenoma-dysplasia-carcinoma sequence

- In line with colorectal adenoma-dysplasia-carcinoma transition, progressive overexpression of 918 genes (p<0.002) and continuous downregulation of 382 genes were detected (p<0.002).

- The expression of 17 potential markers was confirmed using real-time PCR. 15 genes were identified showing significant and progressively increasing gene expression along with the adenoma-dysplasia-carcinoma sequence progression. Ten of them are novel tissue markers which show continuously increasing mRNA expression in line with the colorectal adenoma-dysplasia-carcinoma transition. These are the following: tissue inhibitor of metalloproteinases-1 and -3, von Willenbrand factor, interleukin 8, melanoma cell adhesion molecule, thrombospondin 2, collagen 4A1, matrix Gla protein, interleukin 1 receptor antagonist and calumenin.

- Two genes were found to show significantly decreasing expression (p<0.05). The prostaglandin D2 receptor and the amnionless homolog are novel, validated sequentially downregulated markers of colorectal adenoma-dysplasia-carcinoma sequence.

- Protein expression of two markers (osteopontin and osteonectin) was analysed using tissue microarray immunohistochemistry. The sequential overexpression of both proteins significantly correlates with the progression of colorectal adenoma-dysplasia-carcinoma sequence.

3. Discrimination of colorectal disease subtype

- I have determined the top100 most differentially expressed discriminatory genes which are suitable for molecular-based discrimination of early and advanced colorectal cancer. The metabolic and trasport processes mainly differ between CRC subgroups. In advanced stages of CRC downregulation of apoptosis and immune response was observed, while carbohydrate, fatty acid metabolism and energy metabolism related genes showed higher mRNA expression levels in parallel with CRC progression.

- In adenoma upregulation of proliferation, DNA replication and transcription, and downregulation of immune and defense response were found during the development of dysplastic alterations.
• CD cases are mainly featured by increased expression of carbohydrate metabolism genes, while certain cell proliferation, apoptosis, immune regulation, transport and ubiquitin-dependent protein catabolism genes were found to be overexpressed in UC compared with CD cases.

4. Reducing of the discriminatory mRNA expression marker groups

a. Clontech cDNA microarray system

• 17 genes were found to be significantly differently expressed (p < 0.05) between the colonic sample groups (adenoma, CRC, IBD) using filtered ANOVA method. Hierarchical diagram of 22 colonic cases based on ANOVA genes showed considerable accordance with the conventional histopathological diagnoses.

• Two functional gene groups (two factors with considerable explorative variance) were identified using factor analysis which can differentiate the observed diseases according to their different expression levels.

• According to the expression changes of the following genes identified using discriminant analysis, the three colonic disease groups can be significantly distinguished: HSF1, bystin-like, calgranulin A and TNFR superfamily member 10c.

b. Affymetrix oligonucleotide microarray system

• The top100 most differentially expressed discriminatory genes were determined in each diagnosis, and the overlapping genes of two lists created after application of two different normalization method were identified (overlapping discriminatory genes). During my analyses, I have identified the top 27, 13 and 10 genes associated with adenoma, CRC and IBD, respectively.

• 96.2% of the samples were correctly classified using 7 discriminatory genes determined by step-wise discriminant analysis (indoleamine-pyrrole 2,3 dioxygenase, ectodermal-neural cortex, TIMP3, fucosyltransferase 8, collectin sub-family member 12, carboxypeptidase D, and transglutaminase 2 (Figure 2.).

5. Validation of microarray results

• Similarly with the results of the cDNA microarray analysis, the EGFR and TOP1 mRNA levels of colorectal cancer samples were significantly higher than of their normal corresponding mucosa pairs (relative ratio:3.008, p=0.0487, and relative ratio:2.800, p=0.0492). Elevated EGFR protein expression in CRC compared to
normal samples were in correlation with EGFR mRNA expression data from both microarray and real-time RT-PCR analysis.

- Forty six of the 52 measured genes correlated with the results obtained using Affymetrix microarrays at a significance of p<0.05.

CONCLUSIONS

In conclusion, detection of the mRNA expression levels of marker gene panels gives an opportunity for classification of colonic samples, even in the case of small biopsy specimens. Normal, adenoma, CRC and IBD samples, but also the different stages of CRC can be classified by gene expression analysis of easily-taken biopsy specimens.

In my PhD work, I have also identified the characteristic progression markers of the colorectal adenoma-dysplasia-carcinoma sequence. The osteopontin and the osteonectin are the leading CRC progression markers as validated using RT-PCR and TMA on both overlapping and independent samples. Considering the positive regulatory roles of osteopontin in vascular remodeling and its immunological effects, and of osteonectin in the regulation of cell adhesion and proliferation, these proteins could be potential targets of antineoplastic therapy.

With the help of gene expression patterns associated to histological results a more exact diagnostics may be established. Moreover, in case of histologically non evaluable samples, or uncertain clinical diagnosis, or invasion and micrometastases the histological results may be supported. With a large number of samples one can more accurately establish principal gene lists that characterize distinct conditions. Testing of the potential markers on an independent set of samples is one of the elementary requirements of the mRNA expression-based diagnostics development. In addition, the diagnostic routine will be easier, the prognosis and therapeutic response will also be predictable by creation of disease-specific so-called miniarrays with small number of diagnostic genes.
THE MOST IMPORTANT NEW STATEMENTS

- The oligonucleotide whole genomic microarray analyses of biopsy samples eminently fulfill the Affymetrix quality requirements, and are highly standard, and reproducible.

- The Taqman Microfluidic Card System which offers opportunity for analysis of expression level of 96 genes on 10-100 samples, is particularly suitable for high-throughput, quick and cost efficient RT-PCR validation of gene expression changes detected by microarrays.

- The sequential overexpression of osteopontin and osteonectin mRNAs and proteins significantly correlates with the progression of colorectal adenoma-dysplasia-carcinoma sequence.

- I have identified and validated by RT-PCR ten novel tissue markers which show continuously increasing mRNA expression in line with the colorectal adenoma-dysplasia-carcinoma transition. These are the following: tissue inhibitor of metalloproteinases-1 and -3, von Willenbrand factor, interleukin 8, melanoma cell adhesion molecule, thrombospondin 2, collagen 4A1, matrix Gla protein, interleukin 1 receptor antagonist and calumenin.

- The prostaglandin D2 receptor and the amnionless homolog are novel, validated sequentially downregulated markers of colorectal adenoma-dysplasia-carcinoma sequence.

- During my analyses, I have identified the top 27, 13 and 10 genes associated with adenoma, CRC and IBD, respectively.

- Using whole genomic microarrays I have also determined the top100 most differentially expressed discriminatory genes which are suitable for molecular-based discrimination of early (Dukes A and B stage) and advanced (Dukes C and D) colorectal cancer.

PUBLICATIONS in connection with the PhD dissertation


**PUBLISHED ABSTRACTS in connection with the PhD dissertation**


Galamb O, Sipos F, Molnár B, Kopper L, Dinya E, Tulassay Z. Diagnosis of malignancy in colon biopsy specimen by mRNA expression profiling. *Gastroenterology*, 126(4): A244-A244 Suppl. 2 Apr 2004

ACKNOWLEDGEMENT

Hereby I would like to thank everybody who helped me in my PhD work:
- my program leader, Zsolt Tulassay MD. DSc. and my supervisor, Bélá Molnár MD. PhD. for allowing and supporting my PhD work in Semmelweis University 2nd Department of Internal Medicine;
- Miklós Garami MD. PhD. and Tamás Ponyi PhD. opposers, and Jeanette Molnár MD. PhD. and Katalin Müllner MD. PhD. opposers for the review of my PhD dissertation;
- Tibor Krenács PhD. for the professional support and for his indispensable help in tissue microarray analyses;
- Lajos Berczi MD. PhD. for his useful advices and directions during the personal consultation;
- Norbert Solymosi DVM., Balázs Györffy MD. PhD. and Elek Dinya PhD. for the professional implementation of bioinformatic analyses
- Marica Csorba and Gabriella Kónya for their help in the laboratory experiments;
- every employee of the Cell Analysis Laboratory for their help;
- Ferenc Sipos MD. PhD. for his professional support and for his friendship;
- and last, but not least I thank my Husband and my Family for taking my side every time.