Disease Markers 28 (2010) 1–14
DOI 10.3233/DMA-2010-0677
IOS Press

Applicability of antibody and mRNA expression microarrays for identifying diagnostic and progression markers of early and late stage colorectal cancer

Sándor Spisák[a,b,*], Barnabás Galamb[a], Ferenc Sipos[a], Orsolya Galamb[a,b], Barnabás Wichmann[a], Norbert Solymosi[a], Balázs Nemes[c], Jeannette Molnár[d], Zsolt Tulassay[a,b] and Béla Molnár[a,b]
[a]*2nd Department of Medicine, Semmelweis University, Budapest, Hungary*
[b]*Molecular Medicine Research Unit, Hungarian Academy of Sciences, Budapest, Hungary*
[c]*Transplantation and Surgical Department, Semmelweis University, Budapest, Hungary*
[d]*Department of Nutrition, National Institute of Food and Nutritional Science, Budapest, Hungary*

Abstract. The exact molecular background and the connection between protein and mRNA expression in colorectal cancer (CRC) development and progression are not completely elucidated. Our purposes were the identification of protein markers of colorectal carcinogenesis and progression using protein arrays and validation on tissue microarrays. The connection between antibody and mRNA expression array results was also examined. Using cancerous and adjacent normal samples from 10 patients with early and 6 with advanced CRC, 67 differentially expressed genes were identified between normal and cancerous samples. A marker set containing 6 proteins (CCNA1, AR, TOP1, TGFB, HSP60, ERK1) was developed which could differentiate normal specimen, early and late stage CRC with high sensitivity and specificity. Dukes D stage samples were analyzed on HGU133plus2.0 microarrays. In these samples, mRNA and protein expression of 143 genes showed strong positive correlations ($R^2 > 0.8$), while a negative correlation ($R^2 > 0.9$) was found in case of 95 genes. Based on our results a correlation could be established between transcriptome and antibody array results, hence the former may be used as a high-capacity screening method before applying antibody arrays containing already planned targets. Antibody microarrays may have a fundamental importance in testing of marker combinations and future application in diagnostics of tumorous diseases.

Keywords: Colorectal cancer, progression marker, antibody microarray, gene expression, transcriptomics, proteomics, bioequivalency

1. Introduction

Colorectal cancer (CRC) is the second most frequent cause of tumor related death in Western countries [1,2]. In the molecular biological background of CRC, different genetic pathways are known to be affected: changes in DNA like SNP, insertion, deletion, duplication, methylation, on the other hand changes of the regulation of transcription and translation such as splicing, miRNAs, phosphorylation, glycosylation, complex formation and different interactions [3–11]. One of the best known pathways is the adenoma-dysplasia-carcinoma sequence, in which dysplastic crypts develop from normal epithelium after APC mutation, followed by an adenoma. During the process, chromosomal instability and aneuploidy may also be present [11–13]. Chromosomal or microsatellite insta-

*Corresponding author: Sándor Spisák, 2nd Department of Medicine Semmelweis University, Szentkirályi str. 46, 1088 Budapest, Hungary. Tel.: +36 1 266 09 26; Fax: +36 1 266 08 16; E-mail: spisak@abc.hu.

ISSN 0278-0240/10/$27.50 © 2010 – IOS Press and the authors. All rights reserved

bility and the presence or absence of CpG island methylator phenotype establish the basis of molecular pathological classification. These characteristics may be used successfully during the examination of left and right sided colon tumours or when selecting a chemotherapy protocol [1,2,14]. However, the association between these molecular events and protein expression alterations has not yet been clarified.

Whole genomic arrays provide an excellent opportunity for the development of molecular expression pattern based colorectal cancer classification [3,7,10]. The diagnostic application of these patterns has also emerged as a possibility [15–18] according to the results of our research team. Using array technologies, the discrimination of early and late stage colorectal cancers [14], the prediction of metastatic and nonmetastatic disease [19] as well as the characterization of the adenoma-dysplasia-carcinoma sequence became possible [20,21].

Array techniques can also be used for the protein level based description of molecular changes leading to tumour development [15]. In recent years, besides conventional MS and 2D elfo techniques, protein chips have become available. This technology results not just in qualitative expression patterns, but may also be used to determine the quantity of the examined protein. An additional advantage of protein arrays is their easier sample preparation compared to the other two classical proteomics technologies.

The molecular changes that lead to cancer development may be explored by using several techniques at mRNA and protein level, however the connections between these levels are not fully understood. The lack of permeability between technological platforms hinders the comparison of results. Array techniques allow us to perform mRNA and protein analysis from the same samples, hence the discovery of the coherence between results may become closer.

Regarding the correlation between protein and mRNA expression chips, there are no definite literature data, not even from cell culture studies. In yeast, the detection of protein and mRNA profile similarities failed [22]. This may be explained by the time frame between transcription and translation, the mRNA processing, the half-life time and degradation of these macromolecules. To our knowledge, there are hardly any published studies focusing on the associations between mRNA and protein levels, which may be important in case of tumorous diseases, using only microarray techniques. There are already a few data available from transcript profile analyses that were simultaneously prepared using protein level MS, bead-array, and ELISA examinations. Previously, a positive correlation of 32 percent was reported by the parallel analysis of 1200 proteins and transcripts [23].

In this study we examined how the changes at mRNA level, modified during disease development and progression, influence protein expression, and whether these altered protein expression patterns may be used for the molecular classification of colorectal cancer (Fig. 1).

2. Materials and methods

2.1. Sample collection

Surgically removed tissue samples from the diseased and the healthy parts of the colon of 16 colorectal cancer patients (10 Dukes B, 6 Dukes D stage, Table 1.) were collected. Maximum 10 minutes after resection 500 mg tissue were snap frozen in liquid nitrogen both from the intact mucosa, close to the resection line and from the tumorous tissue. Samples were then stored at -80°C. In parallel, formalin fixed, paraffin embedded tissue samples were also prepared from the same samples for both conventional histopathological diagnosis and tissue microarrays.

2.2. Protein isolation

Frozen samples were homogenised in lysis buffer, then 1 g of Alumina (Sigma Aldrich) was added to each gramm of initial material. After extensive homogenisation, samples were centrifuged at 25000 g for 10 minutes. The supernatant containing the crude extract was stored in 100 μl aliquots at -80°C until utilization.

2.3. Clontech AB 500 array analysis

The protein concentrations of the crude extracts were measured with the BCA method. 1 mg/ml working solutions were prepared from all samples. Cy3 and Cy5 monoreactive dyes were used for protein labelling according to the manufacturer's description (Amersham/GE Healthcare). Labelling was carried out with the dye swap method. In reaction A, the tumorous sample was labelled with Cy5 (red), and the normal sample with Cy3 (green). Afterwards, in reaction B, the tumorous sample was labelled with Cy3 and the normal sample with Cy5. Labelled protein cleaning

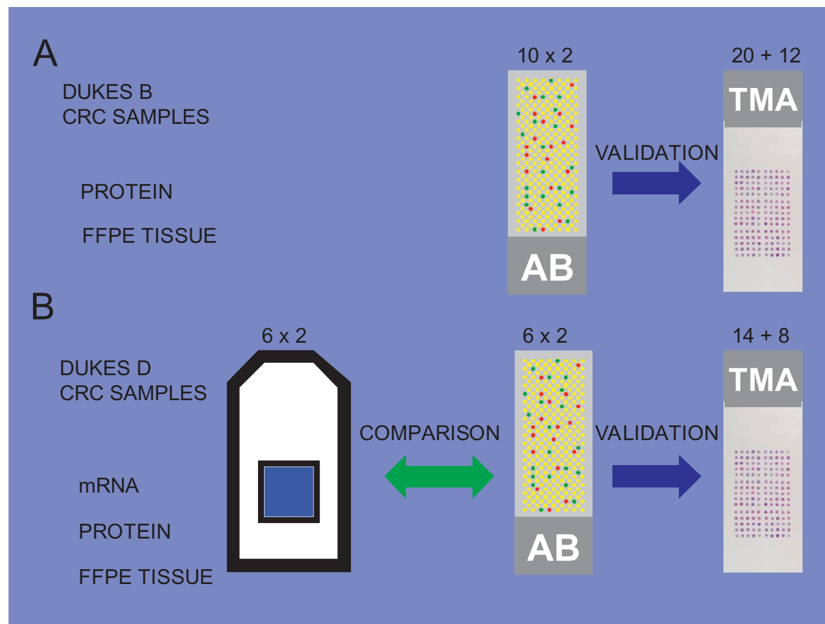

Fig. 1. Depicting of study design. Part A represents the analysis process of Dukes B stage colorectal cancer samples. Fresh frozen surgical tissue samples from tumorous and healthy colonic region of 10 CRC patients were analyzed on antibody arrays. TMA validation was performed on an extended set of 20 tumorous and 12 healthy FFPE samples. Part B shows preparation procedure of 6 Dukes D stage samples. Similarly to the early stage sample analysis, protein expression profiling was done on both protein arrays and TMAs. Furthermore, whole genomic mRNA expression pattern was determined by Affymetrix HGU 133 Plus 2.0 microarrays in order to examine the associations between transcriptome and proteome.

and the removal of redundant dyes were performed on a CD10 column (Amersham/GE Healthcare).

The concentrations of the labelled and cleaned proteins were also measured with the BCA method. Next, equivalent amounts of differently labelled proteins (100–100 μg from each variant) were mixed. The hybridization mix was prepared by using 30 μg of the mixture. The hybridization was performed on Clontech AB 500 (Clontech) arrays according to the manufacturer's description, then the arrays were washed with absolute ethanol to remove non-binding material. After drying centrifugation, protein arrays were stored in dark until scanning.

Scanning was carried out on an Axon 4000B scanner (Axon Instruments, USA) at 532 and 635 nm wavelengths. Array analysis was performed using the GenePix 4.1 software at 33 percent laser force, 560 and 670 PMT gain, and 20 m/pixel resolution. After placing the array net (Lot: 5010317), fluorescence and background intensities were determined on both channels. Then a local background correction was performed. Data were exported in gpr and csv formats. Data analysis was carried out using the R-environment. After processing (RMA, normexp), array results were normalized with the quantile method.

The identification of genes with expression difference was performed by using the normalized M values / $M = \log_2$(R/G); R: red fluorescence intensity, G: green fluorescence intensity/.

The limit of significant difference in case of protein markers is at $M = \pm 0.5$ which represents an alteration in expression of at least 30 percent.

2.4. *Tissue microarray analysis*

The results of the protein array analysis were validated by TMA on an extended set of samples. Apart from 15 healthy tissue samples, 36 colorectal cancer specimen from different localizations, grade of differentiation and Dukes stage were included in the validation process. Cores of 2 mm diameter were collected and placed together into 24 samples recipient blocks. Five μm thick tissue sections were cut from the TMA blocks, mounted on adhesive glass slides and immunostained following endogenous hydrogen peroxidase blocking (0.5% H_2O_2-methanol) and heat-induced epitope retrieval in 150 ml of pH 6.0 TRS buffer (Target Retrieval Solution, S1699) using a commercial microwave oven at 300 W power for 45 minutes. The following immunohistochemistries

Table 1
Patient and histological data

Sample ID	Age	Sex	Localization	Dukes stage	Histology, differentiation	Antibody array	Affymetrix	TMA
					CRC			
5	80	F	rectum	B	well differentiated adenocarcinoma	+		+
8	58	F	sigma	D	moderately differentiated adenocarcinoma	+	+	+
11	66	M	sigma	B	moderately differentiated adenocarcinoma	+		+
12	56	F	sigma	D	moderately differentiated adenocarcinoma	+	+	+
13	69	M	sigma	D	poorly differentiated adenocarcinoma	+	+	+
15	72	M	rectum	D	poorly differentiated adenocarcinoma	+	+	+
16	82	M	rectum	B	moderately differentiated adenocarcinoma	+		+
18	55	M	rectum	B	well differentiated adenocarcinoma	+		+
20	62	M	rectum	B	moderately differentiated adenocarcinoma	+		+
22	64	F	sigma	B	well differentiated adenocarcinoma	+		+
31	52	M	rectum	B	well differentiated adenocarcinoma	+		+
32	59	F	rectum	B	moderately differentiated adenocarcinoma	+		+
33	70	F	rectum	D	moderately differentiated adenocarcinoma	+	+	+
39	62	M	sigma	D	poorly differentiated adenocarcinoma	+	+	+
45	59	M	sigma	B	well differentiated adenocarcinoma	+		+
46	83	M	rectum	B	moderately differentiated adenocarcinoma	+		+
49	56	M	rectum	D	moderately differentiated adenocarcinoma			+
50	69	M	rectum	D	poorly differentiated adenocarcinoma			+
56	63	M	rectum	B	poorly differentiated adenocarcinoma			+
58	72	F	sigma	B	poorly differentiated adenocarcinoma			+
61	58	M	sigma	B	moderately differentiated adenocarcinoma			+
64	73	M	sigma	D	poorly differentiated adenocarcinoma			+
66	67	M	sigma	B	moderately differentiated adenocarcinoma			+
69	57	F	rectum	B	well differentiated adenocarcinoma			+
70	64	F	sigma	B	moderately differentiated adenocarcinoma			+
74	74	M	sigma	D	poorly differentiated adenocarcinoma			+
78	68	F	rectum	B	poorly differentiated adenocarcinoma			+
86	82	M	rectum	B	well differentiated adenocarcinoma			+
91	79	M	rectum	D	moderately differentiated adenocarcinoma			+
94	84	M	rectum	D	moderately differentiated adenocarcinoma			+
97	56	F	sigma	B	poorly differentiated adenocarcinoma			+
100	62	M	sigma	D	poorly differentiated adenocarcinoma			+
110	75	M	rectum	B	well differentiated adenocarcinoma			+
115	65	F	sigma	D	poorly differentiated adenocarcinoma			+
					Normal			
5	80	F	rectum		normal	+	+	+
8	58	F	sigma		normal	+		+
11	66	M	sigma		normal	+	+	+
12	56	F	sigma		normal	+	+	+
13	69	M	sigma		normal	+	+	+
15	72	M	rectum		normal	+		+
16	82	M	rectum		normal	+		+
18	55	M	rectum		normal	+		+
20	62	M	rectum		normal	+		+
22	64	F	sigma		normal	+		+
31	52	M	rectum		normal	+		+
32	59	F	rectum		normal	+	+	+
33	70	F	rectum		normal	+	+	+
39	62	M	sigma		normal	+		+
45	59	M	sigma		normal	+		+
46	83	M	rectum		normal	+		+
49	56	M	rectum		normal			+
50	69	M	rectum		normal			+
56	63	M	rectum		normal			+
58	72	F	sigma		normal			+

were performed: APC, Caveolin, CBP, cyclinA, ERK, HSP60, Cox2, EGFR, C-myc, Cald, Top1, DARPP32, MRE11A, AndrogenR, EPS8.

Immunostained TMA slides and blood smears were digitalized using a high-resolution Mirax Desk instrument (Zeiss, Goettingen, Germany) and the stained spots were analysed with the Mirax TMA Module software (Zeiss). Protein expression was evaluated using an empirical scoring system considering both intensity and frequency in epithelial/lamina propria cells. In case of cytoplasmatic proteins (APC, Caveolin, CBP, ERK, HSP60, Cox2, C-myc, Cald1, DARPP32, MRE11A, AndrogenR and EPS8), the score value was -2 for negative staining; 0 for weak, 1 for moderate and 2 for intensive strong, diffuse cytoplasmatic immunostaining. In case of nuclear proteins (cyclinA1, Top1), the score value was -2 for a negative immunoreaction, 0 if the ratio of immunopositive cells was under 2.5%, 1 for a ratio between 2.5% and 5%, and 2 if positive cells were more than 5%. In case of EGFR, a -2 score represents negative staining, 0–1 represents weak-moderate cytoplasmatic staining, and 2 represents strong cell membrane staining.

To measure the association of the two categorical variables (group and score) contingency tables were constructed and tested by Chi-square test. If the difference was statistically significant ($p < 0.05$), a more detailed analysis was visualized on the basis of the Pearson residuals. These results were summarized in a graphical association plot [http://gap.stat.sinica.edu.tw/Software/GAP].

2.5. Immunohistochemistry

After dewaxing, antigen retrieval was carried out by microwave oven heating (at 750W) in citrate buffer for 20 minutes. Signal conversion was performed with standard indirect immunoperoxidase reaction, where diamino-benzidine was used as chromogen substrate (Dako). For immunohistochemical labellings, Abgene, Dako antibodies were used according to the description of the manufacturer.

2.6. Affymetrix whole genome expression analysis

From the homogenised samples that were collected and snap frozen in RNAlater, total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Quantity and quality of the isolated RNA were tested measuring the absorbance by agarose or capillary gelelectrophoresis using the 2100Bioanalyzer and RNA 6000 Pico Kit (Agilent Inc, Santa Clara, USA). Biotinylated cRNA probes were synthesized from 5–8 μg total RNA and fragmented using GeneChip cDNA synthesis reagents, a sample cleaning kit and the IVT Labeling Kit [http://www.affymetrix.com/support/downloads/manuals/expression_s2_manual.pdf – 1 version] according to the Affymetrix instructions. Ten μg of each fragmented cRNA sample were hybridized onto HGU133 Plus2.0 arrays (Affymetrix) at 45°C for 16 hours. The slides were washed and stained using a Fluidics Station 450 and an antibody amplification staining method was applied according to the manufacturer's instructions. The fluorescent signals were detected by a GeneChip Scanner 3000.

2.7. Statistical analysis

Quality control analyses were performed according to the guidelines of The Tumour Analysis Best Practices Working Group [19]. Scanned images were inspected for artifacts, percentage of present calls ($> 25\%$) and the degree of RNA degradation were evaluated. Based on the evaluation criteria all biopsy measurements fulfilled the minimal quality requirements. The Affymetrix expression arrays were pre-processed by gcRMA with quantile normalization and median polish summarization. For feature selection linear models combined with empirical Bayesian methods were used [24].

2.8. Correlation analysis between Affymetrix and protein array data

Based on the Swissprot ID of the antibody array targets, Affymetrix IDs for 481 genes were assigned by using the Netaffx database. One Affymetrix ID was assigned to each protein. In case of several Affymetrix IDs, the 3' end probe set was preferred. The changes in expression (M values) of both mRNA and protein were plotted in a frame of reference, then correlation analysis was performed. For the selection of the best correlating pairs the distance of the pair from zero and the size of their quotient were taken as a basis.

3. Results

3.1. Discrimination of normal and tumorous groups (normal vs. Dukes B and D)

Tumorous groups were collectively compared to the normal samples, and the list of differentially expressed

Table 2
Top 20 significantly altered protein between normal and colorectal samples

Gene name	SwissProt ID	Biological function	M value	adj,P,Val
ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast)	P51965	protein degradation	−1.319	0.002
sequestosome 1	Q13501	cell differentiation	−1.044	0.005
KH domain containing, RNA binding, signal transduction associated 1	Q07666	cell proliferation	−0.733	0.008
branched chain aminotransferase 1, cytosolic	P54687	cell proliferation	−0.719	0.008
protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein, DARPP-32)	Q9NNW1	signal transduction	−0.616	0.035
syntaxin 8	Q9UNK0	transport	−0.541	0.029
plectin 1, intermediate filament binding protein 500kDa	Q15149	cytoskeletal anchoring	−0.478	0.005
nitric oxide synthase 1 (neuronal)	P29475	cell-cell signaling	−0.408	0.025
adenomatosis polyposis coli	P25054	signal transduction	−0.397	0.026
caldesmon 1	Q05682	muscle contraction	−0.392	0.022
proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki)	Q12920	immune response	0.389	0.033
serine/threonine kinase 24 (STE20 homolog, yeast)	Q9Y6E0	signal transduction	0.431	0.003
topoisomerase (DNA) I	P11387	DNA topological change	0.435	0.008
thioredoxin-like, 32kDa	O43396	apoptosis	0.474	0.001
guanine nucleotide binding protein (G protein), beta polypeptide 1	P04901	signal transduction	0.478	0.004
F11 receptor	Q9Y624	inflammatory response	0.529	0.005
cyclin A1	P20248	cell cycle	0.543	0.002
nuclear mitotic apparatus protein 1	Q14980	cell cycle	0.571	0.004
CDC-like kinase 1	P21127	cell proliferation	0.578	0.003
heat shock 60kDa protein 1 (chaperonin)	P10809	protein folding	0.635	0.004

genes was compiled (Table 1.) Between the two groups (normal vs. CRC) 67 differentially expressed genes were found. Based on their function, they can be classified into apoptosis (5), cell cycle regulation (7), transcription regulation (4), DNA replication (4), molecular transport and cell adhesion (45) related genes. The top 10 over- and underexpressed genes are functionally classified in Table 2.

3.2. *Discrimination of normal and Dukes B groups: protein markers of early cancer progression*

In group Dukes B, 22 genes were found to be significantly altered from normal mucosa (Table 3). Nine of them were underexpressed, while 13 showed upregulation compared to normal. They were classified into the following functional groups: ubiquitin cycle (1), signal transduction (8), meiotic recombination (1), viral envelopment protein (1), cell proliferation (1), mitosis (3), oxidoreductase and lyase activity (1), transcription factor (1), unknown (1), DNA topological change (1), cell motility (1) and transport (1). In some samples, the presence of HPV-16 virus was detected, but as it was not present in all samples, it was excluded of further analysis.

3.3. *Discrimination of normal and Dukes D groups*

In group Dukes D, 25 genes were found to be significantly altered from normal mucosa. Thirteen of them were underexpressed, while 12 showed upregulation compared to normal. They were classified into the following functional groups: ubiquitin cycle (1), signal transduction (7), viral envelopment protein (1), DNA repair (3), RNA binding (1), cell proliferation (3), regulation of translation (1), ubiquitinylation (1), transcription factor (1), cell differentiation (2), cell cycle (1), transport (2) and mitosis (1). The percentual change of expression and the cell function of the identified genes can be seen in Table 4.

3.4. *Identification of progression markers*

The following two strategies were applied for the identification of progression markers. First, results from Dukes B and Dukes D samples were directly compared with each other. Second, data from separately performed analyses were compared in order to determine how the expression of one gene altered already in the early stage of disease changes in the late stage. Using the first method, 58 genes were identified between the early and late stage CRC samples out of which 11 showed a significant expression alteration between the two groups (Table 5).

Using the second method, the underexpressed genes in Dukes D stage could be classified into three subgroups based on their expression status in Dukes B stage. There were 8 genes in the first subgroup (HPV-16, NEDD4, MRE11A, UBE2E1, EPHA4, KHDRBS1, SQSTM1, PPP1R1B), that already showed

Table 3
Significantly altered proteins between healthy colon and early stage cancer samples

	Gene name	Biological function	SwissProt ID	M value
1	ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast)	Ubiquitin cycle, Ubiquitin-dependent protein catabolism	P51965	−1.865995434
2	sequestosome 1	Response to stress, Ubiquitin binding, Endosome transport, Intracellular signaling cascade	Q13501	−1.728747978
3	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	Meiotic recombination, Meiosis, Double-strand break repair via nonhomologous end-joining	P49959	−1.5668813
4	HPV-16 L1	Viral envelopment protein	P03101	−1.170942471
5	protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein, DARPP-32)	Signal transduction, Protein phosphatase inhibitor activity, protein kinase inhibitor activity	Q9NNW1	−1.127870256
6	KH domain containing, RNA binding, signal transduction associated 1	RNA binding, signal transduction	Q07666	−0.99810321
7	branched chain aminotransferase 1, cytosolic	Cell proliferation	P54687	−0.759377286
8	EphA4	RNA binding, signal transduction	P54764	−0.620030199
9	neural precursor cell expressed, developmentally down-regulated 4	Structural molecule activity, Nuclear organization and biogenesis, Mitotic anaphase	P46934	−0.618087113
10	prenylcysteine lyase	Oxidoreductase activity, Lyase activity	Q9UHG3	0.501257698
11	thioredoxin-like, 32kDa	Electron transporter activity, Thiol-disulfide exchange intermediate activity, Apoptosis, Signal transduction	O43396	0.514952034
12	casein kinase 2, beta polypeptide	Wnt receptor signaling pathway, Protein serine/threonine kinase activity	P13862	0.532917948
13	heat shock transcription factor 4	Transcription corepressor activity, Transcription factor activity, Response to unfolded protein, Transcription, Protein folding	Q9ULV5	0.534018124
14	nuclear mitotic apparatus protein 1	Mitosis	Q14980	0.54577123
15	polyamine-modulated factor 1		Q9UBQ3	0.551012235
16	cyclin A1	Mitosis, Regulation of cyclin dependent protein kinase activity	P20248	0.557753319
17	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	Epidermal growth factor receptor signaling pathway, Negative regulation of cell cycle, ATP binding, Transferase activity	P00533	0.564085732
18	CDC-like kinase 1	Cell cycle	P21127	0.574731447
19	topoisomerase (DNA) I	DNA topological change, DNA unwinding	P11387	0.577033093
20	guanine nucleotide binding protein (G protein), beta polypeptide 1	Signal transducer activity, G-protein coupled receptor protein signaling pathway	P04901	0.60497692
21	F11 receptor	Cell motility, Inflammatory response	Q9Y624	0.698497879
22	heat shock 60kDa protein 1 (chaperonin)	Unfolded protein binding, Mitochondrial matrix protein import	P10809	0.776671019

downregulation in early stage CRC. Underexpressed genes in Dukes D stage that showed no significant downregulation in Dukes B stage (EIF4E, HSP90AA1, AR, LCP2) formed the second subgroup. The only gene in the third subgroup was ATP1B2, which was slightly overexpressed in Dukes B stage.

Those genes that were overexpressed in Dukes D stage were also classified into three subgroups. EGFR and NUMA1 were in the first subgroup: they were already overexpressed in early stage of CRC. Genes that were only slightly overexpressed in early CRC (CAV1, TOP2B, EPS8, NPAT, STXBP1, SEMA4D) formed the second subgroup. The third subgroup contained genes that were overexpressed in Dukes D stage but underexpressed in Dukes B stage (GAP43, NCF2, POLE).

All genes that were underexpressed in Dukes B stage were also underexpressed in Dukes D stage, and 9 of the upregulated genes in early CRC were also overexpressed late stage CRC. In case of 4 genes (HSPD1, TOP1, PMF1, HSF4), only protein levels increased during disease progression, their gene expression, however, was not altered.

3.5. *TMA verification of antibody array results*

3.5.1. *Normal vs. tumour*

The upregulation of TOP1, CCNA1 and HSP60, and the downregulation of APC were validated by TMA. In case of v-myc, the expression decrease detected by protein chip was significant ($p = 0.019$), but considerably lower than the expression decrease validated by TMA immunohistochemistry. In case of AR, the results of the protein chip and TMA did not agree.

3.5.2. *Normal vs. early stage CRC*

Six genes (TOP1, HSP60, CYCA1, NUMA1, CALD1, EGFR) were selected for TMA validation

Table 4
Significantly altered proteins between normal mucosa and late stage colorectal cancer samples

	Gene name	Biological function	SwissProt ID	M value
1	ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast) (UBE2E1)	Ubiquitin cycle, Ubiquitin-dependent protein catabolism	P51965	−1.3655365
2	sequestosome 1 ()	Response to stress, Ubiquitin binding, Endosome transport, Intracellular signaling cascade	Q13501	−1.2216673
3	HPV-16 L1		P03101	−1.1267989
4	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)	Meiotic recombination, Meiosis, Double-strand break repair via nonhomologous end-joining	P49959	−1.0716098
5	protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein) (DARPP-32)	Signal transduction, Protein phosphatase inhibitor activity, protein kinase inhibitor activity	Q9NNW1	−0.9386422
6	KH domain containing, RNA binding, signal transduction associated 1	RNA binding, signal transduction	Q07666	−0.8630941
7	heat shock 90kDa protein 1, alpha		P07900	−0.8473879
8	eukaryotic translation initiation factor 4E	RNA cap binding, Regulation of translation	P06730	−0.7378523
9	androgen receptor	Steroid binding, Cell proliferation, Regulation of transcription, Protein dimerization activity	P10275	−0.7205401
10	neural precursor cell expressed, developmentally down-regulated 4 (NEDD4)	protein degradation; protein ubiquitinylation.	P46934	−0.6205902
11	EphA4	RNA binding, signal transduction	P54764	−0.5975661
12	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)		Q13094	−0.5729652
13	ATPase, Na+/K+ transporting, beta 2 polypeptide		P14415	−0.5413946
14	epidermal growth factor receptor pathway substrate 8	SH3/SH2 adaptor activity, Cell proliferation, EGFR signaling pathway, Signal transduction, SH3/SH2 adaptor activity	Q12929	0.5035487
15	growth associated protein 43	Cell differentiation, Regulation of cell growth, Neurogenesis, Calmodulin binding	P17677	0.5042365
16	nuclear protein, ataxia-telangiectasia locus		Q13632	0.5087894
17	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	Epidermal growth factor receptor signaling pathway, Negative regulation of cell cycle, ATP binding, Transferase activity	P00533	0.5192014
18	DNA topoisomerase 1		Q	0.5367019
19	Rho GTPase activating protein 1	SH3/SH2 adaptor activity, GTP binding, Rho protein signal transduction, Cytoskeleton organization and biogenesis, Signal transduction	Q07960	0.5396524
20	syntaxin binding protein 1	Vesicle-mediated transport, Vesicle docking during exocytosis	Q64320	0.5473203
21	caveolin 1	Scaffolding protein, molecular transporting	P51636	0.5955298
22	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D	Receptor activity, Neurogenesis, Anti-apoptosis, Cell differentiation, Immune response	Q92854	0.5980394
23	nuclear mitotic apparatus protein 1	may be a structural component of the nucleus	Q14980	0.6672197
24	polymerase (DNA directed), epsilon	DNA binding, Nucleotide binding, DNA repair	Q07864	0.6765401
25	neutrophil cytosolic factor 2 (65kDa, chronic granulomatous disease, autosomal 2)		P19878	0.7448919

from those genes that showed expression alteration between normal and early CRC (Table 3). The elevated expression of TOP1, HSP60 and CYCA1 could be verified by TMA. In case of NUMA1, the results of the immunohistochemical analysis and the antibody array did not agree. Caldesmon expression correlated significantly with the AB array result, but its expression was detected in stromal cells only. The overexpression of EGFR was also validated in early CRC by both AB array and TMA immunohistochemistry, its expression, however, decreases with the progression of disease. In Dukes D stage, its expression did not differ significantly compared to normal according to both applied methods.

3.5.3. Normal vs. late stage CRC

Comparing results from normal samples with Dukes D stage samples, the following genes were examined with TMA: TOP1, AR, EPS8, CAV1, DARPP32 and NUMA1. According to TMA, TOP1 showed a significant overexpression. In case of AR, an increased expression was detectable by TMA, in contrast to our

Table 5
Selected significantly differentially expressed genes during the tumor progression

Gene name	Biological function	SwissProt ID	M value
ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast)	cell proliferation	P51965	−1.314214341
sequestosome 1	signal transduction	Q13501	−1.043584071
HPV-16 L1	transport	P03101	−0.77024562
KH domain containing, RNA binding, signal transduction associated 1	cytoskeletal anchoring	Q07666	−0.733015297
branched chain aminotransferase 1, cytosolic	cell-cell signaling	P54687	−0.718986189
protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein, DARPP-32)	signal transduction	Q9NNW1	−0.616086234
syntaxin 8	muscle contraction	Q9UNK0	−0.541441325
F11 receptor	immune response	Q9Y624	0.528651479
cyclin A1	cell cycle	P20248	0.542708779
nuclear mitotic apparatus protein 1	DNA topological change	Q14980	0.571409175
CDC-like kinase 1	signal transduction	P21127	0.577723106
heat shock 60kDa protein 1 (chaperonin)	protein folding	P10809	0.634592412

results by AB array. The TMA verification of EPS8, CAV1, DARPP32 and NUMA1 failed, as there was no significant change in immunohistochemical expression between normal and late stage CRC.

3.5.4. *Follow-up of tumour progression*

Three genes that were found to be related to tumour progression in our AB array analysis (CBP (EIF4), ERK (MAPK12), v-myc), and 5 genes whose selection was based on literature data (GSTP, IGFR, TGFB, CALD, COX2) but did not change significantly in this study, were analysed by TMA.

ERK showed overexpression, while v-myc was underexpressed. In case of CBP and EGFR, a tendency of downregulation was detected between early and late stage CRC. The results of TMA validation can be seen in Fig. 2.

3.6. *Validation results*

Our results confirm recently published data, according to which using only one marker is not enough/sufficient for the classification of the different stages of CRC. Therefore, we designed marker sets that are appropriate for such classification.

To test the force of differentiation of the marker sets, an additional, separate set of samples was examined by TMA analysis. Using cluster analysis based on the score values, a set of 6 genes was determined (CycA, Ar, Top1, TGFB, Hsp60, ERK), with which discrimination between normal and tumorous groups became possible (Fig. 3). This marker set was tested by discriminant analysis. The distinction between normal and tumorous groups was 100%, and 90.9% between early and late stage CRC cases (Fig. 4).

3.7. *Correlation between mRNA and protein array results*

Of the 500 genes examined, 465 (93%) showed no expression alteration at protein level. Similarly, the mRNA expression of these 465 genes showed no significant change between normal and Dukes D stage (groups). Both platforms detected the same expression changes, which is visualized in Fig. 5. It can be seen that higher oscillation is detected at mRNA than at protein level. A positive correlation was established in case of 143 transcripts ($R^2 > 0{,}8$), most of which belonged to transport proteins. We also identified 95 genes whose expression changed in opposite directions for mRNA and protein levels ($R^2 > 0.9$); the majority of these genes is involved in the regulation of cell functions.

Next, genes with a highly significant expression alteration on both platforms were examined.

Twelve genes were identified (Fig. 6.) that showed a significant difference between normal and Dukes D stage CRC cases on both platforms. The absolute values of their expression changes (M) was higher than 0.5 on both platforms. These genes could be classified into 3 subgroups. In case of 4 genes (EIF4E, AR, UBE2E1, EPHA4), mRNA underexpression was followed by the downregulation of protein expression. These genes are involved in translation initiation, transcription, intercellular communication, posttranslational modification and signal transduction. The second subgroup contained genes whose increase in mRNA expression was followed by protein overexpression (NCF2, TOP2B, SEMA4D, NUMA1). These genes are involved in cellular defense, electron carrier, DNA topological change, immuneresponse, cell differentiation and mitosis. In the third subgroup, mRNA expression and protein levels changed into opposite directions (GLUL, GAP43, LCP2, PPP1R1B). These

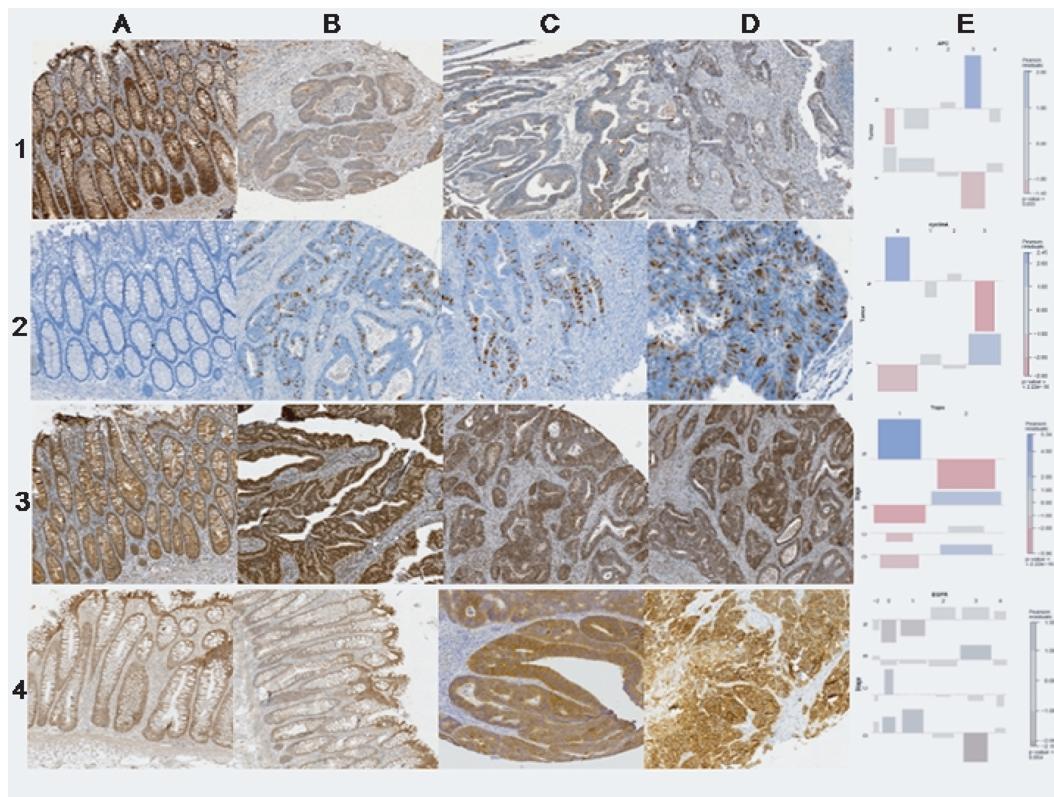

Fig. 2. Validation of selected protein markers on TMA. Protein expression of APC, Cyclin A, TOP1, EGFR. APC immunohistochemistry on normal (1A), Dukes B (1B), Dukes B (1C) and Dukes D (1D) stage colorectal samples and the association plot (1E), showing the frequency of the staining intensities. Cyclin A1 Immunohistochemistry on normal (2A), Dukes B (2B), Dukes B (2C), Dukes D (2D) stage CRC samples, and the association plot (2E). Topoisomerase 1 immunohistochemistry on normal (3A), Dukes B (3B), Dukes B (3C), Dukes D (3D) samples, and the related association plot (3E). EGFR immunohystochemistry on normal (4A), normal (4B), Dukes B (4C), Dukes D (4D) samples, and the related association plot (4E).

genes are involved in metabolism, cell growth, cytokine secretion and signal transduction regulation.

4. Discussion

Protein chips have been used in order to measure changes in the expression of genes involved in signal transduction and to examine the activation status of proteins. While further developing this technology, however, the analysis of expression profiles is getting more and more attention. Using reverse phase arrays performed (RPA) on laser microdissected samples, primary tumours were successfully discriminated from their metastases based on the phosphoproteomics profile of 29 genes involved in signal transduction [25]. It is important to note that in previous AB array analyses the number of genes showing a two-fold expression alteration was very low. Therefore, the intra-group variance expressed by the p value is more useful than the exact values of expression changes.

Other studies also support that M values between 0.5–0.7 represent acceptable protein level changes. In other words, the increase of protein expression to 140 percent, as well as its decrease to 70 percent, represent significant expression alterations [1].

The CCNE1 and CCND1 genes, which were found to be significantly altered in our classification, have previously been reported to be overexpressed in colorectal cancer. Higher levels of CCND1 have also been detected in the plasma of CRC patients. Our TMA-validated AB array results support this observation.

The MRE11A gene codes a protein in the cell nucleus that forms a complex with the RAD50 and NBS1 proteins. This complex is involved in homolog recombination, telomere protection and double stranded DNA repair, hence it can be classified as belonging to the MMR gene family [20]. Our antibody array data agree with the results of Giannini et al. that MRE11

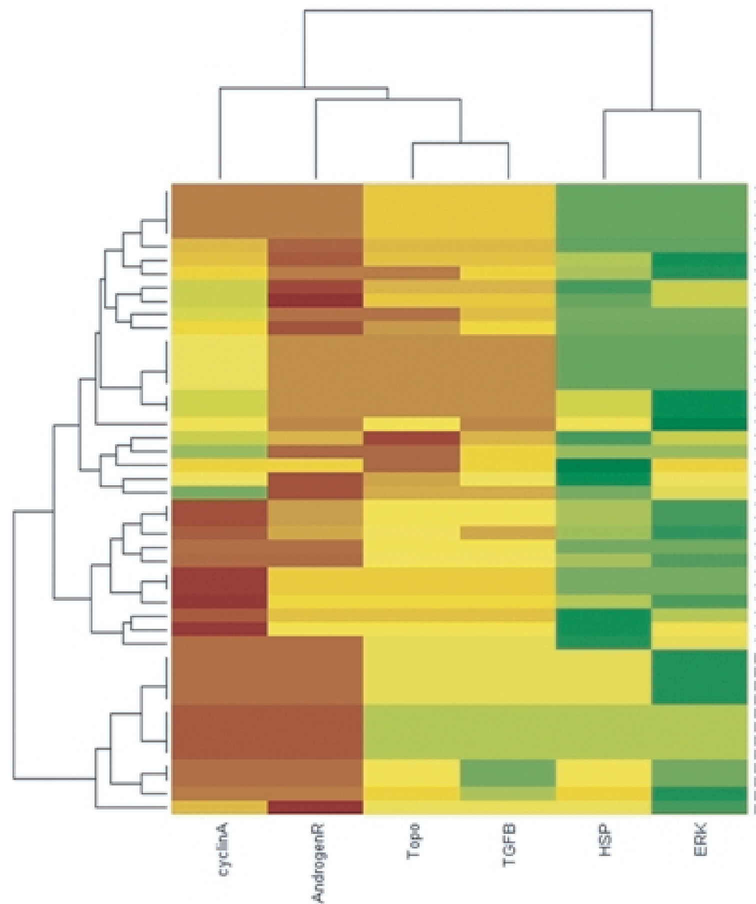

Fig. 3. Hierarchical clustering of 6 differentially expressed genes found to be associated with CRC subtype after TMA validation.

is downregulated in colon cancer, probably due to its mutation which results in a truncated protein. However, this observation made by AB array, could not be validated by TMA as protein expression did not change either in early or in late CRC.

Caldesmon 1 is expressed in low levels in colorectal cancer (GeneAtlas), but it was recently reported that this actin-binding protein plays a crucial role in the invasivity of cancer cells [26]. Ectopic expression of caldesmon 1 arrests extracellular matrix degradation, and decreases podosomes and invadopodia. The examination of caldesmon 1 and related proteins may be important in the assessment of tumour progression. The case of caldesmon 1 highlights a special error of AB arrays: tissue specificity of protein expression and sample homogeneity may influence the results. Caldesmon 1 is expressed mainly in smooth muscle and stromal cells, in cell layers around the crypts, and not in the epithelium. The $M = 0.39$ value measured by AB array suggests that this protein is underexpressed in tumours. However, immunohistochemistry shows that this gene is highly overexpressed in some stromal cell groups of tumours.

The opposing expression changes detected in this study may be caused by a higher expression of the protein in the stromal parts of the tumour, but these histological regions are relatively rare. It is also possible that during sample collection, the healthy colon was contaminated with smooth muscle. This renders the validation of marker localization and tissue specificity even more important.

Although association plots made according to the results of cluster and TMA analyses show no significant molecular alteration between the two disease groups, establishing a molecular-based disease classification may be useful in order to complement conventional clinical disease classification.

During tumour progression, some proteins showed continuous overexpression (TOP1, HSP60), while others (APC) were consistently downregulated in a

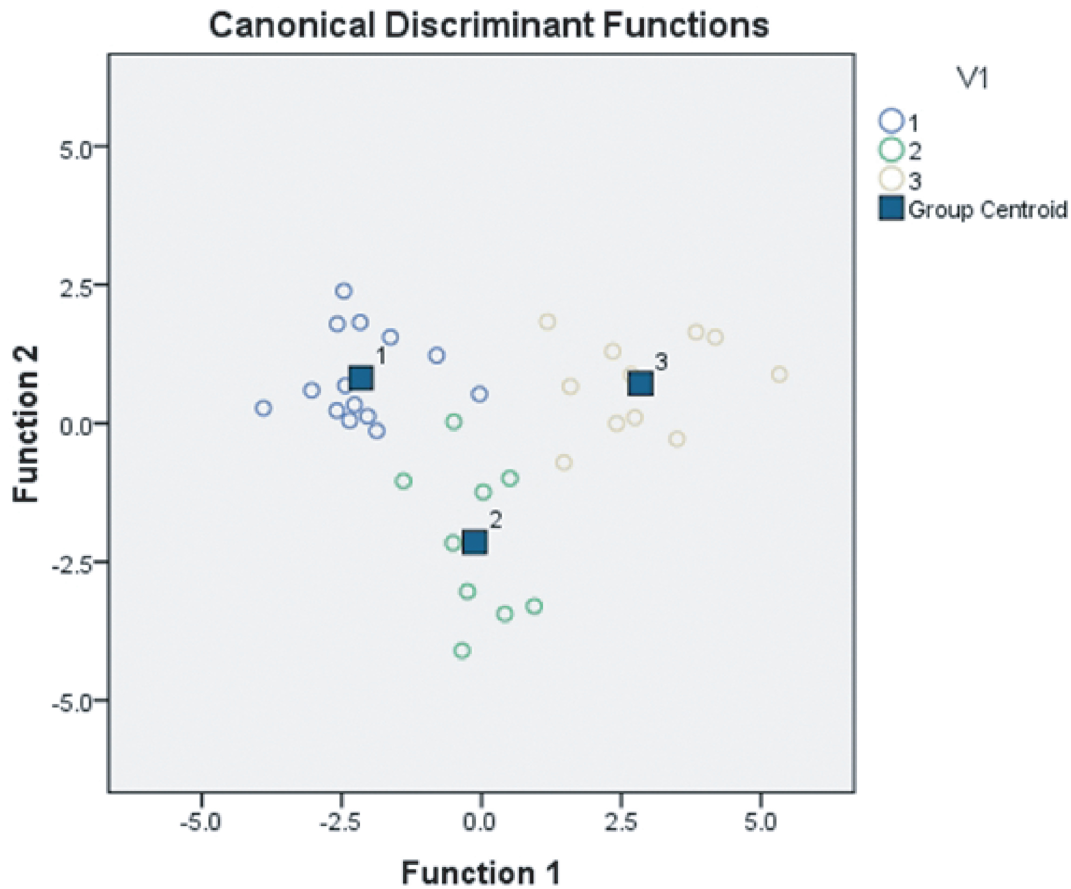

Fig. 4. Discrimination of tumor subtypes. Discriminant analysis of colonic biopsy specimens. Note the clear separation of the single classification groups based on the discriminatory genes detailed in the results section. Group 1 = normal, Group 2 = Dukes B colorectal cancer, Group 3 = Dukes D colorectal cancer.

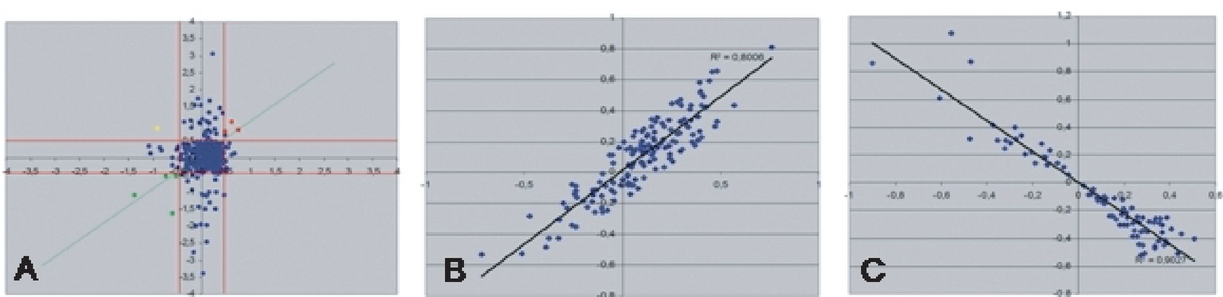

Fig. 5. Associations between mRNA and protein expression levels (A). The selected positively (B) and negatively (C) correlating genes.

trend-like manner. EGFR showed a stage-associated expression, as in the phase of tumour growth it was overexpressed, and after metastasis development, its expression decreased.

Several different factors may help to explain the discrepancies found between protein array and immunohistochemical analysis results. In some extreme cases, antibody array and immunohistochemistry show opposite results (e.g. MRE11A and AR). One of the most important facts, which is also decisive in this study, is the distinct antigen–antibody interaction between protein detection from protein extract and from FFPE tissue. As observed in case of Caldesmon1, both the tissue specific expression and the relative portion of tu-

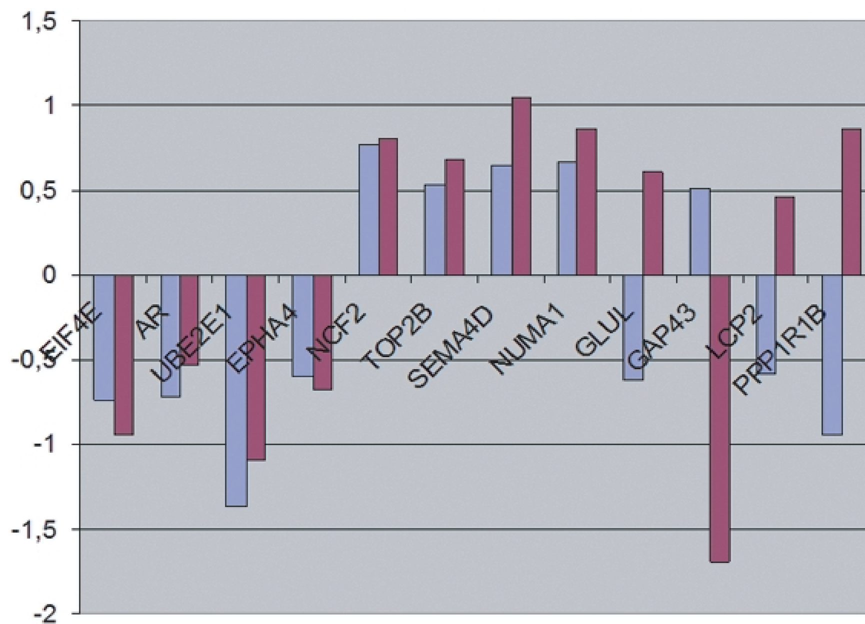

Fig. 6. The association between mRNA and protein products of 12 genes by M values of the two platforms. The first column (black) represents the mRNA and the second column (gray) represents the protein expression.

mor cells may differ in tissue samples. Furthermore, different antibodies were applied on the AB array and the TMA-based immunohistochemistry. A further explanation for the above mentioned discrepancies may be tumor heterogeneity, because the crude extract for the AB array was prepared from a relatively large tumor tissue specimen, while TMAs contain only some 2mm cores from the representative area of the FFPE tissue sample. Another important fact is that each antibody was individually optimized in TMA experiments, while in case of AB microarray analysis similar circumstances are provided for hundreds of antibodies. In contrast to TMA evaluation, which is based on discrete values and may have a subjective component, continuous variables are objectively evaluted during array analysis. During TMA evaluation (scoring) we focused only on the epithelial region. However, our experiences suggest that the different cell types in the tumor microenvironment may affect protein array analysis causing the discrepancies between the results of the two platforms. The strong correlation between protein array and mRNA may be explained by using the homogenized tissue and microarray platforms. Considering these parameters, the protein arrays may become useful tools for the assessment of protein expression in diagnostic practice, at present, however, immunohistochemical biomarker detection on TMAs seems to be more reliable and cost-effective.

To our knowledge, a similarly detailed study to determine whether there is an association between mRNA and protein expression has not been performed before. According to our data and primarily because of the high correlation values, it seems possible to predict protein array results based on whole genome mRNA expression microarrays. These correlations may be increased with the help of better protocols that allow the simultaneous isolation of RNA and protein in high quality and sufficient amounts while using a single set of starting material.

High throughput antibody arrays may be useful for the identification of protein biomarkers and diagnostic protein expression patterns. The identification of marker combinations and their application in everyday practice are important tasks where antibody microarrays may be of fundamental importance. As this technology is extremely sensitive, these marker combinations may be used in screening, diagnostics and follow-up of different stages of tumorous diseases. Today the number of markers on antibody arrays is considerably lower than the number of genes on whole genome mRNA arrays although their price is higher. If a correlation can be established between the results of whole genome mRNA chips and antibody arrays of limited capacity, then the former may be used as a convenient high capacity screening method before applying antibody arrays that contain already planned targets. Based on

our results, it seems that protein level changes can be predicted by the results of mRNA expression analyses. This could be important in case of proteins whose intracellular concentration fluctuates rapidly. Because of biological processes that are not yet fully understood, such as stability, half-lifetime, RNA processing and its regulation or posttranslational modifications, these results must be carefully evaluated. Taking these factors into consideration, antibody arrays may represent a new approach for tumor and biomarker research.

Acknowledgements

We thank hereby the excellent technical assistance of Gabriella Farkas and support of Tibor Kren ács Ph.D. for TMA immunohistochemistry optimalization.

References

[1] J. Madoz-Gurpide, M. Canamero, L. Sanchez, J. Solano, P. Alfonso and J.I. Casal, A proteomics analysis of cell signaling alterations in colorectal cancer, *Mol Cell Proteomics* **6** (2007), 2150–2164.
[2] J.R. Jass, Classification of colorectal cancer based on correlation of clinical, morphological and molecular features, *Histopathology* **50** (2007), 113–130.
[3] S.B. Baylin, J.G. Herman, J.R. Graff, P.M. Vertino and J.P. Issa, Alterations in DNA methylation: a fundamental aspect of neoplasia, *Adv Cancer Res* **72** (1998), 141–196.
[4] G.A. Calin, C.M. Croce, MicroRNA-cancer connection: the beginning of a new tale, *Cancer Res* **66** (2006), 7390–7394.
[5] G.A. Calin and C.M. Croce, MicroRNAs and chromosomal abnormalities in cancer cells, *Oncogene* **25** (2006), 6202–6210.
[6] G.A. Calin and C.M. Croce, MicroRNA signatures in human cancers, *Nat Rev Cancer* **6** (2006), 857–866.
[7] R.Z. Chen, U. Pettersson, C. Beard, L. Jackson-Grusby and R. Jaenisch, DNA hypomethylation leads to elevated mutation rates, *Nature* **395** (1998), 89–93.
[8] R. Garzon, M. Fabbri, A. Cimmino, G.A. Calin and C.M. Croce, MicroRNA expression and function in cancer, *Trends Mol Med,* **12** (2006), 580–587.
[9] H. Hayashi, G. Nagae, S. Tsutsumi, K. Kaneshiro, T. Kozaki, A. Kaneda, H. Sugisaki and H. Aburatani, High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array, *Hum Genet* **120** (2007), 701–711.
[10] P.A. Jones and P.W. Laird, Cancer epigenetics comes of age, *Nat Genet* **21** (1999), 163–167.
[11] B. Vogelstein, E.R. Fearon, S.R. Hamilton, S.E. Kern, A.C. Preisinger, M. Leppert, Y. Nakamura, R. White, A.M. Smits and J.L. Bos, Genetic alterations during colorectal-tumor development, *N Engl J Med* **319** (1988), 525–532.
[12] E.R. Fearon and B.A. Vogelstein, genetic model for colorectal tumorigenesis, *Cell* **61** (1990), 759–767.
[13] A. Leslie, F.A. Carey, N.R. Pratt and R.J. Steele, The colorectal adenoma-carcinoma sequence, *Br J Surg* (2002), 845–860.
[14] D.A. Notterman, U. Alon, A.J. Sierk and A.J. Levine, Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays, *Cancer Res* **61** (2001), 3124–3130.
[15] S. Spisak, Z. Tulassay, B. Molnar and A. Guttman, Protein microchips in biomedicine and biomarker discovery, *Electrophoresis* **28** (2007), 4261–4273.
[16] O. Galamb, F. Sipos, E. Dinya, S. Spisak, Z. Tulassay and B. Molnar, mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray, *World J Gastroenterol* **12** (2006), 6998–7006.
[17] O. Galamb, F. Sipos, B. Molnar, D. Szoke, S. Spisak and Z. Tulassay, Evaluation of malignant and benign gastric biopsy specimens by mRNA expression profile and multivariate statistical methods, *Cytometry B Clin Cytom* **72** (2007), 299–309.
[18] S. Spisak and A. Guttman, Biomedical applications of protein microarrays, *Curr Med Chem* **16** (2009), 2806–2815.
[19] F. Bertucci, S. Salas, S. Eysteries, V. Nasser, P. Finetti, C. Ginestier, E. Charafe-Jauffret, B. Loriod, L. Bachelart, J. Montfort, G. Victorero, F. Viret, V. Ollendorff, V. Fert, M. Giovaninni, J.R. Delpero, C. Nguyen, P. Viens, G. Monges, D. Birnbaum and R. Houlgatte, Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters, *Oncogene* **23** (2004), 1377–1391.
[20] G. Giannini, C. Rinaldi, E. Ristori, M.I. Ambrosini, F. Cerignoli, A. Viel, E. Bidoli, S. Berni, G. D'Amati, G. Scambia, L. Frati, I. Screpanti and A. Gulino, Mutations of an intronic repeat induce impaired MRE11 expression in primary human cancer with microsatellite instability, *Oncogene* **23** (2004), 2640–2647.
[21] G. Giannini, E. Ristori, F. Cerignoli, C. Rinaldi, M. Zani, A. Viel, L. Ottini, M. Crescenzi, S. Martinotti, M. Bignami, L. Frati, I. Screpanti and A. Gulino, Human MRE11 is inactivated in mismatch repair-deficient cancers, *EMBO Rep* **3** (2002), 248–254.
[22] S.P. Gygi, Y. Rochon, B.R. Franza and R. Aebersold, Correlation between protein and mRNA abundance in yeast, *Mol Cell Biol* **19** (1999), 1720–1730.
[23] E. Spooncer, N. Brouard, S.K. Nilsson, B. Williams, M.C. Liu, R.D. Unwin, D. Blinco, E. Jaworska, P.J. Simmons and A.D. Whetton, Developmental fate determination and marker discovery in hematopoietic stem cell biology using proteomic fingerprinting, *Mol Cell Proteomics* **7** (2008), 573–581.
[24] G.K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments, *Stat Appl Genet Mol Biol* **3** (2004), Article3.
[25] C. Belluco, E. Mammano, E. Petricoin, L. Prevedello, V. Calvert, L. Liotta, D. Nitti and M. Lise, Kinase substrate protein microarray analysis of human colon cancer and hepatic metastasis, *Clin Chim Acta* **357** (2005), 180–183.
[26] T. Yoshio, T. Morita, Y. Kimura, M. Tsujii, N. Hayashi and K. Sobue, Caldesmon suppresses cancer cell invasion by regulating podosome/invadopodium formation, *FEBS Lett* **581** (2007), 3777–3782.