

# Potential biomarkers of colorectal adenoma–dysplasia–carcinoma progression: mRNA expression profiling and *in situ* protein detection on TMAs reveal 15 sequentially upregulated and 2 downregulated genes

Orsolya Galamb<sup>a,b,\*,\*\*</sup>, Ferenc Sipos<sup>a,\*</sup>, Sándor Spisák<sup>a</sup>, Barnabás Galamb<sup>a</sup>, Tibor Krenács<sup>c</sup>, Gábor Valcz<sup>a</sup>, Zsolt Tulassay<sup>a,b</sup> and Béla Molnár<sup>a,b</sup>

<sup>a</sup> 2nd Department of Medicine, Semmelweis University, Budapest, Hungary

<sup>b</sup> Molecular Medicine Research Unit, Hungarian Academy of Sciences, Budapest, Hungary

<sup>c</sup> 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

**Abstract.** *Background:* As most colorectal cancers (CRC) develop from villous adenomas, studying alterations in gene expression profiles across the colorectal adenoma–dysplasia–carcinoma sequence may yield potential biomarkers of disease progression.

*Methods:* Total RNA was extracted, amplified, and biotinylated from colonic biopsies of 15 patients with CRC, 15 with villous adenoma and 8 normal controls. Gene expression profiles were evaluated using HGU133Plus2.0 microarrays and disease progression associated data were validated with RT-PCR. The potential biomarkers were also tested at the protein level using tissue microarray samples of 103 independent and 16 overlapping patients.

*Results:* 17 genes were validated to show sequentially altered expression at mRNA level through the normal–adenoma–dysplasia–carcinoma progression. Prostaglandin-D2 receptor (PTGDR) and amnionless homolog (AMN) genes revealed gradually decreasing expression while the rest of 15 genes including osteonectin, osteopontin, collagen IV–alpha 1, biglycan, matrix GLAprotein, and von Willebrand factor demonstrated progressively increasing expression. Similar trends of expression were confirmed at protein level for PTGDR, AMN, osteopontin and osteonectin.

*Conclusions:* Downregulated AMN and PTGDR and upregulated osteopontin and osteonectin were found as potential biomarkers of colorectal carcinogenesis and disease progression to be utilized for prospective biopsy screening both at mRNA and protein levels. Gene alterations identified here may also add to our understanding of CRC progression.

**Keywords:** Colorectal cancer, adenoma–dysplasia–carcinoma sequence, gene expression signature, whole genomic oligonucleotide microarray, tissue microarray, amnionless homolog, prostaglandin D2 receptor

## 1. Introduction

Colorectal cancer (CRC) is one of the most frequent cancers in the world with a very high mortality rate even after surgical resection, radio- and chemotherapy [5]. It seems evident that CRC frequently follows

and develops from adenomatous polyps [22]. Analyzing alterations of gene expression of the colorectal adenoma–dysplasia–carcinoma sequence may help understanding the molecular background of disease progression and reveal biomarkers for prospective diagnostic screening [13].

The colorectal adenoma–carcinoma sequence is featured by the gradual development of *in situ* carcinoma from adenoma through dysplasia with accumulating mutations and associated changes in the expression of genes regulating the Wnt/Wingless, K-ras, TGF-beta, and p53 pathways [4,9,17,22,41]. Microarray-

\*These authors contributed equally to this work.

\*\* Corresponding author: Orsolya Galamb, PhD, 2nd Department of Medicine, Semmelweis University, Szentkirályi str. 46, 1088 Budapest, Hungary. Tel.: +36 1 266 09 26; Fax: +361 266 08 16; E-mail: orsg1@yahoo.com.

based molecular analysis of malignancy in colon adenoma and CRC samples were described using 36 [27], 32 [35,40], 10 [1,20], 9 [23,24], 4 [30], 2 [21] and 1 [28] adenoma samples compared to adenocarcinoma and normal colonic tissues. Consistent with the observation that APC mutation is an early event in colon carcinogenesis, MYC and claudin 1 transcripts displayed an increased expression in adenomas [20]. In general, expression of Wnt target genes was found elevated in adenomatous samples compared to normal tissue [35]. The genes markedly upregulated in adenomas compared to normal tissues were generally associated to pathways of mitosis, DNA replication and spindle organization. Downregulated genes were predominantly involved in host immune defense, inorganic anion transport, organ development, and inflammatory response [35]. Nosho et al. identified gene expression alterations involved in the adenoma–carcinoma sequence including the upregulation of insulin-like growth factor 2 (IGF2), IGF1, Ki-67 and the downregulation of p21, heat shock protein 90, and caspase-7 genes [28, 29]. According to cDNA microarray and immunohistochemical analyses of Mori et al. increased JAK3 kinase, matrix metalloproteinase 13, heat shock protein 60 and MDM2 mRNA and protein expression correlates with the progression of CRC [26]. Several abnormalities (such as upregulated MGSA, BIGH3, matrilysin and downregulated guanylin, hevin) present in the carcinomas were already detectable in adenomas suggesting that these genes and their transcripts may play a role at a relatively early stage of colorectal carcinogenesis [20,30]. A number of genes discriminating carcinoma from adenoma were either relevant to hypoxia [24], or involved in apoptosis regulation and tumor suppression [21].

Genome-wide mRNA expression profiling studies using microarrays may have the potential to reveal the molecular background and support the diagnostic,

prognostic and treatment decisions in human disorders including cancers. In the gastrointestinal tract, biopsy samples are routinely taken during the endoscopic examination with minimal intervention [11,12,14,35,40]. These are suitable for mRNA expression analysis for identifying potential biomarkers of early signs of malignant transformation and progression. Gene expression profiles of the whole colorectal normal–adenoma–dysplasia–carcinoma sequence using biopsy samples and whole genomic microarrays have not been analyzed yet.

The aim of this study was to identify differentially expressed genes associated with colorectal cancer development and progression.

## 2. Material and methods

### 2.1. Patients and samples

After obtaining informed consent of untreated patients – who received neither chemo- nor radiotherapy at the time of sample collection – colon biopsy samples were taken during endoscopic intervention at the 2nd Department of Internal Medicine, Semmelweis University, Budapest, Hungary. Altogether 327 tissue samples of 141 patients were analyzed in this study. Some of the datasets of the 38 patients' samples including 15 tubulovillous/villous adenomas, 15 CRC, and 8 healthy controls, hybridized to Affymetrix microarrays, were used and published in an earlier study investigating different aspects of colorectal disease and cancer [11] and are available in the Gene Expression Omnibus database (series accession number: GSE4183). Furthermore, formalin-fixed and paraffin-embedded tissue samples of 103 independent and 16 overlapping patients were also analyzed using tissue microarrays. The diagnostic groups and the number of patients are summarized in Table 1. Detailed patient specification is described in Supplementary Table 1.

Table 1  
Number of patients per disease group participating in the study

Group	Original set ( $n = 38$ )			Independent set ( $n = 103$ )
	Affymetrix microarray	Taqman RT-PCR	Tissue microarray	Tissue microarray
Adenoma with low-grade dysplasia	6	6	3	17
High-grade dysplastic adenoma	9	6	5	20
CRC Dukes B	7	6	1	20
CRC Dukes C–D	8	4	3	21
Normal	8	5	8	22
Total patient numbers	38	27	16	103

### 2.1.1. mRNA expression microarray analysis

Total RNA was extracted using RNeasy Mini Kit according to the manufacturer's instructions (Qiagen Inc., Germantown, US). DNase I digestion was performed in order to remove the genomic DNA. Quantity and quality of the isolated RNA were tested by measuring the absorbance, agarose gel electrophoresis or/and capillary gel electrophoresis using 2100 Bioanalyzer and RNA 6000 Pico Kit (Agilent Inc., Santa Clara, US). Biotinylated cRNA probes were synthesized from 5–8 µg total RNA and fragmented according to the Affymetrix manual guidelines ([https://www.affymetrix.com/support/downloads/manuals/expression\\_s2\\_manual.pdf](https://www.affymetrix.com/support/downloads/manuals/expression_s2_manual.pdf)), using cDNA synthesis reagents, Sample Cleanup Module and Enzo BioArray High-Yield RNA Transcript Labelling Kit. Total RNA was reverse transcribed using 100 pmol T7-oligo(dT) primer, 10 mM DTT, dNTPs (each 500 µM), 200 U SuperScript II reverse transcriptase for 1 h at 42°C. The second strand cDNA synthesis was performed using 10 U/sample *E. coli* DNA ligase, 40 U/sample *E. coli* DNS-polymerase I, 2 U/sample *E. coli* RNase H, and dNTPs (each 200 µM) for 2 h at 16°C. After the cleanup of the cDNA, biotinylated cRNA probes were synthesized for 4 h at 37°C using T7 RNA polymerase. The cRNA samples were fragmented for 35 min at 94°C to generate 35–200 ribonucleotide length fragments. Ten µg of each fragmented cRNA sample was hybridized into HGU133 Plus2.0 array (Affymetrix Inc., Santa Clara, US) at 45°C for 16 h. The hybridization mix contained 0.05 µg/µl fragmented cRNA samples, hybridization buffer (100 mM MES, 1 M [Na<sup>+</sup>], 20 mM EDTA and 0.01% Tween 20), herring sperm DNA (0.1 mg/ml Promega, Wisconsin, US), acetylated BSA (0.5 mg/ml, Invitrogen), sensitivity controls (1.5 pM BioB, 5 pM BioC, 25 pM BioD and 100 pM Cre), and 50 pM B2 oligonucleotide orientation controls. The slides were washed and stained using Fluidics Station 450 and antibody amplification staining (using EukGE\_Ws\_2v4 fluidic protocol and 10 µg/ml streptavidin-phycoerythrin; Invitrogen/Molecular Probes, Carlsbad, US) according to the manufacturer's instructions (Affymetrix). The fluorescent signals were detected by a GeneChip Scanner 3000 (Affymetrix).

### 2.1.2. Statistical evaluation of mRNA expression profiles

*Pre-processing and quality control.* Quality control analyses were performed according to the suggestions of the Tumor Analysis Best Practices Working Group [39]. Scanned images were inspected for arti-

facts, percentage of present calls (>25%) and control of the RNA degradation were evaluated. Based on the evaluation criteria all biopsy measurements fulfilled the minimal quality requirements. RMA background correction, quantile normalization and median polish summarization were applied. The datasets for further analysis are available in the Gene Expression Omnibus databank (<http://www.ncbi.nlm.nih.gov/geo/>), series accession number: GSE4183.

*Determination of genes involved in the adenoma-dysplasia-carcinoma sequence.* Kendall's rank correlation analysis was performed for quantification of association between the expression level and the disease stages in case of both microarray and Taqman quantitative RT-PCR data. The correlation of the two variables was described by Kendall's tau, its value is between -1 and 1. Higher proximity of the tau value to the terminal values indicates stronger correlation between the two variables. The *p*-value of tau under the null hypothesis of no association is computed by in the case of no ties using an exact algorithm described by Best and Gipps [2]. The R-environment was used for statistical analysis.

### 2.1.3. Taqman quantitative RT-PCR

TaqMan quantitative real-time PCR was used to measure the expression of increasing or decreasing expression tendency showing genes using an Applied Biosystems Micro Fluidic Card System. Using Taqman Reverse Transcription Kit, 400 ng/sample total RNS was reverse transcribed (Applied Biosystems, Foster City, US). The quality of the cDNA samples was checked by CK20/PBGD real-time PCR. The expression analysis of the selected genes was performed from 100 ng/sample cDNA template, using Taqman Low-Density Array for Gene Expression: Format 96a and Taqman Universal PCR Master Mix. The measurements were performed using an ABI PRISM<sup>®</sup> 7900HT Sequence Detection System as described in the products User Guide (<http://www.appliedbiosystems.com>). For data analysis the SDS 2.2 software was used. The extracted delta *Ct* values (which represent the expression normalized to the ribosomal 18S expression) were clustered according to the histological groups. Then the Student's *t*-test was performed to compare the expression values between groups.

### 2.1.4. Tissue microarray analysis

Cores of 1 mm diameter were collected from selected areas of formalin-fixed, paraffin-embedded tissue blocks prepared from 37 colorectal adenoma, 44 dysplastic adenoma, 89 early CRC (stage Dukes B),

57 advanced CRC (stage Dukes C and D) and 53 normal colon samples of 119 patients and inserted into 4 recipient blocks taking 70 samples each. 5  $\mu\text{m}$  thick tissue sections were cut from the blocks and immunostained. Protein products of the 17 significantly up- and downregulated genes (see later) were tested using commercially available antibodies. Four of them, which could be set up to work reliably in formalin-paraffin sections following antigen retrieval, were chosen for systematic screening of normal–adenoma–dysplasia–carcinoma sequence on TMA. Slides were stained using rabbit anti-human antibodies for amnionless homolog (1:200 dilution, Atlas Antibodies, Stockholm, Sweden, code: HPA000817), prostaglandin D2 receptor (1:500 dilution, Chemicon, Temecula, US, code AB9255), osteopontin (1:2000 dilution, Chemicon, Temecula, US) and osteonectin antibody (1:1000 dilution, Chemicon, code AB1858) for 1 h at room temperature. Antibodies were detected by using the En-Vision+ system (Dako, Glostrup, Denmark) followed by DAB-hydrogen peroxidase chromogen-substrate kit (Dako). Immunostained TMA slides were digitalized using high-resolution MIRAX DESK instrument (Zeiss, Gottingen, Germany), and analyzed with the MIRAX TMA Module software (Zeiss). Protein expression was evaluated using an empirical scale considering intensity and proportion of positive cytoplasmic staining of epithelial/carcinoma cells. Scores were given for PTGDR: –2 for no staining; 0 for weak, 1 for moderate, 2 for strong diffuse immunostaining; and for AMN, osteopontin and osteonectin: –2 for no or weak staining, 0 for moderate apical cytoplasmic, 1 for strong apical and weak basal cytoplasmic, 2 for strong diffuse cytoplasmic staining; Pearson's Chi-test and Fischer exact test were done for revealing if the staining difference in progression groups was significant ( $p < 0.05$ ). Also, contingency tables and association plots were constructed from the two categorical variables (group and score) [25].

### 3. Results

#### 3.1. mRNA expression microarray screening

Relative expression values were determined for each gene using the preprocessed microarray data in each disease group (normal, low-grade dysplastic adenoma, high-grade dysplastic adenoma, early stage CRC, advanced stage CRC). The association between gene expression level and disease group was demonstrated

with the Kendall's rank correlation analysis using 0.4 as a cut-off value (when absolute value of tau was 0.4 or higher). Along the transition of colorectal adenoma–dysplasia–carcinoma sequence, the gradual downregulation of 382 genes ( $p < 0.002$ ) and progressive overexpression of 918 genes ( $p < 0.002$ ) were detected. The list of these genes with complete annotation, tau and  $p$ -values are shown in Suppl. Table 2 (<http://www.gub.ac.uk/isco/JCO>).

Several sequential marker genes were represented on the microarray by more than one probe set including some of the progressively differentially expressed genes with different tau and  $p$  values (such as MCAM, collagen type IV–alpha 1, biglycan, interleukin 8, TIMP3, calumenin, SERPINE1).

#### 3.2. Taqman quantitative RT-PCR validation of differentially expressed genes

TaqMan quantitative real-time RT-PCR was performed to measure the expression changes of a selected set of 20 genes (2 continuously downregulated and 18 continuously overexpressed genes). Selection criteria for the genes were the significant progressive under- or overexpression in oligonucleotide microarray analysis in association with the normal–adenoma–dysplasia–carcinoma sequence and the availability of validated TaqMan probes. Although the tau value of VEGF was found to be lower than 0.4, it was also selected, because of its previously reported role as a CRC related gene. The complete results of the TaqMan measurements are presented in Suppl. Table 3 (<http://www.gub.ac.uk/isco/JCO>). The mRNA expression of 17 of the 20 selected genes showed a significant association with the disease stage in both systems, based on the 0.4 or higher absolute value of the tau ( $p < 0.01$ ).

The amnionless homolog and prostaglandin D2 receptor genes showed significantly decreasing expression tendency ( $p < 0.05$ ) (Table 2.) in association with the normal–adenoma–dysplasia–carcinoma sequence progression (Fig. 1).

Fifteen genes were identified showing significant and progressively increasing gene expression along with the adenoma–dysplasia–carcinoma sequence progression in both mRNAs expression analyses ( $p < 0.05$ ) (Table 2). The development and progression of colorectal cancer were characterized by the elevated expression of genes mainly involved in cell proliferation (vascular endothelial growth factor /VEGF/, CXCL1 chemokine ligand, tissue inhibitor of metal-

Table 2

Genes showing a continuous progressively increasing or decreasing expression tendency along with the adenoma–dysplasia–carcinoma sequence progression

Gene symbol	Gene name	Probe set ID	mRNA expression in normal–adenoma–dysplasia–carcinoma	Taqman tau	Taqman <i>p</i> -value	Affymetrix tau	Affymetrix <i>p</i> -value
PTGDR	Prostaglandin D2 receptor	215894_at	DOWN	0.579	0	−0.423	0.001
AMN	Amnionless homolog (mouse)	223587_s_at	DOWN	0.44	0.002	−0.42	0.001
MCAM	Melanoma cell adhesion molecule	210869_s_at	UP	−0.604	0	0.571	0
COL4A1	Collagen IV– $\alpha$ 1	211980_at	UP	−0.604	0	0.631	0
BGN	Biglycan	213905_x_at	UP	−0.586	0	0.549	0
IL8	Interleukin 8	202859_x_at	UP	−0.505	0	0.407	0.001
MGP	Matrix Gla protein	202291_s_at	UP	−0.499	0.001	0.584	0
THBS2	Thrombospondin 2	203083_at	UP	−0.493	0.001	0.455	0
VWF	Von Willebrand factor	202112_at	UP	−0.481	0.001	0.653	0
SPP1	Osteopontin	209875_s_at	UP	−0.47	0.001	0.467	0
VEGF	Vascular endothelial growth factor	210512_s_at	UP	−0.47	0.001	0.376	0.002
TIMP3	Tissue metalloproteinase inhibitor 3	201150_s_at	UP	−0.464	0.001	0.411	0.001
SPARC	Osteonectin	200665_s_at	UP	−0.44	0.002	0.543	0
CXCL1	Chemokine (C-X-C motif) ligand 1	204470_at	UP	−0.429	0.003	0.467	0
CALU	Calumenin	214845_s_at	UP	−0.411	0.005	0.48	0
TIMP1	Tissue metalloproteinase inhibitor 1	201666_at	UP	−0.4	0.006	0.492	0
IL1RN	Interleukin 1 receptor antagonist	212657_s_at	UP	−0.4	0.006	0.47	0

loproteinase 1), apoptosis (osteopontin, VEGF, tissue inhibitor of metalloproteinase 3), angiogenesis (VEGF, von Willebrand factor, interleukin 8), cell adhesion (melanoma cell adhesion molecule, thrombospondin 2, von Willebrand factor, osteopontin, collagen IV– $\alpha$ 1). The expression of several immune regulation-related genes is also changed (interleukin 8, CXCL1, interleukin receptor 1 antagonist).

### 3.3. Protein expression and localization of selected markers in tissue microarray

Downregulated amnionless homolog and prostaglandin D2 receptor proteins were tested with immunohistochemistry in a TMA series of archived tissues. Strong diffuse epithelial cytoplasmic PTGDR and AMN immunostaining was found in the epithelial cells in healthy colon tissue. In adenomas, the number of AMN and PTGDR positive cells gradually decreased near the luminal surface. During disease progression the intensity of immunostaining was further reduced in dysplastic epithelium followed by the appearance of only weak apical cytoplasmic epithelial immunostaining in CRC samples (Fig. 2A–F). Image quantification made on digital slides also confirmed the gradually reduced expression of these antigens along the disease progression (Fig. 3).

Moderate cytoplasmic osteopontin and osteonectin staining was found in TMA samples at the apical cytoplasm of the epithelial cells in healthy colon tissue. In adenomas, the number of osteopontin and osteonectin positive cells gradually increased near the luminal surface accompanied with mild-moderate basal cytoplasmic staining in some of them. The intensity of antigen expression was elevated further during disease progression in CRC samples (Fig. 2G–L).

## 4. Discussion

It is thought that colorectal cancer usually develops from villous adenomas, which transition is associated with obvious phenotypic changes due to altered gene and concomitant protein expression [9,17,20,22,26,28,29,35]. Revealing these molecular changes, they may be exploited for clinical applications, e.g. for predicting potential disease progression in prospective biomarker testing. In this study, we analyzed the gene expression profile of the adenoma–dysplasia–carcinoma sequence using 38 colorectal biopsy specimen hybridized on whole genomic HGU133 Plus 2.0 microarray system to identify disease associated progression biomarkers that show continuous quantitative alter-

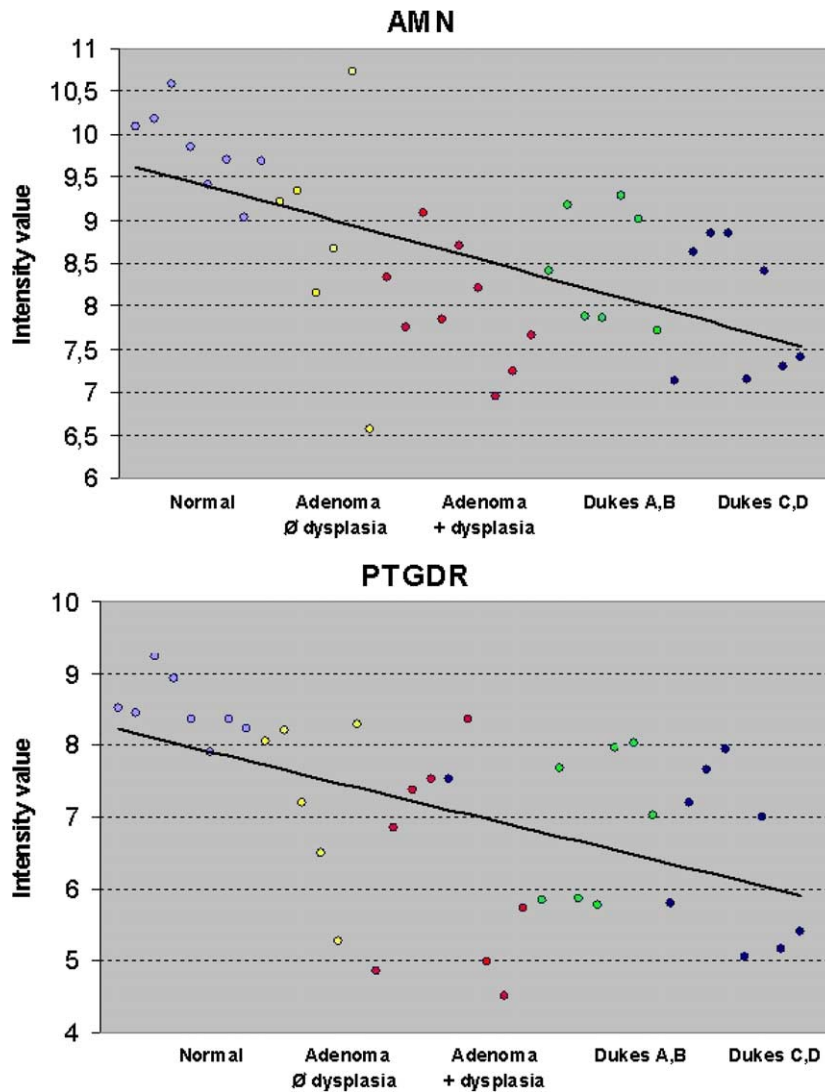


Fig. 1. Amnionless homolog and prostaglandin D2 receptor mRNA expression sequentially decreases during the normal–adenoma–dysplasia–carcinoma transition.

ations, and the results were validated with RT-PCR and on tissue microarrays.

In CRC development, we found a series of potential progression markers at the mRNA level, 17 of which altered significantly and gradually through the normal–adenoma–dysplasia–carcinoma sequence. PTGDR and amnionless homolog mRNA were down-regulated, while 15 others, including, osteonectin, osteopontin, collagen IV–alpha 1, biglycan, matrix GLA protein and von Willebrand factor, were up-regulated, which was also verified with RT-PCR. The progressive alterations in PTGDR and amnionless homolog expression, and of those of osteopontin and osteonectin expression were also verified at the *in situ* protein level,

suggesting that altered levels of these biomarkers may be associated with colorectal cancer development and progression.

The whole genomic microarray analyses of biopsy samples in this study provided highly standard and reproducible results regarding the array sensitivity, present percentage and GAPDH 3'/5' ratio. From the array data, only those transcripts were considered for further analysis as potential progression biomarkers, which could be validated with real-time PCR method. The expression of selected CRC-associated genes was validated also at the protein level, using immunohistochemistry on TMA collection of samples from 16 overlapping patients and an independent set of 103 pa-



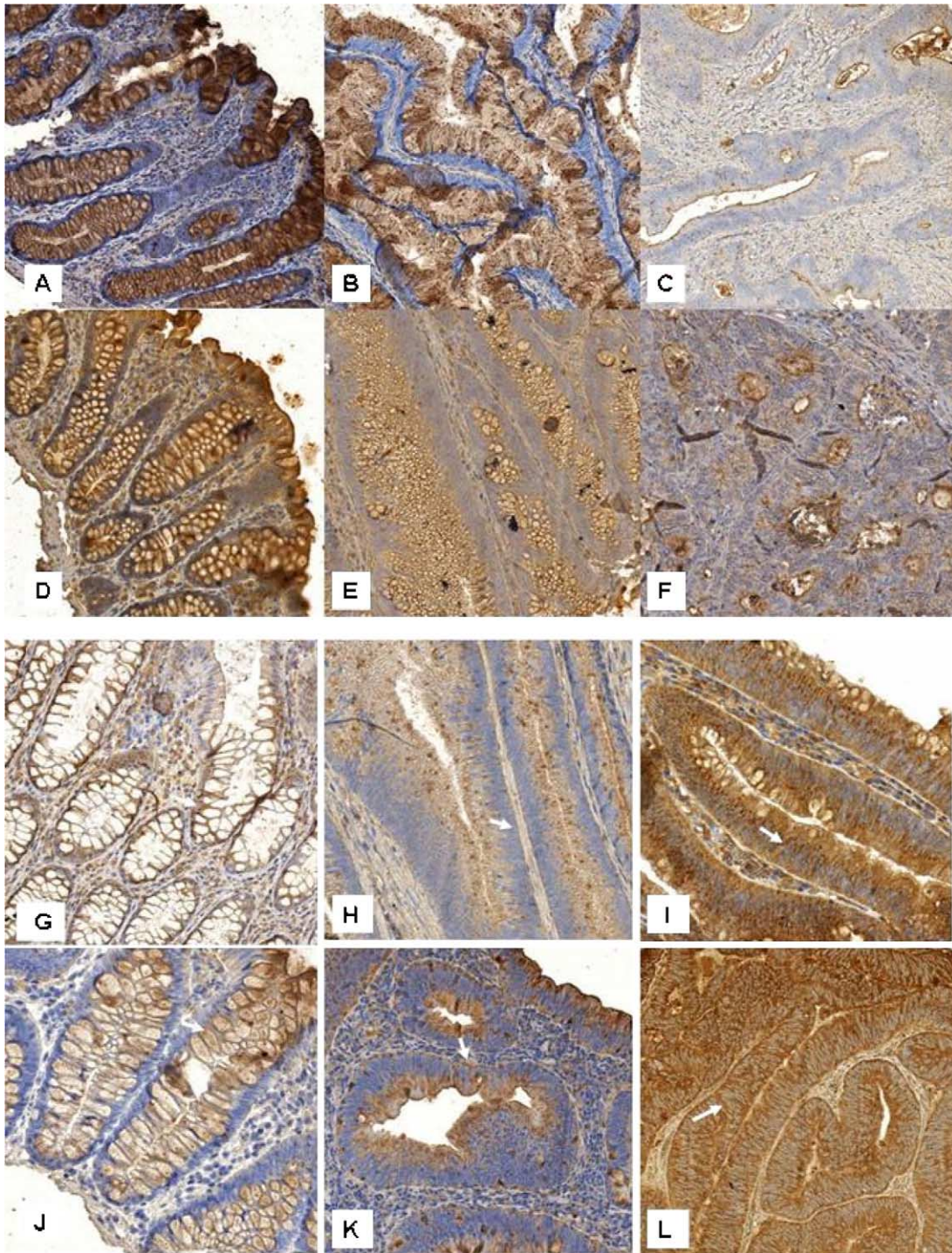


Fig. 2. Immunostaining for downregulated PTGDR and AMN, and upregulated osteopontin and osteonectin in TMA sections. Amnionless homolog/prostaglandin D2 receptor expression in normal colonic (A/D), dysplastic adenomatous (B/E) and CRC (C/F) samples. Osteopontin/osteonectin expression in normal colonic (G/J), dysplastic adenomatous (H/K) and CRC (I/L) samples. White arrows in figure G and J show apical cytoplasmatic staining in normal colonic epithelial cells. White arrows in figure H and K show moderate basal cytoplasmatic staining. Arrows in figure I and L show diffusely stained tumorous epithelium.

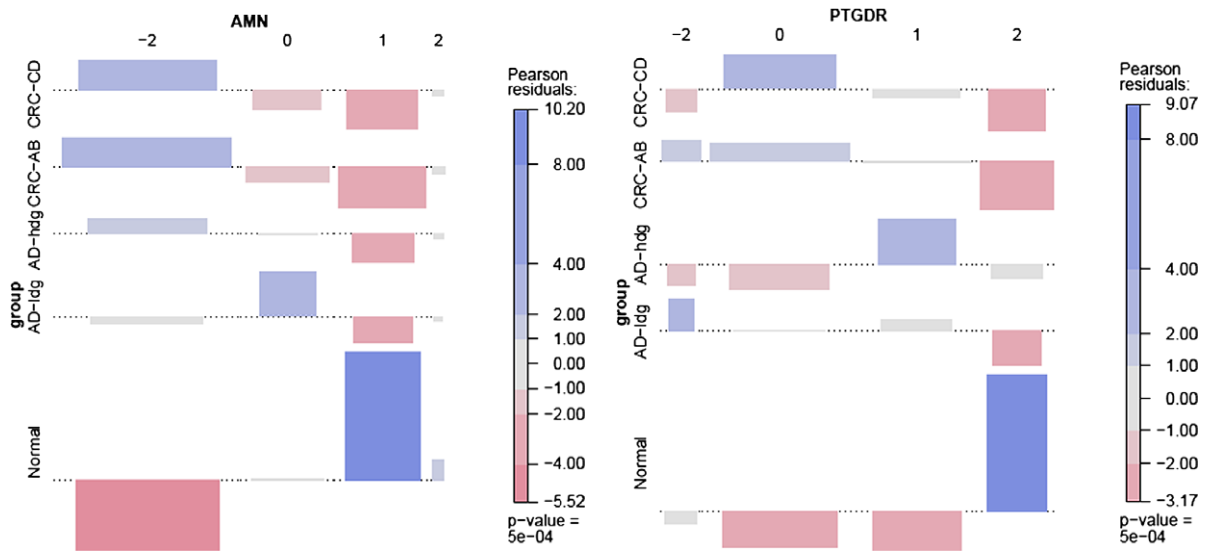


Fig. 3. Association plots representing amnionless homolog and prostaglandin D2 receptor protein expression detected by TMA in the subgroups. To measure the association of two variables (expression and disease stage) the Chi-square test statistic – the sum of the squared Pearson residuals – was used. When the difference was statistically significant ( $p < 0.05$ ), more detailed analysis was visualized on the basis of the Pearson residuals. The darker the intensity of the blue columns, the stronger the statistical significance at the given immunostaining intensity score. The height of the blue columns represents the number of cases belonging to the given score. Immunostaining scores: In case of PTGDR:  $-2$  = no staining;  $0$  = weak staining;  $1$  = moderate staining;  $2$  = strong diffuse epithelial cytoplasmatic immunostaining; in case of AMN:  $-2$  = no or weak apical epithelial cytoplasmatic staining;  $0$  = moderate apical cytoplasmatic staining,  $1$  = strong apical and weak basal cytoplasmatic staining;  $2$  = strong diffuse epithelial cytoplasmatic immunostaining. AMN = amnionless homolog, PTGDR = prostaglandin D2 receptor, CRC-CD = colorectal cancer Dukes C or D stage, CRC-AB = colorectal cancer Dukes A or B stages, AD-hdg = adenoma with high-grade dysplasia, AD-ldg = adenoma with low-grade dysplasia.

tients making up a total of 289 samples. Using digital slides and a dedicated software for scoring, TMA technique allowed the efficient and standardized analysis of large number of samples. Nowadays, antibodies recognizing a wide range of proteins in formalin-fixed paraffin-embedded tissues are available which offer good chances for further validating new biomarkers, such as those we found in this study, in the prospective diagnostic setting.

Genes showing a progressively decreasing expression during colorectal carcinogenesis in this study have received relatively little attention so far. Prostaglandin D2 receptor (PTDGR) G-protein-coupled receptor has been shown to function as a prostanoid DP receptor. The activation of this receptor plays an important role in allergic inflammation by modulating immune cell functions [6]. It is expressed in a series of tissues including several types of leukocytes, the vasculature, retina, nasal mucosa, lung and intestine [36]. PTGDR has been demonstrated to have anti-proliferative activity against human CRC cells *in vitro* [16]. PTGDR mRNA expression has been detected in normal colon but not in several neoplastic colorectal epithelial cell lines [16]. Amnionless homolog transmembrane pro-

tein has been hypothesized to modulate bone morphogenetic protein (BMP) receptor function by serving as an accessory co-receptor to either facilitate or hinder BMP binding [19]. It is known that the mouse AMN gene is expressed in the extraembryonic visceral endodermal layer during gastrulation but it was found to be mutated in amnionless mouse [19]. The function of AMN is otherwise unknown, however it is highly expressed in cubilin-expressing tissues, including the kidney, intestine, and mouse visceral yolk sac [10,19]. This is the first study to show a significant association of the gradually decreasing AMN and prostaglandin D2 receptor mRNA and protein expression with the colorectal adenoma–dysplasia–carcinoma sequence.

The protein expression of two genes showing progressively increasing mRNA levels during colorectal carcinogenesis were also validated. Osteopontin has been shown to bind to cells via integrins as well as CD44 [37]. Although the biological functions of osteopontin are not fully understood, it has been implicated in malignancy, immune function, and vascular remodeling, as well as in bone remodeling [31,33,37]. The matricellular glycoprotein SPARC (osteonectin) has been assigned a major role in the regulation of cell



adhesion and proliferation, as well as tumorigenesis and metastasis. SPARC transcripts were found to be overexpressed in primary CRCs and their liver metastases compared to non-neoplastic mucosa using Northern blot analysis and *in situ* hybridization [34]. However, this is the first study to show a significant association of the progressively increasing osteonectin mRNA and protein expression with the colorectal adenoma–dysplasia–carcinoma sequence.

Additional markers validated at the mRNA level are involved in cell proliferation (VEGF, CXCL1, TIMP1), apoptosis (VEGF, TIMP3), angiogenesis (VEGF, VWF, IL8), cell adhesion (MCAM, THBS2, VWF), transport (COL4A1), cell differentiation (MGP) and immune response (IL8, CXCL1, IL1RN). The exact cellular function of biglycan and calumenin gene products has not yet been determined.

In line with published findings comparing normal and CRC samples, we also found CXCL1, osteopontin, osteonectin, collagen type IV alpha 1, biglycan, interleukin 8, thrombospondin 2, VEGF, von Willebrand factor and TIMP1 overexpressed in CRC [1,3,7,8,15,18,30,32,34,38,42]. However, most studies compared expression features only pairwised and not in the full sequence of disease progression. Only two papers tested the adenoma–dysplasia–carcinoma sequence specific gene expression alterations [1,21]. One of them, also showed the expression of osteopontin, osteonectin, biglycan, TIMP1, –3 and CXCL1 genes to progressively increasing with CRC stage [1]. As far as we are aware, the rest of the markers we identified and validated in this study including amnionless homolog, prostaglandin D2 receptor, collagen IV–alpha 1, matrix GLAprotein, and von Willebrand factor have not been published in relation to the sequence of colorectal adenoma–dysplasia–carcinoma progression.

By testing biopsy and surgical specimens Sabates-Bellver et al. and van der Flier et al. found overexpression of several target genes when using the same microarray platform as we did [35,40]. However, they compared normal vs. adenoma and normal vs. carcinoma samples separately and used disease adjacent normal tissue for control, where gene expression might be influenced by the diseased tissue, as opposed to our approach of including normal healthy control samples. Moreover, in most studies except that published by Sabates-Bellver et al. [35] the degree of dysplasia was not determined within the adenoma group and mixed tubulovillous and tubular adenomas were included together [24,27,28]. Unlike this practice, we used independent samples with determined degree of dysplasia

of tubulovillous/villous adenomas to create low- and high-grade dysplasia groups.

Though histopathology is still the gold standard for differentiating stages of colorectal disease progression, the cut slide may represent only a fragment of the whole sample and disease features, such as dysplastic areas of flat adenomas, may remain hidden due to inappropriate sampling or tissue orientation. Following histological examination, a whole sample-based expression screening for biomarkers of disease progression with RT-PCR based real-time assay may improve diagnostic accuracy. Testing for alterations in biomarker expression at the protein level with prospective immunohistochemistry, may also have a great potential in routine diagnostic applications.

In summary, in this study we searched for potential biomarkers of colorectal adenoma–dysplasia–carcinoma progression using mRNA expression and tissue microarrays in biopsy specimen. 15 genes were identified and validated at mRNA level showing a significantly and progressively increasing expression during the normal–adenoma–dysplasia–carcinoma sequence progression ( $p < 0.01$ ) including osteonectin, osteopontin, collagen IV–alpha 1, biglycan, matrix GLAprotein, and von Willebrand factor. Validated genes of significantly decreasing expression were prostaglandin D2 receptor and amnionless homolog ( $p < 0.01$ ). In line with these, amnionless homolog and prostaglandin D2 receptor protein expression showed a strong negative correlation, while osteopontin and osteonectin protein expression showed a strong positive correlation with the colorectal normal–adenoma–dysplasia–carcinoma sequence, *in situ* on TMAs. Our findings suggest that these factors may be used as biomarkers for predicting colorectal cancer progression, which needs further validation in prospective diagnostic setting. Considering the positive regulatory roles of osteopontin in cell invasion, in vascular remodeling and its immunological (it is a key cytokine regulating tissue repair and inflammation) effects, and of osteonectin in the regulation of cell adhesion and proliferation, these proteins might be candidates for anticancer target treatment. Investigating the regulatory pathways of AMN and PTGDR may also reveal novel molecular targets for the same purpose.

## Acknowledgements

We thank hereby the assistance of Beáta Scholtz PhD with the Taqman RT-PCRs and the help of

Gabriella Farkas with the TMAs. Furthermore we thank Norbert Solymosi DVM for performing the bioinformatical analysis. Many thanks to Jeannette Molnár, MD, PhD, for reviewing the manuscript.

## References

- [1] D. Agrawal, T. Chen, R. Irby et al., Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling, *J. Natl. Cancer Inst.* **94** (2002), 513–521.
- [2] D.J. Best and P.G. Gipps, Algorithm AS 71: The upper tail probabilities of Kendall's tau applied statistics, *Appl. Statist.* **23** (1974), 98–100.
- [3] K. Birkenkamp-Demtroder, S.H. Olesen, F.B. Sorensen et al., Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid, *Gut* **54** (2005), 374–384.
- [4] D. Calistri, C. Rengucci, I. Seymour et al., KRAS, p53 and BRAF gene mutations and aneuploidy in sporadic colorectal cancer progression, *Cell. Oncol.* **28** (2006), 161–166.
- [5] Cancer Research Campaign, *Cancer Statistics: Large Bowel – UK*, CRC, London, 1999.
- [6] T. Chiba, A. Kanda, S. Ueki et al., Possible novel receptor for PGD2 on human bronchial epithelial cells, *Int. Arch. Allergy Immunol.* **143**(Suppl. 1) (2007), 23–27.
- [7] R.S. Croner, T. Foertsch, W.M. Brueckl et al., Common denominator genes that distinguish colorectal carcinoma from normal mucosa, *Int. J. Colorectal Dis.* **20** (2005), 353–362.
- [8] D.C. Damin, M.A. Rosito, P. Gus et al., Von Willebrand factor in colorectal cancer, *Int. J. Colorectal Dis.* **17** (2002), 42–45.
- [9] S. Derks, C. Postma, P.T. Moerkerk et al., Promoter methylation precedes chromosomal alterations in colorectal cancer development, *Cell. Oncol.* **28** (2006), 247–257.
- [10] J.C. Fyfe, M. Madsen, P. Højrup et al., The functional cobalamin (vitamin B12)-intrinsic factor receptor is a novel complex of cubilin and amnionless, *Blood* **103** (2004), 1573–1579.
- [11] O. Galamb, B. Gyorffy, F. Sipos et al., Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature, *Dis. Markers* **25** (2008), 1–16.
- [12] O. Galamb, F. Sipos, E. Dinya et al., mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray, *World J. Gastroenterol.* **12** (2006), 6998–7006.
- [13] O. Galamb, F. Sipos, K. Fischer et al., The results of the expression array studies correlate and enhance the known genetic basis of gastric and colorectal cancer, *Cytometry B Clin. Cytom.* **68** (2005), 1–17.
- [14] O. Galamb, F. Sipos, B. Molnar et al., Evaluation of malignant and benign gastric biopsy specimens by mRNA expression profile and multivariate statistical methods, *Cytometry B Clin. Cytom.* **72** (2007), 299–309.
- [15] M. Guba, H. Seeliger, A. Kleespies et al., Vascular endothelial growth factor in colorectal cancer, *Int. J. Colorectal Dis.* **19** (2004), 510–517.
- [16] G. Hawcroft, S.H. Gardner and M.A. Hull, Expression of prostaglandin D2 receptors DP1 and DP2 by human colorectal cancer cells, *Cancer Lett.* **210** (2004), 81–84.
- [17] D.M. Heijink, J.H. Kleibeuker, M. Jalving et al., Independent induction of caspase-8 and cFLIP expression during colorectal carcinogenesis in sporadic and HNPCC adenomas and carcinomas, *Cell. Oncol.* **29** (2007), 409–419.
- [18] M.N. Holten-Andersen, I.J. Christensen, H.J. Nielsen et al., Total levels of tissue inhibitor of metalloproteinases 1 in plasma yield high diagnostic sensitivity and specificity in patients with colon cancer, *Clin. Cancer Res.* **8** (2002), 156–164.
- [19] S. Kalantry, S. Manning, O. Haub et al., The amnionless gene, essential for mouse gastrulation, encodes a visceral-endoderm-specific protein with an extracellular cysteine-rich domain, *Nat. Genet.* **27** (2001), 351–352.
- [20] K.Y. Kwong, G.C. Bloom, I. Yang et al., Synchronous global assessment of gene and protein expression in colorectal cancer progression, *Genomics* **86** (2005), 142–158.
- [21] S. Lee, Bang, K. Song et al., Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon, *Oncol. Rep.* **16** (2006), 747–754.
- [22] A. Leslie, F.A. Carey, N.R. Pratt et al., The colorectal adenoma-carcinoma sequence, *Br. J. Surg.* **89** (2002), 845–860.
- [23] M. Li, Y.M. Lin, S. Hasegawa et al., Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray, *Int. J. Oncol.* **24** (2004), 305–312.
- [24] Y.M. Lin, Y. Furukawa, T. Tsunoda et al., Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas, *Oncogene* **21** (2002), 4120–4128.
- [25] D. Meyer, A. Zeileis and K. Hornik, The strucplot framework: Visualizing multi-way contingency tables with 'vcd', *J. Stac. Software* **17** (2006), 1–48.
- [26] D. Mori, Y. Nakafus, K. Miyazaki et al., Differential expression of Janus kinase 3 (JAK3), matrix metalloproteinase 13 (MMP13), heat shock protein 60 (HSP60), and mouse double minute 2 (MDM2) in human colorectal cancer progression using human cancer cDNA microarrays, *Pathol. Res. Pract.* **201** (2005), 777–789.
- [27] K. Noshio, H. Yamamoto, Y. Adachi et al., Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis, *Br. J. Cancer* **92** (2005), 1193–1200.
- [28] K. Noshio, H. Yamamoto, H. Takamaru et al., A case of colorectal carcinoma in adenoma analyzed by a cDNA array, *Int. J. Colorectal Dis.* **20** (2005), 287–291.
- [29] K. Noshio, H. Yamamoto, H. Taniguchi et al., Interplay of insulin-like growth factor-II, insulin-like growth factor-I, insulin-like growth factor-I receptor, COX-2, and matrix metalloproteinase-7, play key roles in the early stage of colorectal carcinogenesis, *Clin. Cancer Res.* **10** (2004), 7950–7957.
- [30] D.A. Notterman, U. Alon, A.J. Sierk et al., Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays, *Cancer Res.* **61** (2001), 3124–3130.
- [31] A.J. Oates, R. Barraclough and P.S. Rudland, The role of osteopontin in tumorigenesis and metastasis, *Inv. Metastasis* **17** (1997), 1–15.
- [32] K. Okuno, M. Yasutomi, N. Nishimura et al., Gene expression analysis in colorectal cancer using practical DNA array filter, *Dis. Colon Rectum* **44** (2001), 295–299.

- [33] R. Patarca, R.A. Saavedra and H. Cantor, Molecular and cellular basis of genetic resistance to bacterial infection: the role of the early T-lymphocyte activation-1/osteopontin gene, *Crit. Rev. Immunol.* **13** (1993), 225–246.
- [34] H. Porte, E. Chastre, S. Prevot et al., Neoplastic progression of human colorectal cancer is associated with overexpression of the stromelysin-3 and BM-40/SPARC genes, *Int. J. Cancer* **64** (1995), 70–75.
- [35] J. Sabates-Bellver, L.G. Van der Flier, M. de Palo et al., Transcriptome profile of human colorectal adenomas, *Mol. Cancer Res.* **5** (2007), 1263–1275.
- [36] P. Schratl, J.F. Royer, E. Kostenis et al., The role of the prostaglandin D2 receptor, DP, in eosinophil trafficking, *J. Immunol.* **179** (2007), 4792–4799.
- [37] J. Sodek, B. Zhu, M.H. Huynh et al., Novel functions of the extracellular matrix proteins osteopontin and osteonectin/SPARC, *Connect. Tissue Res.* **43** (2002), 308–319.
- [38] I. Takemasa, H. Higuchi, H. Yamamoto et al., Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer, *Biochem. Biophys. Res. Commun.* **285** (2001), 1244–1249.
- [39] Tumor Analysis Best Practices Working Group, Expression profiling—best practices for data generation and interpretation in clinical trials, *Nat. Rev. Genet.* **5** (2004), 229–237.
- [40] L.G. Van der Flier, J. Sabates-Bellver, I. Oving et al., The intestinal Wnt/TCF signature, *Gastroenterology* **132** (2007), 628–632.
- [41] B. Vogelstein, E.R. Fearon, S.R. Hamilton et al., Genetic alterations during colorectal-tumor development, *N. Engl. J. Med.* **319** (1988), 525–532.
- [42] Y. Wen, S.F. Giardina, D. Hamming et al., GROalpha is highly expressed in adenocarcinoma of the colon and down-regulates fibulin-1, *Clin. Cancer Res.* **12** (2006), 5951–5959.