

Research article

Open Access

Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesis

Esther H Lips¹, Ronald van Eijk¹, Eelco JR de Graaf⁴, Jan Oosting¹, Noel FCC de Miranda¹, Tom Karsten⁵, Cornelis J van de Velde², Paul HC Eilers³, Rob AEM Tollenaar², Tom van Wezel¹ and Hans Morreau^{*1}

Address: ¹Department of Pathology, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands, ²Department of Surgery, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands, ³Department of Medical Statistics, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands, ⁴Department of Surgery, IJsseland Hospital, P.O. Box 690, 2900 AR Capelle a/d IJssel, the Netherlands and ⁵Department of Surgery, Reinier de Graaf Gasthuis, P.O. Box 5011, 2600 GA Delft, the Netherlands

Email: Esther H Lips - estherlips@hotmail.com; Ronald van Eijk - r.van_eijk@lumc.nl; Eelco JR de Graaf - EdGraaf@ysl.nl; Jan Oosting - j.oosting@lumc.nl; Noel FCC de Miranda - n.f.miranda@lumc.nl; Tom Karsten - karsten@rdgg.nl; Cornelis J van de Velde - c.j.h.van_de_velde@lumc.nl; Paul HC Eilers - p.h.c.eilers@uu.nl; Rob AEM Tollenaar - r.a.e.m.tollenaar@lumc.nl; Tom van Wezel - t.van_wezel@lumc.nl; Hans Morreau^{*} - j.morreau@lumc.nl

^{*} Corresponding author

Published: 29 October 2008

Received: 28 May 2008

BMC Cancer 2008, **8**:314 doi:10.1186/1471-2407-8-314

Accepted: 29 October 2008

This article is available from: <http://www.biomedcentral.com/1471-2407/8/314>

© 2008 Lips et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Accurate staging of rectal tumors is essential for making the correct treatment choice. In a previous study, we found that loss of 17p, 18q and gain of 8q, 13q and 20q could distinguish adenoma from carcinoma tissue and that gain of 1q was related to lymph node metastasis. In order to find markers for tumor staging, we searched for candidate genes on these specific chromosomes.

Methods: We performed gene expression microarray analysis on 79 rectal tumors and integrated these data with genomic data from the same sample series. We performed supervised analysis to find candidate genes on affected chromosomes and validated the results with qRT-PCR and immunohistochemistry.

Results: Integration of gene expression and chromosomal instability data revealed similarity between these two data types. Supervised analysis identified up-regulation of *EFNA1* in cases with 1q gain, and *EFNA1* expression was correlated with the expression of a target gene (*VEGF*). The *BOPI* gene, involved in ribosome biogenesis and related to chromosomal instability, was over-expressed in cases with 8q gain. *SMAD2* was the most down-regulated gene on 18q, and on 20q, *STMN3* and *TGIF2* were highly up-regulated. Immunohistochemistry for *SMAD4* correlated with *SMAD2* gene expression and 18q loss.

Conclusion: On basis of integrative analysis this study identified one well known CRC gene (*SMAD2*) and several other genes (*EFNA1*, *BOPI*, *TGIF2* and *STMN3*) that possibly could be used for rectal cancer characterization.

Background

Accurate staging of rectal tumors is essential for choosing the correct treatment. Small pedunculated adenomas can be removed by snare excision, while large sessile adenomas can be cured by transanal endoscopic microsurgery [1]. For carcinomas, total mesorectal excision with preoperative radiotherapy is the gold standard [2]. However, preoperative staging using histology and modern imaging techniques is not always adequate, resulting in either under- or over-treatment. Therefore in current practice, additional markers indicating the aggressiveness of the tumor to be resected are extensively investigated [3,4]. It is of utmost importance to have parameters that can discriminate large benign adenomas from adenomas with a small invasive focus, as well as carcinomas with and without lymph node metastasis.

Recently, studies have investigated the application of microarrays in the diagnosis and prognosis of various stages of colorectal cancer (CRC). Gene expression signatures have been published that discriminate adenomas from carcinomas, Dukes B and C CRC, as well as lymph node positive and negative CRC patients [5-8]. Other studies using array comparative genomic hybridization (aCGH) describe specific genomic alterations related to different stages of colorectal cancer [9-12]. While there is little overlap between gene lists obtained from expression studies, common genomic alterations involved in CRC progression are established [13,14]. Chromosome 8q, 13q, and 20 gain occur early in the establishment of primary CRCs, loss of 4p is associated with the transition from Dukes' A to B-D. Deletion of 8p and gains of 7p and 17q are correlated with the transition from primary tumor to liver metastasis, whereas losses of 14q and gains of 1q, 11, 12p, and 19 are late events (reviewed in [14]).

Several studies have previously integrated gene expression profiles and genomic alterations in CRC and found a good correlation between both data types [15-19]. Tsafir et al. [19] found that often, large chromosomal segments, containing multiple genes, are transcriptionally affected in a coordinated way, and suggested that the underlying mechanism is a corresponding change in DNA content. Furthermore, they showed that these aberrations are absent in normal colon mucosa, appear in benign adenomas, become more frequent as disease advances, and are found in the majority of metastatic samples. In contrast, Platzer et al. found that underexpression was more common than overexpression in amplified regions [20], and a study by Staub et al. found that deleted regions usually show underexpression while amplified regions exhibit heterogeneous expression. For several gene islands of deregulated expression chromosomal aberrations have never been observed [18].

While those studies mainly analyzed how chromosomal aneuploidies affect global gene expression, we used an integrative approach to identify specific candidate genes for staging rectal tumors. In a previous study, we showed that loss of 17p, 18q12-22 and gain of 8q22-24, 13q and 20q could accurately distinguish adenoma from carcinoma tissue, and that gain of 1q23 was correlated to lymph node metastasis [21]. In the present study, we identify target genes on the affected chromosomes and validate the microarray data by means of quantitative RT-PCR and immunohistochemistry. We believe that this integrative approach generates more accurate and robust data than either data type alone.

Methods

Samples

Sixty-six fresh-frozen operated tumor samples were derived from a previous study in which copy number aberrations were determined [21]. In addition, material from 13 other cases was obtained. The samples were from patients treated by TEM from the IJsselland Hospital and Reinier de Graaf Hospital, the Netherlands, or from the TME trial, a Dutch multicenter trial in which 1530 patients were included from 1996 to 1999 [2]. None of the patients received radiotherapy or other adjuvant therapy. All samples were reviewed by a pathologist (H.M.), dysplasia was scored, and tumor percentage was assessed (50-80%) in a frozen section of the tissue. Intramucosal carcinomas were considered as adenoma with high grade dysplasia, as opposed to invasive carcinoma [22]. The local medical ethical committee approved the study (protocol number P04.124).

RNA isolation

Tumors were macrodissected in a cryostat by removing surrounding non-neoplastic tissue. Twenty 30- μ m sections were cut from each tumor. To guide microdissection, a 4- μ m section was cut and haematoxylin and eosin stained, before the first section, and after the tenth and twentieth section, and assessed for the presence of adenoma or carcinoma tissue, or a mixture of both. RNA was isolated with RNazol reagent (Tel-Test Inc., Friendswood, TX) according to the manufacturer's protocol and was purified using the Qiagen RNeasy mini kit with on-column DNase digestion, according to manufacturer's instructions (Qiagen Sciences, Germantown, MD). The quality of the RNA was assessed with lab-on-a-chip using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, California).

Microarray analysis

Two μ g of total RNA was amplified and labeled using Ambion's Amino Alkyl MessageAmp™ aRNA kit and protocol (Ambion Inc., Austin, TX). The quality of each aRNA was checked by lab-on-a-chip (Agilent Technologies). Dye

incorporation was checked with a Nanodrop (Wilmington, DE). For each microarray experiment, 2.0- μ g aliquots of aRNA were labeled with Cy5 (Amersham Biosciences, Buckinghamshire, UK). The labeled aRNAs were mixed with equal amounts of Cy3-labeled reference aRNA, consisting of pooled RNAs isolated from five colorectal cancer cell lines (HCT116, LS411N, SW480, HCT15, Caco2) and five normal rectum samples. To the mixture of labeled reference and sample RNA, 20 μ g human COT-1 DNA (Invitrogen, Carlsbad, CA), 8 μ g yeast tRNA (Invitrogen) and 20 μ g polyadenylic acid (Sigma-Aldrich, St. Louis, MO) were added. Preheated hybridization buffer (25% formamide, 5 \times SSC, 0.1% SDS) was added just before overnight hybridization at 42°C to human 35 K oligo microarrays, manufactured at the Central Microarray Facility (CMF) of the Netherlands Cancer Institute. Protocols, GeneID list and information about arrays are available at the website of the CMF <http://microarrays.nki.nl>. Hybridization slides were washed and scanned using the Agilent G2565BA Microarray Scanner (Agilent Technologies); spot intensities were extracted from the scanned images with Genepix 5.1 (Axon, Baden, Switzerland).

Data analysis

Raw intensity data (.gpr files) were analyzed in the R environment <http://www.r-project.org>. The Limma (linear models for microarray data) package of Bioconductor <http://www.bioconductor.org> was used for importing the data, normalizing the arrays and identifying differentially expressed genes. Control spots and spots with more than 10% saturation, a diameter smaller than 60 μ m or signal intensity less than 20 counts above background were excluded from the analysis. Data were corrected for local background (method normexp) and normalized within arrays by print-tip loess normalization and between arrays by quantile normalization. Duplicate experiments were performed for eight different tumor samples, showing Pearson correlation coefficients ranging from 0.92 to 0.97.

Statistically significant differences in gene expression were assessed using a moderate empirical Bayes test statistic available through Limma. The B-value is the log-odds that a gene is differentially expressed. The obtained *p*-values were controlled for false discovery using the Benjamini and Hochberg procedure. Oligos with corrected *p*-values \leq 0.001 were considered statistically significant.

In the integrated analysis, the gene expression levels were normalized per gene by subtracting the average gene expression of a reference sample set consisting of the adenomas with a limited amount of genomic changes (maximum of two small aberrations). Chromosomal plots of expression values were made in R by smoothing and integrated analysis [21,23]. Heat maps of expression data of

specific chromosomes were generated in Spotfire DecisionSite (Spotfire, Somerville, MA). For supervised analysis, we used Statistical Analyses of Microarrays (SAM) [24]. We analyzed every affected chromosome arm separately in SAM to find specific genes related to that specific chromosomal alteration. Groups were made on the basis of loss or gain and retention of a specific chromosome, determined by SNP array analysis [21]. For the analysis of gene expression levels of individual samples of newly identified and well-known colorectal cancer genes *t*-tests were performed using SPSS 12.0 (SPSS Inc, Chicago, IL, USA).

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/geo/>. The genomic data from the SNP arrays are accessible through GEO Series accession number GSE7946, while the gene expression array data are accessible through GEO Series accession number GSE12225.

Quantitative RT-PCR (qPCR)

Two micrograms of total RNA was reverse-transcribed with AMV Reverse Transcriptase (Roche, Penzberg, Germany). Real-time reverse transcriptase (RT) PCR was carried out in an 7900 HT Real Time PCR System (PE Applied Biosystems, Foster City, CA) in a 10 μ l volume containing 1 \times qPCR SYBR Green/ROX PCR Mastermix (SuperArray, Frederick, MD) and 1 μ l RT² primer set using the following PCR profile: 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. Primers used for real-time RT-PCR were targeted against *SMAD2*, *VEGF*, *EFNA1*, *BOP1*, and *STMN3*. Primer sequences for the target gene *SMAD2* were 5'-ATTGCTGCTCTCTGGCTCAG-3' and 5'-ACTTGTTACCGTTCGCTTCG-3' and for *VEGF* 5'-AAACCTGAGGGAGGCTCC-3' and 5'-TACTTGCA-GATGTGACAAGCCG-3'; for *EFNA1*, *BOP1*, and *STMN3*, we used RT² PCR Primer sets (SuperArray, Frederick, MD). Candidate genes for normalization were selected on the basis of showing the least variation between all samples (*CPSF6*, *GAPDH* and *EEF1A*). In all 79 samples the expression of these three genes was measured. Normalization was based on geometric averaging of the candidate normalization genes, as previously described [25], to acquire a reliable normalization of the qPCR experiments. This method provides a normalization factor (NF), representative of the amount of mRNA in each sample. Subsequently the expression of the gene of interest was divided by this normalization factor. The obtained normalized expression data were log₂ transformed and analyzed in SPSS 12.0 (SPSS Inc, Chicago, IL, USA).

Immunohistochemistry

BOP1 staining was performed on 4 μ m thick fresh frozen tumor sections, using standard procedures. SMAD4 stain-

ing was performed on tissue micro-arrays (TMA), as previously described [26]. For the formalin fixed paraffin embedded tissue antigen retrieval was performed by boiling the slides for 10 min in Tris-EDTA pH 8.0 (SMAD4) using a microwave oven, after which the sections were cooled in this buffer for at least 2 h at room temperature. TMA sections were then rinsed in demineralized water and phosphate buffered saline (PBS). The frozen tissue sections were incubated for one hour with a 1:100 dilution of BOP1 cell supernatant (Ascension, Munich, Germany) and the TMA sections with SMAD4 (clone B-8, sc-7966, Santa Cruz Biotechnology, Santa Cruz, CA; dilution 1:100). Sections were washed in PBS and incubated with biotinylated rabbit anti-rat (1:200; DAKO, Glostrup, Denmark) and streptavidin-biotin complex (1:100; DAKO) (BOP1) or Envision HRP-ChemMate kit (DAKO) (SMAD4) for 30 min. Diaminobenzidine tetrahydrochloride was used as a chromogen for BOP1 staining. All tumor specimens were stained simultaneously to avoid interassay variation. BOP1 staining was categorized as no expression (IHC-score 0), weak expression (1), moderate expression (2) and strong expression (3). SMAD4 was scored in the following categories: no nuclear staining with a positive internal control (total loss) (0), weak nuclear staining (down regulation) (1), and moderate to strong nuclear staining (positive) (2, 3). The mean expression of three punches per patient was assessed for SMAD4.

Results and discussion

Sample description

In a previous study, we built a rectal cancer progression model based on five "malignant" genomic alterations (loss of chromosomes 17p and 18q12-22 and gain of chromosomes 8q22-24, 13q, and 20q) [21]. In addition, gain of 1q23 was associated with lymph node metastasis. We assumed that integrating genomic and gene expression data would allow the identification of important genes for rectal tumor staging. Therefore, we obtained gene expression profiles from 66 samples, which were also typed for LOH and copy number abnormalities in the previous study [21]. From 13 additional samples, only gene expression measurements were available. Adenoma tissue was subdivided into pure adenomas (A/A) and adenoma fractions from cases with a carcinoma focus (A/C). The carcinoma tissue was subdivided into tumor fractions consisting of a mixture of adenoma and carcinoma tissue (AC/C), carcinomas without lymph node metastasis (C/C) and carcinomas with lymph node metastasis (C/C (N+)). Sample characteristics and genomic data are summarized in Table 1.

Global analysis of gene expression and genomic data

First, we performed a global analysis of the similarity between gene expression and genomic data as a verification of the data. We determined if gene expression

changes between tumor stages were also differently regulated on the chromosomal level. Analysis of the chromosomal location of 2853 genes that showed a trend in expression over the subsequent tumor stages (A/A- A/C- AC/C- C/C- C/C(N+)) ("progression genes", Additional file 1) revealed that genes on chromosome 18q were most frequently down-regulated and genes on chromosome 20q were most frequently up-regulated (Figure 1A), which was expected based on our genomic data [21]. Heat maps of gene expression patterns for particular chromosomes were made. A representative heat map for all the genes on chromosomes 18q showed that samples with 18q loss had a lower gene expression than samples with 18q retention (Figure 1B). Also concordance in individual samples was seen: gene expression data was plotted along the chromosome and compared to the patterns obtained by the genomic arrays (Figure 1C). Although the patterns are not exactly similar, for many chromosomes a clear resemblance is observed.

As verification we also determined if gene expression levels of newly identified genes from the progression analysis and well-known colorectal cancer gene sets were also differently regulated on the chromosome level. We focused on the genes on chromosome 1q, 8q, 13q, 17p, 18q and 20q. We compared gene expression values for the samples with gain or loss of the respective chromosome versus the samples with retention. The progression genes on chromosomes 1, 8, 13, and 20 respectively, show a higher expression in the samples with gain, while genes on chromosome 18 show a lower expression in the samples with 18q loss (Additional files 1 and 2). Some of the colorectal cancer genes show a significant change in the same direction as the chromosomal change (e.g. *SMAD4*, *MUC1*, *BCL2*) while for others this is not evident (i.e. *p53*, *BMP7*). It is worth mentioning that for several of these well known colorectal cancer genes possible differential expression cannot be detected since their expression level is too low on the microarrays.

Further identification of potential candidate genes in altered genomic regions

To identify specific genes of pathologic relevance in the affected chromosomal regions, we performed supervised analysis using the Significance Analysis of Microarrays package [24], with groups based on the specific chromosomal alterations. This analysis was done for the five "malignant" chromosomes and 1q. We examined whole chromosome arms as many patients had lost the whole arm. With a minimal fold change of 1.5 and a false discovery rate (FDR) <10%, we identified, respectively, 39, 30, 38, 20, 36, and 32 significant genes in relation to 1q gain, 8q gain, 13q gain, 17p loss, 18q loss and 20q gain (Additional file 1). All expression changes were in the expected direction, with the gain of 8q as an exception, showing

Table 1: Summary of clinical and pathological data of 79 tumor samples

	A/A	A/C	AC/C	C/C	C/C(N+)
Tissue fraction analyzed	Adenoma	Adenoma	Adenoma and carcinoma mixture	Carcinoma	Carcinoma
Tumor stage	Adenoma	Carcinoma	Carcinoma	Carcinoma	Carcinoma
Treatment					
TEM	24	6	5	3	1
TME	4	3	2	20	11
Sex (M/F)	15/13	2/7	3/4	10/13	10/2
Age (mean)	69	71	64	66	62
Dysplasia (adenoma)					
Low	15				
high	9				
Stage (carcinoma)					
T1		7	3	11	
T2		2	3	12	12
T3			1		
Size(cm) (mean)	5.9	4.4	4.2	3.2	5.1
Genomic data					
Samples typed(n)	21	8	7	19	11
Abberations (%)					
1q23 gain	0	9	11	14	62
8q22-24 gain	9	18	44	50	62
13q gain	4	36	67	59	85
17p loss	17	18	44	91	62
18q12-22 loss	17	36	56	86	77
20q gain	17	27	78	86	92

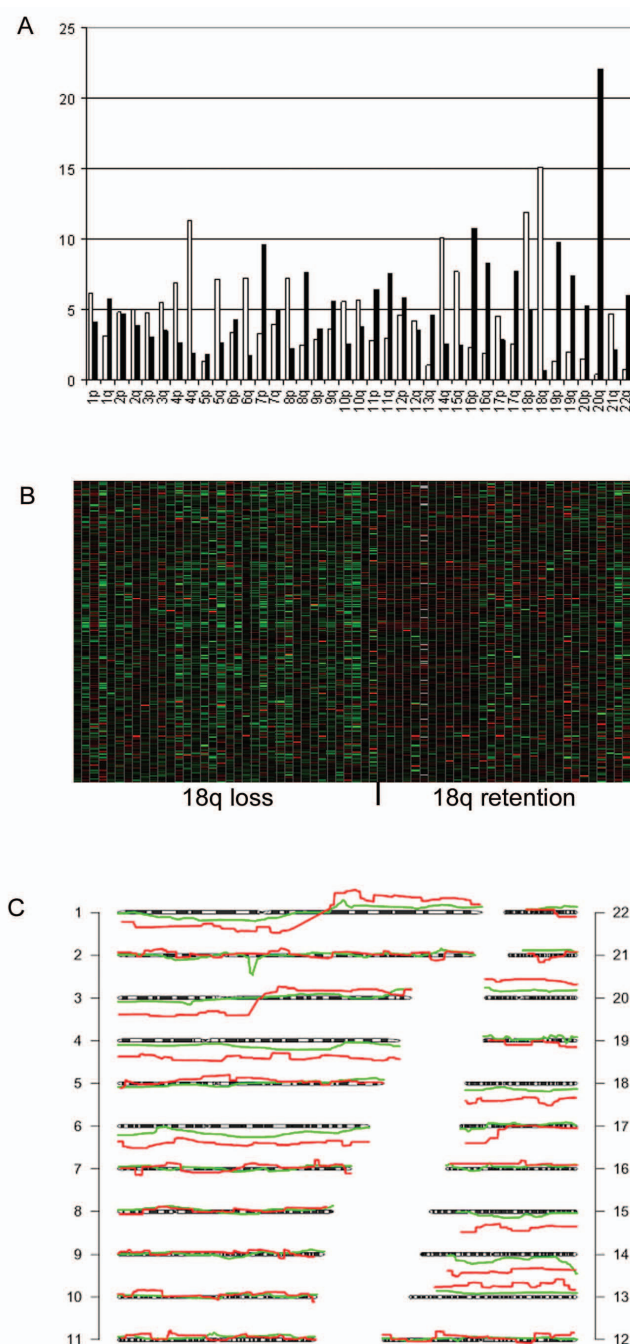
not only 30 up-regulated genes, but also 3 genes that were down-regulated. The genes on chromosome 20q had the highest fold change.

We next focused on four genes identified in the supervised analyses. *EFNA1* on 1q, *BOP1* on 8q, *SMAD2* on 18q and *STMN3* on 20q, were selected, based on a high fold change and a low false discovery rate (Additional file 1). These genes were in the specific listed regions (*SMAD2* at 18q21, *BOP1* at 8q24), with an exception for 1q, as most tumors with gene expression values had lost chromosome 1q entirely. Moreover, these genes were previously shown to be involved in (colorectal) cancer [27-30]. qPCR and immunohistochemistry were used to validate their expression in individual cases.

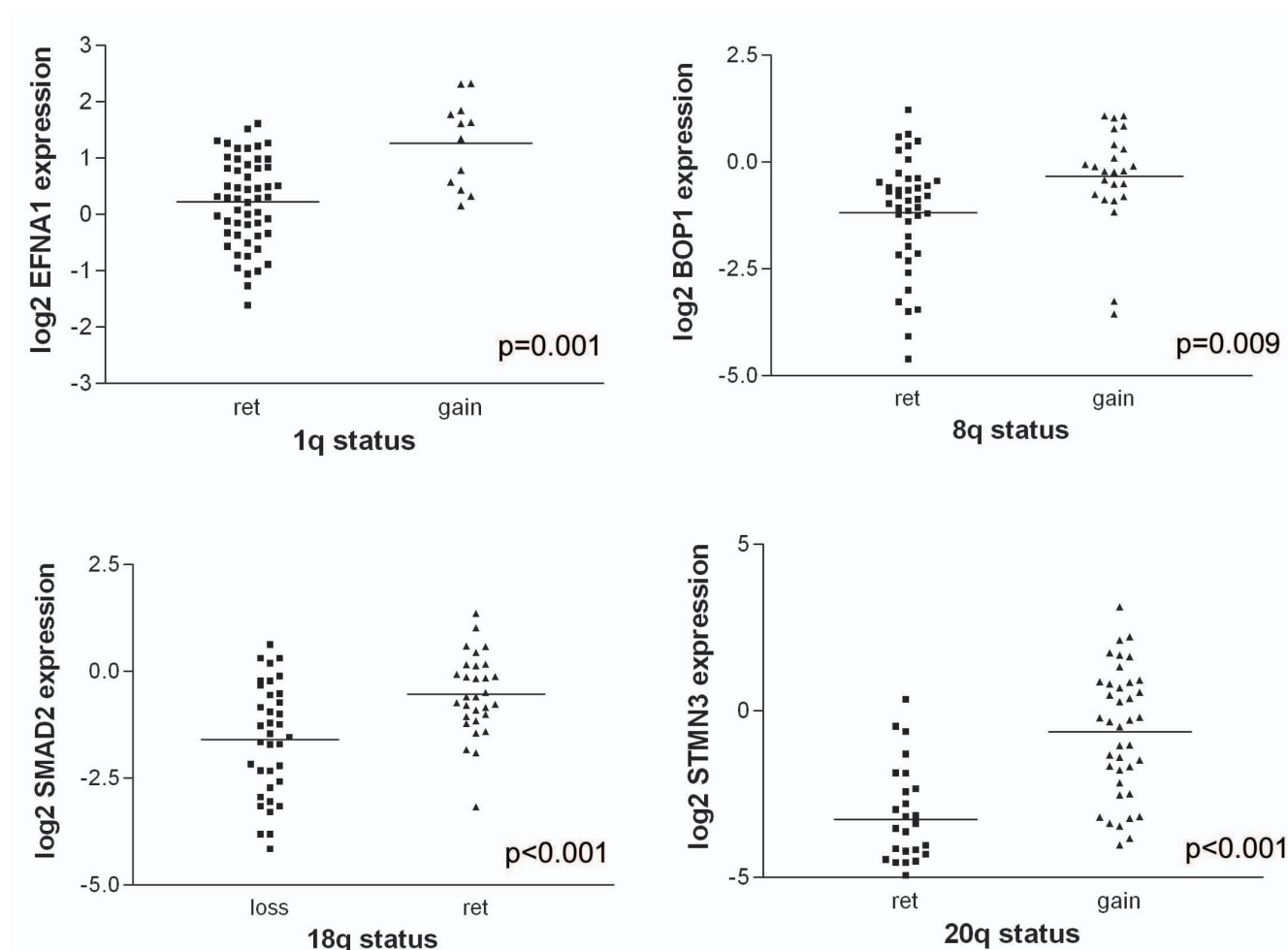
Validation by qPCR

To confirm the association between chromosomal aberrations and specific genes, we performed validation of expression data by qPCR. Correlation coefficients

between expression array data and qPCR data were 0.71, 0.49, 0.81 and 0.91 ($p < 0.001$, for all four genes) for *EFNA1*, *BOP1*, *SMAD2* and *STMN3*, respectively. Additionally the relation between specific genes and genomic regions was validated: *SMAD2* was less expressed in the samples with 18q genomic loss ($p < 0.001$), while *EFNA1*, *BOP1* and *STMN3* were all more expressed in samples with gains of 1q, 8q and 20q, respectively ($p = 0.001$, $p = 0.009$ and $p < 0.001$) (Figure 2). When the different sample groups were compared for expression of these four genes, their pattern of expression accompanied the genomic alterations: *SMAD2* was less expressed in the carcinomas, while *EFNA1*, *BOP1* and *STMN3* all showed an increased expression in the malignant tumor fractions; *EFNA1* was also notably expressed in the A/C fractions (Figure 3). The chromosomal changes and accompanying gene expression changes show thus a good trend with the increasing malignancy of the tumor stages.

**Figure 1**

Visual depiction of the similarity between gene expression and genomic data. A. Summary of gene expression data for all 79 samples. The distribution of 2853 progression genes over the chromosome arms is shown. The x-axis shows all chromosome arms, the y-axis shows the percentage of genes for a certain chromosome arm that is differentially expressed. White bars represent downregulated genes, black bars represent upregulated genes. B. Heat map of all gene expression data for the 66 samples with 18q genomic data available. Every column represents a single sample. Samples on the left side show loss of 18q, while samples on the right side show retention of chromosome 18q, both measured by SNP arrays. The y-axis shows all 18q genes from the centromere to telomere. Red lines indicate genes with a higher expression, green lines a lower expression. C. Chromosomal plot of sample 203 (carcinoma with lymph node metastasis) based on SNP array data (red) and gene expression array data (green).

**Figure 2**

Validation of array data by RT-PCR. Plots of relative gene expression (log2 values) measured with RT-PCR are shown. The x-axis shows the samples, the y-axis shows the log2 relative gene expression value. Samples with retention are compared with samples with loss (18q) or gain (1q, 8q and 20q) of a specific chromosome arm. The line indicates the mean. P-values were computed by Student's t-test.

1q gain and EFNA1

Previously, we found that gain of 1q might be related to lymph node metastasis in rectal cancer [21]. Samples with gain of 1q showed a higher expression of two probes for *EFNA1* than samples with 1q retention (Additional file 1 and Figure 2). *EFNA1* is a ligand for Eph receptor tyrosine kinases and plays a key role in the migration and adhesion of cells during development [31]. Recently, it was found to be related to tumor-induced neovascularization [27]. Brantley-Sieders et al. found that *EFNA1* over-expression elevated vascular endothelial growth factor (VEGF) levels, suggesting that *EFNA1*-mediated modulation of the VEGF pathway is a mechanism by which *EFNA1* regulates angiogenesis [32]. VEGF plays a key role in angiogenesis during tumor growth and metastasis [33]. We measured

VEGF mRNA expression by qPCR and found that it was correlated with *EFNA1* mRNA expression ($r = 0.353$, $p = 0.002$) (Figure 4). *EFNA1* and VEGF showed increased expression in the carcinomas in comparison to adenomas. This was expected, as neo-angiogenesis is an important factor in malignant transformation. However, in lymph node positive carcinomas VEGF expression appears to be down-regulated as compared to node negative rectal cancers whereas such a trend is not observed for *EFNA1*. An explanation for this observation does not seem obvious. As discussed, the contribution of stromal cells to the tumor cell gene expression profile seems small [34]. A possible increase in stromal content cannot provide an explanation for this result. The effect at the protein level should also be assessed. Furthermore an alternative might

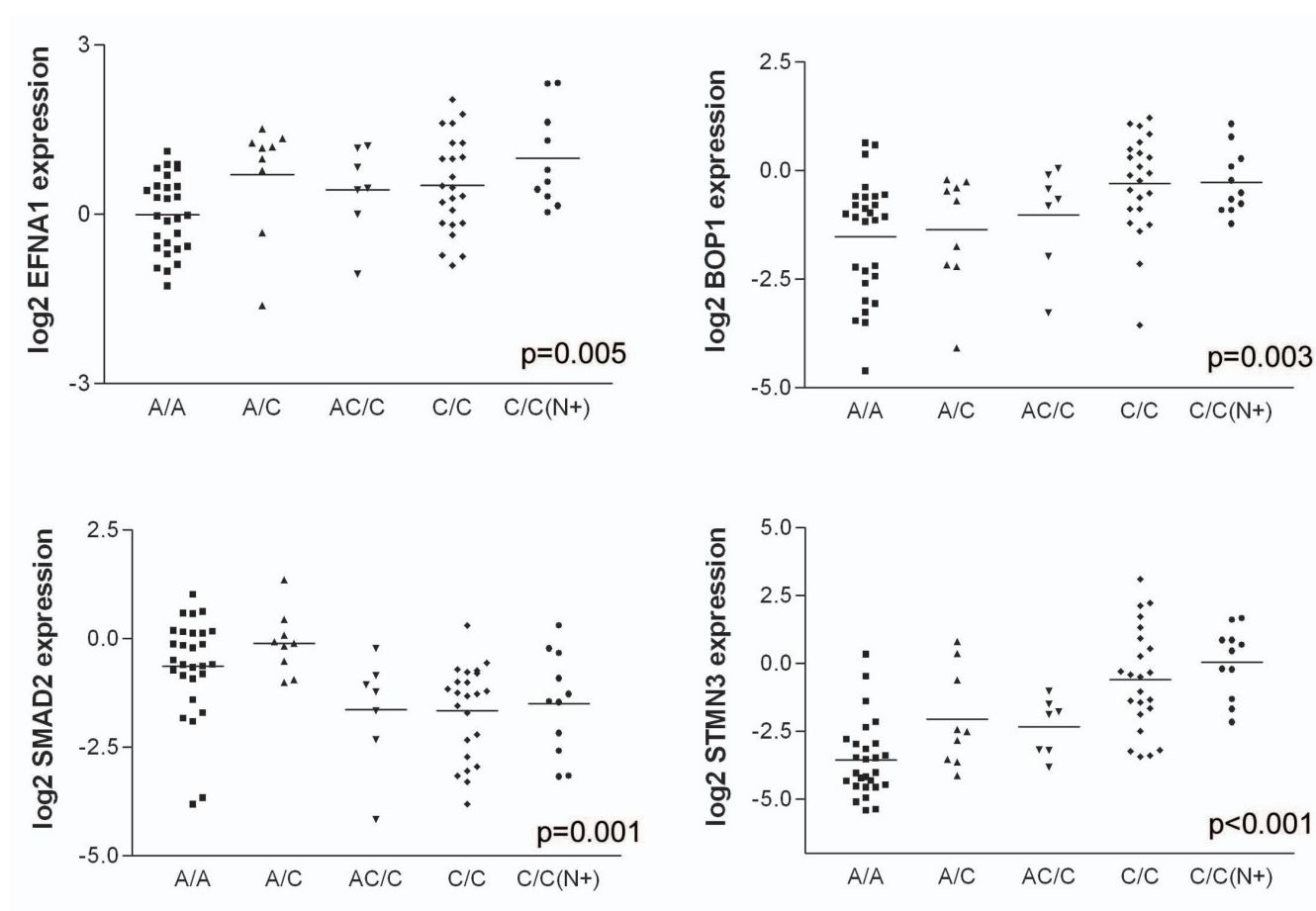


Figure 3

Expression of *EFNA1*, *BOP1*, *SMAD2* and *STMN3* in the different patient groups. Plots of relative gene expression (log2 values) measured with RT-PCR are shown. The x-axis shows the different sample groups. The y-axis shows the relative log2 gene expression data. Lines indicate the means. According to ANOVA analysis, the expression of all genes was significantly different between the groups.

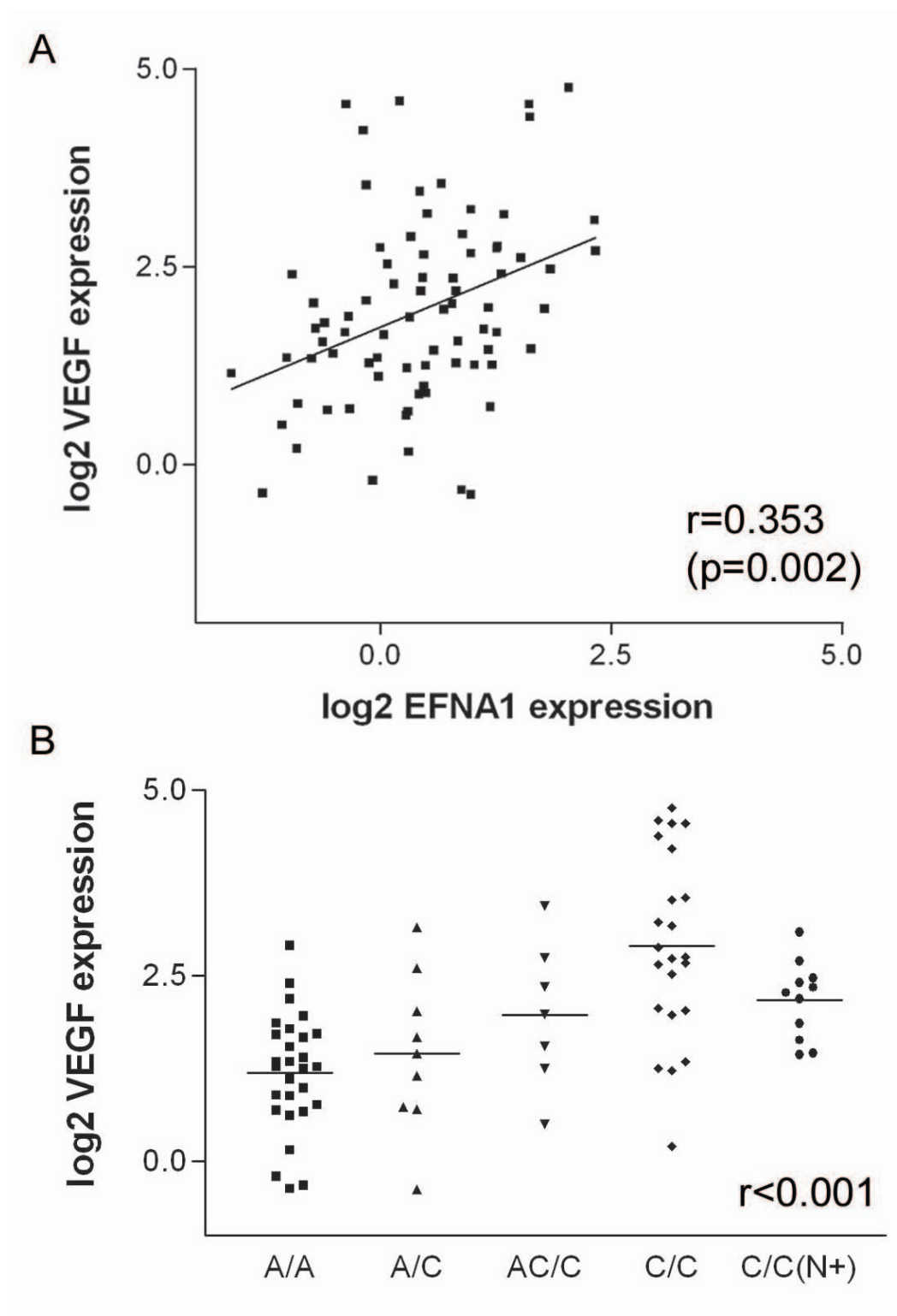
be experimental bias due to the small sample size of node positive rectal cancer cases.

8q gain and *BOP1*

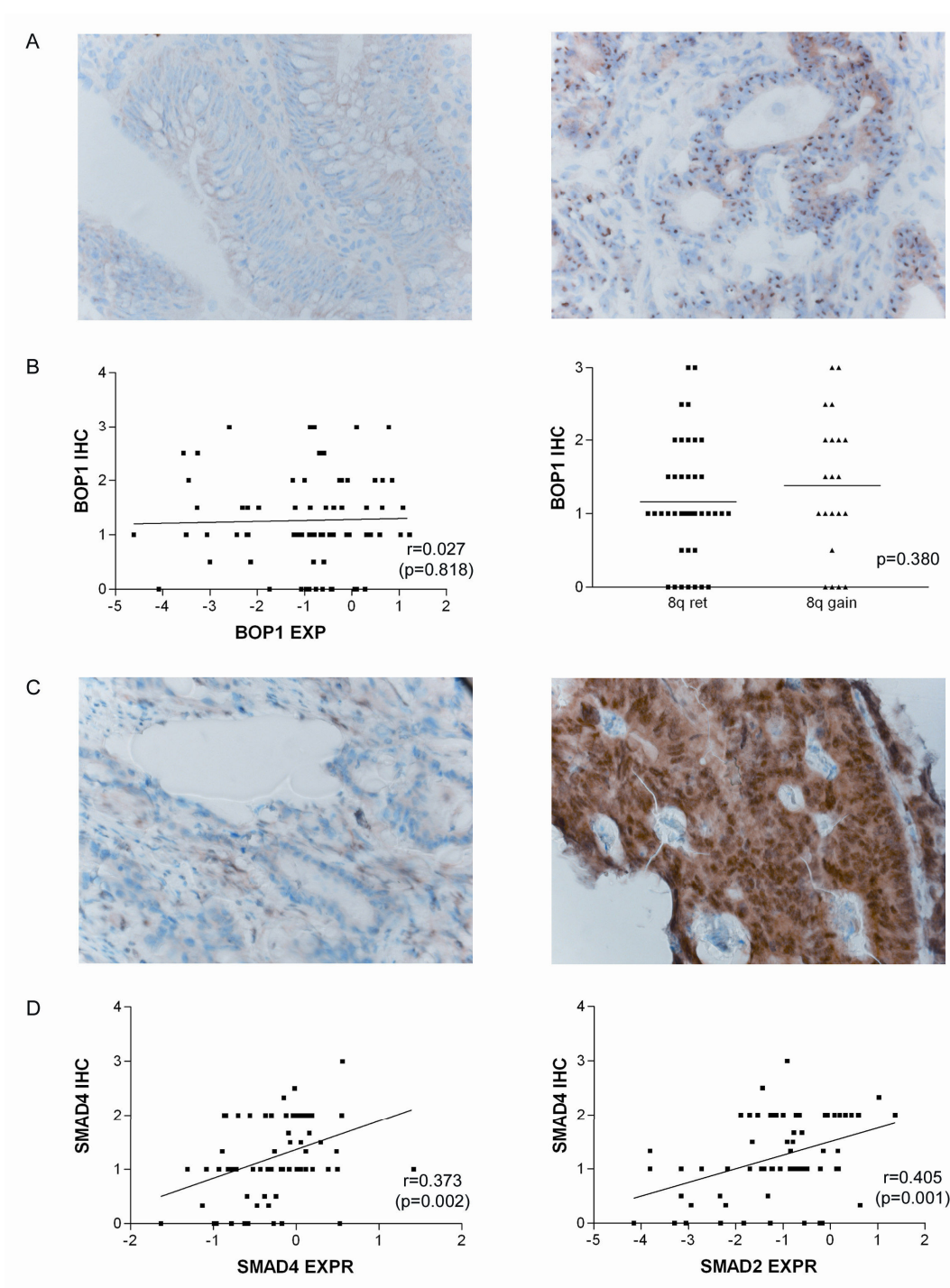
Gain of chromosome 8q and a higher expression of *BOP1* were both observed in most carcinomas (Figure 2). *BOP1* is a member of the Pes1-Bop1 complex, involved in ribosome biogenesis [35]. Killian *et al.* proposed that *BOP1* deregulation leads to altered chromosome segmentation and chromosomal instability in colorectal cancer [29]. They showed that *BOP1* copy number increase was associated with *BOP1* gene over-expression, in concordance with our results. *BOP1* is located on 8q24, close to the *MYC* oncogene. However, gene dosage increase of *BOP1* was independent from that of *MYC* and was more frequent than *MYC* over-expression, suggesting that *BOP1* over-expression may be one of the main oncogenic consequences of 8q24 amplification in colorectal cancer. In our

data series, gain of 8q, and consequently *BOP1* over-expression, was predominantly observed in cases with high chromosomal instability.

BOP1 protein expression was measured through immunohistochemistry on rectal tumor tissue slides. Specific cases with high *BOP1* mRNA expression showed very intense nucleolar *BOP1* staining (Figure 5A), but a direct correlation between both parameters was not established (Figure 5B). Comparing the mean *BOP1* protein expression between samples with 8q retention and 8q gain revealed a slight, although not significant, increase in the samples with gain (1.38 vs 1.16 relative protein expression) (Figure 5B). Post-transcriptional and post-translational mechanisms are likely to influence protein expression, possibly blurring the correlation between mRNA and protein levels for this gene. In such a case, gene expression data and immunohistochemistry results must

**Figure 4**

Correlation of *EFNA1* with *VEGF* expression values. A. Correlation plot of relative log2 *EFNA1* mRNA expression (x-axis) and relative log2 *VEGF* mRNA expression (y-axis). B. Relative log2 *VEGF* expression (y-axis) in the different clinical groups (x-axis).

**Figure 5**

BOP1 and SMAD4 immunohistochemical staining. A. Example of BOP1 immunohistochemical staining in two carcinomas; the left picture shows a weak expression (IHC score 1), the right picture, a very strong expression (score 3). B. Correlation plot of *BOP1* mRNA expression (x-axis) and BOP1 protein expression (y-axis) (left) and BOP1 immunohistochemical staining (y-axis) related to 8q gain(x-axis) (right). C. Example of SMAD4 immunohistochemical staining in two carcinomas; the left picture shows loss of SMAD4 expression (score 1), the right picture positive expression (score 3). D. Correlation plot of *SMAD4* gene expression (x-axis) and SMAD 4 protein expression (y-axis) (left) and correlation plot of *SMAD2* gene expression (x-axis) and SMAD 4 protein expression (right) (y-axis).

be considered independently because each can provide clinically meaningful information [36]. Alternatively, the difference in gene expression might be too subtle to detect with immunohistochemistry.

18q loss and SMAD2

SMAD2, *SMAD4* and *DCC* are indicated as the prominent tumor suppressor genes on 18q [30,37]. We found that *SMAD2* was significantly less expressed in the cases with 18q loss (Figure 2), and *SMAD2* and *SMAD4* were both down-regulated in the more advanced tumor stages (Limma-analysis, data not shown). SMAD proteins mediate TGF- β signaling to regulate cell growth and differentiation [38]. LOH in combination with *SMAD4* mutations is a well studied phenomenon in CRC, and *SMAD4* gene mutations are related to advanced tumor stage [39,40]. Immunohistochemistry was technically not feasible for *SMAD2* but was feasible for *SMAD4*, which is also located on chromosome 18q21.1. Therefore, we tested whether *SMAD4* protein expression was correlated to *SMAD4* and *SMAD2* gene expression, which was indeed the case ($r = 0.373$ ($p = 0.002$) and $r = 0.405$, ($p = 0.001$)) (Figure 5C, D).

According to Knudson's "two hit hypothesis", both copies of a tumor suppressor gene should be deleted by a mutation or allelic loss to reduce protein dosage [41]. In our study, half of the cases showed physical loss of 18q and thus deletion of one of the *SMAD2* alleles. An additional hit such as a mutation can then be expected, leading to the observed reduction in protein expression. However, mutation analysis for *SMAD2* did not reveal any mutations in this sample series (data not shown). In the literature, mutation rates vary between 0 and 30% for *SMAD2* and *SMAD4* [37,42]. Recently, Alberici et al. showed haploinsufficiency for the *SMAD4* locus in mouse models for colorectal cancer, giving an explanation for the relatively low mutation rate observed [43]. Consequently, the loss of one allele already leads to reduced *SMAD4* protein expression and altered TGF- β signaling. The same principle might apply to *SMAD2* and explain our findings, where only one copy of 18q is lost and no mutation is found in the *SMAD2* gene, but reduced gene expression is observed.

Genes on chromosome 20q

Genes on chromosome 20 showed the highest fold change in expression in comparison with genes on the other chromosomes (Additional file 1). Two interesting genes were in the top five overexpressed genes: *STMN3* and *TGIF2*. *STMN3* is overexpressed in various human malignancies and plays a role in regulation of the cell cycle [28,44]. In oral squamous-cell carcinoma, the overexpression of *STMN3* was correlated with tumor progression and poor prognosis. Kouzu et al. emphasized the

potential role of *STMN3* as a biomarker and therapeutic target for oral squamous-cell carcinoma [28]. *TGIF2* was shown to interact with TGF- β -activated SMADS and repress TGF- β responsive transcription [45]. Limma-analysis revealed that *TGIF2* was among the ten most significant genes in the adenoma-carcinoma comparison. In ovarian cancer cell lines, amplification of 20q correlated strongly with *TGIF2* over-expression [46]. A recent study subtracted a chromosomal instability gene expression profile from 12 different cancer data sets. This 25 gene set contained, among others, *TGIF2*, indicating that this gene plays a role in chromosomal instability [47].

Conclusion

Several studies have integrated global patterns of gene expression and genomic data in colorectal cancer, with divergent results. While some studies reported a direct correlation between gene expression and chromosomal aberrations [16,17,19,48], others reported that amplification does not necessarily lead to expression up-regulation [20]. Similarly, genomic regions with deletions show reduced expression while amplified regions exhibit heterogeneous expression [18]. We first determined whether copy number alterations have an effect on gene expression and saw that changes in genomic regions and gene expression are usually in the same direction. We then performed supervised analysis to find target genes on the affected chromosomes. We identified one well known CRC gene (*SMAD2*) and several other genes (*EFNA1*, *BOP1*, *TGIF2* and *STMN3*) that possibly could be used for rectal cancer characterization. It will be interesting to determine if these molecules can discriminate benign adenomas from adenomas containing an invasive focus and carcinomas with lymph node metastasis. To validate the impact of these genes in rectal carcinogenesis it will be worthwhile to study them for mRNA expression and at the immunohistochemical level in representative cases with normal, adenoma and carcinoma tissue from single patients. Unfortunately frozen material of all those tumor stadia was not available for the subjects of this study. Studies in other cancer types have successfully applied supervised methods [49,50]. Garraway et al. performed supervised analysis on cell lines with and without 3p amplifications and identified *MITF* as a new melanoma oncogene.

In conclusion, gene expression values in regions with genomic changes are altered in the same direction. We analyzed gene expression data in relation to specific chromosomal aberrations involved in the progression from adenoma to carcinoma. By not focusing directly on tumor stages, but rather on genomic aberrations related to tumor stages, we identified several specific genes on altered chromosomes in rectal cancer. Specific genes, identified by such integration methods, could be of additional value to further explain rectal tumorigenesis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EL performed the experiments, data analysis and preparation of the manuscript. RE and NM assisted in the experiments. JO assisted in data analysis. EG, TK, CV provided the tumor material. PE, RT, TW, HM initiated the study and supervised the data generation and analyses. All authors read and approved the final manuscript.

Additional material

Additional file 1

Supplementary Tables. Table 1 contains the data of the Limma analysis for comparison of the different tumor groups (a) and the "progression" analysis (b). Supplementary Table 2 contains p-values for genes of the progression analysis on chromosome comparisons. Supplementary Table 3 contains the differentially expressed genes for the six genomic regions of interest.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-314-S1.xls>]

Additional file 2

Figure. Contains graphs of analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-314-S2.ppt>]

Acknowledgements

This project was supported by Dutch Cancer Society grant UL 2003-2807. We would like to thank Debora Lima for technical assistance and Michael Hölzel for providing the BOP1 protocol.

References

- Buess G, Mengtes B, Manncke K, Starlinger M, Becker HD: **Technique and results of transanal endoscopic microsurgery in early rectal cancer.** *Am J Surg* 1992, **163**:63-69.
- Kapiteijn E, Marijnen CA, Nagtegaal ID, Putter H, Steup WH, Wiggers T, Rutten HJ, Pahlman L, Glimelius B, van Krieken JH, Leer JW, Velde CJ van de: **Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer.** *N Engl J Med* 2001, **345**:638-646.
- Anwar S, Frayling IM, Scott NA, Carlson GL: **Systematic review of genetic influences on the prognosis of colorectal cancer.** *Br J Surg* 2004, **91**:1275-1291.
- Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, Somerfield MR, Hayes DF, Bast RC Jr: **ASCO 2006 Update of Recommendations for the Use of Tumor Markers in Gastrointestinal Cancer.** *J Clin Oncol* 2006.
- Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ: **Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays.** *Proc Natl Acad Sci USA* 1999, **96**:6745-6750.
- Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginstier C, Charafe-Jauffret E, Liorid B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D, Houlgatte R: **Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters.** *Oncogene* 2004, **23**:1377-1391.
- Ghadimi BM, Grade M, Difilippantonio MJ, Varma S, Simon R, Montagna C, Fuzesi L, Langer C, Becker H, Liersch T, Ried T: **Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy.** *J Clin Oncol* 2005, **23**:1826-1838.
- Notterman DA, Alon U, Sierk AJ, Levine AJ: **Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays.** *Cancer Res* 2001, **61**:3124-3130.
- Hermesen M, Postma C, Baak J, Weiss M, Rapallo A, Sciutto A, Roemen G, Arends JW, Williams R, Giaretti W, De Goeij A, Meijer G: **Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability.** *Gastroenterology* 2002, **123**:1109-1119.
- Hoglund M, Gisselsson D, Hansen GB, Sall T, Mitelman F, Nilbert M: **Dissecting karyotypic patterns in colorectal tumors: two distinct but overlapping pathways in the adenoma-carcinoma transition.** *Cancer Res* 2002, **62**:5939-5946.
- Leslie A, Stewart A, Baty DU, Mechan D, McGreavey L, Smith G, Wolf CR, Sales M, Pratt NR, Steele RJ, Carey FA: **Chromosomal changes in colorectal adenomas: relationship to gene mutations and potential for clinical utility.** *Genes Chromosomes Cancer* 2006, **45**:126-135.
- Ried T, Knutzen R, Steinbeck R, Blegen H, Schrock E, Heselmeyer K, du MS, Auer G: **Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors.** *Genes Chromosomes Cancer* 1996, **15**:234-245.
- Cardoso J, Boer J, Morreau H, Fodde R: **Expression and genomic profiling of colorectal cancer.** *Biochim Biophys Acta* 2007, **1775**:103-137.
- Diep CB, Kleivi K, Ribeiro FR, Teixeira MR, Lindgjaerde OC, Lothe RA: **The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes.** *Genes Chromosomes Cancer* 2006, **45**:31-41.
- Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF: **Frequent occurrence of uniparental disomy in colorectal cancer.** *Carcinogenesis* 2007, **28**:38-48.
- Grade M, Ghadimi BM, Varma S, Simon R, Wangsa D, Barenboim-Stapleton L, Liersch T, Becker H, Ried T, Difilippantonio MJ: **Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/beta-catenin signaling pathway in human rectal carcinomas.** *Cancer Res* 2006, **66**:267-282.
- Habermann JK, Paulsen U, Roblick UJ, Upender MB, McShane LM, Korn EL, Wangsa D, Kruger S, Duchrow M, Bruch HP, Auer G, Ried T: **Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis.** *Genes Chromosomes Cancer* 2007, **46**:10-26.
- Staub E, Groene J, Mennerich D, Roepcke S, Klamann I, Hinzmann B, Castanos-Velez E, Mann B, Pilarsky C, Brummendorf T, Weber B, Buhr HJ, Rosenthal A: **A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer.** *Mol Cancer* 2006, **5**:37.
- Tsafir D, Bacolod M, Selvanayagam Z, Tsafir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, Gerald WL, Paty PB, Domany E, Notterman DA: **Relationship of gene expression and chromosomal abnormalities in colorectal cancer.** *Cancer Res* 2006, **66**:2129-2137.
- Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T, Markowitz S: **Silence of chromosomal amplifications in colon cancer.** *Cancer Res* 2002, **62**:1134-1138.
- Lips E, de Graaf E, Tollenaar R, van Eijk R, Oosting J, Suzhai K, Karsten T, Nanya Y, Ogawa S, Velde CJ van de, Eilers P, van Wezel T, Morreau H: **Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas.** *J Pathol* 2007, **212**:269-277.
- Rex DK, Ulbright TM, Cummings OW: **Coming to terms with pathologists over colon polyps with cancer or high-grade dysplasia.** *J Clin Gastroenterol* 2005, **39**:1-3.
- Kloth JN, Oosting J, van Wezel T, Suzhai K, Knijnenburg J, Gorter A, Kenter GG, Fleuren GJ, Jordanova ES: **Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer.** *BMC Genomics* 2007, **8**:53.

24. Tusher VG, Tibshirani R, Chu G: **Significance analysis of micro-arrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
25. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** *Genome Biol* 2002, **3**:RESEARCH0034.
26. Lips EH, van Eijk R, de Graaf EJ, Doornebosch PG, de Miranda NF, Oosting J, Karsten T, Eilers PH, Tollenaar RA, van Wezel T, Morreau H: **Progression and tumor heterogeneity analysis in early rectal cancer.** *Clin Cancer Res* 2008, **14**:772-781.
27. Kataoka H, Igarashi H, Kanamori M, Ihara M, Wang JD, Wang YJ, Li ZY, Shimamura T, Kobayashi T, Maruyama K, Nakamura T, Arai H, Kajimura M, Hanai H, Tanaka M, Sugimura H: **Correlation of EPHA2 overexpression with high microvessel count in human primary colorectal cancer.** *Cancer Sci* 2004, **95**:136-141.
28. Kouzu Y, Uzawa K, Koike H, Saito K, Nakashima D, Higo M, Endo Y, Kasamatsu A, Shiiba M, Bukawa H, Yokoe H, Tanzawa H: **Overexpression of stathmin in oral squamous-cell carcinoma: correlation with tumour progression and poor prognosis.** *Br J Cancer* 2006, **94**:717-723.
29. Killian A, Sarafan-Vasseur N, Sesboue R, Le PF, Blanchard F, Lamy A, Laurent M, Flaman JM, Frebourg T: **Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis.** *Genes Chromosomes Cancer* 2006, **45**:874-881.
30. Eppert K, Scherer SW, Ozcelik H, Pirone R, Hoodless P, Kim H, Tsui LC, Bapat B, Gallinger S, Andrulis IL, Thomsen GH, Wrana JL, Attisano L: **MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma.** *Cell* 1996, **86**:543-552.
31. Poliakov A, Cotrina M, Wilkinson DG: **Diverse roles of eph receptors and ephrins in the regulation of cell migration and tissue assembly.** *Dev Cell* 2004, **7**:465-480.
32. Brantley-Sieders DM, Fang WB, Hwang Y, Hicks D, Chen J: **Ephrin-A1 facilitates mammary tumor metastasis through an angiogenesis-dependent mechanism mediated by EphA receptor and vascular endothelial growth factor in mice.** *Cancer Res* 2006, **66**:10315-10324.
33. Hicklin DJ, Ellis LM: **Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis.** *J Clin Oncol* 2005, **23**:1011-1027.
34. de Bruin EC, Pas S van de, Lips EH, van Eijk R, Zee MM van der, Lombaerts M, van Wezel T, Marijnen CA, van Krieken JH, Medema JP, Velde CJ van de, Eilers PH, Peltenburg LT: **Macrodissection versus microdissection of rectal carcinoma: minor influence of stroma cells to tumor cell gene expression profiles.** *BMC Genomics* 2005, **6**:142.
35. Grimm T, Holzel M, Rohrmoser M, Harasim T, Malamoussi A, Gruber-Eber A, Kremmer E, Eick D: **Dominant-negative Pes1 mutants inhibit ribosomal RNA processing and cell proliferation via incorporation into the PeBoW-complex.** *Nucleic Acids Res* 2006, **34**:3030-3043.
36. Ginestier C, Charafe-Jauffret E, Bertucci F, Eisinger F, Geneix J, Bechlian D, Conte N, Adelaide J, Toiron Y, Nguyen C, Viens P, Mozziconacci MJ, Houlgatte R, Birnbaum D, Jacquemier J: **Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers.** *Am J Pathol* 2002, **161**:1223-1233.
37. Thiagalingam S, Lengauer C, Leach FS, Schutte M, Hahn SA, Overhauser J, Willson JK, Markowitz S, Hamilton SR, Kern SE, Kinzler KW, Vogelstein B: **Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers.** *Nat Genet* 1996, **13**:343-346.
38. Shi Y, Massague J: **Mechanisms of TGF-beta signaling from cell membrane to the nucleus.** *Cell* 2003, **113**:685-700.
39. Miyaki M, Iijima T, Konishi M, Sakai K, Ishii A, Yasuno M, Hishima T, Koike M, Shitara N, Iwama T, Utsunomiya J, Kuroki T, Mori T: **Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis.** *Oncogene* 1999, **18**:3098-3103.
40. Miyaki M, Kuroki T: **Role of Smad4 (DPC4) inactivation in human cancer.** *Biochem Biophys Res Commun* 2003, **306**:799-804.
41. Knudson AG Jr: **Mutation and cancer: statistical study of retinoblastoma.** *Proc Natl Acad Sci USA* 1971, **68**:820-823.
42. Takenoshita S, Tani M, Mogi A, Nagashima M, Nagamachi Y, Bennett WP, Hagiwara K, Harris CC, Yokota J: **Mutation analysis of the Smad2 gene in human colon cancers using genomic DNA and intron primers.** *Carcinogenesis* 1998, **19**:803-807.
43. Alberici P, Jagmohan-Changur S, de Pater E, van d V, Smits R, Hohenstein P, Fodde R: **Smad4 haploinsufficiency in mouse models for intestinal cancer.** *Oncogene* 2006, **25**:1841-1851.
44. Rubin CI, Atweh GF: **The role of stathmin in the regulation of the cell cycle.** *J Cell Biochem* 2004, **93**:242-250.
45. Melhuish TA, Gallo CM, Wotton D: **TGIF2 interacts with histone deacetylase I and represses transcription.** *J Biol Chem* 2001, **276**:32109-32114.
46. Imoto I, Pimkhaokham A, Watanabe T, Saito-Ohara F, Soeda E, Inazawa J: **Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines.** *Biochem Biophys Res Commun* 2000, **276**:264-270.
47. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z: **A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers.** *Nat Genet* 2006, **38**:1043-1048.
48. Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF: **Frequent occurrence of uniparental disomy in colorectal cancer.** *Carcinogenesis* 2006.
49. Adler AS, Lin M, Horlings H, Nuyten DS, Vijver MJ Van de, Chang HY: **Genetic regulators of large-scale transcriptional signatures in cancer.** *Nat Genet* 2006, **38**:421-430.
50. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhi R, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML, Fisher DE, Sellers WR: **Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma.** *Nature* 2005, **436**:117-122.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/8/314/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

