COLON CANCER

Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid

K Birkenkamp-Demtroder, S H Olesen, F B Sørensen, S Laurberg, P Laiho, L A Aaltonen, T F Ørntoft

Gut 2005;**54**:374–384. doi: 10.1136/gut.2003.036848

See end of article for authors' affiliations

Correspondence to: Professor T F Ørntoft, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital/Skejby, Brendstrupgaardsvej 100, DK- 8200 Aarhus N, Denmark; orntoft@ki.au.dk

Revised version received 8 June 2004
Accepted for publication 23 June 2004

Background and aims: There are epidemiological, morphological, and molecular differences between normal mucosa as well as between adenocarcinomas of the right and left side of the large bowel. The aim of this study was to investigate differences in gene expression.

Methods: Oligonucleotide microarrays (GeneChip) were used to compare gene expression in 45 single samples from normal mucosa and sporadic colorectal carcinomas (Dukes' B and C) of the caecum compared with the sigmoid and rectosigmoid. Findings were validated by real time polymerase chain reaction.

Results: Fifty eight genes were found to be differentially expressed between the normal mucosa of the caecum and the sigmoid and rectosigmoid ($p<0.01$), including pS2, S100P, and a sialyltransferase, all being expressed at higher levels in the caecum. A total of 118 and 186 genes were differentially expressed between normal and right or left sided tumours of the colon, showing more pronounced differences in Dukes' C than B tumours. Thirty genes differentially expressed in tumour tissue were common to adenocarcinomas of both sides, including known tumour markers such as the matrix metalloproteinases. Keratins 8, 19, and 20 as well as carbonic anhydrases (II, IV, VII) showed side specific expression and were downregulated in left sided tumours whereas teratocarcinoma growth factor and cyclooxygenase 2 (COX-2) were upregulated in left sided adenocarcinomas. Immunohistochemical analysis confirmed differences in side specific expression for cytokeratin 20 and COX-2.

Conclusions: Differences in gene expression between normal mucosa as well as between adenocarcinomas of the caecum and sigmoid or rectosigmoid exist and should be taken into account when examining new targeted therapeutic regimens.

Multiple differences between right sided (RCC) and left sided (LCC) sporadic colon adenocarcinomas with regard to epidemiological, morphological, and molecular characteristics suggest that the mechanisms of sporadic colorectal carcinogenesis may differ according to tumour location.[1] Cancers of the right and left colon may form different but related groups of tumours because of their different embryological origin (midgut and hindgut, respectively) and different exposure to bowel content. Colon cancer has a different prevalence at varying ages, in high and low incidence nations, and in men and women. RCCs are more common in females, LCCs in males.[2] There is also a difference in clinical presentation, in prognosis, and possibly in genetic and environmental epidemiology (see review by Iacopetta[3]). Furthermore, it has been suggested that a mechanism exists that promotes the progression of mucosal lesions to invasive cancers in the left colon and rectum whereas a de novo pathway from depressed type lesions may be implicated in cancers of the right colon.[4] No difference has been found in the distribution of Dukes' stages or in operative mortality between right and left sided sporadic colon cancers. Despite their higher tumour diameter and twofold higher rate of undifferentiated carcinomas, the prognosis of right sided tumours is relatively better than that of left sided tumours, and it has been hypothesised that this could be due to the better blood and lymph supply providing more efficient local tumour defence.[5] Recurrence and survival are similar between RCC and LCC[6] whereas response to 5-fluorouracil treatment is significantly better in RCC.[7]

Two studies suggest that molecular differences in gene expression exist between right and left sided colon cancers. Kapiteijn *et al* showed significantly higher expression of nuclear β-catenin and p53 in rectal cancers compared with proximal cancers.[8] Fric *et al* showed significantly higher expression of cytoplasmic c-erbB2, epidermal growth factor receptor (EGFR), proliferating cell nuclear antigen (PCNA), and dipeptidylpeptidase IV (DPP IV) in right sided sporadic colon cancers compared with left sided cancers.[9] Distal tumours display a higher frequency of 17p and 18q allelic loss, p53 accumulation, c-myc expression, and aneuploidy than proximal tumours. Recently, Glebov *et al* distinguished proximal from distal normal colon mucosa based on gene expression analysis.[10]

To the best of our knowledge there are no expression data available on differences between adenocarcinomas originating from the proximal or distal part of the colon. As this could have a strong impact on molecularly targeted cancer treatment, we wished to elucidate this aspect and gain insight into differential expression of approximately 7000 human genes of right sided and left sided Dukes' stage B and C adenocarcinomas as well as normal colon mucosa.

MATERIALS AND METHODS

Tissue samples, patient information, and RNA isolation

Tissue samples, patient information, and RNA isolation are provided in detail as supplementary data (these data can be viewed on the *Gut* website at http://www.gut.com/

Abbreviations: LCC, left sided colon cancer; RCC, right sided colon cancer; UG cluster, UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene); EGFR, epidermal growth factor receptor; PCNA, proliferating cell nuclear antigen; DPP IV, dipeptidylpeptidase IV; RT-PCR, reverse transcription-polymerase chain reaction; COX-2, cyclooxygenase 2; MSS, microsatellite stable; CA, carbonic anhydrase; MMP, matrix metalloproteinase

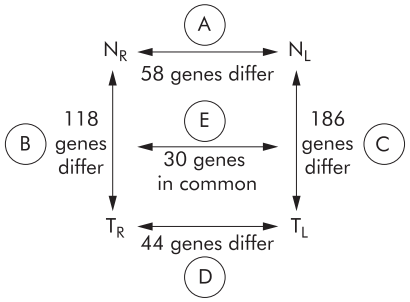

Figure 1 Schematic overview of the five different comparison groups (A–E). Comparison (A): normal right sided colon mucosa (N_R) from the caecum versus normal left sided (N_L) colon mucosa from the sigmoid and rectosigmoid. Comparison (B): normal right sided mucosa from the caecum (N_R) versus matched right sided tumours from the caecum (T_R). Comparison (C): normal left sided mucosa (N_L) from the sigmoid and rectosigmoid versus left sided tumour (T_L) from the same region of the colon. Comparison (D): right sided tumours (T_R) from the caecum versus left sided tumours (T_L) from the sigmoid and rectosigmoid. Comparison (E): expression differences from normal mucosa to adenocarcinomas that are common between caecum tumours and left sided tumours in the sigmoid and rectosigmoid (comparing (B) versus (C)).

supplemental). Samples from the caecum and rectosigmoid or sigmoid were obtained fresh from surgery and immediately transferred to a solution containing sodium dodecyl sulphate and guanidinium isothiocyanate, snap frozen in liquid nitrogen, and stored at −80°C.

Samples consisted of biopsies from the superficial non-necrotic part of tumours and/or normal mucosa biopsies taken from the oral resection margin. All tumour samples were staged as either Dukes' B (eight from the left colon, five from the caecum) or Dukes' C (seven from the left colon, five from the caecum).

Supplementary table 1 shows detailed clinicopathological information—for example, location of samples in the colon and TNM status (see the *Gut* website at http://www.gut.com/supplemental). All 15 left sided and 9/10 right sided tumours (90%) were invasive adenocarcinomas; one was an invasive mucinous adenocarcinoma. Six of 10 right sided tumours (60%) were moderately differentiated, 3/10 (30%) were poorly differentiated, and 1/10 (10%) was well differentiated. Ten of 15 left sided tumours (67%) were moderately differentiated, 4/15 (27%) were poorly differentiated, and 1/15 (6%) was well differentiated. The approximate percentages of the volume fractions of tumour cells and stromal cells were semi quantitatively estimated using paraffin embedded diagnostic tissue sections. More than half of the tumour samples showed more than 70% malignant cells. We hypothesise that the percentage of tumour cells is probably higher in the arrayed samples than in the screened paraffin embedded diagnostic histological tissue sections, as the latter represents the whole invasive tumour in the bowel wall. Informed consent was obtained from all patients. All tumours were sporadic. The local scientific ethics commission approved the project.

Total RNA was isolated from approximately 50 mg of single tissue samples using a Polytron homogeniser followed by treatment with Trizol (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instructions. GeneChip (Affymetrix Inc., Santa Clara, California, USA) analysis of single samples was carried out on 10 samples from the caecum (65B, 66B, 73B, 120B, 137B, 90C, 126C, 145C, 138C, and 162C), five Dukes' stage B (median age 76 years) and five Dukes' stage C (median age 66 years). Each of the tumours was accompanied by a corresponding matched normal mucosa sample at the same location from the same patient (median age 70 years). Matched samples were given the same sample number, differentiated by ''N'' for normal and ''B'' or ''C'' for Dukes' B or Dukes' C tumours. Left sided colon samples comprised eight Dukes' stage B (median age 76 years), seven Dukes' stage C (median age 68 years), and 10 ''normal mucosa'' samples (median age 69 years). Five of these tumours (201C, 202B, 203B, 204C, and 208C) were accompanied by a corresponding matched normal sample at the same location from the same patient. The remaining five normal mucosa samples (157N, 161N, 179N, 195N, and 205N) and 10 tumour samples (16B, 237B, 239B, 54B, 127B, 58C, 74C, 85C, 91C, and 96C) were obtained from an independent set of samples of individual patients who underwent resection of the sigmoid or rectosigmoid colon.

cRNA preparation, array hybridisation and scanning, and RT-PCR

cRNA preparation, array hybridisation, and scanning are provided in detail as supplementary data, including supplementary tables 1–7 (see the *Gut* website at http://www.gut.com/supplemental).

Data analysis and selection of genes

Data analysis and selection of genes is provided in detail as supplementary data (see the *Gut* website at http://www.gut.com/supplemental). Comparison analysis was done using Microarray Suite 5.0 (MAS 5.0), MicroDB 3.0 (MDB 3.0), and Datamining Tool 3.0 (DMT 3.0) (Affymetrix) applying the Affymetrix specific software ''Statistical Expression Algorithms''. Five different comparison groups (A–E) were established and a schematic overview in given in fig 1 and described in detail in the supplementary data (see the *Gut* website at http://www.gut.com/supplemental).

For all comparisons, several filterings were made to obtain solid and consistent data. To exclude genes with minor or only individual importance, genes were excluded if more than 80% (comparison A) or 70% (B and C) of all datasets were accompanied by a ''detection'' call of ''absent''. Genes were included if more than 80% (B and C) or 70% (D) of the comparisons were accompanied by a ''change'' call of increased or decreased. For statistical analysis, an Affymetrix software integrated Mann-Whitney U test was applied to the signal data of the groups compared with each other. Significance was set at a p value of $p \leqslant 0.05$.

Real time PCR, normalisation of RT-PCR data, and microsatellite analysis

Real time PCR, normalisation of RT-PCR data, and microsatellite analysis are described in detail as supplementary data (see the *Gut* website at http://www.gut.com/supplemental).

Immunohistochemistry

Formalin fixed paraffin embedded sections from the normal mucosa and matched tumour tissue were stained with monoclonal mouse antihuman cyclooxygenase 2 (COX-2) (cat No 35-8200; Zymed, AH-diagnostisk, Denmark), diluted 1:300, or monoclonal mouse antihuman cytokeratin 20 (cat. No M7019; Dako Cytomation, Denmark), diluted 1:100, as described in detail in the supplementary data (see the *Gut* website at http://www.gut.com/supplemental).

RESULTS

Using Affymetrix GeneChip oligonucleotide microarrays, we analysed gene expression of 45 colonic samples. The expression profile of 10 sporadic adenocarcinomas of Dukes' B and C from the right side and 15 from the left side were compared with 20 normal colon mucosa samples, 10 matched samples from the right and 10 partly matched samples from the left side. Gene expression differences were determined between: (A) normal mucosa of the right and left

Table 1 Fifty eight genes differentially expressed more than threefold ($p<0.01$), comparing normal mucosa from the caecum to that of the sigmoid or rectosigmoid

Probe set ID	Gene name	Symbol	UG cluster	Cyto band	Ncae med†	Nsig med‡	FC§	p Value¶
D13897_rna2_at	DNA peptide YY		Hs.169249	17q21.1	54	103	1.9	0.002
D14662_at	Antioxidant protein 2 (non-selenium glutathione peroxidase, acidic calcium independent phospholipase A2)	KIAA0106	Hs.120	1q24.1	471	918	2.0	0.005
D37931_at	Ribonuclease, RNase A family, 4	RNASE4	Hs.283749	14	214	337	1.6	0.001
D42043_at	KIAA0084 protein	KIAA0084	Hs.79123	3p24.3	220	142	−1.5	0.003
D84454_at	Solute carrier family 35 (UDP-galactose transporter), member 2	SLC35A2	Hs.21899	Xp11.23	152	247	1.6	0.001
HG1067-HT1067_r_at	Mucin (Gb:M22406)				196	479	2.4	0.001
HG2348-HT2444_s_at	Peptide Yy		Hs.169249		284	575	2.0	0.007
HG273-HT273_s_at	Fibrinogen A alpha polypeptide alt. splice 3 E*		Hs.351593	4q28	37	136	3.7	0.007
J03600_at	Arachidonate 5-lipoxygenase	ALOX5	Hs.89499	10q11.2	92	51	−1.8	0.006
J04164_at	Interferon induced transmembrane protein 1 (9–27)	IFITM1	Hs.146360	11	1494	963	−1.6	0.007
J04809_rna1_at	Cytosolic adenylate kinase (AK1) gene .		Hs.76240	9q34.1	75	116	1.5	0.003
J05036_s_at	Cathepsin E*	CTSE	Hs.1355	1q31	39	138	3.5	0.002
J05582_s_at	Mucin 1, transmembrane	MUC1	Hs.89603	1q21	586	899	1.5	0.005
K02765_at	Complement component 3	C3	Hs.284394	19p13.3	472	216	−2.2	0.007
L42379_at	Quiescin Q6	QSCN6	Hs.77266	1q24	529	1150	2.2	0.000
L77701_at	COX17 (yeast) homologue, cytochrome c oxidase assembly protein	COX17	Hs.16297	3q13.32	270	429	1.6	0.003
M11433_at	Retinol binding protein 1, cellular	RBP1	Hs.101850	3q23	54	25	−2.1	0.008
M12529_at	Apolipoprotein E	APOE	Hs.169401	19q13.2	922	489	−1.9	0.007
M16364_s_at	Creatine kinase, brain*	CKB	Hs.173724	14q32	614	2300	3.7	0.008
M16938_s_at	Homeo box C6	HOXC6	Hs.820	12q12	55	27	−2.0	0.002
M27281_at	Vascular endothelial growth factor	VEGF	Hs.73793	6p12	42	14	−2.1	0.002
M36341_at	ADP-ribosylation factor 4	ARF4	Hs.75290	3p21.2	693	1086	1.6	0.005
M77144_rna1_at	3-Beta-hydroxysteroid dehydrogenase*		Hs.825	1p13.1	109	18	−5.4	0.005
M80244_at	Solute carrier family 7, member 5	SLC7A5	Hs.184601	16q24.3	129	63	−2.0	0.002
M84424_at	Cathepsin E (CTSE) gene				17	53	2.7	0.005
M86849_at	Gap junction protein, beta 2, 26kD (connexin 26)	GJB2	Hs.5566	13q11	89	56	−1.6	0.007
M97925_rna1_at	Defensin 5*		Hs.72887	8pter-p21	215	35	−6.1	0.002
M98539_at	Prostaglandin D2 synthase gene				569	221	−2.6	0.001
S80562_at	Calponin 3, acidic	CNN3	Hs.194662	1p22	127	79	−1.6	0.001
U03057_at	Actin bundling protein (HSN)	SNL	Hs.118400	7p22	312	200	−1.6	0.002
U24576_at	LIM domain only 4; breast tumour autoantigen complete sequence	LMO4	Hs.3844	1p22.3	97	52	−1.9	0.007
U33632_at	Potassium channel, subfamily K, member 1 (TWIK-1)	KCNK1	Hs.79351	1q42	52	87	1.7	0.002
U50553_at	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3	DDX3	Hs.147916	Xp11.3	63	42	−1.5	0.008
U61262_at	Neogenin (chicken) homolog 1	NEO1	Hs.90408	15q22.3	139	210	1.5	0.003
U66661_at	Gamma-aminobutyric acid (GABA) A receptor, epsilon	GABRE	Hs.22785	Xq28	60	32	−1.9	0.007
U73514_at	Hydroxyacyl-coenzyme A dehydrogenase, type II	HADH2	Hs.171280	Xp11.2	298	462	1.6	0.009
U75679_at	Stem-loop (histone) binding protein	SLBP	Hs.75257	4p16.3	65	43	−1.5	0.007
U81599_at	Homeo box B13*	HOXB13	Hs.66731	17q21.2	33	240	7.2	0.001
U90065_s_at	Potassium channel, subfamily K, member 1 (TWIK-1)	KCNK1	Hs.79351	1q42	160	293	1.8	0.002
U90911_at	Clone 23652 sequence		Hs.171807	–	463	304	−1.5	0.000
X00371_rna1_at	Myoglobin gene (exon 1)		Hs.118836	22q13.1	213	108	−2.0	0.004
X52003_at	pS2 protein; trefoil factor 1*	TFF1	Hs.1406	21q22.3	313	2413	7.7	0.000
X59770_at	Interleukin 1 receptor, type II	IL1R2	Hs.25333	2q12	160	427	2.7	0.001
X61118_rna1_at	LIM domain only 2 (rhombotin-like 1) TTG-2	LMO2	Hs.184585	11p13	63	38	−1.7	0.005
X63187_at	WAP four-disulfide core domain 2*	WFDC2	Hs.2719	20q12	134	714	5.3	0.002
X64072_s_at	Integrin, beta 2 (antigen CD18), (mac-1)	ITGB2	Hs.83968	21q22.3	123	64	−1.9	0.005
X65614_at	*S100 calcium-binding protein P*	S100P	Hs.2962	4p16	170	1293	7.6	0.001
X74570_at	Sialyltransferase 4C (beta-galactosidase alpha-2,3-sialytransferase)*	SIAT4C	Hs.75268	11q23	87	486	5.6	0.001
X75042_at	v-rel avian reticuloendotheliosis viral oncogene homolog	REL	Hs.44313	2p13–p12	64	33	−2.0	0.005
X78924_at	Zinc finger protein 266	ZNF266	Hs.118281	19	43	24	−1.9	0.002
X85545_at	Protein kinase, X-linked	PRKX	Hs.147996	Xp22.3	39	24	−1.6	0.007
X87159_at	Sodium channel, non-voltage-gated 1, beta (Liddle syndrome)*	SCNN1B	Hs.37129	16p12.2	79	393	5.0	0.008
X97324_at	Adipose differentiation-related protein	ADFP	Hs.3416	9p21.3	194	112	−1.7	0.001
Y00503_at	Keratin 19	KRT19	Hs.182265	17q21	1422	2968	2.1	0.001
Y11251_at	Splicing factor, arginine/serine-rich 2, interacting protein	SFRS2IP	Hs.51957	12q12	69	41	−1.7	0.004
Z29090_at	Phosphoinositide-3-kinase, catalytic, alpha polypeptide	PIK3CA	Hs.85701	3q26.3	33	18	−1.7	0.004
Z35278_at	Runt-related transcription factor 3; PEBP2aC1 acute myeloid leukaemia	RUNX3	Hs.170019	1p36	47	30	−1.6	0.008
Z48541_at	Protein tyrosine phosphatase, receptor type, O*	PTPRO	Hs.258609	12p13.3	19	61	3.0	0.008

*Twelve of the 58 genes showed fold changes more than threefold.
†Ncae med, median derived from ''signal'' of 10 normal mucosae of the caecum.
‡Nsig med, median derived from ''signal'' of 10 normal mucosae of the sigmoid or rectosigmoid.
§FC, fold change, corresponding to the ''signal ratio'' of Nsig med/Ncae med, was calculated from the ''signal log ratio''.
¶p value, probability that a variant would assume a value greater than or equal to the observed value strictly by chance.
UG cluster, UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene).

Table 2 Twenty two genes differentially expressed more than fourfold ($p<0.01$), comparing normal mucosa to matched Dukes' B or C adenocarcinomas of the caecum

Probe set ID	Gene name	Symbol	UG cluster	Cyto band	Ncae med*	IQ N†	Bcae med‡	IQ B	Ccae med§	IQ C	Avg FC¶ NvB	Avg FC NvC	p Value
AF001548_rna1_at	Chromosome 16 BAC clone		Hs.78344	16p13.13	1165	1437	99	125	261	118	−8.0	−5.8	0.000
D10667_s_at	Smooth muscle myosin heavy chain				77	153	18	3	18	18	−10.2	−11.2	0.000
J03507_at	Complement protein component C7	C7	Hs.78065	5p13	88	16	10	10	11	8	−6.4	−4.6	0.000
J05096_rna1_at	NaK-ATPase alpha 2 (ATP1A2)		Hs.34114	1q21–q23	15	7	4	3	9	5	−5.5	−1.2	0.002
M14539_at	Coagulation factor XIII, A1 polypeptide	F13A1	Hs.80424	6p25.3	159	60	42	59	52	11	−4.6	−3.1	0.000
M63379_at	TRPM-2 protein gene				918	1022	285	328	255	144	−5.1	−4.0	0.000
M63603_at	Phospholamban	PLN	Hs.85050	6q22.1	29	19	7	6	20	12	−7.1	−1.4	0.003
M77349_at	Transforming growth factor	TGFBI	Hs.118787	5q31	253	118	582	654	1933	199	2.1	5.9	0.003
S67156_at	Aspartoacylase (aminoacylase 2)	ASPA	Hs.32042	17pter	25	10	7	6	2	7	−5.0	−4.4	0.000
U18018_at	Ets variant gene 4	ETV4	Hs.77711	17q21	67	33	340	92	261	79	3.3	4.4	0.000
U20758_rna1_at	Osteopontin		Hs.313	4q21–q25	15	2	80	128	120	298	4.6	12.8	0.000
U37283_at	Microfibril associated glycoprotein-2	MAGP2	Hs.58882	12p13.1	53	28	17	10	7	9	−6.2	−2.5	0.008
U70663_at	Kruppel-like factor 4; hEZF	KLF4	Hs.7934	9q31	587	395	124	127	71	34	−3.0	−10.3	0.001
U71207_at	Eyes absent (Drosophila) homologue 2	EYA2	Hs.29279	20q13.1	44	36	20	4	6	4	1.1	−6.4	0.016
U77180_at	Small inducible cytokine subfamily A	SCYA19	Hs.50002	9p13	104	50	3	7	3	7	−10.4	−13.6	0.002
X00371_rna1_at	Myoglobin gene (exon 1)		Hs.118836	22q13.1	213	120	44	32	65	48	−7.1	−3.4	0.000
X03350_at	Alcohol dehydrogenase 1B (class I)	ADH1B	Hs.4	4q21–q23	85	43	12	5	5	3	−9.0	−17.2	0.000
X05232_at	MMP3/stromelysin 1	MMP3	Hs.83326	11q22.3	12	10	165	96	243	544	15.5	33.6	0.000
X07820_at	MMP10/stromelysin 2	MMP10	Hs.2258	11q22.3	4	4	28	42	20	10	10.1	7.1	0.001
X54162_at	Leiomodin 1 (smooth muscle)	LMOD1	Hs.79386	1	126	84	21	11	36	9	−5.3	−4.5	0.000
X54925_at	Matrix metalloproteinase 1	MMP1	Hs.83169	11q22.3	10	5	248	120	244	1929	16.9	36.7	0.000
X65614_at	S100 calcium binding protein P	S100P	Hs.2962	4p16	170	304	1392	1336	2087	371	7.4	6.1	0.000

*Ncae med, median derived from ''signal'' of 10 normal mucosae of the caecum.
†IQ interquartile, difference between the 75th and 25th percentiles.
‡Bcae med, median derived from ''signal'' of five Dukes' B adenocarcinomas of the caecum.
§Ccae med, median derived from ''signal'' of five Dukes' C adenocarcinomas of the caecum.
¶FC, fold change, corresponding to the ''signal ratio'' of Ncae med/Bcae med or Ncae med/Ccae med, was calculated from ''the signal log ratio''.
UG cluster, UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene).

side; (B) normal mucosa and Dukes' B or C adenocarcinomas of the right side; (C) normal mucosa and Dukes' B or C adenocarcinomas of the left side; (D) Dukes' B or C adenocarcinomas of the right and left side; and finally (E), differentially expressed genes in the right sided colon from comparison B with those in the left sided colon from comparison C (fig 1).

Comparison A: normal caecum versus sigmoid/ rectosigmoid

By comparing normal mucosa samples from 20 different patients—namely, 10 right sided from the caecum to 10 left sided from the sigmoid or rectosigmoid—we identified 160 genes showing site specific differential gene expression, being increased or decreased more than 1.5 fold ($p<0.05$, Mann-Whitney U test). Fifty eight genes with a p value of <0.01 are shown in table 1; 12 of these genes with fold changes more than threefold the median signal are labelled with an asterisk.

The gene encoding the pS2 protein, maintaining the mucosal surface barrier and stimulating repair processes, showed 7.7-fold higher expression in the left than in the right colon. Other differentially expressed genes with a consistent difference were calcium binding protein S100P (7.6-fold), homeodomain protein HOXB13 (7.2-fold), defensin 5 (6.1-fold), Gal-beta (1-3/1-4) GlcNAc alpha-2.3-sialyltransferase (5.6-fold), 3-beta-hydroxysteroid dehydrogenase gene (5.4-fold), and HE4 extracellular proteinase inhibitor homologue (5.3-fold). Also, the beta subunit of creatine kinase-B, fibrinogen A alpha polypeptide alt. splice 3 E (3.7-fold), cathepsin E, and protein tyrosine phosphatase were among the genes showing more than threefold significantly different expression between the two groups.

Comparison B: normal versus tumour caecum

By comparing normal mucosa of the caecum to matching caecum adenocarcinomas staged as Dukes' B or C and derived from the same patient, we identified 118 genes significantly up or downregulated more than 2.8-fold ($p<0.05$, Mann-Whitney U test) in adenocarcinomas compared with normal mucosa (see supplementary table on the *Gut* website at http://www.gut.com/supplemental). Seventy three showed fold changes of more than fourfold, and of these, 22 genes with a p value of <0.01 are shown in table 2.

A characteristic finding was that most genes ($n = 15$) were downregulated in carcinomas compared with normal mucosa, and only a few were upregulated ($n = 7$). Several matrix metalloproteinases, such as MMP1, MMP3, and MMP10, located in the extracellular space and involved in proteolysis and peptidolysis were highly upregulated in carcinomas, as well as E1A enhancer binding protein (E1A-F) (fourfold) and calcium binding protein S100P. TRPM-2 protein (fivefold), complement protein component C7 (five-fold), and NAD+ dependent 15 hydroxyprostaglandin dehydrogenase (PGDH; 16-fold) showed decreased expression.

Comparison C: normal versus tumour sigmoid/ rectosigmoid

We compared normal mucosa from the left side of the colon to matching adenocarcinomas of Dukes' B and C from the

Table 3 Forty two genes differentially expressed more than fourfold ($p<0.01$), comparing normal mucosa to Dukes' B or C tumours from the sigmoid or rectosigmoid

Probe set ID	Gene name	Symbol	UG cluster	Cyto band	Nsig med*	IQ N†	Bsig med‡	IQ B	Csig med§	IQ C	Avg FC¶ NvB	Avg FC NvC
D84239_at	Fc fragment of IgG binding protein	FCGBP	Hs.111732	19q13.1	2819	726	233	525	171	408	−21.0	−23.4
HG2981-HT3125	Epican Alt. Splice 1				13	7	58	53	75	43	4.3	5.6
J03910_rna1_at	Metallothionein-IG gene	MT1G	Hs.173451	16q13	2252	1934	409	581	148	105	−6.9	−12.3
J03915_s_at	Chromogranin A	CHGA	Hs.172216	14q32	448	153	62	62	41	20	−5.0	−6.1
J04040_at	Glucagon	GCG	Hs.1460	2q36	314	514	27	35	11	6	−8.7	−15.5
J04093_s_at	UDP glycosyltransferase 1 family	UGT1A6	Hs.284239	2q37	204	78	41	34	35	18	−4.0	−5.2
J04152_rna1	M1S1 gene		Hs.23582	1p32–p31	6	11	66	30	144	509	6.9	19.0
J05257_at	MDP4 MDP7 microsomal dipeptidase	DPEP1	Hs.109	16q24.3	26	10	847	790	388	366	20.1	7.9
L10373_at	(Clone CCG-B7) sequence		Hs.82749	Xq11	310	91	74	77	61	27	−4.6	−5.6
L10955_cds1	Carbonic anhydrase IV gene	CAIV			1186	365	194	195	30	19	−7.2	−32.7
L11708_at	Hydroxysteroid (17-beta) dehydrogenase 2	HSD17B2	Hs.155109	16q24.1	414	71	41	65	22	10	−5.9	−6.4
L12760_s_at	Phosphoenolpyruvate carboxykinase 1	PCK1	Hs.1872	20q13.31	755	566	106	145	40	31	−4.2	−11.0
L21998_at	Mucin 2, intestinal/tracheal	MUC2	Hs.315	11p15.5	4189	1013	1175	2355	243	206	−4.1	−14.8
L22524_s_at	Matrilysin gene				5	5	37	17	570	683	6.1	83.2
L76465_at	Hydroxyprostaglandin dehydrogenase 15	HPGD	Hs.77348	4q34–q35	213	48	61	45	21	22	−4.5	−7.7
M12963_s_at	Alcohol dehydrogenase 1A (class I), alpha	ADH1A	Hs.73843	4q21–q23	1724	645	260	649	59	69	−7.1	−38.8
M14758_at	ATP binding cassette, sub-family B	ABCB1	Hs.21330	7q21.1	162	66	37	60	33	28	−4.1	−6.1
M16364_s_at	Creatine kinase-B	CKB	Hs.173724	14q32	2300	720	501	693	314	275	−4.8	−13.2
M16801_at	Nuclear receptor subfamily 3, group C	NR3C2	Hs.1790	4q31.1	161	34	32	26	13	2	−4.0	−10.3
M18079_at	Fatty acid binding protein 2, intestinal	FABP2	Hs.282265	4q28–q31	184	126	31	23	20	18	−4.9	−10.2
M60047_at	Heparin binding growth factor binding protein	HBP17	Hs.1690	4p16–p15	143	99	33	32	28	12	−4.4	−4.8
M77349_at	Transforming growth factor, beta induced	TGFBI	Hs.118787	5q31	307	145	1735	787	2618	1643	5.3	8.0
M87860_at	S-lac lectin L-14-II (LGALS2) gene				232	129	40	39	12	18	−5.3	−7.4
M97496_at	Guanylate cyclase activator 2A (guanylin)	GUCA2A	Hs.778	1p35–p34	1931	279	147	150	62	47	−15.8	−44.5
U14528_at	DTD sulfate transporter	SLC26A2	Hs.29981	5q31–q34	620	410	137	182	23	12	−4.3	−20.1
U17077_at	BENE protein	BENE	Hs.185055	2q13	1499	544	214	108	207	212	−4.3	−6.2
U70663_at	Kruppel-like factor 4, hEZF	KLF4	Hs.7934	9q31	605	362	76	98	44	56	−6.2	−8.8
X52001_at	Endothelin 3	EDN3	Hs.1408	20q13.2	100	44	34	32	6	7	−4.6	−16.1
X53800_s_at	GRO3 oncogene	GRO3	Hs.89690	4q21	21	13	73	49	75	95	5.3	5.0
X54489_rna1_at	(MGSA)	MGSA	Hs.789	4q21	56	19	262	235	484	601	7.8	8.4
X54925_at	(Interstitial collagenase)	MMP1	Hs.83169	11q22.3	12	7	129	73	945	1041	7.3	31.8
X57579_s_at	Inhibin, beta A (activin A)	INHBA	Hs.727	7p15–p13	7	9	60	68	433	332	5.1	26.7
X59766_at	Alpha-2-glycoprotein 1, zinc	AZGP1	Hs.71	7q22.1	15	19	333	125	114	83	18.8	4.4
X59770_at	Interleukin 1 receptor, type II	IL1R2	Hs.25333	2q12–q22	427	185	48	25	65	42	−6.4	−4.5
X63597_at	Sucrase-isomaltase	SI	Hs.2996	3q25.2	37	36	12	13	3	4	−8.0	−14.1
X63629_at	Cadherin 3, type 1, P-cadherin (placental)	CDH3	Hs.2877	16q22	14	8	214	76	134	146	11.2	6.6
X73501_at	Cytokeratin 20	KRT20	Hs.84905	17q21.1	1553	635	185	130	107	110	−5.5	−12.0
X87159_at	Sodium channel, nonvoltage-gated 1	SCNN1B	Hs.37129	16p12.2	393	180	29	51	10	5	−8.3	−22.1
X98311_at	Carcinoembryonic antigen-related	CEACAM7	Hs.74466	19q13.2	3817	1389	462	475	145	114	−4.1	−16.3
Y00339_s_at	Carbonic anhydrase II (EC 4.2.1.1)	CA2	Hs.155097	8q22	1096	465	58	112	28	12	−20.8	−27.4
Y00787_s_at	Interleukin 8/MDNCF	IL8	Hs.624	4q13–q21	53	260	614	326	1794	2320	5.4	15.5
Z70295_at	GCAP-II (uroguanylin)	GUCA2B	Hs.32966	1p34–p33	453	64	9	5	7	3	−25.0	−39.0

*Nsig med, median derived from ''signal'' of 10 normal mucosae of the sigmoid and rectosigmoid.
†IQ interquartile, difference between the 75th and 25th percentiles.
‡Bsig med, median derived from ''signal'' of eight Dukes' B adenocarcinomas of the sigmoid and rectosigmoid.
§Csig med, median derived from ''signal'' of seven Dukes' C adenocarcinomas of the sigmoid and rectosigmoid.
¶FC, fold change, corresponding to the ''signal ratio'' of Nsig med/Bsig med or Nsig med/Csig med, was calculated from the ''signal log ratio''.
UG cluster, UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene).

same patient in five cases, and in those 10 cases where a matching normal sample was not present, we compared each of the 10 tumours to each of five single normal samples (for details see material and methods). We identified 186 genes significantly differentially expressed more than 2.8 fold ($p<0.05$, Mann-Whitney U test) from the normal mucosa to Dukes' B or Dukes' C tumours (see supplementary table 3 on the *Gut* website at http://www.gut.com/supplemental). The majority confirmed our recently published findings made on pools of colorectal cancer samples[11]; for example, down-regulation of nuclear encoded mitochondrial genes such as TST thiosulfate sulfurtransferase (rhodanese) (4.5-fold) and the SCAD gene 5′ UTR exon 1 and 2 (sevenfold). The 42 most important genes with a fold change ≥ 4 ($p<0.01$) in both Dukes' B and Dukes' C are shown in table 3. Other genes (for example, osteopontin) which showed changes have been omitted here because changes were found to be ≥ 4 fold in either Dukes' C or Dukes' B but not in both.

Thirty genes were found to be downregulated in cancer, such as GCAP-II (33-fold), carbonic anhydrase IV (33-fold), and DTD sulfate transporter gene (20-fold). Only 12 genes were upregulated, among these microsomal dipeptidase (MDP4, MDP7; 20-fold) and interleukin 8/MDNCF (15-fold). As a novel finding we found that carbonic anhydrase VII

Table 4 Sixteen genes differentially expressed more than threefold ($p<0.05$), comparing Dukes' B and C adenocarcinomas of the caecum with those of the sigmoid or rectosigmoid

					Dukes' B			Dukes' C		
Probe set ID	Gene name	Symbol	UG cluster	Cyto band	Bcae med*	Bsig med†	Avg FC Bcae *v* Bsig‡	Ccae med§	Csig med¶	Avg FC Ccae *v* Csig
D00654_at	Enteric smooth muscle gamma-actin gene	ACTG2	Hs.78045	2p13	88	164	−2.7	87	465	−5.6
D13643_at	24-dehydrocholesterol reductase	DHCR24	Hs.75616	1p33–p31.1	453	377	1.4	503	197	3.2
D17408_s_at	Calponin 1, basic, smooth muscle	CNN1	Hs.21223	19p13.2–p13.1	47	105	−2.7	79	281	−4.5
D90279_s_at	Collagen, type V, alpha 1	COL5A1	Hs.146428	9q34.2–q34.3	3	7	−1.6	21	158	−4.1
HG2743-HT2846_s_at	Caldesmon 1 Alt. Splice 6 Non-Muscle (M64110)	CALD1	Hs.325474	7q33	42	88	−2.2	57	306	−4.4
HG2743-HT3926_s_at	Gamma-glutamyltransferase 1 (J04131)	GGT1	Hs.284380	22q11.1–q11.2	10	29	−2.9	30	96	−4.4
M26679_at	Homeo box A5	HOXA5	Hs.37034	7p15–p14	30	17	1.8	59	15	3.4
M58459_at	Ribosomal protein S4, Y-linked	RPS4Y	Hs.180911	Yp11.3	6	157	−8.8	9	560	−23.3
M83216_s_at	Caldesmon 1	CALD1	Hs.286238	7q33	27	106	−2.8	84	466	−4.2
M84526_at	D component of complement (adipsin)	DF	Hs.155597	19p13.3	249	74	2.0	39	112	−5.4
M95787_at	Transgelin 11 (SM22-alpha)	TAGLN	Hs.75777	11q23.2	360	663	−2.0	542	3600	−5.3
U28368_at	Inhibitor of DNA binding 4	ID4	Hs.34853	6p22–p21	4	16	−2.0	4	38	−5.4
U35139_at	Necdin (mouse) homolog	NDN	Hs.50130	15q11.2–q12	23	33	−1.6	9	58	−5.7
U48959_at	Myosin, light polypeptide kinase	MYLK	Hs.211582	3q21	83	168	−2.3	146	618	−5.8
U52191_s_at	SMC (mouse) homolog, Y chromosome	SMCY	Hs.80358	Yq11	2	15	−4.6	2	48	−12.2
X51405_at	Carboxypeptidase E (EC 3.4.17.10)	CPE	Hs.75360	4q32.3	10	22	−2.5	13	46	−4.1

*Bcae med, median derived from ''signal'' of five Dukes' B adenocarcinomas of the caecum.
†Bsig med, median derived from ''signal'' of eight Dukes' B adenocarcinomas of the sigmoid and rectosigmoid.
‡FC, fold change, corresponding to the ''signal ratio'', was calculated from the ''signal log ratio''.
§Ccae med, median derived from ''signal'' of five Dukes' C adenocarcinomas of the caecum.
¶Csig med, median derived from ''signal'' of seven Dukes' C adenocarcinomas of the sigmoid and rectosigmoid.
UG cluster, UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene).

(CA VII) was decreased more than fourfold from normal to Dukes' B and C adenocarcinomas.

Comparison D: tumours from the caecum versus sigmoid/rectosigmoid

Within each of the Dukes' B and C stages, we compared all adenocarcinomas from the left side with all of those from the right side of the colon. We identified five genes in Dukes' B, 39 in Dukes' C, and five genes in both B and C, that showed significant differences in expression levels ($p<0.05$) with an average fold change of 2.8, corresponding to a total of 44 genes differentially expressed in left and right sided tumours (see supplementary table 4 on the *Gut* website at http://www.gut.com/supplemental). Among these 44 genes, 16 showed more than threefold upregulation or more than fourfold downregulation (table 4).

Differential gene expression was more common in Dukes' C than in Dukes' B, and among the genes were caldesmon 1, involved in cellular mitosis and receptor capping, modulator recognition factor 2 (a DNA binding factor), ARHB, involved in signal transduction, transgelin 11 (SM22-alpha), and D component of complement (adipsin, involved in proteolysis and peptidolysis), all five showing higher expression in left sided carcinomas. In contrast, homeobox A5 protein, a sequence specific transcription factor, was more strongly expressed in Dukes' C adenocarcinomas of the right side of the colon.

Comparison E: comparison to identify genes in common or differentially expressed in right sided versus left sided tumours

A total of 186 genes previously identified to be differentially expressed from normal mucosa to tumour in the left side of the colon were compared with 118 genes identified in the right colon. This resulted in 30 common cancer genes being significantly differentially expressed more than threefold (accompanied by a p value of <0.05) in at least one of the Dukes' in both right sided as well as left sided tumours. These may make ideal colonic tumour markers (table 5).

Validation of the results by real time PCR applied to aminopeptidase N/CD13, SCAD, and PCK1 is shown in fig 2 where single GeneChip analyses were compared with real time PCR analyses. Additionally, we identified cancer genes being characteristic for one side of the colon only. Eighty eight genes shown in supplementary table 5 (on the *Gut* website at http://www.gut.com/supplemental) were significantly differentially expressed exclusively in right sided tumours, such as factor XIII subunit a and calcium binding protein S100P (fig 2), suggesting a more crucial role in caecal adenocarcinomas. A total of 156 genes shown in supplementary table 6 (on the *Gut* website at http://www.gut.com/supplemental) were significantly differentially expressed only in left sided tumours. Among these were MDP4/MDP7 and the interferon inducible protein ''9-27''. Differences in expression in most of the growth factors were seen in the left colon such as upregulation of teratocarcinoma derived growth factor (>7 fold). Furthermore, the COX-2 gene was more than sixfold higher in Dukes' C tumours of the left colon, and did not show a significant difference in right sided tumours. Most strikingly, expression of keratins 8, 19, and 20 was severely reduced in the left colon but did not show significant differences in the caecum.

Microsatellite analysis

Microsatellite analysis was performed on microdissected tumour tissue, as described in materials and methods in the supplementary data (on the *Gut* website at http://www.gut.com/supplemental). Of 10 samples, where the amount of tissue allowed microdissection, only one sample (No 120B) was found to be highly microsatellite instable, the

Table 5 Thirty common cancer genes differentially expressed more than threefold accompanied by a p value of <0.05 in at least one of the Dukes' compared with normal mucosa

				Left colon: sigmoid and rectosigmoid							Right colon: caecum						
Probe set ID	Gene name	UG cluster	Cyto band	Nsig med*	Bsig med†	Csig med‡	Avg FC NvB§	p Value NvB¶	Avg FC NvC	p Value NvC††	Ncae med**	Bcae med††	Ccae med‡‡	Avg FC NvB§§	p Value NvB	Avg FC NvC	p Value NvC
D87292_at	TST thiosulfate sulfurtransferase (rhodanese)	Hs.248267	22q13.1	2492	1166	489	−2.1	0.002	−4.5	0.001	2649	1089	749	−1.0	0.766	−3.6	0.008
HG2614-HT2710_at	Collagen type viii alpha 1			35	96	244	1.7	0.006	5.6	0.001	19	38	69	1.3	0.766	5.2	0.032
HG2755-HT2862_at	T-Plastin			58	163	321	2.0	0.010	3.8	0.001	47	135	284	1.4	0.365	3.7	0.016
HG2981-HT3125_s_at	Epican alt. splice 1			13	58	75	4.3	0.001	5.6	0.001	16	78	75	1.7	0.484	4.1	0.032
HG4312-HT4582_s_at	Transcription factor IIia			268	858	462	3.0	0.000	1.8	0.001	245	556	781	1.9	0.075	3.1	0.032
J03910_rna1_at	Metallothionein-IG (MT1G)	Hs.173451	16q13	2252	409	148	−6.9	0.010	−12.3	0.001	1417	219	279	−1.6	0.766	−9.7	0.032
J04970_at	Carboxypeptidase M	Hs.169765	12q15	60	21	21	−3.9	0.000	−2.8	0.001	55	21	19	−1.8	0.266	−3.3	0.032
L12350_at	Thrombospondin 2 (THBS2)	Hs.108623	6q27	22	85	448	2.0	0.004	13.2	0.001	24	53	136	−1.6	0.921	6.1	0.032
L12760_s_at	Phosphoenolpyruvate carboxykinase (PCK1)	Hs.1872	20q13.31	755	106	40	−4.2	0.001	−11.0	0.001	510	35	12	−3.2	0.187	−22.4	0.032
L22524_s_at	Metalloproteinase; MMP-7			5	37	570	6.1	0.001	83.2	0.001	7	62	179	6.0	0.044	23.0	0.032
L76465_at	NAD+ dependent 15 hydroxyprostaglandin dehydrogenase (PGDH)	Hs.77348	4q34–q35	213	61	21	−4.5	0.001	−7.7	0.001	211	41	11	−4.5	0.012	−16.9	0.095
M10942_at	Metallothionein-Ie gene (hMT-Ie)	Hs.74170	16q13	989	134	302	−5.2	0.002	−3.7	0.001	551	228	274	−1.7	0.266	−3.6	0.016
M12759_at	Ig J chain gene			687	117	95	−3.3	0.013	−7.9	0.001	1189	227	60	1.1	0.484	−22.1	0.032
M12963_s_at	Class I alcohol dehydrogenase (ADH1) alpha subunit	Hs.73843	4q21–q23	1724	260	59	−7.1	0.008	−38.8	0.001	1647	282	31	−3.8	0.044	−46.7	0.095
M22324_at	Aminopeptidase N/CD13	Hs.1239	15q25–q26	565	117	198	−3.6	0.002	−2.9	0.001	2169	357	128	−2.8	0.266	−9.0	0.016
M26576_cds2_at	Alpha-1 collagen type IV			105	332	764	2.0	0.010	5.0	0.001	107	233	429	1.4	0.187	−0.3	0.016
M77349_at	Transforming growth factor-beta induced gene product (BIGH3)	Hs.118787	5q31	307	1735	2618	5.3	0.000	8.0	0.001	253	582	1933	2.1	0.187	−0.2	0.008
M87860_at	S-lac lectin L-14-II (LGALS2)			232	40	12	−5.3	0.001	−7.4	0.001	57	35	7	−1.0	0.766	−7.8	0.016
S78187_at	CDC25Hu2 = cdc25+ homologue	Hs.153752	20p13	129	612	304	4.2	0.000	1.6	0.006	218	814	640	3.3	0.024	3.4	0.032
U05861_at	Hepatic dihydrodiol dehydrogenase			173	54	22	−3.8	0.041	−6.8	0.008	99	43	32	−3.4	0.044	−2.4	0.548
U14528_at	Sulfate transporter (DTD)	Hs.29981	5q31–q34	620	137	23	−4.3	0.001	−20.1	0.001	679	85	18	−1.8	0.619	−28.1	0.032
U18018_at	E1A enhancer binding protein (E1A-F)	Hs.77711	17q21	54	350	226	9.4	0.000	3.7	0.025	67	340	261	3.3	0.012	4.4	0.008
U20758_rna1_at	Osteopontin gene	Hs.313	4q21–q25	11	46	463	2.2	0.010	15.1	0.001	15	80	120	4.6	0.004	12.8	0.056
U70663_at	Zinc finger transcription factor hEZF (EZF)	Hs.7934	9q31	605	76	44	−6.2	0.001	−8.8	0.001	587	124	71	−3.0	0.123	−10.3	0.008
U77643_at	K12 protein precursor	Hs.95655	17q25	422	160	148	−3.0	0.000	−3.4	0.001	536	146	165	−1.7	0.187	−3.1	0.016
U78551_at	Gall bladder mucin MUC5B	Hs.102482	11p15	465	75	26	−3.8	0.010	−5.5	0.003	154	421	111	4.7	0.044	−3.4	0.548
X54925_at	MMP1 matrix metalloproteinase 1; type I interstitial collagenase	Hs.83169	11q22.3	12	129	945	7.3	0.001	31.8	0.005	10	248	244	16.9	0.004	36.7	0.056
X59770_at	IL-1R2 type II interleukin-1 receptor	Hs.25333	2q12–q22	427	48	65	−6.4	0.000	−4.5	0.001	160	54	66	−1.7	0.365	−2.9	0.016
X63629_at	p cadherin	Hs.2877	16q22	14	214	134	11.2	0.000	6.6	0.001	15	218	142	5.1	0.123	8.4	0.016
Z80345_rna1_s_at	SCAD	Hs.127610	12q22–qter	285	121	41	−3.0	0.001	−7.5	0.001	399	181	55	−1.3	0.484	−7.0	0.032

*Nsig med, median derived from "signal" of 10 normal mucosae of the sigmoid and rectosigmoid.
†Bsig med, median derived from "signal" of eight Dukes' B adenocarcinomas of the sigmoid and rectosigmoid.
‡Csig med, median derived from "signal" of seven Dukes' C adenocarcinomas of the sigmoid and rectosigmoid.
§Avg FC NvB, fold change, corresponding to the "signal ratio" of Nsig med/Bsig med or Nsig med/Csig med, was calculated from the "signal log ratio".
¶p value NvB, probability that a variant would assume a value greater than or equal to the observed value strictly by chance.
**Ncae med, median derived from "signal" of 10 normal mucosae of the caecum.
††Bcae med, median derived from "signal" of five Dukes' B adenocarcinomas of the caecum.
‡‡Ccae med, median derived from "signal" of five Dukes' C adenocarcinomas of the caecum.
§§Avg FC NvB, fold change, corresponding to the "signal ratio" of Ncae med/Bcae med or Ncae med/Ccae med, was calculated from the "signal log ratio".
UG cluster, UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene).

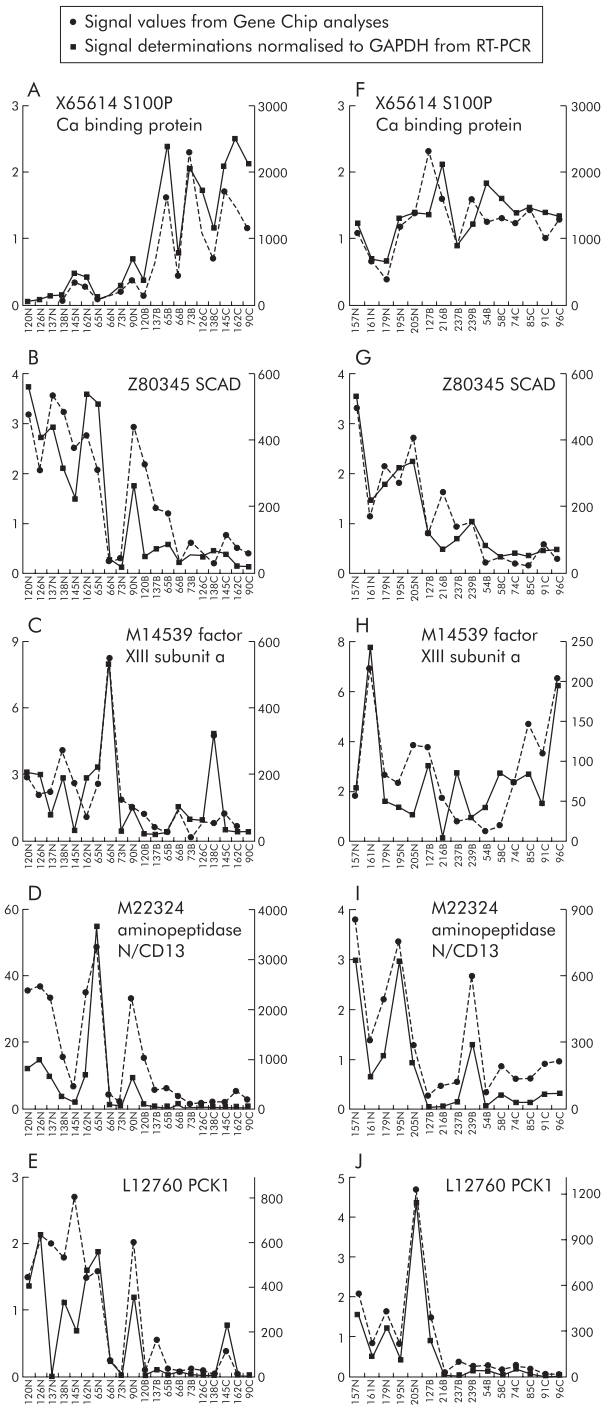

Figure 2 Comparison of single GeneChip analyses with real time polymerase chain reaction (PCR) analyses. Expression analyses of five selected genes using single samples of normal colon mucosa and adenocarcinomas of Dukes' stages B and C from the right (A–E) and left (F–J) sides of the colon. Left y axis shows expression intensities ''normalised to GAPDH'' obtained from reverse transcription (RT)-PCR and the right y axis shows expression intensities ''signal'' derived from GeneChip analysis. (A, F) X65614 S100P Ca-binding protein; (B, G) Z80345 SCAD; (C, H) M14539 factor XIII subunit a; (D, I) M22324 aminopeptidase N/CD13; (E, J) L12760 PCK1 (phosphoenolpyruvate carboxykinase).

other nine samples being microsatellite stable (MSS), as listed in supplementary table 1 (on the *Gut* website at http://www.gut.com/supplemental). The fact that all except one of the tumours were stable with regard to microsatellites BAT25 and BAT26 (MSS) strongly supports the conclusion that the differences described here do not result from differences in microsatellite stability but have to be regarded as differences characterising the function and behaviour of tumours originating from the caecum or sigmoid and rectosigmoid.

Immunohistochemical analysis

Immunostaining was applied to paraffin embedded specimen from eight of the 10 right sided and 11 of the 15 left sided tumours where snap frozen material had been previously analysed on microarrays to enable a comparison of RNA and protein expression. The 19 tumours where selected based on the availability of their matching normal mucosa from the oral resection edge.

Figure 3 (A, B) shows five right and five left sided tumours with their matching normal mucosa stained with COX-2. In the right colon, COX-2 was moderately to strongly expressed in normal mucosa, mostly throughout the entire epithelium as well as in right sided tumours. Comparing normal tissue with tumour, we detected upregulation (in one of eight tissue sections), downregulation (1/8), or about equal expression in normal tissue and tumour (6/8). In the left side of the colon, COX-2 was not or only very weakly expressed in normal mucosa and was upregulated from normal mucosa to tumour. Comparing normal mucosa to tumour, we observed strong upregulation in more than 50% of cells (3/11), moderate upregulation in more than 50% of cells (5/11), and very strong upregulation in single cell groups corresponding to less than 10% of cells (3/11).

Figure 3 (C, D) shows five right and five left sided tumours with their matching normal mucosa stained with cytokeratin 20 (KRT20). KRT20 was strongly expressed in the luminal epithelium of normal mucosa of both sides. Comparing normal tissue to tumour of the right side, we detected strong upregulation with staining of more than 50% of cells (2/8), downregulation (4/8) with staining of less than 10% of cells, or about equal expression in normal and tumour with staining of approximately 30–40% of tumour cells (2/8). Comparing normal mucosa to tumour on the left side, we observed upregulation with staining of more than 50% of cells (2/11), downregulation with staining of less than 10% of cells (7/11), or about equal expression in normal mucosa and tumour with staining of approximately 30–40% of tumour cells (2/11). Staining of tumour cells was very heterogeneous in most of the tumours.

DISCUSSION

While published data on right sided versus left sided colon cancers are lacking, colon cancers per se have previously been compared with normal mucosa. In this study, we identified differences in gene expression in the colon that characterised left and right sided normal mucosa and adenocarcinomas. Using statistical algorithms provided by the Affymetrix software, we identified sets of genes differentially expressed, as well as genes in common, between right sided and left sided adenocarcinomas.

In this study, we analysed a total of 45 samples (20 normal and 25 tumour samples). The complexity of our study is comparable with colon cancer expression analyses previously described by Alon *et al*, analysing 22 normal and 40 tumour samples, and by Notterman *et al*, analysing 18 adenocarcinomas and four adenomas with paired normal tissue, both using Affymetrix GeneChips, as previously discussed.[12 13] The reliability of our data (for example, with regard to comparisons of left sided normal mucosa to Dukes' B and C tumours) is supported by the fact that we confirmed identification of various genes previously identified by other techniques. Metallothionein, fibronectin, and SPARC, for example, had previously been shown to be differentially

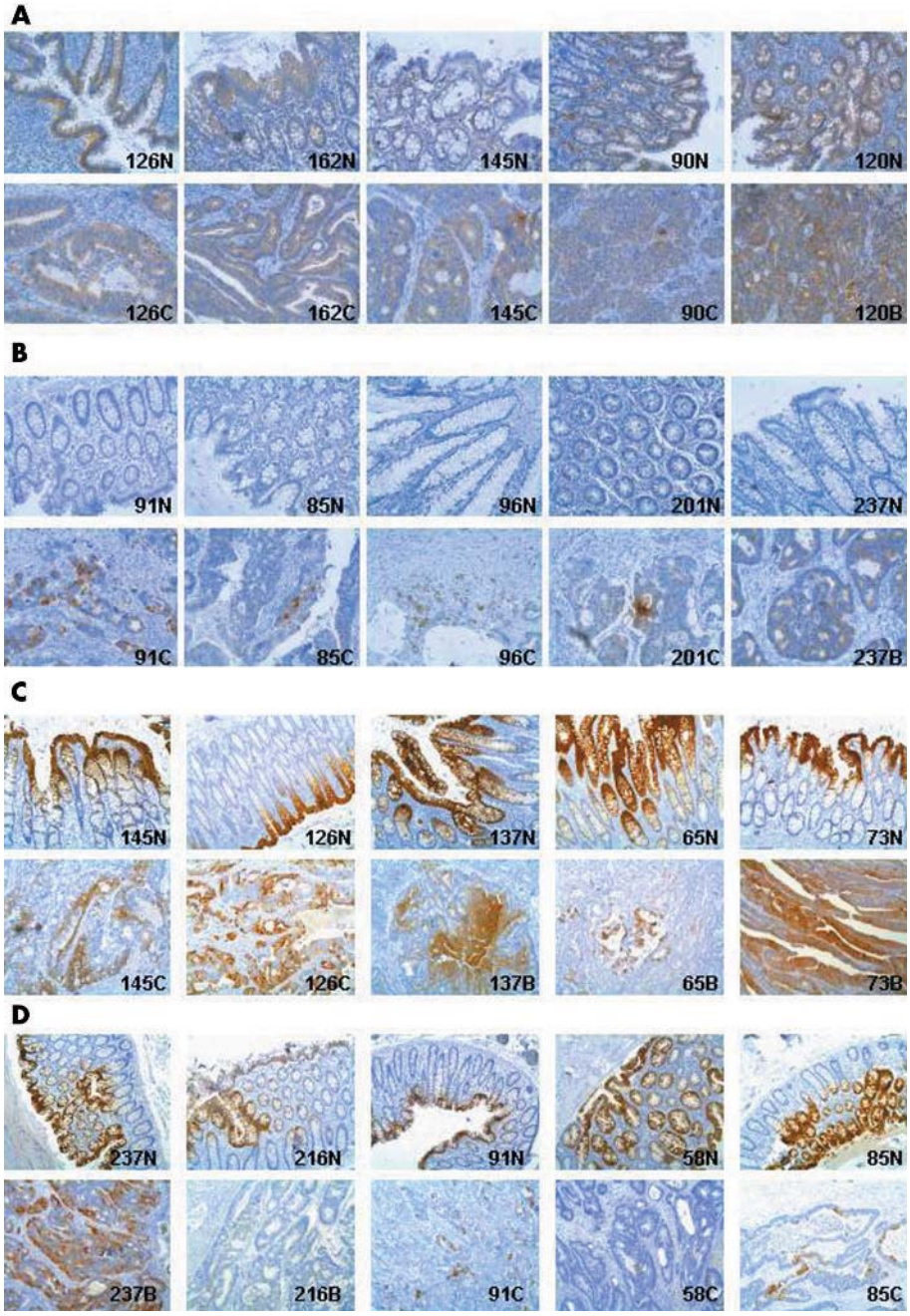

Figure 3 Immunohistochemistry of formalin fixed paraffin embedded sections of Dukes' B and C adenocarcinomas and their matching normal mucosa (N). Sample numbers of tumours refer to samples previously analysed on microarrays. Cyclooxygenase 2 (COX-2) was moderately expressed in right sided normal mucosa as well as in matching tumours (A). COX-2 was not or very weakly expressed in left sided normal mucosa but moderately to strongly expressed in matching tumours showing a very heterogeneous staining pattern (B) (magnification 20×). Cytokeratin 20 (KRT20) was highly expressed in the luminal epithelium of normal mucosa of both sides. KRT20 was downregulated in only 50% of right sided tumours (C) whereas it was strongly downregulated in 80% of left sided tumours (D) (magnification 10×).

expressed in normal tissue and tumour by Zhang *et al*, using the SAGE technique on two normal and two tumour samples.[14] Furthermore, we confirmed differential expression of more than 70% of genes previously identified by GeneChip analyses of pooled samples of the left side of the colon.[11] In addition, these results were also highly comparable with data previously published by Notterman *et al* (for example, upregulation of MGSA from normal to tumour tissue and downregulation of guanylin or chromogranin A).[13] Reliability of the results with regard to differences between the right and left colon was further supported by expression analysis using RT-PCR showing high reproducibility of expression levels detected by the arrays. Previous studies have, in most cases, not taken into account the Dukes' stage or location within the colon where the samples originated. Obviously, adding more subclasses to the material inevitably leads to fewer samples per class and the main findings of this paper should be repeated on larger material.

A grouped Mann-Whitney test intrinsic to the Affymetrix software DMT 3.0 was used for statistical analyses. As some of the data were paired (different tissues from the same patient) a Wilcoxon matched pairs test may have been more

appropriate for these cases and this may have been a limitation of our statistical analyses. On the other hand, some of the samples were grouped (tissue from different patients) and a Mann-Whitney test had to be applied. However, such ''breaking of matching'' is more likely to make the results more, rather than less, comparable between tumour and normal tissue, and so this limitation is not of major importance as it is not likely to explain any of the observed differences.

We focused the analysis on adenocarcinomas of Dukes' stages B and C as these are the most challenging stages in colon cancer, with the possibility of curative treatment. Most of the factors that may influence gene expression were taken into account but for array analysis it was not possible to match all samples with their normal mucosa (as could be achieved for immunostainings) or to match samples with regard to sex, as most of our right sided colon cancer patients were female. In general, colon cancer affects males and females equally but some studies indicate that right sided colon cancer affects more women than men.[2] Comparison of right versus left sided normal mucosa showed 58 genes differentially expressed, 12 with fold changes more than threefold and none located on the Y chromosome. A comparison of right and left sided adenocarcinomas showed two genes located on the Y chromosome and significantly higher expressed in the group of left sided Dukes' C but not Dukes' B. RPS4Y and SMCY show high fold changes of 23- and 12-fold but SMCY increases only up to a signal of 50, which is close to the detection level. From these data there is no evidence that the imbalance between males and females influences the results profoundly.

Genes such as β-catenin, c-erbB2, EGFR, PCNA, or DPP IV, previously shown to be differentially expressed in right and left colon cancers,[9] were not identified as significant in this study but did show side differences when less stringent selection criteria were applied. There are many factors affecting gene expression analysis, such as ischaemic delay, defined as the period of time from clamping of blood vessels to snap freezing, ratio of tumour versus non-tumour cells, RNA extraction method and quality of RNA (28s/18s ratio), type of array used (c-DNA arrays, nylon membrane, oligonucleotide arrays), amplification, labelling (Cy3/Cy5 or d-UTP-Biotin/SAPE) and labelling efficiency, sensitivity and detection threshold, software used for analysis, and statistical significance criteria.

The most predominant differences between normal left and right colon mucosa were higher expression in left sided mucosa of genes such as pS2 protein, calcium binding protein S100P, HOXB13, SIAT4C, and WFDC2. This agrees with previous findings of a 7.7-fold higher expression of pS2 protein[15] and approximately fourfold higher expression of HOXB13 and S100P[10] in the left colon. This agreement is remarkable because different platforms have been used for analysis and two thirds of the samples in the study of Glebov *et al* were HNPCC samples. Homeobox proteins such as HOXB13 or HOXA5 encode transcription factors and upregulate tumour suppressor p53 and may therefore be involved in side specific tumorigenesis.

Defensin 5 was found to be expressed sixfold higher in right sided mucosa which matches the proposal that the right colon provides more efficient local tumour defence, maintaining the mucosal barrier.[5 16] We hypothesise that the right sided colon mucosa provides protection against carcinogens by defensin 5 expression, leading to less frequent carcinogenesis compared with the left side. Remarkably, the site on chromosome 8p housing the defensin gene is frequently lost in liver metastases from primary colon cancers.[15]

The majority of genes found to be differentially expressed from normal mucosa to Dukes' B or C of the left side confirmed our recently published findings performed on pooled samples.[11] Genes such as MDP4/MDP7 and interleukin 8/MDNCF were strongly upregulated, and several nuclear encoded mitochondrial proteins such as rhodanese or SCAD were strongly downregulated in tumours. We also identified reduced levels of several carbonic anhydrases (CA) such as CAVII or CAIV which have not previously been described indepth in colon cancer. CAIV, downregulated by up to 33-fold in left sided tumours, is responsible for maintenance of pH and ion equilibrium. Takenawa *et al* showed that low level expression of CAIV and aquaporin 1 in renal cell carcinomas was associated with poor survival.[17]

Notterman *et al* analysed differential gene expression between the normal colon and tumour, without discriminating between the right and left side.[13] In terms of expression differences between normal mucosa and tumour of the left colon, our study is highly comparable with that of Notterman *et al*. In both studies, prior to analyses samples were defined with regard to Dukes' stage, snap frozen bulk tissue samples yielded high quality RNA, identical labelling and GeneChips were used, and the data were analysed using the Mann-Whitney U test. Notterman *et al* identified CAIV as being downregulated by 38-fold from normal colon to tumour, which is identical to our results. In summary, this strongly supports the hypothesis that a decrease in CAIV expression is linked to carcinogenesis and colon cancer progression.

Expression of genes such as COX-2, caldesmon 1, adipsin, transgelin 11, and ARHB was found to be higher in left sided compared with right sided adenocarcinomas. A previous study showed a better effect of chemoprevention with non-steroidal anti-inflammatory drugs on right sided than on left sided adenocarcinomas,[3] and the sixfold higher expression of COX-2 may explain failure to prevent this, as a higher dose may be needed to inhibit the high levels of this molecule in left sided malignant lesions. Loss of transgelin gene expression may be an important early event in tumour progression as a consequence of deregulation of RAS gene expression through RAF independent pathways.[18] Interestingly, ARHB (RhoB), located on chromosome 2pter-p12, is one of three RAS homologue gene family members and is known as an oncogene.[19]

From normal mucosa to Dukes' B and C of the caecum, we found that TRPM-2 (clusterin) and PGDH were strongly downregulated whereas several matrix metalloproteinases such as MMP1, MMP3, and MMP10 were upregulated, as seen previously in left sided tumours. MMPs are enzymes responsible for extracellular matrix degradation, playing a role in cancer progression and metastatic spreading. MMP1 expression is associated with a poor prognosis in colorectal cancer.[20] One possible therapeutic approach for patients with colon cancer, mainly Dukes' C, could therefore be administration of specific MMP inhibitors to prevent distant metastases and prolong survival,[21] as has been shown by inhibition of MMP2 expression in mouse xenograft experiments.[22] Circulating proenzymes of MMPs have been described as possible serum markers, and proMMP-9, but not proMMP-2 identified here, was found to be significantly higher in cancer sera versus normal sera.[23] From a clinical approach, we suggest analysis of sera levels of the MMPs identified here, as these molecules seem to be ideal general colonic tumour markers reflecting the presence of both left and right sided colonic tumours.

In conclusion, the 30 genes identified in adenocarcinomas of both sides have to be regarded as general tumour markers. The present data and our previously published LOH analyses[11] strongly support the hypothesis that genes such as aminopeptidase N (CD 13), sulfate transporter DTD, SCAD, or PCK1 should be regarded as potential new tumour suppressors requiring further investigation.

Paraffin embedded tissue sections from Dukes' B and C tumours and their matching normal mucosa were subjected to immunohistochemical analysis for COX-2 and cytokeratin 20 (KRT20). Microarray analysis showed a significant decrease in KRT20 from normal mucosa to tumour in the left side of the colon. Immunostaining confirmed the difference seen between the two sides of the colon as KRT20 was strongly downregulated in 80% of left sided tumours compared with 50% of right sided tumours. In general, KRT20 staining was found to be very heterogeneous within the tumours.

COX-2 microarray data showed that COX-2 was upregulated from normal to Dukes' C in right as well as left sided tumours, but the increase was significant only for left side ($p<0.005$). Immunostainings support the microarray based findings to date, that COX-2 is not or very weakly expressed in left sided mucosa but upregulated in matching tumours. In contrast, COX-2 is expressed with the same intensity in right sided normal mucosa compared with matching tumours. COX-2 is heterogeneously expressed within a tumour, as only some groups of cells within a tumour are stained. In conclusion, the microarray based findings were confirmed by immunohistochemistry but an absolute quantitative comparison between RNA expression on microarrays and protein expression on tissue specimen is not possible for KRT20 and COX-2 due to their heterogeneous staining patterns.

The existence of a side specific expression difference for COX-2, having been identified by microarray analysis and confirmed by immunohistochemistry in this study, has recently been reported by Nasir and colleagues.[24] Immunohistochemical staining applying a COX-2 polyclonal antibody on 18 right sided versus 18 left sided adenocarcinomas showed that COX-2 positivity was significantly higher for left compared with right sided tumours.

We conclude that differences in gene expression between normal mucosa as well as adenocarcinomas of the caecum and sigmoid and rectosigmoid colon clearly exist, and we hypothesise that the difference in gene expression could be related to differences in tumour development and the prognosis of patients.

The emerging treatments directed towards specific molecular targets should emphasise the differences seen in right and left sided tumours of the colon. We suggest that some of the highly expressed molecules that are in both left and right sided colonic adenocarcinomas may be promising new potential serum markers and therapy targets.

ACKNOWLEDGEMENTS

We thank Bente Devantier, Ing Lis Thorsens, and Annette B. Nielsen for their technical assistance, and also as project-nurse Edith Kirkedahl Nielsen at Aarhus Sygehus for collection of colon tissue samples. The study was supported by funds from the Karen Elise Jensen Foundation, the Danish Research Council, AROS Applied Biotechnology Aps, Aarhus, the University and County of Aarhus, the Nordic Cancer Union, and the European Union's 5th frameprogram (European Community, No QLG2-CT-2001-01861).

All supplementary data and supplementary tables 1–7 can be viewed on the *Gut* website at http://www.gut.com/supplemental.

Authors' affiliations

K Birkenkamp-Demtroder, S H Olesen, T F Ørntoft, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital/Skejby Hospital, Aarhus N, Denmark
F B Sørensen, Institute of Pathology, Aarhus University Hospital, Aarhus Sygehus, Aarhus C, Denmark
S Laurberg, Surgical Department L, Aarhus University Hospital, Aarhus Sygehus, Aarhus C, Denmark
P Laiho, L A Aaltonen, Department of Medical Genetics, Biomedicum, University of Helsinki, Helsinki, Finland

Conflict of interest: None declared.

REFERENCES

1 **Campbell F**, Appleton MA, Shields CJ, *et al.* No difference in stem cell somatic mutation between the background mucosa of right- and left-sided sporadic colorectal carcinomas. *J Pathol* 1998;**186**:31–5.
2 **Distler P**, Holt PR. Are right- and left-sided colon neoplasms distinct tumors? *Dig Dis* 1997;**15**:302–11.
3 **Iacopetta B**. Are there two sides to colorectal cancer? *Int J Cancer* 2002;**101**:403–8.
4 **Konishi K**, Fujii T, Boku N, *et al.* Clinicopathological differences between colonic and rectal carcinomas: are they based on the same mechanism of carcinogenesis? *Gut* 1999;**45**:818–21.
5 **Reifferscheid M**, Fass J, Hartung R, *et al.* Special aspects of right colon cancer. *Langenbecks Arch Chir* 1987;**371**:193–200.
6 **Tomoda H**, Taketomi A, Baba H, *et al.* The clinicopathological characteristics and outcome of patients with right colon cancer. *Oncol Rep* 1998;**5**:481–3.
7 **Elsaleh H**, Joseph D, Grieu F, *et al.* Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet* 2000;**355**:1745–50.
8 **Kapiteijn E**, Liefers GJ, Los LC, *et al.* Mechanisms of oncogenesis in colon versus rectal cancer. *J Pathol* 2001;**195**:171–8.
9 **Fric P**, Sovova V, Sloncova E, *et al.* Different expression of some molecular markers in sporadic cancer of the left and right colon. *Eur J Cancer Prev* 2000;**9**:265–8.
10 **Glebov OK**, Rodriguez LM, Nakahara K, *et al.* Distinguishing right from left colon by the pattern of gene expression. *Cancer Epidemiol Biomarkers Prev* 2003;**12**:755–62.
11 **Birkenkamp-Demtroder K**, Christensen LL, Olesen SH, *et al.* Gene expression in colorectal cancer. *Cancer Res* 2002;**62**:4352–63.
12 **Alon U**, Barkai N, Notterman DA, *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1999;**96**:6745–50.
13 **Notterman DA**, Alon U, Sierk AJ, *et al.* Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 2001;**61**:3124–30.
14 **Zhang L**, Zhou W, Velculescu VE, *et al.* Gene expression profiles in normal and cancer cells. *Science* 1997;**276**:1268–72.
15 **Paredes-Zaglul A**, Kang JJ, Essig YP, *et al.* Analysis of colorectal cancer by comparative genomic hybridization: evidence for induction of the metastatic phenotype by loss of tumor suppressor genes. *Clin Cancer Res* 1998;**4**:879–86.
16 **Wehkamp J**, Schwind B, Herrlinger KR, *et al.* Innate immunity and colonic inflammation: enhanced expression of epithelial alpha-defensins. *Dig Dis Sci* 2002;**47**:1349–55.
17 **Takenawa J**, Kaneko Y, Kishishita M, *et al.* Transcript levels of aquaporin 1 and carbonic anhydrase IV as predictive indicators for prognosis of renal cell carcinoma patients after nephrectomy. *Int J Cancer* 1998;%20, **79**:1–7.
18 **Shields JM**, Rogers-Graham K, Der CJ. Loss of transgelin in breast and colon tumors and in RIE-1 cells by Ras deregulation of gene expression through Raf-independent pathways. *J Biol Chem* 2002;**277**:9790–9.
19 **Madaule P**, Axel R. A novel ras-related gene family. *Cell* 1985;**41**:31–40.
20 **Murray GI**, Duncan ME, O'Neil P, *et al.* Matrix metalloproteinase-1 is associated with poor prognosis in colorectal cancer. *Nat Med* 1996;**2**:461–2.
21 **Aparicio T**, Kermorgant S, Dessirier V, *et al.* Matrix metalloproteinase inhibition prevents colon cancer peritoneal carcinomatosis development and prolongs survival in rats. *Carcinogenesis* 1999;**20**:1445–51.
22 **Oba K**, Konno H, Tanaka T, *et al.* Prevention of liver metastasis of human colon cancer by selective matrix metalloproteinase inhibitor MMI-166. *Cancer Lett* 2002;**175**:45–51.
23 **Pucci-Minafra I**, Minafra S, La Rocca G, *et al.* Zymographic analysis of circulating and tissue forms of colon carcinoma gelatinase A (MMP-2) and B (MMP-9) separated by mono- and two-dimensional electrophoresis. *Matrix Biol* 2001;**20**:419–27.
24 **Nasir A**, Kaiser HE, Boulware D, *et al.* Cyclooxygenase-2 expression in right- and left-sided colon cancer: a rationale for optimization of cyclooxygenase-2 inhibitor therapy. *Clin Colorectal Cancer* 2004;**3**:243–7.