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

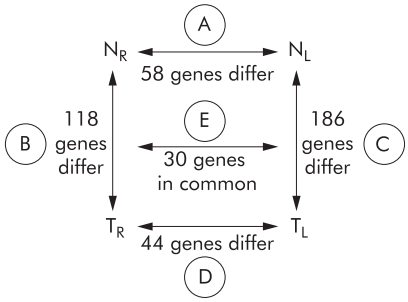

Figure 1 Schematic overview of the five different comparison groups (A–E). Comparison (A): normal right sided colon mucosa (N_R) from the caecum versus normal left sided (N_L) colon mucosa from the sigmoid and rectosigmoid. Comparison (B): normal right sided mucosa from the caecum (N_R) versus matched 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℃.

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).

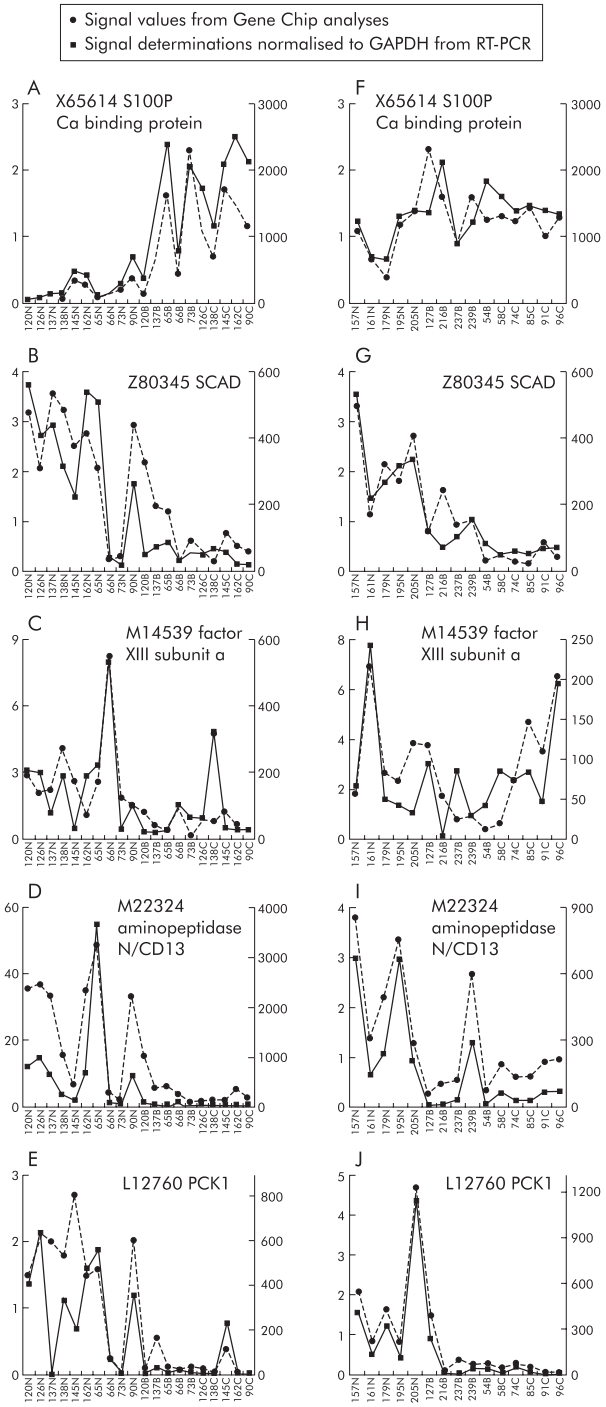

Figure 2 Comparison of single GeneChip analyses with real time polymerase chain reaction (PCR) analyses. Expression analyses of five selected genes using single samples of normal colon mucosa and adenocarcinomas of Dukes' stages B and C from the right (A–E) and left (F–J) sides of the colon. Left y axis 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

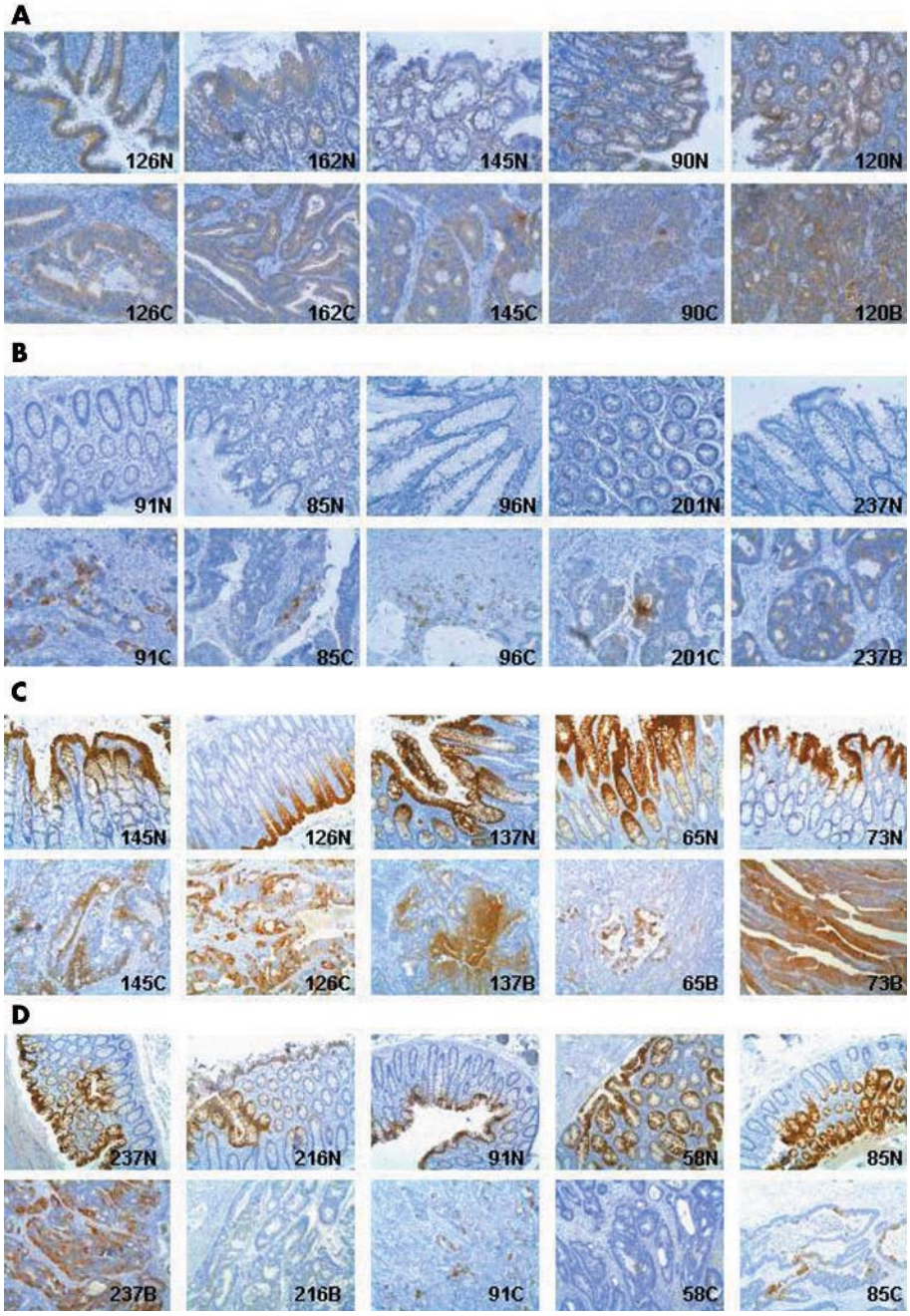

Figure 3 Immunohistochemistry of formalin fixed paraffin embedded sections of Dukes' B and C adenocarcinomas and their matching normal mucosa (N). Sample numbers of tumours refer to samples previously analysed on microarrays. Cyclooxygenase 2 (COX-2) was moderately expressed in right sided normal mucosa 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.

SUPPLEMENTARY DATA

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

Abbreviations: LCC, left-sided Colon Cancer. RCC, right-sided Colon Cancer. UG Cluster, Uni Gene Cluster http://www.ncbi.nlm.nih.gov/UniGene.

Tissue samples, patient information and RNA isolation. RCC samples from the caecum and LCC samples from the rectosigmoid or sigmoid thereby excluding the rectum in Dukes stages B and C were obtained fresh from surgery, taken from the luminal aspect of the tumours in the surgical specimens, immediately transferred to a solution containing SDS (sodiumdodecylsulfate) and Guanidinium isothiocyanate, snap frozen in liquid nitrogen and stored at -80°C. The location of samples in the colon is shown in detail in supplementary table1. The Dukes classification of the clinical stage of disease was applied according to the following criteria: Dukes B: Tumour has penetrated the muscle wall and possible infiltrated the pericolic or –rectal fat, with no detectable metastatic lymph nodes (equivalent to stage II or Astler-Collier As-Co stage B_2); Dukes C: Metastatic lymph nodes detectable (equivalent to stage III or As-Co stages C_1-C_2, irrespective of the depth of infiltration by the cancer in the colonic wall); Paired control samples regarded as “normal mucosa” were obtained from the oral resection margins of the operative specimens, by taking mucosal biopsies from the luminal aspect of the bowel wall. Informed consent was obtained from patients to use their specimens and clinico-pathological data for research purposes. All tumours were sporadic as none of the patients belonged to families with heritable colon cancer or other cancers and no patients had a previous tumour. The local scientific ethical commission approved the project.

Total RNA was isolated from about 50 mg of single tissue samples using a Polytron homogeniser followed by treatment with Trizol (Gibco Life Technologies, Invitrogen Corporation, Carlsbad, CA) according to the manufacturer's instructions. The total RNA was of high quality according to the photometrical measurements and calculation of the OD-ratio followed by optical inspection of the 28s and 18s RNA bands on agarosegel. GeneChip® (Affymetrix Inc., Santa Clara, CA) analysis of single samples was carried out on five right-sided Dukes stage B (median age 76), five Dukes stage C (median age 66) and ten matched "normal" samples (tissue from the oral resection margin, median age 70) and on eight left-sided Dukes stage B (median age 76), seven Dukes stage C (median age 68) and ten "normal mucosa" samples, with five of these patients being matched. Supplementary Table 1 shows detailed clinico-pathologic information as well as the approximate percentages of the volume-fractions of tumour cells and stromal cells in the tissue samples, estimated semi-quantitatively by an experienced pathologist, using paraffin embedded, diagnostic tissue sections (4 through 12 sections *per* tumour). The diagnostic samples from paraffin sections contained normal and tumour tissue as well as submucosal tissue from the colon wall. The estimated percentage of tumour cells is a conservative estimate as tissue used for RNA extraction was from the most superficial tumour rich areas avoiding most of the deeper stroma containing layers. 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.

Supplementary Table 1 *Clinical disease stage of the samples used for RNA expression analysis*

							surgical specimen				paraffin sections							
sample No.	matching	sex[b]	age	loc[c]	tissue[d]	TNM status[f]	length [cm][g]	Ø [mm][h]	distance [mm][i]	microsatellite stability status[j]	type WHO[k]	diff. Grade[l]	necrotic fraction [%] x20[m]	vital tumour fraction [%] n x20	vital tumour nuclear fraction[%] o x200	vital stromal nuclear fraction [%] x200	vital tumour cell fraction [%] p x40	vital stroma cell fraction [%][q] x40
Left sided Coloncancer (Sigmoid and rectosigmoid)																		
Normal (n = 10; mean age = 69, age range 52-83)																		
157N		f	75	7	N	T1N0M0												
161N		f	63	10	N	T3N1M0												
179N		f	74	10	N	T4N0M0	TNM status given for normal samples											
195N		f	76	7	N	T3N0M0	refers to the corresponding tumour sample.											
205N		m	54	10	N	T3N2M1												
203N	[a]	m	77	9	N	T3N0M0												
208N	[a]	m	83	9	N	T3N0M0												
202N	[a]	m	52	7	N	T3N0M0												
201N	[a]	m	78	9	N	T3N1M0												
204N	[a]	m	58	8	N	T3N3M0												
Dukes B (n = 8; mean age = 76, age range 52-83)																		
216B		f	79	7	B	T3N0M0	13	30	40		1	3	25	75	50	15	91	9
237B		f	82	7	B	T3N0M0	16	60	50	MSS	1	1	<50	50	50	20	71	29
239B		m	77	7	B	T2N0M0	15	15	5	MSS	1	2	<50	50	25	5	83	17
54B		f	81	7	B	T2N0M0	14	25	45		1	2	<50	50	75	50	91	9
127B		m	78	10	B	T3N0M0	30	30	20	MSS	1	2	75	25	75	5	83	17
203B	[a]	m	77	9	B	T3N0M0	19	55	40	MSS	1	2	20	80	80	20	70	30
208B	[a]	m	83	9	B	T3N0M0	20	70	30		1	2	<10	>90	70	25	75	25
202B	[a]	m	52	7	B	T3N0M0	25	80	40	MSS	1	3	10	90	75	30	50	50
Dukes C (n = 7; mean age = 68, age range 53-90)																		
58C		m	81	7	C	T3N1M0	30	40	45		1	2	<50	50	50	10	83	17
74C		m	61	10	C	T1N2M0	28	45	25		1	2	75	25	50	10	63	37
85C		m	56	9	C	[e] T4N2M0	26	130	70	MSS	1	3	75	25	>75	<5	83	17
91C		f	53	10	C	T2N1M0	30	20	10		1	3	75	25	50	10	63	37
96C		m	90	7	C	T2N1M0	14	60	60		1	2	75	25	50	10	63	37
201C	[a]	m	78	9	C	[e] T3N1M0	31	40	40	MSS	1	2	10	90	50	30	40	60
204C	[a]	m	58	8	C	[e] T3N3M0	15	50	50		1	3	30	70	60	20	40	60
Right sided Coloncancer (Caecum)																		
Normal (n = 10; mean age = 70, age range 47-92)																		
65N	[a]	f	80	1	N	T3N0M0												
66N	[a]	f	66	1	N	T3N0M0												
73N	[a]	f	78	1	N	T3N0M0												
120N	[a]	f	73	1	N	T4N1M0												

137N	[a]	m	79	1	N		T2N0M0	TNM status given for normal samples											
90N	[a]	f	58	1	N		T3N1M0	refers to the corresponding tumour sample.											
126N	[a]	f	47	1	N		T3N2M0												
145N	[a]	f	72	1	N		T3N1M0												
138N	[a]	m	60	1	N		T3N1M0												
162N	[a]	f	92	1	N		T3N1M0												
Dukes B Single (n = 5; mean age = 75, age range 66-80)																			
65B	[a]	f	80	1	B		T3N0M0	15	100	70		1	2	20	80	75	40	85	15
66B	[a]	f	66	1	B		T3N0M0	-	55	50		1	2	<10	>90	50	25	50	50
73B	[a]	f	78	1	B		T3N0M0	25	35	40		1	2	<5	>95	60	30	70	30
120B	[a]	f	73	1	B		T4N1M0	25	20	70	MSI-H	1	3	50	50	60	20	75	25
137B	[a]	m	79	1	B		T2N0M0	40	90	120	MSS	1	1	10	90	60	20	75	25
Dukes C Single (n = 5; mean age = 66 age range 47-92)																			
90C	[a]	f	58	1	C		T3N1M0	17	70	45		1	3	25	75	80	25	70	30
126C	[a]	f	47	1	C		T3N2M0	49	50	80		1	2	20	80	75	10	60	40
145C	[a]	f	72	1	C		T3N1M0	25	30	100	MSS	1	2	25	70	75	20	50	50
138C	[a]	m	60	1	C	[e]	T3N1M0	60	14	140		1	2	50	50	75	10	60	40
162C	[a]	f	92	1	C		T4N1M0	28	90	30		2	3	20	80	60	20	60	40

[a] Matching normal and tumour samples have the same number, thereby differing in the letter code (N= normal tissue; B and C = tumours Dukes B and C.

[b] sex m=male; f=female

[c] location of samples in the colon (1=caecum; 7-9 sigmoid; 10=rectosigmoid)

[d] N= normal tissue; B and C = tumours Dukes B and C; defined in material and methods

[e] primary tumour characterized as Dukes C, control CT-scan after 3-12 month showed development of distant metastases

[f] "T0" noninvasive. "T1-T4" invasive tumours. (T1 submucosa; T2 tunica muscularis; T3 subserosa; T4 peritoneum or other organs)
"N0" no malign lymphnodes; "N1" 1-3 lymphnode metastases; "N2" >4 lymphnode metastases; "N3" metastatic apical node (marked by surgeon)
"M0" no distant metastases; "M1" distant metastasis, in e.g. liver or lung)

[g] length of surgically removed colon part (cm)

[h] tumours largest diameter (mm)

[i] tumour distance (mm) to nearest colon resection margin

[j] microsatellite stability status was determined with BAT 25 og BAT 26: MSS = microsatellite stable; MSI-H = microsatellite highly instable

[k] Histological type according to WHO (1=adenocarcinoma NOS; 2=mucinous adenocarcinoma)

[l] Predominant histologiocal differentiation grade (1=high; 2= moderate; 3= low)

[m] applied magnification

[n] including both, tumour and stroma

[o] corresponds to the volume fraction of nuclei in vital tumour areas ; including fibroblasts, endothelial and inflammatory cells

[p] percentage volume-fraction of tumour cells (including myxoid and mucinous tumour areas)
calculated as [(100 x ((vital tumour x nuclei)/((vital stroma x nuclei)+(vital tumour x nuclei))]

[q] including collagene and hyalinized areas

Percentages given in the table apply to the luminal aspect of the adenocarcinomas, in that deep parts of the tumours, showing pronounced hyalinization have not been included in the biopsies used for molecular biological investigation.

cRNA preparation. The first- and second-strand cDNA synthesis was performed using the SuperScript Choice System (Invitrogen Corporation, Carlsbad, CA) and 12µg of total RNA according to the manufacturer's instructions. An exception was to use an T7-oligo $(dT)_{24}$-primer containing a T7 RNA polymerase promoter site (DNA Technology A/S, Aarhus C, Denmark), as described by Affymetrix Inc. (Santa Clara, CA). Second strand cDNA synthesis was followed by incubation in 50 mM NaOH/0,1 mM EDTA for 10 min at 65°C leading to degradation of rRNA and tRNA. Labeled cRNA was prepared using the BioArray High Yield RNA Transcript Labeling kit (Enzo Biochem, Inc., Farmingdale, NY). Biotin-labeled CTP and UTP (Enzo Biochem, Inc., Farmingdale, NY) were used in the reaction, together with unlabeled nucleotide triphosphates. Unincorporated nucleotides were removed from the in vitro transcript using RNeasy spin columns (QIAGEN GmbH, Hilden, Germany).

Array Hybridisation and Scanning. Fifteen µg of cRNA were fragmented at 94°C for 35 min in a fragmentation buffer containing 40 mM Tris-acetate (pH 8.1), 100 mM potassium acetate, 30 mM magnesium acetate. Before hybridisation, the fragmented cRNA in a 6 x SSPE-T hybridisation buffer [1 M NaCl, 10 mM Tris (pH 7.6), and 0.005% Triton-X 100] was heated to 95°C for 5 min and subsequently to 40°C for 5 min before loading onto the Affymetrix probe array cartridge. Probe arrays "HuGeneFL" with 7,129 datasets (Affymetrix Inc., Santa Clara, CA) were incubated for 16 h at 45°C at constant rotation (60 rpm). The washing and staining procedures were performed in the Affymetrix Fluidics Station. The probe array was exposed to 10 washes in 6 x SSPE-T at 25°C, followed by four washes in 0.5 x SSPE-T at 50°C. The biotinylated cRNA was stained with a SAPE (streptavidin-phycoerythrin) conjugate (final concentration, 2 µg/µl; Molecular Probes, Eugene, OR) in 6xSSPE-T for 30 min at 25°C, followed by 10 washes in 6 x SSPE-T at 25°C. An antibody amplification step was added using normal goat IgG (final concentration, 0.1 mg/ml; Sigma Chemical Co., St. Louis, MO) and anti-streptavidin antibody (goat) biotinylated (final concentration, 3 µg/ml; Vector Laboratories, Burlingame, CA). This was followed by a staining step with a SAPE conjugate (final concentration, 2 µg/µl; Molecular Probes, Eugene, OR) in 6 x SSPE-T for 30 min at

25°C and 10 washes in 6xSSPE-T at 25°C. The probe arrays were scanned at 560 nm using a confocal laser-scanning microscope with an argon ion laser as the excitation source (Hewlett Packard GeneArray Scanner G2500A). The readings from the quantitative scanning were analysed by the Affymetrix Gene Expression Analysis Software and scaled to a global intensity of 150, as published previously [(1)].

Data analysis and selection of genes GeneChip® (Affymetrix Inc., Santa Clara, CA) analysis of single samples was carried out on ten samples from the caecum (65B, 66B, 73B, 120B, 137B, 90C, 126C, 145C, 138C, 162C), five Dukes stage B (median age 76), five Dukes stage C (median age 66). Each of the tumours was accompanied by the corresponding matched normal mucosa sample at the same location from the same patient . Matched samples are characterised by the same sample number, thereby discriminated 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), seven Dukes stage C (median age 68) and ten “normal mucosa” samples (median age 69). Five of these tumours (201C, 202B, 203B, 204C and 208C) were accompanied by the corresponding matched normal sample at the same location from the same patient. The remaining five normal mucosa samples (157N, 161N, 179N, 195N, 205N) and ten tumour samples (16B, 237B, 239B, 54B, 127B, 58C, 74C, 85C, 91C, 96C) were obtained from individual patients who underwent resection of the sigmoid or rectosigmoid colon, respectively (supplementary table 1).

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”.

Terms used for analysis are defined as follows: “Median” is the middle value of a set of values, “Signal” is a measure of the abundance of the transcript, “Detection Call” indicates whether the transcript is detected (P, present), undetected (A, absent), or at the limit of detection (M, marginal). “Signal log ratio” is the change in the expression level of a transcript between a baseline and an experiment array. This change is expressed as the log2 ratio. A log2 ratio of 1 is equal to a fold change of 2. The “Change Call” indicates the change in the transcript level between a baseline and

experiment (Increase (I), Marginal increase (MI), No Change (NC), Marginal Decrease (MD), Decrease (D)). “Fold Change” corresponds to "signal ratio" and is obtained by calculation from "signal log ratio". Five different comparison groups A-E were established given as schematic overview in figure 1.

In comparison “A” we compared the “Median” of ten samples of right-sided normal mucosa to the “Median” of ten samples of left sided normal mucosa. The analysis of 100 single comparison of each left-sided to each right-sided sample was not practicable in this case as the analysis software used here is restricted with regard to the number of comparisons. In comparison “B” expression of each of the five Dukes B and five Dukes C adenocarcinomas was compared to its matching normal mucosa from the same patient. In comparison “C” each of the 5 non-matched normal samples of the left side was compared to each of five Dukes B and five Dukes C yielding 50 comparisons and in addition three Dukes B and two Dukes C were compared to their matching normal mucosa from the same patient (5 comparisons). In comparison “D” each of the five Dukes B or Dukes C, of the right side was compared to each of the Dukes B (8 samples) or Dukes C (7 samples) of the left side yielding 75 comparisons in total. In comparison “E” 118 genes from comparison “B” were compared to 186 genes from comparison “C” to identify common cancer genes of both sides of the colon.

Comparison A – Normal CAECUM VS SIGMOID/RECTOSIGMOID. We compared ten samples of right-sided normal colonic mucosa (caecum exclusively) to ten samples of left sided normal mucosa (sigmoid and rectosigmoid). Comparisons were based on the “Median” of the two groups using MAS 5.0, MDB 3.0 and DMT 3.0, in total 7129 datasets. **Filter 1:** 3194 datasets were excluded comprising Affymetrix-Markers or having a “Detection”-call “absent” in more than 80 % of the cases (16 of 20 arrays) or more than 70% in each normal caecum (7 of 10 arrays) and normal sigmoid and rectosigmoid (7 of 10 arrays). **Filter 2:** A Mann–Whitney U-Test with an exclusion limit of $p<0.05$ using normal caecum as control baseline was applied to the remaining 3935 datasets resulting in 523 genes significantly changing. Of these, 127 genes had a p-value <0.01. **Filter 3:**

Fold changes were determined using the median of the “Signal”-calls from each group resulting in 160 genes showing expression differences of ≥ 1.5 fold or ≤ -1.5 fold, 58 genes with a $p<0.01$ (Table 1 in the main text).

Comparison B – Normal vs Tumour CAECUM. Each one of the five Dukes B and five Dukes C tumour samples was compared to its matching normal tissue from the same patient (10 in total) using MAS 5.0, MDB 3.0 and DMT 3.0. **Filter 1:** Affymetrix markers, genes which were absent in more than 90% (18 of 20) of all arrays of normal, Dukes B and Dukes C from the right sided Colon as well as datasets with a “Change”-call of “not changed” in more than 70% (7 of 10) of the pair wise comparisons were excluded from analysis resulting in 3026 datasets. **Filter 2:** Genes were included if more than 80% (4 of 5 comparisons) of the pair wise comparisons showed a “Change”-call of increased or decreased, resulting in 508 genes in Dukes B, 987 genes in Dukes C and 403 genes in both Dukes B and C in total. **Filter 3:** Genes were included with an average “signal log ratio” of ≥ 1.5 or ≤ -1.5 based on single comparisons (corresponding to fold changes of ≥ 2.8 or ≤ -2.8) and those showing a significant up- or down regulation with an exclusion limit of $p<0.05$ in a Mann-Whitney U-Test, resulting in 118 genes (Supplementary Table 2).

Supplementary Table 2 One-hundred-eighteen genes differentially expressed more than 2,8-fold comparing normal mucosa to matched Dukes B or C a

Probe set ID	Gene name	UniGene Cluster	Cyto Band	Ncae med[a]	Bcae med[b]	Ccae med[c]	avg FC[d] NvsB	avg FC NvsC	p-value
AB002409_at	SLC Secondary Lymphoid-Tissue Chemokine	Hs.57907	9p13	187	34	44	-4,8	-2,6	0,001
AF001294_at	IPL (IPL) .	Hs.154036	11p15.5	178	557	402	3,7	2,1	0,001
AF001548_rna1_	complete sequence.	Hs.78344	3-p13.12	1165	99	261	-8,0	-5,8	0,000
D00654_at	enteric smooth muscle gamma-actin			420	88	87	-6,0	-3,6	0,005
D10667_s_at	smooth muscle myosin heavy chain			77	18	18	-10,2	-11,2	0,000
D13168_at	gene endothelin-B receptor (hET-BR)			11	6	4	-2,0	-3,4	0,002
D17408_s_at	calponin	Hs.21223	p13.1	257	47	79	-5,2	-2,9	0,001
D83777_at	KIAA0193 gene	Hs.75137	7	71	70	191	-1,5	3,2	0,174
D86062_s_at	KNP-Ib	Hs.182423	21q22.3	64	45	20	-1,0	-3,9	0,034
D87292_at	TST Thiosulfate sulfurtransferase (Rhodanese)	Hs.248267	22q13.1	2649	1089	749	-1,0	-3,6	0,034
HG2197-HT2267_	Collage Type Vii Alpha 1			65	118	182	-1,0	2,9	0,023
HG2614-HT2710_	Collagen Type Viii Alpha 1			19	38	69	1,3	5,2	0,010
HG2755-HT2862_	T-Plastin			47	135	284	1,4	3,7	0,016
HG2981-HT3125_	Epican Alt. Splice 1			16	78	75	1,7	4,1	0,023
HG2981-HT3127_	Epican Alt. Splice 11			33	537	165	3,5	4,9	0,008
HG2981-HT3938_	Epican Alt. Splice 12			5	219	75	5,4	9,9	0,007
HG3431-HT3616_	Decorin Alt. Splice 1			295	78	350	-3,8	1,2	0,406
HG4263-HT4533_	Nkr-P1a Protein			43	3	3	-3,0	-8,7	0,007
HG4312-HT4582_	Transcription Factor Iiia			245	556	781	1,9	3,1	0,001
HG4582-HT4987_	Beta-1-Glycoprotein 1 M20882			44	12	14	-3,5	-1,8	0,004
J00306_at	somatostatin I gene and flanks	Hs.12409	3q28	73	8	7	-4,3	-11,8	0,006
J02854_at	20-kDa myosin light chain (MLC-2)	Hs.9615	-	472	73	167	-4,1	-1,7	0,007
J02874_at	adipocyte lipid-binding protein	Hs.83213	8q21	126	13	9	-9,1	-6,9	0,001
J03507_at	complement protein component C7	Hs.78065	5p13	88	10	11	-6,4	-4,6	0,000
J03910_rna1_at	(clone 14VS) metallothionein-IG (MT1G) gene	Hs.173451	16q13	1417	219	279	-1,6	-9,7	0,016
J04970_at	carboxypeptidase M 3' end	Hs.169765	12q15	55	21	19	-1,8	-3,3	0,019
J05096_rna1_at	NaK-ATPase subunit alpha 2 (ATP1A2)	Hs.34114	1q21	15	4	9	-5,5	-1,2	0,002
K02765_at	subunits	Hs.284394	p13.2	472	171	335	-6,7	-1,2	0,059
L07615_at	neuropeptide Y receptor Y1 (NPYY1)			23	12	4	-1,8	-5,8	0,007
L08895_at	MADS/MEF2-family transcription factor (MEF2C)	Hs.78995	5q14	29	9	16	-3,5	-2,1	0,007
L12350_at	thrombospondin 2 (THBS2)	Hs.108623	6q27	24	53	136	-1,6	6,1	0,049
L12760_s_at	with repeats	Hs.1872	1	510	35	12	-3,2	-22,4	0,003
L22524_s_at	matrilysin gene			7	62	179	6,0	23,0	0,001
L29008_at	L-iditol-2 dehydrogenase	Hs.878	15q15.3	145	525	491	1,5	3,1	0,016
L29433_at	factor X (blood coagulation factor) gene			57	24	16	-3,2	-2,1	0,013
L38486_at	microfibril-associated glycoprotein 4 (MFAP4)	Hs.118223	17p11.2	206	49	39	-3,2	-3,3	0,000

L43821_at	enhancer of filamentation (HEF1)	Hs.80261	-	72	28	30	-1,7	-3,4	0,008
L76465_at	dehydrogenase (PGDH)	Hs.77348	q35	211	41	11	-4,5	-16,9	0,000
M10942_at	metallothionein-Ie gene (hMT-Ie)	Hs.74170	16q13	551	228	274	-1,7	-3,6	0,007
M12759_at	Ig J chain gene			1189	227	60	1,1	-22,1	0,016
M12963_s_at	subunit	Hs.73843	4q21	1647	282	31	-3,8	-46,7	0,000
M14539_at	factor XIII subunit a 3' end	Hs.80424	6p25.3	159	42	52	-4,6	-3,1	0,000
M21005_at	migration inhibitory factor-related protein 8 (MRP8)	Hs.100000	1q21	54	73	190	-1,3	3,7	0,131
M21305_at	sequence. /gb=M21305 CDS	Hs.247946	-	92	13	315	-6,1	2,8	0,880
M22324_at	N	Hs.1239	15q25	2169	357	128	-2,8	-9,0	0,003
M25753_at	cyclin B 3' end	Hs.23960	5q12	42	106	167	3,2	2,7	0,001
M26311_s_at	cystic fibrosis antigen	Hs.112405	1q21	51	106	360	1,0	11,7	0,034
M26576_cds2_at	alpha-1 collagen type IV gene exon 52.			107	233	429	1,4	3,7	0,002
M31994_at	cytosolic aldehyde dehydrogenase (ALDH1) gene			353	205	84	-1,8	-3,4	0,004
M32053_at	H19 RNA gene (spliced in silico)			68	71	161	-1,6	3,4	0,406
M54927_at	myelin proteolipid protein	Hs.1787	Xq22	16	5	3	-5,6	-4,7	0,001
M60828_at	keratinocyte growth factor	Hs.164568	15q15	10	5	10	-4,3	2,1	0,496
M63379_at	TRPM-2 protein gene			918	285	255	-5,1	-4,0	0,000
M63603_at	phospholamban	Hs.85050	6q22.1	29	7	20	-7,1	-1,4	0,003
M77349_at	(BIGH3)	Hs.118787	5q31	253	582	1933	2,1	5,9	0,003
M80482_at	subtilisin-like protein (PACE4)	Hs.170414	15q26	269	149	60	-1,3	-3,2	0,004
M84526_at	adipsin/complement factor D	Hs.155597	19	872	249	39	-5,0	-21,1	0,002
M87860_at	S-lac lectin L-14-II (LGALS2) gene			57	35	7	1,0	-7,8	0,059
M95787_at	22kDa smooth muscle protein (SM22)	Hs.75777	11q23.2	1383	360	542	-3,9	-2,0	0,003
M97252_at	Kallmann syndrome (KAL)	Hs.89591	Xp22.32	3	4	16	-1,4	5,0	0,130
S45630_at	alpha B-crystallin=Rosenthal fiber component	Hs.1940	q23.1	220	88	60	-3,6	-2,3	0,002
S67156_at	ASP=aspartoacylase [kidney 1435 nt]	Hs.32042	17pter	25	7	2	-5,0	-4,4	0,000
S75256_s_at	OC6 Partial 534 nt]. /gb=S75256			345	3135	1831	4,4	1,6	0,019
S78187_at	CDC25Hu2=cdc25+ homolog [3118 nt]	Hs.153752	20p13	218	814	640	3,3	3,4	0,001
U03688_at	dioxin-inducible cytochrome P450 (CYP1B1)	Hs.154654	2p21	8	3	8	-4,0	1,8	0,406
U05861_at	hepatic dihydrodiol dehydrogenase gene			99	43	32	-3,4	-2,4	0,019
U10485_at	lymphoid-restricted membrane protein (Jaw1)	Hs.40202	-	51	3	9	-9,9	-6,7	0,002
U14528_at	sulfate transporter (DTD)	Hs.29981	5q31	679	85	18	-1,8	-28,1	0,019
U18018_at	E1A enhancer binding protein (E1A-F)	Hs.77711	17q21	67	340	261	3,3	4,4	0,000
U19495_s_at	intercrine-alpha (hIRH)	Hs.237356	10q11.1	186	6	14	-9,4	-5,2	0,002
U20758_rna1_at	osteopontin gene .	Hs.313	4q21	15	80	120	4,6	12,8	0,000
U23143_at	gene nuclear encoded mitochondrion protein	Hs.75069	12q12	126	301	160	2,8	1,4	0,000
U24488_s_at	tenascin-X (XA)	Hs.283750	6p21.3	205	8	8	-20,3	-13,3	0,001
U25997_at	stanniocalcin precursor (STC)	Hs.25590	8p21	2	29	48	4,5	30,5	0,001
U28368_at	Id-related helix-loop-helix protein Id4	Hs.34853	6p22	26	4	4	-5,8	-2,9	0,008
U37283_at	microfibril-associated glycoprotein-2 MAGP-2	Hs.58882	p12.3	53	17	7	-6,2	-2,5	0,008

U39447_at	placenta copper monamine oxidase	Hs.198241	17q21	40	14	11	-3,4	-1,7	0,001
U41518_at	clone AQP-1-2344 partial cds	Hs.74602	7p14	157	76	87	-2,9	-1,2	0,034
U43328_at	link protein	Hs.2799	5q14.3	16	4	1	-1,3	-16,6	0,003
U48959_at	myosin light chain kinase (MLCK)	Hs.211582	q21	386	83	146	-4,0	-3,0	0,000
U54617_at	pyruvate dehydrogenase kinase isom 4	Hs.8364	7q21.3	21	7	7	-3,0	-3,1	0,002
U60115_at	skeletal muscle LIM-protein SLIM1	Hs.239069	Xq26	598	96	93	-6,1	-5,6	0,000
U61374_at	cysteines	Hs.15154	Xp21.1	135	23	12	-4,6	-4,6	0,002
U70663_at	zinc finger transcription factor hEZF (EZF)	Hs.7934	9q31	587	124	71	-3,0	-10,3	0,001
U71207_at	eyes absent homolog (Eab1) .	Hs.29279	20q13.1	44	20	6	1,1	-6,4	0,016
U77180_at	3beta)	Hs.50002	9p13	104	3	3	-10,4	-13,6	0,002
U77643_at	K12 protein precursor	Hs.95655	17q25	536	146	165	-1,7	-3,1	0,007
U78551_at	gallbladder mucin MUC5B partial cds	Hs.102482	11p15	154	421	111	4,7	-3,4	0,734
U81607_at	gravin	Hs.788	6q24	64	24	55	-3,3	1,8	0,173
X00371_rna1_at	myoglobin gene (exon 1) (and joined CDS).	Hs.118836	22q13.1	213	44	65	-7,1	-3,4	0,000
X03350_at	allele)	Hs.4	4q21	85	12	5	-9,0	-17,2	0,000
X05232_at	stromelysin	Hs.83326	11q22.3	12	165	243	15,5	33,6	0,000
X06562_at	growth hormone receptor	Hs.125180	5p13	15	1	2	-6,4	-4,0	0,010
X07820_at	metalloproteinase stromelysin-2	Hs.2258	11q22.3	4	28	20	10,1	7,1	0,001
X13839_at	vascular smooth muscle alpha-actin	Hs.195851	10q23.3	2158	730	1279	-2,9	-1,1	0,023
X51405_at	carboxypeptidase E (EC 3.4.17.10)	Hs.75360	4	22	10	13	-3,4	-1,8	0,002
X52003_at	pS2 protein gene	Hs.1406	21q22.3	313	903	2118	4,7	1,7	0,096
X53331_at	matrix Gla protein	Hs.75742	p12.3	698	95	264	-5,6	-1,9	0,004
X54162_at	ocular muscle	Hs.79386	1	126	21	36	-5,3	-4,5	0,000
X54925_at	type I interstitial collagenase	Hs.83169	11q22.3	10	248	244	16,9	36,7	0,000
X57025_at	IGF-I insulin-like growth factor I	Hs.85112	12q22	28	12	10	-3,4	-2,6	0,001
X59770_at	CB23)	Hs.25333	2q12	160	54	66	-1,7	-2,9	0,010
X63629_at	p cadherin	Hs.2877	16q22	15	218	142	5,1	8,4	0,005
X64559_at	tetranectin	Hs.65424	3p22	262	10	9	-17,0	-19,2	0,001
X65614_at	calcium-binding protein S100P	Hs.2962	4p16	170	1392	2087	7,4	6,1	0,000
X66945_at	N-sam fibroblast growth factor receptor	Hs.748	8p11.2	179	81	109	-3,4	-1,2	0,023
X74837_at	HUMM9	Hs.25253	6q22	44	21	21	-1,7	-2,9	0,016
X82209_at	MN1	Hs.268515	22q12.1	21	3	14	-6,5	-1,1	0,059
X83490_s_at	Fas/Apo-1 (clone pCRTM11-Fasdelta(34))			27	15	13	-1,4	-2,9	0,016
X86693_at	hevin like protein	Hs.75445	-	230	80	132	-3,3	-1,5	0,002
X99133_at	NGAL gene	Hs.204238	9q34	309	1326	1047	3,1	2,0	0,041
Y07707_at	ITBA4 gene.	Hs.119018	-	17	41	41	1,9	2,9	0,000
Y09836_at	3'UTR of unknown protein	Hs.82503	-	82	28	49	-4,4	1,3	0,112
Y10032_at	putative serine/threonine protein kinase	Hs.159640	6q23	210	64	54	-2,8	-3,0	0,002
Z22865_at	dermatopontin	Hs.80552	1q12	74	17	23	-4,5	-2,7	0,001
Z29574_at	gene BCMA peptide	Hs.2556	16p13.1	88	30	14	-2,1	-4,2	0,005

Z49269_at	gene chemokine HCC-1	Hs.20144	17q11.2	80	22	32	-4,7	-3,5	0,002
Z80345_rna1_s_a	SCAD gene 5 UTR exon 1 and 2	Hs.127610	12q22	399	181	55	-1,3	-7,0	0,013

[a] Ncae med: Median derived from "Signal" of 10 Normal mucosae of the caecum

[b] Bcae med: Median derived from "Signal" of 5 Dukes' B adenocarcinomas of the caecum

[c] Ccae med: Median derived from "Signal" of 5 Dukes' C adenocarcinomas of the caecum

[d] FC: Fold Change, corresponding to "signal ratio" of Ncae med/Bcae med or Ncae med/Ccae med was calculated from "signal log ratio".

Comparison C - Normal vs Tumour SIGMOID/RECTOSIGMOID. Each of the 5 non-matched normal samples of the left side was compared to each of five Dukes B and five Dukes C, respectively of the left side yielding 50 comparisons in total. In addition, three Dukes B and two Dukes C were compared to their matching normal mucosa from the same patient (5 comparisons in total), using Affymetrix MAS 5.0, MDB 3.0 and DMT 3.0. **Filter 1:** Affymetrix markers, genes with an "Detection"-call "absent" in > 90% (23 of 25) of all arrays of normal, Dukes B and Dukes C from the left sided Colon as well as datasets with a "Change"-call of "not changed" in more than 70% of the comparisons (39 of 55) were excluded from analysis resulting in 3007 datasets. **Filter 2:** Genes were included if 100% of the pair wise comparisons (535 genes of 3/3 matched samples) and > 80% of the each-to-each comparisons of Dukes B (346 genes of 20/25 each-to-each), Dukes C (906 genes of 2/2 matched samples and 639 of 20/25 each-to-each) showed a "change"-call of increased or decreased. This resulted in a total of 153 genes in common in both B-groups and 329 in common in both C-groups. **Filter 3:** Genes were included with an average "signal log ratio" of ≥ 1.5 or ≤ -1.5 derived from the 55 comparisons resulting in 62 genes in Dukes B, 168 in Dukes C and 37 genes in both Dukes B and C. **Filter 4:** Genes which were significantly up- or down regulated with an exclusion limit of $p<0.05$ in a Mann-Whitney U-Test. We identified 186 genes significantly differential expressed more than 2.8 fold ($p<0.05$) from normal to Dukes B or Dukes C tumours (Supplementary Table 3).

Supplementary Table 3 One-hundred-eighty-six genes differentially expressed more than 2,8-fold comparing normal mucosa

Probe Set ID	Gene name	UniGene Cluster	Cyto Band	Nsig med[a]	Bsig med[b]	Csig med[c]	avg FC[d] NvsB	avg FC NvsC	p-value
AB006781_s_at	galectin-4	Hs.5302	-	4645	2577	1540	-2,2	-3,5	0,000
D00408_s_at	fetal liver cytochrome P-450 (P-450 HFLa)	Hs.172323	7	647	502	196	-1,3	-2,9	0,035
D11151_at	DNA endothelin-A receptor 5' flanking region			17	23	95	1,3	4,8	0,002
D13666_s_at	osteoblast specific factor 2 (OSF-2os)	Hs.136348	13	133	180	1531	-1,4	8,2	0,006
D16294_at	mitochondrial 3-oxoacyl-CoA thiolase	Hs.32500	18	917	447	260	-1,9	-2,9	0,000
D17793_at	KIAA0119 gene	Hs.78183	10p15	354	316	103	-1,0	-2,9	0,023
D21255_at	OB-cadherin-2	Hs.75929	16q22.1	17	42	231	2,0	16,9	0,000
D30037_at	phosphatidylinositol transfer protein (PI-TPbeta)	Hs.7370	22q12.1	28	74	83	3,0	3,6	0,000
D42047_at	KIAA0089 gene partial cds	Hs.82432	3	422	238	163	-1,9	-3,2	0,001
D45917_s_at	TIMP-3 partial cds (C-terminus region)	Hs.245188	22q12.3	41	65	175	1,2	2,9	0,006
D83174_s_at	collagen binding protein 2	Hs.9930	11q13.5	268	424	909	1,1	3,0	0,001
D84239_at	IgG Fc binding protein	Hs.111732	-	2819	233	171	-21,0	-23,4	0,000
D86479_at	AEBP1 gene	Hs.118397	7p13	95	142	485	-1,0	5,1	0,015
D86956_at	KIAA0201 gene	Hs.36927	-	133	422	373	3,0	3,1	0,000
D87258_at	cancellous bone osteoblast serin protease	Hs.75111	10q25.3	130	213	843	-1,0	4,9	0,008
D87292_at	TST Thiosulfate sulfurtransferase (Rhodanese)	Hs.248267	22q13.1	2492	1166	489	-2,1	-4,5	0,000
D87449_at	KIAA0260 gene partial cds	Hs.82635	1	254	108	73	-2,2	-2,8	0,000
D90042_at	liver arylamine N-acetyltransferase	Hs.2	8p22	88	53	27	-1,6	-3,6	0,000
HG2348-HT2444_s_	Peptide Yy			575	133	52	-3,4	-8,4	0,000
HG2614-HT2710_at	Collagen Type Viii Alpha 1			35	96	244	1,7	5,6	0,000
HG2743-HT3926_s_	Gamma-Glutamyltransferase 1 (J04131)			20	29	96	-1,3	3,3	0,017
HG2755-HT2862_at	T-Plastin			58	163	321	2,0	3,8	0,000
HG2797-HT2906_s_	Clathrin Light Polypeptide B Alt. Splice 2			581	253	198	-2,3	-2,9	0,000
HG2850-HT4814_s_	Biliary Glycoprotein Alt. Splice 5 A			814	179	122	-2,4	-3,9	0,000
HG2981-HT3125_s_	Epican Alt. Splice 1			13	58	75	4,3	5,6	0,000
HG3044-HT3742_s_	Fibronectin Alt. Splice 1			440	350	3612	-1,6	5,8	0,202
HG3494-HT3688_at	Nuclear Factor Nf-Il6			459	1345	1739	2,1	3,4	0,000
HG371-HT26388_s_	Insulin-Like Leydig Hormone			784	340	173	-2,2	-4,3	0,000
HG4312-HT4582_s_	Transcription Factor Iiia			268	858	462	3,0	1,8	0,000
HG987-HT987_at	Mac25			531	1059	2749	1,3	3,0	0,002
HG998-HT998_s_at	Sulfotransferase Phenol-Preferring			390	155	128	-2,3	-3,5	0,000
J03040_at	SPARC/osteonectin	Hs.111779	5q31.3	304	809	3768	1,7	10,0	0,001
J03278_at	platelet-derived growth factor (PDGF) receptor	Hs.76144	5q31	74	127	387	1,4	3,3	0,002
J03764_at	plasminogen activator inhibitor-1 gene	Hs.82085	7q21.3	25	61	121	2,2	4,5	0,000
J03910_rna1_at	(clone 14VS) metallothionein-IG (MT1G) gene	Hs.173451	16q13	2252	409	148	-6,9	-12,3	0,000
J03915_s_at	chromogranin A	Hs.172216	14q32	448	62	41	-5,0	-6,1	0,000

J04040_at	glucagon	Hs.1460	2q36	314	27	11	-8,7	-15,5	0,000
J04093_s_at	phenol UDP-glucuronosyltransferase (UDPGT)	Hs.284239	2q37	204	41	35	-4,0	-5,2	0,000
J04152_rna1_s_at	M1S1	Hs.23582	1p32	6	66	144	6,9	19,0	0,000
J04164_at	interferon-inducible protein 9-27	Hs.146360	-	963	3522	2269	3,8	2,3	0,000
J04177_at	alpha-1 type XI collagen (COL11A1)	Hs.82772	1p21	26	74	324	2,4	8,6	0,000
J04456_at	14 kd lectin	Hs.227751	22q13.1	374	422	2063	-1,6	3,1	0,096
J04469_at	CKMT mitochondrial creatine kinase	Hs.153998	15q15	1002	281	159	-2,8	-5,4	0,000
J04970_at	carboxypeptidase M 3' end	Hs.169765	12q15	60	21	21	-3,9	-2,8	0,000
J05257_at	MDP4 MDP7 microsomal dipeptidase	Hs.109	16q24.3	26	847	388	20,1	7,9	0,000
J05412_at	regenerating protein (reg) gene	Hs.1032	2p12	52	588	95	6,9	1,5	0,001
J05582_s_at	pancreatic mucin	Hs.89603	1q21	899	447	243	-2,2	-3,4	0,000
L00058_at	(GH) germline c-myc proto-oncogene			55	218	135	4,1	2,7	0,000
L02785_at	colon mucosa-associated (DRA)	Hs.1650	7q31	1320	386	36	-2,3	-52,6	0,000
L05144_at	phosphoenolpyruvate carboxykinase (PCK1)	Hs.1872	20q13.31	2705	710	165	-3,1	-10,5	0,000
L05779_at	cytosolic epoxide hydrolase	Hs.113	8p21	315	185	48	-2,3	-5,3	0,000
L07597_at	ribosomal protein S6 kinase 2 (RPS6KA2)	Hs.149957	3	452	235	149	-1,6	-3,0	0,000
L09708_at	complement component 2 (C2) gene allele b			129	495	406	4,5	3,3	0,000
L10373_at	(clone CCG-B7) sequence	Hs.82749	Xq11	310	74	61	-4,6	-5,6	0,000
L10955_cds1_s_at	carbonic anhydrase IV gene			1186	194	30	-7,2	-32,7	0,000
L11708_at	17 beta hydroxysteroid dehydrogenase type 2	Hs.155109	16q24.1	414	41	22	-5,9	-6,4	0,000
L12350_at	thrombospondin 2 (THBS2)	Hs.108623	6q27	22	85	448	2,0	13,2	0,000
L12760_s_at	phosphoenolpyruvate carboxykinase (PCK1)	Hs.1872	20q13.31	755	106	40	-4,2	-11,0	0,000
L13923_at	fibrillin	Hs.750	15q21.1	75	82	356	-1,1	3,2	0,031
L16842_at	ubiquinol cytochrome-c reductase core I protein	Hs.119251	3p21.3	1527	777	494	-1,6	-2,8	0,001
L16895_at	lysyl oxidase (LOX) gene exon 7	Hs.102267	5q23.3	16	18	90	1,7	8,8	0,011
L21998_at	intestinal mucin (MUC2)	Hs.315	11p15.5	4189	1175	243	-4,1	-14,8	0,000
L22524_s_at	matrilysin gene			5	37	570	6,1	83,2	0,000
L25286_s_at	alpha-1 type XV collagen	Hs.83164	9q21-q22	29	52	179	1,2	4,3	0,003
L41351_at	prostasin	Hs.75799	16p11.2	945	714	287	-1,0	-3,4	0,002
L76465_at	PGDH	Hs.77348	4q34-q35	213	61	21	-4,5	-7,7	0,000
M10050_at	liver fatty acid binding protein (FABP)	Hs.5241	2p11	4191	1920	1015	-3,1	-5,6	0,000
M10942_at	metallothionein-Ie gene (hMT-Ie)	Hs.74170	16q13	989	134	302	-5,2	-3,7	0,000
M11718_at	alpha-2 type V collagen gene 3' end	Hs.82985	2q14-q32	60	155	792	1,6	11,1	0,000
M11749_at	Thy-1 glycoprotein gene	Hs.125359	11q22.3	67	183	475	1,6	5,2	0,000
M12759_at	Ig J chain gene			687	117	95	-3,3	-7,9	0,000
M12963_s_at	class I alcohol dehydrogenase (ADH1) alpha	Hs.73843	4q21-q23	1724	260	59	-7,1	-38,8	0,000
M13929_s_at	c-myc-P64			53	194	135	4,6	3,1	0,000
M14758_at	P-glycoprotein (MDR1)	Hs.21330	7q21.1	162	37	33	-4,1	-6,1	0,000
M16364_s_at	creatine kinase-B	Hs.173724	14q32	2300	501	314	-4,8	-13,2	0,000
M16801_at	mineralocorticoid receptor (hMR)	Hs.1790	4q31.1	161	32	13	-4,0	-10,3	0,000

M18079_at	intestinal fatty acid binding protein	Hs.282265	4q28-q31	184	31	20	-4,9	-10,2	0,001
M22324_at	Aminopeptidase N/CD13	Hs.1239	15q25-	565	117	198	-3,6	-2,9	0,000
M22430_at	RASF-A PLA2	Hs.76422	1p35	2550	573	89	-3,4	-16,8	0,001
M22489_at	bone morphogenetic protein 2A (BMP-2A)	Hs.73853	20p12	111	22	27	-3,7	-4,0	0,000
M23178_s_at	MIP1/SCI	Hs.73817	17q11-	30	45	183	1,4	7,1	0,003
M25629_at	kallikrein clone clone phKK25	Hs.123107	19q13.3	437	176	47	-2,5	-7,3	0,000
M26576_cds2_at	alpha-1 collagen type IV gene exon 52.			105	332	764	2,0	5,0	0,000
M28130_rna1_s_at	interleukin 8 (IL8) gene	Hs.624	4q13-q21	23	104	271	3,0	11,4	0,001
M29877_at	alpha-L-fucosidase	Hs.576	1p34	1693	623	315	-2,3	-3,6	0,000
M32886_at	sorcin CP-22	Hs.117816	7q21.1	1121	542	325	-2,6	-2,8	0,000
M55593_at	collagenase type IV (CLG4) gene			236	260	856	-1,2	4,5	0,020
M60047_at	heparin binding protein (HBp17)	Hs.1690	4	143	33	28	-4,4	-4,8	0,000
M62505_at	C5a anaphylatoxin receptor	Hs.2161	19q13.3	38	46	162	1,1	3,8	0,023
M63835_at	IgG Fc receptor I gene			7	16	84	1,9	12,2	0,002
M68840_at	monoamine oxidase A (MAOA)	Hs.183109	Xp11.4-	445	72	40	-2,7	-4,7	0,000
M69013_at	guanine nucleotide-binding regulatory protein	Hs.1686	19p13.3	749	486	344	-1,8	-3,2	0,000
M69181_at	nonmuscle myosin heavy chain-B (MYH10)			59	63	339	1,1	3,3	0,015
M69203_s_at	cytokine (SCYA2) gene			37	55	145	1,2	5,4	0,007
M76424_at	carbonic anhydrase VII (CA VII) gene			423	115	90	-2,9	-4,8	0,000
M77349_at	transming growth factor-beta induced (BIGH3)	Hs.118787	5q31	307	1735	2618	5,3	8,0	0,000
M80244_at	E16	Hs.184601	16q24.3	63	325	246	3,7	2,6	0,000
M82962_at	PPH alpha	Hs.179704	6p12	726	390	72	-1,7	-8,7	0,000
M83216_s_at	aorta caldesmon	Hs.286238	7q33	43	106	466	-1,1	3,4	0,013
M87860_at	S-lac lectin L-14-II (LGALS2) gene			232	40	12	-5,3	-7,4	0,000
M93221_at	macrophage mannose receptor (MRC1) gene			36	33	110	-1,3	3,3	0,149
M97496_at	guanylin	Hs.778	1	1931	147	62	-15,8	-44,5	0,000
S78187_at	CDC25Hu2=cdc25+ homolog [3118 nt]	Hs.153752	20p13	129	612	304	4,2	1,6	0,000
U04636_rna1_at	cyclooxygenase-2 (hCox-2) gene	Hs.196384	1q25.2	18	52	170	1,4	6,3	0,009
U05861_at	hepatic dihydrodiol dehydrogenase gene			173	54	22	-3,8	-6,8	0,006
U06863_at	follistatin-related protein precursor	Hs.285717	7q21.2	99	160	425	1,3	4,0	0,001
U07563_cds1_at	proto-oncogene tyrosine-protein kinase (ABL)			22	89	67	3,3	3,7	0,000
U08021_at	nicotinamide N-methyltransferase (NNMT)	Hs.76669	11q23.1	112	201	799	1,2	4,8	0,001
U09278_at	fibroblast activation protein	Hs.418	2q23	2	19	124	3,5	29,9	0,000
U10550_at	Gem GTPase (gem)	Hs.79022	8q13	52	119	281	1,2	3,0	0,005
U11862_s_at	HP-DAO1 diamine oxidase copper	Hs.75741	7q34	1109	564	211	-1,7	-3,6	0,000
U13616_at	ankyrin G (ANK-3)	Hs.75893	10q21	70	40	24	-1,5	-3,9	0,006
U14528_at	sulfate transporter (DTD)	Hs.29981	5q31	620	137	23	-4,3	-20,1	0,000
U16306_at	chondroitin sulfate proteoglycan	Hs.81800	5q14.3	119	190	999	1,1	8,4	0,002
U16660_at	ECH1 Delta3,5-Delta2,4-Dienoyl-Coa-Isomerase	Hs.196176	19q13.1	1914	897	535	-1,8	-3,0	0,000
U17077_at	BENE partial cds	Hs.185055	2q13	1499	214	207	-4,3	-6,2	0,000

U18018_at	E1A enhancer binding protein (E1A-F)	Hs.77711	17q21	54	350	226	9,4	3,7	0,000
U20758_rna1_at	osteopontin	Hs.313	4q21	11	46	463	2,2	15,1	0,000
U21128_at	lumican	Hs.79914	12q21.3	126	273	645	1,4	4,7	0,001
U26726_at	11-beta-hydroxysteroid dehydrogenase type 2	Hs.1376	16q22	2295	634	119	-3,5	-16,7	0,000
U28249_at	11kd protein	Hs.92323	-	207	63	53	-2,4	-3,8	0,000
U29091_at	selenium-binding protein (hSBP)	Hs.7833	1q21-q22	1031	456	165	-1,4	-4,4	0,001
U29680_at	A1 protein	Hs.227817	15q24.3	28	40	83	1,5	4,8	0,002
U30521_at	P311 HUM -3.1	Hs.142827	-	87	100	337	1,0	3,3	0,013
U39840_at	HNF-3 alpha	Hs.105440	14q12	130	40	32	-3,6	-3,8	0,001
U51095_at	homeobox protein Cdx1	Hs.1545	5q31-q33	643	329	204	-1,6	-3,0	0,002
U53445_at	Doc1	Hs.15432	3	67	101	276	1,1	2,9	0,020
U53786_at	envoplakin (EVPL)	Hs.25482	17q25	60	174	101	2,9	1,4	0,001
U63824_at	transcription factor RTEF-1 (RTEF1)	Hs.94865	12p13.2	32	99	71	2,9	2,2	0,000
U65932_at	extracellular matrix protein 1 (ECM1)	Hs.81071	1q21	55	85	293	1,3	3,3	0,001
U66661_at	GABA-A receptor epsilon subunit	Hs.22785	Xq28	32	100	69	3,1	1,8	0,001
U70426_at	A28-RGS14p	Hs.183601	1q25-q31	50	78	174	1,3	2,9	0,001
U70663_at	zinc finger transcription factor hEZF (EZF)	Hs.7934	9q31	605	76	44	-6,2	-8,8	0,000
U70732_rna1_at	glutamate pyruvate transaminase (GPT) gene .	Hs.103502	8q24.3	373	164	62	-2,2	-7,8	0,000
U73379_at	cyclin-selective ubiquitin carrier protein	Hs.93002	20	267	827	526	3,0	1,5	0,003
U77643_at	K12 protein precursor	Hs.95655	17q25	422	160	148	-3,0	-3,4	0,000
U78551_at	gallbladder mucin MUC5B partial cds	Hs.102482	11p15	465	75	26	-3,8	-5,5	0,001
U79725_at	A33 antigen precursor	Hs.143131	-	2855	998	327	-2,3	-7,0	0,000
U81599_at	homeodomain protein HOXB13 .	Hs.66731	17q21.2	240	94	48	-1,9	-3,4	0,000
U83246_at	copine I	Hs.166887	-	325	1109	549	3,2	1,7	0,000
U89942_at	lysyl oxidase-related protein (WS9-14)	Hs.83354	8p21.3	37	79	280	2,1	7,1	0,000
X02419_rna1_s_at	uPA gene	Hs.77274	10q24	55	136	289	2,6	6,0	0,000
X02761_s_at	fibronectin (FN precursor)	Hs.118162	2q34	681	620	3340	-1,5	4,0	0,108
X04602_s_at	interleukin BSF-2 (B-cell differentiation factor)	Hs.93913	7p21	5	31	60	2,4	6,1	0,002
X06700_s_at	3' region pro-alpha1(III) collagen	Hs.119571	2q31	241	367	1383	1,2	5,1	0,001
X14253_s_at	teratocarcinoma-derived growth factor 1	Hs.75561	3p21.31	16	156	57	7,3	3,0	0,000
X16354_at	transmembrane carcinoembryonic antigen BGPa	Hs.50964	19q13.2	1871	458	219	-2,6	-4,6	0,000
X52001_at	endothelin 3	Hs.1408	20q13.2	100	34	6	-4,6	-16,1	0,000
X52022_at	RNA type VI collagen alpha3 chain	Hs.80988	2q37	352	563	2144	1,0	4,1	0,004
X53800_s_at	MIP2beta	Hs.89690	4q21	21	73	75	5,3	5,0	0,000
X54489_rna1_at	melanoma growth stimulatory activity (MGSA)	Hs.789	4q21	56	262	484	7,8	8,4	0,000
X54925_at	type I interstitial collagenase	Hs.83169	11q22.3	12	129	945	7,3	31,8	0,000
X57579_s_at	activin beta-A subunit (exon 2)	Hs.727	7p15-p13	7	60	433	5,1	26,7	0,000
X59766_at	Zn-alpha2-glycoprotein	Hs.71	7q22.1	15	333	114	18,8	4,4	0,000
X59770_at	IL-1R2 type II interleukin-1 receptor	Hs.25333	2q12-q22	427	48	65	-6,4	-4,5	0,000
X59871_at	TCF-1 T cell factor 1 (splice m C)	Hs.169294	5q31.1	7	36	25	5,9	3,6	0,000

X63187_at	HE4 extracellular proteinase inhibitor homologue	Hs.2719	20q12	714	124	96	-3,0	-3,5	0,000
X63597_at	si sucrase-isomaltase	Hs.2996	3q25.2	37	12	3	-8,0	-14,1	0,000
X63629_at	p cadherin	Hs.2877	16q22	14	214	134	11,2	6,6	0,000
X64177_f_at	metallothionein	Hs.2667	16q13	1019	313	254	-4,3	-3,5	0,000
X73501_at	gene cytokeratin 20	Hs.84905	17	1553	185	107	-5,5	-12,0	0,000
X74570_at	sialyltransferase	Hs.75268	11q23	486	92	88	-3,2	-4,1	0,000
X74929_s_at	KRT8 keratin 8	Hs.242463	12q13	3547	2083	1642	-1,8	-3,1	0,000
X77777_s_at	intestinal VIP receptor related protein	Hs.198726	3p22	138	46	16	-3,2	-5,1	0,000
X82153_at	cathepsin O	Hs.83942	1q21	64	111	490	1,3	7,2	0,000
X83618_at	HMG Coa Synthase	Hs.59889	1p13	1743	909	385	-2,9	-3,6	0,011
X87159_at	beta subunit of epithelial amiloride-sensitive	Hs.37129	16p12.2	393	29	10	-8,3	-22,1	0,000
X91911_s_at	RTVP-1 protein	Hs.64639	12	27	40	80	1,2	2,8	0,002
X93036_at	MAT8 protein	Hs.92323	-	4683	1804	944	-2,3	-3,5	0,000
X95632_s_at	Arg protein tyrosine kinase-binding protein	Hs.256315	2q33	17	54	62	2,2	2,9	0,000
X95677_at	ArgBPIB protein. /gb=X95677			13	32	37	3,1	3,7	0,003
X98311_at	carcinoembryonic antigen CGM2	Hs.74466	19q13.2	3817	462	145	-4,1	-16,3	0,000
Y00318_at	complement control protein factor I			23	48	75	1,9	3,0	0,000
Y00339_s_at	carbonic anhydrase II (EC 4.2.1.1)	Hs.155097	8q22	1096	58	28	-20,8	-27,4	0,000
Y00503_at	carbonic anhydrase II (EC 4.2.1.1)	Hs.182265	17q21	2968	1120	1001	-2,5	-3,5	0,000
Y00787_s_at	MDNCF	Hs.624	4q13-q21	53	614	1794	5,4	15,5	0,002
Y08136_at	ASM-like phosphodiesterase 3a	Hs.42945	6	156	36	28	-3,8	-4,4	0,000
Y09616_at	putative carboxylesterase	Hs.282975	-	2490	1759	579	-1,7	-4,2	0,001
Z11793_at	selenoprotein P	Hs.3314	5q31	535	272	96	-1,4	-2,8	0,001
Z37976_at	LTBP-2	Hs.83337	14q24	28	70	102	1,8	3,1	0,000
Z48482_at	membrane-type matrix metalloproteinase 2	Hs.80343	16q13-	267	142	69	-1,5	-5,8	0,000
Z69881_at	adenosine triphosphatase calcium	Hs.5541	17p13.3	352	135	64	-3,2	-4,9	0,000
Z70295_at	GCAP-II gene	Hs.32966	1p34-p33	453	9	7	-25,0	-39,0	0,000
Z74615_at	prepro-alpha1(I) collagen	Hs.172928	17q21.3	137	543	2000	2,2	11,1	0,000
Z74616_s_at	prepro-alpha2(I) collagen	Hs.179573	7q22.1	201	836	2784	2,6	10,9	0,000
Z80345_rna1_s_at	SCAD gene 5 UTR exon 1 and 2	Hs.127610	12q22	285	121	41	-3,0	-7,5	0,000

[a] Nsig med: Median derived from "Signal" of 10 Normal mucosae of the sigmoid and rectosigmoid

[b] Bsig med: Median derived from "Signal" of 8 Dukes' B adenocarcinomas of the sigmoid and rectosigmoid

[c] Csig med: Median derived from "Signal" of 7 Dukes' C adenocarcinomas of the sigmoid and rectosigmoid

[d] FC: Fold Change, corresponding to "signal ratio" of Nsig med/Bsig med or Nsig med/Csig med was calculated from "signal log ratio".

Comparison D: RCC versus LCC in stage Dukes B and C. Each of the five Dukes B or Dukes C, respectively of the caecum was compared to each of the Dukes B (8 samples) or Dukes C (7 samples), respectively of the sigmoid and rectosigmoid using MAS 5.0, MDB 3.0 and DMT 3.0 yielding 75 comparison in total. **Filter 1:** Affymetrix markers, genes with a "Detection"-call "absent" in $\geq$ 80% (20 of 25) of all arrays as well as 43 datasets with "Detection"-call "absent" in the combination of 3 of 5 Dukes B-right, 3 of 5 Dukes C-right, 6 of 8 Dukes B-left and 5 of 7 Dukes C-left to diminish the number of false positives further. In addition all comparisons with a "Change"-call of "not changed" in more than 60% of Dukes B or Dukes C comparisons were excluded resulting in total in 2761 datasets. **Filter 2:** Genes were included if more than 70% of the Dukes B comparisons (28 of 40 comparisons) or Dukes C comparisons (25 of 35 comparisons) showed a "Change"-call of increased or decreased, resulting in 71 genes in Dukes B, 228 in Dukes C and 22 in Dukes B and C with expressional differences from RCC to LCC. **Filter 3:** Genes were included with an average "signal log ratio" of $\geq$ 1.5 or $\leq$ -1.5 obtained from the each-to-each comparisons resulting in 10 genes in Dukes B, 44 genes in Dukes C and 5 genes in common for B and C differing between RCC and LCC. **Filter 4:** Of the previous filtered genes, 5 genes in Dukes B, 39 in Dukes C and 5 genes in total for B and C showed significant expression differences with $p<0.05$ in a Mann-Whitney U-Test (Supplementary Table 4).

Supplementary Table 4 Forty-four genes showing significant expression differences (p<0.05) between RCC and LCC adenocarcinomas of clinical stage Dukes B and Dukes C, accompanied by average fold changes of >= 2.8 or <= -2.8.

Probe Set ID	Gene name	Symbol	UG Cluster	Cyto Band	Bcae med[a]	Bsig med[b]	avg FC Bcae vs Bsig[c]	p-value	Ccae med[d]	Csig med[e]	avg FC Ccae vs Csig	p-value
					Dukes B				Dukes C			
D00654_at	enteric smooth muscle gamma-actin	ACTG2	Hs.78045	2p13	88	164	-2,7	0,573	87	465	-5,6	0,039
D13643_at	24-dehydrocholesterol reductase	DHCR24	Hs.75616	1p33	453	377	1,4	0,395	503	197	3,2	0,008
D17408_s_at	calponin 1, basic, smooth muscle	CNN1	Hs.21223	19p13.2	47	105	-2,7	0,149	79	281	-4,5	0,008
D21255_at	cadherin 11, type 2, OB-cadherin	CDH11	Hs.75929	16q22.1	20	42	-1,1	0,319	80	231	-3,4	0,024
D78014_at	dihydropyrimidinase-like 3	DPYSL3	Hs.74566	5q32	94	117	-1,3	0,573	144	408	-3,4	0,004
D86479_at	AE-binding protein 1	AEBP1	Hs.118397	7p13	102	142	-1,0	0,673	151	485	-3,3	0,024
D87258_at	protease, serine, 11 (IGF binding)	PRSS11	Hs.75111	10q25.3	162	213	-1,1	0,888	263	843	-3,0	0,014
D90279_s_at	collagen, type V, alpha 1	COL5A1	Hs.146428	9q34.2	3	7	-1,6	0,079	21	158	-4,1	0,048
HG1428-HT1428_s	Modulator Recognition Factor 2	MRF2	Hs.355963	10	775	1747	-2,1	0,024	429	1442	-3,1	0,024
HG2614-HT2710_a	Collagen Type Viii Alpha 1	COL8A1	Hs.114599	3q12-q13	38	96	-2,1	0,024	69	244	-3,0	0,039
HG2743-HT2845_a	Caldesmon 1 Alt. Splice 4	CALD1	Hs.325474	7q33	47	95	-2,4	0,037	109	354	-3,6	0,001
HG2743-HT2846_s	Caldesmon 1 Alt. Splice 6	CALD1	Hs.325474	7q33	42	88	-2,2	0,037	57	306	-4,4	0,001
HG2743-HT3926_s	GGT1 Gamma-Glutamyltransferase 1	GGT1	Hs.284380	22q11.1	10	29	-2,9	0,055	30	96	-4,4	0,001
J02854_at	20-kDa myosin light chain (MLC-2)	MYRL2	Hs.9615	20pter	73	179	-2,6	0,037	167	639	-3,8	0,008
J04080_at	complement component 1	C1S	Hs.169756	12p13	747	824	-1,2	0,779	958	3056	-2,9	0,008
K02765_at	complement component 3	C3	Hs.284394	19p13.3	171	192	-1,6	0,888	335	1261	-3,8	0,014
L38486_at	microfibrillar-associated protein 4	MFAP4	Hs.118223	17p11.2	49	65	-1,4	0,395	39	112	-3,0	0,004
M11749_at	Thy-1 cell surface antigen	THY1	Hs.125359	11q22.3	136	183	-1,3	0,573	124	475	-3,2	0,008
M12174_at	ras homolog gene family, member B	ARHB	Hs.204354	2pter-p12	79	192	-1,9	0,009	68	239	-2,9	0,001
M21305_at	alpha satellite and satellite 3 junction DNA		Hs.247946		13	259	-11,9	0,037	315	336	1,2	0,816
M26679_at	homeo box A5	HOXA5	Hs.37034	7p15-p14	30	17	1,8	0,253	59	15	3,4	0,004
M33493_s_at	tryptase beta 1	TPSB1	Hs.250700	16p13.3	125	121	-1,1	1,000	80	276	-2,9	0,008
M58459_at	ribosomal protein S4, Y-linked	RPS4Y	Hs.180911	Yp11.3	6	157	-8,8	0,079	9	560	-23,3	0,039
M62402_at	insulin-like growth factor binding protein 6	IGFBP6	Hs.274313	12q13	29	28	-1,1	0,888	22	88	-3,5	0,001
M73720_at	mast cell carboxypeptidase A	MC-CPA	Hs.646	3q21-q25	74	95	-1,3	0,479	63	194	-2,8	0,014
M83216_s_at	caldesmon 1	CALD1	Hs.286238	7q33	27	106	-2,8	0,037	84	466	-4,2	0,001
M84526_at	D component of complement (adipsin)	DF	Hs.155597	19p13.3	249	74	2,0	0,253	39	112	-5,4	0,001
M95787_at	transgelin	TAGLN	Hs.75777	11q23.2	360	663	-2,0	0,110	542	3600	-5,3	0,004
M97252_at	Kallmann syndrome 1 sequence	KAL1	Hs.89591	Xp22.32	4	17	-2,7	0,110	16	74	-3,8	0,014
U06863_at	follistatin-like 1	FSTL1	Hs.285717	3q13.33	72	160	-1,6	0,055	138	425	-2,8	0,008
U16306_at	chondroitin sulfate proteoglycan 2 (versican)	CSPG2	Hs.81800	5q14.3	200	190	1,1	0,779	425	999	-3,3	0,024
U28368_at	inhibitor of DNA binding 4	ID4	Hs.34853	6p22-p21	4	16	-2,0	0,395	4	38	-5,4	0,008
U29091_at	selenium binding protein 1	SELENBF	Hs.7833	1q21-q22	175	456	-3,1	0,015	71	165	-1,7	0,128
U31382_at	guanine nucleotide binding protein 4	GNG4	Hs.32976	1q42.3	22	91	-3,2	0,009	49	55	-1,1	0,938
U35139_at	necdin (mouse) homolog	NDN	Hs.50130	15q11.2	23	33	-1,6	0,253	9	58	-5,7	0,008

U40490_at	nicotinamide nucleotide transhydrogenase	NNT	Hs.18136	5p13.1	8	35	-3,3	0,037	30	27	1,1	0,388
U48959_at	myosin, light polypeptide kinase	MYLK	Hs.211582	3q21	83	168	-2,3	0,253	146	618	-5,8	0,001
U52191_s_at	SMC (mouse) homolog, Y chromosome	SMCY	Hs.80358	Yq11	2	15	-4,6	0,055	2	48	-12,2	0,008
U60115_at	four and a half LIM domains 1	FHL1	Hs.239069	Xq26	96	114	-1,5	0,673	93	183	-3,3	0,014
U77594_at	retinoic acid receptor responder (TIG2)	RARRES2	Hs.37682	7q36.1	33	63	-1,5	0,197	74	173	-3,7	0,004
X14253_s_at	teratocarcinoma-derived growth factor 1	TDGF1	Hs.75561	3p21.31	16	156	-5,3	0,015	53	57	1,7	0,816
X51405_at	carboxypeptidase E (EC 3.4.17.10)	CPE	Hs.75360	4q32.3	10	22	-2,5	0,030	13	46	-4,1	0,008
X53331_at	matrix Gla protein	MGP	Hs.75742	12p13.1	95	218	-1,8	0,253	264	595	-3,3	0,024
X53416_at	filamin A, alpha (actin-binding protein-280)	FLNA	Hs.195464	Xq28	333	224	-1,1	0,673	236	787	-3,1	0,008

[a] Bcae med: Median derived from "Signal" of 5 Dukes' B adenocarcinomas of the caecum

[b] Bsig med: Median derived from "Signal" of 8 Dukes' B adenocarcinomas of the sigmoid and rectosigmoid

[c] FC: Fold Change, corresponding to "signal ratio" was calculated from "signal log ratio".

[d] Ccae med: Median derived from "Signal" of 5 Dukes' C adenocarcinomas of the caecum

[e] Csig med: Median derived from "Signal" of 7 Dukes' C adenocarcinomas of the sigmoid and rectosigmoid

Comparison E: Common expression differences from normal mucosa to adenocarcinoma in RCC and LCC. 118 genes of Dukes B or C of RCC obtained by comparison 2 were compared to 186 genes of LCC from comparison 3. Additionally we identified cancer genes being characteristic for one side of the colon only. 88 genes shown in supplementary table 5 were significantly differential expressed exclusively in the right-sided tumours and suggesting a more crucial role in caecal adenocarcinomas. 156 genes shown in supplementary table 6 were significantly differential expressed only in left sided.

Supplementary Table 5 Eighty-eight genes showing significant different expression in the right-sided colon exclusively (p<0.05).

Probe Set ID	Gene name	UG Cluster	Cyto Band	Ncae med[a]	Bcae med[b]	avg FC[c]	p-value [d]	Ccae med [e]	avg FC	p-value
					Dukes B			Dukes C		
AB002409_at	SLC Secondary Lymphoid-Tissue Chemokine	Hs.57907	9p13	187	34	-4,8	0,012	44	-2,6	0,016
AF001294_at	IPL	Hs.154036	11p15.5	178	557	3,7	0,044	402	2,1	0,095
AF001548_rna1_at	chromosome 16 BAC clone CIT987SK-815A9	Hs.78344	16p13.13	1165	99	-8,0	0,004	261	-5,8	0,095
D00654_at	enteric smooth muscle gamma-actin gene 5'			420	88	-6,0	0,024	87	-3,6	0,421
D10667_s_at	smooth muscle myosin heavy chain			77	18	-10,2	0,004	18	-11,2	0,095
D13168_at	gene endothelin-B receptor (hET-BR)			11	6	-2,0	0,123	4	-3,4	0,032
D17408_s_at	calponin	Hs.21223	19p13.2	257	47	-5,2	0,012	79	-2,9	0,222
D83777_at	KIAA0193 gene	Hs.75137	7	71	70	-1,5	0,187	191	3,2	0,016
D86062_s_at	KNP-Ib	Hs.182423	21q22.3	64	45	-1,0	0,766	20	-3,9	0,008
HG2197-HT2267_s	Collage Type Vii Alpha 1			65	118	-1,0	0,921	182	2,9	0,032
HG2981-HT3127_s	Epican Alt. Splice 11			33	537	3,5	0,187	165	4,9	0,016
HG2981-HT3938_s	Epican Alt. Splice 12			5	219	5,4	0,123	75	9,9	0,016
HG3431-HT3616_s	Decorin Alt. Splice 1			295	78	-3,8	0,044	350	1,2	0,421
HG4263-HT4533_a	Nkr-P1a Protein			43	3	-3,0	0,266	3	-8,7	0,016
HG4582-HT4987_a	Beta-1-Glycoprotein 1 M20882			44	12	-3,5	0,024	14	-1,8	0,095
J00306_at	somatostatin I gene and flanks	Hs.12409	3q28	73	8	-4,3	0,266	7	-11,8	0,032
J02854_at	20-kDa myosin light chain (MLC-2)	Hs.9615	-	472	73	-4,1	0,024	167	-1,7	0,841
J02874_at	adipocyte lipid-binding protein	Hs.83213	8q21	126	13	-9,1	0,024	9	-6,9	0,056
J03507_at	complement protein component C7	Hs.78065	5p13	88	10	-6,4	0,004	11	-4,6	0,056
J05096_rna1_at	NaK-ATPase subunit alpha 2 (ATP1A2)	Hs.34114	1q21-q23	15	4	-5,5	0,004	9	-1,2	0,841
K02765_at	complement component C3 alpha and beta	Hs.284394	19p13.3	472	171	-6,7	0,024	335	-1,2	0,841
L07615_at	neuropeptide Y receptor Y1 (NPYY1)			23	12	-1,8	0,187	4	-5,8	0,032
L08895_at	MADS/MEF2-family transcription factor (MEF2C)	Hs.78995	5q14	29	9	-3,5	0,004	16	-2,1	0,222
L29008_at	L-iditol-2 dehydrogenase	Hs.878	15q15.3	145	525	1,5	0,484	491	3,1	0,032
L29433_at	factor X (blood coagulation factor) gene			57	24	-3,2	0,044	16	-2,1	0,151
L38486_at	microfibril-associated glycoprotein 4 (MFAP4)	Hs.118223	17p11.2	206	49	-3,2	0,012	39	-3,3	0,095
L43821_at	enhancer of filamentation (HEF1)	Hs.80261	-	72	28	-1,7	0,266	30	-3,4	0,032
M14539_at	factor XIII subunit a 3' end	Hs.80424	6p25.3	159	42	-4,6	0,004	52	-3,1	0,151
M21005_at	migration inhibitory factor-related protein (MRP8)	Hs.100000	1q21	54	73	-1,3	0,619	190	3,7	0,032
M21305_at	alpha satellite and satellite 3 junction DNA	Hs.247946	-	92	13	-6,1	0,012	315	2,8	0,095
M25753_at	cyclin B 3' end	Hs.23960	5q12	42	106	3,2	0,012	167	2,7	0,095
M26311_s_at	cystic fibrosis antigen	Hs.112405	1q21	51	106	1,0	0,619	360	11,7	0,032
M31994_at	cytosolic aldehyde dehydrogenase (ALDH1) gene			353	205	-1,8	0,484	84	-3,4	0,008
M32053_at	H19 RNA gene (spliced in silico)			68	71	-1,6	0,365	161	3,4	0,032
M54927_at	myelin proteolipid protein	Hs.1787	Xq22	16	5	-5,6	0,012	3	-4,7	0,095

M60828_at	keratinocyte growth factor	Hs.164568	15q15	10	5	-4,3	0,044	10	2,1	0,222
M63379_at	TRPM-2 protein gene			918	285	-5,1	0,004	255	-4,0	0,008
M63603_at	phospholamban	Hs.85050	6q22.1	29	7	-7,1	0,004	20	-1,4	0,841
M80482_at	subtilisin-like protein (PACE4)	Hs.170414	15q26	269	149	-1,3	0,484	60	-3,2	0,032
M84526_at	adipsin/complement factor D	Hs.155597	19	872	249	-5,0	0,075	39	-21,1	0,016
M95787_at	22kDa smooth muscle protein (SM22)	Hs.75777	11q23.2	1383	360	-3,9	0,024	542	-2,0	0,222
M97252_at	Kallmann syndrome (KAL)	Hs.89591	Xp22.32	3	4	-1,4	0,619	16	5,0	0,016
S45630_at	alpha B-crystallin	Hs.1940	11q22.3	220	88	-3,6	0,044	60	-2,3	0,056
S67156_at	ASP=aspartoacylase [kidney 1435 nt]	Hs.32042	17pter	25	7	-5,0	0,004	2	-4,4	0,056
S75256_s_at	HNL=neutrophil lipocalin			345	3135	4,4	0,024	1831	1,6	0,548
U03688_at	dioxin-inducible cytochrome P450 (CYP1B1)	Hs.154654	2p21	8	3	-4,0	0,044	8	1,8	0,421
U10485_at	lymphoid-restricted membrane protein (Jaw1)	Hs.40202	-	51	3	-9,9	0,044	9	-6,7	0,032
U19495_s_at	intercrine-alpha (hIRH)	Hs.237356	10q11.1	186	6	-9,4	0,075	14	-5,2	0,016
U23143_at	mitochondrial serine hydroxymethyltransferase	Hs.75069	12q12	126	301	2,8	0,004	160	1,4	0,095
U24488_s_at	tenascin-X (XA)	Hs.283750	6p21.3	205	8	-20,3	0,012	8	-13,3	0,095
U25997_at	stanniocalcin precursor (STC)	Hs.25590	8p21	2	29	4,5	0,044	48	30,5	0,016
U28368_at	Id-related helix-loop-helix protein Id4	Hs.34853	6p22	26	4	-5,8	0,044	4	-2,9	0,151
U37283_at	microfibril-associated glycoprotein-2 MAGP-2	Hs.58882	12p13.1	53	17	-6,2	0,004	7	-2,5	0,421
U39447_at	placenta copper monamine oxidase	Hs.198241	17q21	40	14	-3,4	0,004	11	-1,7	0,310
U41518_at	channel-like integral membrane protein (AQP-1)	Hs.74602	7p14	157	76	-2,9	0,044	87	-1,2	0,548
U43328_at	link protein	Hs.2799	5q14.3	16	4	-1,3	0,484	1	-16,6	0,016
U48959_at	myosin light chain kinase (MLCK)	Hs.211582	3cen	386	83	-4,0	0,012	146	-3,0	0,056
U54617_at	pyruvate dehydrogenase kinase isom 4	Hs.8364	7q21.3	21	7	-3,0	0,024	7	-3,1	0,310
U60115_at	skeletal muscle LIM-protein SLIM1	Hs.239069	Xq26	598	96	-6,1	0,024	93	-5,6	0,095
U61374_at	novel protein	Hs.15154	Xp21.1	135	23	-4,6	0,044	12	-4,6	0,032
U71207_at	eyes absent homolog (Eab1) .	Hs.29279	20q13.1	44	20	1,1	0,619	6	-6,4	0,008
U77180_at	macrophage inflammatory protein 3 (MIP-3beta)	Hs.50002	9p13	104	3	-10,4	0,075	3	-13,6	0,008
U81607_at	gravin	Hs.788	6q24-q25	64	24	-3,3	0,044	55	1,8	0,032
X00371_rna1_at	myoglobin gene (exon 1)	Hs.118836	22q13.1	213	44	-7,1	0,004	65	-3,4	0,095
X03350_at	(ADH1-2 allele)	Hs.4	4q21-q23	85	12	-9,0	0,004	5	-17,2	0,056
X05232_at	stromelysin	Hs.83326	11q22.3	12	165	15,5	0,004	243	33,6	0,056
X06562_at	growth hormone receptor	Hs.125180	5p13-p12	15	1	-6,4	0,044	2	-4,0	0,421
X07820_at	metalloproteinase stromelysin-2	Hs.2258	11q22.3	4	28	10,1	0,004	20	7,1	0,151
X13839_at	vascular smooth muscle alpha-actin	Hs.195851	10q23.3	2158	730	-2,9	0,012	1279	-1,1	0,056
X51405_at	carboxypeptidase E (EC 3.4.17.10)	Hs.75360	4	22	10	-3,4	0,012	13	-1,8	0,222
X52003_at	pS2 protein gene	Hs.1406	21q22.3	313	903	4,7	0,044	2118	1,7	0,222
X53331_at	matrix Gla protein	Hs.75742	12p13.1	698	95	-5,6	0,024	264	-1,9	0,421
X54162_at	a 64 Kd autoantigen	Hs.79386	1	126	21	-5,3	0,004	36	-4,5	0,095
X57025_at	IGF-I insulin-like growth factor I	Hs.85112	12q22	28	12	-3,4	0,044	10	-2,6	0,056
X64559_at	tetranectin	Hs.65424	3p22	262	10	-17,0	0,012	9	-19,2	0,095

X65614_at	calcium-binding protein S100P	Hs.2962	4p16	170	1392	7,4	0,004	2087	6,1	0,056
X66945_at	N-sam fibroblast growth factor receptor	Hs.748	8p11.2	179	81	-3,4	0,044	109	-1,2	0,421
X74837_at	HUMM9	Hs.25253	6q22	44	21	-1,7	0,365	21	-2,9	0,016
X82209_at	MN1	Hs.268515	22q12.1	21	3	-6,5	0,044	14	-1,1	0,690
X83490_s_at	Fas/Apo-1			27	15	-1,4	0,187	13	-2,9	0,008
X86693_at	hevin like protein	Hs.75445	-	230	80	-3,3	0,012	132	-1,5	0,310
X99133_at	NGAL gene	Hs.204238	9q34	309	1326	3,1	0,024	1047	2,0	0,841
Y07707_at	ITBA4 gene.	Hs.119018	-	17	41	1,9	0,024	41	2,9	0,016
Y09836_at	3'UTR of unknown protein	Hs.82503	-	82	28	-4,4	0,024	49	1,3	0,841
Y10032_at	putative serine/threonine protein kinase	Hs.159640	6q23	210	64	-2,8	0,123	54	-3,0	0,016
Z22865_at	dermatopontin	Hs.80552	1q12	74	17	-4,5	0,024	23	-2,7	0,222
Z29574_at	gene BCMA peptide	Hs.2556	16p13.1	88	30	-2,1	0,266	14	-4,2	0,032
Z49269_at	gene chemokine HCC-1	Hs.20144	17q11.2	80	22	-4,7	0,044	32	-3,5	0,151

[a] Ncae med: Median derived from "Signal" of 10 Normal mucosae of the caecum

[b] Bcae med: Median derived from "Signal" of 5 Dukes' B adenocarcinomas of the caecum

[c] avg FC NvsB: Fold Change, corresponding to "signal ratio" of Ncae med/Bcae med was calculated from "signal log ratio".

[d] p-value NvsB : probability that a variant would assume a value greater than or equal to the observed value strictly by chance.

[e] Ccae med: Median derived from "Signal" of 5 Dukes' C adenocarcinomas of the caecum

Supplementary Table 6 Onehundred-fifty-six genes showing significant different expression in the left-sided colon exclusively (p<0.05).

Probe Set ID	Gene name	UG Cluster	Cyto Band	Nsig med [a]	Bsig med [b]	avg FC [c]	p-value [d]	Csig med [e]	avg FC	p-value
					Dukes B			Dukes C		
AB006781_s_at	galectin-4	Hs.5302	-	4645	2577	-2,2	0,001	1540	-3,5	0,001
D00408_s_at	fetal liver cytochrome P-450 (P-450 HFLa)	Hs.172323	7	647	502	-1,3	0,328	196	-2,9	0,008
D11151_at	DNA endothelin-A receptor 5' flanking region			17	23	1,3	0,076	95	4,8	0,001
D13666_s_at	osteoblast specific factor 2 (OSF-2os)	Hs.136348	13	133	180	-1,4	0,183	1531	8,2	0,001
D16294_at	mitochondrial 3-oxoacyl-CoA thiolase	Hs.32500	18	917	447	-1,9	0,002	260	-2,9	0,001
D17793_at	KIAA0119 gene	Hs.78183	10p15	354	316	-1,0	0,477	103	-2,9	0,001
D21255_at	OB-cadherin-2	Hs.75929	16q22.1	17	42	2,0	0,008	231	16,9	0,001
D30037_at	phosphatidylinositol transfer protein (PI-TPbeta)	Hs.7370	22q12.1	28	74	3,0	0,001	83	3,6	0,001
D42047_at	KIAA0089 gene partial cds	Hs.82432	3	422	238	-1,9	0,008	163	-3,2	0,002
D45917_s_at	TIMP-3 partial cds (C-terminus region)	Hs.245188	22q12.3	41	65	1,2	0,131	175	2,9	0,001
D83174_s_at	collagen binding protein 2	Hs.9930	11q13.5	268	424	1,1	0,021	909	3,0	0,001
D84239_at	IgG Fc binding protein	Hs.111732	-	2819	233	-21,0	0,001	171	-23,4	0,001
D86479_at	AEBP1 gene	Hs.118397	7p13	95	142	-1,0	0,374	485	5,1	0,001
D86956_at	KIAA0201 gene	Hs.36927	-	133	422	3,0	0,000	373	3,1	0,002
D87258_at	cancellous bone osteoblast serin protease	Hs.75111	10q25.3	130	213	-1,0	0,183	843	4,9	0,001
D87449_at	KIAA0260 gene partial cds	Hs.82635	1	254	108	-2,2	0,002	73	-2,8	0,001
D90042_at	liver arylamine N-acetyltransferase (EC 2.3.1.5)	Hs.2	8p22	88	53	-1,6	0,003	27	-3,6	0,001
HG2348-HT2444_s_a	Peptide Yy			575	133	-3,4	0,000	52	-8,4	0,001
HG2743-HT3926_s_a	Gamma-Glutamyltransferase 1 (Gb:J04131)			20	29	-1,3	0,214	96	3,3	0,005
HG2797-HT2906_s_a	Clathrin Light Polypeptide B Alt. Splice 2			581	253	-2,3	0,001	198	-2,9	0,001
HG2850-HT4814_s_a	Biliary Glycoprotein Alt. Splice 5 A			814	179	-2,4	0,002	122	-3,9	0,001
HG3044-HT3742_s_a	Fibronectin Alt. Splice 1			440	350	-1,6	0,328	3612	5,8	0,001
HG3494-HT3688_at	Nuclear Factor Nf-Il6			459	1345	2,1	0,001	1739	3,4	0,002
HG371-HT26388_s_a	Insulin-Like Leydig Hormone			784	340	-2,2	0,003	173	-4,3	0,001
HG987-HT987_at	Mac25			531	1059	1,3	0,033	2749	3,0	0,001
HG998-HT998_s_at	Sulfotransferase Phenol-Preferring			390	155	-2,3	0,005	128	-3,5	0,001
J03040_at	SPARC/osteonectin	Hs.111779	5q31.3	304	809	1,7	0,041	3768	10,0	0,001
J03278_at	platelet-derived growth factor (PDGF) receptor	Hs.76144	5q31	74	127	1,4	0,041	387	3,3	0,001
J03764_at	plasminogen activator inhibitor-1 gene exons 2 to 9	Hs.82085	7q21.3	25	61	2,2	0,004	121	4,5	0,001
J03915_s_at	chromogranin A	Hs.172216	14q32	448	62	-5,0	0,001	41	-6,1	0,001
J04040_at	glucagon	Hs.1460	2q36	314	27	-8,7	0,002	11	-15,5	0,001
J04093_s_at	phenol UDP-glucuronosyltransferase (UDPGT)	Hs.284239	2q37	204	41	-4,0	0,004	35	-5,2	0,001
J04152_rna1_s_at	M1S1	Hs.23582	1p32-p31	6	66	6,9	0,001	144	19,0	0,001
J04164_at	interferon-inducible protein 9-27	Hs.146360	-	963	3522	3,8	0,000	2269	2,3	0,001
J04177_at	alpha-1 type XI collagen (COL11A1)	Hs.82772	1p21	26	74	2,4	0,001	324	8,6	0,001

J04456_at	14 kd lectin	Hs.227751	22q13.1	374	422	-1,6	0,929	2063	3,1	0,002
J04469_at	CKMT mitochondrial creatine kinase	Hs.153998	15q15	1002	281	-2,8	0,001	159	-5,4	0,001
J05257_at	MDP4 MDP7 microsomal dipeptidase (MDP)	Hs.109	16q24.3	26	847	20,1	0,000	388	7,9	0,001
J05412_at	regenerating protein (reg) gene	Hs.1032	2p12	52	588	6,9	0,000	95	1,5	0,032
J05582_s_at	pancreatic mucin	Hs.89603	1q21	899	447	-2,2	0,001	243	-3,4	0,002
L00058_at	(GH) germline c-myc proto-oncogene 5' flank			55	218	4,1	0,001	135	2,7	0,005
L02785_at	colon mucosa-associated (DRA)	Hs.1650	7q31	1320	386	-2,3	0,001	36	-52,6	0,001
L05144_at	phosphoenolpyruvate carboxykinase (PCK1)	Hs.1872	20q13.31	2705	710	-3,1	0,003	165	-10,5	0,001
L05779_at	cytosolic epoxide hydrolase	Hs.113	8p21	315	185	-2,3	0,001	48	-5,3	0,001
L07597_at	ribosomal protein S6 kinase 2 (RPS6KA2)	Hs.149957	3	452	235	-1,6	0,010	149	-3,0	0,001
L09708_at	complement component 2 (C2) gene allele b			129	495	4,5	0,000	406	3,3	0,001
L10373_at	(clone CCG-B7) sequence	Hs.82749	Xq11	310	74	-4,6	0,001	61	-5,6	0,001
L10955_cds1_s_at	carbonic anhydrase IV gene			1186	194	-7,2	0,001	30	-32,7	0,001
L11708_at	17 beta hydroxysteroid dehydrogenase type 2	Hs.155109	16q24.1	414	41	-5,9	0,001	22	-6,4	0,001
L13923_at	fibrillin	Hs.750	15q21.1	75	82	-1,1	0,657	356	3,2	0,001
L16842_at	ubiquinol cytochrome-c reductase core I protein	Hs.119251	3p21.3	1527	777	-1,6	0,016	494	-2,8	0,001
L16895_at	lysyl oxidase (LOX) gene exon 7	Hs.102267	5q23.3	16	18	1,7	0,328	90	8,8	0,001
L21998_at	intestinal mucin (MUC2)	Hs.315	11p15.5	4189	1175	-4,1	0,006	243	-14,8	0,001
L25286_s_at	alpha-1 type XV collagen	Hs.83164	9q21	29	52	1,2	0,076	179	4,3	0,001
L41351_at	prostasin	Hs.75799	16p11.2	945	714	-1,0	0,051	287	-3,4	0,001
M10050_at	liver fatty acid binding protein (FABP)	Hs.5241	2p11	4191	1920	-3,1	0,000	1015	-5,6	0,002
M11718_at	alpha-2 type V collagen gene 3' end	Hs.82985	2q14	60	155	1,6	0,008	792	11,1	0,001
M11749_at	Thy-1 glycoprotein gene	Hs.125359	11q22.3	67	183	1,6	0,010	475	5,2	0,001
M13929_s_at	c-myc-P64			53	194	4,6	0,000	135	3,1	0,011
M14758_at	P-glycoprotein (MDR1)	Hs.21330	7q21.1	162	37	-4,1	0,001	33	-6,1	0,008
M16364_s_at	creatine kinase-B	Hs.173724	14q32	2300	501	-4,8	0,000	314	-13,2	0,001
M16801_at	mineralocorticoid receptor (hMR)	Hs.1790	4q31.1	161	32	-4,0	0,000	13	-10,3	0,001
M18079_at	intestinal fatty acid binding protein gene	Hs.282265	4q28	184	31	-4,9	0,013	20	-10,2	0,001
M22430_at	RASF-A PLA2	Hs.76422	1p35	2550	573	-3,4	0,026	89	-16,8	0,001
M22489_at	bone morphogenetic protein 2A (BMP-2A)	Hs.73853	20p12	111	22	-3,7	0,001	27	-4,0	0,001
M23178_s_at	MIP1/SCI	Hs.73817	17q11	30	45	1,4	0,076	183	7,1	0,001
M25629_at	kallikrein clone clone phKK25	Hs.123107	19q13.3	437	176	-2,5	0,002	47	-7,3	0,001
M28130_rna1_s_at	interleukin 8 (IL8) gene	Hs.624	4q13	23	104	3,0	0,008	271	11,4	0,002
M29877_at	alpha-L-fucosidase	Hs.576	1p34	1693	623	-2,3	0,003	315	-3,6	0,001
M32886_at	sorcin CP-22	Hs.117816	7q21.1	1121	542	-2,6	0,004	325	-2,8	0,001
M55593_at	collagenase type IV (CLG4) gene			236	260	-1,2	0,534	856	4,5	0,001
M60047_at	heparin binding protein (HBp17)	Hs.1690	4	143	33	-4,4	0,000	28	-4,8	0,001
M62505_at	C5a anaphylatoxin receptor	Hs.2161	19q13.3	38	46	1,1	0,286	162	3,8	0,005
M63835_at	IgG Fc receptor I gene			7	16	1,9	0,051	84	12,2	0,001
M68840_at	monoamine oxidase A (MAOA)	Hs.183109	Xp11.4-p1	445	72	-2,7	0,001	40	-4,7	0,001

M69013_at	guanine nucleotide-binding regulatory protein	Hs.1686	19p13.3	749	486	-1,8	0,006	344	-3,2	0,001
M69181_at	nonmuscle myosin heavy chain-B (MYH10)			59	63	1,1	0,374	339	3,3	0,001
M69203_s_at	cytokine (SCYA2) gene			37	55	1,2	0,183	145	5,4	0,001
M76424_at	carbonic anhydrase VII (CA VII) gene			423	115	-2,9	0,000	90	-4,8	0,001
M80244_at	E16	Hs.184601	16q24.3	63	325	3,7	0,000	246	2,6	0,001
M82962_at	PPH alpha	Hs.179704	6p12-p11	726	390	-1,7	0,002	72	-8,7	0,001
M83216_s_at	aorta caldesmon	Hs.286238	7q33	43	106	-1,1	0,155	466	3,4	0,005
M93221_at	macrophage mannose receptor (MRC1) gene			36	33	-1,3	0,594	110	3,3	0,002
M97496_at	guanylin	Hs.778	1	1931	147	-15,8	0,000	62	-44,5	0,001
U04636_rna1_at	cyclooxygenase-2 (hCox-2) gene	Hs.196384	1q25.2	18	52	1,4	0,131	170	6,3	0,003
U06863_at	follistatin-related protein precursor	Hs.285717	7q21.2	99	160	1,3	0,021	425	4,0	0,001
U07563_cds1_at	proto-oncogene tyrosine-protein kinase (ABL)			22	89	3,3	0,000	67	3,7	0,008
U08021_at	nicotinamide N-methyltransferase (NNMT)	Hs.76669	11q23.1	112	201	1,2	0,026	799	4,8	0,001
U09278_at	fibroblast activation protein	Hs.418	2q23	2	19	3,5	0,006	124	29,9	0,001
U10550_at	Gem GTPase (gem)	Hs.79022	8q13-q21	52	119	1,2	0,051	281	3,0	0,005
U11862_s_at	clone HP-DAO1 diamine oxidase copper	Hs.75741	7q34-qter	1109	564	-1,7	0,004	211	-3,6	0,001
U13616_at	ankyrin G (ANK-3)	Hs.75893	10q21	70	40	-1,5	0,076	24	-3,9	0,003
U16306_at	chondroitin sulfate proteoglycan versican	Hs.81800	5q14.3	119	190	1,1	0,051	999	8,4	0,001
U16660_at	ECH1 Delta3,5-Delta2,4-Dienoyl-Coa-Isomerase	Hs.196176	19q13.1	1914	897	-1,8	0,000	535	-3,0	0,001
U17077_at	BENE partial cds	Hs.185055	2q13	1499	214	-4,3	0,001	207	-6,2	0,001
U21128_at	lumican	Hs.79914	12q21.3	126	273	1,4	0,016	645	4,7	0,001
U26726_at	11-beta-hydroxysteroid dehydrogenase type 2	Hs.1376	16q22	2295	634	-3,5	0,001	119	-16,7	0,001
U28249_at	11kd protein	Hs.92323	-	207	63	-2,4	0,001	53	-3,8	0,001
U29091_at	selenium-binding protein (hSBP) . /gb=U29091	Hs.7833	1q21-q22	1031	456	-1,4	0,041	165	-4,4	0,001
U29680_at	A1 protein	Hs.227817	15q24.3	28	40	1,5	0,056	83	4,8	0,001
U30521_at	P311 HUM -3.1	Hs.142827	-	87	100	1,0	0,374	337	3,3	0,001
U39840_at	hepatocyte nuclear factor-3 alpha (HNF-3 alpha)	Hs.105440	14q12	130	40	-3,6	0,016	32	-3,8	0,001
U51095_at	homeobox protein Cdx1	Hs.1545	5q31-q33	643	329	-1,6	0,062	204	-3,0	0,001
U53445_at	(Doc1)	Hs.15432	3	67	101	1,1	0,286	276	2,9	0,003
U53786_at	envoplakin (EVPL)	Hs.25482	17q25	60	174	2,9	0,000	101	1,4	0,064
U63824_at	transcription factor RTEF-1 (RTEF1)	Hs.94865	12p13.2	32	99	2,9	0,000	71	2,2	0,002
U65932_at	extracellular matrix protein 1 (ECM1)	Hs.81071	1q21	55	85	1,3	0,037	293	3,3	0,001
U66661_at	GABA-A receptor epsilon subunit	Hs.22785	Xq28	32	100	3,1	0,001	69	1,8	0,036
U70426_at	A28-RGS14p	Hs.183601	1q25-q31	50	78	1,3	0,013	174	2,9	0,001
U70732_rna1_at	glutamate pyruvate transaminase (GPT)	Hs.103502	8q24.3	373	164	-2,2	0,001	62	-7,8	0,001
U73379_at	cyclin-selective ubiquitin carrier protein	Hs.93002	20	267	827	3,0	0,000	526	1,5	0,172
U79725_at	A33 antigen precursor	Hs.143131	-	2855	998	-2,3	0,001	327	-7,0	0,001
U81599_at	homeodomain protein HOXB13 .	Hs.66731	17q21.2	240	94	-1,9	0,004	48	-3,4	0,001
U83246_at	copine I	Hs.166887	-	325	1109	3,2	0,000	549	1,7	0,019
U89942_at	lysyl oxidase-related protein (WS9-14)	Hs.83354	8p21.3	37	79	2,1	0,001	280	7,1	0,001

X02419_rna1_s_at	uPA gene	Hs.77274	10q24	55	136	2,6	0,001	289	6,0	0,001
X02761_s_at	fibronectin (FN precursor)	Hs.118162	2q34	681	620	-1,5	0,594	3340	4,0	0,001
X04602_s_at	interleukin BSF-2 (B-cell differentiation factor)	Hs.93913	7p21	5	31	2,4	0,016	60	6,1	0,003
X06700_s_at	3' region pro-alpha1(III) collagen	Hs.119571	2q31	241	367	1,2	0,021	1383	5,1	0,001
X14253_s_at	teratocarcinoma-derived growth factor 1	Hs.75561	3p21.31	16	156	7,3	0,000	57	3,0	0,002
X16354_at	transmembrane carcinoembryonic antigen BGPa	Hs.50964	19q13.2	1871	458	-2,6	0,001	219	-4,6	0,001
X52001_at	endothelin 3	Hs.1408	20q13.2	100	34	-4,6	0,001	6	-16,1	0,001
X52022_at	RNA type VI collagen alpha3 chain	Hs.80988	2q37	352	563	1,0	0,131	2144	4,1	0,001
X53800_s_at	macrophage inflammatory protein-2beta (MIP2beta)	Hs.89690	4q21	21	73	5,3	0,000	75	5,0	0,001
X54489_rna1_at	melanoma growth stimulatory activity (MGSA)	Hs.789	4q21	56	262	7,8	0,000	484	8,4	0,015
X57579_s_at	activin beta-A subunit (exon 2)	Hs.727	7p15-p13	7	60	5,1	0,001	433	26,7	0,001
X59766_at	Zn-alpha2-glycoprotein	Hs.71	7q22.1	15	333	18,8	0,000	114	4,4	0,011
X59871_at	TCF-1 T cell factor 1 (splice m C)	Hs.169294	5q31.1	7	36	5,9	0,000	25	3,6	0,019
X63187_at	HE4 extracellular proteinase inhibitor homologue	Hs.2719	20q12	714	124	-3,0	0,001	96	-3,5	0,001
X63597_at	si sucrase-isomaltase	Hs.2996	3q25.2	37	12	-8,0	0,003	3	-14,1	0,001
X64177_f_at	metallothionein	Hs.2667	16q13	1019	313	-4,3	0,006	254	-3,5	0,001
X73501_at	gene cytokeratin 20	Hs.84905	17	1553	185	-5,5	0,000	107	-12,0	0,001
X74570_at	Gal-beta(1-3/1-4)GlcNAc alpha-2.3-sialyltransferase	Hs.75268	11q23	486	92	-3,2	0,003	88	-4,1	0,001
X74929_s_at	KRT8 keratin 8	Hs.242463	12q13	3547	2083	-1,8	0,003	1642	-3,1	0,002
X77777_s_at	intestinal VIP receptor related protein	Hs.198726	3p22	138	46	-3,2	0,003	16	-5,1	0,001
X82153_at	cathepsin O	Hs.83942	1q21	64	111	1,3	0,013	490	7,2	0,001
X83618_at	HMG Coa Synthase	Hs.59889	1p13-p12	1743	909	-2,9	0,076	385	-3,6	0,011
X87159_at	amiloride-sensitive sodium channel	Hs.37129	16p12.2	393	29	-8,3	0,001	10	-22,1	0,001
X91911_s_at	RTVP-1 protein	Hs.64639	12	27	40	1,2	0,075	80	2,8	0,001
X93036_at	MAT8 protein	Hs.92323	-	4683	1804	-2,3	0,003	944	-3,5	0,001
X95632_s_at	Arg protein tyrosine kinase-binding protein	Hs.256315	2q33	17	54	2,2	0,004	62	2,9	0,001
X95677_at	ArgBPIB protein. /gb=X95677			13	32	3,1	0,013	37	3,7	0,015
X98311_at	carcinoembryonic antigen CGM2	Hs.74466	19q13.2	3817	462	-4,1	0,001	145	-16,3	0,001
Y00318_at	complement control protein factor I			23	48	1,9	0,004	75	3,0	0,001
Y00339_s_at	carbonic anhydrase II (EC 4.2.1.1)	Hs.155097	8q22	1096	58	-20,8	0,000	28	-27,4	0,001
Y00503_at	keratin 19	Hs.182265	17q21	2968	1120	-2,5	0,000	1001	-3,5	0,001
Y00787_s_at	MDNCF	Hs.624	4q13-q21	53	614	5,4	0,021	1794	15,5	0,005
Y08136_at	ASM-like phosphodiesterase 3a	Hs.42945	6	156	36	-3,8	0,001	28	-4,4	0,001
Y09616_at	putative carboxylesterase	Hs.282975	-	2490	1759	-1,7	0,016	579	-4,2	0,001
Z11793_at	selenoprotein P	Hs.3314	5q31	535	272	-1,4	0,021	96	-2,8	0,001
Z37976_at	LTBP-2	Hs.83337	14q24	28	70	1,8	0,003	102	3,1	0,001
Z48482_at	membrane-type matrix metalloproteinase 2	Hs.80343	16q13	267	142	-1,5	0,008	69	-5,8	0,001
Z69881_at	adenosine triphosphatase calcium	Hs.5541	17p13.3	352	135	-3,2	0,000	64	-4,9	0,001
Z70295_at	GCAP-II gene	Hs.32966	1p34-p33	453	9	-25,0	0,000	7	-39,0	0,001
Z74615_at	prepro-alpha1(I) collagen	Hs.172928	17q21.3	137	543	2,2	0,001	2000	11,1	0,001

Z74616_s_at	prepro-alpha2(I) collagen	Hs.179573	7q22.1	201	836	2,6	0,008	2784	10,9	0,001

[a] Nsig med: Median derived from "Signal" of 10 Normal mucosae of the sigmoid and rectosigmoid

[b] Bsig med: Median derived from "Signal" of 8 Dukes' B adenocarcinomas of the sigmoid and rectosigmoid

[c] avg FC NvsB: Fold Change, corresponding to "signal ratio" of Nsig med/Bsig med or Nsig med/Csig med was calculated from "signal log ratio".

[d] p-value NvsB : probability that a variant would assume a value greater than or equal to the observed value strictly by chance.

[e] Csig med: Median derived from "Signal" of 7 Dukes' C adenocarcinomas of the sigmoid and rectosigmoid

Real-time PCR (RT-PCR). cDNA was synthesised from single samples previously analysed on GeneChips. Reverse transcription was performed using Superscript II RT (Invitrogen Corporation, Carlsbad, CA). 1 μg total RNA and 1 μL 50 pmol/μL $(dT)_{24}$-primer in a total volume of 12 μl was incubated 10 min. at 70°C and chilled on ice. After adding 4 μl 1st Strand Buffer (from supplier), 1 μl DTT (0,1M), 2 μl dNTP mix (10mM), and 1 μl SuperScript RT II (200U/μl) the reaction was incubated 1 hour at 42°C, and finally 5 min. at 95°C. The cDNA was diluted 1:20 for use in Real Time PCR. Real-time PCR analysis was performed on selected genes using the primers shown in Supplementary Table 7.

Supplementary Table 7 Primers used for Real-time PCR

Accession no.	Gene name	Forward primer	Reverse Primer	Size (bp)	Tm [a] (°C)		Position [b]
M14539	Factor XIII subunit	GATAGCCAGCATGAGCAGTGAC	TGCCAGGGTTCATCTCAGCT	112	59	59	2208-2319
M22324	Aminopeptidase N	CTGAACCCGGACTTAATCCG	GAAGGAGAACGAGCCACCAC	150	58	58	2665-2814
X65614	S100P Ca-binding protein	TGCTGATGGAGAAGGAGCTACC	GGCATCCTTGTCTTTTCCACTCT	56	59	60	128-183
Z80345	SCAD	GGAACCAAAGCCTGGATCAC	GGACCAGGAAGGCACTGATG	106	58	59	541-646
L05144	PCK1	CTTTGGAGGCCGTAGACCTG	CCGCTGTGGCCTCTGATC	104	58	59	1417-1520

a Melting temperature
b The position of the amplified fragment in the cDNA sequence

Primers were designed in the PrimerExpress software from Applied Biosystems (Foster City, CA). Single determinations were performed on ABI PRISM® 7000 Sequence Detection System using the SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA). The PCR reaction consisted of 12.5 μl SYBR Green PCR Master Mix, 300 nM of forward and reverse primers, and 2.5 μl 1:20 diluted template cDNA in a total volume of 25 μl. The reaction was thermocycled using the default settings of "ABI Prism 7000 SDS Software 1.0": 2 min. at 50°C, 10 min. at 95°C, followed by 40 rounds of 15 sec. 95°C and 1 min at 60°C. A dissociation protocol was added after thermocycling, determining dissociation of the PCR products from 65°C to 95°C. All samples were normalised to GAPDH as described below. According to GeneChip data GAPDH is consistently expressed in our samples, maximum variation is two-fold. Signals from GeneChip analyses were compared to the normalised RT-PCR data.

Normalisation of Real-time PCR (RT-PCR). The Normalisation gene is presumed to be approximately equally expressed in all the samples. To normalise the gene expression for a specific sample, the expression value is divided with the expression value for the Normalisation gene in the same sample. To determine the standard deviation of the normalised value (R), the following formula is used:

R=gene expression / Normalisation gene expression

$$StdevR = R \times \sqrt{\left(\frac{stdev\_gene}{average\_gene}\right)^2 + \left(\frac{stdev\_normalizationgene}{average\_normalizationgene}\right)^2}$$

Microsatellite analysis. Ten of the tumours analyzed on GeneChips were snap frozen in Tissue-Tek-II (Sakura Finetechnical Co., Ltd., Tokyo, Japan) and thirty cryo-sections each were cut at 10μm and stained with haematoxylin. The first and last section was cut at 4 μm, stained with haematoxylin, and routinely mounted. These two sections were used for the identification of tumour and normal cells from each sample. Regions enriched in tumour cells (more than 90%) were microdissected from these sections and DNA was extracted using a Puregene DNA extraction kit

(Gentra Systems, Minneapolis, MN). DNA from blood samples was used as control. From the other 15 tumours used for array analysis it was not possible to obtain either microdissected tumour tissue or control DNA. The samples were analysed for microsatellite instability using markers BAT25 and BAT26 as previously described [(2)].

Immunohistochemistry. For staining, 4μm formalin-fixed and paraffin-embedded sections from normal mucosa and tumour tissue were transferred to Menzel Superfrost® -PLUS-slides. After equilibration to room temperature and deparaffining at 37°C overnight, sections were washed 3 times 5min in Tissue Clear (Sakura), 3 times 5min in 99% EtOH, 2 times 5min in 96% EtOH, 5 min in 70% EtOH, 5 min in running tap water. Endogen Peroxidase was blocked with 1.5% H_2O_2 for 10min. Epitope demasking was performed by HIER (Heat induced epitope retrieval) for 20min in TEG-buffer pH 9.0. The primary antibody was diluted in Antibody Diluent with background reducing agent (Cat No. S3022, DakoCytomation, DK), applied to the tissue sections and incubated for 60min at room temperature in a Magnetic Immuno Staining Tray (Cell Path, plc, UK). The primary antibodies used were monoclonal mouse anti-human COX-2 (Cat No. 35-8200, Zymed, AH-diagnostisk, DK), diluted 1:300 and monoclonal mouse anti-human Cytokeratin 20 (Cat. No. M7019, Dako Cytomation, DK), diluted 1:100. Sections were rinsed three times 3min in PBS buffer and a secondary antibody solution was applied for 30min (undiluted labelled polymer, HRP solution with anti mouse polyclonal antibodies K4001, DakoCytomation, DK). Sections were rinsed with TBS buffer and stained for 10 min with DAB solution (1 DAB tablet (KEM-EN-TEC, Copenhagen, Denmark) was dissolved in 10ml dest. H_2O and 10µl H_2O_2 (35%). Sections were counterstained for 30sec with Mayer's Hematoxylin and mounted with Faramount**,** aqueous mounting medium (cat. No. 3025, DakoCytomation, DK).

References

(1) Thykjaer T, Workman C, Kruhoffer M *et al*. Identification of gene expression patterns in superficial and invasive human bladder cancer. *Cancer Res* 2001;**61**(6):2492-9.

(2) Loukola A, Eklin K, Laiho P *et al*. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res* 2001;**61**(11):4545-9.