npg

ONCOGENOMICS

Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis

P Laiho[1,6], A Kokko[1,6], S Vanharanta[1], R Salovaara[1], H Sammalkorpi[1], H Järvinen[2], J-P Mecklin[3], TJ Karttunen[4], K Tuppurainen[4], V Davalos[5], S Schwartz Jr[5], D Arango[1,5], MJ Mäkinen[4] and LA Aaltonen[1]

[1]Department of Medical Genetics and Molecular and Cancer Biology Research Program, Biomedicum Helsinki, University of Helsinki, Finland; [2]The Second Department of Surgery, Helsinki University Central Hospital, Helsinki, Finland; [3]The Department of Surgery, Jyväskylä Central Hospital, Jyväskylä, Finland; [4]Department of Pathology, University of Oulu, Finland and [5]Molecular Biology and Biochemistry Research Center (CIBBIM), Vall d'Hebron Hospital Research Institute, Barcelona, Spain

Serrated colorectal carcinomas (CRCs) are morphologically different from conventional CRCs and have been proposed to follow a distinct pathway of CRC formation. Despite studies of single molecular events in this tumor type, the diagnosis of serrated CRC relies on morphology and the putative unique biological character of these tumors has not been established. Here we show that the gene expression profiling of 37 CRCs separated serrated and conventional CRCs into two distinct branches in unsupervised hierarchical clustering (P-value 7.8×10^{-7}), and revealed 201 differentially expressed genes representing potential biomarkers for serrated CRC. Immunohistochemistry was utilized to verify the key findings in the 37 CRCs examined by expression profiling, and a separate validation set of 37 serrated and 86 conventional CRCs was examined to evaluate the candidate biomarkers in an extended sample material. Ephrin receptor B2, hypoxia-inducible factor 1-alpha and patched appeared as proteins important for genesis of serrated CRC. This study establishes serrated CRCs as a biologically distinct subclass of CRC and represents a step forward in the molecular classification of these cancers. The study also provides a platform to understand the molecular basis of serrated CRC and in long term may contribute to the development of specific treatment options for this tumor type.
Oncogene advance online publication, 3 July 2006; doi:10.1038/sj.onc.1209778

Introduction

Serrated colorectal carcinomas (CRC) are proposed to arise through a recently introduced serrated neoplasia pathway: the precursor lesions being hyperplastic polyps and serrated adenomas (Hawkins *et al.*, 2002; Higuchi and Jass, 2004). Serrated adenomas were first described in 1990 by Longacre and Fenoglio-Preiser, (1990) on morphological basis. They show dysplasia and sawtoothed architecture similar to hyperplastic polyps. Serrated dysplasia differs from that of conventional adenomas, as serrated adenomas possess more mature epithelium than conventional adenomas (Lazarus *et al.*, 2005). Hyperplastic polyps have been considered to be unrelated to serrated adenomas and to have little or no malignant potential. Recent reports, however, have questioned this view. Some hyperplastic polyps display molecular features as seen in neoplastic lesions, such as microsatellite instability (MSI), aberrant DNA methylation and mutations in *BRAF* and *KRAS* genes. Progression of hyperplastic polyps and serrated adenomas into carcinoma has been reported (Mäkinen *et al.*, 2001). Adenocarcinomas displaying serrated features seem to account for up to 7.5% of all CRCs, making it a significant subgroup of CRC (Mäkinen *et al.*, 2001; Tuppurainen *et al.*, 2005).

Serrated CRCs are morphologically different from conventional CRCs, but whether they are fundamentally biologically different is not clear. Tateyama *et al.* (2002) have proposed that the saw-toothed structure of serrated adenomas may be owing to the inhibition of apoptosis. The decreased expression of CD95 (Fas) in cells of the upper crypt may lead to epithelial hypermaturation when cells continue proliferating but are blocked from ascending to their correct location in the colonic epithelium. This was proposed to explain the characteristic serration of the lesions. Much emphasis has been placed on identifying genetic changes that underlie the serrated pathway. One molecular feature previously associated with serrated lesions is MSI (Iino *et al.*, 1999; Hawkins and Ward, 2001; Mäkinen *et al.*, 2001). In addition, it has been shown that cytosine guanine dinucleotide (CpG) island methylator

phenotype is common in hyperplastic polyps and serrated adenomas (Chan et al., 2002; Park et al., 2003) and that allelic imbalance of the 18q is more common and 5q less common in the serrated adenomas when compared to other polyps (Yashiro et al., 2005). Kambara et al. (2004) found an association between BRAF mutations and DNA methylation in sessile serrated adenomas and suggested that mutations in BRAF are early events in the serrated pathway. The genes frequently altered in conventional CRCs, KRAS, p53 and APC are altered in a subset of serrated lesions as well. Approximately 20% of serrated adenomas display APC/β-catenin abnormalities and KRAS and p53 are mutated in 15 and 8% of serrated adenomas, respectively. This suggests that at least some serrated tumors evolve through the 'classical' pathway involving changes in Wnt signaling (Sawyer et al., 2002).

Despite the growing number of studies examining single molecular events underlying the serrated pathway, the biological background of these tumors is still largely unknown. The hypothesis that serrated CRCs are biologically different from conventional CRCs relies mostly on evidence from morphological data, and considering the likely heterogeneous nature of serrated adenomas, it would be of fundamental importance to clarify whether serrated CRCs indeed form a novel type of CRC with distinct molecular basis. To investigate the molecular basis of serrated CRCs, we have performed expression microarray and immunohistochemical (IHC) analyses of serrated and conventional colorectal adenocarcinomas. The results establish serrated CRCs as a distinct form of colorectal cancer.

Results

Expression microarray analyses
Tumors with serrated histology were screened from two population-based collections consisting of altogether 1508 CRC samples (Aaltonen et al., 1998; Salovaara et al., 2000; Mäkinen et al., 2001) (see Materials and methods). In total, 45 cases of serrated CRC were identified. Eight of these were fresh–frozen, had sufficient quality RNA and were assayed using Affymetrix HG-U133A expression microarrays containing 22 283 probe sets. As a comparison group, expression array data set from 29 identically processed age-, sex- and grade-matched conventional CRCs from the same sample collections was utilized (Arango et al., 2005; unpublished data). Strikingly, gene expression profiling using unsupervised hierarchical clustering with a list of 7928 probes that were expressed in at least 29 of the 37 samples showed a clear distinction between the serrated and the conventional CRCs, as all but one of the eight serrated samples clustered in a single branch in the hierarchical structure ($P = 7.8 \times 10^{-7}$, Fisher's exact test). The only exception to this was one sample (C634), which was grouped in the same branch with the 29 conventional CRCs (Figure 1). Parameters such as sex, site, grade or stage did not cluster together.

Statistical testing was utilized to identify genes and pathways displaying the most significantly altered expression between serrated and conventional CRCs. By using a Student's t-test with correction for multiple testing, we detected 226 differentially expressed genes with an adjusted P-value less than 0.05 representing 201 distinct genes and seven expressed sequences (hypothetical proteins). Fifteen genes were represented by more than one probe (Supplementary Table S1).

We then wanted to identify categories of functionally related genes associated with the serrated histology. For this purpose, Gene Ontology (GO) terms (Ashburner et al., 2000) were used to classify the 7928 filtered probes used in the unsupervised clustering into 685 biological processes, cellular components and molecular functions. Functional group enrichment analysis was used to identify categories with a significant enrichment in the number of genes differentially expressed in tumors with serrated and non-serrated histology. We found nine categories with a P-value <0.01, of which five belonged to categories linked to morphogenesis (morphogenesis and organogenesis) and membrane-associated genes (membrane, integral to membrane and integral to plasma membrane). A detailed description of the groups is provided in Supplementary Table S2.

Having detected a distinguishable expression pattern for the serrated CRCs, we tested whether a small set of genes could be used to predict the tumor subtype. The class prediction method described by Golub et al. (1999) was applied to find a set of differentially expressed genes (probes) between the serrated and conventional CRCs that could predict whether an unknown sample was a serrated or a conventional CRC. Four thousand four hundred and thirteen probes detected in all 37 samples were used to generate a K-nearest neighbors (KNN) classifier that was validated using a leave-one-out cross-validation procedure. This classifier correctly categorized all 37 tumors as serrated or non-serrated histology using the expression data of 10 genes (Supplementary Figure S1). As each one of the 37 rounds of the leave-one-out cross-validation procedure uses a slightly different set of genes, a total of 27 probes were used (Supplementary Table S3).

Clinical and pathological features and immunohistochemistry
The expression microarray results were validated by IHC staining utilizing a training set (TS) and a validation set (VS) of paraffin-embedded tissue blocks. The TS consisted of the 37 (eight serrated and 29 conventional) samples that entered expression array analysis, whereas the VS included a separate set of 37 serrated and 86 conventional CRCs from the same collections. The serrated CRCs represented a mucinous component more often than the conventional CRCs. They also showed more variability in the World Health Organization grade, tended to be more advanced and more frequently located in the proximal colon. Detailed clinical features of the samples used in this study are presented in Table 1.
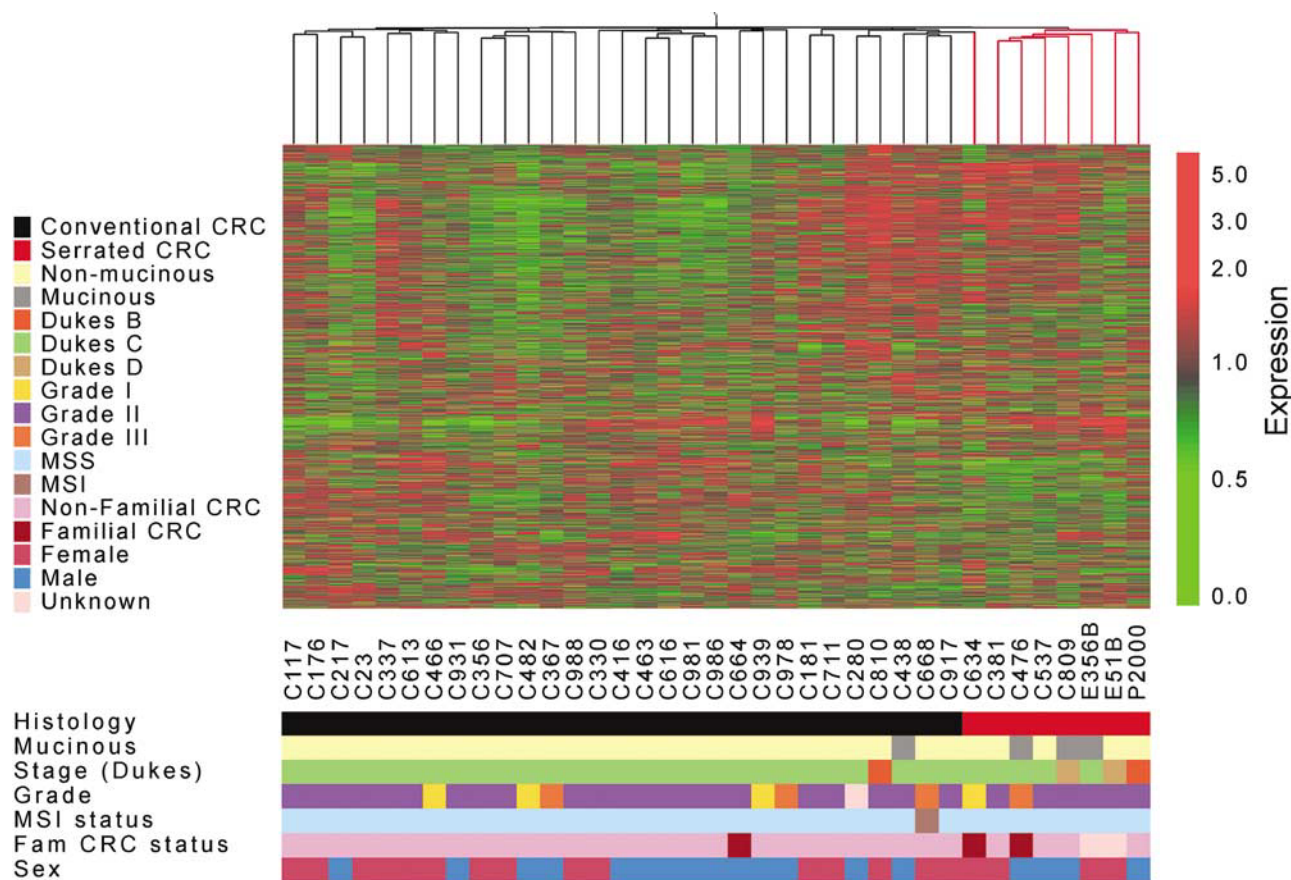
Figure 1 Unsupervised hierarchical clustering of the 37 samples with 7928 probes shows a clear distinction between the serrated CRCs and the conventional CRCs.

The proteins for the IHC were chosen based on observed differential expression at the RNA level, relevance of their biological function and the commercial availability of the antibodies. Based on these criteria, ephrin receptor B2 (EPHB2), patched (PTCH), hypoxia-inducible factor 1-alpha (HIF1α), cyclin T2 (CCNT2), metastasis-associated 1 (MTA1) and methyl-CpG-binding domain protein 4 (MBD4) were selected.

Normal colorectal mucosa was characterized by positive nuclear staining for CCNT2, MTA1 and MBD4. Normal PTCH staining pattern was membranous at the crypt basis, and positive granular cytoplasmic staining was observed in both crypt and luminal epithelium. EPHB2 staining pattern was membranous at the crypt basis, and weaker cytoplasmic immunoreaction was detected in the crypt epithelium. In normal colorectal epithelium, HIF1α was regularly negative. In some tumors, focal positivity was observed at the ulcerated surface of the tumor or around necrotic areas. This pattern, compatible with normal response to ischemia, was not scored in the results. The IHC staining for EPHB2, PTCH and HIF-1α demonstrated statistically significant associations with serrated morphology of the CRCs (Figure 2).

Compatible with a recent report associating EPHB2 expression with CRC grade (Batlle et al., 2005), in conventional CRCs, major loss ($\geqslant 80\%$) of membra-nous EPHB2 staining was observed in 18.8% (3/16) of well-differentiated carcinomas, in 26.4% (23/87) of moderately differentiated carcinomas and in 66.7% (8/12) of poorly differentiated carcinomas ($P = 0.036$, Kruskal–Wallis). However, no significant trend between EPHB2 expression and grade could be observed in the serrated lesions. 46.2% (6/13) of well differentiated, 73.9% (17/23) of moderately differentiated and 66.7% (6/9) of poorly differentiated serrated cancers showed major loss of EPHB2 membranous staining ($P = 0.172$, Kruskal–Wallis).

Serrated CRCs were characterized by more pronounced loss of EPHB2 (TS $P = 0.026$, VS $P = 0.012$, Fisher's exact test; Table 2) and PTCH (TS $P = 0.1$, VS $P = 0.0003$), and frequent positivity for HIF-1α (TS $P = 0.021$, VS $P = 0.00006$). Additionally, serrated CRCs showed a tendency to preserved nuclear immunoreaction for CCNT2. The difference was statistically significant in the TS ($P = 0.004$), confirming the expression array data obtained from the 37 samples but failed to reach a significant level in the VS ($P = 0.21$). A combination of 80–100% loss of membranous EPHB2, loss of PTCH, preserved nuclear CCNT2 and over-expression of HIF1α was useful in distinguishing serrated cancers from conventional cancers: 71.1% (32/45) of the serrated CRCs showed the above-mentioned profile with three or four antibodies, whereas

Table 1 Clinical features of the samples

	Training set				Validation set				All samples				
	S CRCs		C CRCs		S CRCs		C CRCs		S CRCs		C CRCs		P-value
	N	%	N	%	N	%	N	%	N	%	N	%	
MSI													
MSI	0	0	1	3,4	4	10.8	9	10.5	4	8.9	10	8.7	1.0[a]
MSS	8	100	28	96.6	33	89.2	77	89.5	41	91.1	105	91.3	
Age													
Mean (range)	70.9	(46–86)	71.8	(55–88)	65.4	(39–89)	68	(38–88)	66.4	(39–89)	68.6	(38–88)	0.29[b]
Dukes stage													
A	0	0	0	0	7	18.9	17	19.8	7	15.6	17	14.8	0.08[a]
B	1	12.5	1	3.4	13	35.1	31	36	14	31.1	32	27.8	
C	5	62.5	28	96.5	12	32.4	33	38.4	17	37.8	61	53	
D	2	25	0	0	5	13.5	5	5.8	7	15.6	5	4.3	
WHO grade													
1	1	12.5	3	10.3	12	32.4	13	15.1	13	28.9	16	13.9	0.01
2	6	75	23	79.3	17	45.9	64	74.4	23	51.1	87	75.7	
3	1	12.5	3	10.3	8	21.6	9	10.5	9	20	12	10.4	
Location													
Proximal colon	5	62.5	9	31	17	45.9	32	37.2	22	48.9	41	35.7	0.15
Distal colon	3	37.5	20	69	20	54.1	54	62.8	23	51.1	74	64.3	
Gender													
Female	4	50	15	51.7	24	64.9	56	65.1	28	62.2	71	61.7	1.0
Male	4	50	14	48.3	13	35.1	30	34.9	17	37.8	44	38.3	
Type													
Non-mucinous	5	62.5	28	96.6	16	43.2	66	76.7	21	46.7	94	81.7	<0.0001
Mucinous component	2	25	0	0	11	29.7	12	14	13	28.9	12	10.4	
Mucinous	1	12.5	1	3.4	10	27	8	9.3	11	24.4	9	7.8	

Abbreviations: C CRC, conventional CRC, S CRC, serrated CRC. [a] = Fisher's exact test, [b] = t-test, χ^2 test is used for statistical analysis, unless otherwise stated.

only 16.5% (19/115) of the conventional CRCs had a similar profile. The expression of MTA1 and MBD4 did not show correlation between the cancer types after staining the 37 samples hybridized on the microarray chips. After extending the set with 22 samples (six serrated and 16 conventional CRCs) from the VS, the difference was still not statistically significant (for MTA1 in TS $P = 0.557$, in VS $P = 1.0$; for MBD4 in TS $P = 0.124$, in VS $P = 0.351$). Thus, the whole sample set was not stained with these two antibodies. Detailed IHC staining results are presented in Table 3.

EPHB2 *mutation screening and promoter hypermethylation analysis*
To examine possible causes of reduced EPHB2 expression, the genomic DNA of the available 24 serrated tumors was sequenced for somatic mutations in the *EPHB2* gene. Successful sequencing of 98% of the coding region yielded negative results. Twelve samples displayed informative single-nucleotide polymorphisms (SNPs) and three of them (25%) showed loss of heterozygosity (LOH) as compared to the respective normal tissue DNA sequence. The *EPHB2* promoter was hypermethylated in five out of the eight (63%) tumors.

Survival statistics
An average 5-year survival was calculated for all cases. The 5-year survival was 50.1% for the serrated CRCs, and 60.5% for the conventional CRCs ($P = 0.201$, log-rank). The serrated CRCs showed a tendency for worse prognosis. For instance, in male patients and in microsatellite stable cancers, serrated CRCs tended to have worse prognosis. In moderately differentiated cancers, this difference was evident ($P = 0.014$, Log-rank), and a less clear tendency was seen in well-differentiated cancers ($P = 0.134$, log-rank). The detailed survival statistics are presented in Supplementary Table S4.

Discussion

The serrated neoplasia pathway (Hawkins et al., 2002; Jass et al., 2002; Higuchi and Jass, 2004; Goldstein, 2006) has been proposed to represent a previously underrecognized route leading to CRC, but firm molecular evidence for this has been lacking. In this study, we performed expression microarray and IHC analyses to serrated and conventional CRC samples and showed that serrated CRCs differ from conventional
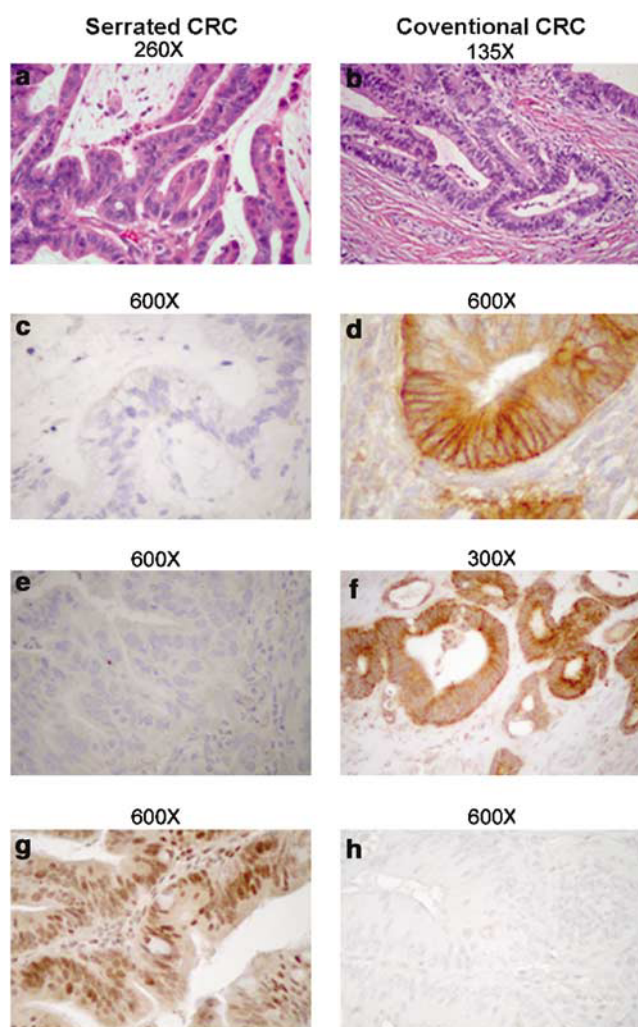
Figure 2 (**a–b**) Morphology of serrated (**a**) and conventional (**b**) CRCs. Serrated CRC consists of small papillary epithelial tufts, accompanied with extracellular mucin production in the depicted case. A conventional CRC is composed of tubular formations with no evidence of papillary growth. (**c–d**). EPHB2 IHC in serrated (**c**) and conventional CRCs (**d**). EPHB2 immunoreaction is absent in serrated CRC (**c**), whereas in conventional CRC membranous immunoreaction for EPHB2 is uniform (**d**). (**e–f**) Patched IHC in serrated (**e**) and conventional CRCs (**f**). Serrated CRC is negative for patched (**e**), whereas in conventional CRC there is preserved cytoplasmic immunoreaction (**f**). (**g**)–(**h**). HIF1α IHC in serrated (**g**) and conventional (**h**) CRCs. HIF1α is overexpressed in a serrated CRC, with uniform nuclear immunoreaction (**g**). A conventional CRC shows the absence of nuclear immunoreaction (**h**).

CRCs on both morphological and molecular level and that these lesions indeed form a distinct subclass of colorectal carcinomas. All samples came from well-characterized population-based collections of CRC samples (Aaltonen et al., 1998; Salovaara et al., 2000; Mäkinen et al., 2001), which minimized sources of bias introduced by sample selection.

Many clinical and pathological features suggested that serrated CRCs may be more aggressive than conventional CRCs. Metastases and poorly differentiated carcinomas were more frequent among the serrated CRCs, and the survival of patients with serrated CRC tended to be worse than that of patients having conventional CRC within many studied parameters, for example, in male patients, in microsatellite stable cancers and in moderately differentiated cancers. In a smaller set of serrated CRCs, we have previously observed a similar trend for distal and microsatellite stable serrated cancers (Mäkinen et al., 2001).

The serrated CRCs had a distinct gene expression profile differing from the conventional adenocarcinomas, which was demonstrated by the unsupervised hierarchical clustering of these two tumor types into separate branches ($P = 7.8 \times 10^{-7}$, Fisher's exact test). This provides strong evidence for fundamental biological differences between these tumors. It should be noted that all the serrated CRCs available for expression profiling were MSS and thus we cannot predict how MSI-serrated cancers would cluster. Nevertheless, our study demonstrates that at least a subset of serrated carcinomas represent yet another pathway to colorectal cancer. The one conventional MSI cancer clustered with other conventional CRCs.

Expression microarray data were further utilized to find differentially expressed genes in the serrated CRCs. Two hundred and one genes and seven sequences encoding predicted proteins were found to have significantly (adjusted P-value < 0.05) altered expression between the two classes (Supplementary Table S1). In addition, GO terms were utilized to investigate whether any functional group had been enriched with genes differentially expressed between the serrated and conventional CRCs. The most prominent categories were morphogenesis, organogenesis and membrane-associated genes. Although these groups are very general, they are of some interest, because morphological characteristics define the serrated tumor phenotype.

Table 2 Results of the immunohistochemistry for EPHB2

	Training set (P = 0.026)				Validation set (P = 0.012)				All samples (P = 0.0004)			
	G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4
EPHB2												
C CRC	3	12	7	7	17	23	19	27	20	35	26	34
%	10.3	41.4	24.1	24.1	19.8	26.7	22.1	31.4	17.4	30.4	22.6	29.6
S CRC	1	0	1	6	3	4	7	23	4	4	8	29
%	12.5	0	12.5	75	8.1	10.8	18.9	62.2	8.9	8.9	17.8	64.4
Total	4	12	7	7	20	27	26	50	24	39	34	63
%	10.8	32.4	21.6	35.1	16.3	22	21.1	40.7	15	24.4	21.3	39.4

Abbreviations: P = Fisher's exact test P-value, G1 = 80–100%, G2 = 50–80%, G3 = 20–50% and G4 = 0–20% of EPHB2-positive cells, C CRC, conventional CRC, S CRC, serrated CRC. The staining was assessed as percentage of positive cells.

Table 3 Results of the immunohistochemistry for PTCH, HIF1α, CNNT2 (cytoplasmic and nuclear), MTA1 and MBD4

	Training set			VS			All samples		
Staining	−/− D	+/+ G	P	−/− D	+/+ G	P	−/− D	+/+ G	P
PTCH									
C CRC	8	21	0.1	27	59	0.0003	35	80	<5E-05
%	27.6	72.4		31.4	68.6		30.4	69.6	
S CRC	5	3		25	12		30	15	
%	62.5	37.5		67.6	32.4		66.7	33.3	
Total	13	24		52	71		65	95	
%	35.1	64.9		42.3	57.7		40.6	59.4	
HIF1α									
C CRC	24	5	0.021	66	20	6.00E-05	90	25	<1E-05
%	82.8	17.2		76.7	23.3		78.3	21.7	
S CRC	3	5		14	23		17	28	
%	37.5	62.5		37.8	62.2		37.8	62.2	
Total	27	10		80	43		107	53	
%	73	27		65	35		66.9	33.1	
CNNT2 cytoplasmic									
C CRC	7	22	0.013	55	31	0.294	62	53	0.019
%	24.1	75.9		64	36		53.9	46.1	
S CRC	6	2		28	9		34	11	
%	75	25		75.7	24.3		75.6	24.4	
Total	13	24		83	40		96	64	
%	35.1	64.9		67.5	32.5		60	40	
CNNT2 nuclear									
C CRC	21	8	0.004	18	68	0.21	39	76	0.005
%	72.4	27.6		20.9	79.1		33.9	66.1	
S CRC	1	7		4	33		5	40	
%	12.5	87.5		10.8	89.2		11.1	88.9	
Total	22	15		22	101		44	116	
%	59.5	40.5		17.9	82.1		27.5	72.5	
MTA1 (N = 59)									
C CRC	4	25	0.557	9	7	1	13	32	0.738
%	13.8	86.2		56.3	43.8		28.9	71.1	
S CRC	0	8		3	3		3	11	
%	0	100		50	50		27.1	72.9	
Total	4	33		12	10		16	43	
%	10.8	89.2		54.5	45.5		27.1	72.9	
MBD4 (N = 59)									
C CRC	12	17	0.124	7	9	0.351	19	26	0.76
%	41.4	58.6		43.8	56.3		42.2	57.8	
S CRC	6	2		1	5		7	7	
%	75	25		16.7	83.3		50	50	
Total	18	19		8	14		26	33	
%	48.6	51.4		36.4	63.6		44.1	55.9	

Abbreviations: C CRC = conventional CRC, S CRC = serrated CRC; D, diffuse, G, granular, −/− = negative or weak positive, +/+ = moderate or strong positive, P = Fisher's exact test P-value.

To validate the expression microarray results at the protein level, six genes that were differentially expressed on microarrays, *EPHB2*, *PTCH*, *HIF1α*, *CCNT2*, *MTA1* and *MBD4*, were selected for IHC analyses in a set of 160 tumors.

EPHB2 is a member of a receptor tyrosine kinase family that regulates several signaling pathways involved for example, in cell growth and migration (Kullander and Klein, 2002). In the intestinal tract, EPHB2 maintains the crypt–villus axis, and the disruption of mouse *Ephb2* and *Ephb3* genes results in the loss of normal organization of the cells in the crypts and in abnormal mixing of different cell types (Batlle *et al.*, 2002). *EPHB2* maps to chromosome 1p36, which is a chromosomal region reported to be lost in 13% of hyperplastic polyps (Rashid *et al.*, 2000). This gene is also mutated in prostate cancer (Huusko *et al.*, 2004). In our IHC analyses, a preserved membranous staining of

EPHB2 was observed more often in the conventional CRCs when compared to serrated lesions; EPHB2 was lost in more than half of the serrated CRCs but only in a quarter of their conventional counterparts. In accordance with this, Batlle *et al.* (2005) observed complete loss of EPHB2 in 25% of conventional CRCs of different Dukes' stages. They also reported that EPHB2 plays a critical role in CRC progression which makes EPHB2 a strong tumor suppressor candidate. We observed *EPHB2* LOH in 25% (3/12) and promoter hypermethylation in 63% (5/8) of serrated tumors.

Our study on EPHB2 inactivation in CRC (Alazzouzi *et al.*, 2005) has showed that *EPHB2* promotor methylation indeed is causally related to the reduced expression of EPHB2 protein in a cell line and we found frequent frameshift mutations in *EPHB2* in MSI adenomas and carcinomas (21 and 41%, respectively). This finding is interesting and well compatible with our present data on serrated CRCs; indeed, previous literature (Hawkins and Ward, 2001; Mäkinen *et al.*, 2001) has connected MSI phenotype and serrated morphology, and early loss of EPHB2 may play a role in this association. That reduction of EPHB2 in serrated CRCs is not associated with grade, whereas such an association clearly is present in conventional CRCs ((Batlle *et al.*, 2005), this study), is also compatible with this hypothesis.

Another gene differentially expressed in the microarray analysis was *PTCH*, which is a member of the Hedgehog (HH) signaling pathway. During development, HH signaling is essential for the patterning of the gastrointestinal tract, as well as the maintenance of stem cells (Oldak *et al.*, 2001). The HH signal regulates the transcription of a number of target genes, for example *BMP-2*, *BMP-4* and the *GLI* genes (Narita *et al.*, 2000; Mullor *et al.*, 2002). Germline mutations in *PTCH* underlie the Gorlin syndrome, which is characterized by the early-onset multiple basal cell carcinomas and increased rate of some other tumors such as medulloblastomas (Hahn *et al.*, 1996; Johnson *et al.*, 1996). Thus far, *PTCH* defects have not been reported in primary CRCs, but hamartomatous polyposis is a rare feature of Gorlin syndrome (Schwartz, 1978). In our IHC analyses, PTCH showed significantly more preserved granular immunoreaction pattern in the conventional CRCs when compared to the serrated CRCs. Based on these findings, *PTCH* could be a candidate gene associated to the genesis of serrated CRC. In addition, a subsequent data analysis of *PTCH* downstream targets (Taipale and Beachy, 2001) indicated *SMO* to have reduced expression level, perhaps through some feedback mechanism. This did not, however, survive correction for multiple testing.

*HIF*1α, a gene overexpressed in serrated CRCs on the expression microarray analysis, activates the transcription of genes that are involved in crucial aspects of cancer biology, including angiogenesis, cell survival, glucose metabolism and invasion (Harris, 2002; Semenza, 2003). HIF1α induces the expression of its target genes, such as *vascular endothelial growth factor* (VEGF) (Pages and Pouyssegur, 2005), in response to hypoxia.

Overexpression of HIF1α has been associated with cell proliferation in colon carcinoma (Zhong *et al.*, 1999). In our IHC results, the HIF1α immunoreaction was seen as diffuse positivity throughout the tumor, which was significantly more frequent in serrated CRCs than in conventional CRCs. Consequently, the abnormal morphology and cell proliferation in the serrated tumors could be associated to the overexpression of *HIF1α*. A subsequent data analysis of HIF downstream targets (Semenza, 2003) indicated that four genes (*BNIP3L*, *AK3*, *FN1*, *VIM*) had differential expression ($P < 0.05$) in CRCs. Of these, *BNIP3L* survived correction for multiple testing. Of some interest was that overexpression (measured as serrated vs non-serrated mean expression ratio > 1.2) was detected in 22 genes, whereas underexpression (ratio < 0.8) was observed in only one gene.

Antibodies against CCNT2, MTA1 and MBD4 did not produce statistically significant results in the IHC VS, although significant differences were observed in the IHC staining pattern of CCNT2 in the TS. Thus, these reagents failed to be useful for the detection of serrated CRC.

The gene expression data together with the IHC analyses strongly indicate that serrated and conventional CRCs display significant molecular differences, observed both on molecular and histopathological level. Interestingly, our expression microarray and IHC studies indicate that suppression of *EPHB2* and *PTCH* and upregulation of *HIF1α*, genes involved in biological processes closely related to the colon cell morphology, differentiation, gastrointestinal tract patterning and proliferation, are associated with the serrated phenotype and thus could be candidate genes involved in the serrated neoplasia pathway.

This work provides a platform for identifying markers for diagnosing serrated CRCs, some of which may be useful in evaluating malignant potential of benign serrated lesions of varying origin. The distinct expression profile of serrated CRCs strongly suggests that these cancers form a novel subtype of CRC, and the underlying biological differences are likely to affect the optimal treatment. Further work is needed to elucidate the key therapeutic targets in the serrated carcinoma pathway, but this work is likely to be rewarded in the form of more specific treatment options for this tumor type.

Materials and methods

Patient material

To identify CRC cases with serrated morphology, samples were screened from two previously described, consecutively colleted and population-based Finnish collections consisting of a set of 1042 samples (Aaltonen *et al.*, 1998; Salovaara *et al.*, 2000) and another set of 466 cases (Mäkinen *et al.*, 2001). Altogether, 160 tumor samples, of which 45 were serrated and 115 conventional CRCs, were included in the study. The histology of the samples was independently confirmed by two pathologists by reanalyzing hematoxylin- and eosin-stained paraffin sections. Samples were diagnosed as serrated CRCs, when the cancer tissue was composed of epithelial proliferation reminiscent of serrated adenoma, that is, the cytoplasm of these cells was clear or eosinophilic, and when the cellular

changes were accompanied with serrated growth pattern, that is, cells forming pseudopapillary or pseudocribriform structures with tufting of the cells into the lumen without true papillary structures with a fibrovascular core (Tuppurainen et al., 2005). After screening the subset of the population-based collections containing both paraffin-embedded and fresh–frozen material, eight fresh–frozen serrated CRCs were identified, which yielded sufficient quality RNA compatible with expression microarray approach. Seven out of the eight had metastasized at the time of diagnosis. As a comparison set, we used expression microarray data from 29 age-, sex- and grade-matched conventional Dukes C CRCs from the same population-based series processed identically and in tandem with the serrated samples (Arango et al., 2005; unpublished data) (Table 1). Five of the serrated CRCs were located in the proximal and three in the distal colon, whereas nine of the conventional CRCs were located in the proximal and 20 in the distal colon. All samples selected for microarray analysis had more than 60% of tumor cells. Notably, only one of the 37 samples that entered the microarray analysis was microsatellite unstable (Table 1). Patient information and samples were collected after obtaining informed consent. The study was approved by the Ethics review committees of the corresponding institutions.

Microarray preparation and hybridization
Total RNA was extracted from fresh–frozen tumors with Trizol reagent (GibcoBRL, Long Island, NY, USA) and further purified using RNeasy spin columns (Qiagen, Valencia, CA, USA). The RNA quality was analysed using a spectrophotometer and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Biotin-labeled and fragmented cRNA was prepared from 8 μg of total RNA with procedures recommended for the HG-U133 GeneChip expression analysis (Affymetrix Inc., Santa Clara, CA, USA). The HG-U133A chips were hybridized, scanned and analysed with Microarray Suite 5.0 software according to the manufacturer's instructions (Affymetrix). The hybridization quality of the chips met the guidelines for the control parameters suggested by The Tumor Analysis Best Practices Working Group (Hoffman et al., 2004). All microarray raw data tables are available in the Gene Expression Omnibus (GEO) database under the Accession number GSE4045.

Statistical analysis of microarray data
The quantitative expression data were normalized by truncating small values to 0.01 and centering both array and gene intensities to the corresponding median. The normalized expression data were filtered to remove expression values below detectable levels using the Affymetrix Detection Algorithm assigned flag calls. Subsequent analyses were restricted to genes with detectable expression in at least 29 of 37 samples (78%), as this allowed inclusion of genes absent only in the eight serrated samples while removing noise from the data. A total of 7928 probe sets fulfilled these criteria. The unsupervised hierarchical clustering used Spearman's rank correlation as the similarity metric. Genes differentially expressed between the serrated and non-serrated CRC groups were identified by performing a Student's t-test for unequal variances and a Benjamini and Hochberg multiple test correction (Benjamini and Hochberg, 1995). Data processing and analyses were performed with GeneSpring 6.2.1 software (Silicon Genetics, Redwood City, CA, USA). For the functional group enrichment analysis, GO terms (http://www.geneontology.org) were used to annotate genes with detectable expression in at least 29 of 37 samples (7928 genes).

GO annotations were found for 3346 genes. The list of 226 differentially expressed genes was used to identify categories that were enriched in serrated vs conventional CRCs. These analyses were implemented using GoMiner software (Zeeberg et al., 2003) and a Fisher's exact test was used to identify significantly enriched categories.

The predictor was built in GeneCluster2 software (Tamayo et al., 1999) using three nearest neighbors with signal-to-noise feature selection (mean class estimate) and an $1/k$ weighting method. The leave-one-out cross-validation procedure consisted of sequential rounds, in which one sample is initially removed from the analysis and the remaining 36 samples are used to identify the genes most differentially expressed between the two groups. The expression of these genes is used to identify the samples with the closest expression profiles to the sample initially left out. The predominant class label (serrated or conventional CRC) in the nearest neighbors is then given to the sample initially left out, and compared to the actual class label to assess the accuracy of the classifier. This process is iteratively repeated 37 times, each time leaving out a different sample. In the end, the observed and predicted histologies of all 37 patients are compared.

Immunohistochemistry
The antibodies were selected on the basis of availability and documentation of reactivity on human tissues and in paraffin-embedded material. Details of the antibodies and concentrations used are provided in Supplementary Table S5. Five-micrometer-thick sections were cut from the paraffin blocks. After deparaffinization and rehydration, sections were pretreated in either 0.01 M citrate (pH 6.0) buffer in a microwave oven at 800 W for 2 min and at 300 W for 10 min or in 0.01 M Tris-EDTA (pH 6.0) buffer in a microwave oven at 800 W for two minutes and at 300 W for 15 min. A positive antibody reaction was detected with diaminobenzidine (DAB; Dako, Copenhagen, Denmark) with hematoxylin counterstain. HIF1α, PTCH and MBD4 were scored as positive or negative, other antibodies were semiquantitatively scored as being negative, or showing weak, moderate or strong positivity. For EPHB2, the IHC staining and scoring of the results was performed as described in the literature (Batlle et al., 2005).

Mutation screening and promoter hypermethylation analysis of EPHB2
The 24 serrated tumors, from which DNA was available, were analysed for somatic mutations in EPHB2 by direct sequencing of the coding exons. LOH was scored in cases displaying SNPs.

The DNA methylation status of the promoter-associated CpG islands was determined by chemical conversion of unmethylated cytosines to uracil in eight serrated tumors with ample tumor and normal tissue DNA available. EPHB2 promoter region of 250 bp was defined by analysing a 1400 bp long CpG island upstream of the transcription site with promoter prediction softwares (MethPrimer, Promoter Scan) (Prestridge, 1995; Li and Dahiya, 2002). In vitro methylated DNA (CpG Genome Universal Methylated DNA from Chemicon International, Temecula, CA, USA) was used as a positive control for methylated alleles, whereas paired normal tissue DNA was used as negative control. One microgram of DNA was denatured by NaOH and modified by sodium bisulfite. DNA samples were then purified using the Wizard DNA purification kit (Promega Corp., Madison, WI, USA), again treated with NaOH, precipitated with ethanol and resuspended in water. Methylation-specific PCR (MSP) (Herman et al., 1996) reactions using primers specific for either the methylated or the modified unmethylated DNA were carried

out to determine the methylation profile of each sample. The PCR conditions and primer sequences for mutation screening and promoter hypermethylation analysis are provided in Supplementary Table S6.

References

Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P et al. (1998). N Engl J Med 338: 1481–1487.

Alazzouzi H, Davalos V, Kokko A, Domingo E, Woerner SM, Wilson AJ et al. (2005). Cancer Res 65: 10170–10173.

Arango D, Laiho P, Kokko A, Alhopuro P, Sammalkorpi H, Salovaara R et al. (2005). Gastroenterology 129: 874–884.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al. (2000). Nat Genet 25: 25–29.

Batlle E, Bacani J, Begthel H, Jonkheer S, Gregorieff A, van de Born M et al. (2005). Nature 435: 1126–1130.

Batlle E, Henderson JT, Beghtel H, van den Born MM, Sancho E, Huls G et al. (2002). Cell 111: 251–263.

Benjamini Y, Hochberg Y. (1995). J Roy Stat Soc Ser B 57: 289–300.

Chan AO, Issa JP, Morris JS, Hamilton SR, Rashid A. (2002). Am J Pathol 160: 529–536.

Goldstein NS. (2006). Am J Clin Pathol 125: 146–153.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP et al. (1999). Science 286: 531–537.

Hahn H, Wicking C, Zaphiropoulous PG, Gailani MR, Shanley S, Chidambaram A et al. (1996). Cell 85: 841–851.

Harris AL. (2002). Nat Rev Cancer 2: 38–47.

Hawkins NJ, Bariol C, Ward RL. (2002). Pathology 34: 548–555.

Hawkins NJ, Ward RL. (2001). J Natl Cancer Inst 93: 1307–1313.

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. (1996). Proc Natl Acad Sci USA 93: 9821–9826.

Higuchi T, Jass JR. (2004). J Clin Pathol 57: 682–686.

Hoffman EP, Awad T, Palma J, Webster T, Hubbell E, Warrington JA et al. (2004). Nat Rev Genet 5: 229–237.

Huusko P, Ponciano-Jackson D, Wolf M, Kiefer JA, Azorsa DO, Tuzmen S et al. (2004). Nat Genet 36: 979–983.

Iino H, Jass JR, Simms LA, Young J, Leggett B, Ajioka Y et al. (1999). J Clin Pathol 52: 5–9.

Jass JR, Whitehall VL, Young J, Leggett BA. (2002). Gastroenterology 123: 862–876.

Johnson RL, Rothman AL, Xie J, Goodrich LV, Bare JW, Bonifas JM et al. (1996). Science 272: 1668–1671.

Kambara T, Simms LA, Whitehall VL, Spring KJ, Wynter CV, Walsh MD et al. (2004). Gut 53: 1137–1144.

Kullander K, Klein R. (2002). Nat Rev Mol Cell Biol 3: 475–486.

Lazarus R, Junttila OE, Karttunen TJ, Makinen MJ. (2005). Am J Clin Pathol 123: 349–359.

Li LC, Dahiya R. (2002). Bioinformatics 18: 1427–1431.

Longacre TA, Fenoglio-Preiser CM. (1990). Am J Surg Pathol 14: 524–537.

Mäkinen MJ, George SM, Jernvall P, Makela J, Vihko P, Karttunen TJ. (2001). J Pathol 193: 286–294.

Mullor JL, Sanchez P, Altaba AR. (2002). Trends Cell Biol 12: 562–569.

Narita T, Saitoh K, Kameda T, Kuroiwa A, Mizutani M, Koike C et al. (2000). Development 127: 981–988.

Oldak M, Grzela T, Lazarczyk M, Malejczyk J, Skopinski P. (2001). Int J Mol Med 8: 445–452.

Pages G, Pouyssegur J. (2005). Cardiovasc Res 65: 564–573.

Park SJ, Rashid A, Lee JH, Kim SG, Hamilton SR, Wu TT. (2003). Am J Pathol 162: 815–822.

Prestridge DS. (1995). J Mol Biol 249: 923–932.

Rashid A, Houlihan PS, Booker S, Petersen GM, Giardiello FM, Hamilton SR. (2000). Gastroenterology 119: 323–332.

Salovaara R, Loukola A, Kristo P, Kaariainen H, Ahtola H, Eskelinen M et al. (2000). J Clin Oncol 18: 2193–2200.

Sawyer EJ, Cerar A, Hanby AM, Gorman P, Arends M, Talbot IC et al. (2002). Gut 51: 200–206.

Schwartz RA. (1978). N Engl J Med 299: 49.

Semenza GL. (2003). Nat Rev Cancer 3: 721–732.

Taipale J, Beachy PA. (2001). Nature 411: 349–354.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E et al. (1999). Proc Natl Acad Sci USA 96: 2907–2912.

Tateyama H, Li W, Takahashi E, Miura Y, Sugiura H, Eimoto T. (2002). Am J Surg Pathol 26: 249–256.

Tuppurainen K, Makinen JM, Junttila O, Liakka A, Kyllonen AP, Tuominen H et al. (2005). J Pathol 207: 285–294.

Yashiro M, Laghi L, Saito K, Carethers JM, Slezak P, Rubio C et al. (2005). Cancer Epidemiol Biomarkers Prev 14: 2253–2256.

Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M et al. (2003). Genome Biol 4: R28.

Zhong H, De Marzo AM, Laughner E, Lim M, Hilton DA, Zagzag D et al. (1999). Cancer Res 59: 5830–5835.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc).