Gene Expression Patterns Distinguish Colonoscopically Isolated Human Aberrant Crypt Foci from Normal Colonic Mucosa

Oleg K. Glebov,[1] Luz M. Rodriguez,[1] Peter Soballe,[2] John DeNobile,[2] Janet Cliatt,[1] Kenneth Nakahara,[1] and Ilan R. Kirsch[1]

[1]Genetics Branch, Center for Cancer Research, National Cancer Institute and [2]Surgery Department, Uniformed Services University, National Naval Medical Center, Bethesda, Maryland

Abstract

Aberrant crypt foci (ACF) are considered the earliest identifiable preneoplastic colonic lesions; thus, a greater understanding of the nature of genetic changes underlying the transformation of normal colonic mucosa (NM) into ACF may provide insight into the mechanisms of carcinogenesis. ACF were identified by indigo carmine spraying onto colonic mucosa during colonoscopy and isolated as standard pinch biopsies of the mucosal areas containing the ACF. RNAs isolated from ACF and matched NM biopsies from the ascending and descending colons of 13 patients were analyzed on arrays containing 9128 cDNAs. Thirty-four differentially expressed ($P < 0.001$) genes were found in a paired comparison of the ACF and NM samples, and 25 of 26 matched pairs of ACF and NM could be correctly classified in leave-one-out cross-validation. Differential expression for seven of eight genes was confirmed by real-time reverse transcription-PCR. Furthermore, ACF and NM samples, including six pairs of ACF and NM samples that had not previously been analyzed by array hybridization, can be correctly classified on the basis of the overexpression in ACF of three selected genes (*REG4*, *SRPN-B5*, and *TRIM29*) evaluated by real-time reverse transcription-PCR. In a separate analysis of 13 biopsy pairs from either ascending or descending colon, ACF and NM samples could also be correctly classified by the gene expression patterns. Analysis of gene expression differences in ACF from the ascending and descending colon versus NM samples indicates that ACF from these distinct colonic locations are converging toward similar gene expression profiles and losing differences in gene expression characteristic of NM from the ascending versus descending colon. (Cancer Epidemiol Biomarkers Prev 2006;15(11):2253–62)

Introduction

An aberrant crypt focus consists of aggregates of abnormal crypts that are larger in size than normal colonic crypts and have a thickened layer of epithelial cells, increased pericryptal space, and irregular (frequently oval or slit-like) lumens. ACF are elevated above adjacent normal colonic crypts and are routinely identified as more densely or less densely stained outcroppings on colonoscopy using methylene blue or indigo carmine staining, respectively.

ACF were originally described in methylene blue–stained whole-mount preparations of colons from carcinogen-treated mice and rats (1). Mutations in *K-ras* and *β-catenin* genes, changes in expression of several genes implicated in colorectal tumorigenesis (such as *APC, p53, c-fos*, and *c-Myc*; ref. 2), and other features characteristic of colon cancer, including altered mucin secretion (3) and microsatellite instability (4), have been found in rodent ACF. It has been shown that the frequency of ACF (and the number of crypts per focus) was increased by carcinogens in a dose- and time-dependent manner and decreased on treatment with agents known to inhibit dimethylhydrazine- or azoxymethane-induced colon cancer in rodents (reviewed in refs. 5, 6). Similarities in the characteristics of ACF and colon cancers allow consideration of ACF as the earliest preneoplastic colonic lesions, and rodent ACF are increasingly used as biomarkers for evaluation of carcinogenic and chemopreventive agents (7-11).

Human colonic ACF are similar to rodent ACF (12-14) and characterized by increased cell proliferation with some expansion of cell proliferative activity to the upper part of aberrant crypts (15-17). Most, if not all, ACF are monoclonal (18, 19) and frequently have mutations in K-ras (reviewed in refs. 20-22). Mutations in the *APC* gene have also been found in ACF but with low frequencies (23, 24). Other genetic and epigenetic alterations that are present in colon tumors have also been found in ACF: mutations in the gene *BRAF* (25); altered DNA fingerprints indicative of gross DNA instability (26); methylation of CpG islands in several genes implicated in colon cancer, particularly the *p16* gene (27); silencing by methylation of the tumor suppressor gene *SLC5A8* (28); and overexpression of carcinoembryonic antigen (29), P-cadherin (30), and E-cadherin (31). Interpretation of data on gene expression changes in human ACF is, however, complicated because gene expression was assessed mostly by immunohistochemical methods and therefore is rather qualitative. In addition, the pattern of altered gene expression usually varies throughout the aberrant crypts (from the base to the apex) and depends on ACF size (see, e.g., refs. 29-31).

The prevalence of ACF is correlated with colorectal cancer risk: the number of ACF is higher in the colons of patients with colorectal adenomas and carcinomas as compared with patients without cancer (17, 32-37), in colons of patients from a population with higher colorectal cancer incidence rate (38), in older patients (17, 39, 40), and in tobacco users (40). It has been found that the number of ACF is increased from proximal to distal colon (33, 34, 38, 39, 41, 42), corresponding to the known distribution of sporadic colorectal cancer. Thus, ACF

Received 9/8/05; revised 7/31/06; accepted 8/17/06.

Grant support: Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

O.K. Glebov and L.M. Rodriguez contributed equally to this work.

Requests for reprints: Ilan R. Kirsch, Oncology Research, Amgen, 1201 Amgen Court West, AW1-J4144, Seattle, WA 98119-3105. Phone: 206-265-7316; Fax: 206-216-5930. E-mail: lkirsch@amgen.com

doi:10.1158/1055-9965.EPI-05-0694

can be used as a potential biomarker for evaluation of colorectal cancer risk. It has been shown that treatment of patients with the nonsteroidal anti-inflammatory drugs sulindac (17) and aspirin (10) significantly reduces the number of ACF in the colon, even after short (2-10 months) treatments (43), indicating that ACF can be used as biomarkers for the identification of cancer preventive agents (44).

The monoclonal origin of ACF, the altered control of cell proliferation reflected in dysplasia, and the presence of gene mutations characteristic of colon neoplasia in human and rodent ACF support the role of ACF as the earliest precursors to colorectal cancer. The study of changes in gene expression in ACF may contribute to an understanding of the nature of ACF and the identification of genes of which the alteration in expression is related to cancer risk and cancer-preventive drug effects.

In this study, we wanted to explore genome-wide changes in gene expression that are characteristic of ACF. As a first step, we asked if ACF identified during routine optical colonoscopy and isolated as pinch biopsies can be distinguished from normal colonic mucosa (NM) by their pattern of gene expression.

Materials and Methods

Methods of RNA extraction and amplification, cDNA labeling and microarray hybridization, data acquisition, and statistical analysis were described in detail earlier (45) and will only be outlined here.

Patients and Biopsy Samples. Protocols used in the study were approved by Institutional Review Boards of the National Cancer Institute and National Naval Medical Center. Before enrollment, informed consent was obtained from all patients who were undergoing medically indicated screening or surveillance colonoscopy. ACF were identified after spraying indigo carmine (0.5%) onto the colonic mucosa during routine colonoscopy. Standard pinch biopsies of colonic mucosal areas containing ACF (with 10 to 50 crypts per focus) and, in addition, NM biopsies were taken, flash frozen in liquid nitrogen, and stored at −80°C. More than 80% of the cells in the biopsies were colonic crypt cells as determined by H&E staining. Matched ACF and NM samples from ascending and descending colons of 13 patients were selected for RNA isolation and array hybridization experiments (Table 1). In addition, matched pairs of ACF and NM samples from ascending or descending colons of six patients were used for validation purposes. All patients were free of malignancy at the time of biopsy collection. Five of 13 patients in the main set and three of six patients in the validation set had previous colorectal cancers. Cancer-related colon surgery was done more than 2 years before the taking of biopsies for all but one of the patients who had undergone colorectal surgery 1 year before the collection of biopsies.

RNA Extraction and Amplification. Total RNA was isolated from flash-frozen specimens (usually two standard ACF or NM pinch biopsies per one RNA sample), homogenized with a Disposable Generator and Micro-H Omni Homogenizer in lysis buffer RLT (Qiagen), and purified using Qiagen RNeasy Mini Kit columns (Qiagen) according to the instructions of the manufacturer. mRNA was amplified according to a modified Eberwine's protocol (46).

cDNA Labeling and Microarray Hybridization. Fluorescently labeled cDNA was synthesized by reverse transcription of amplified colon antisense RNA and human testis antisense RNA [prepared from total testis RNA (Clontech, Inc., Mountain View, CA) as described for colon antisense RNA] with random oligonucleotide primers in the presence of Cy3-dUTP or Cy5-dUTP (Amersham Pharmacia Biotech, Piscataway, NJ), respectively. For each hybridization experiment, 5 μg of colon antisense RNA and 6 μg of testis antisense RNA (common reference) were used to prepare a mixture of labeled cDNAs. Microarrays containing cDNA clones (from Research Genetics, Inc., Huntsville, AL) spotted on lysine-coated glass slides were obtained from the Advanced Technology Center. Detailed information on printed cDNA can be found on the mAdb web site.[3] Microarrays contain 9,128 sequence-verified cDNAs, among which 7,102 represent named genes and 1,179 EST clusters, including 8,556 cDNAs with UniGene cluster ID (mapping to 7,777 unique UniGene clusters).

Data Acquisition and Analysis. Microarrays were scanned with an Axon 4000 laser scanner and image analysis was done with GenePix Pro 3.0 Software (Axon Instruments, Inc., Sunnyvale, CA). Data analysis was done with the BRB-ArrayTools (version 3.2) software package developed by the Biometric Research Branch of the Division of Cancer Treatment and Diagnosis of the National Cancer Institute and The EMMES Corporation (Rockville, MD; ref. 47). All arrays used were printed on the same day (as a one printing batch); before statistical analysis, background intensities were subtracted and data were filtered for minimal spot intensity (100 units) in one of the two channels (if in the other channel, background-subtracted intensity was <100, it was adjusted to 100 units) and for missing values (not in >20% of arrays). Fluorescence intensity ratio data were log transformed and normalized by loess smoother.

The Class comparison and Class Prediction modules of BRB-ArrayTools were used to determine if the pattern of gene expression allowed discrimination of ACF and NM samples.

First, a paired t test was done on average differences in normalized gene expression log-ratios in pairs of samples (ACF and NM) from the same patients to select genes that showed univariately statistically significant differences ($P < 0.001$) in expression between ACF and NM samples. The multivariate class label permutation test was used to compute the number of false-positives among selected genes (48, 49).

Several multivariate classification methods available in BRB ArrayTools 3.2 (compound covariate predictor, diagonal linear discriminant analysis, nearest neighbor and nearest centroid predictors, and support vector machine) were used for classification of ACF and NM samples and gave comparable results, which will be illustrated with the Compound Covariate Predictor (CCP) and Support Vector Machine Predictor (SVMP) classifiers. CCP was calculated as a linear combination of log-ratio differences weighted by univariate t values (50). A positive sign was assigned to t-test values for genes that show higher log-ratios in NM and a negative sign for genes that show higher log-ratios in ACF. The CCP value was calculated for each pair of samples and, as a classification threshold, the sign (positive or negative) of the CCP was used for classifying. SVMP was calculated as a linear function of the log-ratio differences for genes selected in paired t test with weights estimated by linear kernel support vector machine algorithm to minimize the number of misclassified samples.

The misclassification rate was estimated by leave-one-out cross-validation. Specifically, a pair of ACF and NM samples (from one patient) was omitted and a CCP and a SVMP were developed from scratch using the remaining samples. The cross-validation procedure includes performing t tests for a selection of genes on arrays for which differences in expression are significant at a stated P value for remaining samples, recalculating the CCP, SVMP, and classification threshold for the remaining pairs of samples, and applying the new CCP, SVMP, and threshold values to classify the omitted pair of samples. This was done independently for each omitted pair of

[3] http://nciarray.nci.nih.gov/.

Table 1. Characteristics of study patients

Patient ID	Age (y)	Gender	Previous cancer	No. ACF		No. crypts per focus*		RNA analyzed by array hybridization	RNA analyzed by RT-PCR
				Ascending colon	Descending colon	Ascending colon	Descending colon		
Main set of patients									
62901	74	M	Yes	1	9	25	18	Yes	
71301	70	M	Yes	2	17	32	14	Yes	
81701	58	F	No	1	19	12	24	Yes	
101201	41	F	No	1	8	25	16	Yes	
110901	64	F	No	2	8	25	13	Yes	
113001	75	F	Yes	1	13	30	21	Yes	
11102	38	F	No	2	10	16	20	Yes	
22202	63	F	Yes	1	2	30	38	Yes	Yes
40302	69	M	Yes	1	14	30	16	Yes	Yes
50802	46	M	No	3	7	30	14	Yes	Yes
08902	58	M	No	1	24	30	16	Yes	Yes
081402	55	F	No	2	8	35	14	Yes	Yes
082302	45	F	No	2	7	21	19	Yes	Yes
Patients whose RNAs were selected for data validation									
82401[a]	66	M	No	1	16	20	16	Yes[c]	Yes
91401[b]	62	M	No	2	3	22	37		Yes
111601[b]	76	M	Yes	1	1	16	15		Yes
122101[a]	75	F	Yes	0	1		50	Yes[c]	Yes
21302[a]	57	M	Yes	0	12		18	Yes[c]	Yes
32202[a]	71	M	No	0	18		24		Yes

NOTE: For the main set of 13 patients, RNAs were isolated from ACF and NM biopsies from ascending and descending colon. For the set of six patients selected for validation, RNAs were isolated from ACF and NM biopsies either from descending colon ([a]) or ascending colon ([b]). After validation of real-time PCR data, ACF and NM RNAs from three patients ([c]) were amplified and used for array hybridization.
*Average number for patients with more than one ACF.

samples. The ratio of pairs of samples correctly classified in cross-validation to the total number of sample pairs yields the misclassification rate. Permutation P value for a classifier was calculated by performing 2,000 random permutations of class labels and repeating the cross-validation procedure for each permutation. The proportion of random permutations that gave the same or smaller misclassification rate as was obtained with the true class labels is presented as a (permutation) P value for the classifier, and a value of $P < 0.0005$ was reported when no random permutation of the class label was found out of 2,000 permutations with the same or smaller misclassification rate as for the true class labeling. The effects of multiple factors on gene expression were estimated by ANOVA for fixed and mixed effect linear models, which was done as implemented in R-Plugins of BRB-ArrayTools.

Real-time PCR. One-step TaqMan real-time reverse transcription-PCR (RT-PCR) was done to study expression of several cDNAs (genes) using an ABI Prism 7700 Sequence Detection System. Primers and hybridization FAM-labeled probes were selected with PrimerExpress software (Applied Biosystems, Foster City, CA) by using complete cDNA sequences that have the same UniGene cluster ID as cDNAs printed on array (Supplementary Table S1). TaqMan Gold RT-PCR kit (Applied Biosystems) and the protocol of the manufacturer (30 minutes at 48°C for RT reaction, 10 minutes at 95°C for activation of TaqGold Polymerase, and 40 cycles consisting of 15 seconds at 95°C and 1 minute at 60°C) were used, and 20 ng (for gene-specific PCR) or 5 ng (for ribosomal 18S RNA–specific PCR) of total RNA were assayed in 25 μL of one-step RT-PCR reaction mixture with gene-specific or ribosomal 18S RNA–specific primers and probes, in triplicates for each sample and each gene. Serial dilutions of a mixture of colon total RNAs from several other colon biopsies were used to calculate PCR efficiency for a gene in the range of 5 to 40 ng of total RNA per reaction. Gene expression was normalized to the amount of ribosomal 18S RNA in a sample, for which RT-PCR reactions were done on the same 96-well plate in separate wells, together with serial dilutions of colon total RNA (from 1.25 to 10 ng RNA per reaction) to calculate PCR efficiency for ribosomal 18S RNA. Data were analyzed, and normalized (by ribosomal 18S RNA quantity) gene expression was calculated with Q-Gene software (51).

Results

Among patients undergoing screening or surveillance colonoscopy, we have selected 13 patients who have ACF in ascending as well as descending colon (Table 1). As might be expected (33, 34, 39, 41), the number of ACF found in the descending colon (11.23 ± 1.653, $N = 13$) of these patients is higher than the number of ACF in the ascending colon (1.54 ± 0.1831, $N = 13$; $t = 5.827$, $P < 0.0001$). However, the number of crypts per focus in the ascending colon (26.05 ± 2.370, $N = 20$) is higher than in the descending colon (17.60 ± 0.8479, $N = 146$; $t = 3.447$, $P = 0.007$). The tendency of ACF from ascending colon to be larger than the ACF from the descending colon was noted before (38). Addition of patients selected for validation into analysis does not change the conclusion that there is prevalence of ACF in descending compared with ascending colon, although the size of ACF (number of crypts per ACF) in ascending colon is bigger than in the descending colon in studied patients.

In a combined set of 19 patients, eight patients had previous colorectal cancer. There are no statistically significant differences in number or size of ACF in ascending and descending colon of patients with previous colorectal cancer compared with patients without cancer. Matched pairs of ACF and NM samples from ascending and descending colons of 13 patients were analyzed on a total of 52 arrays, with each sample analyzed on one array. The effect on gene expression of various factors, including the tissue type (ACF or NM), biopsy location (ascending or descending colon), history of previous colon cancer, gender, and age of patient, was studied by ANOVA using linear models with only fixed effects (blocked by array ID or patient ID) or mixed effect models where array ID or patient ID was treated as a random factor. All of the listed factors affect gene expression, with biopsy location being the most dominant factor followed by effect of individual

between patient variations. However, in this study, only two factors are balanced: tissue type and biopsy location, whereas other factors are unbalanced, and comparisons of gene expression can be biased by confounding factors. For example, in a main group of 13 patients, there is a higher ratio of male patients with previous colon cancer than of female patients, and patients with previous colon cancer are older (average age, 70 years) than patients without previous cancer (51 years). Given the small number of patients (13) and possible confounding factor effects, we restricted our analysis to comparison of gene expression in ACF versus NM.

ACF and NM samples can be successfully classified by their gene expression profiles without regard to their origin from the ascending or descending colon (data available on request). However, the patterns of gene expression in NM from ascending (right) and descending (left) colon are different (ref. 45, and see below). Therefore, we believe that it is more appropriate to compare gene expression in ACF and NM samples by paired tests that take into consideration correlation in gene expression in ACF and NM from the same side of the colon [e.g., ACF and NM from the right colon (ACFR versus NMR) and ACF and NM from the left colon (ACFL versus NML), resulting in two pairs of samples per patient].

Thirty-four genes show differential expression at a univariate $P < 0.001$ in ACF and NM on analysis of 26 pairs of ACF and NM according to a paired t test (Table 2). The probability of getting by chance 34 differentially expressed genes out of 8,983 genes (which passed the filtering criteria) is 0.015 as calculated by a multivariate class-label permutation test. Twenty-five of 26 and 24 of 26 sample pairs were correctly classified in leave-one-out cross-validation using the CCP or SVMP. Random permutations of class labels showed that the probability of getting by chance as small a cross-validated misclassification rate as that obtained with the true class labels (4% for CCP or 8% for SVMP) is less than 5×10^{-4} and 0.001 for CCP and SVMP, respectively.

In a separate analysis of ACF and NM samples from the ascending colon (13 pairs), 46 genes show differential expression at $P < 0.001$ (Table 3). As calculated by multivariate class-label permutation test, the probability of getting by chance 46 differentially expressed genes out of 8,983 genes is <0.001. All 13 pairs of ACFR and NMR were correctly classified in leave-one-out cross-validation using CCP and SVMP, with probabilities of getting by chance a 0% misclassification rate 0.005 and 0.004, respectively. Similar results were obtained in an analysis of the 13 pairs of ACF and NM samples from the descending colon. Forty-two genes show differential expression at $P < 0.001$ (Table 4; probability of getting by chance 42 differentially expressed genes out of 8,983 genes is 0.006), and 12 of 13 and 13 of 13 sample pairs were correctly classified in

Table 2. Genes differentially expressed ($P < 0.001$) in ACF versus NM

Clone	UG cluster	Gene symbol	Gene*	Ratio NM/ACF†	Parametric P
IncytePD:489032	Hs.250712	*CACNB3*	Calcium channel, voltage dependent, β3 subunit	0.324	1.00e−07
IncytePD:1628341	Hs.55279	*SERPINB5*	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5	0.27	1.00e−07
IncytePD:2132487	Hs.171480	*REG4*	Regenerating islet-derived family, member 4	0.391	1.70e−06
IncytePD:460034	Hs.55279	*SERPINB5*	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5	0.465	7.20e−06
IncytePD:2825369	Hs.494261	*PSAT1*	Phosphoserine aminotransferase 1	0.652	1.53e−05
IncytePD:65463	Hs.148330	*ARF4*	ADP-ribosylation factor 4	0.786	4.62e−05
IncytePD:2344817	Hs.444948		Transcribed locus	1.382	4.68e−05
IncytePD:2595612	Hs.523395	*MUC5B*	Mucin 5, subtype B, tracheobronchial	0.552	5.68e−05
IncytePD:2060823	Hs.2962	*S100P*	S100 calcium binding protein P	0.505	8.45e−05
IncytePD:1699587	Hs.2256	*MMP7*	Matrix metalloproteinase 7 (matrilysin, uterine)	0.519	0.000108
IncytePD:2906971	Hs.551523	*C3F*	Putative protein similar to nessy (*Drosophila*)	1.287	0.000111
IncytePD:2622181	Hs.21160	*ME1*	Malic enzyme 1, NADP(+) dependent, cytosolic	0.61	0.000118
IncytePD:1527755	Hs.436973	*ZNF516*	Zinc finger protein 516	1.381	0.000130
IncytePD:1600442	Hs.516032	*HADHA*	Hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), α subunit	1.262	0.000134
IncytePD:1960889	Hs.93675	*C10orf10*	Chromosome 10 open reading frame 10	0.715	0.000182
IncytePD:1402127	Hs.497674	*LPGAT1*	Lysophosphatidylglycerol acyltransferase 1	0.749	0.000202
IncytePD:1384190	Hs.474982	*ACO2*	Aconitase 2, mitochondrial	1.238	0.000203
IncytePD:972390	Hs.21160	*ME1*	Malic enzyme 1, NADP(+)-dependent, cytosolic	0.673	0.000211
IncytePD:4244645	Hs.113227	*DAO*	D-Amino-acid oxidase	1.51	0.000235
IncytePD:241643	Hs.461285	*ATBF1*	AT-binding transcription factor 1	1.289	0.000257
IncytePD:1968413	Hs.504115	*TRIM29*	Tripartite motif-containing 29	0.377	0.000285
IncytePD:2917432	Hs.317593	*C13orf11*	Chromosome 13 open reading frame 11	0.748	0.000295
IncytePD:2364392	Hs.91521	*DKFZP761M1511*	Hypothetical protein DKFZP761M1511	1.3	0.000318
IncytePD:1300835	Hs.523848	*MYEOV*	Myeloma overexpressed gene (in a subset of t(11;14) positive multiple myelomas)	0.622	0.000356
IncytePD:1417443	Hs.380277	*DAPK1*	Death-associated protein kinase 1	0.747	0.000359
IncytePD:1910469	Hs.418123	*CTSL*	Cathepsin L	1.241	0.000463
IncytePD:182802	Hs.364941	*HSD3B1*	Hydroxy-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 1	1.738	0.000501
IncytePD:1707025	Hs.83313	*C7orf36*	Chromosome 7 open reading frame 36	0.783	0.000589
IncytePD:1919233	Hs.300887	*ELYS*	ELYS transcription factor-like protein TMBS62	0.803	0.000619
IncytePD:1511120	Hs.189641	*SEC24D*	SEC24 related gene family, member D (*S. cerevisiae*)	0.719	0.000653
IncytePD:2258791	Hs.465607	*APBA3*	Amyloid β (A4) precursor protein-binding, family A, member 3 (X11-like 2)	1.186	0.000736
IncytePD:2025677	Hs.93842	*STARD4*	START domain containing 4, sterol regulated	0.771	0.000756
IncytePD:1626232	Hs.546367	*SPINK4*	Serine protease inhibitor, Kazal type 4	0.553	0.000838
IncytePD:942100	Hs.825	*HSD3B2*	Hydroxy-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 2	1.645	0.000902

*Genes are ordered by P values.
†Geometric mean ratio.

Table 3. Genes differentially expressed ($P < 0.001$) in ACFR versus NMR

Clone	UG cluster	Gene symbol	Gene*	Ratio NMR/ACFR†	Parametric *P*
IncytePD:2015871	Hs.471034	*ELA3A*	Elastase 3A, pancreatic (protease E)	2.309	0.000004
IncytePD:1966455	Hs.519884	*GCNT2*	Glucosaminyl (*N*-acetyl) transferase 2, I-branching enzyme	1.689	0.000005
IncytePD:182802	Hs.364941	*HSD3B1*	Hydroxy-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 1	2.974	0.000008
IncytePD:2595728	Hs.511872	*CYP2C18*	Cytochrome *P*450, family 2, subfamily C, polypeptide 18	2.119	0.000009
IncytePD:2718565	Hs.240056	*ETNK1*	Ethanolamine kinase 1	2.013	0.000014
IncytePD:1628341	Hs.55279	*SERPINB5*	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5	0.223	0.000015
IncytePD:2234609	Hs.194777	*MEP1B*	Meprin A, β	2.053	0.000022
IncytePD:489032	Hs.250712	*CACNB3*	Calcium channel, voltage dependent, β3 subunit	0.348	0.000034
IncytePD:238333	Hs.519884	*GCNT2*	Glucosaminyl (*N*-acetyl) transferase 2, I-branching enzyme	1.565	0.000036
IncytePD:2906971	Hs.551523	*C3F*	Putative protein similar to nessy (*Drosophila*)	1.512	0.000048
IncytePD:1985104	Hs.3416	*ADFP*	Adipose differentiation-related protein	1.615	0.000066
IncytePD:2317638	Hs.125039	*CROT*	Carnitine *O*-octanoyltransferase	1.498	0.000068
IncytePD:4073012	Hs.112916	*SLC3A1*	Solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1	1.627	0.000095
IncytePD:2344817	Hs.444948		Transcribed locus	1.548	0.000099
IncytePD:2825369	Hs.494261	*PSAT1*	Phosphoserine aminotransferase 1	0.557	0.000105
IncytePD:2055569	Hs.372914	*NDRG1*	N-myc downstream regulated gene 1	1.548	0.000169
IncytePD:3172982	Hs.486228	*C6orf4*	DKFZP586G0522 protein	1.636	0.000248
IncytePD:2364392	Hs.91521	*DKFZP761M1511*	Hypothetical protein DKFZP761M1511	1.467	0.000269
IncytePD:1806219	Hs.32966	*GUCA2B*	Guanylate cyclase activator 2B (uroguanylin)	2.057	0.000336
IncytePD:1443766	Hs.482390	*TGFBR3*	Transforming growth factor, β receptor III (β-glycan, 300 kDa)	0.54	0.000377
IncytePD:80275	Hs.460019	*ERCC4*	Excision repair cross-complementing rodent repair deficiency, complementation group 4	1.458	0.000435
IncytePD:1798585	Hs.464137	*ACOX1*	Acyl-CoA oxidase 1, palmitoyl	1.453	0.000439
IncytePD:4173045	Hs.432898	*RPL4*	Ribosomal protein L4	1.528	0.000451
IncytePD:1963819	Hs.195080	*ECE1*	Endothelin converting enzyme 1	1.425	0.000461
IncytePD:168865	Hs.282871	*CYP2C8*	Cytochrome *P*450, family 2, subfamily C, polypeptide 8	1.746	0.000463
IncytePD:942100	Hs.825	*HSD3B2*	Hydroxy-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 2	2.384	0.000498
IncytePD:2228063	Hs.98547	*ACCN3*	Amiloride-sensitive cation channel 3	1.975	0.000523
IncytePD:1911142	Hs.247362	*DDAH2*	Dimethylarginine dimethylaminohydrolase 2	1.423	0.000578
IncytePD:2796468	Hs.529117	*CYP2B7P1*	Cytochrome *P*450, family 2, subfamily B, polypeptide 7 pseudogene 1	2.067	0.000604
IncytePD:1879811	Hs.517581	*HMOX1*	Heme oxygenase (decycling) 1	1.5	0.000610
IncytePD:778212	Hs.446077	*SLC38A4*	Solute carrier family 38, member 4	1.534	0.000611
IncytePD:3070110	Hs.28309	*UGDH*	UDP-glucose dehydrogenase	1.536	0.000636
IncytePD:4544094	Hs.151710	*PDE6A*	Phosphodiesterase 6A, cyclic guanosine 3′,5′-monophosphate specific, rod, α	1.799	0.000639
IncytePD:645584	Hs.435036	*GPC3*	Glypican 3	2.082	0.000670
IncytePD:195142	Hs.282624	*CYP2C9*	Cytochrome *P*450, family 2, subfamily C, polypeptide 9	1.906	0.000672
IncytePD:567292	Hs.463439	*SPAG9*	Sperm-associated antigen 9	1.496	0.000696
IncytePD:1600442	Hs.516032	*HADHA*	Hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), α subunit	1.395	0.000704
IncytePD:1699587	Hs.2256	*MMP7*	Matrix metalloproteinase 7 (matrilysin, uterine)	0.38	0.000727
IncytePD:2127868	Hs.429879	*EHHADH*	Enoyl-CoA, hydratase/3-hydroxyacyl CoA dehydrogenase	1.759	0.000750
IncytePD:1813156	Hs.37558	*RFK*	Riboflavin kinase	1.373	0.000753
IncytePD:5033671	Hs.498514	*AKR1C4*	Aldo-keto reductase family 1, member C4 (chlordecone reductase; 3α-hydroxysteroid dehydrogenase, type I; dihydrodiol dehydrogenase 4)	1.403	0.000754
IncytePD:4073867	Hs.498732	*PHYH*	Phytanoyl-CoA hydroxylase (Refsum disease)	1.339	0.000837
IncytePD:1698889	Hs.468415	*PIGF*	Phosphatidylinositol glycan, class F	0.687	0.000839
IncytePD:1511120	Hs.189641	*SEC24D*	SEC24 related gene family, member D (*S. cerevisiae*)	0.58	0.000964
IncytePD:2912830	Hs.32949	*DEFB1*	Defensin, β1	1.838	0.000973
IncytePD:1818744	Hs.516700	*CYP27A1*	Cytochrome *P*450, family 27, subfamily A, polypeptide 1	1.583	0.000983

*Genes are ordered by *P* values.
†Geometric mean ratio.

leave-one-out cross-validation using CCP and SVMP, respectively. Random permutations of class labels showed that the probability of getting by chance 8% and 0% misclassification rate is 0.036 and 0.007 for CCP and SVMP, respectively.

To check if the differences in gene expression defined by array hybridization can be used for predictive classification of ACF and NM samples, we applied classifiers developed on the basis of the 26 pairs of ACF and NM, as well as on the basis of

the 13 pairs of ACFL and NML, to array data for newly obtained ACF and NM samples from descending colons of three patients that were not included in the main group of 13 patients. All six samples (three ACFL and three NML) were correctly classified in both cases.

Thus, ACF can be distinguished from NM regardless of the site of origin based on common differences in gene expression. In addition, ACF from the ascending colon can be distinguished from the normal ascending colonic mucosa, and ACF from the descending colon can be distinguished from the normal descending colonic mucosa.

To validate data on differential gene expression in ACF and NM, we selected eight genes for real-time RT-PCR analysis. Twelve pairs of RNAs from ACF and NM samples from the ascending and descending colons of six patients (that were analyzed by array hybridization) were analyzed by real-time RT-PCR (data for three genes are presented in Supplementary Figs. S1 and S2). In general, there is good agreement between array and real-time RT-PCR data on differences in expression of the eight selected genes in ACF and NM: Spearman rank correlation coefficient (r) for ratios of mean normalized gene expression values obtained by real-time RT-PCR and geometric mean ratios from array data are 0.833, 0.762, and 0.857 for ACF/NM, ACFR/NMR, and ACFL/NML log 2 ratios. Perhaps due to the high individual variability in gene expression and the low number of samples, only some of the eight genes show statistically significant differences (paired Wilcoxon signed rank test, $P < 0.05$) in expression between ACF and NM when their expression is analyzed individually, although all but one display a trend in differential expression expected from the array data (Table 5).

In addition, real-time RT-PCR data can also be used to identify ACF and NM: all 12 pairs of ACF and NM samples can be correctly classified on the basis of overexpression of at least two of three genes (*REG4*, *SRPN-B5*, and *TRIM29*) in ACF when compared with NM from the same location and the same

Table 4. Genes differentially expressed ($P < 0.001$) in ACFL versus NML

Clone	UG cluster	Gene symbol	Gene*	Ratio NML/ACFL†	Parametric P
IncytePD:2132487	Hs.171480	*REG4*	Regenerating islet-derived family, member 4	0.291	0.000001
IncytePD:489032	Hs.250712	*CACNB3*	Calcium channel, voltage dependent, β3 subunit	0.303	0.000002
IncytePD:2622181	Hs.21160	*ME1*	Malic enzyme 1, NADP(+)-dependent, cytosolic	0.531	0.000007
IncytePD:1628341	Hs.55279	*SERPINB5*	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5	0.327	0.000010
IncytePD:2171401	Hs.525339	*ERO1L*	ERO1-like (*S. cerevisiae*)	0.623	0.000039
IncytePD:2595612	Hs.523395	*MUC5B*	Mucin 5, subtype B, tracheobronchial	0.527	0.000082
IncytePD:1903267	Hs.508113	*OLFM4*	Olfactomedin 4	0.58	0.000105
IncytePD:1626232	Hs.546367	*SPINK4*	Serine protease inhibitor, Kazal type 4	0.417	0.000106
IncytePD:1300835	Hs.523848	*MYEOV*	Myeloma overexpressed gene (in a subset of t(11;14) positive multiple myelomas)	0.464	0.000142
IncytePD:2054678	Hs.18844	*PCSK9*	Proprotein convertase subtilisin/kexin type 9	0.374	0.000143
IncytePD:1661184	Hs.113094	*CORO2A*	Coronin, actin binding protein, 2A	0.739	0.000152
IncytePD:740878	Hs.417962	*DUSP4*	Dual specificity phosphatase 4	0.51	0.000160
IncytePD:1252644	Hs.105269	*SC4MOL*	Sterol-C4-methyl oxidase-like	0.679	0.000177
IncytePD:460034	Hs.55279	*SERPINB5*	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5	0.513	0.000190
IncytePD:2912879	Hs.356076	*BIRC4*	Apoptosis inhibitor 3	0.669	0.000204
IncytePD:1707025	Hs.83313	*C7orf36*	Chromosome 7 open reading frame 36	0.722	0.000273
IncytePD:1922164	Hs.13291	*CCNG2*	Cyclin G2	0.62	0.000315
IncytePD:1421929	Hs.191842	*CDH3*	Cadherin 3, type 1, P-cadherin (placental)	0.607	0.000345
IncytePD:927392	Hs.368077	*SERPINB8*	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 8	0.626	0.000350
IncytePD:2771046	Hs.1239	*ANPEP*	Alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)	0.504	0.000359
IncytePD:2079317	Hs.494538	*PTCH*	Patched homologue (*Drosophila*)	1.377	0.000381
IncytePD:1308333	Hs.370858	*FUCA1*	Fucosidase, α-L-1, tissue	1.3	0.000396
IncytePD:797919	Hs.65641	*SAMD9*	Sterile α motif domain containing 9	0.625	0.000400
IncytePD:1963245	Hs.26225	*GABRP*	γ-Aminobutyric acid A receptor, π	1.448	0.000404
IncytePD:2734139	Hs.83465	*HOXD1*	Homeobox D1	0.586	0.000427
IncytePD:1811583	Hs.193725	*PSMD5*	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 5	0.751	0.000429
IncytePD:1900031	Hs.491582	*PLAT*	Plasminogen activator, tissue	1.407	0.000434
IncytePD:1402127	Hs.497674	*LPGAT1*	Lysophosphatidylglycerol acyltransferase 1	0.693	0.000457
IncytePD:2082211	Hs.515223		Homo sapiens, similar to unnamed HERV-H protein, clone IMAGE:3996038, mRNA	0.564	0.000485
IncytePD:1960889	Hs.93675	*C10orf10*	Chromosome 10 open reading frame 10	0.661	0.000532
IncytePD:519653	Hs.81134	*IL1RN*	Interleukin-1 receptor antagonist	0.47	0.000543
IncytePD:2515733	Hs.106242	*CYP4F3*	Cytochrome *P*450, family 4, subfamily F, polypeptide 3	0.61	0.000552
IncytePD:1368173	Hs.75160	*PFKM*	Phosphofructokinase, muscle	1.317	0.000604
IncytePD:1479255	Hs.208544	*KCNK1*	Potassium channel, subfamily K, member 1	0.734	0.000659
IncytePD:2900277	Hs.532634	*IFI27*	IFN, α-inducible protein 27	0.658	0.000677
IncytePD:1942550	Hs.1706	*ISGF3G*	IFN-stimulated transcription factor 3, γ 48 kDa	0.728	0.000680
IncytePD:1968413	Hs.504115	*TRIM29*	Tripartite motif-containing 29	0.375	0.000698
IncytePD:674211	Hs.369819	*TBC1D16*	TBC1 domain family, member 16	1.414	0.000711
IncytePD:1803721	Hs.318894	*GPR126*	G protein-coupled receptor 126	0.629	0.000718
IncytePD:2308302	Hs.152983	*HUS1*	HUS1 checkpoint homologue (*S. pombe*)	0.74	0.000795
IncytePD:1798209	Hs.76224	*EFEMP1*	Epidermal growth factor–containing fibulin-like extracellular matrix protein 1	1.362	0.000839
IncytePD:1635864	Hs.125715	*MBNL2*	Muscleblind-like 2 (*Drosophila*)	0.752	0.000907

*Genes are ordered by P values.
†Geometric mean ratio.

Table 5. Comparison of array and real-time RT-PCR data on differential gene expression

Clone ID	Gene	UniGene ID	Ratio ACF/NM		Ratio ACFR/NMR		Ratio ACFL/NML	
			Arrays*	RT-PCR†	Arrays*	RT-PCR†	Arrays*	RT-PCR†
IncytePD:2079317	patched homologue (*Drosophila*)	Hs.159526	0.956	0.964	1.259	1.413	**0.726**	0.657
IncytePD:489032	calcium channel, voltage dependent, β3 subunit	Hs.250712	**3.086**	1.010	**2.874**	1.058	**3.300**	0.954
IncytePD:2132487	regenerating islet-derived family, member 4	Hs.105484	**2.558**	***19.020***	1.908	*31.470*	**3.436**	***11.490***
IncytePD:1628341	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5	Hs.55279	**3.704**	***7.054***	**4.484**	***12.910***	**3.058**	***3.853***
IncytePD:1968413	tripartite motif-containing 29	Hs.82237	**2.653**	***4.906***	2.632	*9.969*	**2.667**	*2.416*
IncytePD:1910469	cathepsin L	Hs.418123	**0.806**	*0.896*	0.803	1.029	*0.808*	*0.780*
IncytePD:1845046	ephrin-A1	Hs.399713	0.795	0.812	*0.573*	0.701	1.133	0.940
IncytePD:182802	hydroxy-δ-5-steroid dehydrogenase, 3β- and steroid δ-isomerase 1	Hs.364941	**0.575**	***0.490***	**0.336**	***0.508***	0.985	0.473

NOTE: Genes that are identified in array analysis as differentially expressed at $P < 0.001$ or $P < 0.005$ are shown in bold and italic, respectively. Genes that are found to be differentially expressed (paired Wilcoxon signed rank test) at $P < 0.05$ and $P < 0.07$ in real-time PCR are shown in bold italic and italic, respectively.

*Geometric mean ratios; ACF/NM, 26 pairs of ACF and NM; ACFR/NMR and ACFL/NML, 13 pairs of ACF and NM from ascending and descending colon, respectively.

†Geometric mean ratios; ACF/NM, 12 pairs of ACF and NM; ACFR/NMR and ACFL/NML, 6 pairs of ACF and NM from ascending and descending colon, respectively.

patient (Supplementary Figs. S1 and S2). To confirm this observation, we selected six pairs of ACF and NM samples (that had not been analyzed previously by array hybridization) from the ascending (two pairs) and the descending (four pairs) colons of six patients. As shown in Fig. 1, all samples can be correctly classified by the pattern of expression of *REG4*, *SRPN-B5*, and *TRIM29* genes: in all six pairs of samples, expression of each of the three genes is at least twice higher in ACF than in the corresponding matched NM.

Thus, there is a good correlation for data on differences in gene expression in ACF and NM obtained in array hybridization and in real-time RT-PCR experiments, and real-time RT-PCR data can be used to correctly classify ACF and NM.

A paired *t* test identified 1,340 genes that showed statistically significant (at a univariate $P < 0.001$) differences in expression in ascending versus descending NM (Supplementary Table S2), confirming our previous results obtained on analysis of a different patient group (45). Differences in expression of 312 genes were also found to be significant at $P < 0.001$ in a paired *t* test between ACF from ascending versus descending colon (Supplementary Table S3). However, the observed differences in gene expression between ACFR and ACFL likely reflect for the most part the presence of NM cells in pinch biopsies of ACF. Most (270) of the 312 genes are present in the list of genes differentially expressed in NMR and NML samples, being overexpressed or underexpressed in both NMR relative to NML and ACFR relative to ACFL, and the magnitude of differences in expression of the 270 genes is lower in ACFR versus ACFL compared with NMR versus NML. In addition, when genes differentially expressed between NMR and NML are omitted from the analysis, ACFR and ACFL can no longer be successfully distinguished in leave-one-out cross-validation using CCP or SVMP (Supplementary Table S4).

Thus, the comparisons of differential gene expression in ACFR versus ACFL and NMR versus NML show that ACF are losing differences in gene expression characteristic for NM from the ascending and the descending colon. Most of the differential gene expression observed between ACFR and ACFL is not acquired *de novo* but is related to the differential gene expression between NM from the ascending versus the descending colon, and this can be a reflection of the presence of NM in ACF samples.

We have used the Ingenuity Pathway Analysis (IPA)[4] program to map genes affected in ACF to categories (''High Level Functions'') and known gene networks available in the IPA database. In this analysis, we have used lists of genes that show differences in expression between ACF and NM samples at $P < 0.005$ (available on request), as well as at $P < 0.001$ (Tables 2-4). The Fisher exact probability test is used in IPA to identify categories containing more affected genes than may be expected by chance given the genes present on the arrays. According to IPA global function analysis, there is enrichment in ACF of genes implicated in cancer, lipid metabolism, cell death, tumor and tissue morphology, cell growth and proliferation, and other categories (Supplementary Table S5). Genes affected in ACF are mapped with scores >10 to twelve known gene networks that control cell growth and proliferation, cell death and cell movement, cellular and tissue morphology, and hematologic, skeletal, and muscular system development and function (Supplementary Table S6; the score is the anti-logarithm of the probability of getting by chance a number of genes equal to or greater than the actual number of affected genes in a network, given the genes present on the array; a score of 2 means $P = 0.01$).

Discussion

The histopathologic identification and classification of aberrant crypt foci has not achieved a consensus standard and is the subject of a variety of distinct classification schemes reported in the literature (12-14, 52-56). Classification is complicated by the spectrum of foci that include appearances only subtly distinguishable from NM at one end of the histopathologic spectrum and indistinguishable from frank adenomas at the other end. We chose in this study to approach ACF beginning with the modality that is the most likely to have an immediate clinical effect, optical colonoscopy.

We thus applied indigo carmine staining during routine colonoscopic surveillance examinations and isolated ACF containing 10 to 50 aberrant crypts as standard pinch biopsies. We can now conclude that these colonoscopically distinct foci are distinguishable from the NM from which they arose on the basis of their pattern of gene expression. A recent paper (57) studied gene expression profiles within azoxymethane-induced ACF derived from a high-risk versus a low-risk inbred strain of mice. Whereas some of the specific genes that distinguished these two strains overlap with the broadly defined pathways that we found to distinguish NM from ACF,

[4] http://analysis.ingenuity.com/pa.

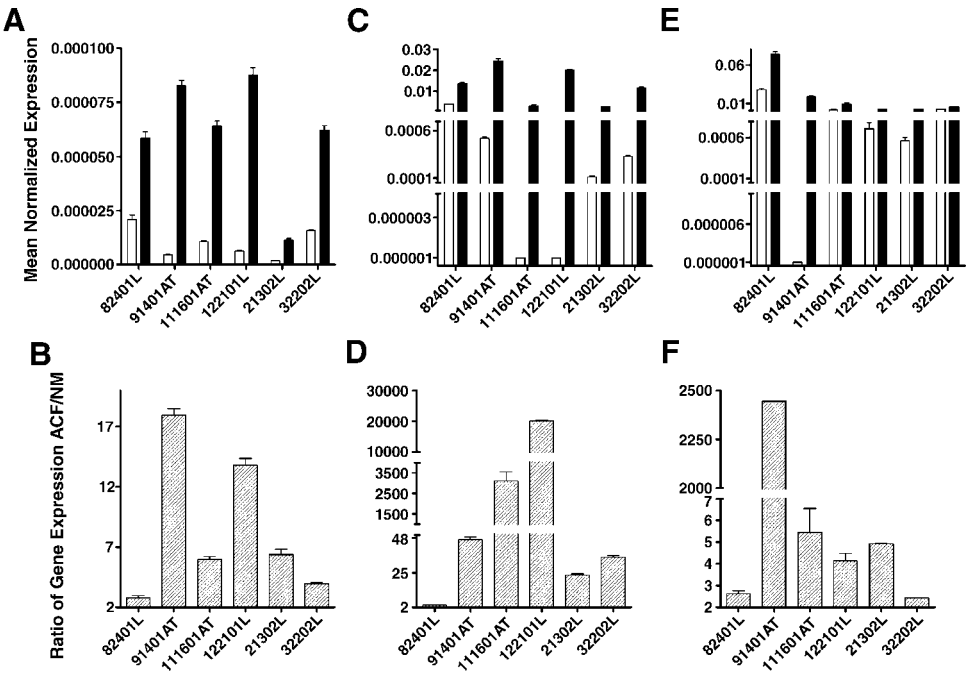

Figure 1. Real-time RT-PCR analysis of *SRPN-B5* (**A** and **B**), *REG4* (**C** and **D**), and *TRIM* (**E** and **F**) gene expression in ACF and NM isolated from ascending and descending colon of six patients from the validation group. *Abscissa,* patient ID; *ordinate, a, c,* and *e,* mean normalized expression; *b, d,* and *f,* ratio of gene expression in ACF to NM. Mean normalized gene expression in pairs of ACF (*open columns*) and NM (*filled columns*) and ratio of gene expression in ACF to corresponding NM (*crossed columns*); *bars,* SE.

these pathways are quite broad in their definition and we could not pinpoint a specific gene or genes that were of particular relevance in this comparison.

Changes in Gene Expression in ACF from Ascending and Descending Colon Are Directed toward Acquiring a Common Pattern of Gene Expression. Correct classification of ACF and NM samples can be achieved using gene expression differences between ACF and NM samples combined from the ascending and descending colon. Classification can be also achieved for ACF versus NM from either the ascending or the descending colon. However, the lists of genes emerging in those classifiers are different. Only 11 and 12 of the 34 genes that are found differentially expressed at $P < 0.001$ in a comparison of the combined set of 26 pairs of ACF and NM samples (i.e., genes that show differences in expression common between all ACF and NM from both ascending and descending colon) show differential expression at $P < 0.001$ in separate analyses of ACF and NM samples from the ascending or the descending colon, respectively. The reason for this is mainly the decrease in the number of sample pairs from 26 to 13 pairs. The differences in expression for most of these 34 genes are significant at a less stringent P value ($P < 0.05$) in separate comparisons of ACFR versus NMR (29 genes) and ACFL versus NML (28 genes), with 23 genes being present in both lists of significant genes.

As previously shown (45), the ascending and descending colon can be considered as two separate entities on the basis of their gene expression pattern. We observed the tendency of ACF from the ascending and the descending colon to converge toward similar gene expression profiles. Most of the genes differentially expressed in ascending versus descending NM, of which the expression is changed in ACFR or ACFL, show smaller or no differences in expression in ACFR versus ACFL. Among 46 genes that are differentially expressed in ACFR versus NMR (Table 3), 40 genes are also expressed differentially at $P < 0.001$ in NM samples from ascending versus descending colon (Supplementary Table S2). Five genes of which the expression is increased in ACFR relative to NMR are underexpressed in NMR relative to NML, and 35 genes that are underexpressed in ACFR relative to NMR are overexpressed in NMR relative to NML. Fourteen of 42 genes differentially expressed in ACFL versus NML (Table 4) also show differential expression at $P < 0.001$ in NM from ascending and descending colon (Supplementary Table S2). Eleven of the 14 genes follow the above-mentioned trend: seven genes overexpressed in ACFL are underexpressed in NML relative to NMR and four genes underexpressed in ACFL are overexpressed in NML relative to NMR.

Many Genes Affected in ACF Are Connected into Networks by *MYC, FOS/JUN*, and *TP53* Oncogenes. Eight of the 12 gene networks, to which genes differentially expressed in ACF ($P < 0.005$) have been mapped with a score >10, are linked by the *MYC, FOS/JUN*, and/or *TP53* genes (Supplementary Table S6). Although it indicates the importance of this particular subset of gene networks for ACF emergence and/or maintenance, the possible role of *MYC, AP1*, and *TP53* proteins by themselves remains unclear. For example, the *MYC* gene is overexpressed in ACFR compared with NMR (1.7-fold; $P = 0.002$), and expression of several genes known to be directly controlled by MYC protein is changed in ACFR in the predicted direction. In Network 5 (Supplementary Table S6), *CPT1A, NDRG1*, and *TAT* genes are underexpressed in ACFR (*MYC* decreases expression of these genes), and *RHOB* is overexpressed in ACFR (*MYC* increases expression of this gene). In ACFL, the *MYC* gene is expressed at the same level compared with ACFR, but its expression is not changed in ACFL compared with NML (*MYC* is overexpressed in NML compared with NMR 2.12-fold; $P = 1.4 \times 10^{-6}$).

We did not confirm overexpression in ACF of several genes that had been reported in the literature: those encoding carcinoembryonic antigen (31), E-cadherin (31), protein kinase C βII (58), inducible nitric oxide synthase (59), c-fos (60, 61), and K-ras (61, 62). Some of the genes of which the expression was found altered in ACF, such as glutathione *S*-transferase π (62) and sodium transporter SLC5A8 (28), were not present on our arrays. The earlier described overexpression of P-cadherin (*CDH3*; ref. 30) and gastric *MUC5B* genes (3, 63, 64) was confirmed by our data, although only for ACF from the descending colon if $P < 0.001$ is used. Most of the published data were obtained by analyzing expression on the protein level, mainly by immunohistochemical methods. It is known that RNA and protein levels only moderately correlate (65, 66), and discordances in RNA and protein levels may affect conclusions about differences in gene expression. In addition, it is possible that differences in expression of some genes do not reach statistical significance due to the presence of NM in

our ACF biopsies, which may artificially decrease the magnitude of differences or increase the variance in apparent RNA levels in ACF and NM. For example, the *NOS2A* gene seems to be overexpressed in ACF (ratios: ACF/NM 1.6, ACFR/NMR 1.76, ACFL/NML 1.46) as may be expected from published data (59), but the differences are only significant at $P = 0.0062$ for the ratio ACF/NM.

It is known that nonsteroidal anti-inflammatory drugs, and particularly celecoxib, decrease the frequency of ACF in humans as well as in rodent models of carcinogen-induced colon cancer (67-70). We compared the changes in gene expression found in normal descending colonic mucosa of patients treated with celecoxib for 1 year (71) to the differences seen in gene expression between ACF and NM from the descending colon. It is provocative that for 21 of the 22 genes that are differentially expressed in ACFL versus NML and for which expression was found to be affected by celecoxib treatment, the celecoxib-driven changes in expression are in the opposite direction: genes with increased expression in ACF compared with NM show decreased expression in celecoxib-treated NM compared with untreated (data available on request).

References

1. Bird RP. Observation and quantification of aberrant crypts in the murine colon treated with a colon carcinogen: preliminary findings. Cancer Lett 1987;37:147–51.
2. Takahashi M, Wakabayashi K. Gene mutations and altered gene expression in azoxymethane-induced colon carcinogenesis in rodents. Cancer Sci 2004; 95:475–80.
3. Bara J, Forgue-Lafitte ME, Maurin N, Flejou JF, Zimber A. Abnormal expression of gastric mucin in human and rat aberrant crypt foci during colon carcinogenesis. Tumour Biol 2003;24:109–15.
4. Canzian F, Ushijima T, Serikawa T, Wakabayashi K, Sugimura T, Nagao M. Instability of microsatellites in rat colon tumors induced by heterocyclic amines. Cancer Res 1994;54:6315–7.
5. Bird RP. Role of aberrant crypt foci in understanding the pathogenesis of colon cancer. Cancer Lett 1995;93:55–71.
6. Bird RP, Good CK. The significance of aberrant crypt foci in understanding the pathogenesis of colon cancer. Toxicol Lett 2000;112–113:395–402.
7. Li H, Schut HA, Conran P, et al. Prevention by aspirin and its combination with α-difluoromethylornithine of azoxymethane-induced tumors, aberrant crypt foci and prostaglandin E2 levels in rat colon. Carcinogenesis 1999;20: 425–30.
8. Pereira MA, Barnes LH, Steele VE, Kelloff GV, Lubet RA. Piroxicam-induced regression of azoxymethane-induced aberrant crypt foci and prevention of colon cancer in rats. Carcinogenesis 1996;17:373–6.
9. Rao CV, Hirose Y, Indranie C, Reddy BS. Modulation of experimental colon tumorigenesis by types and amounts of dietary fatty acids. Cancer Res 2001; 61:1927–33.
10. Shpitz B, Klein E, Buklan G, et al. Suppressive effect of aspirin on aberrant crypt foci in patients with colorectal cancer. Gut 2003;52:1598–601.
11. Wargovich MJ, Jimenez A, McKee K, et al. Efficacy of potential chemopreventive agents on rat colon aberrant crypt formation and progression. Carcinogenesis 2000;21:1149–55.
12. Otori K, Sugiyama K, Hasebe T, Fukushima S, Esumi H. Emergence of adenomatous aberrant crypt foci (ACF) from hyperplastic ACF with concomitant increase in cell proliferation. Cancer Res 1995;55:4743–6.
13. Di Gregorio C, Losi L, Fante R, et al. Histology of aberrant crypt foci in the human colon. Histopathology 1997;30:328–34.
14. Fenoglio-Preiser CM, Noffsinger A. Aberrant crypt foci: a review. Toxicol Pathol 1999;27:632–42.
15. Roncucci L, Pedroni M, Fante R, Di Gregorio C, Ponz de Leon M. Cell kinetic evaluation of human colonic aberrant crypts. (Colorectal Cancer Study Group of the University of Modena and the Health Care District 16, Modena, Italy). Cancer Res 1993;53:3726–9.
16. Shpitz B, Bomstein Y, Mekori Y, et al. Proliferating cell nuclear antigen as a marker of cell kinetics in aberrant crypt foci, hyperplastic polyps, adenomas, and adenocarcinomas of the human colon. Am J Surg 1997;174:425–30.
17. Takayama T, Katsuki S, Takahashi Y, et al. Aberrant crypt foci of the colon as precursors of adenoma and cancer. N Engl J Med 1998;339:1277–84.
18. Siu IM, Robinson DR, Schwartz S, et al. The identification of monoclonality in human aberrant crypt foci. Cancer Res 1999;59:63–6.
19. Sakurazawa N, Tanaka N, Onda M, Esumi H. Instability of X chromosome methylation in aberrant crypt foci of the human colon. Cancer Res 2000;60: 3165–9.
20. Roncucci L, Pedroni M, Vaccina F, Benatti P, Marzona L, De Pol A. Aberrant crypt foci in colorectal carcinogenesis. Cell and crypt dynamics. Cell Prolif 2000;33:1–18.
21. Renehan AG, O'Dwyer ST, Haboubi NJ, Potten CS. Early cellular events in colorectal carcinogenesis. Colorectal Dis 2002;4:76–89.
22. Mori H, Yamada Y, Kuno T, Hirose Y. Aberrant crypt foci and β-catenin accumulated crypts; significance and roles for colorectal carcinogenesis. Mutat Res 2004;566:191–208.
23. Smith AJ, Stern HS, Penner M, et al. Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. Cancer Res 1994;54: 5527–30.
24. Otori K, Konishi M, Sugiyama K, et al. Infrequent somatic mutation of the adenomatous polyposis coli gene in aberrant crypt foci of human colon tissue. Cancer 1998;83:896–900.
25. Beach R, Chan AO, Wu TT, et al. BRAF mutations in aberrant crypt foci and hyperplastic polyposis. Am J Pathol 2005;166:1069–75.
26. Luo L, Li B, Pretlow TP. DNA alterations in human aberrant crypt foci and colon cancers by random primed polymerase chain reaction. Cancer Res 2003;63:6166–9.
27. Chan AO, Broaddus RR, Houlihan PS, Issa JP, Hamilton SR, Rashid A. CpG island methylation in aberrant crypt foci of the colorectum. Am J Pathol 2002;160:1823–30.
28. Li H, Myeroff L, Smiraglia D, et al. SLC5A8, a sodium transporter, is a tumor suppressor gene silenced by methylation in human colon aberrant crypt foci and cancers. Proc Natl Acad Sci U S A 2003;100:8412–7.
29. Pretlow TP, Roukhadze EV, O'Riordan MA, Chan JC, Amini SB, Stellato TA. Carcinoembryonic antigen in human colonic aberrant crypt foci. Gastroenterology 1994;107:1719–25.
30. Hardy RG, Tselepis C, Hoyland J, et al. Aberrant P-cadherin expression is an early event in hyperplastic and dysplastic transformation in the colon. Gut 2002;50:513–9.
31. Wargovich MJ, Chang P, Velasco M, Sinicrope F, Eisenbrodt E, Sellin J. Expression of cellular adhesion proteins and abnormal glycoproteins in human aberrant crypt foci. Appl Immunohistochem Mol Morphol 2004;12: 350–5.
32. Pretlow TP, O'Riordan MA, Pretlow TG, Stellato TA. Aberrant crypts in human colonic mucosa: putative preneoplastic lesions. J Cell Biochem Suppl 1992;16G:55–62.
33. Shpitz B, Bomstein Y, Mekori Y, et al. Aberrant crypt foci in human colons: distribution and histomorphologic characteristics. Hum Pathol 1998;29: 469–75.
34. Nascimbeni R, Villanacci V, Mariani PP, et al. Aberrant crypt foci in the human colon: frequency and histologic patterns in patients with colorectal cancer or diverticular disease. Am J Surg Pathol 1999;23:1256–63.
35. Adler DG, Gostout CJ, Sorbi D, Burgart LJ, Wang L, Harmsen WS. Endoscopic identification and quantification of aberrant crypt foci in the human colon. Gastrointest Endosc 2002;56:657–62.
36. Hurlstone DP, Karajeh M, Sanders DS, Drew SK, Cross SS. Rectal aberrant crypt foci identified using high-magnification-chromoscopic colonoscopy: biomarkers for flat and depressed neoplasia. Am J Gastroenterol 2005;100: 1283–9.
37. Rudolph RE, Dominitz JA, Lampe JW, et al. Risk factors for colorectal cancer in relation to number and size of aberrant crypt foci in humans. Cancer Epidemiol Biomarkers Prev 2005;14:605–8.
38. Roncucci L, Modica S, Pedroni M, et al. Aberrant crypt foci in patients with colorectal cancer. Br J Cancer 1998;77:2343–8.
39. Yamashita N, Minamoto T, Ochiai A, Onda M, Esumi H. Frequent and characteristic K-ras activation and absence of p53 protein accumulation in aberrant crypt foci of the colon. Gastroenterology 1995;108:434–40.
40. Moxon D, Raza M, Kenney R, et al. Relationship of aging and tobacco use with the development of aberrant crypt foci in a predominantly African-American population. Clin Gastroenterol Hepatol 2005;3:271–8.
41. Roncucci L, Stamp D, Medline A, Cullen JB, Bruce WR. Identification and quantification of aberrant crypt foci and microadenomas in the human colon. Hum Pathol 1991;22:287–94.
42. Bouzourene H, Chaubert P, Seelentag W, Bosman FT, Saraga E. Aberrant crypt foci in patients with neoplastic and nonneoplastic colonic disease. Hum Pathol 1999;30:66–71.
43. Niitsu Y, Takayama T, Miyanishi K, et al. Chemoprevention of colorectal cancer. Cancer Chemother Pharmacol 2004;54:S40–S43.
44. Takayama T, Miyanishi K, Hayashi T, et al. Aberrant crypt foci: detection, gene abnormalities, and clinical usefulness. Clin Gastroenterol Hepatol 2005; 3:S42–5.
45. 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.
46. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A 1990;87:1663–7.
47. Simon R, Lam A. Brb-Arraytools user guide, version 3.2. Biometric Research Branch, National Cancer Institute. Available from: http://linus.nci.nih.gov/brb. 2004.
48. Korn EL, McShane LM, Troendle JF, Rosenwald A, Simon R. Identifying pre-post chemotherapy differences in gene expression in breast tumours: a statistical method appropriate for this aim. Br J Cancer 2002;86:1093–6.
49. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 2003; 19:368–75.
50. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. J Comput Biol 2002;9:505–11.
51. Muller PY, Janovjak H, Miserez AR, Dobbie Z. Processing of gene expression data generated by quantitative real-time RT-PCR. Biotechniques 2002;32: 1372–4, 1376, 1378–9.

52. Caderni G, Femia AP, Giannini A, et al. Identification of mucin-depleted foci in the unsectioned colon of azoxymethane-treated rats: correlation with carcinogenesis. Cancer Res 2003;63:2388–92.

53. Paulsen JE, Steffensen IL, Loberg EM, Husoy T, Namork E, Alexander J. Qualitative and quantitative relationship between dysplastic aberrant crypt foci and tumorigenesis in the Min/+ mouse colon. Cancer Res 2001; 61:5010–5.

54. Pretlow TP, Bird RP. Correspondence Re: Y. Yamada Et Al., Frequent β-catenin gene mutations and accumulations of the protein in the putative preneoplastic lesions lacking macroscopic aberrant crypt foci appearance, in rat colon carcinogenesis. Cancer Res 2000;60:3323–7; and Sequential analysis of morphological and biological properties of β-catenin-accumulated crypts, provable premalignant lesions independent of aberrant crypt foci in rat colon carcinogenesis. Cancer Res, 61:1874–1878, 2001. Cancer Res 2001;61:7699–701.

55. Yamada Y, Mori H. Pre-cancerous lesions for colorectal cancers in rodents: a new concept. Carcinogenesis 2003;24:1015–9.

56. Yamada Y, Yoshimi N, Hirose Y, et al. Frequent β-catenin gene mutations and accumulations of the protein in the putative preneoplastic lesions lacking macroscopic aberrant crypt foci appearance, in rat colon carcinogenesis. Cancer Res 2000;60:3323–7.

57. Nambiar PR, Nakanishi M, Gupta R, et al. Genetic signatures of high- and low-risk aberrant crypt foci in a mouse model of sporadic colon cancer. Cancer Res 2004;64:6394–401.

58. Gokmen-Polar Y, Murray NR, Velasco MA, Gatalica Z, Fields AP. Elevated protein kinase C βII is an early promotive event in colon carcinogenesis. Cancer Res 2001;61:1375–81.

59. Xu MH, Deng CS, Zhu YQ, Lin J. Role of inducible nitric oxide synthase expression in aberrant crypt foci-adenoma-carcinoma sequence. World J Gastroenterol 2003;9:1246–50.

60. Stopera SA, Davie JR, Bird RP. Colonic aberrant crypt foci are associated with increased expression of c-fos: the possible role of modified c-fos expression in preneoplastic lesions in colon cancer. Carcinogenesis 1992;13:573–8.

61. Shpitz B, Bomstein Y, Shalev M, et al. Oncoprotein coexpression in human aberrant crypt foci and minute polypoid lesions of the large bowel. Anticancer Res 1999;19:3361–6.

62. Miyanishi K, Takayama T, Ohi M, et al. Glutathione *S*-transferase-π overexpression is closely associated with K-ras mutation during human colon carcinogenesis. Gastroenterology 2001;121:865–74.

63. Perrais M, Pigny P, Buisine M-P, Porchet N, Aubert J-P, Van Seuningen-Lempire I. Aberrant expression of human mucin gene Muc5b in gastric carcinoma and cancer cells. identification and regulation of a distal promoter. J Biol Chem 2001;276:15386–96.

64. Pinto-de-Sousa JJ, Reis CCA, David LL, Pimenta AA, Cardoso-de-Oliveira MM. Muc5b expression in gastric carcinoma: relationship with clinico-pathological parameters and with expression of mucins Muc1, Muc2, Muc5ac and Muc6. Virchows Arch 2004;444:224–30.

65. Merrick BA, Madenspacher JH. Complementary gene and protein expression studies and integrative approaches in toxicogenomics. Toxicol Appl Pharmacol 2005;207:189–94.

66. Tian Q, Stepaniants SB, Mao M, et al. Integrated genomic and proteomic analyses of gene expression in mammalian cells. Mol Cell Proteomics 2004;3: 960–9.

67. Buecher B, Thouminot C, Menanteau J, et al. Fructooligosaccharide Associated with celecoxib reduces the number of aberrant crypt foci in the colon of rats. Reprod Nutr Dev 2003;43:347–56.

68. Brown WA, Skinner SA, Malcontenti-Wilson C, et al. Non-steroidal anti-inflammatory drugs with different cyclooxygenase inhibitory profiles that prevent aberrant crypt foci formation but vary in acute gastrotoxicity in a rat model. J Gastroenterol Hepatol 2000;15:1386–92.

69. Yamada Y, Yoshimi N, Hirose Y, et al. Suppression of Occurrence and Advancement of β-Catenin-Accumulated Crypts, Possible Premalignant Lesions of Colon Cancer, by Selective Cyclooxygenase-2 Inhibitor, Celecoxib. Jpn J Cancer Res 2001;92:617–23.

70. Rao CV, Indranie C, Simi B, Manning PT, Connor JR, Reddy BS. Chemopreventive properties of a selective inducible nitric oxide synthase inhibitor in colon carcinogenesis, administered alone or in combination with celecoxib, a selective cyclooxygenase-2 inhibitor. Cancer Res 2002;62: 165–70.

71. Glebov OK, Rodriguez LM, Lynch P, et al. Celecoxib treatment alters the gene expression profile of normal colonic mucosa. Cancer Epidemiol Biomarkers Prev 2006;15:1381–91.