

Distinguishing Right from Left Colon by the Pattern of Gene Expression¹

Oleg K. Glebov, Luz M. Rodriguez, Kenneth Nakahara, Jean Jenkins, Janet Cliatt, Casey-Jo Humbyrd, John DeNobile, Peter Soballe, Richard Simon, George Wright, Patrick Lynch, Sherri Patterson, Henry Lynch, Steven Gallinger, Aby Buchbinder, Gary Gordon, Ernest Hawk, and Ilan R. Kirsch²

Genetics Branch [O. K. G., L. M. R., K. N., J. J., J. C., C.-J. H., I. R. K.] and Biometric Research Branch [R. S., G. W.], Center for Cancer Research, National Cancer Institute (NCI), and Division of Cancer Prevention [E. H.], NCI, Bethesda, Maryland 20892; Surgery Department, National Naval Medical Center, Bethesda, Maryland 20889 [J. D., P. S.]; Department of Gastrointestinal Oncology, University of Texas-M. D. Anderson Cancer Center, Houston, Texas 77030 [P. L., S. P.]; Department of Preventive Medicine, Creighton University, Omaha, Nebraska [H. L.]; University of Toronto, Toronto, Ontario, M5G 1X5 Canada [S. G.]; Clinical Oncology Research and Development, Pharmacia, Peapack, New Jersey 07977 [A. B.]; and Ovation Pharma, Lincolnshire, Illinois 60015 [G. G.]

Abstract

Distinct epidemiological and clinicopathological characteristics of colorectal carcinomas (CRCs) based on their anatomical location suggest different risk factors and pathways of transformation associated with proximal and distal colon carcinogenesis. These differences may reflect distinct biological characteristics of proximal and distal colonic mucosa, acquired in embryonic or postnatal development, that determine a differential response to uniformly distributed environmental factors. Alternatively, the differences in the epidemiology of proximal and distal CRCs could result from the presence of different procarcinogenic factors in the ascending versus descending colon, acting on cells with either similar or distinct biological characteristics. We applied cDNA microarray technology to explore the possibility that mucosal epithelium from adult proximal and distal colon can be distinguished by their pattern of gene expression. In addition, gene expression was studied in fetal (17–24 weeks gestation) proximal and distal colon. More than 1000 genes were expressed differentially in adult ascending versus descending colon, with 165 genes showing >2-fold and 49 genes showing >3-fold differences in expression. With almost complete concordance, biopsies of adult colonic epithelium can be correctly classified as proximal or distal by gene expression profile. Only 87 genes were expressed

differently in ascending and descending fetal colon, indicating that, although anatomically relevant differences are already established in embryonic colon, additional changes in gene expression occur in postnatal development.

Introduction

The incidence of proximal (right) and distal (left) CRCs³ varies depending on the total incidence of CRCs and the geographic location. This indicates that there are different risk factors associated with proximal and distal colon carcinogenesis (1, 2). These differences are likely to be related to the different ratios of mismatch repair-defective versus mismatch repair-competent CRCs in the right versus left colon. Whereas mismatch repair-competent CRCs predominate throughout the colon, mismatch repair-defective malignancies show a notable predisposition to occur in the right colon. Approximately 30% of all right-sided CRCs are mismatch repair defective as recognized by the presence of a MSI-H, whereas only 2% of left-sided malignancies manifest the MSI-H phenotype (3, 4). This suggests that there may be an inherent or acquired difference between the right and the left colon that makes the right side more susceptible to the initiation or progression of cellular transformation via a pathway that commences with a cell that has lost its ability to recognize and repair nucleotide mismatches. Relative distinguishing characteristics between right- and left-sided CRCs that may reflect the increased number of right-sided MSI-H tumors include several clinicopathological and genetic features. Right-sided CRCs more frequently display undifferentiated and mucinous phenotypes, have pseudo-diploid karyotypes, and exhibit a lower frequency of *TP53* and *K-RAS* gene mutations and *c-MYC* expression. Left-sided tumors, on the other hand, are more uniformly aneuploid with significant loss of heterozygosity (including genes on the 18q chromosome) and rearranged chromosomes, and have a higher frequency of *TP53* and *K-RAS* gene mutations and *c-MYC* expression (reviewed in Refs. 1, 5, and 6).

The proximal and distal colon have a different embryological origin: cecum, ascending colon, and the proximal two thirds of the transverse colon [proximal, or right colon (RC)] derive from embryonic midgut; the distal third of the transverse colon, descending colon, sigmoid colon and rectum [distal, or left colon (LC)] derive from hindgut (7). Reflecting their origin, the proximal and distal colon also have a different vascular supply, with the right colon served by the superior mesenteric artery and the left colon by the inferior mesenteric artery. In

Received 1/27/03; revised 4/17/03; accepted 4/27/03.

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.

¹ Please see Supplementary Data for this article at <http://cebp.aacrjournals.org>.

² To whom requests for reprints should be addressed, at Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda MD 20892. Fax: (301) 496-0047; E-mail: kirsch@exchange.nih.gov.

³ The abbreviations used are: CRC, colorectal carcinoma; MSI-H, high level of microsatellite instability; NCI, National Cancer Institute; NNMC, National Naval Medical Center; HNPCC, hereditary nonpolyposis colorectal cancer; aRNA, antisense RNA; RT-PCR, reverse transcription-PCR; CCP, compound covariate predictor.

addition, right colon and left colon differ in expression of several antigens, metabolism of glucose, polyamines and butyric acid, as well as in bile acid concentrations, and composition and density of the bacterial population (1, 5, 6).

Thus, the differences in incidence of CRC and predilection to microsatellite stable (MSS) or microsatellite unstable (MSI) type of carcinogenesis may: (a) reflect different biological characteristics of proximal and distal colon cells, acquired in embryonic or postnatal development, which in turn determine different responses to common environmental factors; or (b) result from the presence of different procarcinogenic factors in proximal and distal colon (bile acids, bacteria), acting on similar mucosa; or (c) be a combination of a distinct environment and distinct target cells.

To begin our exploration of this issue, we asked whether colonic mucosal epithelium from the right side can be distinguished from colonic mucosal epithelium obtained from the left side of the colon. Our approach has been to define patterns of gene expression for each side of the colon, and compare the right and left colon to determine whether there is a significant difference in these patterns.

Materials and Methods

Patients and Biopsy Samples. Different adult populations of patients were recruited for this study on protocols approved by Institutional Review Boards of both the NCI and the NNMC. Patients with HNPCC, or Lynch syndrome, were enrolled during a multi-institutional (NNMC, NCI, M. D. Anderson Medical Center, and Creighton University) Phase I-II multiple-dose safety and efficacy study of a selective inhibitor of cyclooxygenase-2 (celecoxib) in HNPCC patients and carriers. In this article, we report the analysis of only the colon biopsies that were taken at the beginning of the study, before treatment of the patients started. Patients without known predisposition to colon cancer were enrolled in the course of a routine screening colonoscopy at the NNMC. Before enrollment, informed consent was obtained from all of the patients. All HNPCC patients were known to carry germ-line mutations in the *hMLH1* or *hMSH2* genes. All of the patients were free of malignancy at the time of enrollment. On colonoscopy, standard pinch biopsies of normal colonic mucosa were taken from the ascending (AT-) and descending (L-) colon, flash-frozen in liquid nitrogen, and stored at -80°C . Human fetal colon specimens were obtained from the Human Fetal Tissue Repository of the Albert Einstein College of Medicine as approved by its IRB under protocol no. 93-42. The tissues were obtained after therapeutic abortion from 19 fetuses of gestational age 17–24 weeks. Fetal colons were flash-frozen in liquid nitrogen immediately after resection and were kept at -80°C . The mucosal cell layer was viewed under a dissecting microscope, and pinch biopsies from ascending and descending parts of the frozen fetal colons were taken and kept on dry ice. More than 80% of the cells in standard colon pinch biopsies and in fetal colon samples were colonic crypt cells as determined by H&E staining.

RNA Extraction and Amplification. Total RNA was isolated from flash-frozen specimens, homogenized with a Disposable Generator and Micro-H Omni Homogenizer in either lysis buffer RLT (Qiagen) or in Tri-Reagent (Molecular Research Center, Inc.), and purified using Qiagen's RNeasy Mini Kit columns (Qiagen) according to the manufacturer's instructions.

mRNA was amplified according to a modified Eberwine's protocol (8). SuperScript II reverse transcriptase (Life Technologies, Inc.) was used for first-strand cDNA synthesis, starting with 5 μg of total RNA. First-strand cDNA synthesis was

performed with an oligo(dT)₂₄ primer containing 5'-T7 promoter-primer, and the second strand was synthesized with a mixture of RNase H, DNA polymerase, and DNA ligase.

cDNA was treated with a mixture of phenol-chloroform-isoamyl alcohol, ethanol-precipitated, washed with 80% ethanol, and dried in a Speed-Vac. Purified cDNA was transcribed *in vitro* with a T7 MEGAscript kit (Ambion) according to the manufacturer's instructions, and aRNA was purified using RNeasy mini spin columns (Qiagen). Concentration of the eluted aRNA was determined by spectrophotometric measurement of absorbance at 260 nm.

Using this protocol, 50–100 μg of aRNA were obtained after starting with 5 μg of total RNA, indicating an approximately 1000-fold linear amplification of mRNA.

cDNA Labeling and Microarray Hybridization. Fluorescent-labeled cDNA was synthesized by reverse transcription of colon aRNA and human testis aRNA [prepared from total testis RNA (Clontech, Inc.) as described for colon aRNA] with Superscript II reverse transcriptase (Invitrogen) and random oligonucleotide primers in the presence of Cy3-dUTP or Cy5-dUTP (Amersham Pharmacia Biotech), respectively. For each hybridization experiment, 5 μg of colon aRNA and 6 μg of testis aRNA (common reference) were used to prepare a mixture of labeled cDNAs. After alkaline lysis of aRNA, labeled cDNA was purified with AutoSeq G50 columns (Amersham) and QIAquick PCR purification columns (Qiagen). Purified labeled colon and testis cDNAs were mixed, ethanol precipitated, dried in a Speed-Vac, and dissolved in hybridization buffer containing $3.7\times$ SSPE, $2.5\times$ Denhardt's Solution (Sigma), 0.5 $\mu\text{g}/\mu\text{l}$ Poly(A)_{40–60} (Amersham Pharmacia Biotech), human C₁t1 DNA (Invitrogen), and 0.125 $\mu\text{g}/\mu\text{l}$ yeast tRNA (Sigma) in TE buffer, with or without 50% formamide. After denaturing at 100°C for 2 min., snap-cooling on ice, and warming to room temperature, 10% SDS was added to the hybridization mixture to a final concentration of 0.25%.

Microarrays containing cDNAs spotted on lysine-coated glass slides were obtained from the Advanced Technology Center (NCI). Three different versions of microarrays were used, containing 6500, 6400, and 9000 distinct cDNA clones supplied from Research Genetics, Inc. (Huntsville, AL). Detailed information and comparison of printed cDNA sets can be found on the mAdb web site.⁴ 9k-microarrays contain 9120 sequence-verified cDNAs, including 8281 unique UniGene clusters, among which 7102 represent named genes and 1179 represent EST clusters.

Most of the experiments were performed by incubating slides at 65°C for 16–18 h in sealed hybridization chambers with 20 μl (for 6.5k- and 6.4k-microarrays) or 40 μl (for 9k-microarrays) of hybridization mixture (without formamide) placed under the coverslips. A small number of 6.4k-microarrays were prehybridized in $5\times$ SSPE- $5\times$ Denhardt's Solution-0.1% SDS at 42°C for 1 h, washed in distilled H₂O for 4 min at room temperature, dehydrated in isopropanol for 1 min., and dried by centrifugation at $150\times g$ for 5 min at room temperature. Twenty μl of a hybridization mixture containing 50% formamide were placed under the coverslip, and slides were incubated in sealed hybridization chambers at 42°C for 16–18 h.

After hybridization, slides were washed at room temperature in $2\times$ SSC-0.1% SDS (5 min), $2\times$ SSC (2 min), $1\times$ SSC (1 min), and $0.2\times$ SSC (30 s), and were dipped in 0.05% SSC.

⁴ Internet address: <http://nciarray.nci.nih.gov/>.

Slides were dried by centrifugation at $150 \times g$ for 5 min. at room temperature.

Data Acquisition and Analysis. Microarrays were scanned with an Axon 4000 laser scanner, and primary image analysis was performed with GenePix Pro 3.0 Software (Axon Instruments, Inc.). Images were also visually inspected, and questionable spots were flagged and excluded from the analysis. Multidimensional scaling, clustering of genes and arrays, and class prediction of arrays were performed with the BRB-ArrayTools (version 2.1) software package developed by the Biometric Research Branch of the Division of Cancer Treatment and Diagnosis of the NCI and The EMMES Corporation (9).⁵

Briefly, background intensities were subtracted, and data were filtered for minimal spot intensity (500 units) in each of the two channels and for missing values (not more than in 20% of arrays). Fluorescence intensity-ratio data were log-transformed and normalized by subtracting the median log-ratio from all of the log-ratios on the array. Average linkage hierarchical clustering of genes and samples and multidimensional scaling of samples were performed by using one minus the Pearson correlation coefficient as a distance metric. Test of statistical significance of sample clustering was accessed when possible by using Euclidian distance as the metric, and the first three principal components as the axes, with the null hypothesis that data samples came from the same multivariate Gaussian distribution (10).

The Class Prediction module of the BRB-ArrayTools and CCP (11–13) were used to determine whether the pattern of gene expression allowed classification of samples from proximal and distal colon. First, the *t* test was used to select genes that showed univariately statistically significant ($P < 0.001$) differences in expression between the two classes of samples. In batch analysis, each sample was considered as independent, and a *t* test was performed on average log-ratios of gene expression for two classes, whereas in paired analysis, average differences in gene expression log-ratios of pairs of samples (ascending colon and descending colon biopsies) from the same patient were analyzed by *t* test. Then a CCP was calculated as a linear combination of log-ratios (or log-ratio differences) weighted by univariate *t* values (13). The CCP value was calculated for each sample (or pair of samples) and, as a classification threshold, the mean of the average CCP for two classes was used for batch analysis, and the sign (positive or negative) of the CCP was used for paired analysis. The misclassification rate was estimated by leave-one-out cross-validation. Specifically, a sample (or a pair of samples in case of paired analysis) was omitted and a CCP was developed from scratch using the included samples. This step included performing the *t* tests, recalculating the CCP and classification threshold for the remaining samples, and applying the new CCP values to classify the omitted sample (or pair of samples). This was done independently for each omitted sample (or pair of samples). The ratio of “samples incorrectly classified in cross-validation”: “the total number of samples” yields the misclassification rate. Permutation *P* for the misclassification rate for the CCP was calculated by 2000 random permutations of class labels and a repeated cross-validation procedure for each permutation. The proportion of the random permutations that gave the same or smaller misclassification rate as was obtained with true class labels is presented as a (permutation) *P* for CCP, and

a value of $P < 0.0005$ was reported when no random permutation of class label was found among 2000 with the same or smaller misclassification rate as for true class labeling.

For quality assurance purposes, real-time RT-PCR was used to study expression of several cDNA sequences present on our arrays. RT-PCR was performed in LightCycler with LightCycler-RNA Amplification Kit SYBR-Green I using gene-specific primers according to the manufacturer’s instructions (Roche). RT-PCR data on the relative expression of these cDNAs were in general agreement with microarray hybridization results.

Results

Paired analysis was performed by applying the *t* test to the differences in gene expression log-ratios between ascending colon and descending colon biopsies from each individual patient averaged across the patient set. This seemed to be more sensitive than batch analysis and resulted in a larger number of genes that showed statistically significant differences in expression and a smaller number of misclassified samples in leave-one-out cross-validation. Because, except for the differences noted below, the results of batch analysis were concordant with the results of paired analysis, we will focus on the results of paired analysis.

Biopsies from 50 HNPCC patients were analyzed on a total of 118 microarrays (with sets of ascending-colon and descending-colon biopsies from two patients analyzed twice on 6.5/6.4k-arrays and from six patients analyzed on 6.5/6.4k- and 9k-arrays; on analysis, data for the same biopsies were averaged and considered as one sample). As shown, on analysis of 42 9k-microarray data files generated with biopsies from 21 HNPCC patients, there are reproducible and significant differences in the gene expression profile in ascending-colon and descending-colon biopsies. After filtering, 744 of 7810 genes showed statistically significant ($P < 0.001$) differences in expression (Fig. 1; Supplementary Table 1).¹ All of the biopsies were correctly classified in leave-one-out cross-validation.

At the outset of this study we used 6.5k- and 6.4k-arrays that were printed with sets of cDNAs that only partially overlapped with each other and with the cDNA set printed on 9k-arrays. The differences in composition of cDNAs printed on 6.5k- and 6.4k-arrays, as well as hybridization in different conditions, introduced additional variation in the data: 6.5k-, 6.4k-, and 9k-array data files and samples hybridized in the presence or absence of formamide are easily differentiated by class prediction analysis. Only 2620 genes met filtering criteria when all three array sets, 6.5k-, 6.4k-, and 9k-arrays, were analyzed as one set (116 data files for 50 patients). Despite these confounding variables, the *t* test revealed 468 genes the expression of which differs between ascending-colon and descending-colon biopsies; the corresponding CCP correctly classifies 98 of 100 (independent) samples (with 1 pair misclassified), whether they were hybridized on 6.5k-, 6.4k-, or 9k-arrays, or in the presence or absence of formamide ($P < 0.0005$). Thus, whether relying on only the newest and most validated version of the cDNA array (9k) or on all three (6.5k, 6.4k, and 9k) combined together, there are multiple and reproducible differences in gene expression that allow correct identification of biopsies as from either the ascending or the descending colon of HNPCC patients.

Analysis on 9k-arrays of 26 colon biopsies from 13 patients without predisposition to colon cancer (non-HNPCC patients) also showed differences in gene expression between the ascending and descending colon. Of 8818 genes that met the

⁵ Internet address: <http://linus.nci.nih.gov/~brb>.

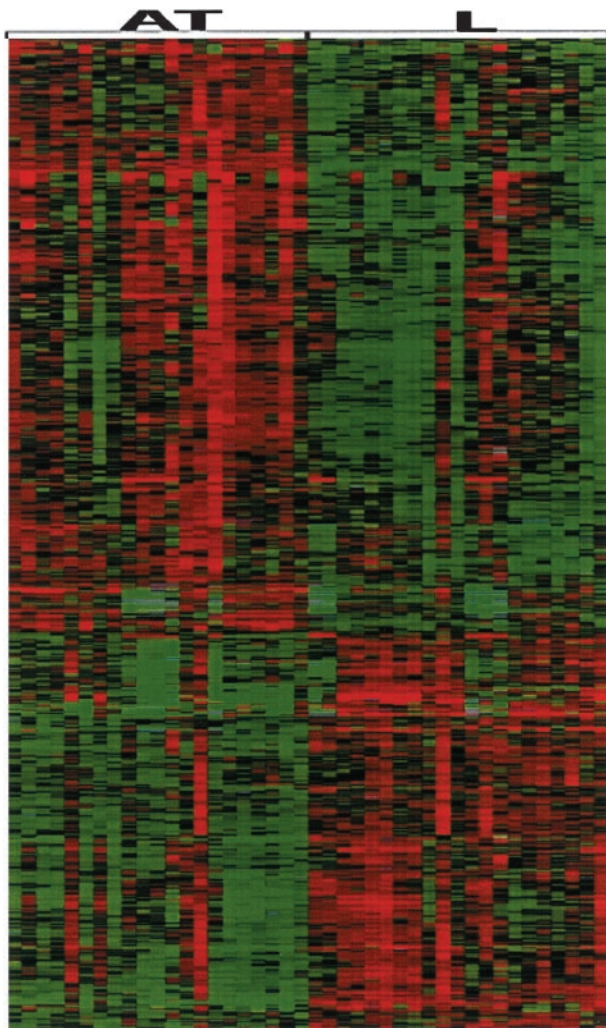


Fig. 1. Hierarchical clustering of genes that are expressed differentially in ascending and descending colon of HNPCC patients. CCP genes (744) were clustered by using average linkage and Pearson correlation coefficient as a distance metric. AT, ascending-colon biopsies; L, descending-colon biopsies. Relative expression: green, overexpression; red, underexpression; black, equal expression; gray, missing value.

analysis criteria, 658 genes showed statistically significant ($P < 0.001$) differences in expression between ascending-colon and descending-colon biopsies (Supplementary Fig. S1; Supplementary Table 2).¹ In cross-validation, the corresponding CCPs allowed correct classification of all of the samples ($P < 0.0005$).

A CCP that was built on 9k-array data for ascending-colon and descending-colon biopsies from HNPCC patients allowed correct classification of all ascending-colon and descending-colon biopsies from non-HNPCC patients ($P < 0.0005$), and a classifier built on samples from non-HNPCC patients predicted correctly 20 of 21 pairs of biopsies from HNPCC patients ($P < 0.0005$; data not shown). Thus, there are common differences in gene expression between the ascending and descending colon of HNPCC and non-HNPCC patients.

Samples from HNPCC and non-HNPCC patients, all hybridized on the same 9k-array platform, were then analyzed together (Fig. 2). In the combined set of all colon biopsies from

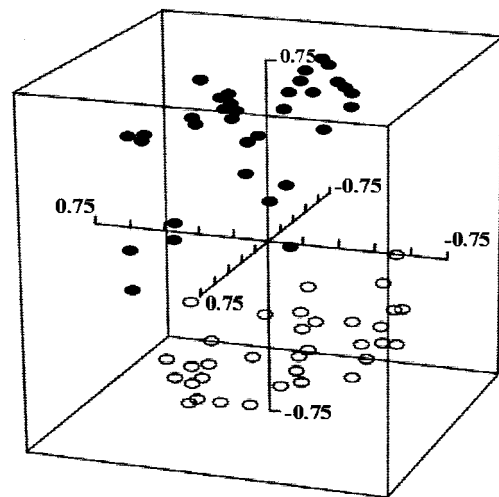


Fig. 2. Multidimensional scaling of ascending- and descending-colon biopsies from HNPCC and non-HNPCC patients. Axes, relative distance (the magnitude of the 1-Pearson correlation coefficient distance in each of the first three principal components). Sixty-eight ascending (filled oval) and descending (open oval) colon biopsies were clustered according to expression of 1336 genes.

HNPCC patients (42 ascending-colon and descending-colon biopsies) and non-HNPCC patients (26 biopsies), statistically significant differences between ascending-colon and descending-colon samples were found in the expression of 1336 genes (of 7994 genes that passed filtering criteria; Supplementary Fig. S2; Supplementary Table 3).¹ Most of the genes that were found to be expressed differentially in ascending-colon and descending-colon biopsies in separate analyses of samples from HNPCC patients (93.7%) and from non-HNPCC patients (78.4%) were present in the list of 1336 genes obtained on analysis of the combined samples. All of the samples were classified correctly on cross-validation, with a pair of biopsies from a patient with a right hemicolectomy having the only borderline CCP value.

To exclude the possible impact of surgery, we analyzed a combined set of 50 ascending-colon and descending-colon biopsies from 14 HNPCC and 11 non-HNPCC patients without surgery. Differences in expression between ascending-colon and descending-colon biopsies were found (at $P = 0.001$ level) for 1349 genes (Supplementary Fig. S3; Supplementary Table 4).¹ All of the samples were classified correctly in cross-validation ($P < 0.0005$). One hundred sixty-five genes showed more than 2-fold and 49 showed more than 3-fold differences in expression between the ascending and descending colon (Table 1). Most of the genes (85%) that were found on the analysis of samples from patients without surgery were also identified as expressed differentially in the analysis of the combined group of patients, and *vice versa*.

There were 9 patients with colon surgery in the combined set of HNPCC (7 of 21) and non-HNPCC (2 of 13) patients: 5 with right hemicolectomy, 2 with left hemicolectomy, 1 with transverse colon removal, and 1 with sigmoid colon removal. Although some of the biopsies from patients with right hemicolectomy were misclassified (in batch analysis; data not shown) or had a borderline CCP value (in paired analysis), most of the biopsies from patients with right or left hemicolectomies were classified according to their postoperative position, *i.e.*, proximal or distal, in the colon. This suggests that the colonic mucosa is capable of some adaptation to its postoperative

Table 1 Genes that show more than three-fold differences in expression in ascending versus descending adult colon

Gene	Clone	UniGene cluster	Fold-difference L:AT ^a	t value
<i>homeo box D13</i>	IncyePD:3462557	Hs.158309	8.15	5.932
<i>solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 2</i>	IncyePD:1634540	Hs.102307	5.71	6.656
<i>serine protease inhibitor, Kazal type, 5</i>	IncyePD:776793	Hs.331555	5.39	6.606
<i>ESTs</i>	IncyePD:1671596	Hs.109438	5.23	17.146
<i>secretory leukocyte protease inhibitor (antileukoproteinas)</i>	IncyePD:2510171	Hs.251754	4.52	15.136
<i>homeo box B13</i>	IncyePD:1861743	Hs.66731	4.31	7.999
<i>BRCA1 associated protein</i>	IncyePD:1623214	Hs.122764	4.06	9.395
<i>connective tissue growth factor</i>	IncyePD:1674454	Hs.75511	4.03	8.561
<i>EGF-containing fibulin-like extracellular matrix protein 1</i>	IncyePD:1798209	Hs.76224	4.03	11.436
<i>protease inhibitor 3, skin-derived (SKALP)</i>	IncyePD:2503913	Hs.112341	3.93	7.209
<i>NADPH oxidase 1</i>	IncyePD:2793437	Hs.132370	3.88	6.904
<i>solute carrier family 28 (sodium-coupled nucleoside transporter), member 2</i>	IncyePD:1632811	Hs.193665	3.73	4.161
<i>retinoic acid receptor responder (tazarotene induced) 2</i>	IncyePD:1988108	Hs.37682	3.72	11.839
<i>sialyltransferase 1 (β-galactoside α-2,6-sialyltransferase)</i>	IncyePD:3211926	Hs.374528	3.66	6.651
<i>osteoblast specific factor 2 (fasciclin I-like)</i>	IncyePD:1994715	Hs.136348	3.51	4.614
<i>cysteine-rich protein 1 (intestinal)</i>	IncyePD:2121863	Hs.3192	3.4	11.104
<i>S100 calcium binding protein P</i>	IncyePD:2060823	Hs.2962	3.39	5.728
<i>CMP-NeuAC:(β)-N-acetylgalactosaminide (alpha)2,6-sialyltransferase member VI</i>	IncyePD:2504517	Hs.109672	3.32	10.432
<i>family with sequence similarity 3, member C</i>	IncyePD:1926135	Hs.29882	3.32	12.324
<i>cysteine-rich, angiogenic inducer, 61</i>	IncyePD:1514989	Hs.8867	3.28	5.845
<i>phospholipase A2, group IIA (platelets, synovial fluid)</i>	IncyePD:699410	Hs.76422	3.27	6.014
<i>protein phosphatase 1, regulatory (inhibitor) subunit 12B</i>	IncyePD:1998571	Hs.130760	3.18	18.759
<i>somatostatin</i>	IncyePD:2494617	Hs.12409	3.11	4.037
<i>chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)</i>	IncyePD:1669647	Hs.237356	3.08	11.331
<i>alpha-actinin-2-associated LIM protein</i>	IncyePD:1924344	Hs.135281	3.05	14.464
<i>ribonuclease, RNase A family, 1 (pancreatic)</i>	IncyePD:3086929	Hs.78224	3.01	9.321
<i>3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)</i>	IncyePD:1807407	Hs.59889	0.33	-8.154
<i>acyl-Coenzyme A oxidase 1, palmitoyl</i>	IncyePD:1798585	Hs.379991	0.33	-13.088
<i>carnitine deficiency-associated gene expressed in ventricle 1</i>	IncyePD:1494531	Hs.333120	0.33	-3.798
<i>glucosaminyl (N-acetyl) transferase 2, I-branching enzyme</i>	IncyePD:1966455	Hs.934	0.32	-7.262
<i>MAWD binding protein</i>	IncyePD:1812030	Hs.16341	0.32	-4.875
<i>glypican 3</i>	IncyePD:645584	Hs.119651	0.32	-6.347
<i>ATP-binding cassette, sub-family G (WHITE), member 2</i>	IncyePD:1501080	Hs.194720	0.32	-8.077
<i>solute carrier family 38, member 4</i>	IncyePD:778212	Hs.165655	0.3	-6.679
<i>homeo box B6</i>	IncyePD:606814	Hs.98428	0.3	-6.918
<i>solute carrier family 16 (monocarboxylic acid transporters), member 1</i>	IncyePD:1981569	Hs.75231	0.29	-13.71
<i>differentially expressed in hematopoietic lineages</i>	IncyePD:1903267	Hs.273321	0.28	-5.393
<i>solute carrier family 16 (monocarboxylic acid transporters), member 1</i>	IncyePD:733668	Hs.75231	0.26	-12.558
<i>hydroxy-δ-5-steroid dehydrogenase, 3 β- and steroid δ- isomerase 1</i>	IncyePD:182802	Hs.38586	0.24	-6.071
<i>apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1</i>	IncyePD:1634063	Hs.560	0.24	-5.824
<i>cytochrome P450, subfamily IIC (mephenytoin 4-hydroxylase), polypeptide 18</i>	IncyePD:2595728		0.22	-7.491
<i>solute carrier family 20 (phosphate transporter), member 1</i>	IncyePD:1846463	Hs.78452	0.2	-16.061
<i>hydroxy-δ-5-steroid dehydrogenase, 3 β- and steroid δ- isomerase 2</i>	IncyePD:942100	Hs.825	0.2	-5.632
<i>nuclear receptor subfamily 1, group H, member 4</i>	IncyePD:214180	Hs.171683	0.16	-9.444
<i>alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)</i>	IncyePD:2771046	Hs.1239	0.15	-7.824
<i>paired-like homeodomain transcription factor 2</i>	IncyePD:2794019	Hs.92282	0.13	-8.926
<i>regenerating islet-derived 1 α (pancreatic stone protein, pancreatic thread protein)</i>	IncyePD:2923150	Hs.1032	0.11	-8.678
<i>meprin A, β</i>	IncyePD:2234609	Hs.194777	0.08	-15.037
<i>ethanolamine kinase</i>	IncyePD:2718565	Hs.120439	0.07	-19.742

^a L:AT, ratio of descending colon to ascending colon.

environment (in terms of gene expression), but this response may not be instantaneous, complete, or universal.

To determine whether the differences in gene expression in adult ascending and descending colon are established in embryonic or postnatal development, 26 samples of embryonic colonic mucosal cells (13 paired samples from ascending and descending colon) were analyzed. Statistically significant differences between ascending colon and descending colon sam-

ples from the fetal colon were found in the expression of 87 genes (Fig. 3; Supplementary Fig. S4; Supplementary Table 5).¹ The probability of having, by chance, at least 87 of 9013 genes pass the filtering criteria, as defined by showing differences in expression at $P = 0.001$, is 0.005 by permutation analysis. Twenty-four of 26 samples are classified correctly by using the CCP ($P < 0.0005$). Twenty-four of these 87 genes are included in the list of CCP genes that showed differences in

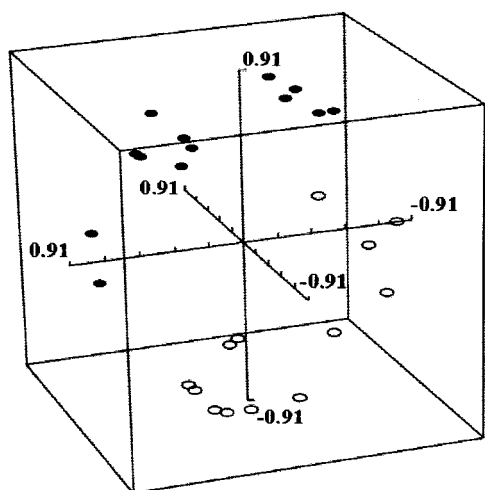


Fig. 3. Multidimensional scaling of ascending and descending fetal colon biopsies. Axes, relative distance (the magnitude of the projection of the 1-Pearson correlation coefficient distance in each of the first three principal components). Twenty-six ascending (filled oval) and descending (open oval) fetal colon biopsies were clustered according to expression of 87 CCP genes.

expression between ascending-colon and descending-colon samples from adult colon. However, only 11 of these 24 show the same directional (*i.e.*, increased expression in right *versus* left colon and *vice versa*) differences in fetal and adult mucosa.

Discussion

A total of 1349 genes show differences in expression between adult ascending (AT) and descending (L) colon biopsies of patients without colon surgery. Seventy % (947) of these genes are expressed at a higher level in descending colon and 30% (402) in ascending colon, suggesting higher transcriptional activity in this overall set of genes in the descending colon. Comparison with functional gene lists that are present in the Cancer Genome Anatomy Project (CGAP; <http://cgap.nci.nih.gov/>), in Gene Ontology Browser and BioCarta Pathways) shows that genes expressed differently in ascending and descending colon are involved in the control of many cellular functions (Table 2). These include cell cycle, proliferation, cell death, response to external stimuli, stress response, and DNA replication and damage repair. Differentially expressed genes are implicated in major signaling pathways important for colon tumorigenesis, including EGF, TGF β , Wnt, Ras, insulin, and integrin signaling. For most of the functional groups, the relative number of genes that are expressed higher in the left compared with the right colon corresponds to a 2:1 ratio that is characteristic of the total set of the CCP genes, but for some groups the biases are more extreme. For example, 90% of the genes from cell cycle and DNA metabolism groups are overexpressed in the left colon (see below), whereas the number of overexpressed genes that are involved in cell death, Wnt, and Ras signaling is approximately the same in the ascending and the descending colon.

There are only a few publications in which gene expression in the right and left colon has been compared (14–23), and the data are primarily derived from the determination of protein levels or enzymatic activities. Some of our observed differences in gene expression in ascending- *versus* descending-colon biopsies correspond to published data: phosphosulfotransferases

that are overexpressed in ascending-colon biopsies were found to be more active in the right colon compared with the left colon (14); and phospholipase A2, the message of which is overexpressed in the left colon, was found to be more abundant at the protein level in the left colon as well (15). Genes for several other proteins that have been previously described as showing differential right *versus* left colon expression include *TGF α* (16, 17), *Bak* (18), *cErbB2* (19), *EGFR* (20). These did not show variation in mRNA expression in the right and left colon segments in our experiments. *c-Ha-ras* was found to show differential right *versus* left expression in inflamed colon mucosa (21) but did not emerge in our analysis of normal mucosa. A handful of other genes [*Cdx2* (22) and genes for gastric M1 and intestinal M3 antigens (23)] were not included on our arrays.

Masys *et al.* (24) have developed an approach to the interpretation of differences in gene expression through the analysis of associations between genes in a particular set and medical subject headings (MeSH) terms in published literature (*e.g.*, presented in the Medline database), combined with an estimate of the probability of getting a number of associations by chance. We used the publicly available interface to the High-density Array Pattern Interpreter (HAPI) system⁶ to analyze genes expressed differently in the ascending and the descending colon. The analysis is somewhat hampered by the fact that only ~30% of genes in the Incyte UniGEM set of cDNAs printed on 9k-arrays have one or more linked citations (Ref. 24; this is also true for the set of genes expressed differentially in right and left colon), and by the current limitations in the number of genes that can be analyzed (250) in one set. Nevertheless, a few conceptual associations were immediately evident when the set of 165 genes that showed a ≥ 2 -fold difference in expression in ascending-colon and descending-colon biopsies was analyzed. For example, there is a significant association of these genes with inorganic and organic chemicals subject areas. Separate analysis of genes overexpressed in right and left colon indicates that inorganic chemical subject area terms (ions, anions, and electrolytes) are associated mainly with genes overexpressed in the right colon, whereas organic chemical terms (carboxylic acids and acyclic acids) are linked to both sets of genes. These associations suggest differences in the physiology of the ascending and the descending colon. The large bowel is considered an appendage of the digestive tract with a relatively simple principal function in conservation of salt and water and disposal of waste material. The proximal colon is involved in the solidification of fecal contents through the absorption of water and electrolytes. The distal colon is used for transient storage of feces and is less involved in ion, anion, and electrolyte transport.

Genes overexpressed in the right colon are also associated with terms in heterocyclic compounds, polycyclic hydrocarbon, and steroids subject areas. Among these genes, several members of the cytochrome P450 family (*CYP2C8*, *CYP2C18*, and *CYP4F12*) are expressed 1.7- to 4.5-fold higher in the right colon, and so are genes for glutathione *S*-transferases Z1 (*GST Z1*, 1.2 \times), 3- β -hydroxysteroid dehydrogenase 1 (*HSD3C4*, 1.7 \times) and hydroxysteroid (17 β) dehydrogenase (2.4 \times). This pattern of gene expression indicates that, if overexpression of genes is equally reflected in protein level and function, the right colon compared with the left colon may be better protected against procarcinogenic heterocyclic compounds in food. It is

⁶ Internet address: <http://array.ucsd.edu/>.

Table 2 Genes expressed differentially in ascending and descending colon control all main cellular functions

CGAP ^a functional group of genes	No. of genes presented on 9k-array	No. of genes expressed differentially in colon ^b		
		Total	Overexpressed in ascending colon	Overexpressed in descending colon
Cell proliferation	234	39	10 (9.1×)	29 (3.3×)
Cell cycle	282	59	7 (1.5×)	52 (2.5×)
Cell death	138	23	11 (1.9×)	12 (3.1×)
Cell-cell signalling	134	13	5 (1.5×)	8 (3.1×)
Cell adhesion	122	21	3 (1.8×)	18 (3.1×)
Embryogenesis and morphogenesis	66	5	1 (3.1×)	4 (3.3×)
Developmental processes	126	27	7 (3.1×)	20 (8.1×)
Signal transduction	503	74	24 (6.2×)	50 (3.2×)
Angiogenesis	93	14	1 (6.7×)	13 (3.1×)
DNA replication, damage, and adducts	166	33	3 (6.7×)	30 (3.1×)
Response to external stimulus	496	75	19 (2.9×)	56 (5.4×)
Stress response	285	38	6 (2.9×)	32 (5.4×)
Oncogenes	177	36	8 (2.6×)	28 (3.3×)
Tumor suppressor genes	101	23	6 (1.8×)	17 (3.3×)
Transcription factors	217	31	10 (6.2×)	21 (3.0×)
Receptors, ligands, and receptor signalling proteins	172	22	7 (6.7×)	15 (3.1×)
EGF signalling	21	3	0	3 (1.8×)
TGFβ signalling	10	0	0	0
Wnt signalling	18	5	2 (1.6×)	3 (1.6×)
Insulin signalling	16	3	0	3 (1.8×)
Ras signalling	16	4	2 (1.4×)	2 (1.8×)
Integrin signalling	23	4	1 (1.3×)	3 (1.8×)
EGF, TGFβ, Wnt, insulin, Ras, and integrin signalling (one set, no repeating genes)	76	15	5 (1.6×)	10 (1.8×)

^a CGAP, Cancer Genome Anatomy Project.

^b At 0.001 level; maximal fold difference shown in parentheses.

also possible that overexpression of these genes reflects greater exposure of the right colon to certain procarcinogenic compounds and concomitant induction of responsive genes rather than constitutive protection from such compounds. Better protection of the right colon against DNA damaging carcinogens is also evidenced in the lower level of *O*⁶-methylguanine in DNA of the proximal colon as compared with DNA derived from the distal colon (25). On a population level, the higher level of expression of 3-β-hydroxysteroid dehydrogenase also serves in protection against tumorigenesis, as was shown in a study of prostate cancer susceptibility (26).

Most of the genes implicated in cell cycle control (notably genes encoding cyclins D1, D2, and G1; cyclin-dependent kinase 2; and PCNA), and all but three genes involved in DNA replication, DNA damage repair, and DNA-adduct metabolism, are overexpressed in the left colon. This suggests that control of cell proliferation and DNA damage repair may be different in the right and left colon, reflecting perhaps a higher proliferative activity in the distal colon (27). Together with the differences in metabolism of procarcinogenic and carcinogenic compounds in the right and left colon mentioned above, these differences in cell cycle and DNA damage control may suggest a basis for the distinct susceptibilities of right *versus* left colon to certain pathways of tumorigenesis. It is provocative that several members of the cytochrome P450 and apolipoprotein families of genes whose polymorphic variants (with different enzymatic activities) have been shown to be associated with variations in colon cancer risk (28) show differential expression in ascending and descending colon.

Future studies of gene expression in the ascending and descending colons of patients from different geographic regions with different patterns (compared with this North American population) of CRC incidence, and different ratios of microsatellite-stable to microsatellite-unstable CRC may help to clar-

ify the relevance of these observed differences in gene expression to colorectal carcinogenesis.

The number of genes expressed differentially in ascending and descending fetal colon is substantially smaller than the number of genes expressed differentially in adult colon, indicating that a more robust and distinctly different pattern between left and right colon is acquired during postnatal life. Three stages are distinguished in human colon development: appearance of stratified epithelium (8 to 10 weeks of gestation), conversion of this epithelium to a villous type with partially developed crypts (12 to 14 weeks), and establishment of the adult type crypt epithelium at ~30 weeks of fetal development (29). Because gene expression was compared in 17–24-week-old fetuses, it is possible that the pattern of gene expression that is characteristic of adult right and left colon is established later, concurrent with the transformation to the adult type colonic epithelium at ~30 weeks of gestation, or subsequently, in response to exposure of the gastrointestinal tract to food. The existence of a developmental program that leads to a distinct pattern of gene expression in ascending and descending colon is nevertheless suggested to be in place by 24 weeks of age.

This suggestion of a fetal developmental program creating distinct gene expression patterns in ascending and descending colon is based on a relatively few genes. We wished to gain an appreciation of whether such small differences might still be informative. To address this question, we compared patterns of gene expression in female *versus* male colon. Only 16 genes were found to be differentially expressed in a batch analysis of 28 female ascending-colon and descending-colon biopsies and 22 male colon biopsies, all from patients without surgery (the probability of getting at least 16 genes significant at the 0.001 level of 8296 filtered genes by chance is 0.088). Nevertheless, all 50 biopsies were classified correctly with the CCP ($P < 0.0005$). Eleven of 16 genes are localized on the sex chromo-

somes: 8 genes overexpressed in male colon are Y-chromosome linked, and 3 genes overexpressed in female colon are X-chromosome linked. The observed pattern offers some reassurance that even small differences in gene expression patterns between two distinct sample sets can be meaningful and informative.

An analysis of biopsies from patients with colon surgery indicates that the gene expression profile of colonic mucosal cells embryologically derived from right and left colon can be modulated after surgery to correspond to the new proximal (in case of right hemicolectomy) or distal (after left hemicolectomy) location. To what extent differences in gene expression in ascending and descending colon are programmed in development, and to what extent they are established postnatally and can be modified by interaction with aging and diet, or by surgery, is one area for future investigation.

References

- Bufill, J. A. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann. Intern. Med.*, *113*: 779–788, 1990.
- Bonithon-Kopp, C., and Benchamiche, A. M. Are there several colorectal cancers? Epidemiological data. *Eur. J. Cancer Prev.*, *8*: S3–S12, 1999.
- Kim, H., Jen, J., Vogelstein, B., and Hamilton, S. R. Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am. J. Pathol.*, *145*: 148–156, 1994.
- Forster, S., Sattler, H-P., Hack, M., Romanakis, K., Rohde, V., Seitz, G., and Wullich, B. Microsatellite instability in sporadic carcinomas of the proximal colon: association with diploid DNA content, negative protein expression of p53, and distinct histomorphologic features. *Surgery*, *123*: 13–18, 1998.
- Pocard, M., Salmon, R. J., Muleris, M., Remvikos, Y., Bara, J., Dutrillaux, B., and Poupon, M. F. Deux colons—deux cancers? Adenocarcinomes coliques proximal ou distal: arguments en faveur d'une cancérogenèse distincte. *Bull. Cancer*, *82*: 10–21, 1995.
- Distler, P., and Holt, P. R. Are right- and left-sided colon neoplasms distinct tumors? *Dig. Dis.*, *15*: 302–311, 1997.
- Sadler, T. W. *Langman's Medical Embryology*. Baltimore: Williams and Wilkins, 1985.
- Van Gelger, R. N., von Zastrow, M. E., Yool, A., Dement, W., Barchas, J. D., and Eberwine, J. H. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA*, *87*: 1663–1667, 1990.
- Simon, R., and Peng, A. BRB-ArrayTools Users Guide, National Cancer Institute, Biometric Research Branch Technical Report 008. Bethesda, MD: National Cancer Institute, 2002.
- McShane, L. M., Radmacher, M. D., Freidlin, B., Yu, R., Li, M-C., and Simon, R. Methods for Assessing Reproducibility of Clustering Patterns Observed in Analyses of Microarray Data. National Cancer Institute Technical Report 002, Feb. 1, 2002, pp. 1–30. Bethesda, MD: National Cancer Institute, 2002.
- Tukey, J. W. Tightening the clinical trial. *Controlled Clin. Trials*, *14*: 266–285, 1993.
- Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O-P., Wilfond, B., Borg, A., and Trent, J. Gene expression profiles in hereditary breast cancer. *N. Engl. J. Med.*, *344*: 539–548, 2001.
- Radmacher, M. D., McShane, L. M., and Simon, R. A paradigm for class prediction using gene expression profiles. *J. Comp. Biol.*, *9*: 505–511, 2002.
- Harris, R. M., Picton, R., Singh, S., and Waring, R. H. Activity of phenol-sulfotransferases in the human gastrointestinal tract. *Life Sci.*, *67*: 2051–2057, 2000.
- Edhemovic, I., Snoj, M., Kljun, A., and Golouh, R. Immunohistochemical localization of group II phospholipase A2 in the tumors and mucosa of the colon and rectum. *Eur. J. Surg. Oncol.*, *27*: 545–548, 2000.
- Hormi, K., Cadiot, G., Kermorgant, S., Dessirier, V., Le Romancer, M., Lewin, M. J. M., Mignon, M., and Lehy, T. Transforming growth factor- α and epidermal growth factor receptor in colonic mucosa in active and inactive inflammatory bowel disease. *Growth Factors*, *18*: 79–91, 2000.
- Messa, C., Russo, F., Caruso, M. G., and Di Leo, A. EGF, TGF- α , and EGF-R in human colorectal adenocarcinoma. *Acta Oncol.*, *37*: 285–289, 1998.
- Liu, L. U., Holt, P. R., Krivosheye, V., and Moss, S. F. Human right and left colon differ in epithelial cell apoptosis and in expression of Bak, a pro-apoptotic Bcl-2 homologue. *Gut*, *45*: 45–50, 1999.
- Fric, P., Sovova, V., Sloncova, E., Lojda, Z., Jirasek, A., and Cermak, J. Different expression of some molecular markers in sporadic cancer of the left and right colon. *Eur. J. Cancer Prev.*, *9*: 265–268, 2000.
- Koenders, P. G., Peters, W. H. M., Wobbes, Th., Beex, L. V. A. M., Magengast, F. M., and Benraad, Th. J. Epidermal growth factor receptor levels are lower in carcinomatous than in normal colorectal tissue. *Br. J. Cancer*, *65*: 189–192, 1992.
- Meltzer, S. J., Zhou, D., and Weinstein, W. Tissue-specific expression of c-Ha-ras in premalignant gastrointestinal mucosae. *Exp. Mol. Pathol.*, *51*: 264–274, 1989.
- Miroshita, T., Inada, K-i., Tsukamoto, T., Kodera, Y., Yamamura, Y., Hirai, T., Kato, T., Joh, T., Itoh, M., and Tasematsu, M. Expression of *Cdx1* and *Cdx2* mRNAs and relevance of this expression to differentiation in human gastrointestinal mucosa—with special emphasis on participation in intestinal metaplasia of the human stomach. *Gastric Cancer*, *4*: 185–191, 2001.
- Bara, J., Nardelli, J., Gadenne, C., Prade, M., and Burtin, P. Differences in the expression of mucos-associated antigens between proximal and distal human colon adenocarcinomas. *Br. J. Cancer*, *49*: 495–501, 1984.
- Masys, D. R., Welsh, J. B., Fink, J. L., Gribskov, M., Klacansky, I., and Corbeil, J. Use of keyword hierarchies to interpret gene expression patterns. *Bioinformatics*, *17*: 319–326, 2001.
- Povey, A. C., Hall, C. N., Badawi, A. F., Cooper, D. P., and O'Connor, P. J. Elevated levels of the pro-carcinogenic adduct, *O*⁶-methylguanine, in normal DNA from the cancer prone regions of the large bowel. *Gut*, *47*: 362–365, 2000.
- Chang, B-l., Zheng, S. L., Hawkins, G. A., Isaacs, S. D., Wiley, K. E., Turner, A., Carpten, J. D., Bleecker, E. R., Walsh, P. C., Trent, J. M., Meyers, D. A., Isaacs, W. B., and Xu, J. Joint effect of *HSD3B1* and *HSD3B2* genes is associated with hereditary and sporadic prostate cancer susceptibility. *Cancer Res.*, *62*: 1784–1789, 2002.
- Ma, Q-Y., Williamson, K. E., and Rowlands, B. J. Variability of cell proliferation in the proximal and distal colon of normal rats and rats with dimethylhydrazine induced carcinogenesis. *World J. Gastroenterol.*, *8*: 847–852, 2002.
- Gertig, D. M., and Hunter, D. J. Genes and environment in the etiology of colorectal cancer. *Semin. Cancer Biol.*, *8*: 285–298, 1998.
- Montgomery, R. K., Mulberg, A. E., and Grand, R. J. Development of the human gastrointestinal tract: twenty years of progress. *Gastroenterology*, *116*: 702–731, 1999.

Distinguishing Right from Left Colon by the Pattern of Gene Expression

Oleg K. Glebov, Luz M. Rodriguez, Kenneth Nakahara, et al.

Cancer Epidemiol Biomarkers Prev 2003;12:755-762.

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/12/8/755
Supplementary Material	Access the most recent supplemental material at: http://cebp.aacrjournals.org/content/suppl/2003/09/12/12.8.755.DC1

Cited articles	This article cites 24 articles, 4 of which you can access for free at: http://cebp.aacrjournals.org/content/12/8/755.full#ref-list-1
Citing articles	This article has been cited by 19 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/12/8/755.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/12/8/755 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.