

## Standardization for Transcriptomic Molecular Markers to Screen Human Colon Cancer

FARID E. AHMED<sup>1</sup>, PAUL W. VOS<sup>2</sup>, STEPHANIE IJAMES<sup>3</sup>, DONALD T. LYSLE<sup>3</sup>, GORDON FLAKE<sup>4</sup>, DENNIS R. SINAR<sup>5</sup>, WADE NAZIRI<sup>6</sup> and STEFAN P. MARCUARD<sup>7</sup>

*Departments of <sup>1</sup>Radiation Oncology, Leo W. Jenkins Cancer Center and <sup>5</sup>Internal Medicine, The Brody School of Medicine (BSOM) at East Carolina University (ECU), Greenville, NC;*

*<sup>2</sup>Department of Biostatistics, School of Allied Health Sciences, ECU, Greenville, NC;*

*<sup>3</sup>Department of Psychology, University of North Carolina, Chapel Hill, NC;*

*<sup>4</sup>Laboratory of experimental Pathology, National Institute of Environmental Health Sciences, Research Triangle Park, NC;*

*<sup>6</sup>Carolina Physicians PA, Greenville, NC; <sup>7</sup>Carolina Digestive Diseases PA, Greenville, NC, U.S.A.*

**Abstract.** *Establishing test performance criteria for a transcriptomic colon cancer marker approach must be carried out in a standardized fashion in order to ensure that the test will perform the same way in any laboratory, anywhere. Condition of sample preservation and shipping prior to total RNA extraction is critical, and we recommend preserving stool samples in an appropriate preservative and shipping them in cold packs so as to keep stools at 4°C. It is not necessary to isolate colonocytes to obtain adequate RNA for testing. It is, however, important to obtain samples from both mucin-rich and non-mucin rich to have a good representation of both left- and right-side colon cancers. Employing a commercial total RNA extraction kit that contains an RLT buffer from Qiagen Corporation (Valencia, CA, USA) removes bacterial RNA from stool preparations and results in a*

*high yield of undegraded RNA of human origin. Genes selected based on the enormous resources of NCI's Cancer Genome Anatomy project give good results. Primers for PCR should span more than one exon. Use of semiquantitative PCR, preferably with several reference housekeeping genes of various copy numbers, depending on tested genes, should enhance confidence in the quantitative results. Having standardized the testing conditions in our ongoing work, it is now imperative that a larger prospective randomized clinical study utilizing stool and tissue samples derived from several control and colon cancer patients, to allow for statistically valid analyses, be conducted in order to determine the true sensitivity and specificity of the transcriptomic screening approach for this cancer whose incidence is on the rise worldwide.*

*Abbreviations:* Ab, antibody; cDNA, copy deoxyribonucleic acid; CD, Crohn's Disease; CGAP, Cancer Genome Anatomy Project; CP, comparative cross point; CRC, colorectal cancer; DGED, Digital Gene Expression Display; E-Method, also referred to as Second Derivative Maximum or CP method; FOBT, fecal occult blood test; GI, gastrointestinal; GLS, Gene Library Summarize; IBD, inflammatory bowel disease; LMM, laser microdissection; mRNA, messenger ribonucleic acid; NCI, National Cancer Institute; RT-qPCR, reverse transcriptase quantitative polymerase chain reaction; rRNA, ribosomal ribonucleic acid; SAGE, serial analysis of gene expression; ss, single stranded; TPC, test performance criteria; UC, ulcerative colitis; UDG, uracil-DNA glycosylase.

*Correspondence to:* Dr. Farid E. Ahmed, Director, General, Environmental and Molecular Toxicology (GEM Tox) Consultants, Inc., 2607 Calvin Way, Greenville, NC, 27834, U.S.A. Tel/Fax +1252 321 7261, e-mail: GEMToxConsultants@yahoo.com

*Key Words:* Adenocarcinoma, colonocyte, diagnosis, inflammatory bowel diseases, laser capture microdissection, RT-qPCR, RNA, staging.

This is our second article that deals with standardization employing transcriptomic markers as screening tools for the detection of colon cancer, particularly at the early stages. While the first article (1) addressed the major experimental design issues and some standardizations, this article deals with other test performance characteristics not addressed there.

Because of the desirability to use noninvasive methods that employ exfoliated stool samples, which represent the entire colon, and that is acceptable to the target population as an initial screen, many investigators have turned to developing molecular screening approaches because the inexpensive fecal occult blood test (FOBT) (both guaic and immunological) have quite a low sensitivity for polyp detection (2-4). Important considerations for developing sound molecular tests include: the number of stool specimens required (one sample, or multiple samples over consecutive days), size of the specimen (whole stool or a sample only), how the specimens are stored and delivered (e.g. room temperature, cold pack, or frozen), and automation as any test that can be automated would be practical and cheaper to apply on a large scale as it reduces the cost of assay implementation (5).

There are many other advantages to using exfoliated stool samples for colorectal cancer (CRC) screening. For example, unlike sigmoidoscopy, a stool reflects the full length of the colorectum, and colonocytes are continuously shed into the fecal stream. Moreover, colonocytes from the colon cancer are 4- to 5-fold greater in number than from normal colonic mucosa (6); therefore, these colonocytes are more abundant than their normal counterpart and, unlike blood, are present in all stool samples from colon cancer patients, therefore partially obviating the use of an enrichment technique to separate tumorigenic from normal colonocytes, thus reducing work and cost of testing. Additionally, because tumorigenic cells contain more RNA than normal ones (7) then stool from colon cancer patients contains RNA that can be adequately measured by a sensitive method such as real-time, quantitative polymerase chain reaction (qPCR) in spite of the dilution exerted by normal colonocytes. Furthermore, because testing can be performed on mail-in-specimens, geographic access to stool screening is not hindered. Recently, use of commercial preparations to conveniently isolate high quality, undegraded RNA that can be reverse transcribed (RT) and used for qPCR, has made the transcriptomic option a feasible and convenient choice, as RNA is perceived to be preferable than either epigenetic, DNA- or protein-based markers for CRC screening (1).

## Materials and Methods

*Acquisition of clinical specimens.* Stool and tissue samples were obtained from ten control subjects and fifteen patients with various stages of colon adenocarcinoma (Dukes' stages 0 to 3), five patients with ulcerative colitis (UC) and five patients with Crohn's disease (CD), according to an approved Medical Center Institutional Review Board (IRB) protocol. All Laboratory work was carried out and standardized under blind conditions and followed ECU's guidelines for handling biohazardous material established by its Biological Safety & Hazardous Substance Committee.

### *Control stool and tissue samples.*

(i) *Fecal specimens.* Control stool samples (20 g) were collected from consenting individuals visiting our GI Clinic/Endoscopy Lab who did not show any polyps or inflammatory bowel diseases, such as colitis or diverticulitis. Stool samples were stored overnight at 4°C in a bacteriostatic preservative S.T.A.R medium (Roche Diagnostics, Indianapolis, IN, USA), followed by RNA extraction the following day. In situations where longer storage of fecal specimens was desired, the preservative RNALater® (Ambion, Austin, TX, USA) was added at 2.5 ml per 1 g of stool, to allow longer storage at 4°C and RNA extraction within 72 h. Extracted RNA can be stored safely and without fear of degradation at -70°C for lengthy time periods.

(ii) *Tissue specimens.* Normal tissues were usually obtained from a small piece of colon tissue (about 0.5 cm<sup>3</sup>) removed >10 cm away from diseased patient tissue at surgery (8), or from biopsies taken during colonoscopy from non-diseased areas of consenting

individuals. For UC or CD patients, a small piece of tissue taken further away from the inflamed or diseased tissue was considered normal. Tissues were flash frozen in liquid nitrogen and stored at -70°C for subsequent laser capture microdissection (LCM) work. Longitudinal sectioning of the tissue before LCM use was employed to pick up epithelial cells that would eventually be shed as colonocytes into stools from bottom cells among the proliferative enterocyte crypt lineage (1).

### *Experimental stool and tissue samples from cancerous or inflamed patients.*

(i) *Fecal specimens.* A 20 g sample of feces (bowel movement) was collected the night before surgery or earlier, before administering any laxative, in a plastic container containing either: a) bacteriostatic preservative S.T.A.R® medium, which was then covered and either the total RNA was extracted immediately, or the sample was kept overnight at 4°C then processed the next day for RNA extraction, or b) RNALater® to allow longer storage at 4°C, then the sample was processed followed by either storage of the extracted RNA at -70°C or by RT and PCR analysis.

(ii) *Tissue specimens.* A small piece of tissue sample (about 0.5 cm<sup>3</sup>) was obtained after colonoscopy for adenoma, or at surgery for carcinoma. Samples were processed after flash freezing in liquid nitrogen and storage at -70°C for subsequent microdissection. Longitudinal LCM sectioning was performed and the marked areas of the crypt indicated where the transformed cells (*i.e.*, adenoma, carcinoma) were to be captured by LMD for subsequent RNA extraction.

For the current study, stool and normal tissue samples were obtained from 10 non-cancerous non IBD control individuals, 5 patients having adenomatous polyps ≥1 cm with high-grade dysplasia (stage 0-1), 5 patients with stage 2 carcinoma, 5 patients with stage 3 carcinoma, 5 non-cancerous patients with severe UC, and five non-cancerous patients with severe CD. Each patient provided a stool sample, thus totaling 35 stool samples. Tissue samples were obtained from only one of the UC patients and only one of the CD patients, but were obtained for each of the remaining patients, providing a total of 27 tissue samples.

*Isolation of exfoliated colonocytes from stool.* Approximately 5 to 10 g of freshly-collected wet feces were homogenized with 200 ml of a buffer consisting of Hank's solution, 10% fetal bovine serum (FBS) and 25 mmol/L Hepes buffer (pH 7.35) at 200 rpm for 1 minute. The homogenates were filtered through a nylon filter (pore size 512 µm), followed by dividing the homogenate into five portions (~40 ml each) in 50 ml polystyrene tubes. Subsequently, 40 µl of Dynal superparamagnetic polystyrene beads (4.5 µm diameter) (Invitrogen, Carlsbad, CA, USA) coated with a mouse IgG1 monoclonal antibody (Ab) Ber-Ep4 (Dako, Glostrup, Denmark) specific for an epitope on the protein moiety of the glycopolypeptide membrane antigen Ep-CAM, which is expressed on the surface of human epithelial cells, including colonocytes and colon carcinoma cells (9, 10), were added to the each tube *i.e.* at a final concentration of 12 ng of Ab/mg magnetic beads (1 µg Ab/10<sup>6</sup> target cells). The mixtures were incubated for 30 minutes under gentle rolling at 15 rounds/minute in a mixer at room temperature. The tubes were then placed on a magnet (Dynal MPC-1) and incubated on a shaking platform for 15 minutes at room temperature.

A few drops of the solution were spread on a glass slide, dried and stained with Diff-Quick stain (Fisher Scientific, Pittsburgh, PA, USA) to visualize the colonocyte preparation. The supernatant was then removed, a fresh 1 ml of Hanks solution added to each tube, followed by transferring the bead suspension to a new 1.5 ml microcentrifuge tube that was placed into a Dynal magnet MPC-S. The new supernatant was also removed and the pellet containing colonocytes were stored at  $-70^{\circ}\text{C}$  until RNA extraction (11).

Colonocytes were retrieved from the stool of 15 patients with CRC and from 10 control individuals. The primary tumors were located in the following sites: sigmoid colon in 2 patients, descending colon in 4, transverse colon in 3 and ascending colon in 6. The clinical stage of the patient was: Dukes' stage A in 5 patients, stage B in 5, and stage C in 5.

*Selection of cancerous or inflamed cells from colon tissue of patients by LCM.* LCM was employed as an enrichment technique for tumors isolated from colon adenocarcinoma patients to separate transformed cells from nonneoplastic stromal and inflammatory cells (12). Details of the procedure are described elsewhere (1).

*Extraction of total RNA from LCM cells and ss-cDNA preparation.* This procedure was used for extracting RNA from a small number of LCM-captured cells. It was carried out according to the manufacturer's specifications using the RNeasy isolation Kit<sup>®</sup> from Qiagen, Valencia, CA, USA, as previously described (13, 14). The quality of RNA was determined on an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc, Palo Alto, CA, USA) utilizing the RNA 6000 Nano LabChip<sup>®</sup>, or by electrophoresis on Superload (Viagen, Austin, TX, USA) agarose gels (15) and RNA quantified with RiboGreen quantification reagent (16) (Molecular Probes, Eugene, OR, USA). The "Sensiscript RT Kit<sup>®</sup> from Qiagen was then employed for making a copy of ss-DNA, resulting in 40 ml of ss-cDNA, of which 2-3 ml was subsequently amplified by PCR. One hundred thousand captured cells on 5 plastic LCM caps (each accommodating 20,000 cells) were sufficient to test all 9 genes of interest, considering that each cell contains  $\sim 20$  pg total RNA or 0.4 pg mRNA (equivalent to 0.36 pg ss-cDNA), as only a few picograms of cDNA are needed per PCR reaction (17).

*Extraction of total RNA from stool samples and from colonocytes.* The following procedure was used to extract RNA from stool samples (13, 14) that have been preserved overnight either in Roche bacteriostatic S.T.A.R<sup>®</sup> medium (Indianapolis, IN, USA) or preserved in Ambion's RNALater<sup>®</sup> for longer storage periods, both at  $4^{\circ}\text{C}$ . This method was also employed for extraction of total RNA from colonocytes.

The RNeasy Mini Kit from Qiagen was used for extraction of RNA from stool samples as follows: to about 0.25 g of stool, 1 ml of Buffer RLT supplemented with 10  $\mu\text{l}$  of  $\beta$ -mercaptoethanol and 1 ml of 70% ethanol were added, allowed to remain for 10 min to stimulate differential lysis of bacterial RNA present in the preparation, and the tube was vortexed. The sample was then applied to a mini spin column sitting in a 2 ml collection tube, and centrifugation carried out for 15 s at  $\geq 8000\text{g}$ . Three consecutive washes, each containing 670  $\mu\text{l}$  of RW1 buffer, were pipetted onto the column. Each wash was centrifuged for 15 s at  $\geq 8000\text{g}$ . After washing, the column was transferred to a new 2-ml tube. Two washes of RPE buffer, each 500  $\mu\text{l}$ , were pipetted onto the column and followed by the elution of the RNA with 30  $\mu\text{l}$  of RNase-free water.

Another kit from Qiagen (Sensiscript Reverse Transcriptase Kit) that produces high quality ss-cDNA from  $<50$  ng of total RNA was employed for making cDNA from the small amount of RNA isolated from stool and LCM samples. A master mix containing  $<50$  ng total RNA and final concentrations of 1x RT buffer, 0.5 mM of each dNTP, 1  $\mu\text{M}$  oligo-dT primer, 10 units of RNase inhibitor per reaction and 1  $\mu\text{l}$  of Sensiscript reverse transcriptase per 20  $\mu\text{l}$  reaction volume was used. The reaction was incubated at  $37^{\circ}\text{C}$  for an hour followed by inactivation of the reaction by heating at  $95^{\circ}\text{C}$  for 5 min, after which the reaction tube was rapidly chilled on ice.

For extraction of RNA from colonocytes, the same procedure described above was used on 1,000 isolated colonocytes. Total RNA yield was  $\sim 2$   $\mu\text{g}$ , sufficient for RT-qPCR on 9 genes.

*Determination of quality and content of extracted RNA.* The quality and yield of the RNA extraction for each sample was determined in an Agilent 2100 Bioanalyzer and quantitated using RiboGreen RNA reagent. Good preparations showed two sharp ribosomal 18S and 28S rRNA bands, which corresponded well to the expected sizes of 1.9 kb and 3.7 kb, respectively, when electrophoresed on Superload formaldehyde-agarose gels (Viagen), in addition to smaller peaks of 5S rRNA and other micro rRNAs below 0.2 kb (1).

*Two-step polymerase chain reaction (PCR) on ss-cDNA.* Both conventional (qualitative endpoint PCR) and semi-quantitative real-time PCR were used to study the expression of selected genes in a two-step RT-PCR as detailed elsewhere (1).

The E-method method analyzes the amplification efficiency of the target and reference gene using relative standards of serial dilutions of a single normal sample (*e.g.* undiluted, 1:10, 1:100, etc). Furthermore, the method normalizes for run-to-run differences and uses one sample of the relative standards as a common reference point for comparison of all experiments within the series (18-21). It was selected to calculate CPs because it produces more accurate results than the  $2^{-\Delta\Delta\text{CT}}$  method (22) as it compensates for differences in target and reference gene amplification efficiency [ $E=10^{-1/\text{slope}}$ ] (23) either within an experiment, or between experiments (1).

*Selection of primer, probes, genes and PCR conditions.* A web-based assay design software (ProbeFinder) was used for selecting target-specific PCR primers. It is accessible at Exiqon site (<http://www.universalprobelibrary.com>), or *via* the Roche Applied Science home page ([www.roche-applied-science.com](http://www.roche-applied-science.com)). The locked nucleic acid, LNA, probes are short (only 8-9 nucleotide long) in which the four nucleotide bases have been substituted with high affinity nucleotide analogs (*e.g.* which are conformationally locked in a C3'-endo/N-type sugar conformation that leads to reduced conformation flexibility). These conformational restraints raise melting temperatures to  $\sim 80^{\circ}\text{C}$  under standard hybridization conditions, which assures duplex stability and specificity in real-time RT-qPCR assays (24). Primers have been selected to span more than one exon. The selected primers were also validated using the Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/BLAST/>). It was also important to determine whether folding of the mRNA might interfere with primer access during the RT step using the Mfold program

(<http://www.bioinfo.rpi.edu/applications/mfold/old/rna>) (25). Validated HPLC purified primers for this study were obtained from Invitrogen.

We selected 9 genes for our present study using two expression analysis tools on the CGAP site: the cDNA Digital DGED (<http://cgap.nci.nih.gov/Tissues/GXS>) and the SAGE DGED (<http://cgap.nci.nih.gov/SAGE/SDGED?METHOD=SS10,LS10&ORG=Hs>) (26). These tools allowed us to create arbitrary pools of libraries and then find genes that are overexpressed in one pool in contrast to the other pool (*i.e.* measure relative gene expression). Because of great differences in the underlying specimens and in the method of mapping tags/sequences to genes, the two tools often give different results (1).

Primers were prepared for nine genes showing increased expression by the serial analysis of gene expression (SAGE) method or microarrays [Insulin-like growth factor II (*IGF2*), FilaminAa (*FLNA*), transforming growth factor  $\beta$ -*igh* induced (*TGF $\beta$ -igh*), CDC28 protein kinase regulatory subunit 2 (*CKS2*), chromosome segregation 1-like (*CSEIL*), chemokine (C-X-C motif) ligand 3 (*CXCL3*), dipeptidase 1 (*DPEP1*), kallikerin 10 (*KLK10*)]; in addition to a standard housekeeping gene (hypoxanthine phosphoribosyl transferase, *HPRT*). The rationale and parameters employed for gene selection and primer sequences are detailed elsewhere (1).

The PCR Mix for thirty-three, 20- $\mu$ l reactions was prepared by adding to a 1.5 ml reaction tube on ice: 10.4  $\mu$ l of water; primers at a final concentration of 200 nM in a volume of 0.2  $\mu$ l each for the forward and reverse primers; probe concentration of 100 nM in a volume of 0.2  $\mu$ l; 4  $\mu$ l of ready-to-use Roche LightCycler<sup>®</sup> FastStartTaqMan<sup>®</sup> Probe Master and 5  $\mu$ l of cDNA template at 7  $\mu$ g/ml (50 ng template). The PCR running conditions were: one cycle at 95°C for 10 min to activate the polymerase, 25-40 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 30 s and extension at 72°C for 1 s. Control samples (negative control) to exclude contamination, in which cDNA was replaced by H<sub>2</sub>O, were run in parallel with each experiment.

**Amplification specificity of studied genes.** The amplification specificity on all nine genes studied was evaluated by running 1% agarose gels on products of endpoint PCR in parallel with real-time PCR to: (i) confirm and determine the analytical specificity of the RT-PCR reaction, and (ii) verify the ability of the Universal probes, specific for studied genes, to bind the PCR product. We performed a conventional 25  $\mu$ l qualitative endpoint PCR reaction, running 10  $\mu$ l of the reaction product on an agarose gel, followed by transfer of the DNA into a Biotran<sup>™</sup> Nylon membrane (ICN, Irvine, CA, USA) using a downward capillary transfer. After crosslinking the DNA to membranes by UV at 100 mJ/cm<sup>2</sup>, a short hybridization probe that is specific for the internal sequence of the PCR product end-labeled with digoxigenin using terminal deoxynucleotidyl transferase (Promega Corporation, Madison, WI, USA) was prepared and hybridized. The signal was detected by chemiluminescence using alkaline phosphatase-conjugated anti-digoxigenin antibody and CDP-star substrate (Roche Diagnostics). Digital capture of light emission was carried out using Alpha Innotech chemiluminescent imaging instrument (San Leandro, CA, USA) (13).

**Quality control (QC) considerations and optimization for the PCR.** QC procedures (27-29) were employed as described elsewhere (1).

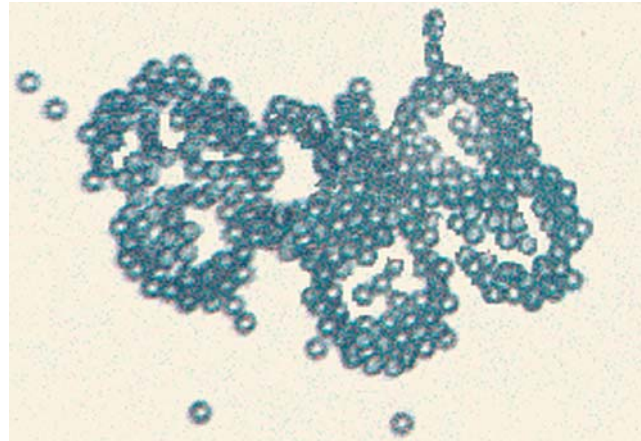


Figure 1. Colonocytes isolated from a stool of a healthy control individual by the immunological paramagnetic bead separation method. Diff-Quick staining (x100).

## Results

Figure 1 shows colonocytes isolated immunologically by the paramagnetic polystyrene bead method. Colonocytes isolated by this method appeared round, measuring about 4  $\mu$ m in size, and the yield was  $\sim 10^5$  per g of wet stool weight. Although no atypical cells were observed in colonocytes of healthy controls, atypical cells were observed in colonocytes from stools of 3 out of 15 patients.

We estimate that using the method outlined herein, we are able to retrieve between 75-250  $\mu$ g of human RNA per gram of stool from a cancer patient, depending on stage (the higher the stage the higher the amount of RNA extracted), and  $\sim 25$   $\mu$ g human RNA per gram of stool from control individuals. Because our method employed RLT lysis buffer (Qiagen), which was reported to lyse bacterial RNA if treatment lasts for 5-10 min (30), our samples were free from bacterial RNA.

We determined the quality of RNA isolated from stools and tissue from a total of 11 individuals (normal, with cancer and with inflammation) using both gels and electrophoretograms obtained from the Agilent 2100 Bioanalyzer (Agilent Corporation, Palo Alto, CA, USA) (Figures 2 and 3). There is no difference in the profile between RNA extracted from LCM tissue or stool. It is evident that the entire total RNA extracted (from either LCM tissue, stool of normal or of various stages of colon cancer, or from patients with inflammation) is of high quality, is of human origin and is intact. Although there may be daily variations between runs as shown by comparing the first 6 gels (A-F) that were processed on one day with the other 5 gels (G-K) that were processed in the next day, the overall electrophoretogram patterns

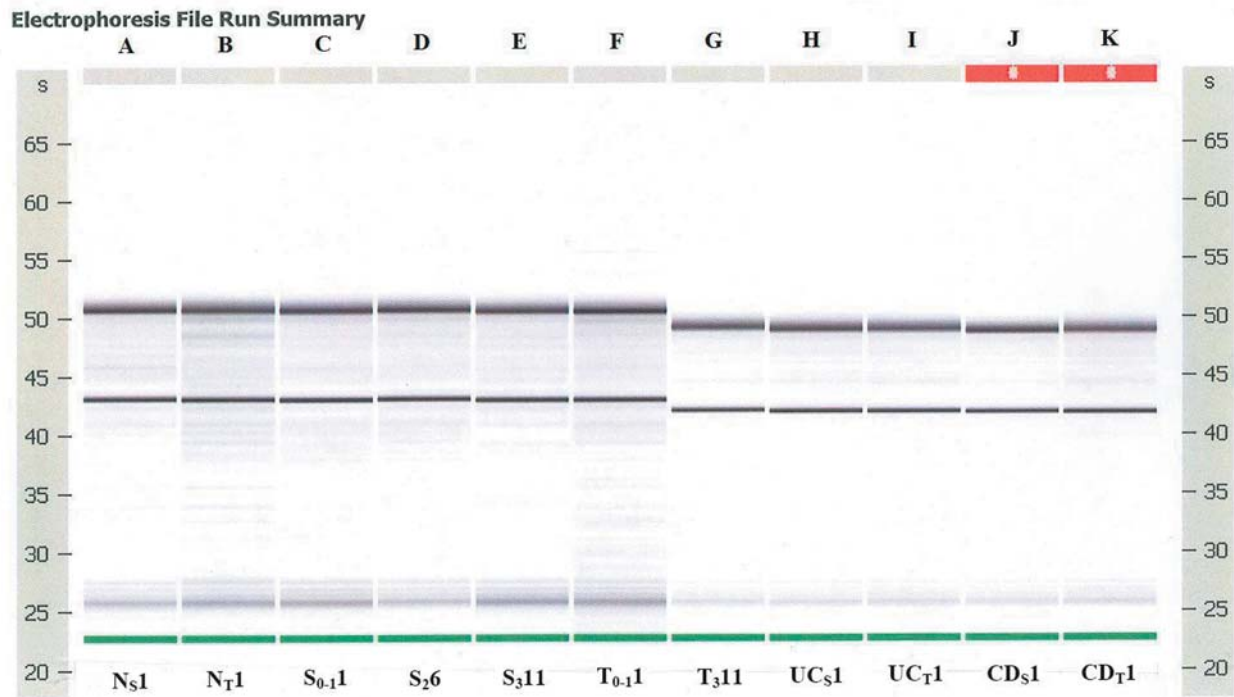


Figure 2. Representative Superload agarose gels for stool and LCM tissue samples showing nondegraded RNA from stool of 11 individuals who are either normal, with various stages of colon cancer or IBD.  $N_s1$ , stool from a normal individual;  $N_t1$ , tissue from a normal individual;  $S_{0-1}$ , stool from a patient with adenomatous colon polyp 1 cm (stage 0-1);  $S_{26}$ , stool from colon carcinoma patient (stage 2);  $S_{311}$ , stool from colon carcinoma patient (stage 3);  $T_{0-1}$ , LCM tissue from a patient with adenomatous colon polyp 1 cm (stage 0-1);  $T_{211}$ , LCM tissue from colon carcinoma patient (stage 2);  $T_{311}$ , LCM tissue from colon carcinoma patient (stage 3);  $UC_s1$ , stool from a patient with severe ulcerative colitis (UC);  $UC_t1$ , LCM tissue from a patient with severe UC;  $CD_s1$ , stool from a patient with Chron's disease(CD);  $CD_t1$ , LCM tissue from a patient with severe CD.

show that the 18S and 28S bands (third and fourth band from the left of the y axis, respectively) migrate to the same extent in all eleven total RNA samples taken from stool or tissue.

Gene expression was measured on colonocytes kept in different preservatives and stool preparation conditions. As seen in Table I and Figures 4 and 5, no change in gene expression pattern was observed in stool samples preserved overnight at 4°C in S.T.A.R<sup>®</sup> medium or 72 h in RNALater<sup>®</sup>. Gene expression values were also not different if determined from isolated colonocytes, or if total RNA was directly extracted from stool as shown in Table II and Figure 6. However, it was not possible to isolate colonocytes or extract RNA from stool if samples were frozen at -70°C even if only overnight, because of colonocyte rupture due to ice crystal formation as observed earlier (13).

As in our previous publication (1), the data presented in Tables I and II, and Figures 4-6 show that the selected genes can distinguish between non-cancerous and cancerous patients, and can also separate between various Dukes' stages.

## Discussion

*Convenience of improved methods and better commercial preparations.* Use of immunological paramagnetic beads for colonocyte isolation is a convenient method that yielded satisfactory results, and colonocytes were less distorted than for example, when compared with those obtained by the Percoll centrifugation methods, which results in only about 2% of the yield obtained by the present method (31, 32).

Our yield of extracted high quality RNA was at least 10 times more than that reported about 9 years ago by Alexander and Raicht (7) due to our use of better preservatives (Roche's S.T.A.R<sup>®</sup> or Ambion's RNALater<sup>®</sup>) and improved commercial preparations for RNA extraction (Qiagen RNeasy Mini Kit and Sensiscript Reverse Transcriptase).

*Standardizing stool preservation.* Our results have shown that it does not matter if the samples were preserved in Roche's S.T.A.R<sup>®</sup> medium overnight, or in RNALater<sup>®</sup> for 72 h provided that samples are kept at 4°C, rather than at room (21°C) or a higher temperature (*e.g.* 37°C). We also recommend that stool samples be shipped in cold packs

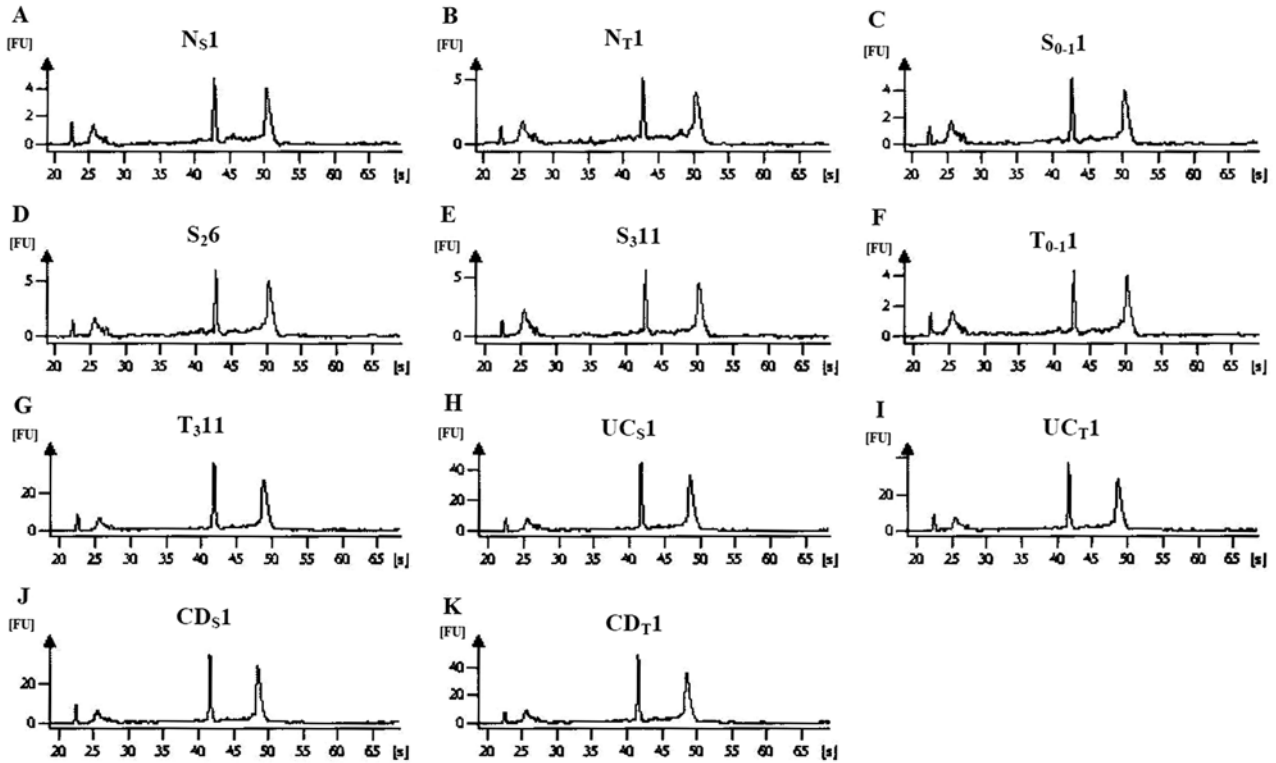


Figure 3. An Agilent 2100 electrophoretogram showing the 28S, 18S and tRNA, 5.8S and 5S bands (from left to right) for the same stool and tissue samples as in Figure 1.

since this will not add much to the cost of sample shipping, and will keep stool samples at the 4°C temperature critical for preserving colonocytes without lysis in stool samples. In agreement with our findings, other investigators have also reported that they are able to obtain almost the same number of colonocytes from stools 3 days after fecal evacuation on fresh sampling, if the material has been kept at 4°C (11).

*Choice of reference housekeeping gene in semi-quantitative PCR for marker development.* In this study, data normalization was carried out using the endogenous housekeeping gene (*HPRT*) with a low copy number because the efficiencies of amplification of the control and experimental genes appeared to be equal, as judged by similar slopes (21, 33). In a recent study of the effect of RNA quality on reference housekeeping gene stability, *HPRT1* was ranked as relatively stable in intact RNA of nasal polyposis and chronic rhinosinusitis samples, but was quite unstable in degraded samples. Therefore, the quality of employed RNA affects qPCR results and it is advisable to use only high quality nondegraded RNA. In our study, we measured RNA integrity on every sample tested and discarded/replaced degraded ones, since we did not wish to

miss the small changes in expression encountered with some of the selected genes.

However, for optimum results, it would be instructive to use more than one housekeeping gene for comparison of expression values. For example, use could be made of  $\beta$ -amyloid and prothobinogen deaminase genes with intermediate copy numbers (the latter also being pseudogene-free),  $\beta$ -actin with a high-copy number, and 18S ribosomal RNA (rRNA) with an invariable copy number. Thus, for a low copy number target gene in a colon sample, an external standard that is also present at a low level in the same sample should be employed, and so on. Selection of various reference genes should further improve on the utility of selected genes to distinguish between cancerous *versus* normal patients, and between various Dukes' stages.

Although semi-quantitative PCR methods were thought to be inferior to quantitative competitive ones, side-by-side comparisons showed both assays produced equivalent measures of template abundance (35). This is because methods employing real-time qPCR measure amplification in the logarithmic phase, whereas quantitative competitive measurements determine amplification during the linear phase (33). While the choice of valid markers for colon cancer using a transcriptomic approach has been difficult

Table I. Comparative CPs of gene expression in stool and tissue from normal, cancer and IBD individuals after two preservatives.

Group	S.T.A.R Medium									RNA Later								
	HPRT	IGF2	FLNA	TGFβ <sub>high</sub>	CKS2	CSE1L	DPEP1	KLKI0	CXCL3	HPRT	IGF2	FLNA	TGFβ <sub>high</sub>	CKS2	CSE1L	DPEP1	KLKI0	CXCL3
NegCont <sup>a</sup>	37.01	37.15	37.45	36.99	37.16	37.71	37.87	37.63	37.13	37.15	36.87	36.96	37.19	37.37	37.56	37.69	37.88	36.96
Ns1	37.9	36.85	37.06	37.37	37.77	36.92	37.25	37.43	36.87	36.99	37.22	37.11	36.79	37.14	37.61	36.83	37.18	37.36
2	36.93	37.14	36.55	36.81	37.17	37.58	37.39	37.77	36.79	37.67	36.66	37.12	36.96	36.79	37.32	37.19	36.99	37.56
3	37.45	37.61	36.79	36.92	37.46	37.84	37.38	37.46	36.48	37.41	37.53	37.26	36.75	36.84	37.63	37.28	37.67	37.24
4	36.55	37.48	37.91	36.76	37.24	37.33	37.47	36.65	37.45	36.91	37.46	36.86	36.99	37.06	37.38	37.83	37.54	37.22
5	37.56	37.42	36.76	37.21	36.64	36.79	37.52	37.43	37.11	37.09	36.78	37.03	37.63	36.79	36.94	37.33	36.87	36.88
6	37.43	37.15	37.73	36.66	37.32	37.24	37.41	36.59	37.37	37.36	37.58	37.49	37.09	36.99	37.32	37.55	37.69	37.41
7	37.41	37.96	37.51	37.09	36.96	37.83	37.26	36.37	37.08	36.77	36.89	37.33	37.11	37.84	36.93	36.59	37.21	37.63
8	36.95	37.32	37.45	37.39	37.88	37.56	37.32	36.62	36.55	37.39	37.83	37.57	37.39	37.93	37.42	37.27	37.21	36.84
9	37.56	37.23	36.89	36.72	37.24	37.86	37.12	36.69	37.73	36.83	37.35	36.59	37.94	36.85	37.27	37.74	37.09	37.66
10	36.99	37.37	37.29	37.39	37.22	37.85	37.39	37.21	37.85	37.08	36.89	36.69	37.94	37.79	36.55	36.79	37.55	37.32
NT1	37.75	36.49	36.79	37.42	37.08	37.63	37.29	36.85	36.78	37.63	37.53	37.42	36.86	37.48	37.56	37.35	36.73	37.61
2	37.33	36.87	36.93	37.55	37.46	37.08	37.62	37.82	36.69	37.41	37.82	37.58	37.25	36.84	37.35	37.44	37.48	37.41
3	36.87	36.72	37.55	37.37	36.74	36.39	37.31	37.22	37.11	36.76	36.94	37.74	36.63	37.33	37.81	37.44	37.18	36.59
4	37.31	37.53	37.58	36.49	36.62	37.29	37.09	36.69	37.42	37.13	37.49	37.25	36.95	36.79	37.55	37.39	37.44	37.32
5	37.33	37.07	36.79	36.88	37.63	37.59	37.69	36.84	36.79	36.76	37.79	37.49	37.11	36.94	36.52	37.06	37.77	37.39
6	37.43	37.02	36.84	36.79	37.55	37.28	37.61	36.95	37.22	36.89	37.34	36.88	37.55	37.06	37.47	37.51	37.39	37.48
7	36.79	36.83	37.52	37.32	37.39	36.69	36.49	37.26	37.35	36.74	37.59	37.59	37.52	36.93	36.95	36.84	36.49	37.31
8	37.57	37.82	36.62	36.58	36.89	37.22	37.45	37.07	37.29	37.66	37.27	36.89	36.84	37.15	37.38	37.94	37.09	37.42
9	37.98	37.83	37.81	36.78	36.39	36.28	36.59	37.93	37.29	36.87	37.55	36.38	36.83	37.28	37.01	36.69	37.47	37.32
10	37.37	37.48	36.97	36.69	37.11	37.55	36.59	36.85	37.34	37.11	37.27	37.85	37.73	36.36	36.29	36.33	37.15	37.28
S0-11	37.21	24.01	26.15	28.08	29.56	31.46	32.5	33.98	35.67	37.57	23.89	25.55	27.68	30.67	31.85	32.54	33.69	34.95
2	37.87	23.88	26.01	27.98	29.21	31.14	32.78	34.01	35.86	36.88	24.01	25.67	26.99	29.88	31.78	32.34	33.89	35.31
3	37.91	23.79	25.96	27.81	29.56	31.67	32.35	33.87	35.59	37.09	24.23	25.64	27.32	30.07	30.99	31.98	34.32	35.29
4	36.98	24.36	26.12	27.67	29.24	31.55	32.15	33.79	35.42	37.54	24.66	25.06	27.09	30.22	31.24	32.25	34.57	35.43
5	37.23	24.09	25.85	28.27	29.89	31.06	32.88	33.95	35.13	36.86	23.98	25.71	27.55	30.16	31.47	32.44	34.07	35.32
S26	37.22	22.84	23.59	25.17	26.55	27.99	29.15	31.02	32.68	37.11	22.92	23.66	25.07	26.14	28.88	29.07	30.89	32.33
7	36.89	22.91	23.44	25.81	26.25	28.01	29.31	30.79	32.71	36.98	22.12	23.24	25.32	26.58	29.01	29.55	30.89	31.87
8	37.47	22.79	23.64	25.89	26.78	27.69	28.99	31.81	32.19	37.21	22.74	23.49	25.89	26.47	28.77	29.26	30.71	32.22
9	36.73	22.53	23.69	25.37	26.49	28.23	29.21	31.44	32.45	36.79	22.87	23.27	25.09	26.79	29.15	30.01	30.98	32.55
10	37.36	22.32	23.59	25.89	26.11	28.15	29.11	30.88	31.99	36.78	22.55	23.14	25.59	26.33	28.77	29.45	31.03	32.16
S311	37.11	19.05	21.19	24.22	26.67	27.05	28.38	29.01	30.25	36.77	19.21	21.21	25.54	26.21	27.21	28.66	29.17	31.94
12	37.36	18.85	21.11	24.49	26.99	27.85	28.08	28.14	30.45	37.14	18.34	20.79	24.48	26.95	27.78	28.49	29.69	30.89
13	37.52	19.36	20.88	24.74	26.86	27.99	28.12	28.53	30.78	36.84	19.04	22.01	24.79	26.85	27.59	28.08	29.34	31.21
14	36.64	19.08	20.58	25.36	26.23	27.13	28.61	28.38	30.87	37.11	18.8	21.37	25.21	26.29	27.69	28.37	28.75	31.33
15	37.29	18.75	21.06	24.43	26.57	27.87	28.24	28.78	30.34	37.41	18.5	20.89	24.87	26.48	27.23	28.28	29.22	30.99
T0-11	37.34	22.01	23.67	25.93	28.75	29.24	30.87	31.08	32.21	37.78	21.16	22.85	26.04	28.94	30.01	30.59	30.99	33.07
2	36.87	21.89	23.23	26.01	28.97	29.89	31.09	32.01	32.68	37.11	21.34	22.26	25.98	29.03	30.15	31.22	31.88	33.01
3	37.53	22.11	23.26	26.11	28.88	29.78	30.89	32.11	32.97	37.21	20.87	23.04	25.65	29.01	29.99	31.14	31.56	32.87
4	36.87	22.42	23.87	26.11	29.06	30.09	31.11	31.89	33.15	36.98	21.44	22.76	25.49	28.98	30.19	30.98	31.49	32.98
5	37.31	22.19	23.43	25.98	29.11	30.22	30.98	31.88	33.09	37.11	21.36	22.99	26.09	29.11	30.78	31.25	32.07	32.79
T26	36.88	19.09	21.69	24.67	25.67	26.78	27.67	29.15	31.89	37.23	18.96	21.23	24.33	25.23	26.88	27.78	29.33	30.67
7	37.11	19.19	20.57	25.87	25.88	26.23	27.66	29.56	31.56	36.99	18.88	21.01	25.67	25.14	26.99	27.53	28.89	30.14
8	36.78	18.85	21.09	24.99	25.49	26.18	27.78	30.01	32.21	37.22	18.36	20.99	25.12	25.68	26.67	27.34	28.99	30.78
9	37.09	18.95	21.14	25.11	25.87	26.41	27.07	29.78	32.11	36.88	18.7	19.99	24.89	25.16	26.46	27.67	29.36	30.79
10	37.22	18.67	21.07	25.8	25.58	26.59	27.46	29.54	30.99	37.08	18.82	20.89	25.07	25.88	26.36	27.99	29.34	30.19
T311	36.89	14.69	19.01	22.56	24.46	25.23	26.05	28.01	29.17	37.22	19.04	18.88	22.99	24.13	25.06	26.56	28.11	29.36
12	37.34	15.89	18.78	22.78	24.22	24.89	26.79	27.79	29.36	37.05	15.33	18.55	22.67	24.66	25.15	26.24	28.01	29.18
13	36.88	15.66	18.23	22.89	24.77	25.11	26.65	28.08	29.16	37.26	15.55	18.06	22.78	24.78	25.71	26.71	28.23	29.47
14	37.23	14.78	18.56	22.86	24.64	25.05	26.31	27.88	29.38	37.22	14.92	18.22	22.42	24.34	25.36	26.55	27.57	29.78
15	36.75	15.01	18.09	23.01	24.61	25.13	26.45	28.11	29.47	36.78	14.69	18.13	23.03	24.79	25.23	26.76	27.88	28.96
UCs1	37.12	36.78	36.99	37.15	36.37	36.39	37.11	36.54	36.28	37.01	36.55	36.49	37.28	37.89	37.07	38.11	36.84	36.95
2	37.36	37.32	37.21	37.63	37.36	37.28	37.16	37.88	36.67	36.69	37.64	37.23	36.32	36.89	37.51	37.33	37.04	37.11
3	36.42	37.37	37.15	37.22	36.53	36.67	37.21	37.88	36.87	37.22	36.55	37.34	36.77	36.78	37.42	37.11	36.73	36.79
4	37.08	36.88	37.23	36.74	37.29	36.77	36.32	37.02	37.35	36.59	37.48	37.53	36.59	37.32	37.28	36.87	37.22	37.23
5	37.29	37.55	36.79	36.75	37.49	37.69	36.84	36.59	36.99	37.22	37.16	36.88	37.32	36.79	37.27	37.35	36.69	37.08
UCT1	36.75	36.93	37.58	37.27	36.93	36.38	36.64	37.08	37.11	37.15	36.87	37.24	37.38	36.59	36.48	36.78	37.33	36.91

Table I. continued

Table I. *continued*

Group	S.T.A.R Medium										RNA Later							
	HPRT	IGF2	FLNA	TGFβ <sub>high</sub>	CKS2	CSE1L	DPEP1	KLKI0	CXCL3	HPRT	IGF2	FLNA	TGFβ <sub>high</sub>	CKS2	CSE1L	DPEP1	KLKI0	CXCL3
CDs1	37.55	37.73	37.23	37.51	36.87	37.32	37.55	36.76	37.55	37.76	37.13	37.47	37.26	37.89	37.11	37.55	36.83	37.07
2	37.23	36.59	37.84	36.95	37.38	36.84	37.39	37.44	36.92	36.73	36.81	37.77	36.74	36.82	37.39	37.32	37.39	37.37
3	37.42	37.51	37.51	37.68	37.32	37.89	36.88	36.49	37.27	37.26	37.34	36.69	37.11	37.27	37.48	37.08	37.25	36.99
4	36.69	37.34	37.27	36.95	37.16	37.08	37.77	37.49	36.83	37.45	37.06	37.27	36.56	37.14	36.86	37.53	37.24	37.05
5	37.34	37.55	37.28	37.39	36.15	37.26	37.83	36.38	37.69	36.83	37.29	37.27	37.16	37.38	37.02	36.77	38.01	37.25
CDT1	36.69	36.45	36.29	36.77	36.55	37.18	36.69	37.21	36.93	36.55	36.74	36.08	37.03	36.48	37.07	36.84	36.39	37.02

<sup>a</sup>No DNA added to reaction (negative control). This table was obtained from 35 individuals: 10 non-cancerous controls, 15 patients with colon cancer, 5 patients with severe ulcerative colitis (UC), and 5 patients with Crohn's disease(CD). Ns1 (NT1) to Ns10 (NT10) , A stool (Tissue) from normal individuals 1 to 10; S0-11 (T0-11) to S0-15 (T0-15)= Stool (LCM Tissue) from patients 1 to 5 with adenomatous colon polyp ≥ 1 cm (stage 0-1); S26 (T26) to S210 (T210) = Stool (LCM Tissue) from colon carcinoma patients 6 to 10 (stage 2); S311 (T311) to S315 (T315) = Stool (LCM Tissue) from colon carcinoma patients 11 to 15 (stage 3); UCS1 (UCT1) =Stool (LCM Tissue) from first patient with severe UC; UCS2 to UCS5 = Stool from patients 2 to 5 with severe UC; CDs1 (CDT 1) = Stool (LCM Tissue) from first patient with CD; CDS2 to CDS5 = Stool from patients 2 to 5 with CD. Samples were preserved in S.T.A.R<sup>®</sup> medium overnight at 4°C. For RNALater<sup>®</sup>, preservation was carried out at 4°C for 27 h before RNA extraction.

Table II. *Comparative CPs of gene expression in isolated colonocytes and stool samples from normal individuals and cancer patients.*

Group	Isolated colonocytes										Stool samples							
	HPRT	IGF2	FLNA	TGFβ <sub>high</sub>	CKS2	CSE1L	DPEP1	KLKI0	CXCL3	HPRT	IGF2	FLNA	TGFβ <sub>high</sub>	CKS2	CSE1L	DPEP1	KLKI0	CXCL3
NegCont <sup>a</sup>	37.36	36.48	37.79	37.83	37.55	36.49	37.57	36.49	36.33	37.32	37.07	37.55	36.69	36.88	37.22	36.98	37.36	37.11
NorC/S1	36.66	37.46	36.31	36.09	37.14	37.81	36.79	37.22	37.43	37.22	36.58	36.94	37.35	37.07	36.97	37.58	36.52	37.08
NorC/S2	37.48	37.29	37.03	37.42	37.77	36.92	37.25	37.43	36.87	36.99	37.22	37.11	36.79	37.14	37.61	36.83	37.18	37.36
C/S0-13	37.29	23.55	25.01	27.56	29.89	31.68	32.67	33.82	35.17	37.16	23.21	24.78	27.22	30.01	31.15	32.22	33.69	34.99
C/S0-14	36.79	24.01	25.11	27.87	30.03	31.58	32.91	33.79	35.01	37.52	23.69	25.09	27.59	29.37	30.99	33.02	33.83	35.15
C2/S25	37.22	22.84	23.59	25.17	26.55	27.99	29.15	31.02	31.68	37.11	22.92	23.66	25.07	26.14	28.02	29.07	30.89	32.33
C2/S26	36.79	22.41	23.36	25.48	26.89	27.69	29.09	30.87	31.99	37.07	23.04	23.57	25.61	26.55	27.89	29.18	30.71	31.87
C2/S27	37.37	22.15	23.75	25.08	26.33	27.45	29.57	30.51	31.44	36.78	22.45	23.49	25.45	26.11	27.61	29.26	30.43	32.04
C3/S38	37.26	19.45	21.36	24.68	25.78	26.89	28.58	29.41	30.99	37.11	19.23	21.42	24.52	25.61	26.71	28.42	29.13	30.84
C3/S39	36.15	19.51	21.11	23.98	25.43	26.76	28.43	29.32	31.03	37.82	18.98	21.33	24.09	25.33	26.19	28.21	29.45	30.88
C3/S310	37.09	18.98	21.79	23.35	25.13	26.35	28.27	29.67	31.27	37.31	18.75	21.44	23.67	25.63	26.69	29.01	29.59	31.46

<sup>a</sup>No DNA added to reaction (negative control). This table was obtained from 10 individuals: 2 non-cancerous controls and 8 patients with colon cancer. NorC/S1 and NorC/S2=Isolated colonocytes/Stool from normal individuals 1 and 2; C<sub>0-1</sub>/S<sub>0-1</sub> 4 and C<sub>0-1</sub>/S<sub>0-1</sub> 5=Isolated colonocytes/Stool from patients 3 and 4 with adenomatous colon polyp ≥ 1 cm (stage 0-1); C<sub>2</sub>/S<sub>2</sub> 5 to S/C<sub>2</sub>7=Isolated colonocytes/Stool from colon carcinoma patients 5 to 7 (stage 2); C<sub>3</sub>/S<sub>3</sub> 8 to 10=Isolated colonocytes/Stool from colon carcinoma patients 8 to 10 (stage 3). All samples were preserved overnight at 4°C in S.T.A.R<sup>®</sup> preservative before RNA extraction.

because there are no known colon cancer specific genes, and carrying out a reproducible RT-qPCR requires numerous validations due to issues of RNA degradation, PCR inhibitors, template quality, biological replicas and issues of sensitivity and specificity (36), the use of RT-qPCR has been shown to result in a sensitive and specific measure for developing diagnostic markers for CRC, particularly at the early Dukes' stages, by such an inclusive transcriptomic approach (1) and present work.

*Use of transcriptomic approach for molecular staging of colon cancer.* Although we have used relatively few patients, it is evident from Figures 4 through 6 that this transcriptomic approach could result in an improved molecular noninvasive

screening method for staging colon cancer patients over the current histopathological procedures that use qualitative parameters such as depth of tumor penetration, histopathological grade of the tumor, presence of lymphovascular invasion, and extracapsular extension beyond the lymph nodes (37). Naturally, we need more CRC patients to ascertain whether all Dukes' stages can be routinely detected with higher sensitivity and specificity with this approach than with current measures.

*Differences between mutation and expression measurements in feasibility assessments.* To show how biological measurements may be used to derive numerical values for feasibility,

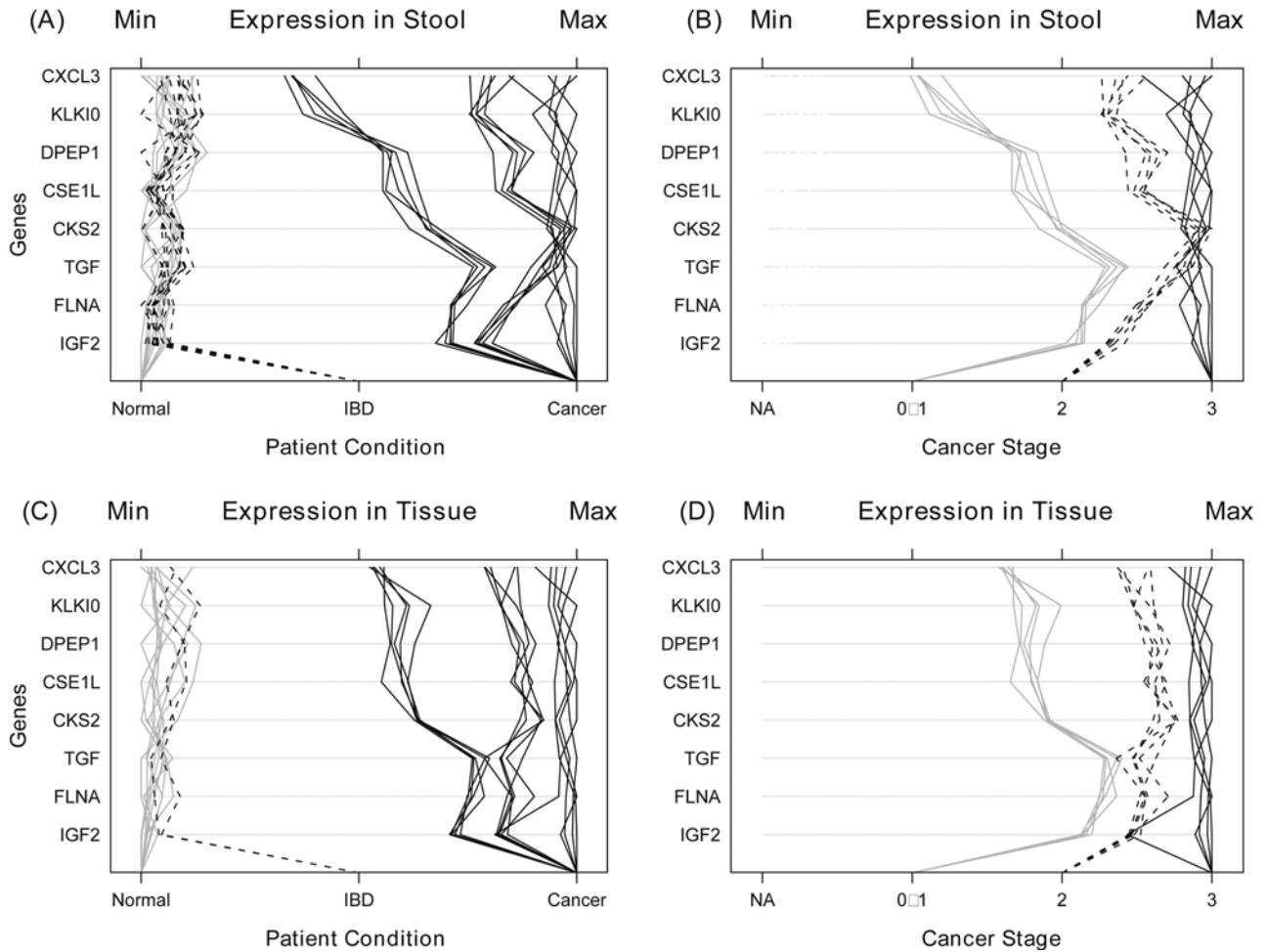


Figure 4. Gene expression in stool samples taken from 35 individuals preserved in Roche's S.T.A.R. medium. A) The condition of the patient is indicated by the bottom row of the panel and by the type of line. There were 10 normal patients (gray lines), 10 with inflammatory bowel disease (IBD) (dashed lines), and 15 with cancer (black lines). Instances of high expression appear on the right and those with low expression on the left. Expression was measured by CP and scales were chosen so that minimum values line up on the Min mark labeled at the top of the panel. The same is true for the maximum values which line up under the mark labeled Max. B) This panel displays gene expression for stool samples taken from 15 cancer patients. Stage of cancer is indicated by the bottom row of the panel and by the type of line. There were 5 patients with stage 0 or 1 (gray lines), 5 with stage 2 (dashed lines), and 5 with stage 3 (black lines) cancer. The 10 noncancerous patients (stage NA) are not shown. C) Gene expression in tissue samples taken from 35 individuals. Conditions of the patient are the same as in panel A. D) This panel displays gene expression for tissue samples taken from 15 cancer patients as in panel B. Stages of cancer are also indicated as in panel B.

consider two samples of 100 cells each. In the first sample (non cancerous), all 100 cells are normal, while in the second sample (cancerous) 1 cell is cancerous while 99 are normal. If for simplicity we assume that normal cells have a mutation rate of 1 mutation unit (mu) and the cancerous cell have mutation rates of 100 mu (a 100-fold increase), then in the normal sample, mutation yield is  $100 \times 1 \text{ mu} = 100 \text{ mu}$ . On the other hand, in the cancerous sample, the total mutation yield is  $99 \times 1\text{mu} + 1 \times 100\text{mu} = 199 \text{ mu}$ , which is about a two fold increase over the normal rate. Numerical values for CRC mutations have been provided for *APC* (38), *k-ras* (39), and *TP35* genes (40).

In the case of gene expression, the situation is, however, different as there is no definitive yes or no answer, and it is not possible to derive a transcription unit as easily. For example, cancer genes are grouped into two classes: class I genes that are mutated or deleted, and Class II genes that are not altered at the DNA level, but that affect the phenotype by expression changes (47). To further complicate the picture, some genes may be in class I in a certain situation and class II in others *e.g.* the breast cancer *BRCA1* gene that can be mutated in some familial breast cancer patients (48) making it a class I gene, while nonsporadic mutations were found in some patients without a family history of breast cancer, but

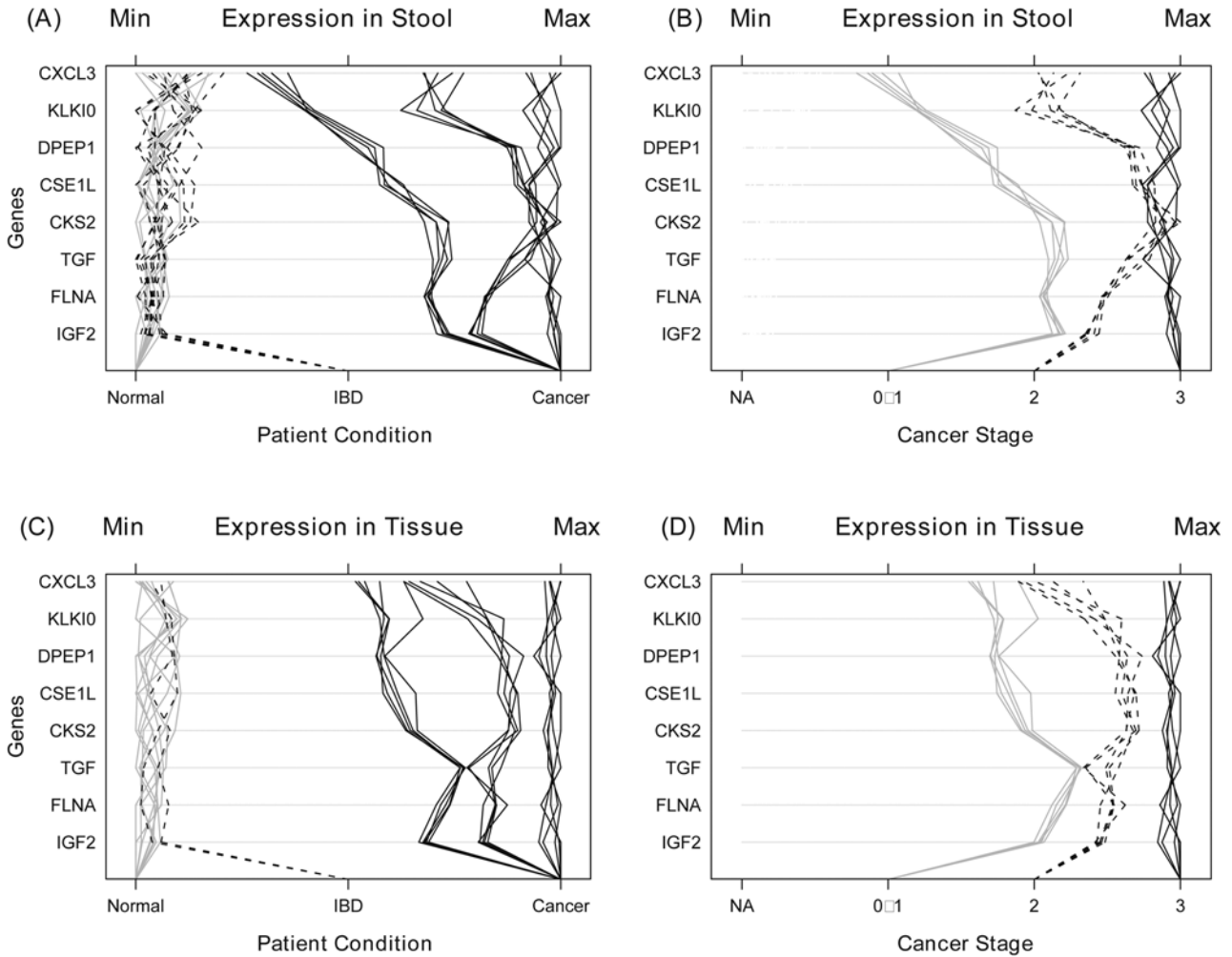


Figure 5. Gene expression in stool samples taken from 35 individuals preserved in Ambion's RNALater. The legend as in Figure 4 for panels A-D.

rather a reduced amount of *BRCA1* mRNA representing down-regulation of the wild-type gene (49) making it a class II tumor suppressor gene. Add to that the presence of many splice variants and the degree of complexity of gene expression patterns compared to mutation becomes clear. Therefore, in the case of gene expression, one cannot simply quantify the neoplasia-derived RNA as a fraction of total human cellular RNA, but rather an expression index, which is defined as a product of two or more genes divided by the expression of another gene or a product of two or more genes, is commonly derived (41-45). In that situation, we look at how much that index can predict tumors at a certain sensitivity and specificity. We have shown in this article and a previous one (1) that using a transcriptomic approach, we can detect changes close to >95% sensitivity and >95% specificity, which is better than anything else on the market, including mutations, proteomics and epigenetic changes.

Due to inadequacies in reporting results of many tumor markers and the difficulty in interpreting and comparing data from different articles, the National Cancer Institute-European Organization for Research and Treatment of Cancer (NCI-EORTC) have recently published guidelines for tumor marker studies in order to encourage transparency and adequate reporting so that the relevant information becomes available (46). These guidelines have been observed while carrying out this discovery project in order to set the ground for an independent validation set. Having achieved that aim, the next step will naturally be to design a prospective randomized study using enough number of patients (600 control individuals and 600 colon cancer patients) to be able to have statistical confidence in its outcome using the methods that we have outlined in our publications (1), (13) & current study) as proof of principle.

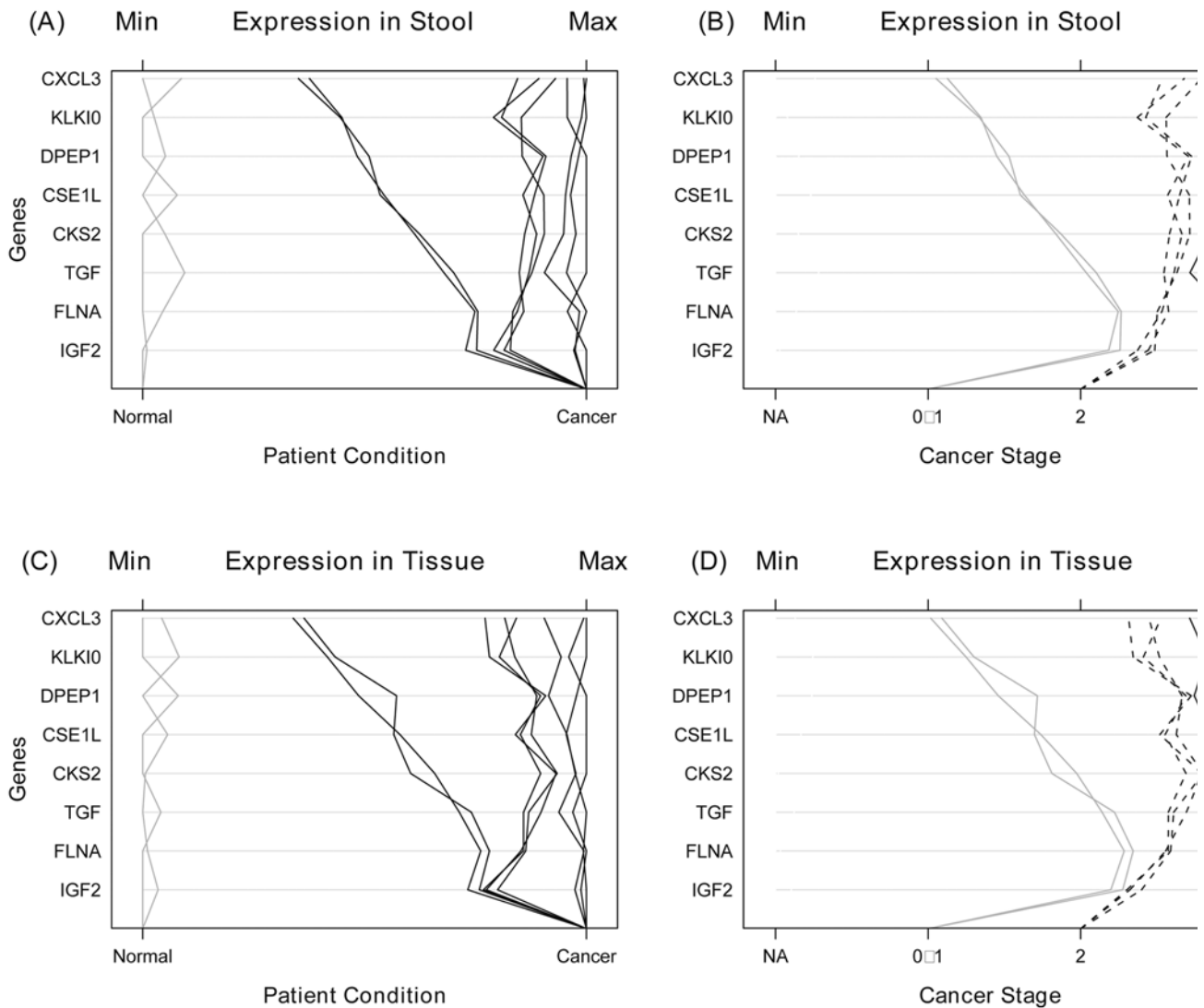


Figure 6. A) Gene expression in colonocytes extracted from stool preserved in Roche's S.T.A.R. medium, which was taken from 10 individuals. The condition of the patient is indicated by the bottom row of the panel and by the type of line. There were 2 normal patients (gray lines), and 8 with cancer (black lines). Instances of high expression appear on the right and those with low expression on the left. Expression was measured by CP and scales were chosen so that minimum values line up on the Min mark labeled at the top of the panel. The same is true for the maximum values which line up under the mark labeled Max. B) This panel displays gene expression for colonocytes extracted from 8 cancer patients. Stage of cancer is indicated by the bottom row of the panel and by the type of line. There were 2 patients with stage 0 or 1 (gray lines), 3 with stage 2 (dashed lines), and 3 with stage 3 (black lines) cancer. The noncancerous patient (stage NA) is not shown. C) Gene expression in stool samples, without colonocyte extraction, preserved in Roche's S.T.A.R. medium, which was taken from the same 10 individuals as in A. Conditions of the patient are the same as in panel A. D) This panel displays gene expression for stool samples, without colonocyte extraction, taken from 8 cancer patients as in panel B. Stages of cancer are also indicated as in panel B.

## Acknowledgements

We express our gratitude to Gerard J. Nuovo of Ohio State University's Department of Pathology; Michelle Moore and Rodney Pennington of Roche Applied Science for technical advice and supplies; and Carl F. Schafer of the NCI's Center for Bioinformatics for compiling the CGAP database and advice on gene selection.

## References

- 1 Ahmed FE, Vos P, James S, Lysle DT, Allison RR, Flake G, Sinar DR, Naziri W, Marquard ST and Pennington R: Transcriptomic molecular markers for screening human colon cancer in stool and tissue. *Cancer Genom Proteom* 4: 1-20, 2007.
- 2 Morikawa T, Kato J, Yamaji T, Wada R, Miksushima T and Shiratori Y: Comparison of the immunochemical fecal occult

- blood test and total colonoscopy in the asymptomatic population. *Gastroenterology* 125: 422-428, 2006.
- 3 Song K, Fendrick AM and Ladabaum U: Fecal DNA testing compared with commercial colorectal cancer screening methods: a decision analysis. *Gastroenterology* 126: 1270-1279, 2004.
  - 4 Allison JE: Colon cancer screening guidelines 2005: the Fecal Occult Blood Test option has become a better FITT. *Gastroenterology* 125: 745-748, 2005.
  - 5 Davies R J, Miller R and Coleman N: Colorectal cancer screening: prospects for molecular stool analysis. *Nature Rev Cancer* 5: 199-209, 2005.
  - 6 Osborn NK and Ahlquist DA: Stool screening for colorectal cancer: molecular approaches. *Gastroenterology* 128: 192-206, 2005.
  - 7 Alexander RJ and Raicht RF: Purification of total RNA from human stool samples. *Dig Dis Sci* 43: 2652-2658, 1998.
  - 8 Deschner EE. Early proliferative changes in gastrointestinal neoplasia. *Am J Gastroenterol* 77: 207-211, 1982.
  - 9 Salem RR, Wolf BC, Sears HF, Lavin PT, Ravikumar TS, DeCoste D, D'Emilia JC, Herlyn M, Schlom J and Gottlieb LS: Expression of colorectal carcinoma-associated antigens in colonic polyps. *J Surg Res* 55: 249-255, 1993.
  - 10 Winter MJ, Nagtegaal ID, van Krieken JH and Litvinov SV: The epithelial cell adhesion molecule (Ep-CAM) as a morphoregulatory molecule is a tool in surgical pathology. *Am J Pathol* 163: 2139-2148, 2003.
  - 11 Matsushita HM, Matsumura Y, Moriya Y, Akasu T, Fujita S, Yamamoto S, Onouchi S, Saito A, Sugito M, Ito M, Kozu T, Minowa T, Momura S, Tsunoda H and Kakizoe T: A new method for isolating colonocytes from naturally evacuated feces and its clinical application to colorectal cancer diagnosis. *Gastroenterology* 129: 1918-1927, 2005.
  - 12 Ahmed FE: Laser capture microdissection: application to carcinogenesis. *Cancer Genom Proteom* 3: 217-226, 2006.
  - 13 Ahmed FE, IJames S, Lysle DT, Dobbs LJ Jr, Johnke RM, Flake G, Stockton P, Sinar DR, Naziri W, Evans MJ, Kovacs CJ and Allison RR: Improved methods for extracting RNA from exfoliated human colonocytes from stool and RT-PCR analysis. *Dig Dis Sci* 49: 2650-2659, 2004.
  - 14 Ahmed FE and Vos P: Molecular markers for human colon cancer in stool and blood identified by RT-PCR. *Anticancer Res* 24: 4127-4134, 2004.
  - 15 Gregg K, Zhou W and Davis S: Enhanced sensitivity RNA gel loading buffer that enables efficient RNA separation on native gels. *BioTechniques* 36: 334-336, 2004.
  - 16 Hashimoto JG, Beadles-Bohling AS and Wiren KM: Comparison of RiboGreen and 18S rRNA quantitation for monitoring real-time RT-PCR expression analysis. *BioTechniques* 36: 54-60, 2004.
  - 17 Kuecker SJ, Jin L, Kulig E, Oudrago GL, Roche PC and Lloyd RV: Analysis of *PRL*, *PRL-R*, *TGFβ1* and *TGFβ-RII* gene expression in normal neoplastic breast tissue after laser capture microdissection. *Appl Immunol Molec Morph* 7: 193-200, 1999.
  - 18 Tellman G: The E-method: a highly accurate technique for gene-expression analysis. *Nature Methods* 3: 1-2, 2006.
  - 19 [http://www.lightcycler-online.com/lc\\_sys/soft\\_nd.htm#quant](http://www.lightcycler-online.com/lc_sys/soft_nd.htm#quant).
  - 20 LightCycler Software®, Version 3.5. Roche Molecular Biochemicals, Mannheim, Germany, pp. 64-79, 2001.
  - 21 Pfaffl M W: Quantification strategies in real-time PCR. In: *A-Z of Quantitative PCR*. Bustin SA (ed.). International University Line, La Jolla, CA, USA, pp. 89-120, 2004.
  - 22 Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC<sub>T</sub></sup> method. *Methods* 25: 402-408, 2001.
  - 23 Tichopad A, Dilger M, Schwarz G and Pfaffl MW: Standardised determination of real-time PCR efficiency from a single reaction setup. *Nucleic Acids Res* 31: 21-33, 2003.
  - 24 Kennedy B, Arar K, Reja V and Henry RJ: Locked nucleic acids for optimizing displacement probes for quantitative real-time PCR. *Anal Biochem* 348: 294-299, 2006.
  - 25 Ahmed FE: qRT-PCR: application to carcinogenesis. *Cancer Genom Proteom* 2: 317-332, 2005.
  - 26 Lal A, Lash AE, Allschul SF, Velculescu V, Zhang L, McLendon RE, Marra M, Prange C, Morin P J, Polyak K, Papadopoulos N, Vogelstein B, Kinzler KW, Strausberg RL and Riggins GJ: A public database of gene expression in human cancers. *Cancer Res* 59: 5403-5407, 1999.
  - 27 Muller R, Padmabandu G and Taylor RH: Standardization of qPCR and qRT-PCR assays. In: *A-Z of Quantitative PCR*. Bustin S A (Ed.). International University Line, La Jolla, CA, USA, 2004.
  - 28 Longo MC, Berninger MS and Hanley JL: Use of uracil DNA glycosylase to control carryover contamination in polymerase chain reaction. *Gene* 93: 125-128, 1990.
  - 29 Thornton CG, Hartley JL and Rashtchian A: Utilizing uracil DNA glycosylase to control carryover contamination in PCR: characterization of residual UDG activity following thermal cycling. *BioTechniques* 13: 181-183, 1992.
  - 30 Di Cello F, Xie Y, Paul-Satyaseela M and Kim KS: Approaches to bacterial RNA isolation and purification for microarray analysis of *Escherichia coli* K1 interaction with human brain microvascular endothelial cells. *J Clin Microbiol* 43: 4197-4199, 2005.
  - 31 Albaugh GP, Igenyar V, Lohani A, Malayeri M, Bala S and Nair PP: Isolation of exfoliated colonic epithelial cells, a novel non-invasive approach to the study of cellular markers. *Int J Cancer* 52: 347-350, 1992.
  - 32 Yamao T, Matsumura Y, Shimada Y, Moriya Y, Sugihara KKKI, Akasu T, Fujita S and Kakizoe T: Abnormal expression of CD44 variants in the exfoliated cells in feces of patients with colorectal cancer. *Gastroenterology* 114: 1196-1205, 1998.
  - 33 Ahmed FE: Molecular methods for studying differential gene expression in carcinogenesis. *J Env Sci Health C20: 77-116*, 2002.
  - 34 Pérez-Novo CA, Claeys C, Speleman F, Van Cauwenberge P, Bachert C and Vandesoimplele J: Impact of RNA quality on reference gene expression stability. *BioTechniques* 39: 52-56, 2005.
  - 35 Haberbaugen G, Pinsl J, Kuha C-C and Markert-Hahn C: Comparative study of different standardization concepts in quantitative competitive reverse transcription PCR assay. *J Clin Microbiol* 36: 628-633, 1998.
  - 36 Nolan T, Hands R E, Ogunkolade W and Bustin SA: SPUD: A quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Anal Biochem* 351: 308-310, 2006.
  - 37 Greene FL, Page DL, Fleming ID, Fritz AG, Blach CM, Haller DG and Morrow M: *AJCC Cancer Staging Handbook*, 6th edition. Springer, New York, 2002.

- 38 Travreso G, Shuber A, Levin B, Johnson C, Olsson L, Schoetz DJ Jr, Hamilton S-R, Boynton K, Kinzler KW and Vogelstein B: Detection of APC mutations in fecal and DNA from patients with colorectal tumors. *New Eng J Med* 346: 311-320, 2002.
- 39 Vogelstein B and Kinzler KW: Digital PCR. *Proc Natl Acad Sci USA* 96: 9236-9241, 1999.
- 40 Dong SM, Traverso G, Johnson C, Geng L, Favis R, Boynton K, Hibi K, Goodman SN, D'Allesio M, Paty P, Hamilton SR, Sidransky D, Barany F, Levin B, Shuber A, Kinzler KW, Vogelstein B and Jen J: Detecting colorectal cancer in stool with the use of multiple genetic targets. *J Natl Cancer Inst* 93: 858-865, 2001.
- 41 Willey JC, Crawford EL, Jackson CM, Weaver DA, Hoban JC, Khuder SA and DeMuth JP: Expression measurement of many genes simultaneously by quantitative RT-PCR using standardized mixtures of competitive templates. *Am J Resp Cell Mol Biol* 19: 6-17, 1998.
- 42 DeMuth JP, Jackson CM, Weaver DA, Crawford EL, Durzinsky DS, Durham SJ, Zaher A, Phillips ER, Khuder SA and Willey JC: The gene expression index *cmyc* x *E2F-1/p21* is highly predictive of malignant phenotypes in human bronchial epithelial cells. *Am J Respir Cell Mol Biol* 19: 18-29, 1998.
- 43 Ahmed FE: Colon cancer: molecular staging models that characterize risk and enhance progression. *Cancer Genomics Proteomics* 1: 295-310, 2004.
- 44 Ahmed FE: Development of novel diagnostic and prognostic molecular markers for sporadic colon cancer. *Exp Rev Mol Diag* 5: 337-352, 2005.
- 45 Ahmed FE: Colon cancer, prevalence, screening, gene expression and mutation, and risk factors and assessment. *J Env Sci Health C2I*: 65-131, 2003.
- 46 McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M and Clark GM: Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 97: 1180-1184, 2005.
- 47 Sager R: Expression genetics in cancer: shifting the focus from DNA to RNA. *Proc Natl Acad Sci USA* 94: 952-955, 1997.
- 48 Miki Y, Swensen J, Shattuck-Eidens D, Futreal AP, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett ML, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannon-Albright L, Goldgar D, Wiseman R, Kamb A and Skolnick MH: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266: 66-71, 1994.
- 49 Thompson ME, Jensen RA, Obermiller PS, Page DL and Holt JT: Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. *Nature Genet* 9: 444-450, 1995.

*Received August 8, 2007*  
*Revised September 29, 2007*  
*Accepted October 16, 2007*