Absolute dPCR Quantification of MicroRNAs for the Diagnostic Screening of Colon Cancer

Farid E Ahmed1*, Mostafa M Gouda23, and Nancy C Ahmed1

1GEM Tox Labs, Institute for Research in Biotechnology, 2905 South Memorial Drive, Greenville, NC 27834, USA
2Department of Nutrition & Food Science, National Research Center, El-Bohoot Street, Dokki, Cairo, Egypt
3National Research and Development Center for Egg Processing, College of Food Science and Technology, Huazhong Agricultural University, Wuhan, Hubei, PR, China

*Corresponding author: Farid E Ahmed, GEM Tox Labs, Institute for Research in Biotechnology, 2905 South Memorial Dr, Greenville, NC 27834, USA, Tel: 252-689-5682; E-mail: gemtoxconsultants@yahoo.com

Abstract

No validated micro(mi)RNA diagnostic test currently exist. We first carried a microarray study on immunocaptured stool colonocytes on 15 subjects and selected a panel of 14 preferentially expressed mature miRNAs associated with colon cancer (12 Up-Regulated, miR-19a, miR-20a, miR-21, miR-31, miR-34a, miR-96, miR-106a, miR-133a, miR-135b, miR-206, miR-224 and miR-302; and 2 Down-Regulated, miR-143 and miR-145). We then used absolute quantitative digital PCR on these 15 stool samples from stages 0-4 using a chip-based Absolute QuantStudio 3D Digital PCR. To ensure that we have chosen human and not bacterial small total RNA, we have carried out co-extraction protocols with E. coli K1 strain RS18, Agilent electrophoretic patterns, and also sequenced random samples throughout this research using mRNA/miRNA sequencing.

Our quantitative dPCR miRNA data showed that the quantitative changes in the expression of a few mature miRNA genes in stool, associated with right and left colon cancer, provides for a more convenient, sensitive and specific diagnostic screening markers, more useful than test markers currently available on the market, such as the low-sensitivity (<15%) fecal occult blood test (FOBT); result in better compliance; and is more economical than the invasive/expensive colonoscopy exam in colon cancer (CC). Initial test performance characteristics of the miRNA approach showed that the test has a high numerical predictive value in colon cancer. Moreover, underpinning of the miRNA markers as a function of total RNA showed that the test can numerically differentiate between control and CC, particularly at the early stages of that cancer. We propose to extend our results to a larger prospective and randomized five-years nested case-control epidemiological study, to validate the expression of the above 14 miRNAs, in stool of 180 individuals in an epidemiologically designed randomized study. Fewer, or even one miRNA, may suffice to serve as an efficient and a quantitative marker for the non-invasive diagnostic screening of CC in human stool. The above approach when combined with bioinformatics' analysis, to correlate miRNA seed data with our previously published messenger (m)RNA target data in stool, allows for a thorough mechanistic understanding of how miRNA genes regulate mRNA expression.

Keywords: Chips; Diagnosis; Epidemiology; Immunobeads

Received Date: January 18, 2019; Accepted Date: February 22, 2019; Published Date: March 01, 2019


© 2019 Tridha Scholars.
Introduction

The discovery of small non-coding protein sequences, 17-27 nucleotides long RNAs (miRNAs), has opened new opportunities for a non-invasive test for early diagnosis of many cancers [1]. The latest miRBase release 22 on, March 12, 2018 [http://www.mirbase.org] indicates the total number of miRNAs labeled “high confidence” has increased by 168, to 1996, then in the previous release [2].

MiRNA functions seem to regulate development [3] and apoptosis [4], and specific miRNAs are critical in oncogenesis [1,5], effective in classifying solid [6-10] & liquid tumors [11,12], and serve as oncogenes or suppressor genes [13]. MiRNA genes are present abundantly at fragile sites, as well as loss of heterozygosity sites 9r amplification of regional break points, hinting to carcinogenic involvement (14). MiRNAs are good candidates for markers of cancer diagnosis, prognosis or response to therapy [1,15-17]. Profiles of miRNA expression differ between normal tissues and tumor types, and evidence suggests that miRNA expression profiles cluster similar tumor types together more accurately than expression profiles of protein-coding mRNA genes [1,18,19].

A study that looked at global expression of 735 miRNAs in 315 specimens from normal mucosa, tubulovillus adenomas, adenocarcinomas proficient (pMMR), and deficient (dMMR) in DNA mismatch repair, representing sporadic and inherited CRC stages I-IV [20]. Results showed that: a) the majority of miRNAs that were differentially expressed in normal and polyps (miR-1, miR-9, miR-31, miR-99a, miR-135b and miR-137) were also differentially expressed with a similar magnitude in normal versus both the pMMR and dMMR tumors, b) all but one miRNA (miR-99a) demonstrated similar expression differences in normal versus carcinoma, suggesting a stepwise progression from non-malignant to malignant CC. Many miRNAs were associated with CC pathways, such as APC/WNT signaling and cMYC, and miR-31, miR-224, miR-552 and miR-590 showed more than 2 folds changes between pMMR and dMMR tumors. The data suggests involvement of common biologic pathways in pMMR and dMMR tumors, in spite of the presence of numerous molecular differences between them, including differences at the miRNA level [20,21].

Unlike mRNA screening, a small number of miRNAs can differentiate cancer from normal [1,10,16,18-20], and miRNAs in stool remain mostly intact for detection [22,23]. Thus, miRNAs are more suited for developing a reliable noninvasive diagnostic screen for CC because a) Escherichia coli presence does not hinder detection of miRNA by a sensitive technique such as dPCR, as the primers employed are selected to amplify human and not bacterial miRNA genes [24], and b) the miRNA expression patterns are the same in primary tumor, or diseased tissue, as in stool samples [1,21,23]. Colonoscopy - obtained from patients’ medical records - is considered as “The Gold Standard” to which the miRNA test is compared. However, because the low sensitivity guaiac FOBT is still the most commonly used screen in annual checkups [25-29] (www.cancer.org), we will also include this test for comparison with our developed miRNA screening test in stool.

Colonocytes purified from human stool is needed for carrying out cytological test, to give a quantitative estimate of the performance of the proposed miRNA approach. Although we may miss exosomal RNA, a parallel test could also be carried out on miRNAs obtained from stool samples to compare the extent of loss when colonocytes are only used, and an appropriate correction for exsosomal loss can then be made [30].

Biomarker validation as outlined in the miRNAs' approach has been designed to test the theory that “quantitative measurement of a carefully-selected miRNA stool panel by dPCR is a reliable, sensitive as well as specific indicator, for the early non-
invasive diagnostic screening of human CC”. To prove this hypothesis, it must first be validated in a study, as proposed herein, using a nested case control epidemiology design and employing a prospective specimen collection, retrospective blind evaluation (PRoBE) of control subjects and test colon cancer patients [31], as specifically delineated by the National Cancer Institute’s (NCI’s) Early Detection Research Network (EDRN) http://edrn.nci.nih.gov for cancer biomarker discovery studies.

Innovation of the dPCR-miRNA stool screening approach lies in the collective use of many methods in the proposed research, such as: immuno-paramagnetic beads [25,26] to capture colonocytes from the complicated stool environment, whose fragile total small RNA is stabilized immediately after stool shedding by commercially-available preparations kits so it does not fragment, followed by standardized analytical quantitative miRNA dPCR-chip profiling in noninvasive stool samples, which are neither labor intensive, nor require extensive sample preparation, to develop a panel of few stable miRNAs for absolute quantitative diagnostic quantification of early CC, cheaply and with more sensitivity/specificity than available commercial CC tests [1, 21-29].

**Epidemiology of Colon Cancer**

Colorectal cancer (CRC) is considered to be the third most common malignancy in the Worldwide, with about one million new cases and a half million fatalities occurring annually [1,23]. Screening for CRC allows early stage diagnosis of the malignancy and potentially reduces disease mortality [28,29]. The cheap fecal occult blood test (FOBT) test shows low sensitivity and necessitates dietary restrictions, impeding its compliance as well as its use [29]. Currently, CRC is the only cancer for which colonoscopy is a recommended screening test [21-23]. Although colonoscopy is a reliable screening tool, the invasive nature, abdominal pain and high cost have hampered worldwide application of this procedure [25] (www.cancer.org). Compared to the currently used low sensitivity FOBT test, a noninvasive sensitive screening paradigm that does not require dietary restriction, seems a more convenient test.

Epidemiologically, colon cancers (CCs) and rectal cancers (RCs) differ in their morbidities and etiologies [32-38]. RC is common in Asia (China) where it accounts for over 50% of CRC, compared with less than 30% in western countries. Data from Peking Union Medical College Hospital, China, indicated that colon and rectal cancers accounted for 55.6% and 44.4% of CRC, respectively, during the years 1989 through 2008, and are more prevalent in younger Asians [39]. In contrast, colon cancer was shown to account for over 60% of CRC cases in the USA and Europe, and is related to fatty foods, less exercise and a Caucasian ethnic origin [27-29], suggesting differences in carcinogenesis between these two types of cancer. Several structural and molecular studies have indicated differences in etiology, clinical manifestation, pathological features and genetic abnormalities between CC and RC [31-33]. The proximal colon, distal colon and rectum have different embryological origin. Molecular studies have shown that tumor suppressor genes, point mutations and genetic instability differ according to whether the subsite is either the colon or the rectum. CC has been reported to more likely have CpG island methylator phenotype and k-ras mutations, whereas rectal and distal colon tumors are more likely to have p53 and APC mutations [34-37]. Gene hybridization techniques have shown amplification of 20q in CC, compared with amplification of 12p in RC [38].

A study indicated significant differences between rectal and colon cancers in the amplification of genes for cell cycle, as cyclin-A2, -B1, -D1 and –E [40]. A study using Illumina HT-12 V4.0 Expression Beadchip oligonucleotide microarrays, found RC to be more complex than CC as 676 genes related to 11 pathways in CC and 1,789 genes related to 30 signal pathways altered in RC, with 824 common differentially expressed genes up- or down-regulated in both Cancers [40], leading to the conclusion
that colon and rectal cancers are two distinct types of malignancies. We have focused in our research on colon cancer as it is more abundant in the USA, and more CC patients report to our Collaborating Clinics, compared to RC.

Materials and Methods
First, we employed a global microarray expression analysis (41-43) using an exfoliated colonocytes enrichment strategy (44-47), which recruited 15 subjects (three controls, three TNM stage 0-1, three stage 2, two stage 3, and three stage 4 colon cancer) in triplicates, using an Affymetrix GeneChip miRNA 2.0 Array, containing 15,644 probe sets that provides 100% miRBase v15 coverage, and selected a panel of miRNAs for subsequent dPCR studies (41-43). Microarray results showed 180 preferentially expressed miRNA genes that were either increased (124 miRNAs), or decreased (56 miRNAs) in expression in stool samples from CC patients. We then carefully selected 14 miRNAs, 12 of them showed increased expression and 2 showed decreased expression, as presented in Tables 1 & Table 2, and Figure 1 below, for analysis of absolute miRNAs expression by a chip-based digital (d) validation PCR study [48-50].

<table>
<thead>
<tr>
<th>MiRNA</th>
<th>Up-Regulated</th>
<th>Down-Regulated</th>
<th>Chromosome Location</th>
<th>Known Putative Cancer Target Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiR-19a</td>
<td>Yes</td>
<td>No</td>
<td>13q31.3</td>
<td>Undetermined</td>
</tr>
<tr>
<td>MiR-20a</td>
<td>Yes</td>
<td>No</td>
<td>13q31.3</td>
<td>PTEN, TMP1</td>
</tr>
<tr>
<td>MiR-21</td>
<td>Yes</td>
<td>No</td>
<td>17q23.1</td>
<td>PTEN,BCL2,PDCCD4,TIMP3,SPRY2,REC,T1AM1</td>
</tr>
<tr>
<td>MiR-31</td>
<td>Yes</td>
<td>No</td>
<td>9p21.3</td>
<td>T1AM1,AX1N1,FOX2C2,FOX3P3,H1F1AN</td>
</tr>
<tr>
<td>MiR-34a</td>
<td>Yes</td>
<td>No</td>
<td>1p36.22</td>
<td>BCL2,TP53,E2F3,NOTCH1,E2F1,S1RT</td>
</tr>
<tr>
<td>MiR-96</td>
<td>Yes</td>
<td>No</td>
<td>17q32.2</td>
<td>KRAS</td>
</tr>
<tr>
<td>MiR-106a</td>
<td>Yes</td>
<td>No</td>
<td>Xq26.2</td>
<td>PTEN,E2F1,RB1</td>
</tr>
<tr>
<td>MiR-133a</td>
<td>Yes</td>
<td>No</td>
<td>18q11.2/20q13.33</td>
<td>BAX,KRAS</td>
</tr>
<tr>
<td>MiR-155b</td>
<td>Yes</td>
<td>No</td>
<td>1q32.1</td>
<td>MSH2</td>
</tr>
<tr>
<td>MiR-200c</td>
<td>Yes</td>
<td>No</td>
<td>12p13.31</td>
<td>ZEB1</td>
</tr>
<tr>
<td>MiR-224</td>
<td>Yes</td>
<td>No</td>
<td>Xp23</td>
<td>Undetermined</td>
</tr>
<tr>
<td>MiR-30a</td>
<td>No</td>
<td>Yes</td>
<td>6q13</td>
<td>RASA1,ERG,SEMA6D,SEMA3A</td>
</tr>
<tr>
<td>MiR-143</td>
<td>No</td>
<td>Yes</td>
<td>5q32</td>
<td>KRAS,MAPK7,DNMT3A</td>
</tr>
<tr>
<td>MiR-145</td>
<td>No</td>
<td>Yes</td>
<td>5q32</td>
<td>TGFBRE,APC,IRS1,STAT1,YES1,FLI1</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of fourteen Up- or Down-regulated microRNAs in human stool.

Our absolute dPCR data tabulated in Table 2, and presented graphically in Figure 1 below, show 14 preferentially expressed mature miRNAs associated with colon cancer (12 Up-Regulated, miR-19a, miR-20a, miR-21, miR-31, miR-34a, miR-96, miR-106a, miR-133a, miR-135b, miR-206, miR-224 and miR-302; and 2 Down-Regulated, miR-143 and miR-145) in stool samples from healthy controls, and stages 0-1 to 4 individuals with colon cancer.

Table 2: Absolute quantification of up-/down-regulated miRNAs in stool by Quantstudiotm 3D chip-based digital PCR.
We further calculated standard deviations (SD) obtained from the one-way NOVA, using the 5-level factor Type (normal, stage01, stage2, stage3, stage4). The adjusted R-squared values representing the proportion of variation explained by Type are also reported. Type was statistically significant for every gene; all p-values were less than 0.000001 (no adjustments for multiple comparisons). These data are tabulated in Table 3, and shown graphically in Figure 1, below.

<table>
<thead>
<tr>
<th>Type</th>
<th>miR-19a</th>
<th>miR-20a</th>
<th>miR-21</th>
<th>miR-31</th>
<th>miR-34a</th>
<th>miR-96</th>
<th>miR-106a</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd</td>
<td>92.239</td>
<td>111.10331</td>
<td>99.76355</td>
<td>146.64101</td>
<td>209.04905</td>
<td>278.47558</td>
<td>301.87638</td>
</tr>
<tr>
<td>r2</td>
<td>99.4831</td>
<td>99.18486</td>
<td>99.34603</td>
<td>98.65141</td>
<td>97.63002</td>
<td>96.13899</td>
<td>96.19772</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>miR-133a</th>
<th>miR-135b</th>
<th>miR-200c</th>
<th>miR-224</th>
<th>miR-30a</th>
<th>miR-143</th>
<th>miR-145</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd</td>
<td>300.06189</td>
<td>409.67168</td>
<td>449.86741</td>
<td>376.84372</td>
<td>424.99723</td>
<td>132.76331</td>
<td>110.89266</td>
</tr>
<tr>
<td>r2</td>
<td>96.85741</td>
<td>95.49454</td>
<td>96.70427</td>
<td>97.61795</td>
<td>97.95389</td>
<td>99.87075</td>
<td>99.91289</td>
</tr>
</tbody>
</table>

Table 3: Representation of SDs and R2 for miRNAs tested by absolute digital PCR. Chip System Chip System.

Figure 1: Absolute Quantification of Up- or Down-Regulated miRNAs in Human Stool by QuantStudioTM 3D Digital PCR Chip System.

For each gene on the graph in Figure 1, we have shown the min and max in order to make the presentation clearer. At top left is high expression Value of 9985, which is the maximum value for that gene, at the bottom one finds the value for the minimum the colors range from dark blue (control) to orange (stage 4). The groups are also distinguished by line type: control (solid), stage 0-1 (long dash), stage 2 (dash), stage 3 (dot), stage 4 (dash and dot). The figure is a parallel coordinate plot made in R [51], using the package MASS [52].

Stool Collection and Storage
Stool was obtained from 15 participating subjects {three healthy controls and twelve colon cancer patients of all the colon cancer stages [three TNM stage 0-1 (e.g., polyps³ 1 cm, villous or tubvillous, or with high grade dysplasia), three stage 2, three stage 3, and three stage]} [23]. All stools were collected with sterile, disposable wood spatulas in clean containers, after stools were freshly passed, it was then placed for storage into Nalgene screw top vials (Thermo Fisher Scientific, Inc., Palo Alto, CA,
USA), each containing 2 ml of the preservative RNA later (Applied Biosystems/Ambion, Austin, TX, USA), which prevents the fragmentation of the fragile mRNA molecule [22], and vials were stored at – 80°C until samples were ready for further analysis. Total small RNA, containing miRNAs, was extracted from all frozen samples at once, when ready, and there was no need to separate mRNA containing small miRNAs from total RNA, as small total RNA was suitable to make ss c-DNAs.

**Extraction of total small RNA**

A procedure used for extracting small total RNA from stool was carried out using a guanidinium-based buffer, which comes with the RNeasy isolation Kit®, Qiagen, Valencia, CA, USA, as we have previously detailed [22]. DNase digestion was not carried out, as our earlier work had demonstrated no difference in RNA yield or effect on RT or PCR after DNase digestion [23, 41-43, 53-57]. The time taken to purify aqueous RNA from all of the 15 frozen stool samples was ~ two hours. Small RNA concentrations were measured spectrophotometrically at λ 260 nm, 280 nm and 230 nm, using a NanoDrop spectrophotometer (Themo-Fischer Scientific). The integrity of total RNA was determined by an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Palo Alto, CA, USA) utilizing the RNA 6000 Nano LabChip®. RNA integrity number (RIN) was computed for each sample using instrument's software [22, 23, 42, 43, 53-57].

**Preparation of ss-cDNA for molecular analysis**

The RT2 miRNA First Strand Kit® from SABiosciences Corporation (Frederick, MD, USA) was employed for making a copy of ss-DNA in a 10.0 µl reverse transcription (RT) reaction, for each RNA samples in a sterile PCR tube, containing 100 ng total RNA, 1.0 µl miRNA RT primer & ERC mix, 2.0 /µl 5X miRNA RT buffer, 1.0 µl miRNA RT enzyme mix, 1.0 µl nucleotide mix and Rnase-free H2O to a final volume of 10.0 µl. The same amount of total RNA was used for each sample. Contents were gently mixed with a pipettor, followed by brief centrifugation. All tubes were then incubated for 2 hours at 37°C, followed by heating at 95°C for 5 minutes to degrade the RNA and inactivate the RT. All tubes were chilled on ice for 5 minutes, and 90 µl of Rnase-free H2O was added to each tube. Finished miRNA First Strand cDNA synthesis reactions were then stored overnight at -20°C [22, 23, 42, 43, 53-57].

**Experimental Digital Absolute Quantitative PCR Approach**

Because the use of 96- or 384-well plates for a single sample is neither practical or affordable, nor very accurate, widespread implementation of dPCR technology has necessitated the introduction of nanofluidic techniques and/or emulsion chemistry. Three enhancements associated with dedicated instruments have helped promote the use of dPCR: (a) Partition volumes have been lowered to as little as 5 picoliter (pl); (b) The partitioning process has been automated, and (c) The number of partitions has been increased to over 100,000 for a single experiment. These innovative elements have simplified dPCR, and increased its precision, while holding down the total reaction value of a single experiment, compared to that of a conventional qPCR [48-51].

Digital PCR is a new approach to miRNAs quantification that offers alternate method to qPCR for absolute quantification, by partitioning a sample of DNA or cDNA into many individual, parallel PCR reactions; some of these reactions contain the target molecule (positive), while others do not (negative). A single molecule can be amplified a million-fold or more. During amplification, TaqMan chemistry with dye-labeled probes is used to detect sequence-specific targets. When no target sequence is present, no signal accumulates. Following PCR analysis, the fraction of negative reactions is used to generate an absolute
count of the number of target molecules in the sample, without the need for standards or endogenous controls. In conventional qPCR, the signal from wild-type sequences dominates and obscures the signal from rare sequences [55-57]. By minimizing the effect of competition between targets, dPCR overcomes the difficulties inherent to amplifying rare sequences and allows for sensitive & precise absolute quantification of the selected miRNAs.

Applied Biosystem QuantStudio™ 3D instrument used in this research study only performs the imaging and primary analysis of the digital chips. The chips themselves must be cycled offline on a Dual Flat Block GeneAmp® 9700 PCR System or the ProFlex™ 2x Flat PCR System. The QuantStudio™ 3D Digital PCR System can read the digital chip in less than 1 minute, following thermal cycling [48]. It allows for one sample per chip; although, duplexing allows for analysis of two targets per chip. Sample prep for digital PCR is no different than for real-time PCR, when using the QuantStudio™ 3D Digital PCR System. To figure out the concentration of cDNA stock from results, if one includes all of the necessary dilution factors into the Analysis Suite™ software, the software will give the copies/µL in the stock.

There are 2 dilutions that one needs to take into account: (a) The first is the dilution of the sample in the reaction, and (b) The second is the dilution of the stock that one makes before adding it to the digital PCR reaction. For example, if one wants to add 1 µL of a sample that has been diluted 1:10 from the stock. Thus, if one adds 1 µL of his/her sample to a 16 µL (final volume) reaction, the dilution factor of the sample is 1:16 or 1/16 = 0.0625. Since the stock has also been diluted 1:10 (0.1), one also need to factor this in. The final dilution factor to enter into the software is 0.0625 x 0.1 = 0.00625 (1:160). One can use either annotation to indicate the dilution factor in the Analysis Suite™ software. If one enters that value into the “Dilution” column, the software will give the copies/µL in the starting material (stock). The Poisson Plus algorithm for projects that contain QuantStudio™ 3D Chips with target, quantities >2000 copies/µL. The Poisson Plus algorithm corrects for well-to-well load volume variation, on a per Chip basis. This becomes important at higher target concentrations. There is also an option to export the Chip data as XML on the Export tab—thousands of discrete subunits prior to amplification by PCR, each ideally containing either zero or one (or at most, a few) template molecules [50].

Each partition behaves as an individual PCR reaction – as with real-time PCR—fluorescent FAM probes [or others, as VIC fluorescence]. Samples containing amplified products are considered positive (1, fluorescent), and those without product – with little or no fluorescence (i.e., are negative, 0). The ratio of positives to negatives in each sample is the basis of amplification. Unlike real-time qPCR, dPCR does not rely on the number of amplification cycles to determine the initial amount of template nucleic acid in each sample, but it relies on Poisson Statistics to determine the absolute template quantity. The unique sample partitioning step of dPCR, coupled with Poisson Statistics allows for higher precision than both traditional and qPCR methods; thereby allowing for analysis of rare miRNA targets quantitatively and accurately [50,51].

The use of a nanofluidic chip, shown below, provides a convenient and straightforward mechanism to run thousands of PCR reactions in parallel. Each well is loaded with a mixture of sample, master mix, and Applied Biosystems TaqMan Assay reagents, and individually analyzed to detect the presence (positive) or absence (negative) of an endpoint signal. To account for wells that may have received more than one molecule of the target sequence, a correction factor is applied using the Poisson model. It features a filter set that is optimized for the FAM™, VIC®, and ROX™ dyes, available from Life Technologies [49].
A workflow of the dPCR procedure by the QuantStudioTM 3D Digital PCR System Chip is presented in Figure 3, below. Digital PCR, however, has several tips to follow: 1) A rough estimate of the concentration of miRNAs of interest has to be first carried out, in order to make appropriate dilutions, so that not too many partitions will get multiple copies that prevent accurate calculation of the copy number of miRNAs of interest; 2) Non-template controls and a RT negative control must be set up for each miRNA, when using a “primer pool method” for retro-transcription; 3) A chip-based dPCR method requires less pipetting steps, which reduces potential PCR contamination compared to another type of dPCR marketed by Bio-Rad Laboratories, thus called “Bio-Rad's droplet digital PCR”, which requires multiple pipette transfers that potentially increase the risk of contamination [50], and 4) Quant StudioTM 3D chip has 20,000 fixed reaction wells, whereas Bio-Rad's droplet PCR relies upon the generation of droplets; a step that could be extremely variable, as reported by Miotto et al. [11,48].
Acquisition of Patients and Specimens for carrying out the Clinical Study

Our collaborating clinicians are aware of the constraints imposed by working with RNA and the need to preserve it so it does not ever fragment thereafter [22,23,41-43]. Participating clinics will consent prospective individuals when they report to the clinic for consult. Those individuals not showing any polyps, or inflammatory bowel diseases, such as colitis or diverticulitis, will be asked by their physicians if they wish to participate in the study. If they agree, they/their guardian will be consented, each given a stool collection kit and detailed collection instructions. Each study subject will be collected one 10 g stool sample, in a standardized fashion, in a large 40 cc plastic jacket given to each participating, consented individual, prior to any bowel preparation. The study nurse will show and ask participants to brush both the mucinous layer, which is rich in colonocytes, and the non-mucinous parts of stool in order to have a representation of the entire colon (both right and left side colon) [1,27,54-57], to be preserved overnight at room temperature in the fixative RNALater® (Invitrogen) added at 2.5 ml per 1 g of stool, followed by calling the laboratory personnel to pick up the sample by next morning. Samples will be stored at -80°C in small aliquots until needed. International Collaborators will also give study participants these written instructions in their native languages to ensure standardization, and will explain to them what's needed to collect samples representing both right and left side colon. When ready for analysis, samples are defrosted at room temp, filtered through a nylon mesh by laboratory personnel, in order to remove the preservative, and any debris prior to extraction of total small RNA. All laboratory work will be carried out and standardized under blind conditions and, in accordance with organization’s Standard Operating Procedures (SOPs) for handling of biohazardous waste material [1,21-23,41-43,54-57].

Randomized Selection of Control Human Subjects and Case Patients

To avoid bias, and ensure that biomarker selection and outcome assessment will not influence each other, a prospective specimen collection retrospective blinded evaluation (PRoBE) design randomized selection [31] of control subjects and case patients from our consented cohort population, will be employed, without no prior knowledge of who has what diagnosis, and stool specimens collected prior to removal of the lesion on patients undergoing colonoscopy, which will form the cohort.

By the 8th months of each year, we would have a cohort of 135 subjects, who are representative of the entire cohort, to select 6 control subjects and 30 CC patients. This will undoubtedly be the least common of the three groups (normal, adenoma & cancer) by far. We will then match 1 to 1 adenoma cases to the cancer cases for age (+/- 5 years), gender, clinic and month of diagnosis. Similarly, we will then match the normal controls from among the collected specimens to the cancer and adenoma cases. If there is no match, we will liberalize the data restriction to allow +/- 2 months. Thus, we will collect a case-case-control group nested in the overall colonoscopy cohort that is collected. The absolute quantitative dPCR miRNA expression analysis will be carried out on all coded samples at once during the last three months of each study year as shown in the time line Table 3, with the investigators blind to knowledge of the patients’ diagnosis, so that no analytical bias is introduced to the study.

While we believe that the 135 stool samples collected every year are representative of the overall cohort, there may be some volunteer bias, which we will not know how it would affect the studied miRNA markers. Therefore, we will collect demographic & clinical data on both groups (those who participated & those who did not) and compare for the following factors: age, gender, race/ethnicity, reason for colonoscopy, diagnoses, so that an assessment can be made at study conclusion as to what degree selection may have affected the study results.
<table>
<thead>
<tr>
<th>Method-Aim/Months</th>
<th>Aim 1</th>
<th>Aim 2</th>
<th>Aim 3</th>
<th>Aim 4</th>
<th>Aim 5</th>
<th>Aim 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05-08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29-32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37-40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33-36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49-52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53-56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57-60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *a = Refers to potential frequency and/or level of effort needed to accomplish/complete project aim.

Table 3: Timeline for accomplishing research aims during the proposed 5 years study.

**Enrichment & exfoliation strategy of colonocytes from stool for miRNA profiling**

Approximately 1 g of thawed stool is homogenized in a Stomacher® 400 EVO Laboratory Blender (Seward, UK) at 200 rpm for 3 minutes, with 40 ml of a buffer of Hank’s solution, containing 10% fetal bovine serum (FBS) and 25 mmol/L Hepes buffer (pH 7.35). The homogenates is filtered through a nylon filter (pore size 512 µm), followed by addition of 80 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 [58,59], at a final concentration of 12 ng of Ab/mg magnetic beads (1 µg Ab/106 target cells). The mixtures are incubated for 30 min on a shaking platform at room temperature. To visualize colonocytes, a drop of the solution is spread on a glass slide, dried and stained with Diff-Quick stain (Fisher Scientific, Pittsburgh, PA), and another drop is placed on a hemocytometer, and counted under the microscope to estimate the number of colonocytes form which total small RNA will be extracted. The supernatant is removed and the pellet containing colonocytes will be stored at –80°C until small RNA extraction [22,59,60].

By the 9th month of each study year, isolation of colonocytes from stool, and comparing the Agilent electrophoretic (18S and 28S) patterns to those obtained from total RNA extracted from whole stool, and differential lysis of colonocytes by RT lysis buffer (Quagen), could be construed as a validation that the electrophoretic pattern observed in stool (18S and 28S) is truly due
to the presence of human colonocytes, and not due to stool contamination with Escherichia coli (16S and 23S) [24]. One must also take into account that some exosomal RNA [30] will be released from purified colonocytes into stool, and correction is made for that effect.

**Why use a miRNA Assay for Colon Cancer Screening**

The expression of individual genes may be altered by mutations in the DNA, or by a change in their regulation at the RNA or protein levels. Epigenetic silencing is an important mechanism that contributes to gene inactivation in CRC [21]. Analysis of promoter methylation of hypermethylated in cancer 1 (HIC1) gene in stool showed it to be highly specific (98%) for colon adenoma and carcinoma, but sensitivity was quite low (31% for adenoma & 42% for all cancer) [61], which suggested that an epigenetic marker only is not good enough for screening, but a combination of genetic and epigenetic markers would be required to reliably identify CRC at an early stage.

Working with the stable DNA has been relatively easy. A study by scientists affiliated with Exact Sciences Corp., Marlborough, MA, which markets a mutation-based DNA test, assessed a newer version of a fecal DNA test for CRC screening using a vimentin methylation marker and another mutation DY marker plus non-degraded DNA in a limited sample of 44 CRC patients and 122 normal controls. It cited a sensitivity of 88% and a specificity of 82% only for advanced cancer, but not adenoma [62,63]. Besides, DNA mutation tests are not cost-effective, as screening for multiple mutations is expensive because these demanding mutation tests are not automated and are labor intensive. In addition, mutation detection in oncogenes and suppressor genes suffers from: a) the detection of mutations in these genes in fewer than half of large adenomas and carcinomas, b) the detection of gene mutations in non-neoplastic tissues, c) mutations found only in a portion of the tumor, and d) mutations often produce changes in the expression of many other genes [63,64].

Protein-based methods are currently not suited for screening and early diagnosis either because proteins are not specific to one tumor or tissue type (e.g., CEA), their susceptibility to proteases, current lack of means to amplify proteins, no function is known for more than 75% of predicted proteins of multicellular organisms, there is not always a direct correlation between protein abundance and activity, and most importantly because detection of these markers exfoliately often signifies the presence of an advanced tumor stage. The dynamic range of protein expression in minimally-invasive body fluids (e.g., blood) is as large as 1010 [65]. Moreover, mRNA levels do not necessarily correlate with protein expressions [66]. Protein microarray studies revealed that protein expression vastly exceeds RNA levels, and only post-translationally modified proteins are involved in signal transduction pathways leading to tumorigenesis [67]. There is no well-documented protein test that has been shown in clinical trials to be a sensitive and a specific indicator of colon neoplasia, especially in early stages. More recently, a serum proteomic study employing liquid chromatography (LC)-mass spectrometry (MS) carried out in a nonbiased fashion failed to differentiate between individuals with large adenoma (£ 1 cm) and normal individuals [68]. Proteomic research is a relatively new discipline, so it will take considerable time to identify and validate proteins suitable for use as clinical markers, and resolve issues of bias and validations [65,69].

On the other hand, a transcriptomic mRNA approach has shown promise to detect adenomas and colon carcinomas with high sensitivity and specificity in preliminary studies [22], but no randomized, standardized, blinded prospective clinical study has been carried out to validate the superiority of the mRNA approach. A study indicated that a combination of a transcriptomic mRNA and miRNA expression signatures improves biomolecular classification of CRC [69]. Furthermore, not only does miRNAs regulate mRNA, but they also regulate protein expression. Two studies have shown that a single miRNA act as a
rheostat to fine tune the expression of hundreds of proteins [70,71]. Hence, for CRC screening, miRNA markers are much more comprehensive and preferable to a DNA-, epigenetic-, mRNA- or a protein-based marker [23]. An added advantage of the use of the stable, nondegradable miRNAs by PCR expression, or chip-based methods is being automatable, which makes them much more economical and more easily acceptable by laboratory personnel performing these assays [48-51].

Suitability of Stool for Developing a Highly Sensitive Diagnostic Biomarker Screen for Colon Cancer

Links between miRNAs and CRC have been reported in several studies in colon cancer cell lines, cells in culture, blood, colon tissue of CRC patients, and human stool [23,43,54-57,73-86].

Stool testing has several advantages over other colon cancer screening methods as it is truly noninvasive and requires no unpleasant cathartic preparation, formal health care visits, or time away from work or routine activities. Unlike sigmoidoscopy, it reflects the full length of the colorectum and samples can be taken in a way that represents the right and left side of the colon. It is also believed that colonocytes are released continuously and abundantly into the fecal stream [79, 80], contrary to blood that is released intermittently as in guaiac FOBT [25]; therefore, this natural enrichment phenomenon partially obviates the need to use a laboratory-enrichment technique to enrich for tumorigenic colonocytes, as for example when blood is used for testing. Furthermore, because testing can be performed on mail-in-specimens, geographic access to stool screening is essentially unimpeded. The American Cancer Society (ACS) has recognized stool-based molecular testing as a promising screening technology for CRC (www.cancer.org).

Our results and others have shown that even the presence of bacterial E. coli DNA, non-transformed RNA and other interfering substances in stool does not interfere with measuring miRNA expression [1,22, 23,51-57,77-86], when an enrichment method such as the immunological paramagnetic capture method is used [25,26], when good ss-cDNA is produced [87], and when appropriate PCR primers are employed [23,55-57,77,79,81-86], as in this study. Besides, stool colonocytes [88] contain much more miRNA (than that available in free circulation such as in plasma [53,57,84]. Considerable effort has gone into selecting a reasonable number of miRNA genes (fourteen) from among the many mature human miRNA sequences identified in a previous preliminary microarray data generation study, as a number that can be screened reliably by PCR in a subsequent quantitative dPCR study to ultimately validate smaller panel of miRNA diagnostic screening gene markers, preferably 10 or less, for routine use.

**Extraction of Total Small RNA from Stool Samples**

We have routinely carried out RNA isolation procedures (both manual and automatic) from colon tissue, blood and stool samples in our labs, manually, as well as by employing the Roche MagNA pure LC™ automated system, using Qiagen’s RNeasy Isolation Kit® from Qiagen, Valencia, CA, containing RLT buffer (a guanidinium-based solution) and other commercial RNA extraction preparations, which provide the advantage of manufacturer’s established validation and quality control standards, increasing the probability of good results [21,22,43,53-57,82-91], to extract high quality total RNA from an environment as hostile as stool; thus, shattering the myth that it is difficult to employ RNA as a screening substrate. The trick has been to stabilize total RNA shortly after obtaining fresh stool by fixing samples in a chaotropic agent [RNAlater® (Invitrogen)] and observing that RNA does not ever fragment thereafter. Fragmented RNA results in poor cDNA synthesis and ultimately in less than optimal PCR amplification.
We found total small RNA isolated from stool to be suited for dPCR analysis, without further mRNA purification [87] because:

a) purified mRNA involves additional steps, and the increased sensitivity could be balanced by possible loss of material, b) not all mRNA molecules have polyA tail, and c) the concentration of mRNA may be insufficient to allow quality assessment using the Agilent 2100 Bioanalyzer [87]. Good human stool preparations showed two sharps ribosomal 18S and 28S rRNA (28S/18S=0.33), with a small fraction of micro RNA and 5S rRNA or tRNA molecules in the Agilent capillary electrophoresis equipment fitted with a RNA 6000 Nano LabChip [22]. However, E. coli shows 16S and 23S (23S/16S=1.8) [24]. RNA will be quantified spectrophotometrically. Acquiring sufficient small mRNA to analyze from stool or isolated colonocytes is feasible, as each cell contains ~ 20 pg total RNA or 0.4 pg mRNA (equivalent to 0.36 pg ss-cDNA). Only few nanograms of that DNA are needed per PCR reaction [92].

**Reverse Transcription (RT) & Preparation of Single Stranded Copy Deoxy Ribonucleic Acid (ss-cDNA)**

An Applied Biosystem kit (the TaqMan® MicroRNA Reverse Transcription Kit) that makes high quality ss-cDNA from total small RNA, and has been employed in earlier studies, will also be used in this study. It uses 50 nM RT primers that bind to the 3’ portion of miRNA molecules, 1x RT buffer, 0.25 mM each of dNTPs, 3.33 U/µl RT in a 7.5 µl reaction for 30 min at 96°C, 2 min at 56°C, 30 sec at 98°C, 2 min at 60°C and held at 10°C, the chip is then processed, and results expressed in copies µl [48,50], as shown in the workflow in Figure 3.

**Quality control (QC), and good laboratory practices (GLPs) procedures**

Rigid QC considerations are necessary to ensure the uniformity, reproducibility and reliability of dPCR amplification technology. Compared to real-time quantitative PCR (qPCR), dPCR clearly offers more sensitive and considerably more reproducible clinical methods that could lend themselves to diagnostic, prognostic, and predictive tests. But for this to be realized, the technology will need to be further developed to reduce cost and simplify application. Concomitantly the preclinical research will need be reported with a comprehensive understanding of the associated errors [48,50].

**Statistical Methods for Validating the MicroRNA Approach**

If the difference in gene expression dPCR value in copies/µl between healthy and cancer patients and among the stages of cancer at the end of the proposed validation study is as large and informative for multiple miRNA genes as in the limited preliminary results, suggesting that classification procedures could be based on values exceeding a threshold, then sophisticated classification procedures would not be needed to distinguish between these two groups; otherwise, we will use predictive classification, as detailed below. The goal will be to assign cases to predefined classes based on information collected from the cases. In the simplest setting, the classes (i.e., tumors) are labeled cancerous and non-cancerous. Statistical analyses for predictive classification of the information collected (i.e., quantitative PCR results on miRNA genes) attempt to approximate an optimal classifier. Classification can be linear, nonlinear, or nonparametric [94,95].

The miRNA expression data will be analyzed first with parametric statistics such as Student t-test or analysis of variance (ANOVA) test if the data distribution is random, or with nonparametric Kruskall-Wallis, Mann-Whitney and Fisher exact tests if the distribution is not random [96]. If necessary, more complicated models such as multivariate analysis [97] and logistic discrimination will be employed. For the corrected index, cross-validation will be used to protect against overfitting. Efron and Tibshirani [98] suggested dividing the data into 10 equal parts and using one part to assess the model produced by the other nine; this is repeated for each of the 10 parts. Cross-validation provides a more realistic estimate of the misclassification rate.
The area under the ROC curves, [in which sensitivity is plotted as a function of \((1 – \text{specificity})\)], will be used to describe the trade-off between sensitivity and specificity [99]. We will also employ principal component analysis (PCA) method [100], which is a multivariate dimension reduction technique to simplify grouping of genes that show aberrant expression from those not showing expression, or a much-reduced expression.

In cases where several genes by themselves appear to offer distinct and clear separation between control and cancer cases in either stool or tissue samples, a PMI [101,102] may not be needed. If the miRNA gene panel (or a PMI) derived by the end of the study is better than existing screening methods, all of the data generated will be used to assess the model so over-fitting is not a concern.

Cross-validation will be used to protect against over-fitting. The level of gene expression will be displayed using parallel coordinate plots [103-105] produced by the lattice package in R (version 2.9.0, http://cran.r-project.org) [51]. Other packages such as GESS (Gene Expression Statistical System) published by NCSS (www.ncss.com) will also be employed in the study.

Each subject will have his or her medical record number as the key ID for merging various tables in the database. A database will be established using widely available software like MS-Access, which output spreadsheets that will be analyzed with R (version 2.9.0, The R Foundation for Statistical Computing, http://www.r-project.org/) and S-plus software (Insightful Corporation, Seattle, WA).

**Results**

**Accessing Test Performance Characteristics (TPC) of the MiRNA Approach**

The copies/µl values of the miRNA gene panel (or a derived microRNA index, PMI) obtained from stool/colonocyte samples of normal subjects and colon cancer patients with high sensitivity and specificity will be compared to the commonly used guaiac FOBT test and with colonoscopy results obtained from patients’ medical records in 180 subjects (30 controls & 150 CC patients) at Study end to access TPC of the microRNA approach.

False positive discovery rates (expected proportion of incorrect assignment among the accepted assignments) will be assessed in our proposed approach by statistical methods [103-105], as it could reflect on the cost effectiveness of our test. The number of optimal miRNA genes (whether 14 or less) to achieve an optimum miRNAs' gene expression panel is established by appropriate statistics, as detailed below.

**Providing Numerical Underpinning of the Method as a Function of Total RNA**

Cytological methods on purified colonocytes employing Papanicolaou and Giemsa staining, which showed a sensitivity for detecting tumor cells in smears comparable to that found in biopsy specimens (78.1% versus 83.66%), have been employed [106]. A known number of the colonocytes isolated from 1g stool (from normal and neoplastic preparations), extracting total RNA from them to determine the actual amount of total RNA per stool sample, and determining the average copies/µl value from the panel of selected miRNAs from dPCR using the QuantStudioTM 3D Digital PCR Chip instrument will ultimately give an average value per a certain amount (pg or ng) of total RNA.
Determining a Panel of MiRNA genes, or a Predictive MicroRNA Index (PMI)

If results using a nested case-control design that involves prospective collection of specimens before outcome ascertainment from the study cohort are found to provide a clear-cut miRNA expression value, similar to data from the Preliminary Study, one may not need to derive a PMI. It may, however, be necessary to do so if data evaluation dictates the need for that alternative. In this case, the results of the quantitative expression of miRNA genes used to derive the index. Wiley et al. [102,103] considered 15 genes to derive a mRNA gene expression index for lung cancer, and derived an index defined as the product of two genes divided by the expression of a third gene that was 100% successful in identifying cancerous tissue. Derived research data will be used to check the sensitivity and specificity of the index. If the sensitivity falls below 90% or the specificity falls below 95%, forthcoming data using additional miRNA genes may be used with linear or logistic discriminant analysis to refine the index. To determine the usefulness of the PMI as a screening test, the clinical sensitivity (i.e., no or small number of false negatives) and specificity (i.e., no or small numbers of false positives) of the index is used, and a cutoff for a positive or negative index established. All obtained results are classified as either a true positive, false positive, true negative, or false negative by using a two by two matrix [Table 4]. The cutoff is defined so that the specificity is at least 95%, and the corresponding sensitivity is expected to be better than current pue (i.e. ~85%). The relationship between other values of sensitivity and specificity will be described using ROC curves [99]. Once results are classified, the clinical sensitivity and specificity of the PMI are calculated using conventional calculations [107]. The positive and negative predictive value are also calculated, although the population being tested will heavily influence these calculations. To measure the clinical utility of gene expression testing as a screening test, the diagnostic sensitivity and specificity of the miRNA gene panel selected, or a derived PMI, are compared to the published sensitivity and specificity of the commonly employed diagnostic screening test, guaiac FOBT, which for over 3 decades in large adenoma averaged <12%, and in carcinoma averaged ~30%, and the specificity averaged ~95% [108-111], and to the gold screening standard colonoscopy results obtained from participants’ medical records that averaged 87% for sensitivity and 100% for specificity [112]. The limitations of FOBT are biologically inescapable and cannot be reversed by technological advances [113]. Based on our data, we will be able to screen colon cancer, particularly at the pre-malignant stage, with >90% sensitivity and >95% specificity, employing £10 miRNA genes in a functional assay, which is better than any available noninvasive test. Thus, a large number of patients will be spared the discomfort, risk and expense of screening colonoscopy. Only those patients truly at risk of having a colon cancer will need to undergo colonoscopy.

<table>
<thead>
<tr>
<th>Cancer Cases</th>
<th>True Positive (TP)</th>
<th>False Negative (FN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Subjects</td>
<td>False Positive (FP)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

Note: % Sensitivity = TP/TP+FN × 100; % Specificity = TN/ FP + TN × 100

Table 4: Predictive MicroRNA Index (PMI).

Bioinformatic methods to correlate seed miRNA data with mRNA data

To provide information about complex regulatory elements, it is important to correlate miRNA resulting from this study with our mRNA data, which we produced in our earlier published research (22), as well as those data available in the open literature using several computer models [e.g., Target Scan [114], DIANA-micro [115], miRanda [116], PicTar [117], EMBL [118], EIMMo [119], mieWIP [120] or PITA Top [121], each algorithm having its advantages and disadvantages. The authenticity of functional miRNA/mRNA target pair, once identified, will be validated by fulfilling four basic criteria: a) miRNA/mRNA target interaction can be verified, b) the predicted miRNA and mRNA target genes are co-expressed, c) a given miRNA must have a predictable effect on target protein expression [i.e., if a gene is a true target of a given miRNA, its miRNA mimic will
decrease the target gene expression level while a miRNA antisense ss-oligonucleotides (ASO) inhibitor will increase the target gene expression level [122], and d) miRNA-mediated regulation of target gene expression should equate to altered biological function [123]. To examine the significance of the gene-term enrichment, a modified Fisher exact test [EASE score] is used to calculate the p-value & Bonferroni criterion employed to correct for multiple hypothesis testing (threshold 0.05), having the human genome as background. MiRNAs are annotated based on their targets identified via miRDB [124]. A thermodynamic biomarker discovery approach is to apply Shannon’s mathematical theory of communication encompassing normalized Shannon entropy [125] & Jensen-Shannon divergence [126] to trace the transcriptional changes in CC as the disease progress. Information theory measures allow the identification of biomarkers for progressive and relatively sudden transcriptional changes leading to malignant phenotypes on omics-generated data [127].

**Recommended Alternate Methods for Achieving Study Aim**

We have proposed the most practical, least labor-intensive and the most economical approach to accomplish study aims. However, in a few samples (< 5%) in control, pre- or malignant cases, it may be necessary to use methods other than automatic RNA extraction, or dPCR for sample analysis. However, because the error rate is so small and would occur in control and cases, adopting different extraction/analysis methods will not bias results.

**Manual extraction of total RNA from problematic samples using the AGPC method**

In very few samples, inhibitors present in stool may make it difficult to isolate RNA automatically using Qiagen kits that provide the advantage of manufacturer's validation and QC standards, increasing the probability that good results, may not be obtainable. In such cases we will manually isolate RNA by a modification of the classical acid guanidinium thiocyanate-phenol-chloroform (AGPC) extraction method [128] using the chaotropic agent guanidinium thiocyanate (GSC) that inactivates ribonucleases and most microorganisms. Only total small RNA samples with an OD 260/280nm ratio from 1.9 to 2.0, and RNA integrity (RIN) on Agilent 2100 BioAnalyzer of > 7.0, were used in this study.

**Use of RT2 focused PCR arrays to study miRNA expression in conjunction with qPCR System**

Qiagen introduced a focused human PCR array in a 96 well plate containing 88 cancer-related miRNA genes, 4 normalization housekeeping synthetic miRNA genes, 2 RT controls and 2 controls to test the efficiency of the dPCR reaction. These focused arrays could be used to study miRNA expression by a universal multiplex qPCR assay using Roche 480 Light Cycler PCR instrument, in which a single cDNA preparation can quantitatively assay 88 miRNA genes with high specificity due to the use of universal primers containing a modified oligonucleotide [129].

**Use of a plate assay to study miRNA expression**

Signosis, Inc., Sunnyvale, CA (www.signosisinc.com) uses high throughput plate assay to monitor individual miRNAs, without the need to carry out a RT reaction. In that assay one of the bridges oligos is partially hybridized with the miRNA molecule and the capture oligo, and another bridge forms a hybrid between the miRNA molecule and the detection oligo. The product is
immobilized onto plate through hybridization with an immobilized oligo and is detected by a streptavidin-horse radish peroxidase (HRP) conjugate and chemiluminescent substrate, using a plate reader. This hybrid structure is sensitive to the sequence of the miRNA molecule. One oligonucleotide difference will prevent the formation of the hybrid and therefore miRNA isoform could be differentiated.

**MiRNA measurements from exosomes and microvesicles extracted from stool**

MiRNAs are resistant to ribonucleases present in stool, probably by inclusion in lipid or lipoprotein complexes in either microvesicles (up to 1 µm), or in small membrane vesicles of endocytic origin known as exosomes (50-100 nm) [132]. The mechanism of release of miRNA from exosomes and microvesicles is not clear, although an apoptotic delivery candidate is believed to shed from cells during apoptosis [133]. Exosomes that were released from human and murine mast cell lines were show to contain both mRNAs and miRNAs [134]. MiRNAs in microvesicles were shown to regulate cellular differentiation of blood cells and certain metabolic pathways, and to modulate immune functions [135].

MiRNA signatures of tumor-derived exosomes were shown to function as diagnostic markers in ovarian cancer, and tumor-derived miRNA profiles and profiles of exosomal miRNAs were not significantly different [30]. If necessary, exosomal miRNAs extracted from stool colonocytes by differential centrifugation, followed by filtration through 0.22 µm filters, total RNA extracted by Trizol & concentration measured at λ 280 [134].

**Use of Real-Time qPCR to study microRNA expression**

dPCR has the edge over qPCR with the respect to technical reproducibility, because the digital output derived from diluting the sample essentially counts the number of molecules, which is more reproducible than the analog Cq output given by qPCR, thus improves both quantitative and qualitative molecular measurements.

One key advantage of qPCR, however, is its readily scalability. As a result, although dPCR is potentially more sensitive than qPCR when sample volumes are matched, qPCR has the edge if sensitivity is improved by performing a larger-volume reaction [136].

**Conclusion**

**Quantitative milestones expected to be accomplished by the end of the research**

To establish a validated diagnostic miRNAs' stool test for CC, we have first carried out a microarray miRNA experiment on immunocaptured and enriched stool colonocytes of 15 subjects, including healthy controls and CC patients of the four known TNM stages of that cancer. We then selected a smaller panel of 14 expressed mature miRNAs to do further testing by absolute quantitative digital PCR on these same 15 stool samples, using a chip-based Absolute QuantStudio 3D Digital PCR, coextraction protocols with E. coli K1 strain RS18, as well as comparing Agilent electrophoretic patterns, and sequenced random samples throughout this research using mRNA/miRNA sequencing, to be absolutely sure that we have used human and not bacterial small total RNA, which are also very abundant in human stool. Our data showed that the quantitative changes in the expression of a few mature miRNA genes in stool, associated with right and left colon cancer, gives a more sensitive and specific diagnostic screening markers than any test markers currently available on the market (FOBT, and mutation tests), give better compliance, and is more economical than the current gold standard (colonoscopy exam). TPC showed that the test has a high numerical predictive value in CC, and that the test numerically differentiates between control subjects and CC at the early cancer stages. We are planning to extend our initial research results to a larger prospective and randomized five-years nested
case-control epidemiologic study, to validate the expression of the above 14 miRNAs, in stool of 180 individuals in an epidemiologically designed study, using all stages of CC. We could show that fewer miRNAs, or even one miRNA, may suffice to be an efficient and a quantitative marker for the non-invasive diagnostic screening of CC, particularly for the non-invasive early detection of stage 0-1 in human stool, which is very difficult to achieve by current screening methods. The proposed absolute quantitative approach combined with bioinformatics’ analysis also allows for a thorough mechanistic understanding of how miRNA genes regulate mRNA expression. Test standardization will establish test’s performance characteristics ensuring that the assay will perform the same way in any laboratory, by any trained personnel, anywhere in the World for the early diagnostic screening of that curable cancer.

The following three milestones are expected to be achieved by the end of proposed research to judge success:

**Milestone 1:** Derive a workable miRNA gene panel, or a PMI in stool indicative of premalignant & malignant conditions using total small RNA extracted from stool of 150 CC patients and 30 control subjects. This milestone is achieved, if ≥ 114 (95%) of the patients with cancer have a miRNA panel that gives numerical pre- and malignant copies/µl values in stool by QuantStudioTM 3D Digital PCR System.

**Milestone 2:** Access TPC & Provide numerical underpinning of the method as a function of total RNA

Test performance characteristics (TPC) of the miRNA approach are determined by comparing copies/ul values of the miRNA gene obtained from stool samples of normal subjects and colon cancer patients with guaiac FOBT test and with colonoscopy results obtained from patients’ medical records on the 150 subjects. A numerical underpinning of the method are determined by calculating the amount of total small RNA in 1 g of stool, and determining the average copy/ul value for the miRNA gene per a known amount (pg or ng) of total RNA.

**Milestone 3:** Establish the clinical sensitivity and specificity of the miRNA gene panel, or a PMI, using total small RNA extracted from stool of 180 subjects (30 controls and 150 with pre- and malignant CCs)

This milestone is carried out as follows:

1. Guaiac FOBT [Hemoccult II Sensa, Beckman Coulter, Fullerton, CA] standardized at research facility is performed in parallel with the miRNA panel for each stool sample obtained from the 30 normal & 150 colon cancer.
2. Colonoscopy results, which are considered as the “Gold Standard” for CRC screening, are reviewed by Gastroenterologists, as well as blindly checking histopathologic results of biopsies/surgical specimens and final patients’ diagnosis, including those carried out on polyp biopsies, if removed, as obtained from patients’ medical records.
3. Using the copies/ul results from the panel of genes selected (or a PMI) obtained from stool samples of normal, and from stool samples of cancer patients, a 2 × 2 tables (see Predictive MiRNA Index, Table 4) is constructed to determine the clinical sensitivity and specificity of the microRA assay from miRNAs stool specimens' results.
4. The calculated sensitivity/specificity of the miRNA assay is compared to the FOBT assay in all the 180 subjects assessed in the same laboratory by the same investigators, as well as colonoscopy results obtained from patients’ medical records, to establish TPCs. If the results are at least as specific as the FOBT (95%) and the sensitivity ≥95%, which exceeds colonoscopy, then this milestone will have been successfully achieved.
Acknowledgment
We express our thanks to participating volunteers who provided stool samples for study. We thank Dr. Clark D. Jeffries at Renaissance Computing Institute, University of North Carolina at Chapel Hill for his help and insight on Bioinformatics' analysis. All authors declare no conflict of interest during carrying out of this research, or during data analysis. Funds for this research were provided by GEM Tox Labs operating funds.

Conflict of interest
Authors certify that there has been no conflict of interest in carrying out this research, interpretation of data, or in preparing this manuscript for publication.

References


