

# Nanocytology of rectal colonocytes to assess risk of colon cancer based on field cancerization

Dhwanil Damania<sup>1</sup>, Hemant K. Roy<sup>2</sup>, Hariharan Subramanian<sup>1</sup>, David S. Weinberg<sup>3</sup>, Douglas K. Rex<sup>4</sup>, Michael J. Goldberg<sup>2</sup>, Joseph Muldoon<sup>2</sup>, Lusik Cherkezyan<sup>1</sup>, Yuanjia Zhu<sup>1</sup>, Laura K. Bianchi<sup>2</sup>, Dhiren Shah<sup>2</sup>, Prabhakar Pradhan<sup>1</sup>, Monica Borkar<sup>2</sup>, Henry Lynch<sup>5</sup>, Vadim Backman<sup>1</sup>

<sup>1</sup>*Biomedical Engineering Department, Northwestern University, Evanston, IL 60208, USA,*

<sup>2</sup>*Department of Internal Medicine, NorthShore University HealthSystem, Evanston, IL 60208, USA*

<sup>3</sup>*Fox Chase Cancer Center, Philadelphia, PA 19111-2497, USA*

<sup>4</sup>*Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Bloomington, IN 47408, USA*

<sup>5</sup>*Creighton University Medical Center-Saint Joseph Hospital, Omaha, NE 68131, USA*

## **Running title: Colonoscopic prescreening using rectal Nanocytology**

Disclosures: Drs. Roy, Subramanian, Goldberg and Backman are co-founders and/or shareholders in Nanocytomics LLC

**Keywords:** nanocytology, nano-architecture; sub-diffractive sensitivity; field-carcinogenesis, colorectal cancer

## **Corresponding Author:**

Hemant K. Roy MD  
Duckworth Professor of Cancer Research  
NorthShore University HealthSystem  
Clinical Associate Professor  
University of Chicago Pritzker School of Medicine  
[h-roy@northwestern.edu](mailto:h-roy@northwestern.edu) Phone: 847-570-3115

## Abstract

Developing a minimally invasive and cost effective pre-screening strategy for colon cancer is critical, because of the impossibility of performing colonoscopy on the entire at-risk population. The concept of field carcinogenesis, in which normal-appearing tissue away from a tumor has molecular and, consequently, nano-architectural abnormalities, offers one attractive approach to identify high-risk patients. In this study, we investigated whether the novel imaging technique partial-wave spectroscopic (PWS) microscopy could risk-stratify patients harboring precancerous lesions of the colon, using an optically measured biomarker ( $L_d$ ) obtained from microscopically normal but nanoscopically altered cells. Rectal epithelial cells were examined from 146 patients, including 72 control patients, 14 patients with diminutive adenomas, 20 patients with non-advanced-non-diminutive adenomas, 15 patients with advanced adenomas/high-grade dysplasia, 12 patients with genetic mutation leading to Lynch syndrome, and 13 cancer patients. We found that the  $L_d$  obtained from rectal colonocytes was well-correlated with colon tumorigenicity in our patient cohort and in an independent validation set of 39 additional patients. Therefore, our findings suggest that PWS-measured  $L_d$  is an accurate marker of field carcinogenesis. This approach provides a potential pre-screening strategy for risk stratification before colonoscopy.

## Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths both in the United States and worldwide with a low 5-year survival rate (~60%). There are approximately 141,210 new cases and 49,380 deaths reported in 2011 in United States (1). CRC can be cured if detected at an early stage. However, the early-stage disease is mostly asymptomatic; hence approximately two-thirds of patients with CRC are diagnosed at a more advanced stage. According to existing guidelines, every individual above the age of 50 is a candidate for colonoscopy(2). At present, colonoscopy is considered the gold standard for CRC screening because of its high sensitivity (~97%) to advanced neoplastic lesions and ability to reduce CRC incidence by 65% - 90% by removal of precursor lesions (2). In spite of the unequivocal benefits of colonoscopic examination, only about one-fourth of the eligible screening population undergoes endoscopic CRC screening (3). The reasons for the low compliance rate include the discomfort of the endoscopic procedure, expense, and risk of complications. Even if compliance can be improved, it would be implausible to perform colonoscopy on the entire at-risk population ( $\geq 100$  million Americans) given resource constraints. Moreover, it appears that only 20-30% of patients harbor neoplasia and only ~5% of them are screen-relevant, thus resulting in a majority of colonoscopies being retrospectively unnecessary (4). Therefore, developing a “pre-screen” to colonoscopy is critical. However, currently available screening techniques such as flexible recto-sigmoidoscopy, fecal-occult-blood test (FOBT), fecal DNA analysis, etc. are either sub-optimal in their sensitivity to significant neoplasia or suffer from a high false positive or false negative rate (5-7). For example, FOBT has a poor sensitivity (~ 11%) to detect proximal neoplasia (5), while fecal-DNA lacks the sensitivity to advanced adenoma (~ 27%) (7). Similarly, flexible sigmoidoscopy lacks the ability to detect proximal lesions(8), resulting in false negatives. Most

importantly, the inconvenience of bowel preparation can strongly dissuade patients from undergoing alternative techniques such as CT colonography (9). Thus, developing a simple, minimally-intrusive, sufficiently sensitive and cost-effective pre-colonoscopy risk-stratification technique would be of paramount importance.

One emerging modality of cancer risk stratification is via identification of field carcinogenesis. This represents the impact of the “field of injury” concept - that the genetic and environmental risk factors confer a fertile mutational field throughout the organ, and the focal neoplastic lesion results from a stochastic mutational event (e.g. truncation of the adenomatous polyposis coli tumor suppressor gene). This concept is well established in a variety of malignancies (10) such as the diffuse aero-digestive injury associated with smoking-induced lung cancer (11). In the colon, this “condemned” mucosa hypothesis is the rationale for colonoscopic post-polypectomy surveillance (if a patient has had one adenomatous polyp, they are at higher risk of developing others elsewhere in the colon and thus should undergo more frequent colonoscopy). Aside from the adenomatous polyp, there have been a number of putative biomarkers that occur earlier in the pre-dysplastic (i.e., histologically normal) mucosa. These include altered epithelial cell proliferation (12), cell apoptosis (13), gene expression (14), rate of methylation (15), and biochemical (e.g., altered protein kinase C activity) (16) etc. While these all correlate with proximal neoplasia, more accurate markers are needed for clinical implementation.

Our group has been interested in developing nanocytology as a modality of detecting the ultrastructural consequences of the genetic/epigenetic alterations in field carcinogenesis. During field carcinogenesis, the mucosa is normal under light microscopy but this evaluation is limited to structures >500 nm due to the diffraction limitation of light. Thus, conventional light

microscopy is insensitive to structures such as ribosomes, macromolecular complexes, higher order chromatin structure, etc. that are in the order of tens to a few hundred nanometers. In order to probe these nano-scale structures, we have developed a novel optical technology, partial wave spectroscopic microscopy (PWS). PWS is sensitive to structures greater than 10-20 nm through analysis of multiple interferences of light reflected from intracellular spatial variations in refractive index and, in principle, is sensitive to essentially any length scale of these variations (limited by signal-to-noise). Because refractive index is a linear function of local macromolecular mass-density (DNA, RNA, proteins, etc), PWS readout is an image of a cell showing the intracellular distribution of a parameter called disorder strength ( $L_d$ ), which quantifies spatial fluctuations in macromolecular density.  $L_d$  is defined as:  $L_d = \delta n^\alpha \times l_c^\beta$  where  $\delta n$  is the standard deviation of the refractive-index (and thus mass-density) variations, and  $l_c$  is the correlation length of these variations. The coefficient  $\alpha$  depends on the cytology sample preparation ( $\alpha=1$  in our case) while  $\beta$  depends on the configuration of the optical set up and is  $\sim 1$  for the instrumentation used in this study.  $L_d$  is a measure of the spatial variations of macromolecular density and increases with macromolecular condensation (17). The exact nature of the compaction depends on the intracellular location where  $L_d$  is increased. For example, if  $L_d$  is increased at a particular location in the nucleus, this corresponds to chromatin condensation at that specific location. We have previously reported that  $L_d$  is exquisitely sensitive to subtle genetic/epigenetic perturbations in colon carcinogenesis using colon cancer cell lines and animal models (18). Specifically, microscopically identical but genetically altered (via partial knockdown of the proto-oncogenes or tumor suppressor genes) colorectal cancer cells had aggressiveness parallel with  $L_d$  (18). In animal models of intestinal neoplasia,  $L_d$  was elevated at

an early stage (18). Indeed,  $L_d$  increase appears to be a hallmark of field carcinogenesis in colon, lung and pancreatic cancer (19-21).

Herein we investigate the clinical potential of PWS interrogation of rectal colonocytes to detect colon field carcinogenesis and hence serve as a minimally intrusive screening technique for colorectal malignancies. We report the study performed on 146 patients including normal controls, patients harboring adenomatous polyps in their colon, patients having specific genetic mutations leading to Lynch syndrome and cancer patients. Our results show a gradient increase in the nanoarchitectural biomarker, disorder strength ( $L_d$ ), from control to patients having advanced adenoma to cancer patients. The  $L_d$  increase parallels the patient's risk of developing CRC with respect to different pre-malignant stages. We also studied the effect of several demographic factors on  $L_d$ . Furthermore, using rectal  $L_d$  as a single marker, we developed the prediction rule in a training set containing controls and advanced adenoma patients and evaluated its performance on an independent validation set.

## **Materials and Methods**

### **Clinical Sample Preparation**

All studies were performed and samples were collected with the approval of the Institutional Review Board at Northshore University HealthSystem, Fox Chase cancer center, and Indiana University Medical Center. Patients undergoing screening or surveillance colonoscopy were included in the study. The exclusion criteria included incomplete colonoscopy (failure to visualize cecum), poor colonic preparation, coagulopathy, prior history of pelvic radiation, or

systemic chemotherapy. The samples were collected as follows: colonoscopy to cecum was done with standard techniques using Olympus 160 or 180 series or Fujinon colonoscopes. Upon insertion of the colonoscope into the rectum, a cytology brush was passed through the endoscope and gently applied to the visually normal rectum. The brush was then smeared onto a sterile glass slide. The slide was then fixed using 95% Ethanol. Although the cytology slide contained different types of cells including epithelial and inflammatory cells, red blood cells etc., we note that all the measurements reported here were taken from epithelial cells (i.e., colonocytes). This was made possible by staining each patient slide with standardized Hematoxylin and Cytostain staining protocol and directly visualizing the cells before taking the PWS measurements. All the measurements were taken by an operator blinded to the diagnosis and all the colonocytes were selected randomly from the regions not hindered by mucus or cell-debris.

### **PWS system**

The detailed explanation and schematic of the PWS instrument used in this study is reported in reference (22) and in the supplementary information. For a given specimen, after normalizing each pixel by the corresponding incident light profile, a 3-D data-cube  $R(\lambda; x, y)$  is generated, the fluctuating part of the reflection coefficient where  $(x, y)$  refers to a specific pixel in the object plane and  $\lambda$  is the wavelength. The spectral fluctuations in wavelength range 550 - 700 nm are further analyzed by means of 1D mesoscopic light transport theory to obtain  $L_d$ . Thus, a map of disorder strength  $L_d(x, y)$  is obtained from each pixel  $(x, y)$ . Using this 2-D map,  $L_d(x, y)$  for each cell, the mean intracellular disorder strength  $L_d^{(c)}$  (the average over  $x$  and  $y$  pixels), is

obtained. The average of  $L_d^{(c)}$  for a group of cells ( $\sim 50 - 70$  cells for each patient) is calculated and defined as the mean disorder strength per patient,  $L_d^{(p)}$ . In the end, total average is calculated over all the patients of a specific group and that is termed as the group mean of the disorder strength,  $L_d^{(g)}$  while its standard deviation is defined as  $\sigma^{(g)}$ . This average disorder strength  $L_d^{(g)}$  and the standard error calculated from its standard deviation  $\sigma^{(g)}$  are depicted in all the bar-plots in this report.

## Statistical Methods

First, we performed the power analysis in order to determine the sample size for each high-risk group. For example, patients with advanced adenoma (i.e. high-grade dysplasia) are clinically the most-screening relevant population. For this population of patients, we used nQuery Advisor 6.01 software (Statistical Solutions, Saugus, MA) and determined that for a sample size of  $n_1=72$  (Controls) and  $n_2 = 15$  (advanced-adenoma), we would have 80% power to detect a between-group difference quantified as an effect size (mean difference divided by common standard deviation) = 0.804, using a two sample t-test at a 0.05 significance level. According to Cohen (23), an effect size of 0.80 is regarded as a large effect size, which is not unusual to observe in well-controlled experimental studies. In addition, with the sample sizes of 72 and 15 in each of the two groups, control and cases, respectively, an area under the ROC curve of 0.80 would be estimated within  $\pm 0.15$  as the 95% confidence interval. This sample size also has >88% power to detect the difference between 0.75 and an area under the ROC curve of 0.50 in null hypothesis, at a 0.05 significance level. Similarly, we conducted sample-size analysis for the number of columnar epithelial cells that should be measured for each patient. For the



present study, we have found that approximately 40 cells is sufficient to provide a confidence interval on a patient's mean  $L_d$  that is 20% of the difference between control and non-diminutive adenoma patients and 4% for advanced adenomas.

The disorder strength ( $L_d$ ) obtained in this study had a skewed distribution (a long tail) approximating a log-normal function. Hence the data was log-transformed to convert to a normal distribution. The Shapiro-Wilk's test was used to assess the normality assumption of the log-transformed data and a p-value  $> 0.05$  provided evidence that the log-transformed data were indeed normally distributed. In order to accurately measure the performance of  $L_d$ , we calculated the statistical parameters, effect-size and p-values on the log-transformed data. All p-values were calculated using standardized Student's t-test on the total number of patients for each subtype. A two-tailed p-value (assuming unequal variances) of 0.05 or less was considered to be statistically significant in this study. The effect-size between two groups of patients was calculated on the log-transformed average disorder strength,  $L_d^{(g)}$  and its standard deviation,  $\sigma^{(g)}$ . The value of effect-size  $> 0.5$  is statistically considered significant and it is more robust parameter than the mean-difference between two populations. Mean-differences provide the %fold-increase between two populations. Moreover, effect-size has been used to take into account the slide-to-slide variability and to robustly measure statistical significance of the average disorder strength difference, i.e.,  $\Delta L_d$  for control group and higher-risk patient groups. All p-values and effect-sizes were calculated using Microsoft Excel (Microsoft Corporation, Redmond, WA). Statistical Software STATA (StataCorp LP, College Station, Texas) was used to generate ANCOVA (analysis-of-covariance) and AUROC test statistics.

## Results

For each cell, PWS microscopy generates a two-dimensional image of  $L_d(x, y)$  ( $L_d$  as a function of location within the cells). Figure 1 (a) and (b) show representative microscope images of stained rectal colonocytes obtained from a control and a cancer patient, respectively. These images appear microscopically indistinguishable, suggesting no obvious alterations at microscopic length scales ( $>300$  nm). However, when the pseudocolor maps of spatial distribution of rectal  $L_d$  are plotted, there appear regions with higher  $L_d$  (represented by red-color) indicative of nanoscale perturbations [Fig.1: (c),(d),(e) versus (f),(g),(h)] for the colonocytes obtained from the cancer patient compared to control, indicating the nanoscale sensitivity of PWS. Furthermore, the augmentation of rectal  $L_d$  seems to be throughout the cell.

We first investigated whether the  $L_d$  was sensitive to field-effect in histologically normal appearing rectal colonocytes obtained from various patient sub-types. In this study, there were  $N = 146$  patients including controls ( $n = 72$ ), patients harboring diminutive adenoma (polyp size  $< 5$  mm,  $n = 14$ ), non-diminutive-non-advanced adenoma (5 – 9 mm polyps,  $n = 20$ ), advanced adenoma (polyp size  $\geq 10$  mm, high-grade dysplasia or  $>25\%$  villous features,  $n = 15$ ); patients harboring germline mutations for Lynch syndrome ( $n = 12$ ) but without concurrent neoplasia and those having adenocarcinoma in their colon ( $n = 13$ ). Figure 2 and Table 1 show the overall results of rectal  $L_d$  obtained from the reported patient population. As indicated in Figure 2(a), there appears to be a progressive increase in  $L_d$  that correlates with the risk of developing CRC: no neoplasia patients  $<$  non-advanced adenomas (most of which spontaneously regress)  $<$  advanced adenomas (a more aggressive precancerous lesion; CRC progression risk of 2-5% per year (24))  $<$  HNPCC patients (lifetime risk of CRC of 60-80% (25))  $<$  patients with frank CRC.

Furthermore, Figure 2(a) demonstrates that there is no significant difference in the  $\Delta L_d$  (Effect-size = 0.08, %Difference = 12.44%,  $P$ -value  $\sim$  0.68) between control patients and those with diminutive adenoma. This result is consistent with the reported low risk of transformation (<0.1%) of diminutive adenoma to carcinoma(26). However, there was a statistically significant  $\Delta L_d$  (Effect-size = 0.64, %Difference = 42.44%,  $P$ -value  $\sim$  0.0001) between control patients and those harboring intermediate size adenomas (5 – 9 mm polyps).  $\Delta L_d$  increased further (Effect-size = 1.02, %Difference = 113.24%,  $P$ -value  $\sim$  0.000006) for patients with advanced adenomas. These results indicate that  $L_d$  is both significantly different and progressively increasing with higher polyp-size and hence with higher risk of developing CRC (26). Importantly, they imply that PWS is sensitive to the colon field carcinogenesis (14, 20, 27) and can quantify alterations in histologically “normal” appearing rectal colonocytes irrespective of the actual polyp location.

We further assessed the PWS performance in patients with inherited genetic mutations which lead to familial colorectal cancer (CRC), as it accounts for approximately 15-20% of all CRCs (28). Specifically, we investigated hereditary non-polyposis colorectal cancer (HNPCC, i.e., Lynch syndrome) cases which account for approximately 2-3% of all CRC patients (28). This disease is mainly caused by germline mutations in the DNA mismatch repair genes such as MLH1, MSH2, MSH6, etc.(28). PWS results from normal-appearing colonocytes demonstrated a two-fold increase and highly statistically significant  $\Delta L_d$  (Effect-size = 1.17, % Difference = 184.5%,  $P$ -value  $\sim$  0.000015) in these patients as depicted in Figure 2(a). This sharp increase in  $L_d$  parallels the reported elevated life-time risk ( $\sim$  70%) of developing CRC in these patients (28-29). Moreover,  $\Delta L_d$  is the highest (Effect-size = 1.42, % Difference = 281%,  $P$ -value  $\sim$  0.0000006) between the control patients and cancer patients as shown in Table 1. This is expected as the cells obtained from cancer patients would have undergone the most nano-

architectural alterations of all patient categories. Overall, Figure 2 and Table 1 indicate that the disorder strength increase parallels the risk of developing CRC, from control patients to those with neoplastic lesions to patients with a proven history of colon cancer. This increasing trend suggests that rectal  $L_d$  is correlated with the tumorigenicity of colon carcinogenesis and highlights the potential of PWS to risk-stratify patients for CRC based on “field-effect”, underscoring it as a promising pre-screening technique for colon cancer.

We next evaluated the diagnostic performance of PWS. In order to gauge the diagnostic capability of the PWS technique, we calculated the performance characteristics using the single parameter  $L_d$ . It is important to note that we used a single marker to avoid any over-fitting of the dataset presented here. Figure 3 highlights the estimate of the AUROC using  $L_d$  as the only diagnostic marker. The AUROC for PWS analysis of rectal brushings was 0.85 for advanced adenoma/HGD, and it improves further for higher risk patient populations. The AUROC was 0.89 for HNPCC and 0.92 for cancer patients. This result highlights the feasibility of the PWS technique for colon cancer screening and the satisfactory power of the single biomarker rectal  $L_d$ . The values of the sensitivity and specificity depicted in Figure 3 are better (e.g., the sensitivity of FOBT, fecal-DNA is ~ 26% for HGD) than those of other existing screening techniques that are currently (albeit poorly) used as a pre-screen for colonoscopy (6-7). In the future, these performance characteristics could be further improved by including other independent PWS-derived biomarkers.

One of the challenges with CRC screening is the early detection of proximal colonic neoplasia. There is a growing debate about the sensitivity of colonoscopy to proximal lesions (30). Hence, we tested the sensitivity of PWS to proximal adenomas. Figure 4 provides information about the location of the polyp and its effect on  $L_d$ . It appears from Figure 4 that

there is a statistically non-significant difference in  $\Delta L_d$  (Effect-size = 0.30, *P*-value ~ 0.15) between patients having proximal and distal lesions. There were 19 patients having proximal polyps of size > 5 mm and 16 patients with similar distal lesions. These results suggest that PWS is equally sensitive to both proximal and distal lesions. This is an important result with respect to sub-optimal efficacy of colonoscopy (30) and flexible sigmoidoscopy (8) to proximal lesions. We believe that PWS can be a handy tool for clinicians in successfully diagnosing proximal lesions and can potentially reduce the polyp miss rate. Additionally, pre-screening using PWS could filter patient populations that might benefit from further invasive colonoscopic investigation.

For the control group ( $n = 72$ ) in our study, we included patients harboring hyperplastic polyps ( $n = 7$ ) and patients having diverticulitis ( $n = 8$ ) in their colon. We evaluated the confounding effect of these subtypes of patients on the PWS performance. However, there seems to be no statistically significant difference (Effect-size = 0.18, *P*-value ~ 0.25) between  $L_d$  values of the patients with non-neoplastic lesions and those with clean colons. Hence, we combined all these patients in the ‘control’ group.

We next studied the role of confounding demographic risk factors such as age, gender, smoking and drinking history on the sensitivity of the measured biomarker. Age has been implicated as one of the key risk factors for colonic neoplasia and there have been a variety of age-related changes in colonic mucosa (such as methylation (31)). We therefore performed ANCOVA analysis and noted no significant confounding with age ( $P = 0.54$  for  $L_d$ ). As outlined in Table 2, smoking and drinking history also did not have any confounding effect on  $L_d$  ( $P = 0.57$  for current smokers and  $P = 0.99$  for drinking). Similarly, male gender is a well-established risk factor for colonic neoplasia(32). However, ANCOVA analysis indicated that there was no

significant confounding with gender ( $P = 0.29$  for  $L_d$ ). Overall, the non-significant ANCOVA  $P$ -values suggest that  $L_d$  is not confounded by age, gender, smoking or drinking patterns.

The next question we addressed was the performance of  $L_d$  in a prospective study. Although we understand that we have a modest dataset to make any definitive conclusions, we still wanted to gauge the diagnostic power of our approach when tested on an independent training and testing set for any patient having an adenomatous polyp of size  $\geq 10$  mm or with tubulovillous features. We specifically selected advanced adenoma patients as they comprise the most clinically relevant screening population. The training set was comprised of a subset of  $n = 87$  patients (wherein 72-controls and 15-advanced adenoma) out of the  $N = 146$  total patients. However, the validation set ( $N = 39$  patients with 14-control and 25-advanced adenoma) was an independent enriched dataset for case-control recruited from a second clinical site. We developed a cutoff based on the  $L_d$  values of the training set and applied the same to the validation set. The estimates of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) based on this threshold are listed in Table 3. As summarized in Table 3, we obtained promising performance characteristics for the training set (sensitivity = 73%, specificity = 78%, PPV = 41%, NPV = 93%). The 95% confidence intervals for all these parameters were: Sensitivity = 0.733 [0.48 to 0.891]; Specificity = 0.778 [0.669 to 0.858]; PPV = 0.407 [0.245 to 0.593]; NPV = 0.933 [0.841 to 0.974]. We then applied the same  $L_d$  threshold to the validation set which yielded sensitivity = 74% and specificity = 83%, similar to the training set and further supporting the robustness of our conclusions. We again emphasize that this performance is based on a single biomarker ( $L_d$ ) and it is probable that inclusion of additional PWS-measured markers would further improve the performance characteristics.

## Discussion

Herein we show that field-carcinogenesis based PWS analysis of rectal colonocytes has promise as a novel, minimally intrusive colorectal cancer risk-stratification technique. Our results obtained from 146 patients demonstrated rectal  $L_d$  paralleled the risk of developing cancer: diminutive adenoma < non-diminutive-non-advanced adenoma < advanced adenoma. Intriguingly, the rectal  $L_d$  of HNPCC patients without concurrent neoplasia was higher than patients with advanced adenomas but lower than CRC further supporting the notion that  $L_d$  mirrored overall risk of CRC development. The diagnostic promise of this single biomarker was validated with an independent, albeit small, dataset. Rectal  $L_d$  appeared robust and not confounded by co-segregating risk factors (age, gender, smoking-history, alcohol consumptions) and was able to sense proximal and distal neoplasia equivalently.

While colonoscopy is the recommended screening option for CRC (2), it cannot be applied on the entire at-risk population due to cost, resource constraints and possible complications. Moreover, some of the current non-invasive screening options have poor sensitivity to neoplasia. For example, guaiac-based fecal occult blood (FOBT) testing has sensitivity of 10.8% (6), whereas fecalDNA has sensitivity of 27% (5, 7) to clinically significant lesions. In this situation, field cancerization provides a possible solution for developing a minimally invasive pre-screen. There are several reports supporting the biological plausibility of using rectal mucosa to detect CRC (10). Several lines of evidence suggest that there are early genetic/epigenetic and consequently morphological changes that occur in the rectal mucosa prior to development of adenoma-carcinoma (14, 33-34). There have been a few efforts to detect CRC using field-effect, e.g. flexible sigmoidoscopy to detect a sentinel distal adenoma as a marker of

advanced proximal neoplasia, however its sensitivity is low (~33% in women (30)). In contrast, the PWS nanocytology-based approach is minimally invasive, easy, quick, cost effective, more patient-compliant and sufficiently sensitive (AUROC of 0.85 for advanced adenoma patients, 0.89 for HNPCC and 0.92 for cancer) for clinical practice. Further studies using optimized instrumentation and potentially more PWS markers will likely show improved diagnostics. Hence, PWS nanocytology has the promise of translating a field carcinogenesis approach into clinically practical means of colonic risk stratification.

Recently, CT colonography (virtual colonoscopy) has been sanctioned for average-risk screening (9). In comparison, the performance of rectal PWS data had similar per-polyp rate sensitivity (~ 90%) for advanced adenomas to the reported large multicenter trials using CT colonography. However, PWS nanocytology has advantages, including a lack of need for colon purge, less discomfort and expense, and no radiation exposure. This is likely to improve patient compliance with screening which has been a major barrier to reducing the toll of CRC.

Some recent reports discuss the possibility of functional proteomics for diagnosing CRC in clinical practice (35). They describe alterations in the expression of nine proteins in precancerous and neoplastic tissues, e.g., nm23, manganese superoxide dismutase (MnSOD) etc. suggesting their role in colon tumorigenesis(16, 36). Although promising, the proteomic approaches lack reproducibility, suffer from tedious sample analysis (using mass-spectrometry) and conflicting results for different biomarkers(35). Hence, their translation into clinical practice is a big question mark(35). However, in comparison, our PWS-driven screening approach (which investigates changes at similar length-scales) provides compelling results with satisfactory sensitivity and specificity for patients with advanced histologies in the colon.



While the rectal  $L_d$  appears to be a more accurate and robust biomarker than conventional approaches, the biological underpinnings have been incompletely elucidated. Others have shown that there are subtle submicron abnormalities in both localized and more distal field carcinogenesis (34, 37-38). With regards to PWS,  $L_d$  represents changes in the local mass density of cellular building blocks (proteins, RNA, DNA, etc.) with an increase in  $L_d$  corresponding to macromolecular condensation (e.g., condensation leads to higher local mass-density and correlation length, hence higher  $L_d$ ). These changes occur both in the cytoplasm and nucleus. Our studies investigating the cytoplasmic origin has taken a candidate approach focusing on the cytoskeleton because of early proteomic data (39-40) and the observation that many of the key molecules in early colon carcinogenesis actually interact with the cytoskeleton (APC,  $\beta$ -catenin, E-cadherin, Src etc). We have recently reported that pharmacological disruption of the cytoskeleton ameliorated the pro-neoplastic (increase)  $L_d$  in stably transfected colon cancer cell lines (22). From a nuclear perspective,  $L_d$  increase indicates chromatin condensation, which was further corroborated by our group's electron microscopy demonstration that histologically normal rectal biopsies from patients harboring adenomas had altered chromatin architecture (e.g. higher heterochromatin content) and exhibited chromatin clumping, potentially driven by histone deacetylase 2 (HDAC 2) over-expression (17, 41-42). This finding is consonant with the recent interest in high order chromatin structure as a surrogate for cellular transcriptional activity

There are several limitations with our approach that need to be acknowledged. First, we report an outstanding performance of rectal PWS marker,  $L_d$  that is tested only on a modest sample size. We realize that in order to validate our approach, there is a need to conduct studies over a large dataset of the prospective population. Second, there are other rectal diseases (e.g.,

inflammatory bowel disease, ulcerative colitis etc.) which can increase the risk of CRC (43) but are not part of this study. Future studies will include this group of patients. Third, we report performance based on a single marker, rectal  $L_d$ ; however, identifying more PWS-derived markers can improve the performance characteristics.

In conclusion, we provide evidence that interrogation of the rectal epithelium with PWS can quantify nanoscale architectural alterations in colon field carcinogenesis. The current approach appears to have reasonable diagnostic accuracy (comparable to standard techniques) and will likely improve with ongoing technological refinements. If confirmed in large-scale validation trials, we envision that it would be applied as a minimally-invasive and cost-effective “prescreen” technique that can identify patients at high risk and who would likely benefit from further testing (colonoscopy). Employing PWS nanocytology with field carcinogenesis for accurate risk-stratification could be a paradigm shift in CRC screening.

## **Acknowledgments**

This research work was supported by the National Institute of Health grants R01CA128641, R01EB003682, U01 CA111257 and National Science Foundation grant CBET-0937987. We would like to thank Andrew Gomes and Hongyan Du for their valuable suggestions during statistical discussion.

**References:**

1. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin.* 2011;61:212-36.
2. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology.* 2008;134:1570-95.
3. Cooper GS, Doug Kou T. Underuse of colorectal cancer screening in a cohort of Medicare beneficiaries. *Cancer.* 2008;112:293-9.
4. Roy HK, Backman V, Goldberg MJ. Colon cancer screening: the good, the bad, and the ugly. *Arch Intern Med.* 2006;166:2177-9.
5. Morikawa T, Kato J, Yamaji Y, Wada R, Mitsushima T, Shiratori Y. A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. *Gastroenterology.* 2005;129:422-8.
6. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med.* 2004;351:2704-14.
7. Song K, Fendrick AM, Ladabaum U. Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. *Gastroenterology.* 2004;126:1270-9.
8. McCallion K, Mitchell RM, Wilson RH, Kee F, Watson RG, Collins JS, et al. Flexible sigmoidoscopy and the changing distribution of colorectal cancer: implications for screening. *Gut.* 2001;48:522-5.
9. Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, et al. Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med.* 2008;359:1207-17.

10. Kopelovich L, Henson DE, Gazdar AF, Dunn B, Srivastava S, Kelloff GJ, et al. Surrogate anatomic/functional sites for evaluating cancer risk: An extension of the field effect. *Clinical Cancer Research*. 1999;5:3899-905.
11. Bhutani M, Pathak AK, Fan Y-H, Liu DD, Lee JJ, Tang H, et al. Oral Epithelium as a Surrogate Tissue for Assessing Smoking-Induced Molecular Alterations in the Lungs. *Cancer Prevention Research*. 2008;1:39-44.
12. Anti M, Marra G, Armelao F, Percesepe A, Ficarelli R, Ricciuto GM, et al. Rectal epithelial cell proliferation patterns as predictors of adenomatous colorectal polyp recurrence. *Gut*. 1993;34:525-30.
13. Bernstein C, Bernstein H, Garewal H, Dinning P, Jabi R, Sampliner RE, et al. A bile acid-induced apoptosis assay for colon cancer risk and associated quality control studies. *Cancer Research*. 1999;59:2353-7.
14. Chen LC, Hao CY, Chiu YS, Wong P, Melnick JS, Brotman M, et al. Alteration of gene expression in normal-appearing colon mucosa of APC(min) mice and human cancer patients. *Cancer Res*. 2004;64:3694-700.
15. Ushijima T. Epigenetic field for cancerization. *Journal of Biochemistry and Molecular Biology*. 2007;40:142-50.
16. Polley AC, Mulholland F, Pin C, Williams EA, Bradburn DM, Mills SJ, et al. Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. *Cancer Res*. 2006;66:6553-62.
17. Backman V, Kim JS, Pradhan P, Szleifer I. The influence of chromosome density variations on the increase in nuclear disorder strength in carcinogenesis. *Phys Biol*. 2011;8.
18. Subramanian H, Pradhan P, Liu Y, Capoglu IR, Li X, Rogers JD, et al. Optical methodology for detecting histologically unapparent nanoscale consequences of genetic alterations in biological cells. *Proc Natl Acad Sci U S A*. 2008;105:20118-23.

19. Subramanian H, Pradhan P, Liu Y, Capoglu IR, Rogers JD, Roy HK, et al. Partial-wave microscopic spectroscopy detects subwavelength refractive index fluctuations: an application to cancer diagnosis. *Opt Lett.* 2009;34:518-20.
20. Subramanian H, Roy HK, Pradhan P, Goldberg MJ, Muldoon J, Brand RE, et al. Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy. *Cancer Res.* 2009;69:5357-63.
21. Roy HK, Subramanian H, Damanian D, Hensing TA, Rom WN, Pass HI, et al. Optical detection of buccal epithelial nanoarchitectural alterations in patients harboring lung cancer: implications for screening. *Cancer Res.* 2010;70:7748-54.
22. Damanian D, Subramanian H, Tiwari AK, Stypula Y, Kunte D, Pradhan P, et al. Role of cytoskeleton in controlling the disorder strength of cellular nanoscale architecture. *Biophys J.* 2010;99:989-96.
23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.; 1988.
24. Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. *Gut.* 2007;56:1585-9.
25. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. *N Engl J Med.* 2003;348:919-32.
26. Butterly LF, Chase MP, Pohl H, Fiarman GS. Prevalence of clinically important histology in small adenomas. *Clin Gastroenterol Hepatol.* 2006;4:343-8.
27. Kopelovich L, Henson DE, Gazdar AF, Dunn B, Srivastava S, Kelloff GJ, et al. Surrogate anatomic/functional sites for evaluating cancer risk: an extension of the field effect. *Clin Cancer Res.* 1999;5:3899-905.
28. Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet.* 1999;36:801-18.

29. Brand RM, Jones DD, Lynch HT, Brand RE, Watson P, Ashwathnayan R, et al. Risk of colon cancer in hereditary non-polyposis colorectal cancer patients as predicted by fuzzy modeling: Influence of smoking. *World J Gastroenterol.* 2006;12:4485-91.
30. Lewis JD, Ng K, Hung KE, Bilker WB, Berlin JA, Brensinger C, et al. Detection of proximal adenomatous polyps with screening sigmoidoscopy: a systematic review and meta-analysis of screening colonoscopy. *Arch Intern Med.* 2003;163:413-20.
31. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res.* 1998;58:5489-94.
32. Roy HK, Gomes AJ, Ruderman S, Bianchi LK, Goldberg MJ, Stoyneva V, et al. Optical measurement of rectal microvasculature as an adjunct to flexible sigmoidoscopy: gender-specific implications. *Cancer Prev Res (Phila).* 2010;3:844-51.
33. Kukitsu T, Takayama T, Miyanishi K, Nobuoka A, Katsuki S, Sato Y, et al. Aberrant crypt foci as precursors of the dysplasia-carcinoma sequence in patients with ulcerative colitis. *Clin Cancer Res.* 2008;14:48-54.
34. Alberts DS, Einspahr JG, Krouse RS, Prasad A, Ranger-Moore J, Hamilton P, et al. Karyometry of the colonic mucosa. *Cancer Epidemiol Biomarkers Prev.* 2007;16:2704-16.
35. de Noo ME, Tollenaar RA, Deelder AM, Bouwman LH. Current status and prospects of clinical proteomics studies on detection of colorectal cancer: hopes and fears. *World J Gastroenterol.* 2006;12:6594-601.
36. Reymond MA, Steinert R, Kahne T, Sagynaliev E, Allal AS, Lippert H. Expression and functional proteomics studies in colorectal cancer. *Pathol Res Pract.* 2004;200:119-27.
37. Weyn B, Jacob W, da Silva VD, Montironi R, Hamilton PW, Thompson D, et al. Data representation and reduction for chromatin texture in nuclei from premalignant prostatic, esophageal, and colonic lesions. *Cytometry.* 2000;41:133-8.

38. Verhest A, Kiss R, Dolne D, Larsimont D, Salmon I, Delaunoit Y, et al. Characterization of Human Colorectal Mucosa, Polyps, and Cancers by Means of Computerized Morphonuclear Image Analyses. *Cancer*. 1990;65:2047-54.
39. Bernstein C, Bernstein H, Payne CM, Dvorak K, Garewal H. Field defects in progression to gastrointestinal tract cancers. *Cancer Lett*. 2008;260:1-10.
40. Bernstein H, Payne CM, Kunke K, Crowley-Weber CL, Waltmire CN, Dvorakova K, et al. A proteomic study of resistance to deoxycholate-induced apoptosis. *Carcinogenesis*. 2004;25:681-92.
41. Pradhan P, Damania D, Joshi HM, Turzhitsky V, Subramanian H, Roy HK, et al. Quantification of nanoscale density fluctuations using electron microscopy: Light-localization properties of biological cells. *Appl Phys Lett*. 2010;97:243704.
42. Stypula Y, Crawford S, Subramanian H, Tiwari AK, Kunte D, Cornwell ML, et al. Understanding Biological Mechanisms of Nuclear Disorder Strength in Early Carcinogenesis. *Gastroenterology*. 2011;140:S765-S6.
43. Sands BE, Grabert S. Epidemiology of inflammatory bowel disease and overview of pathogenesis. *Med Health R I*. 2009;92:73-7.

### **Table Legends:**

**Table 1:** highlights the statistical performance of the PWS analysis for all these category of patients compared to the controls. The %mean-differences, effect-size and the P-values are calculated on the log normalized  $L_d$  in order to obtain a normal distribution.

**Table 2:** shows the impact of demographic factors on the single biomarker: Demographic characteristics such as age, gender, smoking and drinking history is shown for different patient groups and their effect on  $L_d$ . The P-value is calculated using the analysis of covariance

(ANCOVA) in STATA. It is evident that  $L_d$  is not confounded by age ( $P = 0.54$ ), gender ( $P = 0.20$ ), smoking ( $P = 0.57$ ) and drinking history ( $P = 0.99$ ).

**Table 3: highlights the performance of  $L_d$  on an independent training (n = 87) and testing set (n = 39) developed for controls and patients having advanced adenoma:** Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) are shown for controls and patients having advanced adenoma in their colon using the training set. Based on a single cut-off  $L_d$ -value derived from the training set, we obtained equivalent performance characteristics (i.e., sensitivity & specificity) for the enriched testing set.

Note: As the testing dataset is an enriched population, predictive values (both negative and positive) are uninformative.

### Figure Legends:

**Figure 1: Representative rectal colonocytes from control and cancer patients:** (a) and (b) are the H & E stained microscopic images of rectal epithelium from control and cancer patient respectively. (c), (d), (e) are representative PWS generated pseudocolor heat-map of  $L_d$  for the colonocytes from the circled region of control patient and similarly (f), (g), (h) are for cancer patient. Although the microscopic images of the colonocytes from the control and cancer patient are indistinguishable, the disorder strength ( $L_d$ ) was markedly increased in cancer patients compared to the control.

**Figure 2: Results obtained from N = 146 patients using PWS microscopy:** demonstrates that PWS measured disorder strength ( $L_d$ ) parallels the increasing risk of developing colorectal cancer



for patients harboring different size of precancerous lesions (non-diminutive adenoma) in their colon, having specific genetic mutation leading to Lynch syndrome (HNPCC) and frank cancers.

**Figure 3: shows the diagnostic performance of the single parameter, disorder strength ( $L_d$ ) for various risk-groups.** The performance is excellent for the advanced adenoma patients with AUROC of 0.85, and it improves with AUROC of 0.89 and 0.92 for patients with HNPCC and frank cancer respectively.

**Figure 4:** depicts that  $L_d$  from rectal colonocytes is equally sensitive to both, the patients having proximal lesions (n = 19) and distal lesions (n =16) of size > 5mm with a non-significant P-value ~ 0.15.