

Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

Early detection of colorectal cancer relapse by infrared spectroscopy in “normal” anastomosis tissue

Ahmad Salman
Gilbert Sebbag
Shmuel Argov
Shaul Mordechai
Ranjit K. Sahu

Early detection of colorectal cancer relapse by infrared spectroscopy in “normal” anastomosis tissue

Ahmad Salman,^{a,*} Gilbert Sebbag,^b Shmuel Argov,^c Shaul Mordechai,^{d,*} and Ranjit K. Sahu^e

^aShamoon College of Engineering, Department of Physics, 56 Bialik Street, Beer-Sheva 84100, Israel

^bBen Gurion University, Soroka University Medical Center, Surgery B Department, Faculty of Health Sciences, Beer-Sheva 52900, Israel

^cSoroka University Medical Center, Department of Pathology, Beer-Sheva 84105, Israel

^dBen-Gurion University of Negev, Department of Physics, Beer-Sheva 84105, Israel

^eCenter for Autoimmune and Musculoskeletal Diseases, Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, New York 11030, United States

Abstract. Colorectal cancer is one of the most aggressive cancers usually occurring in people above the age of 50 years. In the United States, colorectal cancer is the third most diagnosed cancer. The American Cancer Society has estimated 96,830 new cases of colon cancer and 40,000 new cases of rectal cancer in 2014 in the United States. According to the literature, up to 55% of colorectal cancer patients experience a recurrence within five years from the time of surgery. Relapse of colorectal cancer has a deep influence on the quality of patient life. Infrared (IR) spectroscopy has been widely used in medicine. It is a noninvasive, nondestructive technique that can detect changes in cells and tissues that are caused by different disorders, such as cancer. Abnormalities in the colonic crypts, which are not detectable using standard histopathological methods, could be determined using IR spectroscopic methods. The IR measurements were performed on formalin-fixed, paraffin-embedded colorectal tissues from eight patients (one control, four local recurrences, three distant recurrences). A total of 128 crypts were measured. Our results showed the possibility of differentiating among control, local, and distant recurrence crypts with more than a 92% success rate using spectra measured from the crypts' middle sites. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.20.7.075007]

Keywords: crypt; metastasis; local and distant recurrence; multivariate analysis; principle component analysis; linear discriminant analysis.

Paper 150059R received Feb. 1, 2015; accepted for publication Jun. 10, 2015; published online Jul. 15, 2015.

1 Introduction

Colorectal cancer is the third most common cancer diagnosed in both men and women in United States. The American Cancer Society's most recent estimates for the number of colorectal cancer cases in the United States for 2013 are: 102,480 new cases for colon cancer, 40,340 new cases for rectal cancer, and 50,830 deaths. The estimated numbers for 2014 are: 96,830 new cases of colon cancer, 40,000 new cases of rectal cancer, causing 50,310 yearly deaths.^{1,2} It is estimated that over 40,000 members of the adult population in the UK are diagnosed with colorectal cancer each year.³

Colorectal cancer, when detected in the early stages, is one of the most curable gastrointestinal cancers. Treatment is mainly surgical, during which the cancerous section of the bowel is removed. Surgery is followed by chemotherapy and radiotherapy, according to disease staging.⁴ Full recovery after five years can theoretically be achieved if the tumor is completely removed before metastases evolve.

Relapse of colorectal cancer is high and has a deep influence on quality of life.⁵

Three kinds of recurrence are diagnosed: (a) local recurrence (~10%), (b) local recurrence with distant metastasis (~10%), and (c) recurrence only by distant metastasis without local

recurrence (~80%). Because the margins of the removed cancerous lesion cannot always be determined very well macroscopically, surgeons may occasionally leave malignant cells during the surgery. The margins of the removed (diseased) colonic section may lead to 83% of the local recurrences. This shortcoming apparently arises due to the normal morphology often observed in the resection margins in standard pathology tests, although an early stage of abnormality may be present.

Colon cancer tumors that have high microvessel density tend to recur or spread.⁶ The likelihood that rectal cancer will recur within five years of treatment is 11% for an original tumor smaller than 3 cm and 28% for an original tumor larger than 3 cm.⁷ Colon and rectal cancers' recurrence also depends on the stage of the disease.⁷ Researchers have confirmed that the more advanced the rectal cancer, the more likely it was to recur. The local recurrence rate of rectal cancer is double that of other parts of the colon. Cancer recurrence rates during a five-year period are as follows: 10% for stage 1 rectal cancer, 24% for stage 2 rectal cancer, and 41% for stage 3 rectal cancer. Thus, careful assignment of the patients into subgroups with different risks of recurrence is very important and enables physicians to more precisely focus their treatment strategies.

Previous studies demonstrate that up to 55% of rectal cancer patients experience a recurrence within five years from the

*Address all correspondence to: Ahmad Salman, E-mail: ahmad@sce.ac.il; Shaul Mordechai, E-mail: shaulm@bgu.ac.il

time of surgery.^{7,8} The high variance between study results may derive from the variety of different cancer cases, original tumor size, the stage of the disease in which the surgery is performed, and the expertise of the surgical team.^{7,9}

The leading cause of high local recurrences of rectal cancer (up to 83% in some studies),^{10,11} is attributed to the fact that the surgical margins of the removed cancerous colorectal part, although macroscopically “clean,” may be microscopically involved with malignant cells. This is why surgery always takes safe margins (5 to 7 cm) away from the tumor limits when needed and feasible.

Colorectal cancer patients undergo follow-up routine monitoring of the colon for any relapse and for diagnosis of malignant polyps. This is accomplished using tests such as colonoscopy, flexible sigmoidoscopy, and CT colonoscopy and barium enema, or biochemical tests such as fecal immunochemical test or sDNA test.¹² A fecal occult blood test is used for screening the healthy population at risk. These protocols are time-consuming and sometimes uncomfortable for the patient. There is a clear need for a new methodology that can lead to early diagnosis of abnormality—one that would result in a better prognosis for patients as well as help the surgeon decide if the resection margin contain foci with recurrence and metastatic potential. Fourier transform infrared (FTIR) spectroscopy has emerged as a powerful tool for chemical analysis because of its ability to provide detailed information on the spatial distribution of chemical composition at the molecular level.¹³ In applications requiring qualitative and quantitative analyses, the potential of IR spectroscopy to identify biochemical components via analysis of their vibrational spectrum fingerprint is unsurpassed.¹³ The biochemical fingerprint of cells, tissues, and fluids that become altered in a diseased state can be detected using FTIR spectroscopy.

Applications of FTIR spectroscopy in the field of medicine, especially diagnosis, have shown encouraging trends. Various biomolecular components of the cell give rise to a characteristic spectrum that is rich in structural and functional aspects.^{13–15}

Distinctions between normal and neoplastic cells and tissues have been made by utilizing spectroscopic methods.^{16,17} FTIR microscopy (FTIR-MSP) has been shown to provide important clues to the changes in the biochemical composition of cells and tissues, especially during carcinogenesis.^{18–21} Many studies have shown that FTIR has a good potential as a diagnostic tool to evaluate different tissues from different organs, such as the colon and cervix,^{22–24} which exhibit spectral variations in the mid-IR region, congruent with the biochemical variations in the tissue. These organs have a well-defined pattern of cell growth in their epithelia, resulting in established spectral variations that can be exploited for detecting changes during disease initiation, progression and remission.

Biological changes associated with abnormal cell proliferation are reliable markers for detecting the initiation of colorectal malignancy.^{25,26} Abnormal maturation and migration of colonic crypt cells indicate premalignant or malignant stages of cancer.²⁷ Thus, the detection of abnormal crypts becomes an integral part of the diagnosis and management of colorectal cancer^{28,29} and the potential of FTIR spectroscopy adds a translational dimension to the field of IR spectroscopy in biomedicine.³⁰

Our previous studies have shown that IR spectroscopy can detect the early development of diseases or cell transformation at a stage when the morphology is still normal.^{31,32} This important observation was named by our group as the

“crypt model.”^{33,34} Our research group has reached important milestones in the overall understanding of the complex changes that occur in single cells as well as tissues on the level of biochemical composition when a cell undergoes changes between normal and abnormal conditions.^{19,20,22,33}

Our overall goal in this study is to evaluate the potential of IR optical technology as a tool for the objective detection and assessment of abnormality in the colon tissue, enabling a “correct” determination of the margins of the diseased part and thus reducing colorectal cancer recurrences.

2 Material and Methods

In this study we included 128 crypts from eight colorectal cancer patients. The biopsies used in this study were cut from blocks of the anastomosis area of colorectal sections of these patients. According to expert pathology review, these blocks have normal morphology. One patient (21 crypts) who has no recurrence will be used as control, four patients had local recurrence,³⁵ and three patients had distant recurrence³⁶ (Table 1).

2.1 Preparation of Samples

Formalin-fixed, paraffin-embedded colorectal tissues were retrieved from the histopathology files of Soroka University Medical Center (SUMC) in Beer-Sheva, Israel. The colorectal cancer patients have undergone surgery since 2007 and their records were retrieved with their consent, after the SUMC Institutional Review Helsinki Board approval. Two paraffin longitudinal tissue sections were cut from each biopsy; one was placed on an IR-transparent zinc-selenium slide, and the other on a glass slide. We followed the method proposed by Argov et. al.¹⁹ for sample preparation and data acquisition, where sections 10-microns thick were used. The second slide was stained with hematoxylin and eosin for histology review by an expert pathologist. Three measurements were taken from each crypt, one from the bottom, the second from the middle, and the third from the top, as can be seen in Fig. 1. The average size of the crypt in normal colonic mucosa is about 0.5 mm.

Table 1 Details about the samples included in this study, including patient numbers, number of biopsies from each patient, number of measured crypts from each biopsy, and the medical files of the patient.

Patient number	Number of biopsies	Number of crypts	Medical files
1	2	21	No recurrence
2	1	13	Local recurrence
3	2	21	Local recurrence
4	1	9	Local recurrence
5	1	10	Local recurrence
6	2	24	Distant recurrence
7	2	20	Distant recurrence
8	1	10	Distant recurrence



Fig. 1 Longitudinal cross section of colonic tissue showing a few crypts. The measurements were performed along the height of the crypt at three sites; B, M, and T.

2.2 FTIR Measurements

Data were collected in the transmission mode, as used in all our previous studies on colonic tissues, because this sampling method provides a good signal to noise ratio. Spectra were collected at 4-cm^{-1} spectral resolution from 600 to 4000 cm^{-1} . The measured spot has a circular shape with a diameter of $100\ \mu\text{m}$ and it takes about one minute to acquire a spectrum of 128 scans. For the analysis, ranges from 600 to 1800 cm^{-1} and 2800 to 2930 cm^{-1} were used. Previous studies have shown that these spectral ranges are most informative in terms of detection of cell activity and proliferation.

2.3 Spectral Manipulation

Spectral manipulations of the measured samples were performed using OPUS 7. The spectra were bisected into high (2800 to 3200 cm^{-1}) and low (900 to 1775 cm^{-1}) regions, respectively. Both regions were subjected to baseline corrected ranges using concave rubber band methods (with 32 and 64 consecutive points for the low region, the high region, respectively, and five iterations in the manual mode). In addition, the spectra were normalized using vector normalization followed by offset correction.

2.4 Multivariate Analysis

Multivariate analysis is commonly used in problems of classification. Principle component analysis (PCA) and linear discriminant analysis (LDA) are extensively used as differentiation tools in issues involving biological samples.^{37–42} The PCA and LDA calculations were done using the software codes developed in-house, and based on MATLAB® (The MathWorks, Natick, Massachusetts) software.

PCA is used for dimensionality reduction of the measured spectra. Using PCA we find PCs with the highest variability

and ignoring the direction with low variability in order to improve the recognition. In the PCA procedure, the data is represented by only a partial set of the new basis vector (in the present work we used PC1-PCK, with $1 \leq K \leq 8$). Moreover, many experiments show that in a significant number of problems the variability is correlated with the reparability, so that properly choosing K will lead to a better characterization of the sample.

The LDA calculations followed the PCA as a linear classifier.⁴³ The assumptions using LDA are that the probability density of each class is assumed to be Gaussian and centered around the mean and the covariance matrix is the same for all classes.

In the leave one out (LOO) algorithm, the training set contained all but one of the measured spectra. The category of the left-out spectrum was then calculated and compared to the known category. This procedure was repeated many times, but in each repetition a different measured spectrum was left out. The LOO algorithm is a good method when the number of points is relatively small, as it uses all the points except one for training. In the LOO algorithm all the data points are included, both for training and validation, exactly in the same manner; therefore randomness does not play a role.

3 Results

Our study is based on 128 crypts from 12 biopsies; all were taken from resection margins of eight colorectal adenocarcinoma patients. The resection margins were identified as normal by standard histology methods. Figure 2 shows the typical IR control absorption spectra measured from three different sites along the height of the longitudinal section of the same crypt-base, middle, and top—as shown in Fig. 1.

These spectra are displayed in the full mid-IR region before manipulation. In the 400 to 4000 cm^{-1} mid-IR region, the main features of the spectra are due to the organic molecules such as phospholipids, proteins, lipids, carbohydrates, and nucleic acids. The absorption bands with centroids at 2967 , 2926 , and 2859 cm^{-1} are due to CH_2 and CH_3 symmetric and anti-symmetric vibrations⁴⁴ and at 1744 cm^{-1} they are due to the

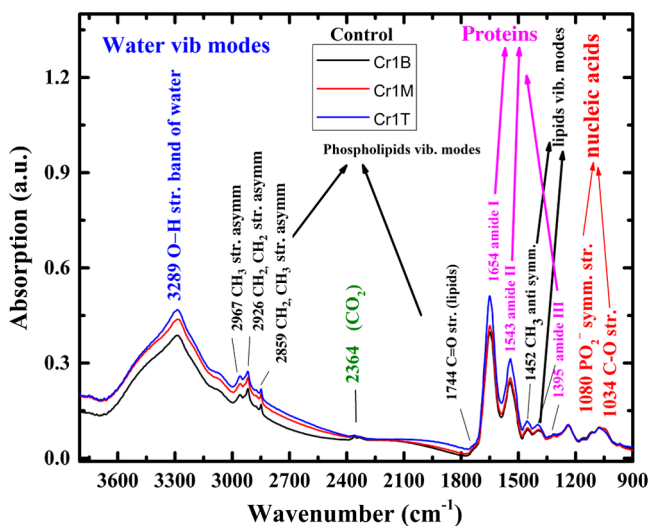


Fig. 2 Infrared (IR) absorption spectra measured from three different sites of the same crypt, taken from the control section in the range 900 to 3700 cm^{-1} before manipulation. The main vibrations are assigned in the figure.

phospholipids and C=O ester vibration.⁴⁵ Amide I (C=O stretching and C–N stretching vibrations at 1648 cm^{-1}), amide II (C–N stretching and a CNH bending vibration at 1553 cm^{-1}), and amide III (1252 cm^{-1}) bands are due to proteins.^{46–49} The peaks at 1241 and 1086 cm^{-1} are attributed to PO_2^- ionized antisymmetric and symmetric stretching, respectively.⁵⁰ The lower range of the spectrum, at 800 to 1185 cm^{-1} , reflects mainly carbohydrates. CH_2 , CH_3 , P=O functional groups of proteins, lipids, and phosphate compounds are the main attributers to the absorption bands in the 1185 to 1485 cm^{-1} range.

Figure 3 shows the absorption spectra from three different crypts in the 950 to 1775 cm^{-1} range after spectral manipulation, baseline correction, bisecting, and vector normalization.

The spectra were measured from bottom, middle, and top sites of the same crypt, respectively: (a) the same control crypt shown in Fig. 2, (b) crypt from a section belonging to a patient who experienced a local recurrence, and (c) crypt from section belonging to a patient who experienced a distant recurrence.

In all the measured crypts, we noticed that carbohydrates and nucleic acids absorption bands in the 900 to 1185 cm^{-1} range have higher absorption intensities for the middle sites when compared to the top and bottom spectra measured from the same crypt. There was no trend in the other sites; in some cases, the bottom was higher and in other cases the top was higher, without regard to the specific crypt type. Our main goal in this study was to classify the investigated crypts into three categories—control, local recurrence, and distant recurrence, based on their IR absorption spectra. We tried to differentiate among the three categories using spectra measured from bottom, middle, and top sites of the crypts. When we plotted the spectra measured from all 128 crypts from the same sites, the spectra were similar and overlapped, making it difficult to classify them using simple methods such as comparison of specific absorbance bands intensities, k-means, and clustering; thus, we used PCA followed by LDA for the classification goal.

We used PCA for dimensionality reduction. In this study the dimensions were reduced from hundreds to just a few. In PCA the data is expressed by a few points in such a way that most of the information in the measured data is preserved. The maximum dimension used in this study was eight. Two-dimensional plots of PC components were tried in order to check the PCA potential in differentiating between the various categories of the crypts. Figure 4 shows the results of PCA performed in the low wavenumber range of 900 to 1775 cm^{-1} for the classifications in the three groups. In this figure, we show the projection of the data on specific planes; the classification was done in couples. In Fig. 4(a), the data was classified into two groups, control and abnormal, the latter containing the crypts from local and distant recurrence patients. In Fig. 4(b), the classification was performed between the control and local recurrence crypts data. Figure 4(c) makes the same comparison using distant instead of local crypts data. Figure 4(d) shows the classification comparisons between local and distant recurrence crypts.

We are using PCA to look for a new basis called loadings or PCs, and each spectrum is represented as a linear combination of its basis. The variances among the classes is captured by the different PCs. PC1 has the largest variance reflected by the ratio of its eigenvalue over the sum of all the eigenvalues of all the PCs $[(\text{eigen value } 1)/(\text{sum of all eigen values})] \times 100$.

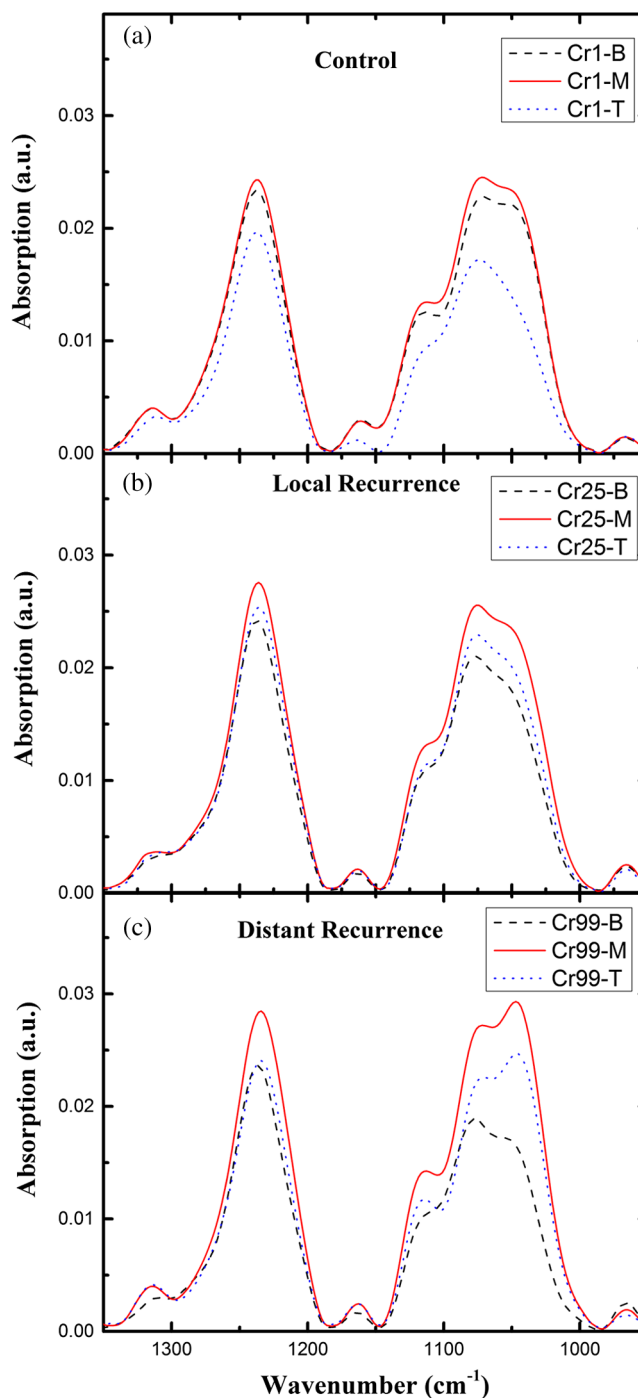


Fig. 3 IR absorption spectra measured from the bottom, middle, and top sites of the same crypt in the range 950 to 1350 cm^{-1} after spectral manipulation: (a) control crypt, (b) local recurrence crypt, and (c) distant recurrence crypt.

As can be seen, there is a clear trend of clustering in the four plots, but nevertheless, some of the data points of the different groups are mixed. In order to improve the classification results we applied LDA to the data after PCA calculation. LDA is a classifier that is applied to the data for classification purposes. The LDA separation rule suggests a larger variance between classes and a small variance within classes.

LDA calculations were applied to three different regions—the low region (950 to 1775 cm^{-1}), the high region (2800 to

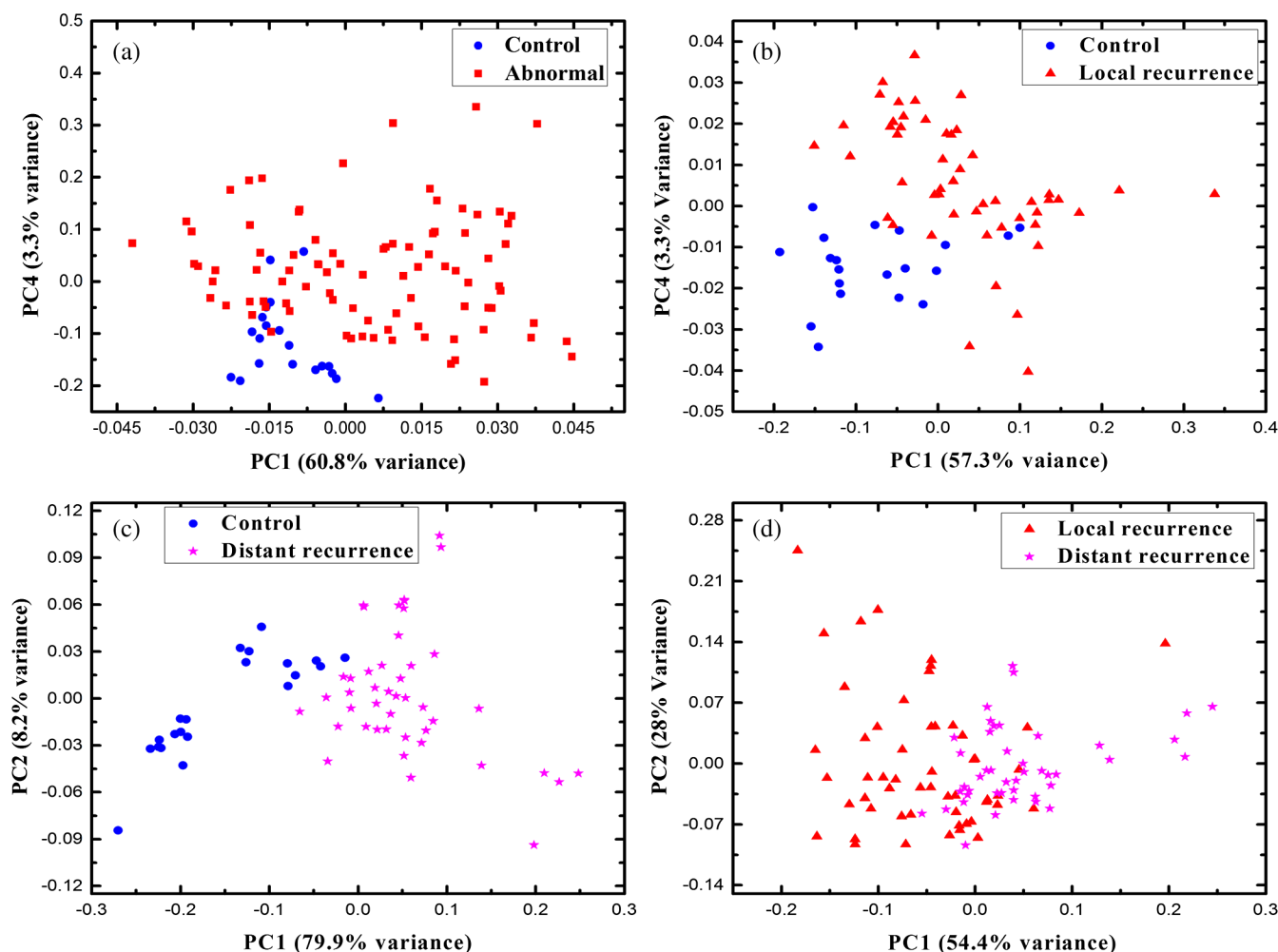


Fig. 4 Two-dimensional plots of the control, local recurrence, and distant recurrence crypts based on the principle component analysis (PCA) calculations in the 950 to 1775 cm^{-1} range. Part (a) represents PC4 versus PC1 for all data, clustered as control and abnormal categories. Part (b) represents PC4 versus PC1 for control and local recurrence crypts. Part (c) represents PC2 versus PC1 for the control and distant recurrence crypts data. Part (d) represents PC2 versus PC1 for the local and distant recurrence crypts data. In these plots, each individual spectrum is represented as one point in the specific projection plane. The variance of each PC is given in percentage.

3200 cm^{-1}), and the combined regions (2800 to 3200 cm^{-1} , 950 to 1775 cm^{-1}), in order to evaluate the classification rates of the spectra into the control, local recurrence, and distant recurrence groups.

For this goal in our spectra analyses, after PCA we used LDA—a statistical multivariate supervised method, with the LOO approach for validation. We repeated this algorithm many times (in many experiments), using different ranges and different sites of the spectra. LDA determines the best differentiation of the groups for each number of PCs, assigning the errors in identification of the spectra to the correct group. These errors are displayed as a function of PC numbers. As examples, we plotted two Figs. 5(a) and 5(b). Figure 5(a) shows the percentage of identification errors versus PC number, with the calculation based on the low wavenumber range 950 to 1775 cm^{-1} . The same calculations were performed three times—in the first we used data from the bottom sites of the crypts, in the second we used data from the middle sites, and in the third we used data from the top sites.

We tried two different strategies of classification. In the first strategy, we required simultaneous classification of the spectra

into three categories—control, local recurrence, and distant recurrence [Fig. 5(b)]. In the second strategy, the classification was done in two stages. In stage 1 the data was classified as control and abnormal, while the abnormal contains both local and distant recurrence crypts [Fig 5(a)]. In stage 2, the abnormal crypts were classified into local recurrence and distant recurrence crypts (data not shown). As can be seen from Figs. 5(a) and 5(b), the middle and bottom sites gave better results than the top. The second strategy gave better results than the first strategy, as reflected in the lower errors achieved using the first five PCs. Using the first strategy, the error was more than 20%, while using the second strategy with the same PCs number the error was less than 7%. We used a specific number of PCs to generate confusion matrixes, which were used to calculate the errors associated with each experiment. In each matrix we listed the success and error rates of the classification procedure.

Table 2 shows the confusion matrices for several experiments for different classifications employing different PCs. We tried the same analysis for the high wavenumber region but the results were poor. In addition, using the combined region also did not improve the results achieved in the low wavenumber region.

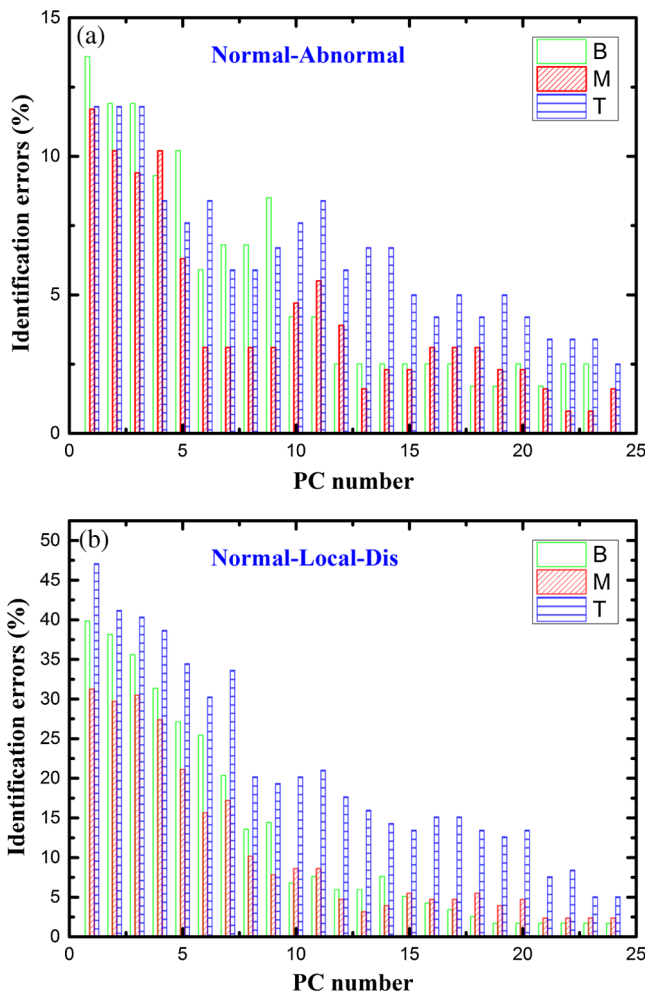


Fig. 5 Identification errors in percentage as a function of PC numbers for data in the (950 to 1775 cm^{-1}) range. (a) The data was classified into three classes, control, local recurrence, and distance recurrence. (b) The data was classified into two classes, control and abnormal (with local and distance recurrence added together). The identification was estimated using linear discriminant analysis calculations, applied to the data after PCA calculations were performed. For both plots, the calculation was based on spectra measured from bottom (green), middle (red), and top (blue) sites of the crypts included in this study.

Thus, we chose to introduce the classification results based on the low wavenumber region alone.

In order to make some speculation about biological changes in the epithelial cells in the colonic tissue in three biological systems, the LDA analysis was performed on three different ranges of the 950 to 1775 cm^{-1} low region—range I (950 to 1350 cm^{-1}), range II (1350 to 1480 cm^{-1}), and range III (1480 to 1775 cm^{-1}). The first range is dominated by carbohydrates and nucleic acids, the second range is dominated by lipids and proteins, and the third range is dominated mainly by the protein absorption bands. The calculations were performed on the spectra measured from the middle sites of the crypts. The error rates of the classification results, together with the number of PCs used in each calculation, are summarized in Table 3.

In our analyses, every spectrum was treated as individual case, but many of them are from the same biopsy. Thus, we repeated our LDA analysis using a “voting procedure.” In this procedure, from each class of the same tissue, several FTIR spectra were obtained from different crypts. The category of

Table 2 Successful identification of the colorectal crypts using linear discriminant analysis (LDA) with the leave one out (LOO) approach. In (a), the identifications were obtained using LDA calculations and the LOO approach for the middle sites spectra, with six PCs. In (b), the identifications were obtained using the middle site spectra with one PC. In (c), the identifications were obtained using the bottom spectra with five PCs. In (d), the identifications were obtained using the middle sites spectra with one PC. In (e), the identifications were obtained using middle sites spectra with eight PCs.

(a) Control, local recurrence, and distant recurrence			
	Control	Local	Distant
Control	20	0	1
Local	0	50	3
Distant	0	0	54

(b) Control and abnormal		
	Control	Abnormal
Control	17	4
Abnormal	0	107

(c) Control and local recurrence		
	Control	Local
Control	18	3
Local	0	53

(d) Control and distant recurrence		
	Normal	Distance
Normal	17	4
Distance	0	54

(e) Distant recurrence and local recurrence		
	Distant	Local
Distant	47	7
Local	1	52

a given tissue was determined by a voting procedure: the training set is taken from all the other tissues. Then we pick each spectrum of the tissue, one by one, and determine its category. The tissue’s category is chosen as the one that was predicted the maximal number of times (hence the name “voting”). Using the voting procedure, an 83.3% success rate was achieved.

4 Discussion

Relapse of colorectal cancer causes great burdens, both mentally and physically, with inconvenience induced by the need for long treatments and follow-up. Determining safe surgical margins of

Table 3 Identification errors in percentage versus PC number, derived using LDA calculation for ranges I (950 to 1350 cm^{-1}), II (1195 to 1726 cm^{-1}), and III (600 to 1195 cm^{-1}). The calculation was performed to differentiate between all couples within the three categories.

Number of PCs	Range I	Range II	Range III
(a) Control—abnormal			
PC1	16.4	8.6	10.2
PC1 to PC2	15.6	9.4	8.6
PC1 to PC3	9.4	10.2	8.6
PC1 to PC4	10.2	10.9	8.6
PC1 to PC5	10.9	7.0	5.5
PC1 to PC6	10.9	7.0	4.7
PC1 to PC7	10.9	7.0	5.5
PC1 to PC8	12.5	7.0	4.7
(b) Control—local recurrence			
PC1	25.7	20.3	17.6
PC1 to PC2	25.7	21.6	14.9
PC1 to PC3	17.6	21.6	12.2
PC1 to PC4	17.6	20.3	14.9
PC1 to PC5	16.2	5.4	9.5
PC1 to PC6	12.2	9.5	8.1
PC1 to PC7	13.5	9.5	6.8
PC1 to PC8	13.5	5.4	4.1
(c) Control—distant recurrence			
PC1	20.0	5.3	6.7
PC1 to PC2	20.0	6.7	8.0
PC1 to PC3	8.0	9.3	4.0
PC1 to PC4	6.7	2.7	5.3
PC1 to PC5	5.3	0	2.7
PC1 to PC6	5.3	0	1.3
PC1 to PC7	5.3	2.7	1.3
PC1 to PC8	4.0	2.7	1.3
(d) Local recurrence—distant recurrence			
PC1	44.9	28.0	18.7
PC1 to PC2	35.5	27.1	21.5
PC1 to PC3	22.4	13.1	19.6
PC1 to PC4	24.3	15.0	16.8
PC1 to PC5	23.4	18.7	15.9
PC1 to PC6	23.4	15.9	15.9
PC1 to PC7	7.5	5.6	13.1
PC1 to PC8	10.3	6.5	6.5

the removed cancerous lesion during colorectal surgery is of great importance, because it will hopefully minimize both the local recurrence and subsequently, any distant cancer disease. As a consequence, it may improve patient survival and quality of life, and may also reduce a patient's distress. The main purpose of this interdisciplinary study is to evaluate the potential of FTIR spectroscopy in tandem with multivariate analysis to detect and predict whether the patient will experience colorectal cancer recurrence based on colorectal tissue taken during surgery from the margins of the removed diseased part of the colon. The promising classification results obtained in this study are very important in this direction. Enlarging the database and the number of patients and intensive studies in this field may lead, in the future, to helping surgeons detect and assess the existence of free micrometastasis margins of the removed section of the colon in order to reduce the recurrence rate.

The final goal of the project is to analyze frozen tissue sections during surgery and this initial study was based on formalin-fixed, paraffin-embedded tissue samples. It is important to emphasize that all the samples have passed the same procedure of formalin fixation and deparaffinization before measurements. This procedure may influence the lipid signatures in the higher wave number region as previously shown.³³ However, all our samples have passed the same procedure of preparation and our analysis was based on all of the wavenumbers of the spectrum where proteins, carbohydrates, and other molecules do exist. These molecules are not influenced by the preparation procedure. Thus, we believe that our results are good and similar to what could be achieved using frozen tissues. Our logical next step should be to perform the analysis on FTIR spectra obtained from frozen tissues. Thus, we may be able to give the surgeon a quick and safe enough FTIR spectrum analysis report during the surgery to decide about the level of resection margins.

The PCs, calculated using PCA, are arranged according to their weights in generating the calculated spectra as linear combinations of these PCs. PC1 has the largest weight; the second PC (PC2), has the largest weight after PC1, and so on. The weights of these PCs are determined by their eigenvalues. Based on the characters of these PCs, it is possible to suggest a correlation between the differences among the various classes and the number of PCs used in order to classify them. In other words, it is possible to suggest that the largest spectral differences between the categories are in those sites where the differentiation is the best, according to a lower number of PCs. According to this suggestion, the best results of the classification were obtained from the middle site spectra, as can be seen from Fig. 5 and Table 2. These results can be explained due to the biochemical changes along the crypt height. The human large intestinal surface consists of small crypts having epithelial cells secreting mucosa at a rate of about three liters per day. According to the "crypt model," the cells in the normal crypt have a high turnover rate, with a life cycle of approximately 48 h. The epithelial cells in a normal crypt continuously undergo mitosis at the bottom of the crypt, move gradually upward along the basement membrane, and are finally disposed into the intestinal lumen.^{27,51–53} In the middle of the crypts the epithelial cells are mature and are different relative to the bottom and top crypt cells. Thus, the maturation and migration of cells in the colonic crypt hold important clues to the origin of the premalignant and malignant stages of cancer.²⁵ Abnormal cell proliferation and biological changes have been known to be an indicator of the

initiation of malignancy.^{35,54} Studies about the growth pattern of the crypts have been undertaken to trace abnormal proliferation. Although certain markers have been used to identify abnormal cell proliferation^{51,52,54} the actual maturation of the crypt in terms of total biochemical changes has not been established. The detection of abnormal crypts is essential for the diagnosis of the spread of colorectal cancer and its effective management.^{16,36} In our previous studies, we have shown the possibility of using FTIR-MSP to help identify abnormal crypt proliferation in histologically normal biopsies.^{30,34} This abnormal growth cannot yet be observed by expert pathologists (based on cell morphology), but may develop later into premalignant and malignant stages.

As can be seen from Fig. 3, in all the crypts, without exception, the middle sites spectra have higher absorption bands in the carbohydrate region of 900 to 1185 cm^{-1} .

It is important to outline that there are significant advantages to investigating the crypts using longitudinal sections over transverse sections, as can be seen from Fig. 3. There are clear differences in the spectra measured from different sites along the height of the crypt. Thus, using transverse cross, the changes due to different sites along the height of the crypt may interrupt and confuse the spectral changes due to different categories such as normal and abnormal.

A correct classification was recorded when the decision method was correlated with the patient's medical files. Our preliminary results showed a good agreement with the relapse of colorectal malignancy, as appeared in the medical files (Fig. 5). The classification procedure was done with two strategies. The second strategy was better than the first, as reflected in Fig. 5.

Using PCA calculations the PCs were set in descending order relative to their weight, as reflected in their eigenvalues. Thus, PC1 has the largest weight because it has the largest eigenvalue. As can be seen from Figs. 5(a) and 5(b), the differentiation based on the middle sites using the first five PCs has the lowest number of errors. This may lead to speculation that major changes may have occurred in the epithelial cells at the middle sites of the crypt. It was possible to differentiate between control and distant recurrence crypts with 81% sensitivity and 100% specificity using the first PC only [Table 2(c)], while we needed five PCs to achieve 90% sensitivity and 100% specificity [Table 2(b)]. We may speculate that the differences between distant recurrent crypts and normal crypts are larger than the differences between the control and local recurrences. The differences between the local and the distant crypts are minor, but our system is still capable of differentiating between them. Using the middle sites' spectra and eight PCs, it was possible to achieve 92.5% accuracy, calculated as the diagonal sum of [Table 2(d)] divided by the total number of spectra.

In choosing the number of PCs in this study, we considered the complexity of the problem in addition to the number of classes to be classified.

PCA enables the use of vectors for features that have a lower dimension while preserving the classification performance. This dimensionality reduction has many advantages because it enables us to simplify the classifier and to improve the processing time.⁵⁵ The choice of the number of PCs depends on the number of classes and the level of classification. In this study, the investigated crypts were identified "morphologically" as normal crypts. This identification was made by an experienced surgeon during surgery. Their spectra are similar and overlap, with

minute differences in intensities. Thus the classification of these spectra into different classes requires the utilization of advanced mathematical and statistical tools to differentiate among them with high accuracy. We, therefore, used the first eight PCs for the classification procedure.

LOO is an approach that is extensively explored in machine learning. This approach was used with LDA calculation for cross-validation using, each time, different sets for prediction so that the test sets would be statistically independent. This approach is commonly used to estimate the error in a small sized population as well as ensuring the validation of results.^{56,57}

Examining the LDA classification results listed in Table 3, using the LOO algorithm for all the ranges in the low region, using PC1 (the highest variance), we observe several conclusions. (1) The changes between the different couples of the three categories are spread in all the cell components. It was possible to differentiate between the different couples with good accuracy. (2) The spectral differences between control crypts and distant crypts are the largest. Using the 1350 to 1480 cm^{-1} and 1480 to 1775 cm^{-1} ranges, it was possible to differentiate between control and distant sites with more than a 93% success rate. (3) The classification errors are high when differentiating between local recurrence and distant recurrence crypts, relative to the other couples. Thus, the biological changes are minor and we need more PCs to achieve good classification results. Moreover, the largest changes between these crypts are in the amide I and II bands, resulting from proteins absorption bands. This is also true for control and local recurrence crypts [Table 3(b)], where amide I and amide II bands gave the best differentiation results. (4) Range I (950 to 1350 cm^{-1}) gave the lowest classification results for all four different pairs. Although the middle sites gave the highest absorbance in the same crypt, for the control, local, and distant crypts (Fig. 2), when comparing the different classes, this range did not show a good differentiation trend. Using the "voting" procedure in LDA calculation with the first three PC's, an 83.9% success rate was achieved. It is important to mention that using this procedure our calculation was based on 12 biopsies while we have 128 crypts. We believe that this is the main reason for the success reduction and that using a higher number of biopsies can improve our success rate.

Typically, three measurements are made on a crypt at the base, middle and apex and about ten crypts were measured per histological section. It takes about 30 min for a quick diagnosis. In frozen sections no staining is involved, alleviating the need for extra processing time. The extra time required will involve transport of the tissue from the OT to the laboratory, time taken for preparing frozen section and data analysis which should be in the range of another 30 min. Therefore, within one hour, while the patient is still under surgical procedure a quick diagnosis can be made and the results conveyed back to the surgeon.

Our method is objective and is fully computerized, fast, and simple to use. As a consequence, it may be possible to develop the method for frozen tissue sections. The success of this method using frozen sections may initiate the process of applying the method on site during colorectal surgery in order to assist the surgeon in determining the safe surgical margins of the removed cancerous lesion.

5 Conclusions

It was possible to differentiate between control, local, and distance recurrence crypts based on their IR absorption spectra and

using multivariate analysis PCA followed by LDA. The longitudinal sections of the crypts contain viable information due to their unique biological structure. Such information does not exist in the transverse sections.

Enlarging the database to include more crypts from the three categories will enable the use of more sophisticated nonlinear classifiers, which may improve the results and achieve even higher success rates.

Acknowledgments

The authors mourn the loss of their esteemed colleague Dr. Shmuel Argov from the Department of Pathology, Soroka University Medical Center, who passed away during the performance of this study. This manuscript is dedicated to him in loving memory. Dr. Elad Shufan is gratefully acknowledged for fruitful discussion concerning pattern recognition issues.

References

- American Cancer Society, "Cancer facts and figures 2014," American Cancer Society, Atlanta, <http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf> (2014).
- American Cancer Society, "Cancer facts and figures 2013," American Cancer Society, Atlanta, <http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-036845.pdf> (2013).
- "Beating bowel cancer," <https://www.beatingbowelcancer.org/facts-and-figures>.
- M. S. Cappell, "From colonic polyps to colon cancer: pathophysiology, clinical presentation, screening and colonoscopic therapy," *Minerva Gastroenterol. Dietol.* **53**(4), 351–373 (2007).
- R. G. Beets-Tan and G. L. Beets, "Rectal cancer: review with emphasis on MR imaging," *Radiology* **232**(2), 335–346 (2004).
- Y. B. Chen et al., "Correlation of tumor microvessel density to metastasis and recurrence of rectal cancer," *Chin. J. Cancer* **23**(10), 1203–1206 (2004).
- S. Manfredi et al., "Population-based study of factors influencing occurrence and prognosis of local recurrence after surgery for rectal cancer," *Br. J. Surg.* **88**(9), 1221–1227 (2001).
- L. Stocchi et al., "Impact of surgical and pathologic variables in rectal cancer: a United States community and cooperative group report," *J. Clin. Oncol.* **19**(18), 3895–3902 (2001).
- A. Wibe et al., "Effect of hospital caseload on long-term outcome after standardization of rectal cancer surgery at a national level," *Br. J. Surg.* **92**(2), 217–224 (2005).
- I. J. Adam et al., "Role of circumferential margin involvement in the local recurrence of rectal cancer," *Lancet* **344**(8924), 707–711 (1994).
- P. Quirke et al., "Local recurrence of rectal adenocarcinoma due to inadequate surgical resection. Histopathological study of lateral tumour spread and surgical excision," *Lancet* **328**(8514), 996–999 (1986).
- American Cancer Society, ACS, "Cancer facts and figures," <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2009/index> (2009).
- H. Mantsch and D. Chapman, *Infrared Spectroscopy of Biomolecules*, Wiley-Liss, New York (1996).
- A. Zwielly et al., "Advanced statistical techniques applied to comprehensive FTIR spectra on human colonic tissues," *Med. Phys.* **37**(3), 1047–1055 (2010).
- M. M. Mariani et al., "Micro-Raman detection of nuclear membrane lipid fluctuations in senescent epithelial breast cancer cells," *Anal. Chem.* **82**(10), 4259–4263 (2010).
- G. I. Zonios et al., "Morphological model of human colon tissue fluorescence," *IEEE Trans. Biomed. Eng.* **43**(2), 113–122 (1996).
- I. Georgakoudi et al., "NAD(P)H and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes," *Cancer Res.* **62**(3), 682–687 (2002).
- L. Chiriboga, H. Yee, and M. Diem, "Infrared spectroscopy of human cells and tissue. Part VI: a comparative study of histopathology and infrared microspectroscopy of normal, cirrhotic, and cancerous liver tissue," *Appl. Spectrosc.* **54**(1), 1–8 (2000).
- S. Argov et al., "Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients," *J. Biomed. Opt.* **7**(2), 248–254 (2002).
- A. Salman et al., "FT-IR microscopic characterization of normal and malignant human colonic tissues," *Cell. Mol. Biol.* **47**, 159–166 (2001).
- E. Bogomolny et al., "Early spectral changes of cellular malignant transformation using Fourier transform infrared microspectroscopy," *J. Biomed. Opt.* **12**(2), 024003 (2007).
- A. Salman et al., "Probing cell proliferation in the human colon using vibrational spectroscopy: a novel use of FTIR-microspectroscopy," *Vib. Spectrosc.* **34**(2), 301–308 (2004).
- L. Chiriboga et al., "Infrared spectroscopy of human tissue. II. A comparative study of spectra of biopsies of cervical squamous epithelium and of exfoliated cervical cells," *Biospectroscopy* **4**(1), 55–59 (1998).
- R. K. Sahu et al., "Prediction potential of IR-micro spectroscopy for colon cancer relapse," *Analyst* **135**(3), 538–544 (2010).
- J. M. Keller et al., "Peroxisome through cell differentiation and neoplasia," *Biol. Eur. Cell Biol. Organ.* **77**(1), 77–88 (1993).
- D. Naumann, "FT-Infrared and FT-Raman spectroscopy in biomedical research," in *Infrared and Raman Spectroscopy of Biological Materials*, H.-U. Gremlich and B. Yan, Eds., pp. 323–377, Marcel Dekker, Inc, New York (2000).
- G. L. Eastwood, "Gastrointestinal epithelial renewal," *Gastroenterology* **72**(5 Pt 1), 962–975 (1977).
- M. J. Wargovich et al., "Aberrant crypts as a biomarker for colon cancer: evaluation of potential chemopreventive agents in the rat," *Cancer Epidemiol. Biomarkers Prev.* **5**(5), 355–360 (1996).
- E. E. Deschner, "Cell proliferation as a biological marker in human colorectal neoplasia," in *Colorectal Cancer: Prevention, Epidemiology, and Screening*, P. S. S. Winawer and D. Schottenfeld, Eds., pp. 13–142, Raven Press, New York (1980).
- R. K. Sahu et al., "Detection of abnormal proliferation in histologically 'normal' colonic biopsies using FTIR-microspectroscopy," *Scand. J. Gastroenterol.* **39**(6), 557–566 (2004).
- A. Salman et al., "FTIR spectroscopic method for detection of cells infected with herpes viruses," *Biopolymers* **67**(6), 406–412 (2002).
- E. Bogomolny et al., "Early detection of premalignant changes in cell cultures using light-induced fluorescence spectroscopy," *Eur. Biophys. J.* **38**(7), 971–980 (2009).
- R. K. Sahu et al., "Can Fourier transform infrared spectroscopy at higher wavenumbers (mid IR) shed light on biomarkers for carcinogenesis in tissues?," *J. Biomed. Opt.* **10**(5), 054017 (2005).
- A. Salman et al., "Probing cell proliferation in the human colon using vibrational spectroscopy: a novel use of FTIR-microspectroscopy," *Vib. Spectrosc.* **34**(2), 301–308 (2004).
- D. F. Griffiths et al., "Demonstration of somatic mutation and colonic crypt clonality by X-linked enzyme histochemistry," *Nature* **333**(6172), 461–463 (1988).
- T. Enomoto, M. Kuranami, and A. Kakita, "Variations in the expression of platelet-derived endothelial cell growth factor in human colorectal polyps," *Surg. Today* **30**(8), 711–717 (2000).
- A. Salman et al., "Detection and identification of cancerous murine fibroblasts, transformed by murine sarcoma virus in culture, using Raman spectroscopy and advanced statistical methods," *BBA* **1830**(3), 2720–2727 (2013).
- A. Salman et al., "Identification of fungal phytopathogens using Fourier transform infrared-attenuated total reflection spectroscopy and advanced statistical methods," *J. Biomed. Opt.* **17**(1), 017002 (2012).
- A. Salman et al., "Utilizing FTIR-ATR spectroscopy for classification and relative spectral similarity evaluation of different *Colletotrichum* coccodes isolates," *Analyst* **137**(15), 3558–3564 (2012).
- A. Pomerantz et al., "Characterization of *Phytophthora infestans* resistance to mefenoxam using FTIR spectroscopy," *J. Photochem. Photobiol. B Biol.* **141C**, 308–314 (2014).
- A. Salman et al., "Classification of *Colletotrichum* coccodes isolates into vegetative compatibility groups using infrared attenuated total reflectance spectroscopy and multivariate analysis," *Methods* **68**(2), 325–330 (2014).

42. A. Salman et al., "Characterization and detection of vero cells infected with herpes simplex virus type 1 using Raman spectroscopy and advanced statistical methods," *Methods* **68**(2), 364–370 (2014).
43. R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.* **7**(2), 179–188 (1936).
44. Z. Movasaghi, S. Rehman, and D. I. ur Rehman, "Fourier transform infrared (FTIR) spectroscopy of biological tissues," *Appl. Spectrosc. Rev.* **43**(2), 134–179 (2008).
45. R. K. Dukor, "Vibrational spectroscopy in the detection of cancer," in *Handbook of Vibrational Spectroscopy*, John Wiley & Sons, Ltd., (2006).
46. G. Kos et al., "A comparative study of mid-infrared diffuse reflection (DR) and attenuated total reflection (ATR) spectroscopy for the detection of fungal infection on RWA2-corn," *Anal. Bioanal. Chem.* **378**(1), 159–166 (2004).
47. N. Stone et al., "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss.* **126**, 141–157 (2004).
48. E. O. Faolain et al., "Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents," *J. Histochem. Cytochem.* **53**(1), 121–129 (2005).
49. D. Naumann, "Infrared and NIR Raman spectroscopy in medical microbiology," *Proc. SPIE* 245–257, (1998).
50. P. G. Andrus and R. D. Strickland, "Cancer grading by Fourier transform infrared spectroscopy," *Biospectroscopy* **4**(1), 37–46 (1998).
51. T. C. Ko, W. A. Bresnahan, and E. A. Thompson, "Intestinal cell cycle regulation," *Prog. Cell Cycle Res.* **3**, 43–52 (1997).
52. C. S. Potten and M. Loeffler, "A comprehensive model of the crypts of the small intestine of the mouse provides insight into the mechanisms of cell migration and the proliferation hierarchy," *J. Theor. Biol.* **127**(4), 381–391 (1987).
53. L. Roncucci et al., "Aberrant crypt foci in colorectal carcinogenesis. Cell and crypt dynamics," *Cell Proliferation* **33**(1), 1–18 (2000).
54. M. Lipkin, "Proliferation and differentiation of normal and diseased gastrointestinal cells," in *Physiology of the Gastrointestinal Tract*, L. R. Johnson, Ed., 2nd ed., pp. 255–284, Raven Press, New York (1987).
55. I. Lapidot and J.-F. Bonastre, "Optimizing feature representation for speaker diarization using PCA and LDA," in *Speech Processing Conf.*, Tel-Aviv, Israel (2012).
56. T. Evgeniou, M. Pontil, and A. Elisseeff, "Leave one out error, stability, and generalization of voting combinations of classifiers," *Mach. Learn.* **55**(1), 71–97 (2004).
57. A. Elisseeff and M. Pontil, "Leave-one-out error and stability of learning algorithms with applications," in *Learning Theory and Practice*, J. Suykens et al., Eds., IOS Press (2002).

Ahmad Salman is a senior lecturer of physics in the Department of Physics, Shamoon College of Engineering, Israel. He received his PhD from Ben-Gurion University, Beer-Sheva, Israel. His research interests include vibrational spectroscopy (Raman and FTIR), and optical diagnostics. He has published many journal papers on biomedical optics.

Gilbert Sebbag received his medical graduation from Faculte de Medecine Rangueil, Toulouse III, France, and his Israel Surgery Board Certification. He received his master's degree in health administration from Ben Gurion University, Beer Sheva. He is the chairman of Surgery B Department, Soroka Medical Center, and a senior lecturer in surgical oncology in the Faculty of Health Sciences, Ben Gurion University, Beer Sheva, Israel.

Shmuel Argov: Biography is not available

Shaul Mordechai is a professor of physics and the head of the Biomedical Spectroscopy Laboratory at the Department of Physics, Ben Gurion University, Israel. He received his PhD from the Racah Institute of Physics, Hebrew University, Jerusalem. His research interests include medical physics, biomedical optics, FTIR-microscopy, FTIR-imaging, mid-IR spectrometry, fluorescence spectroscopy, and optical diagnostics. He has published numerous journal papers and book chapters in biomedical optics and collaborated with scientists from universities and national laboratories.

Ranjit K. Sahu is an adjunct professor at CUTM Paralakhemundi India. He received his PhD from Central Food Technological Research Institute, Mysore, India. He was a postdoctoral fellow at Ben Gurion University Israel and continues to contribute to the field biomedical optics (infrared spectroscopy).