FT-IR Microspectroscopy of Mouse Colon Tissues

Insight into the Chemistry of Carcinogenesis and Diagnostic Potential

Menashi A. Cohenford,* SeungJin Lim,*
Chris Brown,[†] Muhammad A. Chaudhry,[‡]
Saroj Sigdel,[‡] Emily Beckelhimer,^{*§} and
Basil Rigas[¶]

From the Departments of Integrated Science and Technology*
and Biology[§] and the Joan C. Edwards School of Medicine,[‡]
Marshall University, Huntington, West Virginia; the Department
of Chemistry,[†] University of Rhode Island, Kingston, Rhode
Island; and the Division of Cancer Prevention,[¶] Stony Brook
University, Stony Brook, New York

We studied colon carcinogenesis using Fourier-transform infrared (FT-IR) microspectroscopy, an evolving method that allows the nondestructive assessment of the chemical composition of cells and tissues and of the *in situ* relationship between molecules, and assessed its diagnostic potential. Mid-FT-IR spectra were obtained from frozen colon tissue samples of normal (C57BL/6J) and *Min* (*Apc*^{Min} mutant) mice, the latter recapitulating key features of human colon carcinogenesis. Classic spectroscopic analysis demonstrated marked differences in the Mid-FT-IR spectra between normal and dysplastic tissues, especially regarding peak positions and band intensity ratios in the regions 1800 to 985 cm⁻¹ and 3000 to 2700 cm⁻¹, reflecting changes in cellular nucleic acids, phosphates, and carbohydrates. Analysis of the spectra using the multivariate methods backpropagation neural networks, decision tree, adaboost with decision tree, and support vector machine, which interrogated the intrinsic dimensionality of IR spectra, revealed that their sensitivity was between 91.1% and 100% and their specificity between 94.1% and 100%, with the outcomes of the Support Vector Machine algorithm being identical to those of histologic analysis. FT-IR microspectroscopy holds great promise not only as a method of ascertaining changes in the chemistry of the neoplastic cells but also as a diagnostic tool, especially for early stages of carcinogenesis not detectable

by other means. (*Am J Pathol* 2012, 181:1961–1968; <http://dx.doi.org/10.1016/j.ajpath.2012.08.039>)

The mechanism by which a cell progresses from normalcy to malignancy is currently not completely understood. In colon cancer, for example, this sequence includes morphologic transitions from normal to adenoma to carcinoma. These cellular changes reflect complex and largely unknown changes in both the type of molecules that make up a cell and in the physical relationship of molecules to each other. Being nondestructive, Fourier-transform infrared (FT-IR) microspectroscopy provides the opportunity to probe such changes in intact tissues. Our study of cervical scrapings by FT-IR microspectroscopy demonstrated the power of this approach.^{1,2} Analysis of FT-IR spectra from cervical cells representing the various stages of cervical carcinogenesis revealed that the transformation from normalcy to malignancy is a gradual process, which is far more complex than what could be deduced from cytologic findings. Of interest, morphologically normal cervical cells from women with neoplasia differ structurally from normal cells from women with normal cervixes. The latter finding established the intuitive expectation that the chemical composition of a neoplastic cell changes before its morphologic changes occur. Several chemical changes associated with such transitions in the cervix and elsewhere have been inferred.^{2–5}

Treating cancer at its early stages improves patient care and reduces health care expenditures, but it requires objective and reliable methods for its detection. Following our initial description of the FT-IR spectroscopic features of cancer,⁶ investigators have focused on the application of this approach to cancer detec-

Supported by West Virginia NASA Epscor, West Virginia Space Consortium and Sure program, and Stony Brook Cancer Prevention Center.

Accepted for publication August 15, 2012.

Address reprint requests to Basil Rigas, M.D., Division of Cancer Prevention, Stony Brook University, HSC, T-17 Room 080, Stony Brook, NY 11794-8175. E-mail: Basil.Rigas@stonybrookmedicine.edu.

tion.^{7–11} Of the various FT-IR techniques, FT-IR microspectroscopy^{1,12} has received considerable attention because of its ability to sample defined components of a tissue, providing the much-needed focus on informative structural changes associated with carcinogenesis.

The present study applied FT-IR microspectroscopy to the study of colorectal cancer. The few reported FT-IR microspectroscopic studies of colorectal tissues have used mainly formalin fixed samples or samples deparaffinized with organic solvents.^{8–10} We used frozen colorectal sections to minimize specimen manipulation and to avoid potential interferences from extraneous processing reagents. As a result, we preserved the chemical composition of the tissues and enhanced the reliability of the study. This report describes the mid-FT-IR spectral features of normal and neoplastic mouse colorectal tissues and evaluates the diagnostic performance of four multivariate techniques that we used to analyze these spectra.

Materials and Methods

Animals and Tissue Processing

Colon specimens were obtained from two groups of 2-week-old male mice (Jackson Laboratory, Bar Harbor, ME): group 1 ($n = 5$): C57BL/6J mice with no history of *Apc* gene mutations [normal or wild type (WT)], and group 2 ($n = 5$): C57BL/6J-*Apc*^{Min}/J strain mice (referred to here as *Min* mice), which are congenic to C57BL/6J and recapitulate key features of human colon cancer.^{13–15} Mice were placed on a high-fat diet^{13,14} and sacrificed in pairs of one *Min* and one WT mouse, starting 11 weeks after the initiation of the high-fat diet. Colons were removed immediately, washed with saline, cut transversely, rolled on a stick from the cecum to the rectum with the distal tissue part forming the outermost layer of the wrap, immediately frozen in liquid nitrogen, and embedded in optimum cutting temperature medium (Ted Pella, Inc, Redding, CA). The optimum cutting temperature medium touched only the bottom and outermost part of the wrap, circumventing its central portion that was used for acquisition of spectra. Using a cryostat, 7- μ m sections were serially cut from each sample. One cut was placed on a conventional microscope slide, fixed in formaldehyde, stained with H&E, and reviewed by a pathologist who demarcated areas with normal or abnormal morphologic features. The second section was placed on a silver-doped tin oxide-coated IR reflective slide (Kevley Technologies, Chesterland, OH) and used for FT-IR microspectroscopy.

Acquisition and Analysis of FT-IR Spectra

Mid-FT-IR microspectroscopy was performed in reflection mode on unstained frozen tissue sections in a region of 4000 to 600 cm^{-1} using a Jasco IRT 5000 microscope linked to an FT-IR 6100 spectrometer (Jasco International Co., Ltd., Hichioji, Japan). The aperture was adjusted so that the measured area of the sample was 50 $\times$ 50 μm . Spectral resolution was 4 cm^{-1} , and 300 scans were co-added. Spectra were obtained from various areas of

each sample; the histologic features of the studied areas were determined on an adjacent tissue section that was stained and evaluated by a pathologist masked to sample identity.

FT-IR spectra from colorectal tissues were evaluated by classic spectroscopic methods, including determination of peak location, band height (amplitude) and area, and comparison of ratios of these 2 parameters, and multivariate techniques, including back-propagation neural network (BNN), decision tree (DT), Adaboost (AD) with DT, and support vector machine (SVM). Spectral analysis by these methods was performed under identical experimental settings, which involved focusing on the wave number region of 3300 to 900 cm^{-1} and choosing from this spectral range 400 specific wave numbers that were used in their evaluation. These wave numbers were compiled using the feature selection technique of Guyon et al.¹⁶ The wave number region of 3300 to 900 cm^{-1} was selected for the analysis because it best correlated the classification of spectra with biopsy diagnosis.

All spectra were normalized to a minimum and maximum absorbance value of 0.0 and 1.0, respectively. During analysis, spectra from morphologically normal-appearing tissues were assigned a dummy variable of -1.0 and those from preneoplastic (including dysplastic) tissues a dummy variable of $+1.0$. A 10-cross validation approach was then applied to modeling and predicting of diagnoses.

Statistical Analysis

Data were analyzed for normality using normal probability plots and Kolmogorov test statistics. All outcomes were approximately normally distributed. Band peak intensity ratios, areas, and shifts in spectral peaks were compared among the different groups of spectra by one-way analysis of variance followed by Tukey's multiple comparison procedure. Differences were considered significant at the $\alpha < 0.05$ threshold and were analyzed using SAS statistical software version 9.1 (SAS Institute Inc, Cary, NC).

Performance Characteristics of BNN, DT, AD-DT, and SVM

The performance characteristics of BNN, DT, AD-DT, and SVM were compared relative to biopsy results that served as the gold standard. Sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were calculated using the following conventional relationships with the variables true positive (TP), true negative (TN), false positive (FP), and false negative (FN): sensitivity = $\text{TP}/(\text{TP} + \text{FN})$, specificity = $\text{TN}/(\text{TN} + \text{FP})$, accuracy = $(\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$, positive predictive value = $\text{TP}/(\text{TP} + \text{FP})$, and negative predictive value = $\text{TN}/(\text{TN} + \text{FN})$.

Evaluation of Spectral Signatures by BNN, DT, AD-DT, and SVM

To detect the subtle differences among the spectra, focus was placed on different multivariate techniques, with each based on a unique computational principle. Below is a list of the methods used and a brief description of the attributes for each.

BNN has been successfully used to solve classification problems in several domains, especially, the Back-Propagation algorithm, which has been selected often for modeling IR spectra from tissues classified as normal and malignant.¹⁷ The way the algorithm works is that during the training process the Back-propagation algorithm adaptively changes free internal network parameter-based input spectra from the data set. As the training progresses, the weighed sum of the network parameters are iteratively adjusted to the actual class labels. The training continues until no more refinements to the model can be made. After this training phase, the neural network allows discrimination between spectra from normal or malignant tissues.

DT is a popular learning algorithm for classification and regression in data mining. The algorithm builds a DT from the observed data by repeatedly identifying the most predictive attributes consecutively. The most predicting attribute is determined typically by information gain, gain ratio, or Gini index. The motivation for using DT was driven by the following characteristics: i) DT is easy to understand and converts the learned model into production rules, allowing fast evaluation of test examples; ii) no *a priori* knowledge of the data distribution is required in DT¹⁶; and iii) several reports on the use of DT for analyzing IR spectra exist. Because of its simplicity and surprisingly good performance, DT has been frequently used as the first step in classification. In this study, the FT-IR spectra were evaluated by the C4.5 DT algorithm of Quinlan.¹⁸

AD is a metaclassification algorithm used to improve the accuracy of any given learning algorithm. The basic idea behind boosting refers to a general method of producing highly accurate predictions by combining moderately inaccurate (weak) classifiers. AD uses a given weak classification algorithm repeatedly, where at each step the weights of incorrectly classified samples (ie, spectra) are boosted to reinforce the weak classification algorithm to effectively classify the hard-to-analyze spectra. The main motivation for using a metaclassifier as AD was because the method has offered stellar performance with a number of data sets^{19,20} and the technique has yielded impressive results in analyzing IR spectra.²¹ In previous studies, we demonstrated that when used in conjunction with DT, AD enhanced the performance of DT.¹⁹

SVM is a supervised learning algorithm that attempts to find the class boundary between two different classes of data (spectra) by means of the optimal hyperplane separating the two classes.^{22–25} It is based on statistical learning theory and offers a clear theoretical analysis relative to other complex algorithms, such as artificial neural networks, often criticized for the black box nature

Figure 1. Mid-FT-IR spectra of colon tissues. **A:** Superimposed spectra of normal and dysplastic colon tissues in the frequency regions 2000 to 900 cm^{-1} (**upper panel**) and 3700 to 2600 cm^{-1} (**lower panel**) obtained as in *Materials and Methods*. Tissues were from normal WT mice (NW) and *Min* mice with low-grade (LG) or high-grade (HG) dysplasia. Representative tissue sections from normal (**B**), LG dysplasia (**C**), and HG dysplasia (**D**) are shown (H&E staining). Original magnification, $\times 10$ (**B** and **D**); $\times 20$ (**C**).

of the algorithm. Our motivation for using SVM its reportedly high performance with different types of data in a wide range of application domains, including text categorization, face recognition, gene expression measurements, and microarray expression profiles.

Results

FT-IR Spectra of Colon Tissues from WT and *Min* Mice

As expected, tissue sections from WT mice harbored normal tissues only, whereas those from *Min* mice included normal and dysplastic tissues (low or high grade). Indeed, of the five mice in group 1, no animal developed colorectal cancer or colorectal dysplasia. Four of five mice in group 2 developed colorectal dysplastic lesions. Mid-FT-IR spectra were acquired from >700 microscopic sites from the colons of the 10 study mice.

Figure 1A displays typical mid-FT-IR absorption spectra of normal and dysplastic colon tissues at the spectral regions of 2000 to 900 cm^{-1} and 3700 to 2600 cm^{-1} . The spectra from normal colonic tissues exhibited distinct bands peaking at a mean $\pm$ SD of $967.41 \pm 1.03 \text{ cm}^{-1}$, $1083.49 \pm 1.32 \text{ cm}^{-1}$, $1170.40 \pm 1.24 \text{ cm}^{-1}$, $1237.68 \pm 0.84 \text{ cm}^{-1}$, $1394.89 \pm 2.37 \text{ cm}^{-1}$, $1455.12 \pm 1.70 \text{ cm}^{-1}$, $1543.64 \pm 5.53 \text{ cm}^{-1}$, $1657.24 \pm 1.73 \text{ cm}^{-1}$, $2851.39 \pm 0.97 \text{ cm}^{-1}$, $2923.23 \pm 0.83 \text{ cm}^{-1}$, $2959.10 \pm 0.035 \text{ cm}^{-1}$, $3084.03 \pm 3.75 \text{ cm}^{-1}$, and $3292.36 \pm 1.22 \text{ cm}^{-1}$.

Classic spectroscopic analysis demonstrated that the mid-FT-IR spectra from dysplastic tissues typically yielded lower IR absorptions compared with normal tissues. These changes in IR absorption were observed

Table 1. IR Spectra in Regions 1280 to 985 cm^{-1} and 3000 to 2700 cm^{-1}

Study group	AUC, mean $\pm$ SD	
	1280 to 985 cm^{-1}	3000 to 2700 cm^{-1}
NW cells ($n = 246$)	19.82 $\pm$ 4.59*	11.49 $\pm$ 2.32 [†]
NM cells ($n = 244$)	17.38 $\pm$ 5.79 [‡]	11.81 $\pm$ 5.12 [‡]
LG dysplasia ($n = 151$)	12.62 $\pm$ 5.31 [§]	7.70 $\pm$ 2.23 [¶]
HG dysplasia ($n = 136$)	14.03 $\pm$ 5.49 [§]	8.69 $\pm$ 3.62 [¶]
Combined NW cells ($n = 490$)	18.61 $\pm$ 5.36	11.65 $\pm$ 3.97
Combined LG and HD dysplasia ($n = 287$)	13.29 $\pm$ 5.43	8.18 $\pm$ 3.00

* $P < 0.0001$ for NW versus NM, LG, and HG.

[†] $P = 0.77$ for NW versus NM and $P < 0.0001$ for NW versus LG and HG.

[‡] $P < 0.0001$ for NM versus HG and LG.

[§] $P < 0.11$ for LG versus HG.

[¶] $P = 0.11$ for LG versus HG.

^{||} $P < 0.0001$ for CN versus CD.

CD, combined low grade and high grade dysplasia; CN, combined normal wild cells and normal *Min* cells; HG, high grade; LG, low grade; NM, normal *Min* cells; NW, normal wild-type cells.

throughout the region of 1800 to 985 cm^{-1} but most noticeably at the 1083 cm^{-1} and 1237 cm^{-1} bands (Figure 1A) and at the spectral region of 3000 to 2700 cm^{-1} (Figure 1A). To substantiate the variations in IR absorption among the different cell types, we computed the areas for various peaks in the spectra of normal tissues and tissues diagnosed as having low- and high-grade dysplasia. Table 1 summarizes the integrated spectral areas [area under the curve (AUC)] in the regions of 1280 to 985 cm^{-1} and 3000 to 2700 cm^{-1} . These regions were selected for study because of the conspicuous changes on simple inspection of the spectra. For the region of 1280 to 985 cm^{-1} , there was a progressive reduction in the AUC from normal WT tissues to normal tissues from *Min* mice to dysplastic tissues, with the exception that there was no statistically significant difference in this regard between low- and high-grade dysplasia. When the results from all normal tissues combined were compared with those from all dysplastic tissues, the latter showed a 28.6% reduction in AUC ($P < 0.0001$). The AUC for the region of 3000 to 2700 cm^{-1} revealed no difference between normal WT tissues and normal tissue from *Min* mice, but the reduction in dysplastic tissues was significant and pronounced. Again, no difference was found between low- and high-grade dysplasia when the cumulative results of normal and dysplastic tissues were compared, the AUC reduction in the latter was similar (29.8%; $P < 0.0001$).

The region of 1300 to 1000 cm^{-1} is dominated by spectral features of phosphodiester, with the symmetric and asymmetric phosphate stretching modes contributing to the bands at around 1083 cm^{-1} and 1237 cm^{-1} , respectively.^{1,3–8,26} The relative intensity of these bands was typically higher in normal tissues than in dysplastic tissues, an observation that, when combined with the data in Table 1, suggests the likelihood of varied phosphate levels in the two tissue types, with the normal cells exhibiting a higher phosphate content. In a study of human colorectal tissues by IR spectroscopy, Salman and

coworkers⁸ reported a similar pattern of phosphate content showing that the bands for phosphate groups, symmetric and asymmetric vibrations, were higher in normal tissues than malignant tissues.

In the same spectral regions, the spectra from preneoplastic tissues yielded bands identical to those of normal tissues; however, in some instances, their peak frequency was shifted. Noticeable were the statistically significant shifts of the 1170 cm^{-1} and 1237 cm^{-1} peaks, which in the spectra of low-grade dysplastic tissues ($n = 151$) shifted to 1171.42 $\pm$ 2.74 cm^{-1} and 1238.62 $\pm$ 1.02 cm^{-1} ($P < 0.0001$). In the spectra of high-grade dysplastic tissues ($n = 220$) these peaks were shifted further, appearing at 1172.92 $\pm$ 1.09 cm^{-1} and 1238.79 $\pm$ 0.93 cm^{-1} .

The peaks near 2851 cm^{-1} and 2923 cm^{-1} originate from symmetric and asymmetric CH_2 stretching modes and at around 2959 cm^{-1} from asymmetric CH_3 vibrations of proteins, lipids, and DNA.^{8,11,26–28} Our data in Table 1 reveal that the respective band areas differed significantly between normal and dysplastic tissues, a finding similar to observations made previously in human colonic tissues.^{7,9} The other spectral regions evaluated in this study included the bands at 1657 cm^{-1} , corresponding to amide I, and at 1543 cm^{-1} , corresponding to amide II. Table 2 summarizes the intensity ratios of these and other bands and compares them among various cohorts. Overall, a similar pattern of changes is observed: the value of these ratios decreases in dysplastic tissues compared with normal tissues. Some of the ratios (1083 cm^{-1} /amide II, 1237 cm^{-1} /amide II, and amide II/amide I) are significantly decreased in normal tissues derived from *Min* mice compared with those from WT mice. Other than the amide II/amide I ratios, no statistically significant difference between the two subtypes of dysplasia was noted.

In a few FT-IR spectroscopic studies, the ratio of integrated absorbance under amide I/II was used to indirectly measure the DNA content of a cell, with increases in the ratio suggesting a higher content of cellular DNA.^{29,30} Table 2 reports a statistically significant difference between the mean and SDs of amide II/I ratios in normal colon cells and in colon cells with high-grade dysplasia, suggestive of a varied DNA content in the two cell types. An analysis of the reciprocal values of amide II/amide I ratios suggested that the DNA content of cells with high-grade dysplasia was more pronounced than of the normal cells. This result is consistent with previous observations showing that in colorectal biopsy specimens, dysplastic and malignant cells are more commonly aneuploid than nonneoplastic cells.³¹

Next, we computed the distribution curves of the 1083 cm^{-1} /amide II and 1237 cm^{-1} /amide II peak intensity ratios from the spectra of various cohorts. In each instance, the mean and SD of ratios (Table 2) were fitted to a normal distribution curve. Fitting the data to normal curves was reasonable because the ratios from each cohort essentially followed a normal frequency distribution. Figure 2 shows the extent of overlap of each ratio of the four cohorts of spectra. Interestingly, the sequence of

Table 2. Band Intensity Ratios from IR Spectra of Mouse Colon Tissue Samples

Study group	1083 cm ⁻¹ /1237 cm ⁻¹	1083 cm ⁻¹ /AM I	1083 cm ⁻¹ /AM II	1237 cm ⁻¹ /AM I	1237 cm ⁻¹ /AM II	AM II/AM I
NW (n = 246)	1.11 ± 0.14*	0.46 ± 0.49 [†]	0.74 ± 0.15 [‡]	0.41 ± 0.07 [§]	0.66 ± 0.08 [¶]	0.63 ± 0.07
NM (n = 294)	1.08 ± 0.19**	0.46 ± 0.18**	0.63 ± 0.21**	0.43 ± 0.17**	0.58 ± 0.16**	0.73 ± 0.11**
LG (n = 151)	0.88 ± 0.09 ^{††}	0.28 ± 0.07 ^{‡‡}	0.045 ± 0.11 ^{††}	0.32 ± 0.05 ^{§§}	0.50 ± 0.08 ^{††}	0.64 ± 0.07 ^{¶¶}
HG (n = 136)	0.88 ± 0.129	0.27 ± 0.093	0.045 ± 0.17	0.30 ± 0.64	0.50 ± 0.12	0.60 ± 0.06

Tissues were from normal wild-type mice (NW) and *Min* mice that were either histologically normal (NM) or had low-grade (LG) or high-grade (HG) dysplasia. Data are expressed as mean ± SD.

**P* = 0.11 for NW versus NM and *P* < 0.0001 for NW versus LG and HG.

[†]*P* = 0.92 for NW versus NM and *P* < 0.0001 for NW versus LG and HG.

[‡]*P* < 0.0001 for NW versus NM, LG, and HG.

[§]*P* = 0.15 for NW versus NM and *P* < 0.0001 for NW versus LG and HG.

[¶]*P* < 0.0001 for NW versus NM, LG, and HG.

^{||}*P* < 0.0001 for NW versus NM, *P* is nonsignificant for NW versus LG, and *P* = 0.02 for NW versus HG.

***P* < 0.0001 for NM versus LG and HG.

^{††}*P* = 0.99 for LG versus HG.

^{‡‡}*P* = 0.75 for LG versus HG.

^{§§}*P* = 0.34 for LG versus HG.

^{¶¶}*P* = 0.002 for LG versus HG.

AM, amide.

these curves parallels the progression of colorectal carcinogenesis in both cases.

Analysis of spectra from a larger cohort of morphologically normal and dysplastic tissues revealed that the band intensity ratios at 1045 cm⁻¹/amide II, which provide for an estimate of carbohydrate levels,^{8,32} were on an average lower in *Min* mice than in WT mice (eg, 0.59 ± 0.32, 0.47 ± 0.20, and 0.55 ± 0.18 versus 0.76 ± 0.29 for the histologically normal, low grade, high grade, and WT normal cohorts, respectively). This difference was statistically significant even when multiple comparisons were made by analysis of variance between the spectral groups of WT normal and histologically normal (*P* < 0.001) and WT normal with the low-grade and high-grade cohorts combined (*P* < 0.001). Figure 2 displays the

distribution for the mean and SD of the 1045 cm⁻¹/amide II band ratios for each group of spectra that were analyzed.

Analysis of Spectra by Four Multivariate Methods

Multivariate analysis interrogates the intrinsic dimensionality of IR spectra and offers distinct advantages over classic spectroscopic analysis. This approach has been successfully used to assess the diagnostic potential of IR spectra.^{1,2,8,12,33} The spectra from our study were evaluated by four multivariate methods: BNN, DT, AD-DT, and SVM.

Our goal was to separate the spectra from morphologically normal-appearing tissues from those of dysplastic tissues. Thus, two sets of spectra were compared. One set included the spectra obtained from WT and *Min* mice tissues diagnosed as normal, whereas the other included spectra from tissues with dysplasia (low or high grade). The validation criterion relied on cytologic evaluation (morphologic review), which was used as the gold standard. In total, 894 spectra were analyzed of which 523 were from morphologically normal tissues (WT normal = 246 and histologically normal = 277) and 371 from tissues with dysplasia (low grade = 151 and high grade = 220). Table 3 summarizes the performance characteristics of each of the four methods that we used. In general, all methods performed well, exhibiting sensitivity between 91.1% and 100% and specificity between 94.1% and 100%. The most impressive finding was the performance of the SVM method whose outcomes were identical to those of cytologic testing. Thus, these four methods can be ranked in decreasing order as follows: SVM, AD-DT, BNN, and DT.

Figure 3 shows the predicted class boundary between the spectra from every normal and normal-appearing and every dysplastic tissue sample evaluated by SVM. The markings below (x) and above (+) the zero line represent the predicted boundary for the two categories of tissues, respectively. Notably, spectra from dysplastic tissues

Figure 2. The distribution curves of band intensity ratios. The ratios of the amplitude of the 1083 cm⁻¹/amide II band (left panel), the 1237 cm⁻¹/amide II band (right panel), and the 1045 cm⁻¹/amide II band (bottom panel) from the IR spectra from each of the four study groups were fitted into distribution curves, as described in *Materials and Methods*; each mean value was multiplied by 100. Tissues were from normal WT mice (NW) and *Min* mice that were either histologically normal (NM) or had low-grade (LG) or high-grade (HG) dysplasia.

Table 3. Diagnostic Performance of IR Spectra from Mouse Colon Tissues Using Multivariate Methods

Diagnostic status	DT	BNN	AD-DT	SVM
True positive, No. (n = 371)	339	338	357	371
False positive, No.	31	15	15	0
True negative, No. (n = 523)	492	508	508	523
False negative, No.	32	33	14	0
Sensitivity, %	91.4	91.1	96.2	100
Specificity, %	94.1	97.1	97.1	100
False-positive rate, %	5.9	2.9	2.9	0
False-negative rate, %	8.6	8.9	3.8	0
Positive predictive value, %	91.6	95.8	96.0	100
Negative predictive value, %	93.9	93.9	97.3	100
Accuracy, %	93.0	94.6	96.8	100

yielded less variability in their margins than normal tissues.

Because SVM proved highly accurate in distinguishing normal from dysplastic tissues, we studied its ability to discriminate between the subcohorts of normal tissues. Figure 4 shows the predicted class boundaries between the spectra from both normal cohorts, with the zero line serving as the margin for predicting diagnosis status. SVM correctly identified all WT normal spectra, whereas it misidentified as WT normal 7 of 277 histologically normal spectra. These misclassified spectra can be seen in Figure 4 as the seven × markings crossing above the zero line. Thus, the analysis of the normal cohorts yielded an overall accuracy of 98.7%.

Discussion

Our mid-FT-IR microspectroscopy study documents the spectroscopic features of a series of murine colon tissues

Figure 3. Predicted class boundary of IR spectra evaluated using SVM. All of the study spectra (from groups) were evaluated by SVM as described in *Materials and Methods* and plotted as shown. The markings above and below the zero line display the respective predicted margin to class boundary for all dysplastic fields (+, low grade and high grade) and normal fields (x, normal from WT mice and *Min* mice) that were microspectroscopically evaluated. As shown, spectra from dysplastic tissues yielded less variability in their margins than normal tissues.

recapitulating the progression from normal to cancer. Our IR spectroscopic findings capture the continuum of progressively accumulating changes in the colonocyte that ultimately lead to its malignant phenotype. Furthermore, our findings establish the potential of IR microspectroscopy to form the basis of a diagnostic method for early neoplastic changes, including those that precede the onset of overt morphologic changes.

Our experimental system includes four types of tissues: normal colonic tissues; tissues that are morphologically indistinguishable from the previous ones, differing only by a single mutation in the *Apc* gene, which is considered key to initiating the carcinogenic process in the colon^{13–15}; and tissues advancing toward frank malignancy, namely, tissues with low- or high-grade dysplasia. The low- and high-grade tissues are defined by changes in their microscopic appearance, which include substantially overlapping features.

Classic spectroscopic analysis, limited as it is, provided highly informative findings. We demonstrated that in certain areas of the IR spectrum, such as the 1280 to 985 cm^{-1} region, there was a progressive decrease of the AUC values going from normal to the *Apc* mutated cell to the dysplastic colonocyte. However, the two morphologic stages of dysplasia gave values that were not statistically different. In contrast, the second spectral region that we analyzed in a similar manner, 3000 to 2700 cm^{-1} , provided a separation between normal and dysplasia but not between their two respective subgroups. These differences reflect the fluidity of the morphologic assessment and also the multiple (and different) chemical changes in the tissue that are captured by different regions of the IR spectrum. This phenomenon is apparent in the significant overlap of the distribution curves of two spectral band ratios of the four groups, 1083 cm^{-1} /amide II and 1273 cm^{-1} /amide II, that, nevertheless, maintain discrete peaks.

Figure 4. Classification of spectra from normal tissues. Data were generated by the SVM Algorithm. All spectra from *Min* mice were from areas that appeared histologically normal. The markings above and below the zero line display the respective predicted margin to class boundary for every normal field from WT mice (+) and *Min* mice (x) that were microspectroscopically studied.

Our spectral findings reflect the changes in the colono-cyte because altered morphologic features reflects changes in the chemistry of the cell. However, even subtler changes, not yet appreciable morphologically, are also detected by IR microspectroscopy, as is the case in the differences between normal tissues from WT and *Min* mice. However, the precise nature of the chemical changes underlying the spectroscopic changes is difficult to ascertain based on current understanding. Tentative conclusions (or inferences in some cases) have been reported by us^{1,2,4,5} and others^{8,11,32} regarding the chemical basis of IR findings. For example, the increased shift in the peak around 1170 cm^{-1} has been attributed to reduced levels of hydrogen bonding in cellular proteins.¹¹ In colon tissues, the bands around the 1083 cm^{-1} and 1237 cm^{-1} region have been assigned primarily to symmetric and asymmetric phosphate stretching modes, respectively, and speculated to originate from the phosphodiester groups of nucleic acids.^{1,4,6,8,11} Although the peaks around 1083 cm^{-1} and 1237 cm^{-1} can also arise from collagen, the contribution of collagen to these peaks is minimal in colorectal tissues.^{6,8} Finally, the 1045 cm^{-1} band has been ascribed to cellular carbohydrates,³² whereas the ratio of the 1045 cm^{-1} /amide II bands has been used to provide an estimate of cellular carbohydrate levels.^{8,33,34}

Regardless of these limitations, our data underline the potential of IR spectroscopy to map in a noninvasive way the progressive shift of chemical changes during carcinogenesis, generating new insights into this process. Deciphering such changes would be particularly helpful for the earliest stages of the process, preceding the onset of morphologic changes. An improved ability to correlate chemical and spectral changes would provide a critical insight into cancer that may be difficult to obtain otherwise.

An interesting but not totally unexpected finding was the close similarity between the IR spectra of mouse and human colorectal tissues. For example, in human and mouse colorectal tissues, a major distinguishing feature between normal and dysplastic tissues lies in their IR band intensities for peaks around 1083 cm^{-1} and 1237 cm^{-1} arising from the symmetric and asymmetric phosphate stretching vibrations, respectively. Previous studies by us and others^{6,8,34,35} found that malignant human colorectal tissues displayed decreased amplitude of these bands relative to their normal counterparts, similar to what we observed in mice. One of these reports⁸ attributed these differences to the higher total phosphate content of normal tissues compared with malignant types. Finally, a similar trend in the carbohydrate content that we observed in benign and malignant colorectal tissues was also shown in human colonic tissues.⁸

The application of newer methods to IR spectra analysis, unavailable when the initial observations on mid-FT-IR spectroscopic changes associated with cancer were made,^{1,2} revealed that current sophisticated approaches can enhance the diagnostic potential of IR microspectroscopy. The four multivariate methods that we used in these studies, BNN, DT, AD-DT, and SVM, varied in their ability to detect subtle differences between spec-

tra that affect the diagnostic sensitivity, specificity, and other relevant parameters of IR microspectroscopy. Of them, SVM held the highest promise, yielding 100% correlation between spectra classification and diagnosis. The rest in order were AD-DT, BNN, and DT. Consistent with our observation is the recent FT-IR study of Xie et al,³⁶ which demonstrated that when combined with discriminant analysis, SVM yielded sensitivity and specificity values higher than 90% when detecting human colorectal cancer. SVM's superior performance has been attributed to its ability to model complex nonlinear boundary-generating models much less prone to overfitting relative to other multivariate techniques.^{37,38} The performance of these methods, remarkable as it is, ought to be viewed in the context of our study, which is constrained by its size and experimental setting. Nevertheless, others have reported congruent results, strengthening the notion that IR microspectroscopy has a diagnostic potential, with one of them reporting near-perfect classification accuracy in determining normal, dysplasia, and carcinoma in a rat model system.³⁷ However, to be used effectively as a diagnostic tool, larger data sets need to be systematically developed and evaluated by techniques such as SVM.

In summary, FT-IR spectroscopy is technically maturing as a versatile method for the study of tissues and cells. As our results show, it holds great promise not only as a diagnostic tool, especially for the early stages of carcinogenesis not detectable by any other means, but also as a means of ascertaining early changes in the chemistry of the neoplastic cell associated with its dangerous transformation toward malignancy.

Acknowledgements

We thank Kevin Rice for providing assistance with data analysis, Jason Cohenford for help with manuscript review, and the Department of Pathology, St. Mary's Hospital (Huntington, WV), which provided histologic analysis support.

References

1. Cohenford MA, Rigas B: Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology. *Proc Natl Acad Sci U S A* 1998, 95:15327–15332
2. Cohenford M, Bhandare PS, Cahn FR, Krishnan K, Rigas B: Method for the detection of malignant and premalignant stages of cervical cancer. 2000. United States patent US 6031232.3
3. Rigas B, LaGuardia K, Qiao L, Bhandare PS, Caputo T, Cohenford MA: Infrared spectroscopic study of cervical smears in patients with HIV: implications for cervical carcinogenesis. *J Lab Clin Med* 2000, 35:26–31
4. Wong PT, Wong RK, Caputo TA, Godwin TA, Rigas B: Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis. *Proc Natl Acad Sci U S A* 1991, 88:10988–10992
5. Wong PT, Papavassiliou ED, Rigas B: Phosphodiesterstretching bands in the infrared spectra of human tissues and cultured cells. *Appl Spectrosc* 1991, 45:1563–1567
6. Rigas B, Morgello S, Goldman IS, Wong PT: Human colorectal cancers display abnormal Fourier-transform infrared spectra. *Proc Natl Acad Sci U S A* 1990, 87:8140–8144

7. Conti C, Ferraris P, Giorgini E, Rubini C, Sabbatini S, Tosi G, Anastassopoulou J, Arapantoni P, Boukaki E, Konstadoudakis S, Theophanides T, Valavanis C: FT-IR microimaging spectroscopy: a comparison between healthy & neoplastic human colon tissues. *J Mol Structures* 2008, 881:46–51
8. Salman A, Argov S, Ramesh J, Goldstein J, Sinelnikov I, Guteran H, Mordechai S: FT-IR microscopic characterization of normal and malignant human colonic tissues [published online]. *Cell Mol Biol (Noisy-le-grand)* 2001, 47:OL159–OL166
9. Argov S, Ramesh J, Salman A, Sinelnikov I, Goldstein J, Guteran H, Mordechai S: Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients. *J Biomed Opt* 2002, 7:248–254
10. Ramesh J, Salman A, Mordechai S, Argov S, Goldstein J, Sinelnikov I: FTIR microscopic studies on normal, polyp, and malignant human colonic tissues. *Subsurface Sensing Technol Appl* 2001, 2:99–117
11. Yunxiang C, Tiyu G, Jun F, Zhenquan G, Xiu K, Jiangqiang D: Comparative FT-IR spectroscopic analysis of human breast benign and malignant tissues. *Chinese Sci Bull* 1999, 44:804–808
12. Cohenford M, Bhandare PS, Cahn FR, Krishnan K, Rigas B: Method for the detection of cellular abnormalities using Fourier transform infrared spectroscopy. 2003. United States patent 6,620,621
13. Moser A, Pitot H, Dove W: A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* 1990, 247:322–324
14. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* 1992, 256:668–670
15. Lipkin M, Yang K, Edelmann W, Xue L, Fan K, Risio M, Newark H, Kucherlapati R: Preclinical mouse models for cancer chemoprevention studies. *Ann N Y Acad Sci* 1999, 889:14–19
16. Guyon I, Weston J, Barnhill S, Vapnik V: Gene selection for cancer classification using support vector machines. *Machine Learning* 2002, 46:389–422
17. Xiang Y, Tian J, Zhang Z, Dai Y: Diagnosis of endometrial cancer based on back-propagation neural network and near-infrared spectroscopy of tissue. *Fuzzy Systems Knowledge Discovery* 2009, 7:508–512
18. Quinlan R: C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers, 1993
19. Alvarez MA, Lim S: A machine learning approach for one-stop learning. *Data Mining and Knowledge Discovery Technologies*. Edited by D Taniar. Hershey, NY: IGI Publisher, 2008: chapter 14, pp. 333–357
20. Freund Y, Schapire RE: Experiments with a new boosting algorithm. *Thirteenth International Conference on Machine Learning*, San Francisco, CA 1996, pp 148–156
21. Tan C, Li M, Qin X: Study of the feasibility of distinguishing cigarettes of different brands using an Adaboost algorithm and near-infrared spectroscopy. *Anal Bioanal Chem* 2007, 389:667–674
22. Noble WS: Support vector machine applications in computational biology. *Kernel methods in computational biology*. Edited by B Schoelkopf, K Tsuda, JP Vert. Cambridge, MA: MIT Press, 2004, pp 71–92
23. Vapnik VN: *Statistical Learning Theory*. New York, NY: John Wiley & Sons, 1998
24. Chang CC, Lin CJ: LIBSVM-a library for support vector machines. *ACM transactions on intelligent systems and technology*, ch 2. New York, Association for Computing Machinery (ACM), 2001, pp 1–27
25. Noble WS: What is a support vector machine? *Nature Biotechnol* 2006, 24:1565–7
26. Rigas B, Wong PT: Human colon adenocarcinoma cell lines display infrared spectroscopic features of malignant colon tissues. *Cancer Res* 1992, 52:84–88
27. Andrus PG: Cancer monitoring by FTIR spectroscopy. *Technol Cancer Res Treat* 2006, 5:157–167
28. Nafie AA, Essam EA: In vivo spectral analysis of bladder cancer using Fourier transform infrared. spectroscopy, a comparative study. *Aust J Basic Appl Sci* 2011, 5:1734–1739
29. Benedetti E, Bramanti E, Papineschi F, Rossi I: Determination of the relative amount of nucleic acids and proteins in leukemic and normal lymphocytes by means of FT-IR microspectroscopy. *Appl Spectrosc* 1997, 51:792–797
30. Mordechai S, Mordechai J, Ramesh J, Levi C, Huleihel M, Erukhimovitch V, Moser A, Kapelushnik J: Application of FT-IR microspectroscopy for the follow-up of childhood leukemia chemotherapy. *Proc SPIE Subsurface Surface Sensing Technol Appl III* 2001, 4491:243–250
31. Ahnen D: Abnormal DNA content as a biomarker of large bowel cancer risk and prognosis. *J Cell Biochem Suppl* 1992, 16:143–150
32. Parker FS: *Applications of Infrared Spectroscopy in Biochemistry, Biology and Medicine*. New York, NY: Plenum Press, 1971 pp 477–491
33. Yano K, Ohoshimo S, Shimizu Y, Moriguchi T, Katayama H: Evaluation of glycogen level in human lung carcinoma tissue by an infrared spectroscopic method. *Cancer Letters* 1996, 110:29–34
34. Colagar AH, Chaichi MJ, Khadjvand T: Fourier transform infrared microspectroscopy as a diagnostic tool for distinguishing between normal and malignant human gastric tissue. *J Biosci* 2011, 36:669–77
35. Piva JC, Silva JLR, Raniero L, Martin AA, Bohr HG, Jalkanen KJ: Overview of the use of theory to understand infrared and Raman spectra and images of biomolecules: colorectal cancer as an example. *Theoretical Chem Accounts* 2011, 130:1261–1273
36. Xie Y, Liu Q, He F, Guo C, Wang C, Zhao P: Diagnosis of colon cancer with Fourier transform infrared spectroscopy on the malignant colon tissue samples. *Chinese Med J* 2011, 124:2517–2521
37. Cheng CG, Tian TM, Jin WY: A study on the early detection of colon cancer using the methods of wavelet feature extraction and SVM classifications of FTIR. *Spectroscopy* 2008, 22:397–404
38. Han LY, Cui J, Lin HH, Ji ZL, Cao ZW, Li YS, Chen YZ: Recent progresses in the application of machine learning approach for predicting protein functional class independent of sequence similarity. *Proteomics* 2006, 6:4023–4037