Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1758 (2006) 830–845
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.elsevier.com/locate/bbamem

Review

High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data

Rohit Bhargava [a], Daniel C. Fernandez [b], Stephen M. Hewitt [c], Ira W. Levin [b,*]

[a] *Department of Bioengineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[b] *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD 20892, USA*
[c] *Tissue Array Research Program, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA*

Received 10 February 2006; accepted 9 May 2006
Available online 17 May 2006

Abstract

Vibrational spectroscopy allows a visualization of tissue constituents based on intrinsic chemical composition and provides a potential route to obtaining diagnostic markers of diseases. Characterizations utilizing infrared vibrational spectroscopy, in particular, are conventionally low throughput in data acquisition, generally lacking in spatial resolution with the resulting data requiring intensive numerical computations to extract information. These factors impair the ability of infrared spectroscopic measurements to represent accurately the spatial heterogeneity in tissue, to incorporate robustly the diversity introduced by patient cohorts or preparative artifacts and to validate developed protocols in large population studies. In this manuscript, we demonstrate a combination of Fourier transform infrared (FTIR) spectroscopic imaging, tissue microarrays (TMAs) and fast numerical analysis as a paradigm for the rapid analysis, development and validation of high throughput spectroscopic characterization protocols. We provide an extended description of the data treatment algorithm and a discussion of various factors that may influence decision-making using this approach. Finally, a number of prostate tissue biopsies, arranged in an array modality, are employed to examine the efficacy of this approach in histologic recognition of epithelial cell polarization in patients displaying a variety of normal, malignant and hyperplastic conditions. An index of epithelial cell polarization, derived from a combined spectral and morphological analysis, is determined to be a potentially useful diagnostic marker.

Keywords: Fourier transform infrared (FTIR) spectroscopy; Imaging; Biophotonics; Prostate; Tissue microarray; Bayesian statistics; Likelihood classification; Discriminant; Cancer; Histology; Pathology; ROC

Contents

* Corresponding author.
E-mail address: iwl@helix.nih.gov (I.W. Levin).

0005-2736/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2006.05.007

1. Introduction

Morphologic examination of tissue specimens using light microscopy is the method of choice for the definitive detection and grading of most human cancers. The information content of microscopy images, however, is limited to a spatial variation in optical properties and requires extensive human observations to recognize both the constitutive histologic entities and the pathologic state. Correlations of morphologic and biochemical tissue differences in conjunction with assessments based on chemically-specific spectroscopic techniques have been suggested as routes to augment studies in pathogenesis [1]. Non-destructive vibrational, infrared and Raman, spectroscopic techniques [2] are especially attractive as they afford an abundance of chemical-specific information without the requirement of extrinsic contrast agents. Significant numbers of studies have involved the measurement of average infrared (IR) spectra of intact tissues [3], exfoliated cells [4], carefully extracted cellular constituents [5] and bodily fluids [6]. Though spatially-resolved infrared micro-spectroscopic characterization of intact biological tissue has also been a topic of continual interest for over 50 years, the past decade has yielded the most systematic studies of various contributory factors in the microspectroscopic analyses of tissue [7]. Since tissue is microscopically heterogeneous, microspectroscopy is the sampling mode most capable of representing accurately tissue complexity in histopathologic determinations. The development of clinical protocols for the routine examination of tissue histology or of localized tumors using IR microspectroscopic methods, however, has not been possible due to several related issues.

First, microspectroscopic instrumentation has conventionally employed point detection and square apertures, measuring several tens of micrometers in size per edge, to restrict radiation incident upon the sample. The coarse spatial specificity of conventional point detection techniques is typically larger than characteristic dimensions of most cells. Consequently, the measured spectra consist of an unpredictable mixture of spectral contributions from neighboring ensembles of other, neighboring cell types. Increasing spatial specificity by decreasing aperture sizes to typical cell size dimensions (~5–10 μm) requires substantial increases in data acquisition times to offset the decreased throughput [8]. These long measurement times, in turn, limit both the number of spectral observations in a particular sample and the number of samples that can be analyzed in a reasonable time interval. This lack of throughput in the number of spectra and samples that can be analyzed has limited large population validations of spectroscopic studies that report promising initial results and has led to numerous different claims of biospectroscopic markers for a variety of tissue types and diseases. Careful validation of initial results in larger studies and the meticulous delineation of confounding variables [9] is critical to the establishment of robust protocols and may often reveal the failure of biomarkers based on a small number of observations. While limiting the statistical validity of obtained results, the lack of testing with large cohorts has also led to significant confusion regarding the potential of spectroscopy for biomedical decision-making. Specifically, the roles of patient to patient variation, variation due to sampling methodology [10] and the role of intra-patient heterogeneity arising from cellular turnover [11] has led to questions of whether vibrational spectroscopy can, in fact, be employed for histopathologic determinations.

The analysis of spectroscopic data for diagnostic decision-making has also been an area of considerable activity. While many approaches employed in the chemometrics of well-defined systems can be adapted, consensus is emerging that the optimal data analysis algorithm would employ spectral data as input and result in a determination that can be represented in terms of quantitative numbers or images interpretable by personnel across several disciplines. Extensive spectral processing is *de rigueur;* baseline corrections, normalization and derivative or orthogonal factor calculations for every spectrum are typically required. In this manuscript, we present a detailed description of how impediments to the vibrational biospectroscopic imaging of tissues and cells can be addressed through rigorous, quantitative tests by an integrated platform of multichannel, multiplexed spectroscopy, high throughput microarray sampling and fast numerical analysis incorporating explicit user control. The application of this approach to prostate histopathology [12] has been recently reported and we explain here previously unreported aspects of the approach. We further describe the use of a spatial index, formed using spectral data, of specific cell types in prostate tissue to examine pathologic conditions.

2. An integrated, high-throughput platform

Fourier transform infrared (FTIR) spectroscopic imaging, tissue microarrays and pattern recognition technologies form the core of the described approach to cell and tissue analysis. We briefly review the technology that we have employed and describe novel features that allow the work reported in this manuscript.

2.1. FTIR spectroscopic imaging

Recently developed Fourier transform infrared (FTIR) imaging technologies [13] are beginning to provide the capability to measure routinely large numbers of spectra with localization of spectral signals from regions small enough to be useful for biomedical analyses [14]. When a previous study compared the performance of imaging and point mapping instruments [15], mapping instrumentation provided higher quality spectra but lower spatial resolution and throughput. Here, we compare newer instrumentation in which both modalities have been implemented using the same optical train through the interferometer and microscope. While a wide variety of tissues have been examined using mapping techniques, data consist typically of few spectral observations from each sample with studies being restricted to small numbers of samples due to large experimental times, cost constraints and technological capabilities. For example, Fig. 1 demonstrates mapping and imaging results for a small prostate tissue sample. In conventional histopathologic examinations, the sample is stained with dyes as shown in Fig. 1A to visualize the distribution of various cell types. The architecture of the prostate gland is complex: three-dimensional ducts appear in stained pathologic slides as empty regions (lumen) or containing secretory, mineralizing material (stones). Lumen are lined by a thin layer (5–20 μm wide) of dark-stained epithelial cells. The ductal structure within the prostate is supported by structural elements of intervening cell types that are collectively termed "stroma". Epithelial cells, in which over 95% of prostatic carcinomas arise, are of particular interest and must be segmented from stroma before a spectroscopic measurement, reflecting biochemical changes indicative of malignancy, can be determined.

We first employ point detection *mapping* approach to acquire a single spectrum from each point under conditions of moderately high spatial resolution. Fig. 1B through E show both the absorbance maps and representative spectra for different aperture sizes (detailed in Experimental). The same spatial region, examined in an *imaging* approach (6.25 μm per pixel) is shown in Fig. 1F. The imaging representation demonstrates a closer correspondence to the visible microscopy image compared to the point detection maps obtained using larger aperture sizes and provides improved contrast compared to maps obtained using smaller aperture sizes. In examining epithelial cells in prostatic tissue, significant spectral contributions from surrounding stroma can occur for the larger aperture sizes [16]. For the smaller sizes (<10–15 μm) and imaging results, there are a large number of pixels that do not have stromal contributions. Hence, this level of spatial resolution is required to segment definitively the epithelial pixels. While the spatial fidelity of the maps increases with decreasing aperture size, spatial speckle noise arising from spectral noise in individual pixels degrades visualization and will degrade the quality of histopathologic segmentation based on the data.

The typical spectra, obtained from the pixel matching closest to the arrow indicating epithelium in Fig. 1A, are shown in

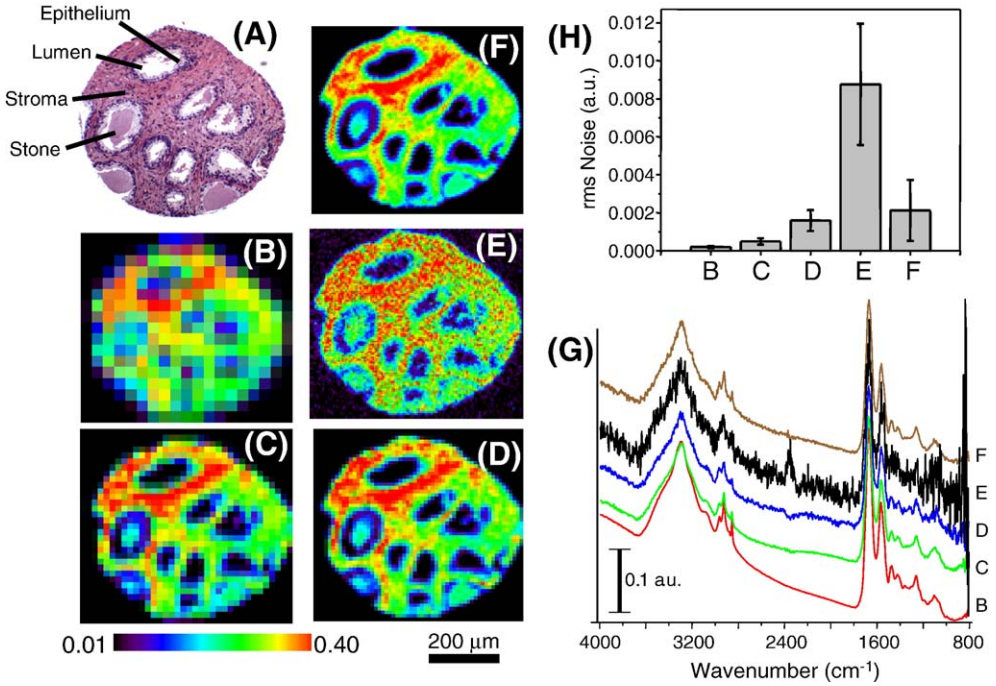

Fig. 1. (A) Hematoxylin and Eosin (H&E) stained image showing several constituents of prostate tissue. The structural elements of the prostate can be clearly observed with darkly-stained epithelial cells lining the lumen, which may contain mineralizing deposits (stones). The stroma envelopes epithelial foci. Spectroscopic mapping with square apertures of edge sizes (B) 40, (C) 20, (D) 12 and (E) 6 micrometer are compared to an image obtained using a 16 element array detector (F). The corresponding chemical image of an unstained microarray made using a subsequent section showing the distribution of the absorbance of the amide I vibrational mode. (G) Spectra, with letters corresponding to the image labels, clearly demonstrate the increase in noise as the aperture size decreases. (H) The rms spectral noise in a non-absorbing region (2100–2000 cm^{-1}) of the spectrum demonstrates that the spectral quality of a single pixel from imaging data is better than the data from an aperture of the same size.

Fig. 1G and demonstrate the increase in spectral noise when the aperture size is decreased. In particular, the spectra acquired using the smallest aperture (6 μm×6 μm) do not allow identification of spectral features in the fingerprint region. Until the employment of a synchrotron as a source [17], these data trends were typical and routine, sensitive analyses in the "fingerprint" region (1900–900 cm^{-1}) were not feasible for aperture sizes smaller than ~15 μm. In this milieu, the contrast between data from point mapping experiments and the imaging data of a small pixel size is especially striking. It must be noted that a given aperture or optical pixel size does not imply that the spatial resolution is determined solely by these geometrical constraints. Since the achieved spatial resolution may be limited by the wavelength, specific optics of the microscope (magnification and point spread function) [18], the aperture (or array pixel size) [19] and the sample (optical effects) [20], reported aperture sizes or pixel sizes for mapping and imaging approaches, respectively, are actually magnification-based and only imply the best resolution that may be achieved under ideal conditions. A comparison of the signal to noise ratio expression reveals, however, that the spectral fidelity advantage of the imaging approach arises from the reduced detector size and effective optimal utilization of interferometer throughput [21,22].

Since our mapping and imaging data were acquired on the same microscope platform, sample and wavelength bandpass data, we can straightforwardly compare data resulting from measurement with an array detector and a point detector mounted on the same dewar. A plot of the rms spectral noise, shown in Fig. 1H, demonstrates the decrease in spectral fidelity quantified in terms of the root-mean-square (rms) noise in a non-absorbing region of the absorption spectra. When normalizing for data acquisition time [23], the typical imaging data quality is approximately equal to that of an aperture size of 20 μm on a side, a typical size employed in many recent studies, while the spatial area is smaller by a factor of ~10 (square of the pixel size ratio). To achieve the SNR of the imaging data with an equivalent sized mapping aperture, ~100 spectral co-additions would be required in accordance with the trading rules of FTIR spectroscopy [23]. Radiation throughput loss, further, will arise from both a reduction in aperture size and from diffraction and when the apertures are reduced to a size of 6 μm×6 μm, close to the wavelength at 2000 cm^{-1}, from 20 μm×20 μm. While the multiplex advantages of interferometry are present in both mapping and imaging approaches, the combined ~100-fold data acquisition time advantage of multichannel and aperture-less detection represents a formidable capability enhancement compared to conventional data acquisition modes. By comparing the typical SNR characteristics, more than 20 years would be required to acquire the array data we report with an imaging approach in this manuscript if a mapping approach were to be employed to attain the same spatial and spectral fidelity. Hence, high throughput infrared imaging microscopy is not merely an extension of previously developed microscopy

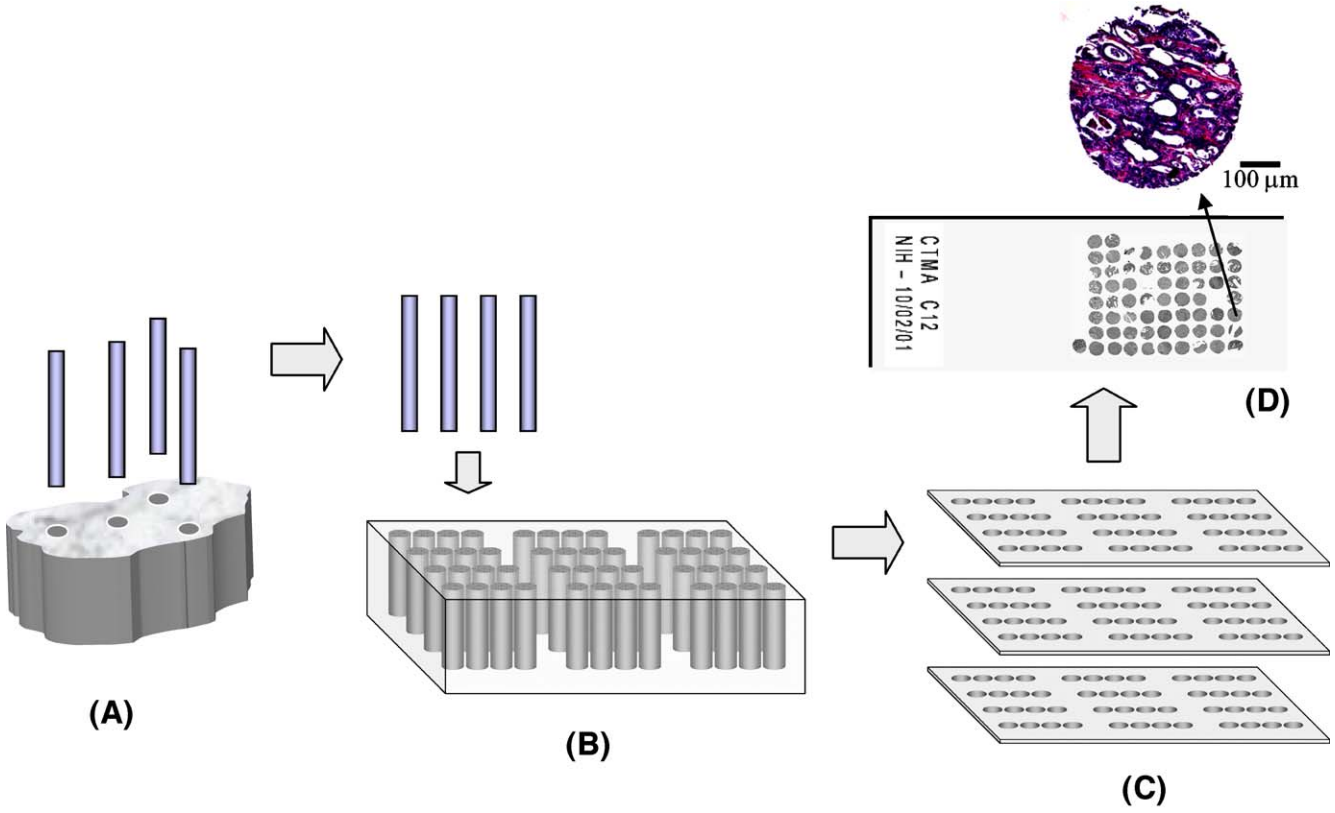

Fig. 2. The process of manufacturing tissue microarrays consists of (A) extracting cylindrical tissue cores from a paraffin tissue block, leaving the specimen largely intact and useful for further analyses, (B) arranging multiple cores in a paraffin recipient block followed by (C) sectioning into thin sections on rigid substrates. (D) A schematic of a typical tissue microarray showing target dimensions of samples reported in this study.

methods, but now provides the capability to record data at sufficiently high spatial selectivity from samples sizes and numbers that are clinically relevant.

2.2. Tissue microarrays

While infrared spectroscopic imaging approaches permit rapid recording of data from large numbers of tissue samples, examinations of large patient populations has other impediments. Since typical experiments are conducted in the transmission configuration [24], the acquisition, storage and processing of hundreds to thousands of infrared transparent salt crystal substrates is expensive. At the same time, the examination of progressively larger number of spectra from a single patient provides a diminishing return in the amount of useful information. As a first estimate, a small sample size, usually one that corresponds to a reasonably large field of view of infrared imaging instrumentation, should suffice. We arbitrarily hypothesized that 5000–10,000 spectra obtained from each patient will provide sufficient numbers of observations for statistical validations. This hypothesis, which can only be examined after histologic classification, will be examined later in this manuscript. At the time of the start of this study, the de facto standards in wide-field imaging were fields of view of 500 μm × 500 μm (64 × 64 array detector) and 750 μm × 750 μm (256 × 256 element array detector) [25]. Hence, we selected a sample size with characteristic dimensions of ~500 μm.

Tissue microarrays [26] (TMAs) consist of multiple tissue samples of uniform dimensions placed on a single substrate. This arrangement facilitates consistent and convenient processing for all samples after the array is constructed. The formation of a tissue microarray is illustrated in Fig. 2. Briefly, small diameter, cylindrical cores biopsies are extracted from specific areas of large tissue blocks. Several cores are then place in an array grid pattern in a paraffin recipient block. The block is subsequently sectioned with a microtome to provide many array sections, each being deposited onto a specific substrate. The use of a precise sample positioning microscope stage then provides a basis for automation of array construction, image acquisition and correlations between different instruments and sampling techniques. Individual specimens can be acquired from hundreds of different donors and may contain multiple samples from the same donor tissue.

We propose that TMAs be employed not only as high throughput molecular screening tools [27], using, for example, fluorescence techniques [28], but also to address some of the common challenges in the vibrational spectroscopic analyses of tissue. Variations observed in spectroscopic measurements affect reproducibility and limit the confidence in results when multiple samples are examined at different times, prepared under varying conditions or analyzed in dissimilar environments. Sample variations have been examined in detail for studies on skin tissue, in which a number of minor [29] to significant differences [30] were observed between patients, as well as in samples from the same individual. Spectral variations dependent on the state of hydration, [29] on chronologically or photoaged skin constituents [31], on sampling area and diurnal and daily variations [32] were also reported. While a large number of samples can be incorporated into an array from physiologically similar patient groups, a larger benefit for spectroscopic studies lies in including all possible patient cohorts. We hypothesize that robust protocols may be formulated by employing a large patient and sample set for building a prediction model. Instead of devising sample handling approaches, the prediction protocol then robustly incorporates these inevitable patient variations. TMAs are an ideal platform for both discovery of spectral biomarkers and validating vibrational spectroscopic studies in this paradigm. Questions regarding the variability in patient spectral response can now be decoupled from those of sample processing history and examined independently. Similarly, a single array can be subject to various processing treatments and the effects examined simultaneously for a large number of patients.

Several potential pitfalls may be encountered in designing vibrational spectroscopy protocols based on data from TMAs. As with all sampling modalities, TMA cores must be representative of the tissue from which they are extracted. A number of studies have concluded that TMA technology is capable of representing overall tissue malignancy [33], while others have suggested a minimum number of cores to assure concordance [34]. The spectral basis set, hence, must be constructed using multiple samples from a number of patients and must be extensively validated for a significantly larger population. Clearly, it is impossible to estimate the size and numbers of samples required for any given problem *a priori*; the refinement process is inevitably iterative. A second major concern is that array-specific spectral impurities may lead to array-to-array variations and lead to sub-optimal prediction models. We have adopted a two-pronged approach toward quality control for individual arrays: first, an adjacent section of our unstained TMA slide is stained and examined by a pathologist. Alternately, the spectroscopically imaged slide is later stained. The goal is to verify results for every core on the array for a small data set and subsequently examine the performance of automated algorithms, comparing results to the manually examined data set. Second, we have written software to allow quantitative measures of differences within arrays and across different arrays for extensive verification. In our opinion, a combination of both visualization, statistical analysis and rigorous, critical human examination of the results at every stage is not only possible, but is essential, when employing TMAs in developing spectroscopy protocols.

2.3. Methodology for automated tissue segmentation

While a combination of imaging technologies and tissue microarrays provides large amounts of data, the extraction of meaningful information requires automated numerical analysis techniques. It is now widely accepted that no single spectral marker, as, for example, a “cancer peak”, can be reliably employed as an indicator of pathology or histology. Hence, in using multivariate techniques for tissue classification, either unsupervised or supervised methods are employed. Clustering techniques are the most prominent unsupervised methods and

are employed generally to determine distinct groups in a data set by dividing the data into homogeneous sub-groups. In particular, hierarchical clustering methods [35], which provide classifications at various levels of similarity, have been demonstrated to be useful in segmenting tissue. Even though cluster analysis is an objective, exploratory data analysis procedure, it requires significant human intervention in analyzing hierarchical organizations. Supervised classification methods employ spectra obtained from specific cell types to form a training data set used in identifying unique patterns through a variety of possible approaches [36,37]. Although the training of the classification model is supervised, the transformation of data into results requires no user input. Pattern recognition techniques are employed to classify given spectral observations according to empirical classification rules developed from a set of spectral observations known to belong to specific sub-groups.

The data analysis paradigm in a supervised classification approach is straightforward. The first step is the selection of a model to analyze data. The second step consists of data preprocessing to optimize results and computations for the chosen model. The third step is the actual classification procedure, while the fourth step may consist of post-processing operations to isolate specific cell types or to enhance the display. The salient features of our approach, comparisons with current best practices and their potential advantages and limitations are described for each of the above four stages.

2.3.1. Model selection

Model selection is the single most important activity in any supervised classification procedure as the entire classification approach is then optimized to yield consistent results. For spectroscopic classification in histopathology, the model must satisfy three major criteria: First, it must reproduce the essential features of the image pertinent to conventional histology. Second, the model must be robust and validated in extensive, double-blinded studies. Third, implementation must be computationally efficient to be clinically applicable. The model for our spectroscopic image analysis of tissues is morphology-based in which we determine functionally distinct cell types and extra-cellular components that have characteristic physical dimensions accessible by our system's spatial resolution. Fulfilling the first two criteria, our task is somewhat simpler than a general chemometric recognition problem as we are guided by the end goal of identifying diagnostic histologic and pathologic classes. In contrast to a previous histologic application where we sought to comprehensively identify all cell types and did not consider specific cellular features, we are additionally interested here in the polarized epithelial cells as potential markers of malignancy. Hence, we have implemented a two-step classification approach. The first step consists of the segmentation of tissue into constituent cell types, as before. The second step now consists of isolating epithelial cells and of examining pixels belonging to this subtype for being predominantly nuclear or cytoplasmic, an additional two classes (*vide infra*). Hence, the model selection and overall approach that is presented is a general approach that is extendable to numerous histopathologic analyses that probe the spectrum in greater detail than a previous step.

Other histopathologic models have been employed [38] for vibrational spectroscopy in which spectra are assumed to be a mixture of pure spectra of morphologic subtypes. Here, we assume that every pixel contains only one histologic type. A complication in model selection is that the histologic subtypes common in clinical practice may not match a spectroscopic representation. In Fig. 3, for example, histopathologic practice identifies the round object as a stone (top row). Spectroscopically, however, the images demonstrate that the one histologic

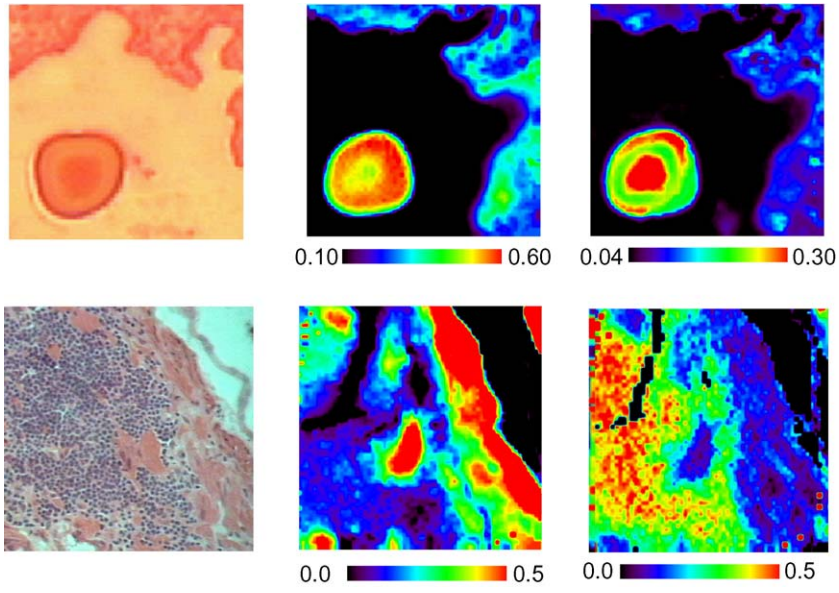

Fig. 3. H&E stained image of a stone (top, left) and corresponding infrared spectral images using different spectral measures: amide II (middle, top) and asymmetric CH stretching (right, top) absorption. H&E stained image (bottom, left), intensity of the Amide II vibrational mode (bottom, middle) and the ratio of absorption of two vibrational modes (1245 cm^{-1}/1645 cm^{-1}) illustrates the differences between muscle and lymphocytes.

sub-type demonstrates at least two distinct types of spectral patterns. Hence, while hierarchical prediction of the clinically useful class may be complicated, our supervised method necessarily defines all pixels as "stone" and seeks a mathematical treatment of spectra allowing for those pixels to be recognized as stone. The same approach will help mitigate the effects of inter-patient or of sample handling variations. In some other cases, however, correlation between spectral metrics and pathology is straightforward. For example, complementary spectral indices depicted in the bottom row of Fig. 3 allows the visualization of protein-rich muscle bands and lymphocytes.

The goal of the study we report here was to segment epithelium from other cell types and to examine whether polarization in epithelial cells is a useful marker for detecting malignancy. Polarization is used here in the histologic, rather than optical, sense. During usual development, epithelial cells develop into a basal zone containing the nucleus and an apical zone containing the cytoplasm. The basal zone is adjacent to the stroma, while the apical zone opens into the lumen. To determine cell polarization, we first have to segment the epithelial cells within the complex prostate tissue. The results for this procedure have previously been reported [12], and here we describe details of the process that were not previously described. The same process is later employed for determining epithelial cell polarization.

We define the model such that **C** is the universal set of classes such that

$$\mathbf{C} = [c_1, c_2, \ldots, c_{n_c}], n_c = 10 \quad (1)$$

where n_c is the number of classes described as [Epithelium, Fibrous Stroma, Mixed Stroma, Smooth Muscle, Stone, Blood, Lymphocytes, Nerve and Ganglion].

2.3.2. Data pre-processing and spectral marker selection

Each pixel in an imaging data set yields a spectrum. All spectral regions, however, do not contain pertinent information. For example, the measurement conditions leading to water vapor or atmospheric carbon dioxide absorbance clearly do not contribute to an understanding of tissue structure. Further, several absorbance features are correlated and inclusion may be marginally useful at best. Thus, there is a need to reduce the total dimensionality of spectra to an intrinsically smaller set that describes the essential features of the spectral profile in a compact and useful manner. Similarly, all pixels in an image do not contain information. In particular, we seek to eliminate from further consideration pixels whose spectra do not demonstrate the presence of tissue, whose signal to noise ratios are unacceptably high and those that demonstrate a wavelength dependent edge-effect distortion [20]. An enormous benefit can be realized in terms of computation time by this initial reduction in both the number of spectral components and spatial elements in an imaging data set.

Vibrational spectra typically consist of several hundred to a few thousand measurements of absorption or scattering. The distribution of intensity of a given vibrational mode, the location of the peak position of each mode and relative intensities from two vibrational modes are all sensitive measures of tissue biochemistry that can be related to model studies. In contrast to other approaches using complex non-linear relationships between spectral features, we employ only measures of these quantities. The use of these simple spectral measures will later allow the correlation of useful spectral features with tissue biochemistry, providing a rational basis for our prediction algorithm.

A typical spectrum gives rise to several tens of measures of different types (absorbance, FWHH or peak position). Consider the spectral profile composed of absorbance measures for a large wavelength range, **U**, defined as

$$\mathbf{U} = [A_1, A_2, \ldots, A_{n_a}], \quad (2)$$

where n_a is the number of spectral resolution elements in a spectrum. **U** is mapped to a smaller set, **S**, defining the dimensionality of **U**, such that

$$\mathbf{S} = [s_1, s_2, \ldots, S_m], \quad (3)$$

where m ($\ll n_s$) is the number of potentially useful spectral measures. We term these measures "spectral metrics" as they are the smallest measurement set of spectral features that allow a representation of the intrinsic dimensionality that is sufficient for classification. The reduction of spectral profiles to this small set of metrics will reduce computation time and allow recognition of related structure in the spectra. While new variables may be constructed from the data in other manner, for example, by employing a principal components analysis, we seek to preserve spectral features to correlate individual spectral contributions to the underlying biochemistry. The utility and need for this data reduction step is well-recognized but there is little consensus on the metrics that may be used to describe the spectra, the method for finding metrics, validation of metrics or their physical interpretation in terms of molecular vibrational modes. Just as a pathologist is employed to identify cell types, a spectroscopist examines average and typical spectra of all classes from numerous patients to deduce **S** manually.

While many studies employ specifically a few narrowly defined spectral regions, large spectral regions, derivative spectra, factors or principal components, we employ a different approach here. Since a large number of features exist in any spectrum, the universal set of potentially useful metrics, $\mathbf{U}_M$, is large. Let the useful subset of metrics be denoted by **M** as

$$\mathbf{M} = [m_1, m_2, \ldots, m_{n_m}], \quad (4)$$

where n_m is the number of potential useful metrics. The reduction of **S** to a set of usable spectral metrics, **M**, requires a selection procedure that depends on the working hypothesis of the classification scheme. While **S** was determined manually, the reduction of **S** to **M** is accomplished objectively and without human input.

2.3.3. Metric Bayesian classification (MBC)

The IR spectral response of biological tissue is remarkably similar for most cell types; the differences arise from the pattern of absorbance rather than from any specific absorption peaks.

The classification task, hence, is to determine the probability that an unknown spectral profile originated from a certain class. This process needs to be applied to every pixel in an imaging data set. We discuss a single metric profile below. The analysis can be extended to the entire imaging data set by repeating the process pixel-by-pixel. Clearly, this approach classifies the image based on the spectral content of each pixel, without regard to its spatial position.

Let $p(c_i|\mathbf{S})$ be the *a posteriori* probability [39] of a metric profile, **S**, belonging to a certain class, c_i. The assignment is to then determine a class, c_j, using a reduced data set, **M**, directly mapped to every spectral profile, **S**, such that

$$p(c_j|\mathbf{M}) > p(c_i|\mathbf{M}), \qquad i = 1, 2, \ldots, n_c, i \neq j. \tag{5}$$

The expression can be shown to result in an optimal condition of Bayes minimum error [40]. The decision rule may also be minimized for error, risk or a custom procedure, for example, by minimizing the maximum cost of misclassification. The application of Eq. (5) is accomplished in a straightforward manner by deriving n_c discriminant functions, g, such that

$$g_k(\mathbf{M}) = \max[g_k(\mathbf{M})] \leftrightarrow p(c_j|\mathbf{M}) > p(c_i|\mathbf{M}). \quad k, i = 1, 2, \ldots, n_c, i \neq j. \tag{6}$$

For example, the general form of the discriminant

$$g_k(\mathbf{M}) = (\mathbf{w}_k)^T \mathbf{M} + w_{0,k}, \qquad 1 \leq k \leq n_c \tag{7}$$

requires an optimization of weights, **w**. While this form of linear discriminant functions is a generic feature of classification, the term linear discriminant analysis (LDA), or Fishers criterion, is a specific decision rule in which the *intra*-class variability is minimized while maximizing *inter*-class separability. LDA-based classification protocols have been applied to IR spectra [41] and maps [42]. A feature of LDA and other discriminant functions is that the decision rule must be explicitly specified and, unlike Bayes' decision rules, does not evolve naturally from the underlying data distributions. While the formulation of explicit discriminant functions allows for flexibility in the choice and assignment of parameters, it is also a complex process to implement and requires a careful, extensive optimization during the classifier training.

We formulate a discriminant function for class k based directly on the Bayesian probability, where Eq. (5) is expanded in terms of measured class conditional probabilities [43],

$$g_k(\mathbf{M}) = \log[p(\mathbf{M}|c_k)] + \log[p(c_k)] - \log[p(\mathbf{M})], \qquad 1 \leq k \leq n_c \tag{8}$$

where $p(\mathbf{M}|c_k)$ can be readily estimated from its true population by employing TMA sampling and extensive accumulation of spectra from investigator identified spatial regions. The probability of occurrence of any given metric pattern is independent of the class and will appear with the same magnitude in all discriminants. The probability of occurrence of a class, $p(c_k)$, cannot be estimated by our approach here due to our investigator-biased selection of tissue spatial regions from which cores were extracted for constructing the micro-arrays. Further, the regions identified by pathologist review as basis for acquiring calibration spectra are biased. Even if a large population of entire prostate tissue sections were imaged, a prohibitively expensive task, it is debatable if classification accuracy would be improved for small core biopsies of individual patients by this determination of ensemble average class probabilities. The inclusion of class probabilities may even be misleading for cell types such as lymphocytes and blood, whose presence in biopsy material is highly unpredictable.

The typical distribution of classes in individual biopsies may range from a fraction to several tens of percent. Hence, the term $\log[p(c_k)]$ ranges from ~ -2.5 to ~ -0.25. It must be noted that the term is independent of the metric vector, **M**, length whereas the conditional probability term, $p(M|c_i)$, contributes an increasing magnitude with the length of the metric vector. For a vector size of $\sim$20 and typical probability values of $\sim$0.2–0.8, the contribution of the vector dependent term is significantly greater than the class probability and does not affect segmentation. It is also obvious from this examination why a univariate analysis may not provide accurate results in a multiclass segmentation procedure. A finer point in this discussion is that the magnitude of the discriminant function for any class is not important in itself but is significant only relative to the magnitudes for all classes. Hence, the reduction of the discriminant magnitude for any class does not necessarily imply an increase in misclassification. Neglecting this term in the calculation may not greatly affect results and can be countered by other means (*vide infra,* classification tuning). Since the logarithms of all terms are negative, it can be instantly deduced that by neglecting the class probability, the more prevalent classes are penalized while the infrequently occurring classes are emphasized. Hence, a systematic error of over-assigning less prevalent classes may be possible.

While sharing the underlying philosophy of linear discriminant functions, our method differs in many respects and must not be confused with the LDA-based approach. Our discriminant function derives directly from the natural abundance of spectral metric density and requires no explicit calculations or fitting to the weights. We must emphasize that this is possible by the large numbers of spectral sampling facilitated by TMAs. A drawback is that the weights are overly dictated by the underlying distributions of components and requires careful and extensive sampling to evolve the decision rules. With this backdrop, the use of TMAs is even more critical to the success of our decision rules. We do not recommend this approach as a general methodology when small data sets are analyzed. The choice of a particular form of the discriminant is not critical and no unique function exists for a given problem. Our approach may also be likened to a Gaussian maximum likelihood (GML) approach, which we have employed on the same data sets and have found useful. Unlike the GML, we do not assume the data distribution to be normal or assume equality of covariance matrices. Similarly, the class membership can be likened to a regression problem and be formulated as such. Yet another approach may be a factor based method. Regression methods, however, require extensive computations. Additionally, the structure in the data observed in factor methods may be complicated to predict without an associated class identification method. In an imaging data set consisting of several tens of thousands to a million

spectra, we cannot afford extensive computations. Further, in spite of the success of conventional numerical analyses, their typical processing times (~1–10 s/spectra) cannot be tolerated in an imaging modality where the data analysis times may far exceed the data acquisition time. Our method (*vide infra*) employs simple dimensionality reduction and is based on lookup tables, providing a computationally efficient route.

At this point, it is instructive to re-examine the reduction of spectral features, **S**, to the metric set, **M**. Our analysis depends not on the ability of a metric to distinguish *all* classes but on its ability to distinguish *at least* two classes. Hence, it is necessary and sufficient to examine pairwise error in classification assignable to every candidate metric. Referring again to the Bayes rule of minimum error, as described by Eq. (5), for a two-class system

$$\begin{aligned} p(c_1|m_1) > p(c_2|m_1) &\rightarrow p(m_i|c_1)p(c_1) \\ &> p(m_1|c_2)p(c_2). \qquad i = 1, 2, \ldots, n_{\mathrm{m}} \end{aligned} \tag{9}$$

A simple pairwise probability discriminant may be constructed such that

$$p_{\mathrm{ratio},i}(m_{\mathrm{k}}) = \frac{p(m_{\mathrm{k}}|c_i)p(c_i)}{p(m_{\mathrm{k}}|c_i)p(c_i) + p(m_{\mathrm{k}}|c_j)p(c_j)}. \qquad i \neq j; i, j = 1, 2, \ldots, n_m \tag{10}$$

For binary values, i.e. $p_{\mathrm{ratio},i}$ is 0 or 1, m_{k} provides error free segmentation while a value of 0.5 provides segmentation no better than due to chance alone [44]. This concept is illustrated in Fig. 4 where the probabilities and two pairwise probability discriminant are plotted for three classes. Clearly, metric 1 is significantly better for distinguishing epithelium from either of the two stroma classes than it is for distinguishing the two stroma classes. Where the pairwise discriminant differs from 0 or 1, there is an error in class assignment. For example, at a metric value of 0.03 effectively indicates that the spectrum did not arise from epithelial cells but, in assigning the spectrum to myxoid stroma there is a non-zero probability that it actually originated from a fibrous stroma pixel. A pairwise classification error is useful in quantifying differences between classes. The magnitude of a classification error for misclassifying a metric profile to class j when it actually belongs to class i is given by

$$\varepsilon_{\mathrm{p}} = \frac{1}{2}\left[1 - \int |p(c_i|\mathbf{M}) - p(c_j|\mathbf{M})| p(\mathbf{M}) d\mathbf{M}\right]. \tag{11}$$

The integral, called the Kolmogorov variational distance, is evaluated for every metric and employed to deduce the order of importance for metrics in classifying data. While the usual goal is to minimize the error of classification, we calculate a pairwise error to first compare all classes as pairs. For example, the shaded box in Fig. 4 indicates a region where the error arises from fibrous stroma being labeled a myxoid type while outside the box, the error is due to myxoid stroma being labeled as fibrous. The sum of these explicitly calculated errors is sought to be minimized. The spectral feature set is ranked on the sum of errors for every class and the candidate metric set is reduced to ones that demonstrate appreciable segmentation capability. It is noteworthy that the cost of errors can be assigned at this stage. For example, the cost of labeling a fibrous stroma cell as another type of stroma is not as significant as labeling it as epithelium when the goal of the segmentation is to extract epithelia for cancer diagnostics. The ordering will then be based on a cost – rather than error – minimization model and may well result in a distinct set of metrics. Hence, the possibility exists in our method of clustering spectra in the manner of hierarchical clustering analyses.

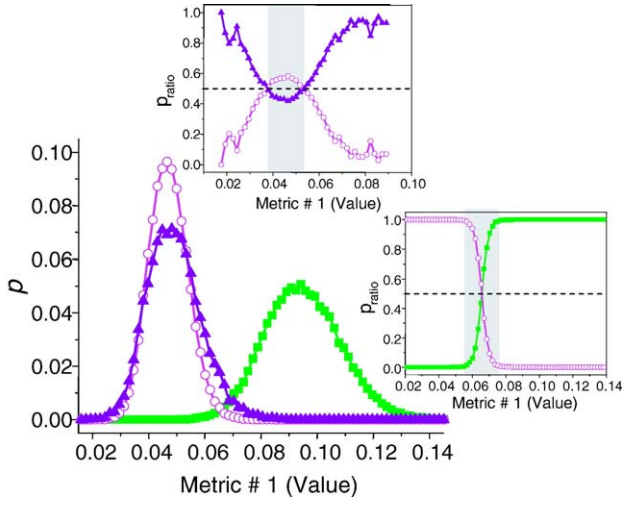

Fig. 4. Probability, $p(m_1/c_i)$, distributions for metric values demonstrate clear differences between epithelium (filled squares) and two classes of stromal cells assigned to fibrous (open circles) and myxoid (filled triangles) types. The insets plot the ratio of the pairwise simple discriminant (PSD). The inset at the top represents a the PSD for correctly partitioning two types of stroma while the lower inset demonstrates a PSD representing an excellent segmentation metric.

2.3.4. Class assignment

The application of the discriminant leads to a "fuzzy" result with each class having a probability. The zeroth order term, $\mathbf{w}_0$, may be adjusted such that

$$\mathbf{g}_i(\mathbf{M}) \begin{cases} > 0 \\ < 0 \end{cases} \rightarrow \mathbf{M} \in \begin{cases} c_j \\ c_j, i \neq j \end{cases} \tag{12}$$

Occasionally, there may exist pixels for whom all $\mathbf{g}_i$ will either be less zero or at least two $\mathbf{g}_i$ will be positive. In such cases, an additional rule may be adopted where the distance to the decision hyperplane, d, may be considered such that

$$d_{\mathrm{k}} = \frac{|g_{\mathrm{k}}(\mathbf{M})|}{|g_{\mathrm{k}}|_{\max}} \tag{13}$$

$$\mathbf{M} \in \begin{cases} \min(d_i), & g_i < 0, \qquad i = 1, 2, \ldots, n_c \\ \max(d_i), & g_i, \ldots, g_{i+x} > 0, \qquad x \geq 1 \end{cases} \tag{14}$$

Realizing that Eq. (9) may be employed in a number of cases, we may wish to calculate the distance vector, ***d***, rather than

first calculate discriminant functions and then apply checks. Hence, we can combine Eqs. (7) and (9) to produce a working algorithm as

$$d_k = f_{1,k}(M_{1,k}) + f_{2,k}(M_{2,k}) + \ldots + f_{n_m,k_m}(M_{n_m,k}) - T_{h,k}. \quad (15)$$

Where, f_k is an array of functions that depend only on the value of **M**. We have stored f_k as a set of lookup tables for all values of **M**. Hence, upon metric calculation, a simple lookup table is referenced to provide individual f values. T_h is the array of adjustable thresholds that can provide explicit control over the classification. The adjustable parameter array, **T**, may be employed for a generalized error correction procedure by iterative minimization of prediction error or for explicit control over classifier performance. This parameter can be employed to either suppress (higher $T_{h,k}$) or enhance (lower $T_{h,k}$) the classification tendency of specific classes by assigning investigator determined values. It is notable that the value of this parameter provides explicit control over the classification process in that the acceptance of specific pixels in a certain class may be facilitated, enhancing the sensitivity of the classifier to selected cell types. Of course, a change in sensitivity is accompanied by a simultaneous change in specificity. It is also notable that this parameter can be employed to mitigate the effects of unknown $p(c_i)$ with or without user input. Since we have normalized the vector length of the discriminant function, d always ranges from 0 to 1. Hence, a threshold range from 0 to 1 is a convenient bound to describe classifier performance as a function of threshold magnitude.

3. Experimental

Data were acquired using the Spotlight 300 (Perkin Elmer Inc.) infrared imaging spectrometer. The system integrates a detector with dual detection elements, a single element unit and a 16-channel array unit, permitting both imaging and mapping studies using the same optical path, detector and interferometric conditions. A strong apodization function and no under-sampling were employed. Data were acquired in transmission mode and the sample was raster scanned to access desired areas, typically 700 μm × 700 μm for each spot. Each pixel sampled a 6.25 μm × 6.25 μm area at the sample plane. Data were acquired at every second zero crossing of a He–Ne laser integrated in the interferometer and were fast Fourier transformed during data acquisition using strong apodization to yield single beam spectra with a resolution of 4 cm^{-1} (2 cm^{-1} data point interval). The single beam spectra were ratioed against an appropriately acquired background and converted to absorbance. A 1641-point subset of each spectrum, spanning a spectral range of 4000 −720 cm^{-1}, was saved. Acquisition time for a typical tissue array core was approximately 40 min and the average single pixel root mean square noise level was greater than 500:1. The microscope was isolated in a vented Plexiglas housing to enable efficient purging with dry nitrogen gas and to eliminate air currents.

Tissue microarrays were constructed from anonymized, formalin-fixed, paraffin-embedded prostate tissue, obtained by the Cooperative Human Tissue Network for the tissue array research program (TARP) laboratory. A Beecher automated tissue arrayer, ATA-27, equipped with a video overlap system and 0.6 mm needles was employed to construct the arrays. Appropriate institutional review board and National Institutes of Health guidelines for the protection of human subjects were followed. The original donor tissue was archival in nature and embedded according to the standard operating procedures of the donor institutions. The array recipient block was constructed from low-melt Paraplast X-tra paraffin, a medium routinely used for embedding tissue sections. A 5 μm section of the tissue array was deposited on a barium fluoride (BaF2) optical window for FTIR spectroscopic imaging and a serial section was placed on standard microscopy slides. Sections for spectroscopic imaging were deparaffinized by immersing in hexane at 40 °C with continuous stirring for 48 h. Every 3−4 h during the deparaffinization process, the immersion vessel was emptied, rinsed thoroughly with acetone followed by hexane and, after thorough drying, refilled with fresh neat hexane to promote dissolution of paraffin embedded in the tissue. Thorough deparaffinization was assured by monitoring the disappearance of the paraffin band at 1462 cm^{-1} at several sites on the tissue arrays. Tissue array sections for visible microscopy-aided determinations by pathologists were stained with H&E and digitally imaged.

All spectral processing and classifications were carried out using software written in-house using IDL (Research Systems). A typical array contained 2.5 million pixels and required 14 GB of storage. Classification time for entire arrays was 20 min and that for a single array spot was 35 s for a 10 class histologic model.

4. Results and discussion

4.1. Methodology validation

4.1.1. Metric evaluation

Spectra may be reduced to a universal set of descriptive features, **S**, from which the final set of metrics is to be derived. The critical issue of interest in reducing **S** to a metric subset, **M**, is whether the observed distributions of two classes for a metric in **M** can be shown to arise from two distinct population distributions [45]. While the mean and variance analysis of the distributions (*t*- and *F*- test, respectively) are useful distribution difference descriptors for a cursory elimination of potential metrics, there is clearly a different probability for every class when the distributions are different. Further, metrics that are useful in differentiating a pair of classes may not be useful for other pairs. Hence, it is important to determine pair-wise differences in class distributions. We have compared binned data renormalized to unit area under the distribution curve using arbitrary numbers of bins to accurately represent data distribution. This results only in the loss of one degree of freedom, which is appropriately accounted for in the calculations. We have arrived as a set of 93 potential metrics, which are formulated in terms of pairwise ratios of absorbance and characteristic spectral features to account for tissue thickness

variations. The set is further refined based on validation optimization (*vide infra*).

To further reduce data dimensionality and order the metric data set to be employed, the potential metric set is reordered based on the lowest pairwise error and the average pairwise error as shown in Fig. 5A. The error is simply the area overlap of the metric distributions, as shown, for example, in Fig. 4. While metrics may be employed in any order for classification, the metric order in Fig. 5A is employed to calculate the probability of each class for every pixel in the basis data set and the accuracy of the prediction is evaluated. We evaluate classification accuracy using receiver operating characteristic (ROC) curves [46]. The area under the curve of ROC plots represents a particularly powerful measure of the "overall" segmentation capability and the average sensitivity; in our paradigm of explicit classification control, it is a particularly useful quantity. Hence, we evaluate the classification accuracy and the contribution of each potential member of the final metric data set to the increase in the AUC arising from the metric. ROC curves are calculated for a single class at a time, thereby eliminating any bias that may arise from the discriminant function.

The AUC for the ROC of each class is shown in Fig. 5B and C. The former demonstrates the ROC curves for the training data set while the latter is for an independent sample set (validation). AUC values for these cases have been reported previously [12]. Fig. 5C shows the average ROC curve obtained by averaging the performance over all classes. The curves demonstrate that the classification accuracy of the protocol achieves a stable value at around 20 metrics for the calibration data and the accuracy actually decreases for the validation data. The decrease in accuracy is expected as it likely represents array specific bias that is present in the developed protocol. The small magnitude of the decrease for validation data and its agreement with the performance of the calibration data is surprising and can be attributed to the robust sampling of a large population of samples afforded by TMAs. The close agreement of the calibration and validation data set performance is also a consequence of the careful selection, rather than random inclusion of large spectral regions, of spectral metrics.

The sequential forward selection (SFS) process we have employed is a bottom-up search procedure that seeks to examine the effect of adding an extra metric on the performance of the classifier. The primary disadvantage of this process is that a feature once included in the calculation is impossible to eliminate if a metric added later warrants it. Hence, the ROC curves must be employed as a penultimate step in the selection of metrics. We have verified our results by a "leave-one-out" procedure after a small set of metrics is selected. For the classification protocol previously reported [12], for example, a set of 20 metrics was selected at this stage. One metric was left out at a time and the classification process was repeated. As opposed to calculating ROC, optimized classifications were performed at this time (*vide infra*, model tuning). Classification

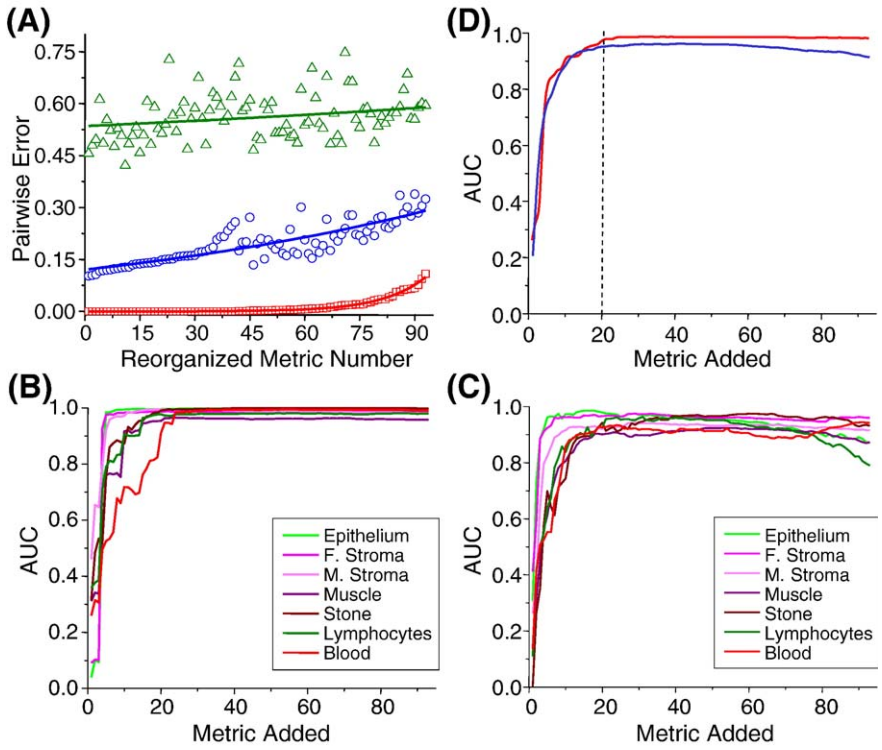

Fig. 5. (A) Pairwise error for metrics arranged in increasing order. The minimum error (red squares), average error (blue circles) and maximum error (green triangles) for all classes indicates that there are a number of metrics useful for segmenting at least two classes. (B) Area under the ROC curve (AUC) is obtained for each class by building successive prediction algorithms based on the number of spectral metrics. The AUC values are calculated for classification on the calibration data. (C) The same exercise as in (B) is repeated for an independent calibration data set. (D) The average AUC values for all classes as a function of the number of metrics. The red curve indicates the values calculated for calibration data while the blue curve denotes the average value of curves for validation data.

accuracy was improved by a few percent when two metrics were left out. Hence, the prediction model was based on 18 metrics.

4.2. Algorithm performance and tuning

Accuracy is the ultimate performance standard in prediction. Classification accuracy, however, has to be considered in the context of the task. For example, if the task is to identify various cell types in prostate tissue, the method above may be followed. If the task is to construct a protocol to isolate epithelium to examine for malignancy markers, then a method is needed to estimate the purity of selected cell populations. Ideally, it would be better to specify cell population purity or detection. A unique opportunity is present in our classification approach for this task. Referring to Eq. (15), the threshold parameter T_h in every expression allows for control over the classification process. The dependence of classification on the value of the threshold is shown in Fig. 6.

The graph in Fig. 6 displays the fraction of epithelial pixels that are correctly classified (correct) and those of other cell types that are classified as epithelium (incorrect). A classified image is shown at the bottom for reference. The image panel to the left demonstrates the pixels that are recognized as epithelium (white) at specific threshold values. As the threshold is decreased, more and more pixels are recognized as belonging to the class. At high thresholds, not all epithelial pixels are recognized (sensitivity) while at lower thresholds impurity in the pixels classified as epithelium increases (specificity). The sensitivity-specificity trade-off involved in the detection is common in many biomedical problems and allows a user either to select the detection capability or the purity of obtained result as dictated by the graph for each class. This is a simplified representation as the actual class in a complete classification involves a trade-off between the threshold corrected values of the discriminant for each class. However, the described curves represent the limits of performance. An impurity level no greater than that demonstrated by the curves can occur. The accuracy and yield curves are the values employed to calculate the ROC curves for histologic accuracy reported above.

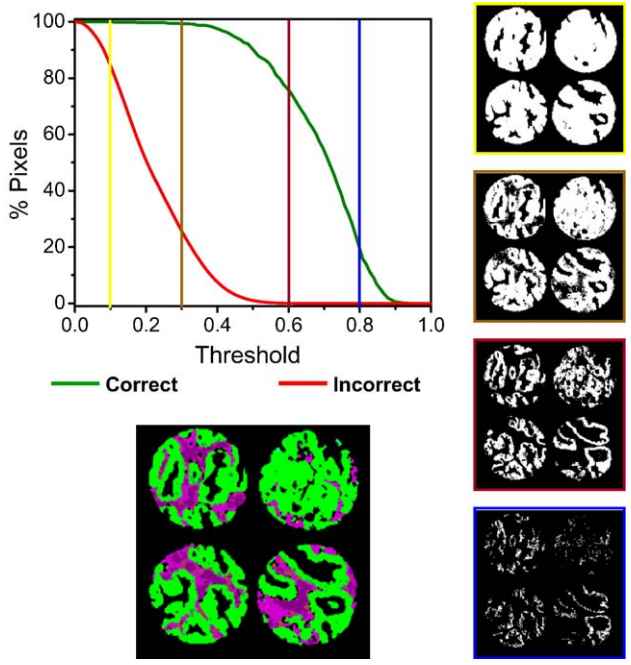

Fig. 6. Explicit control over the classification process is possible using the classification approach described in this manuscript. Consider the classified image (below, left) in which epithelial pixels are denoted by green color. The correct and incorrect fraction of pixels as a function of the adjustable threshold for the epithelial class is (top, left). The classification procedure allows an acceptance of pixels as epithelium as a function of the threshold employed to determine the discriminant function. The panel on the right indicates the pixels that are classified as epithelium (white) and those that are not (black). An increasing number of pixels are classified as the class when the threshold is progressively lowered (bottom to top).

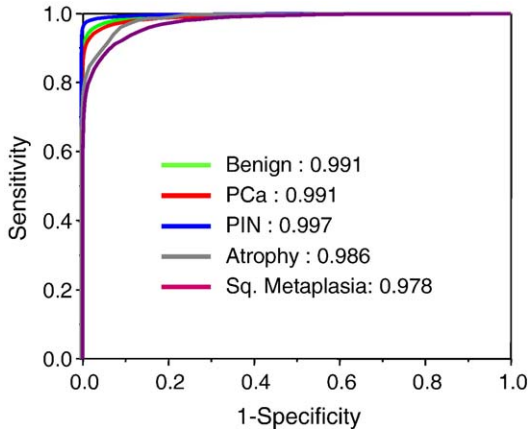

Fig. 7. Comparison of ROC curves for segmentation of epithelium of different pathologies. AUC is indicated on the graph. The number of pixels of benign, carcinoma (PCa), prostatic intraepithelial neoplasia (PIN), atrophy and squamous metaplasia class were 53134, 37342, 15128, 507 and 2566, respectively.

Our classification procedure assumes that the input data, in terms of hand delineated ROI, are absolutely correct. At least four sources of error may be identified. Errors may arise from the misidentification of specific cell types (by a pathologist), marking regions identified on H&E stained sections on corresponding unstained sections (by image analysts), the noise in data measurement and the prediction deficiencies of the classification process. The underlying assumption in labeling is that the human markings are perfect and the confidence in the determination of a specific cell type is absolute. The close supervision of a highly trained pathologist minimizes identification errors. Further, identification errors can be reasonably assumed negligible for prostatic tissue as cell types are morphologically distinct and frequently occur in large ensembles of tens to hundreds of cells. Marking errors in the training set are minimized by identification of large collections of cells away from interfaces, by excluding poorly preserved or artifactual tissue, by rigorous verification with stained sections through software overlaying procedures, by time-lapsed human re-verification and examination of outliers through histograms of every class for every spectral metric. Errors are assumed to be negligible due to the numerous, careful verifications but are cumulative in that both validation and training sets rely on

marked ROI. We do not anticipate these errors to be greater than a fraction of a percent.

The data analysis algorithm also incorporates filters to eliminate pixels with insufficiently high signal to noise ratios, pixels demonstrating significant scattering edge effects [20] and those that do not appear to exhibit absorption features characteristic of biologic materials (for example, dust particles). We did not observe any missing data. Outlier examination is significantly cumbersome for large multivariate data sets. We examined outliers in every class for each metric by extensive, manual examination of scale-expanded histograms. While a small number of pixels (less than 100 or 0.025% pixels in the data set) were eliminated by a subjective examination of absorbance patterns, some seemingly outlier pixels were included. The non-normality may arise from natural variations or from some sampling artifacts. We can only surmise that the contributions of non-natural outliers is small due to our very large number of spectra. Since we followed routine clinical practice in preparing tissue, this impurity level likely arises from coagulated fixing and embedding media, dust and other laboratory contaminants (particulate matter).

4.3. Epithelium classification

An array including several malignant and benign conditions was imaged and tested for accuracy in segmenting epithelium using a previously optimized model [12]. We sought here to examine first if there were any differences in the classification accuracy for epithelium in different pathologic states. ROC curves for epithelium identification are shown in Fig. 7. Notably, these ROC curves are calculated using data that had not been employed for training. Further, training samples consisted of only benign and adenocarcinoma class pixels. Hence, we conclude that the differentiation of epithelium from other prostate cell types is generally independent of pathologic status. The capability arises from the large spectral differences between epithelium and other cell types as well as from the large training set that employed a range of samples with varied pathologic conditions.

While the previous sections demonstrated that epithelium can be separated from other cell types in prostate tissue and that pathologic state of epithelium did not affect the segmentation appreciably, we examine next whether the segmentation approach could be extended to cell level examination of this particular cell type. Since 95% of prostate cancers arise in the epithelium, examination of spectral changes in this cell type is warranted. While one approach could be to examine spectral content of epithelial cells, we attempt to determine here if structural parameters could be obtained from spectral data and whether such parameters could be employed to examine malignancy. We further divided epithelial pixels into two categories: cytoplasm-rich pixels and nucleus-rich pixels. The results of the procedure are shown in Fig. 8 for an array consisting of both malignant and benign samples.

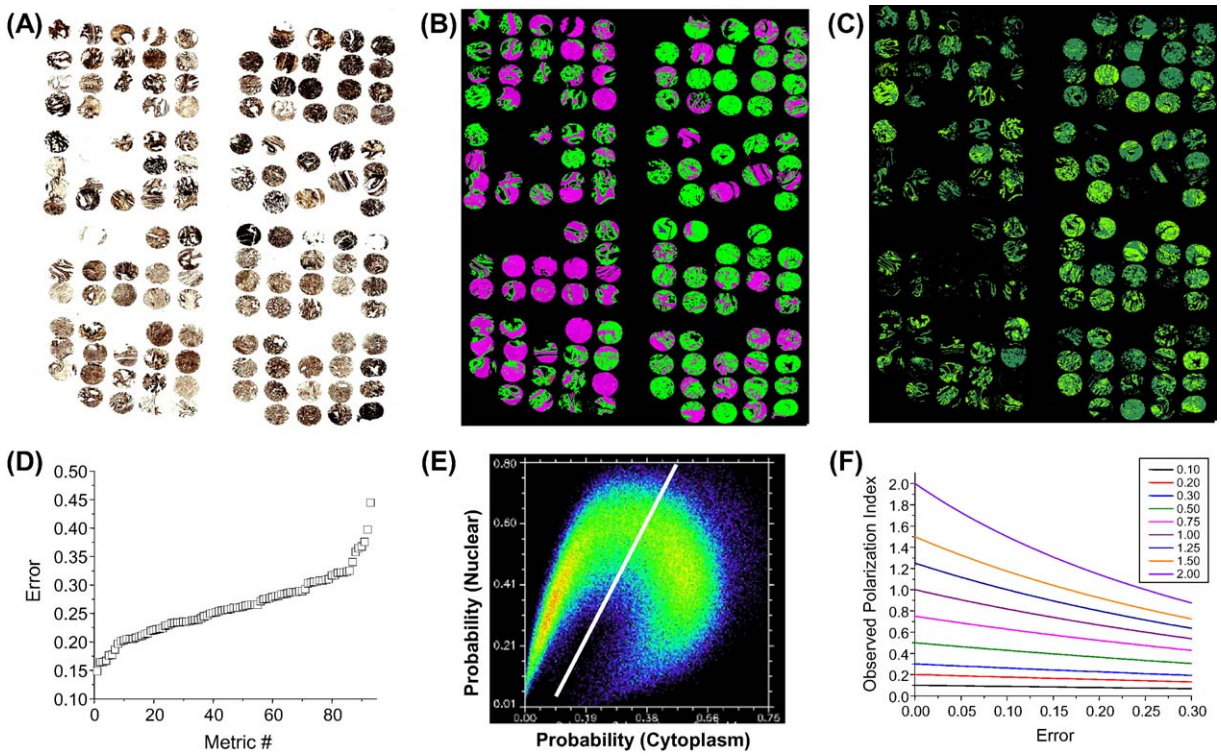

Fig. 8. (A) Visible light microscopy image of a prostate tissue microarray. (B) The array is first segmented into various cell types (green denotes epithelium) using the histologic protocol. (C) Epithelial pixels are further segmented into nuclear-rich (dark green) and cytoplasm-rich (light green) pixels using a second set of 14 metrics. (D) Pairwise error is high for segmenting epithelium due to various factors described in the text. (E) A significant number of pixels, having moderately high probability for both classes (cytoplasm-rich and nuclear-rich), can be misclassified. (F) Misclassification error, as described in the text, decreases the observed polarization index. The decrease is a function of both the error and the true value of the index (box).

Following manual marking, metric organization and classification testing as before, a classification protocol is formulated based on 14 metrics. The area under the ROC curve (0.72), unfortunately, was observed to be smaller than that obtained in segmenting epithelium in prostatic tissue. Similarly the magnitude of the probability of a type in the rule data is not very high. The relative decrease in classification ability is directly related to the high pairwise error in classification (Fig. 8D), which arises from a number of factors. First, the nucleus-rich pixels actually consist of many of the same constituents as the cytoplasmic (apical) region. For the basal region of epithelial cells, there is a larger nuclear contribution to the spectrum. The contribution of nuclear constituents, however, is disproportionately low [47]. Second, the differences in the two classes generally arise in the fingerprint region of the spectrum. Due to the long wavelengths in this spectral region, absorption information is spread over multiple neighboring pixels [48]. Hence, the spatial blurring of information contributes to reducing observable differences. Third, marking unstained sections is significantly more difficult than marking large ensembles of different cell types. Consequently, the marking error is expected to be large. The sum of these factors implies that the probability of a large number of pixels will be high for both classes. A distribution of the relative probability of belonging to the cytoplasm-rich and the nucleus-rich class for each pixel can be seen in Fig. 8E. While a majority of pixels segment into two distinct regions (red-yellow colors), an appreciable fraction is clearly intermediate. To understand the effect of this error on the polarization index, we plot the decrease in the index observed as a function of the error in Fig. 8F (*vide infra*).

A ratio of the apical to basal pixels is calculated. In many cancers, cytologic changes in epithelial cells include the loss of polarization and enlargement of nuclei. Hence, it may be hypothesized that atypical pathologic conditions (prostatic intraepithelial neoplasia or carcinoma) may lead to a decreased cytoplasmic-to-nuclear pixel ratios. In contrast, benign hyperplasia and benign tissue are likely to contain larger cytoplasmic-to-nuclear pixel ratios. There are well-known exceptions to this general rule and numerous complications arise from staining patterns. This is, however, a good test case to analyze cytologic details of a specific cell type in tissue, a previously unreported approach. It is interesting to note that we are trying to automatically measure the morphologic structure of cellular ensembles using spectral data, an interesting development in a field (IR microscopy) in which morphology has traditionally been employed to determine regions from which spectra are to be measured.

The cytoplasmic to nuclear pixel ratio is shown in Fig. 9 for four common Gleason Grades seen in prostate cancer, BPH and benign cores of a tissue microarray. A total of 225 cores from ~125 patients were employed to construct the plot. The model was trained on a second array that did not include any data from these patients. Since there is fairly large scatter in the data, it is more instructive to examine the median and the quartile distribution. The height of the histogram bar in Fig. 9, thus,

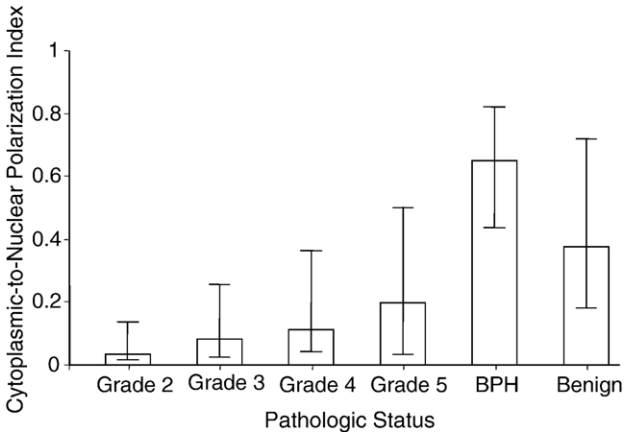

Fig. 9. Epithelial polarization index measured for different pathologies in prostatic tissue.

denotes the median and the range bars denote the quartiles. Although increased nuclear to cytoplasmic ratios are common to malignant cells in comparison to their benign counterparts, this phenomenon does not consistently follow for different grades of malignant cells, as these data suggest. Rather a common feature is increased variability. Clearly, the plot reflects the large scatter in index values but also demonstrates an unmistakable trend. The morphologic index, constructed from spectral identification, correlates with malignancy and appears to be an important parameter in predicting disease in the prostate. Analogous to spectral parameters, in which no single peak predicts malignancy, this one index does not predict the pathologic state of the tissue. However, it is suggested as an additional metric in protocols for differentiating benign and malignant tissue. The classification approach that we have outlined in this manuscript is ideal for employing both spectral and spatial metrics. As with spectral metrics, the spatial index is simply a number that can help distinguish classes of pathology. Its incorporation into spectral prediction algorithms is facilitated with the classification method described here. To our knowledge, that would form the first example of utilizing effectively both morphologic and spectral content of infrared imaging data for disease diagnosis. Caution must be exercised in interpreting results from this procedure. The mean value is strongly dependent on the small error introduced by chance, bias or array-to-array variation. Consider, for example, that the number of nucleus-rich pixels exceed the true value by a fraction ε. Since we are not rejecting any pixels, the excess is obtained at the expense of cytoplasmic pixels. The cytoplasm to nuclear pixel ratio then changes from the ratio P_c/P_n to $\frac{P_c}{P_n}\left[\frac{(1-\varepsilon)}{1+\frac{P_c}{P_n}\varepsilon}\right]$, where P_c is the true number of cytoplasmic pixels and P_n is the true number of nuclear pixels. A small error of ~5% can lead to errors of 5% ($P_c/P_n=0.1$) to 15% ($P_c/P_n=2.0$). Curves relating the observed index to the error are shown in Fig. 8F. Though, the results for epithelial polarization alteration do not provide an unambiguous index, they do indicate that structural parameters derived from spectral data may prove to be an important addition to spectral analyses of tissue.

5. Conclusions

The approaches described in this manuscript significantly enhance the rate and quality of spectroscopic analyses of tissue specimens, allowing realization of the statistical sampling and further numerical analysis to explore associations between molecular changes and clinico-pathologic information. The use of TMAs further allows the original tissue block to be employed for other diagnostic protocols including conventional immunohistochemical staining. The approach is very general as almost all types of tissues or cells can be obtained in a TMA format allowing the translation of spectroscopic sensitivity advantages without barrier to all constituent tissues in the human body. In principle, the approach may be extended to examining fluid or sera by appropriate spotting technology to produce arrays. Our analysis of prostate cancer took well over 2 years with a large effort devoted to designing and implementing analysis protocols. It is expected that the next disease to be analyzed will be completed in a matter of months. We anticipate that imminent advances in IR imaging technology coupled to well-designed microarrays will reduce this time further. The development of a spatial index based on spectral data is demonstrated to be a useful metric of the pathologic state of tissue. It is anticipated that integrative approaches involving both spatial and spectral data will make disease diagnoses using infrared spectroscopy more powerful.

Acknowledgements

This research was supported in part by the intramural Research Programs of the National Institute of Diabetes and Digestive and Kidney Diseases and the National Cancer Institute, Center for Cancer Research.

References

[1] Efforts to provide histologic differentiation using several modalities have been reported (IJ Bigio, SG Bown Cancer Biol Ther 3: 259–67 (2004)).
[2] H.H. Mantsch, L.P. Choo-Smith, R.A. Shaw, Vib. Spectrosc. 30 (2002) 31–41.
[3] B. Rigas, S. Morgello, I.S. Goldman, P.T.T. Wong, Proc. Natl. Acad. Sci. U. S. A. 87 (20) (1990 (Oct.)) 8140–8144.
[4] P.T.T. Wong, R.K. Wong, T.A. Caputo, T.A. Godwin, B. Rigas, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 10988–10992.
[5] D.C. Malins, N.L. Polissar, K. Nishikida, E.H. Holmes, H.S.. Gardner, S.J.. Gunselman, Cancer 75 (1995) 503–517.
[6] H.H. Eysel, M. Jackson, A. Nikulin, R.L. Somorjai, G.T.D. Thomson, H.H. Mantsch, Biospectroscopy 3 (1997) 161–167.
[7] M. Diem, M. Romeo, S. Boydston-White, M. Miljkovic, C. Matthaus, Analyst 129 (2004) 880–885.
[8] R.G. Messerschmidt, in: P.B. Rousch (Ed.), The Design, Sample Handling and Applications of Infrared Microscopes, American Society for Testing and Materials, Philadelphia, PA, 1987, pp. 12–26.
[9] M. Jackson, L.P. Choo, PH. Watson, W.C. Halliday, H.H. Mantsch, Biochim. Biophys. Acta 1270 (1995) 1–6 (See, for example).
[10] R.A. Shaw, F.B. Guijon, V. Paraskevas, S.L. Ying, H.H. Mantsch, Anal. Quant. Cytol. 21 (1999) 292–302.
[11] S. Boydston-White, T. Gopen, S. Houser, J. Bargonetti, M. Diem, Biospectroscopy 5 (1999) 219–227.
[12] D.C. Fernandez, R. Bhargava, S.M. Hewitt, I.W. Levin, Nat. Biotechnol. 23 (2005) 469–474.
[13] E.N. Lewis, P.J. Treado, R.C. Reeder, G.M. Story, A.E. Dowrey, C. Marcott, I.W. Levin, Anal. Chem. 67 (1995) 3377–3381.
[14] I.W. Levin, R. Bhargava, Ann. Rev. Phys. Chem. 56 (2005) 429–474.
[15] R. Bhargava, B.G. Wall, J.L. Koenig, Appl. Spectrosc. 54 (2000) 470–474.
[16] M. Jackson, Faraday Discuss 126 (2004) 1–18.
[17] N. Jamin, P. Dumas, J. Moncuit, W.H. Fridman, J.L. Teillaud, G.L. Carr, G.P. Williams, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 4837–4840.
[18] G.L. Carr, Rev. Sci. Inst. 72 (2001) 1613–1619.
[19] Various effects contribute to the spatial fidelity loss in mapping microscopes as described in Sommer AJ, Katon JE Appl. Spectrosc. 45: 1633–1640 (1991).
[20] R. Bhargava, S.Q. Wang, J.L. Koenig, Appl. Spectrosc. 52 (1998) 323–328.
[21] R. Bhargava, I.W. Levin, Anal. Chem. 73 (2001) 5157–5167.
[22] R. Bhargava, I.W. Levin (Eds.), Spectrochemical Analysis Using Infrared Multichannel Detectors, Blackwell Publishing, Oxford, 2005.
[23] P.R. Griffiths, J.A. de Haseth, Fourier Transform infrared spectrometry, John Wiley and Sons, NY, 1986.
[24] An alternative approach has been suggested, in which reflective glass slides are employed (TJ O'Leary, WF Engler, KM Ventre Appl. Spectrosc. 43: 1095–1097 (1989)). Recently, spectral corrections have been proposed that allow for a better understanding of the distortion of spectra measured using these types of reflective slides (Romeo M, Diem M Vib. Spectrosc. 38: 129–132 (2005)).
[25] J.L. Koenig, S.Q. Wang, R. Bhargava, Anal. Chem. 73 (2001) 360A–369A.
[26] J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter, O.P. Kallioniemi, Nature Med. 4 (1998) 844–847.
[27] T. Braunschweig, J.Y. Chung, S.M. Hewitt, Expert Rev. Proteomics 2 (3) (2005 (Jun)) 325–336.
[28] R.L. Camp, G.G. Chung, D.L. Rimm, Nat. Med. 8 (2002) 1323–1327.
[29] A.C. Williams, B.W. Barry, H.G.M. Edwards, D.W. Farwell, Pharm. Res. 10 (1993) 1642.
[30] P.J. Caspers, G.W. Lucassen, R. Wolthius, H.A. Bruining, G.J. Puppels, Biospectroscopy 4 (1998) S31.
[31] M. Gniadecka, Faurskov, O. Nielsen, S. Wessel, M. Heidenheim, D. Christinsen, H.C. Wulf, J. Invest. Dermatol. 11 (1998) 1129.
[32] L. Knudsen, C.K. Johansson, P.A. Philipsen, M. Gniadecka, H.C. Wulf, J. Raman Spectrosc. 33 (2002) 574.
[33] Y. Hendriks, P. Franken, J.W. Dierssen, W. de Leeuw, J. Wijnen, E. Dreef, C. Tops, M. Breuning, A. Bro¨cker-Vriends, H. Vasen, R. Fodde, H. Morreau, Am. J. Pathol. 162 (2003) 469–477.
[34] M.A. Rubin, R. Dunn, M. Strawderman, K.J. Pienta, Am. J. Surg. Pathol. 26 (2002) 312–319.
[35] P. Lasch, W. Haensch, D. Naumann, M. Diem, Biochim. Biophys. Acta 1688 (2004) 176–186.
[36] P. Lasch, D. Naumann, Cell Mol. Biol. 44 (1998) 189–202.
[37] D.M. Haaland, Appl. Spectrosc. 511 (1997) 340–345.
[38] K.E. Shafer-Peltier, A.S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R.R. Dasari, M.S. Feld, J. Raman. Spectrosc. 33 (2002) 552–563.
[39] p(a|b) is the probability of a occurring given that b has occurred
[40] A.R. Webb, Statistical Pattern Recognition, John Wiley and Sons, New York, 2002.
[41] H.H. Eysel, M. Jackson, A. Nikulin, R.L. Somorjai, G.T.D. Thomson, H.H. Mantsch, Biospectroscopy 3 (1997) 161.
[42] J.R. Mansfield, L.M. McIntosh, A.N. Crowson, H.H. Mantsch, M. Jackson, Appl. Spectrosc. 53 (1999) 1323–1330.
[43] As per Bayes' Theorem, the a posteriori probability $p(A|B)$ may be expressed in terms of the a priori probabilities $p(B|A)$, $p(A)$ and $p(B)$ as $p(A|B) = (p(B|A)p(A))/p(B)$.
[44] The decision rule is equivalent to the classically employed likelihood ratio, $l_r = (p(m_k|c_i)p(c_i))/(p(m_k|c_j)p(c_j))$, but avoids the singularities inherent in disparate regions of occurrence.
[45] It is impossible to prove that the two sets came from the same distribution. Hence, the null hypothesis must be that the distributions came from the

same distribution and, for a metric to be a viable candidate; we need only to disprove the null hypothesis.
[46] J.A. Swets, Science 240 (1988) 1285–1293.
[47] P. Lasch, A. Pacifico, Diem M Biopolymers 67 (2002) 335–338.
[48] A recent study has reported high quality spectra from sub-cellular regions (Gazi E, Dwyer J, Lockyer NP, Miyan J, Gardner P, Hart CA, Brown MD, Clarke NW Vib. Spectrosc. 38: 193–201 (2005)). Unfortunately, the spatial selectivity is unattainable in IR imaging methods and the large scale imaging of tissue is not possible using the study's synchrotron based spectrometer. Emerging coupling of focal plane array detectors with synchrotron sources (G L Carr, O Chubar and P Dumas, Multichannel Detection With A Synchrotron Light Source: Design and Potential in Bhargava R and Levin IW, eds., Spectrochemical Analysis Using Infrared Multichannel Detectors, Blackwell Publishing, pp. 56–84, Oxford (2005)) may provide significantly improved determination of the polarization index in large tissue sections.