

# Applications of Infrared and Raman Microspectroscopy of Cells and Tissue in Medical Diagnostics: Present Status and Future Promises

Max Diem,<sup>1</sup> Miloš Miljković,<sup>1</sup> Benjamin Bird,<sup>1</sup> Tatyana Chernenko,<sup>1</sup>  
Jen Schubert,<sup>1</sup> Ellen Marcsisin,<sup>1</sup> Antonella Mazur,<sup>1</sup> Erin Kingston,<sup>1</sup> Evgenia Zuser,<sup>1</sup>  
Kostas Papamarkakis,<sup>1</sup> and Nora Laver<sup>2</sup>

<sup>1</sup>*Laboratory for Spectral Diagnosis (LSpD), Department of Chemistry & Chemical Biology, Northeastern University, 316 Hurtig Hall, 360 Huntington Avenue, Boston, MA 02115, USA*

<sup>2</sup>*Department of Pathology, Tufts Medical Center, 800 Washington Street, Boston, MA 02111, USA*

Correspondence should be addressed to Max Diem, m.diem@neu.edu

Copyright © 2012 Max Diem et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract.** This paper summarizes the progress achieved over the past fifteen years in applying vibrational (Raman and IR) spectroscopy to problems of medical diagnostics and cellular biology. During this time, a number of research groups have verified the enormous information content of vibrational spectra; in fact, genomic, proteomic, and metabolomic information can be deduced by decoding the observed vibrational spectra. This decoding process is aided enormously by the availability of high-power computer workstations and advanced algorithms for data analysis. Furthermore, commercial instrumentation for the fast collection of both Raman and infrared microspectral data has rendered practical the collection of images based solely on spectral data. The progress in the field has been manifested by a steady increase in the number and quality of publications submitted by established and new research groups in vibrational biological and biomedical arenas.

**Keywords:** Infrared spectroscopy, Raman spectroscopy, medical diagnosis, cytopathology, histopathology, cellular imaging, multivariate methods

## 1. Introduction

The concept of using vibrational spectroscopic method as adjunct medical diagnostic tools dates back over half a century to a time when infrared spectroscopy was itself in its infancy [1, 2]; yet even then, forward-looking spectroscopists thought of the possibility of using the biochemical information obtainable by spectroscopic methods, rather than the morphological information commonly used in classical cytopathology and histopathology, for medical diagnoses. However, it really took until the first decade of the 21st century that the promise for spectral cytopathology (SCP, spectral diagnosis of cells) and spectral histopathology (SHP, spectral diagnosis of tissue) became practical. Notwithstanding a flurry of review articles of a decade earlier which proclaimed spectral diagnostic successes, it took over a

dozen of years of intense efforts to understand even the basic effects that confound infrared spectroscopy of cells and tissues [3, 4], to develop the computational methods to detect the often minute changes in the spectra of cells and tissues with disease [5], and, develop medically acceptable methods for the comparison between spectral and classical diagnostic results. Advances in the spectroscopic efforts were enormously aided by concomitant improvement in measurement technology in the early 2000s and an explosive growth of computational power available to spectroscopists. Interestingly, in the eyes of the authors, the increased computational power, along with the development of some fundamental theoretical underpinnings, was the most important development to propel SCP and SHP toward the commercial realm.

The somewhat sobering consequence of this last statement is the fact that spectral changes between states of disease, or other cell biological event, that are visible to the naked eye are most likely not due to the anticipated effects but due to the aforementioned confounding variations of spectral features based mostly on morphological changes within the tissue or the cells studied. Thus, one universally applicable and highly important result of the spectroscopic studies of cells and tissues is the realization that infrared (micro)spectra are highly dependent on sample morphology: if the sample is not a homogeneous film, but consists of discrete particles, and if the particle size is approximately the same as the wavelength of the infrared light, scattering effects will confound the observed infrared spectra. This scattering and other physical phenomena (*vide infra*) may cause the mixing of absorptive and dispersive line shapes in infrared spectroscopy, which was first documented by researchers in the field of biomedical applications of infrared spectroscopy [4, 6–8].

A typical example of cell-morphology-dependent effects, from the authors' own laboratory, was our first attempt to distinguish normal and cancerous cervical cells. To this end, normal exfoliated cells from the human cervix were to be compared with cultured cervical cancer (HeLa) cells. However, the enormous change in morphology of these cells made this either a trivial or impossible task [9]: the cultured cells have large, relatively thick nuclei which gave good infrared absorption spectra and exhibit strong protein, DNA, and RNA features. Their cytoplasm, on the other hand, is thin and spread-out with pronounced pseudopod features that are common for cultured cells. The cytoplasm of such cell gives very scant spectra dominated by protein features and strong band distortions (see below) at the edges of the cells.

In exfoliated cervical cells, on the other hand, the spectrum of the cytoplasm is stronger and often exhibits pronounced glycogen features. Their pyknotic nuclei exhibit virtually no DNA/RNA features [10]. Thus, the spectral distinction of "cancerous" cervical HeLa cells from normal cervical cells is trivial but medically totally irrelevant. When we attempted to improve the situation by removing the cultured cells from their substrate by trypsination, and mix them with the exfoliated cells, we learned (the hard way) that trypsination although a commonly used procedure in cell biology, changes cell morphology (and possibly the biochemical composition of the cell) drastically; these changes revert when trypsinized cells are subsequently replated in culture flasks and allowed to grow. Upon trypsination, cells typically go from a spread-out morphology, in which a cell can measure up to 25  $\mu\text{m}$  (or larger) in size to near-spherical shape of about 10  $\mu\text{m}$  in size; the accompanying changes in light scattering properties can confound their infrared spectra and produce large shifts in some spectral bands. The resulting spectral changes again are so strong that a comparison between exfoliated (normal) and trypsinized cultured (cancer) cells is totally trivial [9].

Furthermore, spectra of cells or even adjacent tissue pixels do exhibit a natural variance, due to a number of factors (metabolic activity, stage in the cell cycle, tissue architecture, etc.) Thus, any changes in spectral characteristics should be based on a large number of spectra (a spectral “dataset”) of cells or tissue pixels acquired microscopically. In addition, visible images of the cells or tissue pixels must be available to allow correlation of the spectral changes to either confounding morphological causes, contamination, or the actual desired changes in cellular events (disease). Finally, multivariate methods of data analysis should be carried out on the datasets to help differentiate uncorrelated changes (noise) with changes correlated to the desired (or suspected) cause of change.

Thus, this paper differs in the scientific philosophy from that normally seen in the field of “Spectroscopy of Biological Molecules” (the subject of the ECSBM conference series) in that the spectra of cells and tissue pixels are more complicated (they are superpositions of component spectra of unknown abundance) and they are no longer static, but subject to changes that are normally ignored in spectroscopy (morphology, metabolic activity, disease). Thus, those spectroscopists who are accustomed to spectral reproducibility and constancy are forewarned herewith that the remainder of the discussion below will challenge these concepts, but will demonstrate that vibrational spectroscopy of biological systems such as cells and tissue can be interpreted, and valuable diagnostic information can be deduced from spectral results.

This paper will concentrate mostly on recent results in SCP from the authors’ laboratory, the Laboratory for Spectral Diagnosis (LSpD) at the Northeastern University in Boston. The reason for concentrating this review on SCP is that the LSpD has contributed to this field more than to the other fields in spectral diagnostics. Also, the size of the datasets in terms of patient numbers (>250 for oral and cervical cytology) exceeds by far any other datasets investigated by other groups.

Spectral histopathology is being pursued actively at the LSpD as well, with large independent training and test sets available to date for a number of malignancies. These data are not included in this paper, since the work is being carried out under a licensing agreement and cannot yet be divulged. Rather, SHP is introduced from a methodological point of view in which the general procedures and pitfalls are discussed. Raman imaging aspects are treated briefly toward the end of this paper.

## 2. Methods

### 2.1. Instrumental Aspects

All infrared spectroscopic results presented in this paper were acquired microscopically *via* one of three imaging infrared microspectrometers (Spectrum One/Spotlight 400, Perkin-Elmer Corporation, Shelton, CT) at the LSpD, henceforth referred to as the PE400’s. Infrared spectra of cells or tissues were collected in transfection (reflection-absorption) mode from samples mounted on “low-e” (also known as MirrIR) slides (Kevley Technologies, Chesterland, OH, USA) at a spectral resolution of  $4\text{ cm}^{-1}$ . All IR data represented in this paper (both for SCP and SHP) were collected in imaging mode at  $6.25\text{ }\mu\text{m}$  pixel size. The spatial resolution of the PE400 was established using the military resolution targets and was about twice the diffraction limit at  $1600\text{ cm}^{-1}$ , *ca.*  $12\text{ }\mu\text{m}$ . For simplicity’s sake, we may assume that the voxel size interrogated by the instrument is *ca.*  $10 \times 10 \times 5\text{ }\mu\text{m}^3$  (in *x*, *y*, and *z* direction, resp.) in the mid-IR, where the *z*-direction is not so much determined by the diffraction limit, but the maximal thickness of the sample before detector nonlinearity is observed. The pixel size used in these studies allows the detection

of spectral differences of items as small as a cellular nucleus. Since the goals of both SCP and SHP are the detection and diagnosis of individual cancer cells, it is advantageous to operate the spectrometers at a pixel resolution of about the size of a cellular nucleus. In the past, between 4 and 8 interferograms were coadded for each pixel; however, after the implementation of noise-adjusted principal component analysis (see below), only one or two interferograms are coadded. Under these conditions, acquisition of a complete (*ca.* 700 to 4000  $\text{cm}^{-1}$ ) FTIR spectrum for one pixel requires between 5 and 10 ms using the PE400. Spectra were collected at 4  $\text{cm}^{-1}$  spectral resolution and are stored as 800-point intensity vectors with 2  $\text{cm}^{-1}$  data spacing from 800 to 4000  $\text{cm}^{-1}$  in native PE 400 imaging format (.fsm files).

All Raman data were acquired using a confocal Raman microscope (Model CRM 2000, WITec, Inc., Ulm, Germany). In Raman microspectroscopy using mid-visible lasers (*ca.* 500 nm) for excitation, the diffraction limited voxel size is about  $0.3 \times 0.3 \times 1 \mu\text{m}^3$  [11] and proportionally larger for longer wavelength excitation. Laser power at the sample was typically about 10 mW; under these conditions, a Raman spectrum (300 to 3300  $\text{cm}^{-1}$ ) can be acquired in 250 to 300 ms. The raw Raman data are stored as 1024 intensity points with nonlinear wavenumber spacing between data points. The spacing depends on the exciting laser wavelength and grating used and the wavenumber range studied. A 1024-point vector of wavenumber values corresponding to each intensity data point is output with the Raman spectral dataset. Before multivariate analysis of Raman data, all spectra are interpolated to linear wavenumber spacing and corrected for cosmic rays.

## 2.2. Cell Cultures

Most cells grown in the authors' laboratory were purchased from ATCC, Manassas, VA and cultured in 75  $\text{cm}^3$  culture flasks (Corning, Lowell, MA, USA) using minimum essential Eagle's medium (ATCC, Manassas, VA, USA) supplemented with 10%, by volume, fetal bovine serum (FBS, ATCC). Cultured flasks were incubated at 37°C and kept in an atmosphere of 5%  $\text{CO}_2$ . Cells were cultured until confluent and removed from the flasks using trypsin-EDTA (ATCC). Cells were then reseeded onto the windows of choice, "low-e" slides (see below) for infrared measurements or  $\text{CaF}_2$  disks for Raman spectroscopy, immersed in fresh culture medium supplemented with 10% FBS, and placed back into incubation for approximately 12 hours. Cells were fixed with 4% buffered, aqueous formalin for Raman measurements, and in Surepath solution (see below) prior to infrared data acquisition.

## 2.3. Exfoliated Cells

Oral cells exfoliated by LSpD personnel as part of an oral cancer screening program at the Northeastern University (under a local IRB), as well as cells collected from clinical patients at the Tufts Medical Center (TMC) in Boston, were treated exactly the same way. The cells were exfoliated *via* cytobrushes which were immersed into Surepath fixative immediately after exfoliation. This fixation medium consists of 24% aqueous solution of ethanol and 1% each of methanol and isopropanol. We have shown that this fixative changes the cellular spectra minimally, even after prolonged exposure of the cells (1 month) to the fixative, and that spectral changes due to disease are much larger than those produced by fixation protocols and exposure to fixatives (see below) [12]. The Surepath fixation protocol was adapted

at the LSpD since it was the method of choice at TMC, from where all clinical samples were derived. A comparison of formalin- and Surepath-fixed cells showed minimal differences [12].

After repeated wash and centrifugation cycles, cells were spin-deposited *via* cytocentrifugation onto “low-e” slides (see above). Sparse, uniform samples of cells, which adhere to the substrate very strongly, could be produced this way.

#### 2.4. Tissue Sections

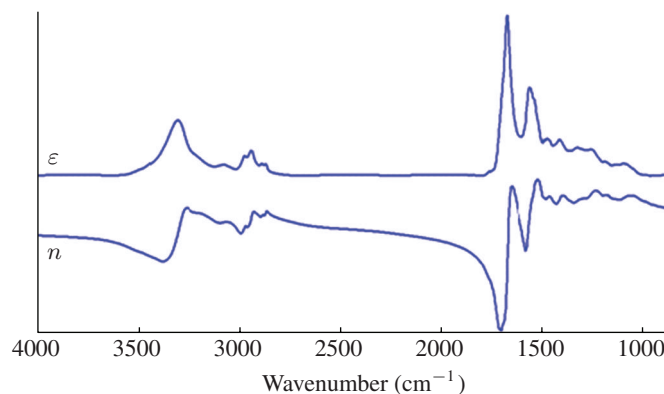
Tissue sections were cut, using a microtome, to a thickness of 5  $\mu\text{m}$  from formalin-fixed, paraffin-embedded tissue blocks from the archives of the Pathology Department at TMC. The sections were mounted on low-e slides and deparaffinized using standard protocols [13]. Some spectra were also obtained from the tissue pixels while still embedded in paraffin. After infrared data acquisition, tissue sections were stained with hematoxylin-eosin (H&E) to permit correlation of visual and infrared spectral images.

#### 2.5. Computational Methods

All data manipulation and analysis was carried out, using software developed in house using the MATLAB (MathWorks, Natick, MA, USA) platform. The analyses start with the raw (Raman or infrared) instrument-based data files. Most of the data analysis routines are contained in a software package referred to as “ViChe” (vibrational chemometrics), which includes all of the preprocessing and multivariate imaging reconstruction algorithms, for example, principal component and hierarchical cluster analysis (HCA) imaging. The latter have recently been discussed in detail [14]. Among the preprocessing routines, noise-adjusted principal component analysis (NA-PCA) was taken from the literature [15, 16], whereas the correction routines for band shape distortion, to be discussed in the next section, were developed in-house [17–19]. The algorithm to construct spectra of individual cells from imaging datasets has been reported [20] and submitted for IP protection. Following earlier arguments [5, 21], all data analysis was carried out on second-derivative spectra.

### 3. Results and Discussion

This paper follows the presentation by the author at ECSBM14 both in subject matter as well as in order. Thus, the first subject to be discussed will be methods for the correction of dispersive band shape distortions that are frequently encountered in infrared spectroscopy of human cells and tissues. This particular problem, the sample morphology-dependent spectral distortion, has plagued this research area since its inception more than 60 years ago, and is not restricted to microscopic data acquisition in transfection mode. (In the very early work on infrared spectroscopy of tissue, Blout and Mellors [1] suggested to put a drop of oil on the tissue sections to match the refractive index of sample and surroundings). The difficulties arising from the dispersive band shapes are so severe that there was widespread pessimism about the future of infrared microspectroscopy as a possible medical diagnostic tool. Only after the SPEC2010 conference in Manchester, UK, where a number of research groups presented their views and approaches to overcome this problem, did the mood swing drastically. At the time of



**Figure 1:** Observed absorption coefficient and computed dispersion of the refractive index for a tissue pixel or cell.

writing of this paper, there are three methods of correcting the dispersive band shapes in the literature or submitted for patent protection.

### 3.1. Correction of Dispersive Band Shapes

In two pioneering papers, the research group around Peter Gardner at the University of Manchester, UK, described [4, 7] how reflection processes and Mie scattering can mediate the mixing of dispersive and absorptive band shapes. However, similar mixing phenomena were well known to occur, for example, in specular reflection [22] and in absorption measurements from metal surfaces [23, 24] as well as in ATR spectroscopy. In all these modalities of IR spectroscopy, the absorption spectra are not measured directly, but *via* methods that depend on the complex refractive index  $\eta$ , which is given by

$$\eta = n - i n \kappa = n - \frac{i \varepsilon \lambda}{4\pi}. \quad (3.1)$$

In (3.1),  $n$  is the real part of the refractive index,  $\kappa$  the absorptivity,  $\varepsilon$  the molar extinction coefficient, and  $\lambda$  the wavelength of light. Whenever  $\kappa$  or  $\varepsilon$  have a maximum (i.e., whenever one observes a peak in the absorption spectrum),  $n$  undergoes anomalous dispersion, shown in Figure 1. The real and imaginary parts of the refractive index are related to each other by the Kramers-Kronig transform:

$$n(\nu_i) - n(\infty) = \left( \frac{2}{\pi} \right) \int_0^\infty \frac{\nu \kappa(\nu)}{\nu^2 - \nu_i^2} d\nu, \quad (3.2)$$

in which the  $\nu_i$  are the frequencies of the spectral peaks, and,  $n(\infty)$  is the refractive index at infinitely long wavelength where no transitions occur. Thus, the dispersion of the refractive index can be calculated from the absorptivities and *vice versa*. A comparison between an absorption spectrum and the corresponding dispersion curve is shown in Figure 1. These phenomena are well known in classical



optics and in chiroptical spectroscopy: circular dichroism and optical rotatory dispersion are typical examples of two effects related by the Kramers-Kronig transform. However, in both classical optics and spectroscopy, one normally shies away from conditions where the two effects, absorption and dispersion, interact: most textbooks of optics treat the refractive index as a quantity that changes very mildly with wavelength, because most optical materials are chosen such that they have no absorption in the spectral range of interest; that is, they are colorless (clear) in the visible spectrum. In absorption spectroscopy, on the other hand, one assumes that the reflection losses at the sample, caused by the refractive index, are small and will not distort absorption spectra noticeably.

However, under certain condition this simplistic situation breaks down, and one observes extensive mixing of reflective and absorptive band profiles. This was first formulated for the case of Mie scattering by Bassan et al. [4] and can be visualized as follows. Mie scattering is not a molecular but rather a macroscopic effect in which spherical or near-spherical metallic or dielectric particles scatter incoming radiation to produce broad, undulating background patterns. This effect predominates if the particle size and the wavelength of light, typically between 5 to 12  $\mu\text{m}$  for mid-IR measurements, are approximately equal. Consequently, small human cells, or the nuclei of cells, can exhibit strong Mie scattering. The classical physical equations for the Mie scattering are quite complicated [25]; however the Mie scattering cross-section can be approximated relatively accurately [26, 27] for a transparent sphere by

$$Q_{\text{sca}} = 2 - \left(\frac{4}{\rho}\right)(\sin \rho) + \left(\frac{4}{\rho^2}\right)(1 - \cos \rho), \quad (3.3)$$

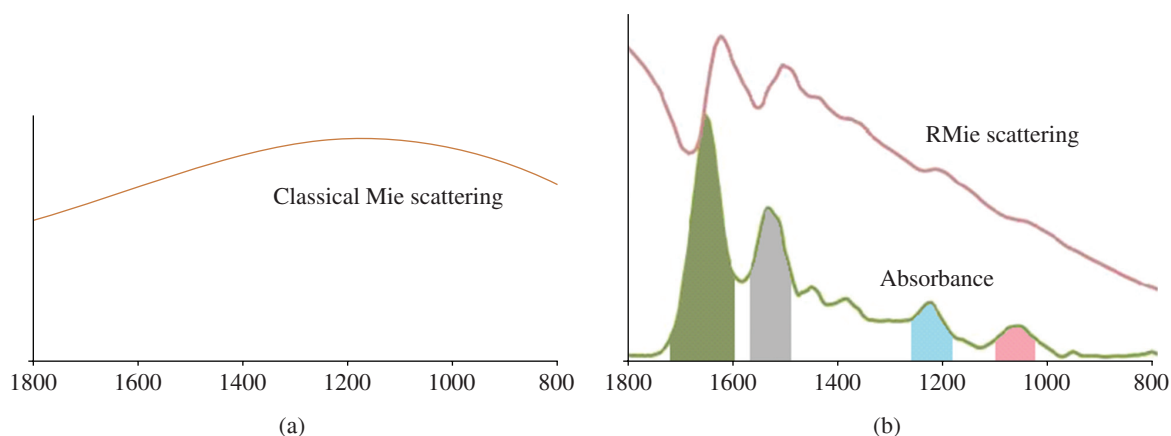
with the scattering factor  $\rho$  given by

$$\rho = \frac{4\pi r(n_{12} - 1)}{\lambda}, \quad (3.4)$$

where  $r$  is the radius of the scattering sphere,  $\lambda$  the wavelength of the light, and  $n_{12}$  the ratio of the refractive index of the scattering sphere and the surroundings. However, if the scattering material exhibits absorptions and therefore a wavelength-dependent refractive index,  $n_{12}$  needs to be replaced by the dispersion curve shown in Figure 1. The resulting Mie scattering, referred to as resonance Mie (RMie) by Gardner's group [4], is shown in Figure 2 for a region in which classical Mie scattering would exhibit a very gently varying profile.

Similar mixing of reflective and absorptive band profiles may be observed, at times, in pure reflection spectroscopy but also appeared in studies where surface-enhanced infrared absorption (SEIRA) was investigated. In these studies, coagulated (coalesced) gold or silver surfaces were prepared by vapor deposition of the respective metals and used as substrates for neat liquids. The "absorption" spectra observed for a number of these liquids showed purely reflective band profile. This can be understood in terms of the near-constant refractive index of the metal particles undergoing Mie scattering, in contact with the neat liquids, whose refractive indices undergo anomalous dispersion [23, 24].

As mentioned before, the distortions observed in infrared spectra of spherical cells (e.g., lymphocytes) can be so severe that interpretation of the spectra is impossible. Also, the use of multivariate



**Figure 2:** Classical Mie, resonance Mie (RMie) and pure absorption spectra of tissue. The olive-shaded area denotes the amide I manifold, the gray the amide II region, and the blue and pink areas the regions of antisymmetric and symmetry phosphodiester stretching vibrations.

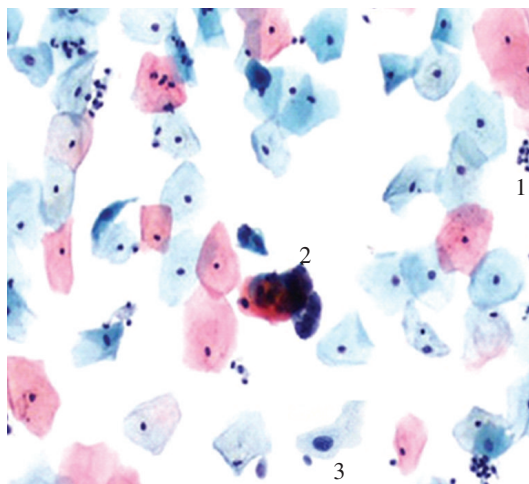
methods of data analysis, such as principal component analysis (PCA), was severely confounded by the intensity distortions and frequency shifts caused by reflective and RMie scattering contributions. Therefore, Bassan et al. [17] proposed a method to correct distorted spectra by fitting a refractive index (interference) spectrum obtained *via* the Kramers-Kronig transform of the absorption spectrum to minimize the distortion. A similar approach, requiring much less computation time but more than one “interference” spectrum, was published by the LSpD group shortly thereafter [18, 28].

However, both of these approaches required “uncontaminated” spectra as reference spectra and their Kramers-Kronig transforms as dispersive interference spectra, and results obtained using these two approaches either reduced [17] or amplified [18] the real variance in the spectra. Thus, we introduced [19] another method based on the well-documented “phase correction” (PC) approach that is widespread in standard FTIR spectroscopy. The phase correction approach is based on the concept that the complex Fourier transform separates the real and imaginary parts of spectra or interferograms by varying the phase angle between them. In classical FTIR spectroscopy, the collected interferogram is generally asymmetric about the zero path difference (ZPD) peak; such a “chirped” interferogram gives, upon forward FFT, a spectrum that contains a mixture of reflective and absorptive band shapes [29]. Nearly all commercial FTIR instruments use the Mertz phase correction method [30] for which the instrumental phase angle is determined experimentally and is used to correct the spectra.

A PC-based approach was attempted by us [6] earlier but worked only intermittently due to some computational and theoretical problems. Recently, a revised phase correction algorithm was implemented for fast, reliable, and elegant correction of reflective band contributions. In short, the distorted spectra, expanded to the desired frequency range and denoised by NA-PCA (see above), are reverse Fourier transformed back into interferogram space. The resulting real (Re) and imaginary (Im) interferograms are zero-filled, and phase shifted by a trial phase according to

$$\begin{pmatrix} \text{Re}' \\ \text{Im}' \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \text{Re} \\ \text{Im} \end{pmatrix}. \quad (3.5)$$





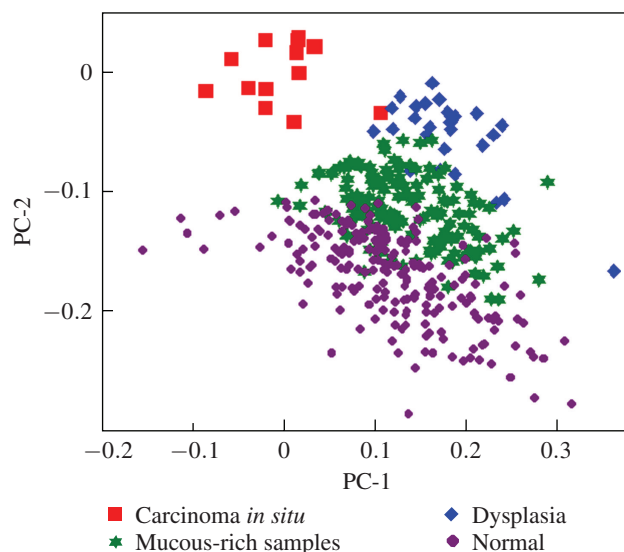
**Figure 3:** Stained cytopathological sample of cervical epithelial cells.

Phase corrected spectra are computed by complex forward FFT of  $(\text{Re}' + i\text{Im}')$ . The “best” phase is assumed to be the one that produces the highest intensity corrected, since addition of a reflective component will always reduce the peak intensity [29]. In principle, a wavenumber-dependent phase angle can also be computed, as in the Mertz algorithm, from a low-resolution interferogram obtained by truncating the FFT to fewer data points.

### 3.2. Spectral Cytopathology

#### 3.2.1. General Remarks

One of the early goals of researchers involved in SCP was the developments of methods to aid in the diagnosis of cervical cell smears [31–34] used for screening for cervical cancer (the so-called Papanicolaou test (or “Pap” test for short) [35, 36]. The reason for selecting this subject was the well-documented high rate of false-positive and false-negative readings of these samples by cytologists and cyto-technicians. Interestingly, one should not take the low sensitivity and specificity of the Pap test as a failure of the method; quite contrary, no single test has reduced the incidence of, and morbidity from a given cancer as much as the Pap test. (The terms sensitivity and specificity are used in this paper in the clinical sense, where sensitivity refers to the ratio of true diagnoses divided by the total number of true-plus false-negative diagnoses, and specificity refers to the ratio of true negative diagnoses, divided by the true negative plus-false-positive of diagnoses.) Yet, for a single reading of a classical “smear,” the overall accuracy (average of sensitivity and specificity) was less than 70%. Improvements were achieved by producing better samples: rather than smearing the exfoliated cells on a microscope slide followed by staining, liquid-based methods were devised [37] that produced sparse monolayers of cells, which, after staining, provided a much clearer picture of individual cells. A small section of such a stained sample is shown in Figure 3.



**Figure 4:** PCA scores plot for macroscopically acquired cervical cell samples.

This figure shows about 50 cervical squamous cells from a small section of a liquid-based sample which stain either pale blue or pale pink; in addition, this sample contains bacteria (1), polymorphonuclear leukocytes (PMNs) (2), cellular debris such as naked nuclei, and one “abnormal” cell (3), indicated by an enlarged nucleus. The difficulty in classical (visual) cytopathology is the detection of as few as a few percent of abnormal cells in a sample that may contain 1,000 to 10,000 cells. Furthermore, an enlarged nucleus *per se* (a larger nucleus/cytoplasm (N/C) ratio) is also observed for cells from the lower layers of the cervical epithelium; thus further criteria, such as the morphology of the nuclear membrane, need to be invoked for a reliable discrimination of normal from abnormal cells. The level of abnormality also needs to be graded; for cervical cytology, the grades (in order of increasing severity) include reactive, low-grade dysplasia (low-grade squamous intraepithelial lesion, LSIL), high-grade dysplasia (high-grade squamous intraepithelial lesion, HSIL), carcinoma in site (CIS), and invasive cancer.

Given the complexity of the problem, it was no surprise that cervical cancer screening was selected as a target for early spectral diagnostics. When first attempts at this goal were made in the early 1990s, infrared microspectroscopy had not progressed to a level that permitted acquisition of spectra of individual cells in reasonable times; thus, cell pellets of unknown composition in terms of the cell types contained in the pellet were used as samples, and the measurements were carried out macroscopically. Results from these early efforts are exemplified by the PCA scores plot [33, 38] shown in Figure 4. In this plot, each symbol represents one spectrum collected macroscopically from a cell pellet; the confirmatory diagnosis was by classical pathology. In retrospect, it is amazing that these early cell pellet results gave encouraging results, and it took quite a while to understand why these crude measurements showed any kind of discrimination.

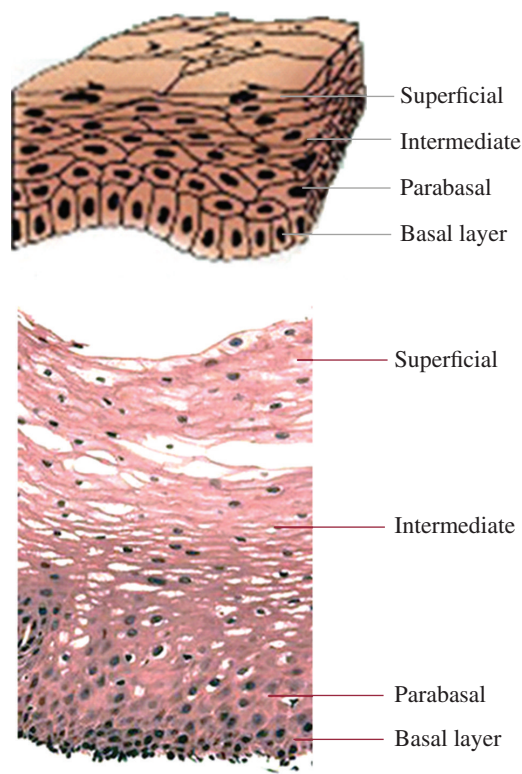
To allow direct assessment of individual cells and to convince cytologists of the value of the spectral method, the author’s laboratory switched to single-cell-based spectral cytopathology in 2002. Although much more time consuming than the pellet-based approach, the single-cell method permits

a direct comparison of spectral results with visual inspection of a cell and thus is of much higher inherent value to a cytologist. To this end, the sample is stained and cover-slipped after infrared data acquisition, and cells can be relocated on the substrate by their stored positions. Early results showed promise: the distinction between squamous cells from the distal urethra from urothelial (bladder) cells by spectral methods proved to be straightforward [39, 40], and the classification of superficial and intermediate cells from cervical epithelium and effects of hormonal influences could be established quite readily [40]. The hormonal effects were originally sought to help classify samples from premenopausal and menopausal women, but needed to be expanded to include menstrual status since the level of hormones influences the maturation of cervical cells. This was first established for canine cervical cells [41] but was found to hold for human cervical cells as well. In order to facilitate the detection and diagnosis of cervical disease to be discussed below, only women using oral contraceptives were used, since they keep the hormonal level constant and thereby eliminate one variable in the process of establishing SCP as a possible diagnostic method. Results of the cervical work will be presented later in this section.

Concurrent with the efforts on cervical cells, development of a screening test for oral cancer was initiated at the LSpD. Oral mucosa, like cervical mucosa, consists of stratified squamous epithelium but, due to digestive enzymes in the saliva, does not exhibit large spectral contributions due to glycogen, which masks a large part of the low-frequency spectrum ( $1000\text{--}1200\text{ cm}^{-1}$ ). The original results for the oral mucosa were extremely intriguing and largely form the basis of our present understanding of spectral cytology. Before presenting these results, a short introduction of stratified squamous tissue will be presented, followed by cursor interpretation of a typical vibrational spectrum of a cell.

Stratified squamous epithelial tissue is a frequently found epithelium in the human body (nasopharyngeal and oral cavities, esophagus, urethra, vagina, cervix, and others). It is a multilayered structure (see Figure 5) consisting of a layer of actively dividing basal cells anchored to the basement membrane, beneath which one finds connective tissue (stroma). The daughter cells created by division of the basal cells form the parabasal layer and mature and migrate to the surface layer. In this process, their morphology and chemical composition change drastically. Whereas the basal cells are roughly cuboidal in shape, about  $15\text{ }\mu\text{m}$  on edge, with a large nucleus and very little cytoplasm, the mature stratified (flat) cells may measure up to  $60\text{ }\mu\text{m}$  on edge and exhibit very small, pyknotic nuclei. They also accumulate glycogen for energy storage (except for the oral mucosal cells, see above); since they are fully differentiated, their chemical composition reflects a reduction of certain compounds in the cytoplasm. Using high spatial resolution synchrotron radiation FTIR imaging and enzymatic digestion studies, we have shown that RNA signatures, for example, are nearly absent in the cytoplasm of mature squamous cells [42, 43]. Furthermore, we have shown that the nuclear DNA is virtually unobservable in pyknotic cells [3] but contributes to the observed spectra in rapidly dividing cells, for example, cancer cells and lymphocytes.

Finally, a very cursory interpretation of a typical spectrum of a cell (or tissue pixel) will be presented. The  $1500\text{--}1700\text{ cm}^{-1}$  region of the spectrum of a cell or tissue pixel is dominated by the protein amide I and amide II bands, shaded olive and gray in Figure 2; both these bands split into subbands in the second derivative spectra and are known to be due to exciton-like coupled states of mainly the  $\text{C=O}$  and  $\text{O=C-N}$  stretching coordinates, respectively [44]. Certain proteins, in particular the proteins of connective tissue (collagen), have sufficiently different infrared spectra to allow detection of their spectral signatures with the naked eye.

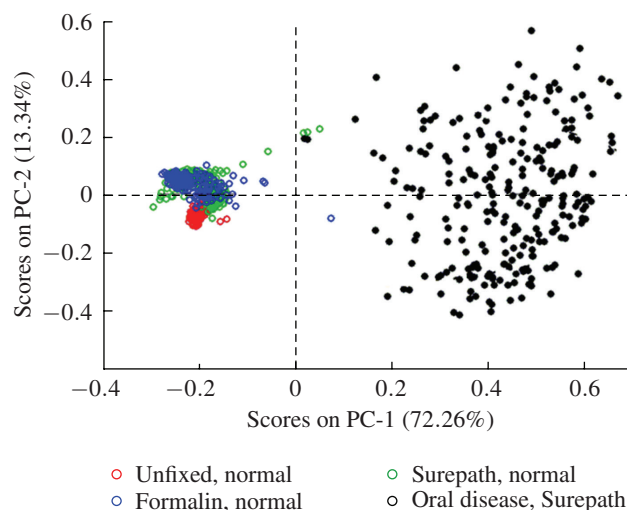


**Figure 5:** Schematic drawing (top) and actual image (bottom) of stained stratified squamous tissue (from Wikipedia).

A high-frequency band, at about  $1740\text{ cm}^{-1}$ , is due to the ester linking of phospholipids [5, 43]. The antisymmetric and symmetric phosphodiester stretching vibrations of DNA, RNA, and phospholipids are observed at *ca.*  $1235$  and  $1090\text{ cm}^{-1}$ . The intensity of these bands varies enormously in disease. The  $\text{C}=\text{O}$  stretching bands of non-hydrogen-bonded nucleotides are observed mostly as high-frequency shoulders of the amide I peak. The amide III vibration occurs superimposed on the antisymmetric phosphodiester stretching vibration. Carbohydrates show strong peaks due to  $\text{C}-\text{O}-\text{H}$  deformation and  $\text{C}-\text{O}$  stretching coordinates between  $1000$  and  $1200\text{ cm}^{-1}$ . One of the most abundant cellular carbohydrate is glycogen, which shows three strong bands at  $1151$ ,  $1078$ , and  $1028\text{ cm}^{-1}$ . Carbohydrate bands are also observed for glycoproteins, particularly in mucus.

### 3.2.2. Spectral Cytopathology of the Oral Mucosa

We start the discussion of SCP by presenting the results of a preclinical trial presently ongoing on the campus of Northeastern University, in collaboration with the Department of Pathology at the Tufts Medical Center (TMC) in Boston. Oral cytology was selected as an initial large-scale target because of the ease of sample collection, the prevalence of viral diseases in the oral cavity (*human papillomavirus*,

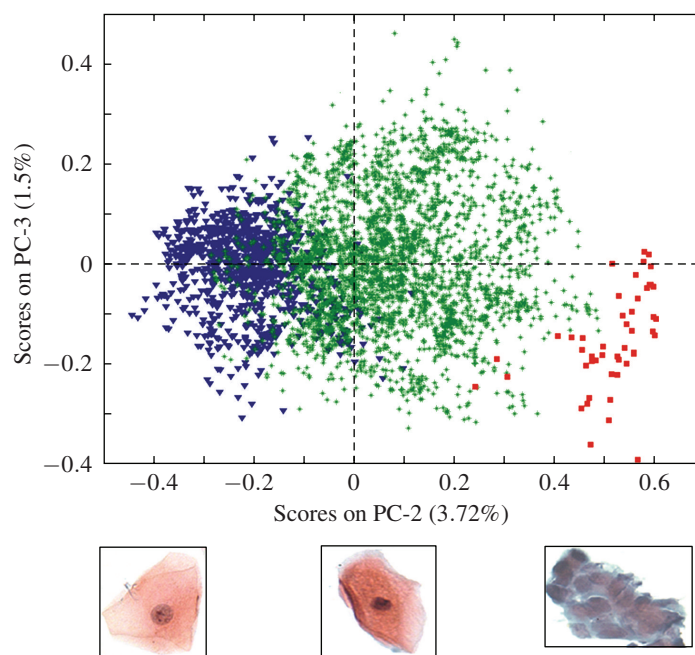


**Figure 6:** PCA scores plot of fixed normal and precancerous oral mucosa cells.

*herpes simplex*, Epstein-Barr), and the prevalence of oral and nasopharyngeal cancers in the Far Eastern population, who represents a large percentage of oral cancer incidence seen at TMC.

Cytological samples were harvested as described above and immediately inserted into a vial filled with Surepath fixative. A frequent criticism, by referees and grant reviewers, of the authors' efforts to use SCP as a diagnostic tool has been regarding the effect of fixation on cells and tissues. We have—hopefully once and forever—answered these issues in a recently published paper which demonstrates that even prolonged exposure to fixative and fixation by different methods (drying, formalin and Surepath fixation) cause changes in spectral features that are significantly less than those caused by disease. This is shown *via* the PCA plot depicted in Figure 6. In the past, enormous spectral changes of cells and tissues upon fixation had been reported [45], which we believe were mostly due to morphological changes upon fixation. In tissue, the changes upon formalin fixation/paraffin embedding are larger, but they do not interfere with vibrational spectral diagnostics if all tissue sections that are compared are treated the same way. This was demonstrated nicely by back-to-back papers published on rat brain gliomas in 2006 [46, 47] which were either flash frozen or formalin-fixed and paraffin embedded and subsequently deparaffinized (see below).

In Figure 7, we present first results on the work on oral mucosa cells [48], which initially were somewhat surprising but followed a finding that had been reported before for cervical cells [33, 34]. The results in this earlier work implied that morphologically normal cells from abnormal samples exhibit abnormal spectra, which represented a progression from normal to cancerous cells. However, the sample set reported then was too small to reach any detailed conclusions. All the cell spectra shown in Figure 7 were from cells harvested from the tongue, since we had demonstrated earlier that there exist small, but reproducible changes in the spectra of cells harvested from different anatomical regions of the oral cavity. In the score plot of Figure 7, the normal cells (from six volunteers) shown as blue symbols form a tight cluster. The cells harvested directly from a cancerous lesion of the tongue, shown in red, form a less homogeneous cluster which is well separated from the normal cells. Most interesting is the diffuse cluster represented by the green symbols. These spectra were from cells, which exhibited normal

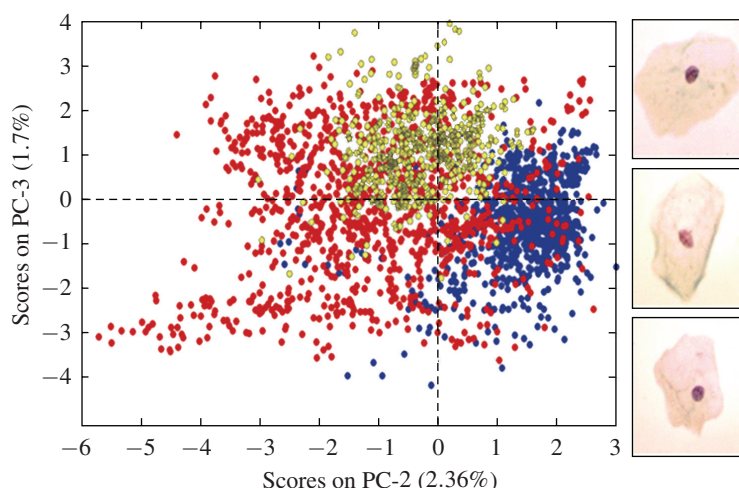


**Figure 7:** PCA score plot of oral SCP. The cell images represent a normal cell (left), a morphologically normal cell from an abnormal sample (middle), and a clump of cancerous cells. All cells were harvested from the tongue.

morphology but were collected from cancer patients from areas quite remote from the cancerous lesions and from patients diagnosed with precancerous disease. This observation, namely that the majority of exfoliated cells that still exhibit normal morphology but exhibit abnormal spectra, can be explained by the fact that most of the area surrounding the cancerous lesion is already affected by a biochemical change or mutation that pathologists refer to as “malignance induced changes” or “field cancerization.” Although the definition of these terms is somewhat vague, it is well known that—particularly in the case of oral cancer—the rate of reoccurrence after treatment of a first cancer incidence is 20-fold higher than that for healthy patients. This implies that there are precancerous changes in the cells that do not manifest themselves morphologically. Completely analogous results were obtained by *in vivo* Raman spectral measurements by the Vanderbilt research group [49] for the ectocervix. Another interpretation of these results will be presented at the end of the SCP section.

In addition, we demonstrated that infection by the *Herpes simplex* virus could be detected by SPC and that the cells collected from different anatomical regions of the oral cavity (cheeks and gums, tongue, and mouth floor under the tongue) can be distinguished by SCP. Furthermore, the spectral patterns of degradation products of pain-killer medication (ibuprofen) could be found in cell spectra, as could be the byproducts of nicotine use [48]. Although the changes produced by these degradation products were too small to be perceived by visual inspection of cells, PCA could easily classify uncontaminated from the contaminated cell spectra. In particular, the PCA loading vectors along which the spectral classes were differentiated often gave a good indication of the spectral changes detected by PCA. This will be discussed further in the next section.





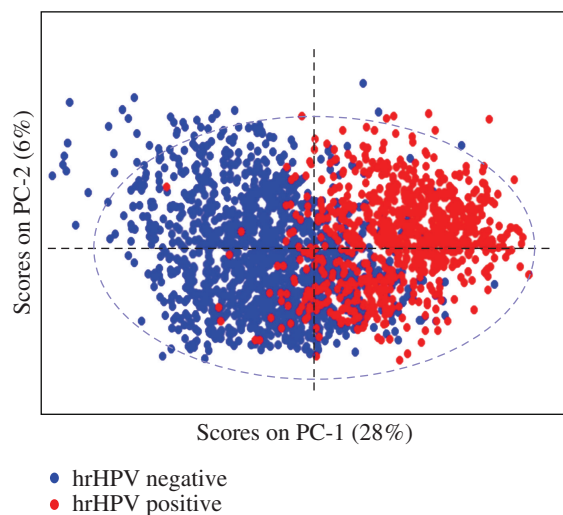
**Figure 8:** PCA score plot of spectra of cervical squamous cells from normal subjects (blue) and from LSIL (red) and history of HSIL (yellow) patients.

### 3.2.3. Spectral Cytopathology of Cervical Mucosa

In cervical cytology, which was the original goal of the research described here (see above), completely analogous results were observed, namely, that the majority of cells from abnormal samples showed spectral abnormality, although they exhibited normal morphology. These results mirror the earlier findings by Cohenford et al., [33] and Cohenford and Rigas [34], but took into account additional confounding factors, as indicated above: the cells are subject to hormone-mediated maturation processes that include, for example, the accumulation of glycogen in the final stage of maturation. Since the glycogen absorption bands mask a large part of the low-frequency ( $1000\text{--}1200\text{ cm}^{-1}$ ) spectrum, valuable information in the nucleic acid phosphate stretching region (*ca.*  $1090$  and  $1235\text{ cm}^{-1}$ ) is rendered inaccessible. Thus, the studies reported below were from subjects using oral contraceptives which prevent complete maturation of the squamous cervical epithelium and therewith reduce the glycogen abundance. The changes in maturation patterns of cervical cells in response to menopause could be demonstrated nicely using SCP [40].

Figure 8 shows the results of PCA analysis of cells exfoliated from normal patients, and patients diagnosed with LSIL/HSIL (see Section 3.2.1) [50]. Here, the results of the oral cytology are repeated in that most of the cells from patients with dysplasia exhibit spectral abnormalities, although the cell morphology is normal (see cell images in Figure 8). Even more surprising is the fact that the cells from a patient with a prior diagnosis of HSIL and subsequent treatment still exhibit abnormal spectral patterns and cluster with the abnormal spectra.

The implications of this observation are quite far reaching in that detection of abnormality is not restricted to the few cells in a cervical exfoliate that exhibit abnormal morphology. Rather, SCP detects abnormal spectral signatures that are exhibited by most of cells, even if they still have normal morphology. The fact that the abnormality persists after treatment led us to explore the possibility that the spectral changes (and therewith the “malignancy associated changes” [51] or “field cancerization” [52] mentioned earlier) are actually due to viral infection. In part, this thought was provoked by results



**Figure 9:** PCA score plot of SCP results of hrHPV positive and negative samples (from [54]).

from a patient with an acute *H. simplex* infection of the oral cavity. In this case, most of the cells showed spectral abnormality, not only those that were so grotesquely deformed by the viral infection that they could be visually diagnosed by a cytologist [48]. Since statistically over 95% of all cervical dysplasia occur along with (and likely are caused by) infection with the human papillomavirus (HPV), the possibility exists that SPC detects the infection by HPV in cervical cytology [53]. Similarly, oral dysplasia could be caused by the Epstein-Barr or HPV as well.

These observations may help explain the positive results reported for cell pellet studies (Figure 4): although the composition of the cell pellet in terms of contributions from superficial, intermediate, and parabasal cells, as well as PMNs and bacteria, was not known for these samples, the abundance of virally infected cells may have been responsible for the distinction of disease states. Efforts to answer the possibility of the involvement of viral infection toward the observed spectral changes will be pursued in the next section.

#### 3.2.4. Viral Effects

In order to explore the sensitivity of SCP toward viral infections, a study was undertaken in which 48 samples were tested for high-risk HPV infection (hrHPV) *via* the Digene hybrid capture test (Qiagen, Valencia, CA, USA). The spectral results were analyzed by SIMCA [54]. The result of a 10 sample training subset, shown in Figure 9, looked extremely promising, with good spectral separation of HPV-positive and HPV-negative samples. When applied to the remaining set of samples, a sensitivity of 88% was achieved; yet the specificity was only 43 percent. This implies that SCP was quite good at detection of hrHPV strain when it was present (as determined by the Digene test) but not accurate when no hrHPV infection was present. We attributed the low specificity to the fact that low-risk HPV (lrHPV) infection is epidemic in the population of women between 20 and 25 years of age, with infection rates of about 30%, or about the same as the infection rate with hrHPV [55]. Thus, it is quite likely that the samples which

tested negative for hrHPV by the Digene test had low-risk HPV infection, which the spectral methods were not (yet) able to differentiate.

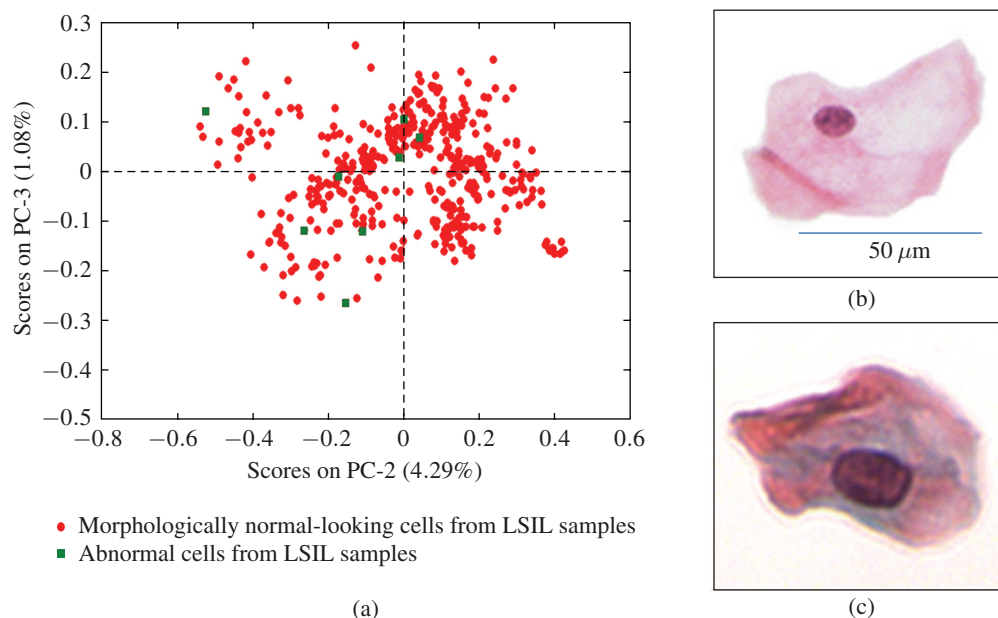
Inspection of the PCA loading vector (not shown) indicated that the spectral changes, along which PCA and SIMCA distinguish the HPV-infected from normal cells, occurred in the protein spectral region by a distinct shift of the amide I band and the appearance of small shoulders. This leads to the conclusion that it is not a change in the viral DNA which is detected, but rather different proteins expressed by the virus. Given the size of the viral genome (*ca.* 5000 base pairs, bp) and the number of copies of the viral genome in a cervical cancer cell (maximally *ca.* 600 in the CaSki cell line, fewer in HeLa cells), one arrives at a number of about 3 million base pairs added to the human genome in the case of HPV infection. The human genome consists of 3 billion bp; thus, it is impossible to detect this change with present methodology. If affected cells produce proteins different from the normal proteome of cervical cells, such changes can be amplified and detected spectroscopically. Efforts are underway to shed light onto the reasons for the observed spectral changes [54, 56].

### 3.2.5. Future Potential of SCP

Aside from the fixation studies, which were carried out for both exfoliated and cultured cells, and the viral load studies, which were carried out on cultured cells only, the majority of the work presented in the sections above has dealt with exfoliated cell; that is, it reported a truly new form of cytology, namely, SCP. To the best of our knowledge, work on exfoliated cells at the cell-by-cell level is now being carried out exclusively at the LSpD, and the size of the datasets at the LSpD far exceeds all previously collected datasets combined [48, 54]. At the time of writing of this paper (Summer 2011), it appears that SCP has matured to a level which allows for detection of cellular abnormalities, such as dysplasia, cancer, and viral infection in exfoliated cells, and thus is poised to be applied to areas where classical cytology has very poor performance, in many cases below 50% accuracy.

The reason that the progress in SCP has been somewhat slower than that in other areas of spectral diagnostics is the fact that the correlation between classical cytopathology and SCP is difficult. In SCP, one has to rely on luck that within an ensemble of cells scrutinized by SCP there is a diagnosable, abnormal cell. After a few thousands of cells from dysplastic patients, however, it will become very likely that some cells are found that display abnormal spectra and can, indeed, be diagnosed. Such a case is shown in Figure 10 which shows a clearly dysplastic cell whose spectrum clustered with other abnormal spectra [54].

For cultured cells, the efforts and research directions are more diversified and represent a number of other research groups as well as the LSpD. These efforts have demonstrated that infrared microspectroscopy can detect the stages of a cells division cycle [57], the effect of drug treatment on cells [58–60], the degree of aggressiveness [61],  $\mu$ RNA expression [62], cancer activation of fibroblasts [63], and a few others. In general, the results of these studies demonstrate that carefully carried-out FTIR studies can reveal an extraordinary amount of information on the complex biochemical changes that occur when cells undergo natural or induced processes. A few general rules seem to apply for carefully planned and executed studies. The raw spectra, whether monitoring drug treatment or any other of the changes listed above, exhibit no or barely visible spectral changes, and multivariate methods of analysis need to be employed to visualize spectral variations. A typical example is the study by the Brussels group [59], which demonstrated elegantly that the spectra of untreated cells and those treated with drugs



**Figure 10:** (a) PCA plot of cervical cells from patients diagnosed with low-grade dysplasia. Red circles are mostly from cells with normal morphology (shown in (b)), whereas the green squares are from “diagnostic” cells with abnormal cytology, shown in (c) (from [54]).

are virtually identical to the naked eye, but that statistical (or, in this case, 2D) analysis of these datasets reveals changes that can be interpreted biochemically. Some of the spectral changes, for example, due to drug interactions, are smaller than the changes due to cancerous disease; thus, the authors believe that the spectroscopy of cells can reveal much more information and can be used to reveal very subtle details. For bacterial cells, for example, infrared spectroscopy, coupled to analysis by neural networks, could predict the mode of action of newly discovered drugs [64].

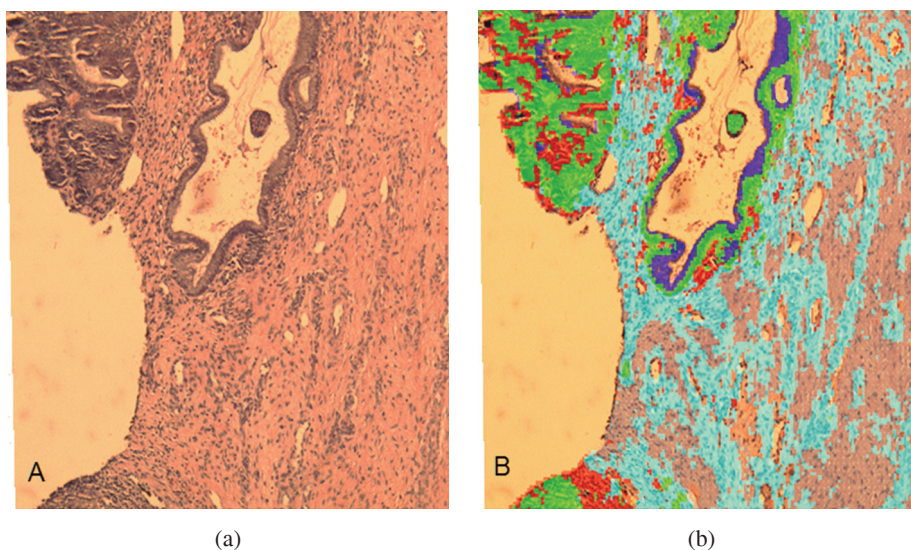
### 3.3. Spectral Histopathology (SHP)

#### 3.3.1. General Remarks

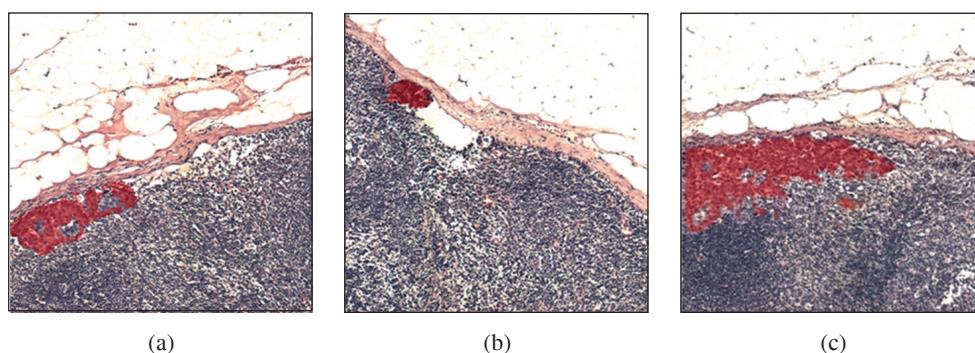
Although classical histopathology is the gold standard of primary medical diagnostics (nearly every cancer diagnosis is initially based on histopathology) and has high sensitivity and specificity in detecting cancers, the method is somewhat more ambiguous when it comes to grading disease. It is also an inherently subjective approach to diagnostics and lacks reproducibility and cannot easily be carried out *via* a quantitative and reproducible measurement. Furthermore, the detection of specific subtypes, for example, the overexpression of cancer genes, requires immunohistochemical stains and subsequent pathological analysis. SHP has the promise to enhance many of these aspects and combine morphological aspects and biochemical compositional information into a novel approach.

SHP has progressed at a faster pace than SCP, mostly for the reason that correlation with classical methods, that is, standard histopathology, is more straightforward, and parallel images from





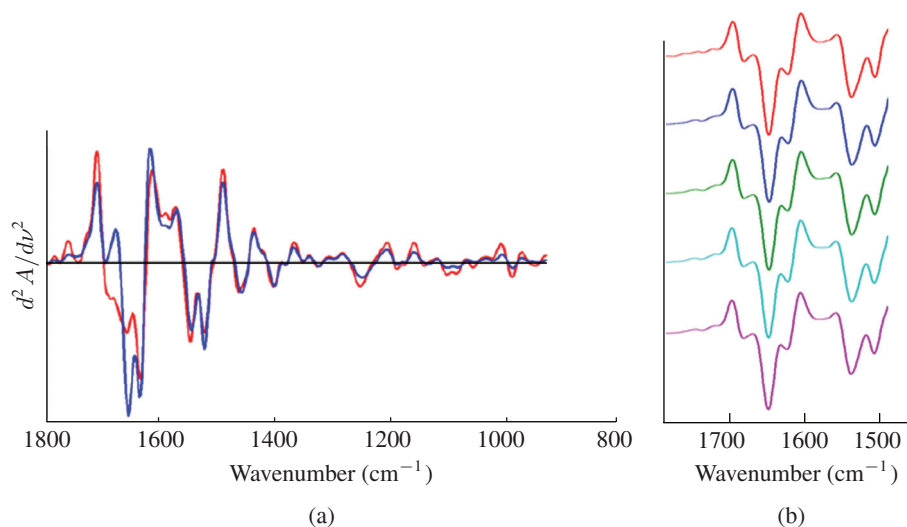
**Figure 11:** (a) Photomicrograph of an H&E-stained tissue section from a sample with cervical adenocarcinoma. (b) Overlay of H&E image and infrared pseudocolor map from hierarchical cluster analysis (HCA). See text for details.



**Figure 12:** Superposition of HCA-based detection of breast cancer micrometastases on H&E-stained tissue sections of sentinel lymph nodes.

histopathology and SHP can readily be compared (see Figures 11–13). In these figures, even a layman can perceive that the tissue morphology and architectural information available from classical histopathology translates directly into tissue structures revealed by SHP. Thus, it becomes obvious that the different biochemical composition indicated by tissue morphological variations is what is also detected in SHP. The similarity of SHP and H&E images allows a detailed “annotation,” that is, the correlation with tissue and cell morphological feature with corresponding spectral features, which, in turn, permits the training of diagnostic algorithms.

The course to be taken for successful SHP studies was first outlined in a series of pioneering papers by the group at the Robert Koch Institut, Berlin [21, 65–67], and involves the following key



**Figure 13:** (a) Comparison of 2nd derivative spectra of two different breast cancer micrometastases. (b) Comparison of 2nd derivative spectra of different breast cancer micrometastases while still paraffin-embedded.

steps: acquisition of very high S/N spectral data (the spatial resolution in the original studies were restricted by instrument performance), preprocessing including computation of 1st or 2nd derivatives and normalization to minimize instrumental and background artifacts, data presegmentation by unsupervised methods such as hierarchical cluster analysis (HCA), very careful annotation of diseased areas by a pathologist, and sufficiently large training datasets to construct a robust diagnostic algorithm. The diagnostic algorithm used in these initial studies was an artificial neural network (ANN) trained on thousands of spectra [67]. This work laid the ground rules in SHP and demonstrated that the patient-to-patient variations of the observed spectra were smaller than those due to disease classification or tissue type [21].

Over the past decade, tissue sections from bladder, bone, brain, breast, cartilage, cervix, colon, esophagus, kidney, liver, skin, spleen, teeth, thyroid, and a few others have been studied, mostly by SHP but more recently also by Raman spectral imaging. For a summary of all these studies, the reader is referred to some recent reviews [68–70]. Unfortunately, many of these studies were carried only up to the presegmentation (HCA) stage, since a sufficient number of samples from different patients with the same disease diagnosis often was not available. This aspect has changed drastically since the introduction of commercial tissue microarrays (TMAs). A TMA consists of between 50 and 120 individual tissue cores, each about 1–1.5 mm in diameter, which have been punched out of paraffin-embedded tissue blocks and may be seen as prototypical examples of a given cancer type. These cores themselves are embedded in paraffin and sectioned to standard thickness. Thus, one can purchase a TMA which contain samples from dozens of patients or disease stages. The use of TMAs was pioneered by the group around I. Levin at the NIH [71] and has been adopted by several groups [68], including the LSpD. Since these archived tissue sections are available with detailed disease diagnoses and often with disease outcome, the author believes that the future of SHP will be tied to TMA methodology for some time to come.



As in the case of SCP, fixation issues have been the source of many questions and criticism of SHP. Early studies [72, 73] have reported large spectral changes upon fixation, which could not be reproduced by other groups. However, there doubtlessly exist spectral changes caused by treatment with some of the harsher fixation protocols; here, only the two most commonly tissue treatment methods will be discussed. The least damaging way of tissue preparation is, of course, flash-freezing and cryo-sectioning the tissue section and performing spectral analysis immediately after thawing and drying the tissue section [21]. The other method involved formalin fixation and embedding the tissue section in paraffin, sectioning the tissue block, and subsequent deparaffination. These procedures, which are commonly used in standard histopathology laboratories, will certainly change protein structure; on the other hand, these changes are sufficiently small that immunohistochemical agents still recognize specific protein structures and binding sites. It is, of course, impossible to directly compare frozen and formalin-fixed and paraffin-embedded tissue sections, but if studies are carried out which do not mix the tissue preparation procedures, both methods yield comparable results. The equivalence of the two approaches was demonstrated when the author of this paper was a guest editor of a special journal issue, and coincidentally, two virtually identical infrared imaging studies on a rat model of glioblastoma multiform were submitted for publication [46, 47]. One study used frozen tissue section, the other formalin fixed and paraffin-embedded sections. Although there were, of course, spectral differences between the two tissue preparations, both studies arrived at images that were quite comparable and reached similar conclusions.

In SHP, different tissue types are frequently found in one section, such as white and gray brain matter, stroma, epithelial layers, inflammatory cells, and of course diseased tissue types. In general, infrared imaging techniques, combined with unsupervised multivariate methods, can detect the different tissue types and allow a biochemical interpretation of the spectral changes between tissue types. A typical example is the spectral detection of the maturation of squamous epithelial tissue *via* the accumulation of glycogen, which is polymer of glucose found as glycogen granules in the cytoplasm of mature squamous cells. Another example is the detection of different protein classes: the stroma and some other tissues contain collagen, which has a very characteristic infrared spectrum and can be realized in the spectra even by visual inspection. Keratin, a structural protein, is often detected in keratinizing squamous cell carcinomas, and the resulting “keratin pearls” were first described by Schultz et al., [74]. Similarly, parakeratosis (the deposition of keratin) in squamous epithelium was described by Wood et al., [75].

### 3.3.2. Cervical Adenocarcinoma

Infection of cervical tissue by the HPV virus is thought to start at the squamous-columnar junction (SCJ) and to proceed within the basal layer of squamous tissue and eventually lead to squamous cell carcinoma of the cervix. Thus, efforts at the LSpD are aimed at following the pathways of the virus in the cervical epithelium.

Several papers have reported the normal spectral changes within the layers of squamous tissue, and the distinction of the underlying stroma from the squamous tissue [75–77]. Spectral detection of cervical dysplasia and squamous carcinoma was reported by Steller et al. [77], but spectral

characterization of cervical adenocarcinoma has not yet been recorded in the literature, partially because of the rare occurrence of this disease. Here, we present selected results from a large section of tissue (*ca.*  $12 \times 2 \text{ mm}^2$ ) that contains normal squamous tissue, the SCJ, areas of normal columnar (glandular) epithelium, and large areas of cervical adenocarcinoma. In particular, we wish to focus on the abundance of inflammatory cells in the vicinity of the adenocarcinoma.

In both squamous and glandular cancer of epithelium (carcinomas and adenocarcinomas, resp.), inflammatory cells are frequently observed. Steller et al., [77] reported spectral changes due to these cells in the stroma underlying a squamous cell carcinoma, but the infiltration of inflammatory cells was relatively mild. Here, we report results on a tissue section that is heavily inflamed; in fact, bands of inflammatory cells can be detected visually in Figure 11(a). These inflammatory cells, shown in light blue and red hues in Figure 11(b), are easily separated by hierarchical cluster analysis from the surrounding stroma. Normal, uninflamed stroma is shown as the transparent regions on the right side of Figure 11(b), and the adenocarcinoma is shown in green. The purple layer denotes the body of the glandular cells, excluding the layer of nuclei closest to the basement membrane. The cell nuclei (green) underlying the purple layer cluster with the cancerous cells, indicating that these cells are abnormal. Like all images based on hierarchical cluster analysis (HCA), no reference dataset is utilized in this image reconstruction process; rather, the image is based entirely on spectral similarities. Spectral classes obtained from HCA images, and pathological diagnoses of the cluster-based regions, have been used to train diagnostic algorithms for the automatic diagnosis of tissue sections.

The tissue sections from this sample of cervical adenocarcinoma have presented significant difficulties for the interpretation of the spectral results, the unsupervised cluster analysis, and the training of diagnostic algorithms, due to the abundance of inflammatory cells. However, the separation of stroma and both squamous and glandular epithelial tissue is trivial by SHP, as is the distinction of the different layers of the squamous tissue. Here, the spectral changes are so reproducible that a diagnostic network, such as an ANN, can be trained to separate these tissue types. However, the regions of inflammation do present some difficulties. First, inflammatory cells are small and nearly spherical in shape and present strong RMie scattering. Once corrected, it appears that the spectral characteristics of these immune cells change with the proximity to the cancer. In some areas of this tissue section, the spectra of the inflammatory cells are nearly indistinguishable from the spectra of the adenocarcinoma cells, such as those in the red areas of Figure 11, while in other (light blue areas in Figure 11), the inflammatory cells separate quite nicely from the cancerous regions. It is possible that the two classes of inflammatory cells are biochemically different: we have shown before that the activation of B lymphocytes can be detected by SHP; furthermore, we have indications that the spectra of certain immune cells (specifically phagocytic histiocytes) change upon interacting with and destroying cancer cells. Thus, the similarity of the adenocarcinoma cells and the histiocytes may have biochemical origins. The only way to address these problems is *via* immunohistochemical staining to further identify the cell types in the sample.

### 3.3.3. Breast Cancer Micrometastases in Lymph Nodes

In this section, we shall present results of infrared imaging studies, combined with hierarchical cluster analysis (HCA), of lymph node tissue sections infiltrated by breast cancer micrometastases [13, 18, 28, 78, 79]. This is a significant medical problem since treatment depends on the presence or absence of cancerous cells in the sentinel lymph nodes, where they form metastatic tumors. Metastases less than

2 mm in size are referred to as micrometastases, which tend to form in the subcapsular sinuses of the lymph nodes. In this study, over 50  $1\text{ mm} \times 1\text{ mm}$  spectral images were collected, each consisting of 25,600 spectra. Here, we were particularly interested whether or not HCA can reliably segment the raw datasets into spectra of capsule, lymphocytes, metastatic cancer, and so forth. In particular, we wished to establish that the spectra of the metastatic cancers were sufficiently similar to permit their detection by a trained, diagnostic algorithm.

Figure 12 shows three typical  $1 \times 1\text{ mm}^2$  images of H&E-stained lymph node tissue sections. In each of them, the capsule of the lymph node, composed of fibroconnective tissue, is shown in pink, whereas the lymphocytes within the body of the lymph node appear dark violet. Depending on the scarcity of the tissue section, the lymphocytes may exhibit strong RMie scattering. The dark red regions in the three panels of Figure 12 are superpositions of the infrared spectral regions indicative of metastatic cancer cells. These regions coregister exactly with the regions of the lymph nodes that show morphological abnormalities consistent with invasive breast cancer cells. Thus, it appeared that the automatic detection and diagnosis of breast cancer micrometastases by infrared spectral imaging methods should be possible using a trained diagnostic algorithm.

However, in spite of the excellent discrimination of the cancerous regions from the surrounding lymphocytes, the original studies showed poor similarity between the metastatic cancer spectra from different samples [78]. This is shown in Figure 13(a), which depicts large spectral differences in the amide I region of the breast cancer regions. Below *ca.*  $1480\text{ cm}^{-1}$ , the spectra are nearly identical, aside from an intensity scale factor. This spectral difference was soon realized to be an artifact due to interference with dispersive line shapes. When spectra from five different micrometastases were collected from tissue sections still embedded in paraffin [79], nearly identical spectra were observed, see Figure 13(b). (Spectral maps of tissue section, still embedded in paraffin, were first reported by the Reims group [80]. They showed that very similar maps could be obtained if the few strong vibrations of paraffin are properly accounted for). In this case, there is much better matching of the refractive index of areas occupied by tissue and areas devoid of sample; index matching enormously reduces the incidence of scattering effects that mix absorptive and reflective line shapes as discussed before (see (3.2) and (3.3)). The dispersive line shapes predominantly affect the high wavenumber side of the amide I band, and cause enormous changes in the 2nd derivative spectra; see Figure 13(a). After correction for these effects, the spectra of micrometastases were found to be sufficiently similar that diagnoses with trained algorithms is possible. These results explained why, in the past, best SHP images were obtained if the amide I spectral region was excluded [81].

Since the seminal papers on the origin of dispersive band contaminations were published [4, 17], these difficulties are understood and are being corrected routinely. This has opened the possibilities for a wide application of SHP for diagnostic purposes, since the major variance of the spectral data has been eliminated. Several research groups have used tissue microarrays to increase the size of datasets, and have found that the spectral signatures for similar disease states are remarkably reproducible. In dense and cohesive tissues sections, for example, from the colon [82] or liver [83], scattering effects were much smaller and thus allowed early studies to be carried out without interference from dispersive band shapes. On the other hand, very sparse tissue, such as normal lung tissue, consists of very few “filaments” of aligned cells which exhibit enormous band shape distortion; in these, no reliable data can be obtained without scatter correction.

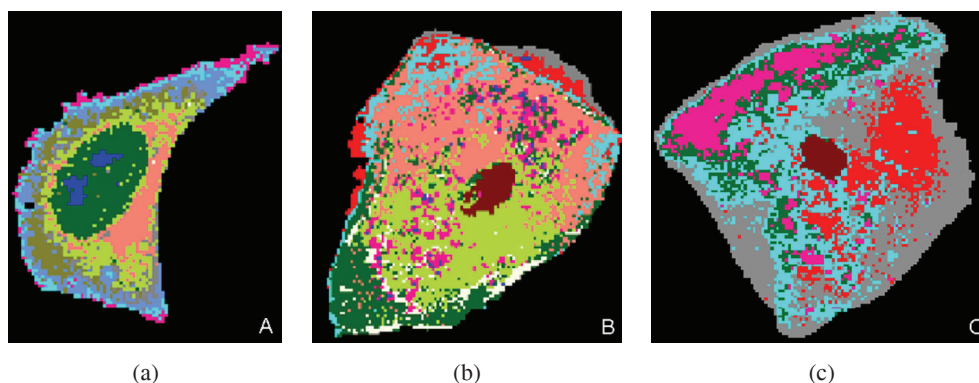
### 3.3.4. Diagnostic Algorithms and Future Prospects of SHP

The inherent sensitivity of vibrational spectral fingerprints toward changes in biochemical composition of tissue pixels makes SHP an ideal candidate for medical diagnostic imaging. With recent advances in data preprocessing, the increasing number of research groups involved in the field, and the generally good agreement between spectral and pathological results, it appears that SHP is poised to enter the mainstream diagnostic arena. The major obstacle to a broad application of SHP is, in the eyes of the authors, the difficulty in obtaining sufficient and reliable annotations to train diagnostic algorithms.

The severity of this problem was first indicated in a review chapter authored by Stone et al. [84], who has been on the forefront of Raman spectral histopathology and *in vivo* Raman diagnostics. He reported that a consensus diagnosis by a group of three pathologists was obtained in only about 30% of all cases presented to them. Anecdotal evidence from collaborators of the authors has indicated that the same tissue section can produce different pathological diagnoses at different times. The approach that appears most successful in obtaining accurate and reproducible annotation involves the use of high-resolution digital images of the H&E-stained tissue sections onto which HCA images can be superimposed. This step requires that the digital H&E and the HCA images are exactly registered and can be zoomed together. This approach permits the pathologist to annotated the spectral images based on single-cell features; that is, the pathologist can select the most typical features in a tissue section and correlate it to spectral features at the level of one or a few cells. Spectra from the pathologist-annotated tissue areas are subsequently extracted from the datasets and used to train diagnostic algorithms. As indicated above, it is imperative that tissue samples from different patients are used in this training phase, because small but systematic differences may exist between the extracted spectra from different patients. At present, it is not known whether or not these differences can be later correlated to special aspects of disease type and progression. These questions can only be answered by parallel spectral and immunohistochemical studies.

At the LSpD, the diagnostic methodology of choice has been artificial neural nets (ANNs) in various implementations. ANNs are self-learning methods modeled after the neural interactions in the human brain. They can be used as binary (two-class) classifiers or to differentiate more than two input classes. They can be “stacked” to operate as hierarchical networks, for example, as several consecutive binary classifiers. Recent studies comparing them to other multivariate classifiers have established that they perform at a similar level of predictive accuracy as, for example, the “random forest” algorithm. The authors’ use of ANNs has been frequently criticized by reviewers of proposals and publications emerging from the LSpD, and “overtraining” and the well-publicized failures of early applications of ANNs are repeatedly cited. These two points will be addressed briefly. First, there exist well-established rules in bioinformatics on the size of training and validation sets required to produce reliable algorithms; failure to adhere to these rules certainly will produce algorithms that can be hopelessly overtrained. Yet, any discriminatory algorithm, including the operation of the human brain, suffers from this shortcoming. In the latter case, the subjectivity of classical histopathology is certainly a manifestation of insufficient training.

Secondly, the well-published failures of discriminate algorithms of any kind can be attributed to insufficient training, such as omitting entire classes of possible inputs or conditions. Also, many tasks to which discriminate algorithms were applied, that is, the morphological discrimination of abnormal from normal exfoliated cells or facial recognition, require the translation of certain features, be it the nuclear-to-cytoplasm ratio of a cell or the height-versus-width ratio of a face, to be collected in metrics which are



**Figure 14:** Raman images, reconstructed *via* hierarchical cluster analysis from raw hyperspectral datasets of (a) a cultured HeLa cell, (b) an exfoliated oral mucosa cell, and (c) a squamous cell from the distal urethra. See text for details.

subsequently analyzed by the ANN. It is conceivable that these metrics lacked specificity for the task at hand and that the discrimination failed, not because of shortcomings of the discriminatory method but a shortcoming of the input data. In SCP and SHP, the form of spectral results—one-dimensional vectors of intensity data at given wavenumber point—is ideal for an ANN or “random forest” and does not involve the constructing of metrics.

### 3.4. Raman Spectral Images of Squamous Cells

The review of Raman data in this paper is somewhat biased toward cellular imaging, rather than diagnostic applications. All the diagnostic work carried out at the LSpD utilizes infrared microspectroscopy which offers much higher speed and a spatial resolution of about the size of a cell. Raman imaging, on the other hand, offers much higher spatial resolution and is therefore highly sensitive to detect biological changes at a much smaller (subcellular) level. Thus, this paper shall concentrate on cell imaging applications of Raman microspectroscopy; however, other groups have used Raman microspectroscopy for diagnostic purposes as well by defocusing the laser beam to larger spot size (2–5  $\mu\text{m}$  in diameter) and sacrificing spatial information. Here, we wish to report information that supplements the cytological efforts at the LSpD by developing label-free methods to visualize cellular organization. Figure 14 shows Raman images of three different squamous epithelial cells obtained by raster scanning the laser beam, focused to a spot of about 300 nm diameter, over the cell and collecting an entire Raman spectrum from each spot. Subsequently, unsupervised hierarchical cluster analysis was used to convert the hyperspectral dataset into a pseudocolor image.

Figure 14(a) shows a Raman image of a cultured cervical cancer (HeLa) cell [85]. It was reconstructed from a dataset containing *ca.* 10,000 spectra *via* HCA. The segmentation of spectra into different classes is based on their similarity; that is, pixels shown in the same color result from very similar spectra. The cellular details available in this image are astounding: the large nucleus (typical for an actively growing cell, see Section 1), shown in dark green, is easily distinguished from the



cytoplasm. Furthermore, two nucleoli shown in dark blue are detected within the nucleus. The spectra of the nucleus (mostly protein and DNA) and nucleolus (mostly protein and RNA) differ very minutely; in a single spectrum, such differences would not be significant. However, HCA detects correlated spectral differences which have physical significance; thus, the mean cluster spectra between nuclei and nucleoli show distinct spectral changes which could be interpreted biochemically [85] and were reproducible between different cells imaged.

The cytoplasm also shows small but significant spectral differences. Using a mitochondria-specific stain, we were able to assign the yellowish-green and salmon-colored clusters in the perinuclear region to be due to high abundance of mitochondria. This was accomplished by adding the stain to the cell in aqueous environment and, without reregistering the sample, rescanning the cell using the Raman microscope as a confocal fluorescence microscope. Such experiments can be carried out staining for other specific cell organelles [86].

The image in Figure 14(b) is from a mature oral mucosa cell with a pyknotic nucleus [43]. Spectra from the pyknotic nucleus separate readily from those of the cytoplasm. Within the cytoplasm, an interesting feature is observed, represented by the occurrence of the purple spots. The mean cluster spectra of these spots indicate a superposition of cytoplasmic protein and phospholipids, indicated by strong aliphatic CH<sub>2</sub> deformation and stretching modes at *ca.* 1445 and 2950 cm<sup>-1</sup>. Natural phospholipid spots could be due to intracellular lipid droplets or due to structures such as the Golgi apparatus, vacuoles, or multilamellar vesicles. When exposing a cell to deuterated phospholipids [87] (e.g., liposomes produced entirely from deuterated phospholipids), we found that the deuterated lipids equilibrate with the naturally occurring lipids within the cytoplasm. Thus, Raman imaging can be used to study transport and exchange phenomena which are difficult to perform by other imaging methods.

Finally, Figure 14(c) depicts a squamous cell from the distal urethra. These cells constitute the majority of cells found in urine cytology. Like most stratified squamous cells, they accumulate glycogen upon maturation. The glycogen is not distributed uniformly within the cytoplasm but forms granules which can be visualized by Raman spectral imaging. In Figure 14(c), the areas shown in red exhibit the signatures of cytoplasmic protein and glycogen, which can be identified by comparison with reference spectra. The purple areas, as in Figure 14(b), are due to phospholipids.

Raman spectral imaging bears the advantage over other cellular imaging methods in that no specific label or dye needs to be added to the cells but that the image is based on an inherent vibrational spectroscopic fingerprint pattern that can be detected with a spatial resolution similar to that of confocal fluorescence microscopy. Sample preparation is trivial for confocal Raman microscopy: a live or fixed cell is grown or placed on a CaF<sub>2</sub> substrate, immersed in buffer solution which is brought in contact with a water immersion objective. This method produces information as closely as possible to an “undisturbed,” noninvasive approach. In particular, the possibility of Raman imaging methods to be carried out in aqueous surroundings on live cells opens the possibility to monitor cells for later medical use, for example, stem cells. Indeed, the early differentiation steps of stem cell colonies [88–93] and embryonic bodies have been detected. Recently, Notingher reported Raman spectra of beating cardiomyocytes [94].

In addition, a number of studies have appeared in the literature that used Raman microspectroscopy (at lower spatial resolution, as pointed out above) for Raman spectral cytopathology. Here, the



ability to observe live cells in their native environment and the high selectivity of vibrational spectroscopy bear enormous promise for the application of Raman spectroscopy for identifying cell types in blood and to use this technique to cell sorting applications, for example, for the detection and isolation of circulating tumor cells. Efforts in this direction have been spearheaded by the Jena (Germany) group around Popp, who has reported the identification of different cell types in blood, even under flow conditions [95–97]. In addition, a study to differentiate cells with different viral infections has been published [56]. Also, the spectral differences due to overexpression of an oncogene were recently reported [98]. At the LSpD, the uptake of drug-loaded targeted and nontargeted nanoparticles, and the subsequent release of the drug inside the cell, has been studied [99, 100]. These studies are aimed at a readership in pharmaceutical science and will not be discussed here any further.

Raman spectroscopy has been used for *in vivo* and *ex vivo* diagnostic applications. In the former category, the efforts for *in vivo* diagnosis of cervical [49, 101] cancer and the work on *in vivo* detection of esophageal and bladder cancer [102–104] have defined the enormous potential of fiberoptic-based Raman diagnostics. Stone's group at the Gloucestershire Hospital (UK) was instrumental in these *in vivo* efforts; they also have explored the possibility of intraoperative assessment of tissue using Raman spectroscopy [105, 106]. *Ex vivo* Raman images of tissue have been collected, using mostly near-IR excitation, for liver [107] and lung [108] and have revealed information similar to images described Section 3.3, albeit with higher spatial resolution.

At this point, it appears that both infrared and Raman microspectroscopy have comparable diagnostic sensitivity and are poised to enter the medical diagnostic fields. The two techniques have somewhat different strengths (speed, ability to measure in aqueous environment) and may complement each other in the same way classical Raman and infrared spectroscopy are complimentary. The next frontier in this field promise to be nonlinear techniques, which have shown to increase data acquisition by orders of magnitude [109, 110].

#### 4. Conclusions

In this paper, aspects of infrared and Raman spectral imaging and medical diagnostics have been presented. Although there are major differences to classical spectroscopy of biomolecules, due to the size and complexity of the systems reported in this paper, it is important to point out that this work is based on and relies on decades of research in biospectroscopy. The aspects most different from “classical” biospectroscopy are the heavy reliance of this new research endeavor on mathematical methods for data analysis, partially necessitated by the fact that the amount of data collected often measures in the gigabyte regime, and visual interpretation of such an amount of data is impossible. Furthermore, the multivariate methods of analysis are highly suitable for extracting small, correlated spectral differences that are often smaller than and buried in uncorrelated noise level.

Thus, the authors hope that this research not only advances the field of medical diagnostics by spectral methods but also helps to usher in new ways to look and process spectral data.

#### Acknowledgment

Partial support of this research by grants CA 090346, CA 111330 and CA 153148 from the National Institutes of Health during the past five years is gratefully acknowledged.

## References

- [1] E. R. Blout and R. C. Mellors, "Infrared spectra of tissues," *Science*, vol. 110, no. 2849, pp. 137–138, 1949.
- [2] D. L. Woernley, "Infrared absorption curves for normal and neoplastic tissues and related biological substances," *Cancer Research*, vol. 12, no. 7, pp. 516–523, 1952.
- [3] B. Mohlenhoff, M. Romeo, M. Diem, and B. R. Wood, "Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy," *Biophysical Journal*, vol. 88, no. 5, pp. 3635–3640, 2005.
- [4] P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas, and P. Gardner, "Resonant Mie scattering in infrared spectroscopy of biological materials—understanding the 'dispersion artefact'," *Analyst*, vol. 134, no. 8, pp. 1586–1593, 2009.
- [5] M. J. Romeo, S. Boydston-White, C. Matthäus et al., "Infrared and Raman microspectroscopic studies of individual human cells," in *Vibrational Spectroscopy for Medical Diagnosis*, M. Diem, P. R. Griffiths, and J. M. Chalmers, Eds., pp. 27–70, John Wiley & Sons, Chichester, UK, 2008.
- [6] M. Romeo and M. Diem, "Correction of dispersive line shape artifact observed in diffuse reflection infrared spectroscopy and absorption/reflection (transflection) infrared micro-spectroscopy," *Vibrational Spectroscopy*, vol. 38, no. 1-2, pp. 129–132, 2005.
- [7] P. Bassan, H. J. Byrne, J. Lee et al., "Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells," *Analyst*, vol. 134, no. 6, pp. 1171–1175, 2009.
- [8] H. J. Gully-Stahl, A. J. Sommer, and A. P. Evan, "Evanescent wave imaging," in *Vibrational Spectroscopic Imaging for Biomedical Applications*, G. Srinivasan, Ed., pp. 99–121, McGraw Hill, New York, NY, USA, 2010.
- [9] M. Diem, M. Romeo, C. Matthäus, M. Miljkovic, L. Miller, and P. Lasch, "Comparison of Fourier transform infrared (FTIR) spectra of individual cells acquired using synchrotron and conventional sources," *Infrared Physics and Technology*, vol. 45, no. 5-6, pp. 331–338, 2004.
- [10] P. Lasch, A. Pacifico, and M. Diem, "Spatially resolved IR microspectroscopy of single cells," *Biopolymers*, vol. 67, no. 4-5, pp. 335–338, 2002.
- [11] C. Otto and J. Greve, "Progress in instrumentation for Raman micro-spectroscopy and Raman imaging for cellular biophysics," *The Internet Journal of Vibrational Spectroscopy*, vol. 2, no. 3, 1998.
- [12] A. I. Mazur, E. J. Marcsisin, B. Bird et al., "Evaluating different fixation protocols for spectral cytopathology, part I," *Analytical Chemistry*, vol. 84, no. 3, pp. 1259–1266, 2011.
- [13] B. Bird, M. Miljkovic, M. J. Romeo et al., "Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology," *BMC Clinical Pathology*, vol. 8, no. 1, article 8, pp. 1–14, 2008.
- [14] M. Miljković, T. Chernenko, M. J. Romeo, B. Bird, C. Matthäus, and M. Diem, "Label-free imaging of human cells: algorithms for image reconstruction of Raman hyperspectral datasets," *Analyst*, vol. 135, no. 8, pp. 2002–2013, 2010.
- [15] R. K. Reddy and R. Bhargava, "Accurate histopathology from low signal-to-noise ratio spectroscopic imaging data," *Analyst*, vol. 135, no. 11, pp. 2818–2825, 2010.
- [16] A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
- [17] P. Bassan, A. Kohler, H. Martens et al., "Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples," *Analyst*, vol. 135, no. 2, pp. 268–277, 2010.
- [18] B. Bird, M. Miljković, and M. Diem, "Two step resonant Mie scattering correction of infrared micro-spectral data: human lymph node tissue," *Journal of Biophotonics*, vol. 3, no. 8-9, pp. 597–608, 2010.

- [19] M. Diem, B. Bird, and M. Miljković, *Phase Correction to Compensate for Reflective Distortions of Optical Spectra*, U.P. Office, 2011.
- [20] J. M. Schubert, A. I. Mazur, B. Bird, M. Miljković, and M. Diem, "Single point vs. mapping approach for spectral cytopathology (SCP)," *Journal of Biophotonics*, vol. 3, no. 8-9, pp. 588–596, 2010.
- [21] P. Lasch, "Computergestuetzte Bildrekonstruktion auf Basis FTIR-mikrospektrometrischer Daten humaner Tumoren," in *Humanmedizin*, p. 124, Freie Universitaet Berlin, Berlin, Germany, 1999.
- [22] J. M. Chalmers and M. W. Mackenzie, "Some industrial applications of FT-IR diffuse reflectance spectroscopy," *Applied Spectroscopy*, vol. 39, no. 4, pp. 634–641, 1985.
- [23] D. A. Heaps and P. R. Griffiths, "Effect of molecular spacers on surface-enhanced attenuated total reflection infrared spectra of bulk liquids," *Vibrational Spectroscopy*, vol. 41, no. 2, pp. 221–224, 2006.
- [24] D. A. Heaps and P. R. Griffiths, "Band shapes in the infrared spectra of thin organic films on metal nanoparticles," *Vibrational Spectroscopy*, vol. 42, no. 1, pp. 45–50, 2006.
- [25] M. Romeo, B. Mohlenhoff, and M. Diem, "Infrared micro-spectroscopy of human cells: causes for the spectral variance of oral mucosa (buccal) cells," *Vibrational Spectroscopy*, vol. 42, no. 1, pp. 9–14, 2006.
- [26] H. C. van de Hulst, *Light Scattering by Small Particles*, Dover, Mineola, NY, USA, 1981.
- [27] P. Walstra, "Approximation formulae for the light scattering coefficient of dielectric spheres," *British Journal of Applied Physics*, vol. 15, no. 12, pp. 1545–1552, 1964.
- [28] B. Bird, M. Miljković, N. Laver, and M. Diem, "Spectral detection of micro-metastases and individual metastatic cells in lymph node histology," *Technology in Cancer Research and Treatment*, vol. 10, no. 2, pp. 135–144, 2011.
- [29] P. R. Griffiths and J. A. De Haseth, "Fourier transform infrared spectrometry," in *Chemical Analysis*, P. J. Elving and J. D. Winnefordner, Eds., vol. 83, John Wiley & Sons, New York, NY, USA, 1986.
- [30] L. Mertz, "Auxiliary computation for Fourier spectrometry," *Infrared Physics*, vol. 7, no. 1, pp. 17–23, 1967.
- [31] P. T. T. Wong, R. K. Wong, T. A. Caputo, T. A. Godwin, and B. Rigas, "Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 24, pp. 10988–10992, 1991.
- [32] P. Wong, S. Lacelle, M. Fung, K. Fung, and M. Senterman, "Characterization of exfoliated cells and tissues from human endocervix and ectocervix by FTIR and ATR/FTIR spectroscopy," *Biospectroscopy*, vol. 1, no. 5, pp. 357–364, 1995.
- [33] M. A. Cohenford, T. A. Godwin, F. Cahn, P. Bhandare, T. A. Caputo, and B. Rigas, "Infrared spectroscopy of normal and abnormal cervical smears: evaluation by principal component analysis," *Gynecologic Oncology*, vol. 66, no. 1, pp. 59–65, 1997.
- [34] M. A. Cohenford and B. Rigas, "Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 26, pp. 15327–15332, 1998.
- [35] G. N. Papanicolaou and H. F. Traut, "Smear diagnosis of carcinoma of the cervix," *American Journal of Obstetrics & Gynecology*, vol. 42, no. 2, pp. 193–206, 1941.
- [36] G. N. Papanicolaou, *The Epithelium of Women's Reproductive Organs*, The Commonwealth Fund, New York, NY, USA, 1948.
- [37] A. M. van Driel-Kulker, J. J. Ploem-Zaaijer, M. van der Zwan-van der Zwan, and H. J. Tanke, "A preparation technique for exfoliated and aspirated cells allowing different staining procedures," *Analytical and Quantitative Cytology*, vol. 2, no. 4, pp. 243–246, 1980.

- [38] M. Diem, Unpublished results.
- [39] B. Bird, M. J. Romeo, M. Diem, K. Bedrossian, N. Laver, and S. Naber, "Cytology by infrared micro-spectroscopy: automatic distinction of cell types in urinary cytology," *Vibrational Spectroscopy*, vol. 48, no. 1, pp. 101–106, 2008.
- [40] M. Diem, C. Matthäus, T. Chernenko et al., "Infrared and Raman spectroscopy and spectral imaging of individual cells," in *Infrared and Raman Spectroscopic Imaging*, R. Salzer and H. W. Siesler, Eds., pp. 173–202, Wiley-VCH, Weinheim, Germany, 2009.
- [41] M. J. Romeo, B. Mohlenhoff, M. Jennings, and M. Diem, "Infrared micro-spectroscopic studies of epithelial cells," *Biochimica et Biophysica Acta*, vol. 1758, no. 7, pp. 915–922, 2006.
- [42] M. Diem, L. Chiriboga, P. Lasch, and A. Pacifico, "IR spectra and IR spectral maps of individual normal and cancerous cells," *Biopolymers*, vol. 67, no. 4-5, pp. 349–353, 2002.
- [43] C. Matthäus et al., "Infrared and Raman microscopy in cell biology," in *Methods in Cell Biology, Volume 89: Biophysical Tools for Biologists*, J. J. Correia and H. W. Detrich, Eds., vol. 2, pp. 275–308, Elsevier, 2008.
- [44] J. W. Brauner, C. R. Flach, and R. Mendelsohn, "A quantitative reconstruction of the amide I contour in the IR spectra of globular proteins: from structure to spectrum," *Journal of the American Chemical Society*, vol. 127, no. 1, pp. 100–109, 2005.
- [45] M. Jackson and H. H. Mantsch, "The use and misuse of FTIR spectroscopy in the determination of protein structure," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 30, no. 2, pp. 95–120, 1995.
- [46] K. R. Bamberg, E. Schültke, B. R. Wood et al., "A Fourier transform infrared microspectroscopic imaging investigation into an animal model exhibiting glioblastoma multiforme," *Biochimica et Biophysica Acta*, vol. 1758, no. 7, pp. 900–907, 2006.
- [47] N. Amharref, A. Beljebbar, S. Dukic et al., "Brain tissue characterisation by infrared imaging in a rat glioma model," *Biochimica et Biophysica Acta*, vol. 1758, no. 7, pp. 892–899, 2006.
- [48] K. Papamarkakis, B. Bird, J. M. Schubert et al., "Cytopathology by optical methods: spectral cytopathology of the oral mucosa," *Laboratory Investigation*, vol. 90, no. 4, pp. 589–598, 2010.
- [49] A. Robichaux-Viehoever, E. Kanter, H. Shappell, D. Billheimer, H. Jones, and A. Mahadevan-Jansen, "Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia," *Applied Spectroscopy*, vol. 61, no. 9, pp. 986–993, 2007.
- [50] J. M. Schubert, B. Bird, K. Papamarkakis et al., "Spectral cytopathology of cervical samples: detecting cellular abnormalities in cytologically normal cells," *Laboratory Investigation*, vol. 90, no. 7, pp. 1068–1077, 2010.
- [51] G. R. Ogden, J. G. Cowpe, and M. W. Green, "The effect of distant malignancy upon quantitative cytologic assessment of normal oral mucosa," *Cancer*, vol. 65, no. 3, pp. 477–480, 1990.
- [52] M. G. C. T. van Oijen and P. J. Slootweg, "Oral field cancerization: carcinogen-induced independent events or micrometastatic deposits?" *Cancer Epidemiology Biomarkers and Prevention*, vol. 9, no. 3, pp. 249–256, 2000.
- [53] H. zur Hausen, *Infections Causing Human Cancer*, Wiley-VCH, Weinheim, Germany, 2006.
- [54] J. M. Schubert, "Spectral cytology of human oral and cervical samples," in *Chemistry & Chemical Biology*, Northeastern University, Boston, Mass, USA, 2011.
- [55] Prevention, C.f.D.C.a, Prevalence of high-risk and low-risk strains of HPV (2003-2004), 2007.
- [56] K. Ostrowska, "Vibrational spectroscopy for cervical cytology," in *School of Physics*, Dublin Institute of Technology, Dublin, Ireland, 2011.
- [57] S. Boydston-White, M. J. Romeo, T. Chernenko, A. Regina, M. Miljković, and M. Diem, "Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: principal component and artificial neural network analysis," *Biochimica et Biophysica Acta*, vol. 1758, no. 7, pp. 908–914, 2006.



- [58] E. J. Marcsisin, C. M. Uttero, M. Miljković, and M. Diem, "Infrared microspectroscopy of live cells in aqueous media," *Analyst*, vol. 135, no. 12, pp. 3227–3232, 2010.
- [59] R. Gasper, T. Mijatovic, R. Kiss, and E. Goormaghtigh, "Time dependence of cellular chemical changes induced in prostate PC-3 cancer cells by two structurally related cardenolides monitored by fourier transform infrared (FT-IR) spectroscopy," *Applied Spectroscopy*, vol. 65, no. 6, pp. 584–594, 2011.
- [60] K. R. Flower, I. Khalifa, P. Bassan et al., "Synchrotron FTIR analysis of drug treated ovarian A2780 cells: an ability to differentiate cell response to different drugs?" *Analyst*, vol. 136, no. 3, pp. 498–507, 2011.
- [61] E. Gazi, M. Baker, J. Dwyer et al., "A correlation of FTIR spectra derived from prostate cancer biopsies with gleason grade and tumour stage," *European Urology*, vol. 50, no. 4, pp. 750–761, 2006.
- [62] N. S. Ozek, S. Tuna, A. E. Erson-Bensan, and F. Severcan, "Characterization of microRNA-125b expression in MCF7 breast cancer cells by ATR-FTIR spectroscopy," *Analyst*, vol. 135, no. 12, pp. 3094–3102, 2010.
- [63] S. E. Holton, M. J. Walsh, and R. Bhargava, "Subcellular localization of early biochemical transformations in cancer-activated fibroblasts using infrared spectroscopic imaging," *Analyst*, vol. 136, no. 14, pp. 2953–2958, 2011.
- [64] G. Schiffer et al., "Application of FTIR spectroscopy in antibacterial drug research," in *Workshop on FTIR Spectroscopy in Microbiological and Medical Diagnostic*, Robert Koch Institute, Berlin, Germany, 2002.
- [65] P. Lasch, W. Haensch, E. N. Lewis, L. H. Kidder, and D. Naumann, "Characterization of colorectal adenocarcinoma sections by spatially resolved FT-IR microspectroscopy," *Applied Spectroscopy*, vol. 56, no. 1, pp. 1–9, 2002.
- [66] P. Lasch, W. Haensch, D. Naumann, and M. Diem, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis," *Biochimica et Biophysica Acta*, vol. 1688, no. 2, pp. 176–186, 2004.
- [67] P. Lasch, M. Diem, W. Hänsch, and D. Naumann, "Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging," *Journal of Chemometrics*, vol. 20, no. 5, pp. 209–220, 2006.
- [68] F. N. Pounder and R. Bhargava, "Toward automated breast histopathology using Mid-IR spectroscopic imaging," in *Vibrational Spectroscopic Imaging for Biomedical Applications*, G. Srinivasa, Ed., pp. 1–26, McGraw Hill, New York, NY, USA, 2010.
- [69] M. J. Romeo, R. K. Dukor, and M. Diem, "Introduction to spectral imaging, and applications to diagnosis of lymph nodes," in *Vibrational Spectroscopy for Medical Diagnosis*, M. Diem, P. R. Griffiths, and J. M. Chalmers, Eds., pp. 1–26, John Wiley & Sons, Chichester, UK, 2008.
- [70] M. J. Romeo, S. Boydston-White, C. Matthäus et al., "Vibrational microspectroscopy of cells and tissues," in *Biomedical Vibrational Spectroscopy*, P. Lasch and J. Kneipp, Eds., pp. 121–147, Wiley-Interscience, Hoboken, NJ, USA, 2008.
- [71] D. C. Fernandez, R. Bhargava, S. M. Hewitt, and I. W. Levin, "Infrared spectroscopic imaging for histopathologic recognition," *Nature Biotechnology*, vol. 23, no. 4, pp. 469–474, 2005.
- [72] E. Ó Faoláin, M. B. Hunter, J. M. Byrne et al., "A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy," *Vibrational Spectroscopy*, vol. 38, no. 1–2, pp. 121–127, 2005.
- [73] H. H. Mantsch and M. Jackson, "Molecular spectroscopy in biondiagnostics (from Hippocrates to Herschel and beyond)," *Journal of Molecular Structure*, vol. 347, pp. 187–206, 1995.

- [74] C. P. Schultz, K. Z. Liu, P. D. Kerr, and H. H. Mantsch, "In situ infrared histopathology of keratinization in human oral/oropharyngeal squamous cell carcinoma," *Oncology Research*, vol. 10, no. 5, pp. 277–286, 1998.
- [75] B. R. Wood, L. Chiriboga, H. Yee, M. A. Quinn, D. McNaughton, and M. Diem, "Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium," *Gynecologic Oncology*, vol. 93, no. 1, pp. 59–68, 2004.
- [76] L. Chiriboga, P. Xie, H. Yee et al., "Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in the human cervix," *Biospectroscopy*, vol. 4, no. 1, pp. 47–53, 1998.
- [77] W. Steller, J. Einenkel, L. C. Horn et al., "Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging," *Analytical and Bioanalytical Chemistry*, vol. 384, no. 1, pp. 145–154, 2006.
- [78] B. Bird, M. Romeo, N. Laver, and M. Diem, "Spectral detection of micro-metastases in lymph node histo-pathology," *Journal of Biophotonics*, vol. 2, no. 1-2, pp. 37–46, 2009.
- [79] B. Bird, K. Bedrossian, N. Laver, M. Miljković, M. J. Romeo, and M. Diem, "Detection of breast micro-metastases in axillary lymph nodes by infrared micro-spectral imaging," *Analyst*, vol. 134, no. 6, pp. 1067–1076, 2009.
- [80] E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard, and M. Manfait, "Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies," *Analyst*, vol. 133, no. 2, pp. 197–205, 2008.
- [81] M. J. Romeo and M. Diem, "Infrared spectral imaging of lymph nodes: strategies for analysis and artifact reduction," *Vibrational Spectroscopy*, vol. 38, no. 1-2, pp. 115–119, 2005.
- [82] P. Lasch and D. Naumann, "FT-IR microspectroscopic imaging of human carcinoma thin sections based on pattern recognition techniques," *Cellular and Molecular Biology*, vol. 44, no. 1, pp. 189–202, 1998.
- [83] L. Chiriboga, H. Yee, and M. Diem, "Infrared spectroscopy of human cells and tissue. Part VI: a comparative study of histopathology and infrared microspectroscopy of normal, cirrhotic, and cancerous liver tissue," *Applied Spectroscopy*, vol. 54, no. 1, pp. 1–8, 2000.
- [84] N. Stone, C. Kendall, and H. Barr, "Raman spectroscopy as a potential tool for early diagnosis of malignancies in esophageal and bladder tissues," in *Vibrational Spectroscopy for Medical Diagnosis*, M. Diem, P. R. Griffiths, and J. M. Chalmers, Eds., pp. 203–230, John Wiley & Sons, Chichester, UK, 2008.
- [85] C. Matthäus, T. Chernenko, J. A. Newmark, C. M. Warner, and M. Diem, "Label-free detection of mitochondrial distribution in cells by nonresonant Raman microspectroscopy," *Biophysical Journal*, vol. 93, no. 2, pp. 668–673, 2007.
- [86] K. Klein, A. M. Gigler, T. Aschenbrenner et al., "Label-free live-cell imaging with confocal Raman microscopy," *Biophysical Journal*, vol. 102, no. 2, pp. 360–368, 2012.
- [87] C. Matthäus, A. Kale, T. Chernenko, V. Torchilin, and M. Diem, "New ways of imaging uptake and intracellular fate of liposomal drug carrier systems inside individual cells, based on Raman microscopy," *Molecular Pharmaceutics*, vol. 5, no. 2, pp. 287–293, 2008.
- [88] E. Zuser, T. Chernenko, J. Newmark, M. Miljković, and M. Diem, "Confocal Raman microspectral imaging (CRMI) of murine stem cell colonies," *Analyst*, vol. 135, no. 12, pp. 3030–3033, 2010.
- [89] H. G. Schulze, S. O. Konorov, N. J. Caron, J. M. Piret, M. W. Blades, and R. F. B. Turner, "Assessing differentiation status of human embryonic stem cells noninvasively using Raman microspectroscopy," *Analytical Chemistry*, vol. 82, no. 12, pp. 5020–5027, 2010.



- [90] I. Notingher, I. Bisson, J. M. Polak, and L. L. Hench, "In situ spectroscopic study of nucleic acids in differentiating embryonic stem cells," *Vibrational Spectroscopy*, vol. 35, no. 1-2, pp. 199–203, 2004.
- [91] I. Notingher, I. Bisson, A. E. Bishop, W. L. Randle, J. M. P. Polak, and L. L. Hench, "In situ spectral monitoring of mRNA translation in embryonic stem cells during differentiation in vitro," *Analytical Chemistry*, vol. 76, no. 11, pp. 3185–3193, 2004.
- [92] J. K. Pijanka, D. Kumar, T. Dale et al., "Vibrational spectroscopy differentiates between multipotent and pluripotent stem cells," *Analyst*, vol. 135, no. 12, pp. 3126–3132, 2010.
- [93] J. W. Chan, D. K. Lieu, T. Huser, and R. A. Li, "Label-free separation of human embryonic stem cells and their cardiac derivatives using Raman spectroscopy," *Analytical Chemistry*, vol. 81, no. 4, pp. 1324–1331, 2009.
- [94] I. Notingher, "Label-free imaging of phenotypic spectral markers in live cells derived from human stem cells," in *Proceedings of the 14th European Conference on the Spectroscopy of Biological Molecules*, Coimbra University, Coimbra, Portugal, 2011.
- [95] U. Neugebauer, T. Bocklitz, J. H. Clement, C. Krafft, and J. Popp, "Towards detection and identification of circulating tumour cells using Raman spectroscopy," *Analyst*, vol. 135, no. 12, pp. 3178–3182, 2010.
- [96] U. Neugebauer, J. H. Clement, T. Bocklitz, C. Krafft, and J. Popp, "Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging," *Journal of Biophotonics*, vol. 3, no. 8-9, pp. 579–587, 2010.
- [97] S. Dochow, C. Krafft, U. Neugebauer et al., "Tumour cell identification by means of Raman spectroscopy in combination with optical traps and microfluidic environments," *Lab on a Chip—Miniaturisation for Chemistry and Biology*, vol. 11, no. 8, pp. 1484–1490, 2011.
- [98] L. Hartsuiker, N. J. L. Zeijen, L. W. M. M. Terstappen, and C. Otto, "A comparison of breast cancer tumor cells with varying expression of the Her2/neu receptor by Raman microspectroscopic imaging," *Analyst*, vol. 135, no. 12, pp. 3220–3226, 2010.
- [99] T. Chernenko, C. Matthäus, L. Milane, L. Quintero, M. Amiji, and M. Diem, "Label-free Raman spectral imaging of intracellular delivery and degradation of polymeric nanoparticle systems," *ACS Nano*, vol. 3, no. 11, pp. 3552–3559, 2009.
- [100] C. Matthäus, T. Chernenko, L. Quintero et al., "Raman micro-spectral imaging of cells, and applications monitoring the uptake of drug delivery systems," in *Confocal Raman Imaging*, O. Hollricher and T. Dieing, Eds., pp. 137–162, Springer, Heidelberg, Germany, 2011.
- [101] U. Utzinger, D. L. Heintzelman, A. Mahadevan-Jansen, A. Malpica, M. Follen, and R. Richards-Kortum, "Near-infrared Raman spectroscopy for in vivo detection of cervical precancers," *Applied Spectroscopy*, vol. 55, no. 8, pp. 955–959, 2001.
- [102] H. Barr, C. Kendall, and N. Stone, "Raman spectroscopy as a potential tool for early diagnosis of malignancies in esophageal and bladder tissues," in *Vibrational Spectroscopy for Medical Diagnosis*, M. Diem, P. R. Griffiths, and J. M. Chalmers, Eds., John Wiley & Sons, Chichester, UK, 2008.
- [103] H. Barr, C. Kendall, and N. Stone, "The light solution for Barrett's oesophagus: photodiagnosis and photodynamic therapy for columnar-lined oesophagus," *Photodiagnosis and Photodynamic Therapy*, vol. 1, no. 1, pp. 75–84, 2004.
- [104] C. Kendall, J. Day, J. Hutchings et al., "Evaluation of Raman probe for oesophageal cancer diagnostics," *Analyst*, vol. 135, no. 12, pp. 3038–3041, 2010.
- [105] J. Horsnell, P. Stonelake, J. Christie-Brown et al., "Raman spectroscopy—a new method for the intra-operative assessment of axillary lymph nodes," *Analyst*, vol. 135, no. 12, pp. 3042–3047, 2010.
- [106] L. E. Orr, J. Christie-Brown, J. C. Hutchings et al., "Raman spectroscopy as a tool for the identification and differentiation of neoplasias contained within lymph nodes of the head and

- neck,” in *Photonic Therapeutics and Diagnostics VI*, Proceedings of SPIE, p. 7548W1—11, January 2010.
- [107] C. Krafft, M. A. Diderhoshan, P. Recknagel, M. Miljkovic, M. Bauer, and J. Popp, “Crisp and soft multivariate methods visualize individual cell nuclei in Raman images of liver tissue sections,” *Vibrational Spectroscopy*, vol. 55, no. 1, pp. 90–100, 2011.
- [108] C. Krafft, D. Codrich, G. Pelizzo, and V. Sergo, “Raman and FTIR imaging of lung tissue: bronchopulmonary sequestration,” *Journal of Raman Spectroscopy*, vol. 40, no. 6, pp. 595–603, 2009.
- [109] C. W. Freudiger, W. Min, B. G. Saar et al., “Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy,” *Science*, vol. 322, no. 5909, pp. 1857–1861, 2008.
- [110] H. Kano, “Molecular vibrational imaging of a human cell by multiplex coherent anti-Stokes Raman scattering microspectroscopy using a supercontinuum light source,” *Journal of Raman Spectroscopy*, vol. 39, no. 11, pp. 1649–1652, 2008.

