High-speed spectral nanocytology for early cancer screening

John E. Chandler
Hariharan Subramanian
Charles D. Maneval
Craig A. White
Richard M. Levenson
Vadim Backman

High-speed spectral nanocytology for early cancer screening

John E. Chandler,[a] Hariharan Subramanian,[a] Charles D. Maneval,[a] Craig A. White,[a] Richard M. Levenson,[b] and Vadim Backman[a]

[a]Northwestern University, Department of Biomedical Engineering, 2145 Sheridan Road, Evanston, Illinois 60208
[b]University of California, Davis Medical Center, Department of Pathology and Laboratory Medicine, PATH Building, 4400 V Street, Sacramento, California 95817

Abstract. High-throughput partial wave spectroscopy (HTPWS) is introduced as a high-speed spectral nanocytology technique that utilizes the field effect of carcinogenesis to perform minimally invasive cancer screening on at-risk populations. HTPWS uses fully automated hardware and an acousto-optic tunable filter to scan slides at low magnification, to select cells, and to rapidly acquire spectra at each spatial pixel in a cell between 450 and 700 nm, completing measurements of 30 cells in 40 min. Statistical quantitative analysis on the size and density of intracellular nanostructures extracted from the spectra at each pixel in a cell yields the diagnostic biomarker, disorder strength (L_d). Linear correlation between L_d and the length scale of nanostructures was measured in phantoms with $R^2 = 0.93$. Diagnostic sensitivity was demonstrated by measuring significantly higher L_d from a human colon cancer cell line (HT29 control vector) than a less aggressive variant (epidermal growth factor receptor knockdown). Clinical diagnostic performance for lung cancer screening was tested on 23 patients, yielding a significant difference in L_d between smokers and cancer patients, $p = 0.02$ and effect size $= 1.00$. The high-throughput performance, nanoscale sensitivity, and diagnostic sensitivity make HTPWS a potentially clinically relevant modality for risk stratification of the large populations at risk of developing cancer. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JBO.18.11.117002]

Keywords: light-scattering spectroscopic microscopy; cancer screening; cellular nanoarchitecture.

Paper 130357RR received May 20, 2013; revised manuscript received Sep. 9, 2013; accepted for publication Sep. 23, 2013; published online Nov. 5, 2013.

1 Introduction

In the United States, cancer continues to be a critical health care issue, ranking as the second leading cause of death behind cardiovascular disease.[1] Early detection of cancerous lesions is widely recognized as one of the most important factors for successful treatment of the disease. However, for most major cancer types, there remains a severe lack of cost-effective and minimally invasive screening procedures that can be performed in a primary-care setting, specifically for early stage detection. The current gold standard for detection of most major cancers remains performing imaging using techniques such as CT, MRI, X-ray, and positron emission tomography on symptomatic individuals followed by a diagnosis confirmed with tissue collection via biopsy or fine-needle aspiration. However, all of these techniques have proven to be either unrealistic due to cost and/or risk or simply ineffective as screening modalities for early detection in at-risk populations.[2] In particular, none of the aforementioned techniques can be feasibly implemented in a primary-care setting for the stratification of a large at-risk population. To combat these drawbacks, development of minimally invasive screening techniques for cancer is critical in order to identify individuals with abnormalities requiring more invasive investigation and to help stratify the high-risk populations that are otherwise subjected to more expensive and higher risk screening procedures.[2]

For the purpose of cancer screening, one of the most successful technologies has been the automated cytology systems used to automate image collection from patient sample slides for visual inspection by a pathologist. The most effective use of this technique has been in automating analysis of Papanicolaou (Pap) smear tests for cervical cancer screening. For this application, several instruments exist including the ThinPrep Imaging System (Hologic, Bedford, Massachusetts) and the Beckton Dickinson Focal Point GS Imaging System (Beckton Dickinson, Franklin Lakes, New Jersey). These systems are capable of transferring cells from solution onto a glass slide, staining the cells, and then imaging specific regions of interest on the slide for interpretation by a pathologist. Automated cytology systems have demonstrated the ability to improve rates of lesion detection during cervical cancer screening.[3,4] In addition to improved diagnostic performance, these techniques also provide a means for high-throughput screening, completing analysis of a single patient slide in 4 to 8 min. Despite these advantages, the initial cost of implementation has made these instruments only viable in clinics that experience high-screening volumes, although costs are decreasing with improvements to the technology. Additionally, due to the drastic increases in productivity that these systems allow, workload limits of less than 100 slides per day have been recommended for cytotechnologists to prevent decreases in diagnostic performance, since all slides showing indications of disease require visual inspection to complete the diagnosis.[5–7] While automated cytopathology has proven highly successful for cervical cancer screening, its

Address all correspondence to: John E. Chandler, Northwestern University, Department of Biomedical Engineering, 2145 Sheridan Road, Evanston, Illinois 60208. Tel: (847)467-3216; Fax: (847)491-4928; E-mail: johnchandler2011@u.northwestern.edu

use has been limited for other major cancer types due to a lack of easily accessible organ sites from which cells can be investigated.

Ongoing research in the field of cytology and histopathology is exploring the use of the technology for other major cancers besides cervical. This research is focused on developing quantitative, fully automated screening techniques and developing new techniques that are sensitive to the earliest changes associated with carcinogenesis, without needing to image or acquire samples at the actual tumor site. As a result, computer-aided diagnosis has become a major subject of research for further automation and quantification. These systems employ quantitative image analysis and segmentation algorithms to detect cancerous cells or indications of disease. Examples of this type of work have demonstrated automated lung cancer detection via sputum cytology, prostate cancer from digitized needle biopsies, and breast cancer from fine-needle biopsies.[8–10] In addition, methods for improving sensitivity to distant tumors have been explored using DNA-specific staining to detect malignancy-associated changes in cells collected away from a tumor site.[11] This approach has been demonstrated in several cancer sites including the lung, breast, and cervix.[12–14]

One of the drawbacks of current automated cytology systems is the diffraction-limited length-scale sensitivity/resolution that cannot detect changes in intracellular nanoarchitecture associated with the earliest stages of carcinogenesis. Currently available imaging techniques that do provide nanoscale resolution sensitive to the intracellular changes associated with carcinogenesis include stochastic optical reconstruction microscopy, stimulated emission depletion microscopy, and photoactivated localization microscopy. While these techniques are capable of imaging structures in individual cells below the diffraction limit ($\sim$200 nm), they have not shown utility for clinical use due to a lack of sample throughput, automation, and the requirement of image interpretation by a trained expert in cell biology. Patient screening requires methods with very high-sample throughput, automated measurements, and analysis with the entire process completed in less than 10 min per patient.

Recently, we have shown that via the field effect of carcinogenesis, nanoarchitectural changes associated with the development of different types of cancers are detectable in cells further away from the location of the actual tumor using partial wave spectroscopic (PWS) microscopy.[15–17] Field carcinogenesis is an early stage of carcinogenesis in essentially all types of carcinomas. It is manifested by the accumulation of genetic, epigenetic, and nanoarchitectural alterations throughout an affected organ site that increase the risk of neoplastic transformation with focal neoplastic lesions emerging as a result of further stochastic molecular events out of this "fertile field of injury." The field effect has been employed previously for research on potential cancer screening tests as shown with the SCM (structuredness of the cytoplasmic matrix) test and for detection of malignancy-associated changes via quantitative image analysis of cells stained with DNA-specific dyes.[12,13,18–20] Our PWS data show that the nanoarchitectural changes in buccal mucosa cells are sensitive to the presence of a distant lung tumor, and this sensitivity is not confounded by demographic factors such as age, smoking, gender, etc.[15] PWS quantitatively differentiates between patients with and without cancer by measuring the disorder strength (L_d), a statistical property proportional to the size and density of macromolecular structures in a patient's cells. Despite previously demonstrated diagnostic performance,

PWS has not been a clinically viable technique for the same reason as other nanoscale-sensitive techniques due to low sample throughput associated with the slow manual process required to perform measurements. As a result, there remains a lack of high-throughput nanoscale-sensitive screening technologies that could prove effective for early cancer detection.

To address the lack of high-throughput nanoscale-sensitive screening techniques available for cancer, we have used automated cytology as a model of a successful screening approach that serves as the foundation for an optical technique that is both high-throughput and diagnostically sensitive to the nanoscale intracellular changes associated with carcinogenesis. More specifically, a new high-throughput PWS (HTPWS) system has been developed allowing automated, rapid collection, and analysis of patient's data. The goals of such a system were to achieve comparable performance to current automated cytopathology machines used in cervical cancer screening ($\sim$4 to 8 min/slide) and to demonstrate a technology that has the potential to be the first level of cancer screening and stratification for large at-risk populations in a primary-care setting.[6] In this work, we present a completed prototype for HTPWS measurements that has the potential to become a tool for automated cancer nanocytology. Development of the new HTPWS instrument included a three-step process: radically modifying the instrumentation to allow for new illumination and collection schemes, development of new and improved software interfaces to control and automate all aspects of data acquisition and processing, and validation experiments to demonstrate the effectiveness and performance of the new instrumentation.

2 HTPWS Instrumentation, Software, and Analysis

2.1 First-Generation PWS Instrumentation for Quantitative Analysis of Intracellular Disorder (L_d)

This initial design of the PWS microscope was based on several key components including a critical illumination system via a fixed low-numerical-aperture (20×, NA = 0.4) objective lens (38-339, Edmund Optics, Barrington, New Jersey), a manual sample stage, and a spectral filtering system using a scanned 10-μm slit spectrometer (SP-2150i, Acton Research, Acton, Massachusetts).[16,21] With this system, collimated broadband light from a Xenon lamp (66902 100W, Oriel Instruments, Stratford, Connecticut) was focused on the sample by the low-numerical-aperture objective lens. Backscattered light was collected by the same objective lens, and the magnified image was focused on the slit of the spectrometer, which was mounted on an automated linear stage (T-LA60A, Zaber Technologies, Vancouver, British Columbia), to allow the slit to be scanned across the entire illuminated field ($\sim$200 μm). Images were collected with a CCD camera (Coolsnap HQ, Photometrics, Tucson, Arizona) with 6.45-μm pixels from the output of the spectrometer at each 10-μm step of the automated stage. Each image collected by the camera yielded an image with the y-axis representing the spatial y-dimension in the image plane and the x-axis representing the wavelength of light. By combining all the images collected at each x-spatial position, a three-dimensional data cube (x, y, λ) was formed representing the diffraction-limited spatial data as well as the sub-diffraction spectral data.

All PWS measurements yield backscattered intensity spectra corresponding to each spatial pixel in an image of a cell, which can be quantitatively analyzed to find a diagnostic biomarker for cancer called the disorder strength or L_d. L_d quantifies the spatial variation of refractive index in the cell, which increases linearly as function of both the length scale and density of the intracellular refractive index fluctuations. Thus, increased L_d values correlate with local macromolecular condensation events such as heterochromatin and euchromatin condensation in the nucleus.[22,23] The computation of disorder strength (L_d) is given by $L_d(x, y) \approx \sigma_n l_c$, in which $\langle l_c \rangle$ is the spatial (x, y) correlation length of the refractive index fluctuations, and $\langle \sigma_n \rangle$ is the standard deviation of the refractive index fluctuations.[15,16] In physical terms, the spatial correlation length, $\langle l_c \rangle$, corresponds to the size of intracellular structures causing refractive index fluctuations, and the standard deviation, $\langle \sigma_n \rangle$, is proportional to the density of the intracellular structures. This calculation generates an image of the L_d values at each spatial position (pixel) in the sample. The total magnification of the system is such that the pixel size at the object plane is less than the diffraction-limited spot size, so L_d values and their spatial distributions are independent of the hardware.

For the experiments described here, L_d was calculated on the spectra between 500 and 700 nm. Prior to the L_d calculation, each signal is normalized by the input illumination spectrum acquired from a mirror measurement, and a sixth-order low-pass Butterworth filter is applied to remove noise. A low-order polynomial subtraction was used to remove low-order slopes in the signal that result from differences in the lamp spectrum, sample roughness, and instrument artifacts. HTPWS L_d values had to be scaled down by a factor of 3.37 to match L_d values from the first-generation system to account for the reduction in spectral sampling frequency. For diagnostic purposes, cells were characterized by their mean intracellular L_d and patients by the mean of all intracellular L_d values from each measured cell. Statistical evaluation of diagnostic performance was performed in Excel (Microsoft, Redmond, Washington) and MATLAB® (MathWorks, Natick, Massachusetts). Mean L_d values between different sample groups were compared using the two-sided Student's t-test, and effect size was computed as Cohen's d.

2.2 Goals of the HTPWS Instrumentation

While the first-generation PWS microscope effectively demonstrated the modality's potential as a clinical diagnostic tool, its design limited its effectiveness for large-scale high-throughput clinical studies and research. As described, the first-generation PWS system collects the backscattered image by linearly scanning the entire cell using the combination of the slit spectrometer and scanning stage, a process which takes approximately 3 min /cell. Measurements for an entire slide are slowed further, since the field of view is limited to approximately 120 μm requiring at least 2 to 3 min to manually find and focus on the cell. As a result, the total time to measure approximately 30 cells per patient is ~4 to 5 h. To improve upon the performance of the first-generation instrument, a new HTPWS system has been developed, as shown in Fig. 1, which reduces the time to measure patient samples and automates the data acquisition process. More specifically, to automate the process of finding and selecting cells for measurement that previously took more than an hour per slide, the HTPWS system used automated hardware to scan an entire slide at low magnification and to analyze the collected images to identify and locate the positions of cells. In addition, the slit-spectrometer approach to spectral filtering was replaced by an acousto-optic tunable filter (AOTF) to decrease data collection times and to eliminate the requirement of spatially scanning each image, significantly reducing data collection times from 3 min /cell to less than 5 s/cell. While the HTPWS system fundamentally operates on the same principles as the first-generation device, the changes are detailed here in the sections pertaining to the tunable
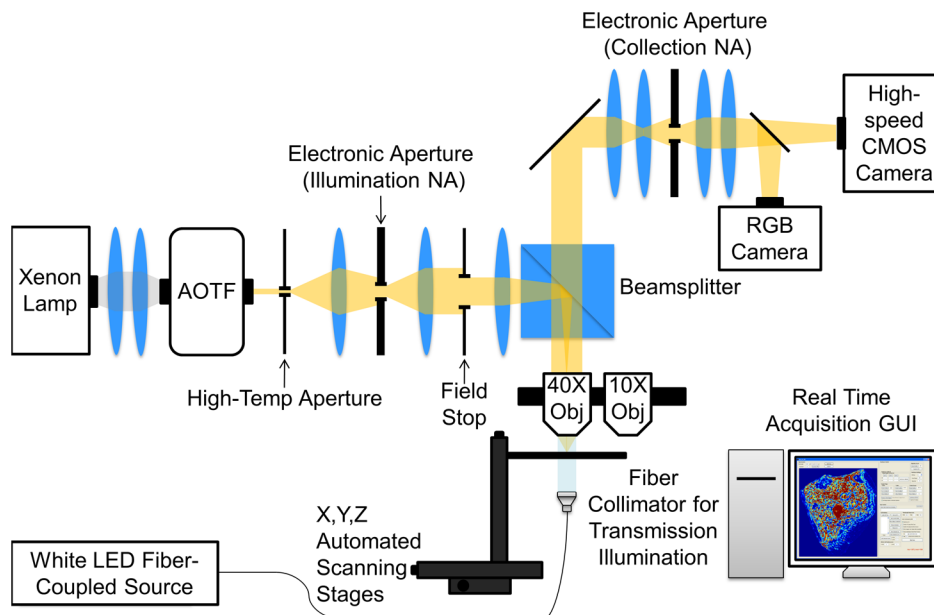
Fig. 1 Schematic of the high-throughput partial wave spectroscopy (HTPWS) instrument. Tunable illumination is incident on the sample from an acousto-optic tunable filter (AOTF) with the illumination numerical aperture set by an electronic aperture. Backscattered reflectance is collected through a second electronic aperture that sets the collection numerical aperture with a high-speed CMOS camera. All data collection is automated via the acquisition GUI. Transmission bright-field image collection is also possible with the fiber-coupled LED source.

illumination hardware and optics, high-speed fully automated hardware, and custom software interfaces and algorithms used to achieve fully automated data acquisition.

2.3 Illumination Design

The HTPWS microscope uses a Köhler illumination alignment with tunable illumination to increase both uniformity of illumination and spectral sampling speed. Light from a Xenon lamp is focused through an AOTF (HSI-300, Gooch & Housego, Orlando, Florida). The AOTF has a minimum switching speed of 50 μs, bandwidth of 3 nm, and a spectral range from 450 to 800 nm. The tuned light exiting the AOTF is focused through an electronic motorized aperture (62281, Newport Corporation, Irvine, California) that sets the illumination numerical aperture. Light exiting the aperture is collimated and passed through the field aperture, after which it is focused onto the back focal plane of the objective lens (40×, NA = 0.6 LUCPlanFL N, Olympus, Center Valley, Pennsylvania). This new illumination scheme achieves uniform intensity across the sample plane due to the Köhler alignment, and wavelength switching is less than 100 μs. In addition, because the incident illumination is tuned to a single monochromatic wavelength, this illumination system will allow future multimodal experiments combining fluorescence with HTPWS. These HTPWS experiments can be performed with molecularly specific dyes, which have been shown to enhance the refractive index of cellular organelles within the sample.[24]

2.4 High-Speed Automated Hardware

To shift from an entirely manual and user-intensive measurement process to automated high-throughput measurements, high-speed automated hardware was added to the system to replace each manual operation performed by the user. The new automated hardware is controlled via custom algorithms written by the authors in MATLAB®. First, the sample stage was upgraded with automated, encoded, linear stages (A-LSQ075B-E01, Zaber Technologies) for the x- and y-axes as well as an automated linear stage for the z-axis (T-LS28, Zaber Technologies). An automated objective turret was added to switch between the high-magnification (40×) and low-magnification (10×, UPlanFL N, Olympus) objectives. The angle of the collected backscattered light was controlled using a second motorized aperture before being focused on an ultra-fast CMOS camera (Hamamatsu ORCA-Flash 2.8, Bridgewater, New Jersey) with 3.63-μm pixels binned 2×2. For maximum speed, synchronization of wavelength tuning and image capturing was done via hardware triggers between the AOTF and the camera; 2×2 binning was enabled with the camera to maximize sensitivity and minimize exposure time for the fastest acquisition.

In addition to the PWS illumination, a transmission illumination arm was added to allow collection of bright-field images alongside PWS measurements. A white-light emitting diode [light emitting diode (LED)] fiber-coupled source (LE-1W-CE, WT&T, Lachine, Quebec) was connected to a fiber collimator, and the output beam was passed through a diffuser. This transmitted light was collected via a high-resolution scientific color CMOS USB camera (DCC1645C, Thorlabs, Newton, New Jersey) that was added to the system for this purpose. This camera was also used for rapid collection of low-magnification/low-resolution transmission images for slide mapping. A flipper

mirror was used to switch between the transmission collection camera and the camera used for PWS measurements.

2.5 Automated Slide Mapping

The first task performed during a HTPWS measurement is generating a large low-magnification (typically 10× to 20×) image of the slide. This is accomplished using an algorithm that rapidly collects many low-resolution images and tiles them together to create the full image of the slide. A user defines the bounds of the region to be mapped by specifying the positions of two diagonal corners. The algorithm then calculates the number of images required to map the entire region specified based on a pixels-to-micron conversion factor specific to the objective and imaging sensor used. The region is then raster scanned, and an image is acquired at each x and y position in order to make a complete image of the region without gaps or overlaps. Finally, all the images are tiled together to form a complete image of the entire region. To maximize the speed of the algorithm, autofocusing (using a custom autofocus algorithm—discussed in Sec. 2.7) is performed only on the first image, and predictive autofocusing is used for all subsequent images. The focus is rechecked and corrected every 10 images as necessary throughout the scan. Images are stored in memory until the end of each scanning line, and then saved in order to prevent the algorithm from overloading the computer's physical memory.

2.6 Semi-Automated Cell Selection

Cell selection is performed using the large low-magnification image of the slide generated by the slide-mapping algorithm. This can be accomplished in one of two ways: a user can manually select cell positions using the mouse and saving them to a list for measurements, or image-segmentation algorithms unique for the cell type on the slide can be employed to automatically generate a list of candidate cells which are then shown to the user for approval. Manual selection allows the user to zoom-in and explore any region of a slide image and select positions anywhere on the image with a crosshair. Each selected position is saved to a positions list that is exported to the acquisition graphical user interface (GUI). In contrast, when an automated segmentation-based selection algorithm is running, the user is immediately prompted with a list of potential cell positions. Images of each potential cell and its local surroundings are displayed to the user, and a prompt asks whether to keep or reject the position. This semi-automated form of position selection has the advantage of being much faster, but it is limited to cell types for which an algorithm has been developed. We have already developed one such algorithm for stained buccal cells.

2.7 Autofocus Algorithm

In order to perform both automated slide mapping and automated PWS measurements, a rapid autofocus algorithm was developed to accurately identify the correct focus plane at any position on the slide. Due to the unique needs of slide mapping and HTPWS measurements, the former requiring higher speed and the latter requiring greater accuracy, a two-step algorithm was developed. The first step is a predictive autofocus algorithm based on the equation of a plane in three dimensions. Three in-focus x, y, z positions on the slide are collected to generate an equation for a plane that predicts the in-focus position anywhere on the slide. Between points where autofocusing is

actually performed, this algorithm is used to predict the in-focus z-position during slide mapping. Thus, based on the equation of a plane, the predicted z-position is given by $z = \frac{d-ax-by}{c}$, where a, b, c, and d are constants defining the equation of the plane in three-dimensional space, and x, y, and z are spatial coordinates.

To more accurately get an in-focus image at a cell position for HTPWS measurements, an algorithm based on edge detection of the field aperture is used. The best focus is determined by the highest number of edges detected at the edge of the field. In this manner, focus consistency, which is critical to prevent variability in quantitative HTPWS L_d analysis, is obtained by focusing on a fixed object that is always in the same position. The approach is similar to one proposed by Hughlett and Kaiser with the exception being that we use the edge of the field aperture instead of a shadow projection wire to quantify focus via an edge-detection technique.[25] Edges corresponding to the aperture are isolated by segmenting the field-of-view and by applying a black/white threshold to a Sobel gradient magnitude image of the field. A slight erosion of the segmented field leaves a mask that can be applied to images to obtain only edges corresponding to the field aperture.

To find an in-focus image, the algorithm must search for the z-position that corresponds to the maximum number of edges from the field aperture. It does this by scanning a user-determined range around the predicted focus position with large incremental steps ($\sim$5 μm). At each position, the number of edges is found using a Sobel edge detector. Because the number of edges forms a Gaussian curve with the in-focus position corresponding to the center of the peak, the algorithm detects when the number of edges switches from increasing to decreasing and stops scanning. The stage then backtracks in fine increments ($\sim$0.3 μm) to find the maximum number of edges corresponding to the aperture at the current x, y coordinate. Figure 2 shows the difference in edges detected at the field aperture outside the mask for in-focus and out-of-focus images.

2.8 Automated Spectral Measurements

Spectral measurements are completed on the HTPWS microscope via the automated measurement and analysis interface. The list of selected positions is automatically loaded, and the user sets the parameters for the scan including the wavelengths to scan, the illumination bandwidth, exposure time, input NA, and collection NA. Typical settings for a high-throughput scan include a spectral range of 450 to 700 nm with a step size of 1 nm. Input NA is typically set with the input aperture at 10%, approximating plane-wave illumination without sacrificing more light than necessary. Output NA is typically not constrained by the electronic aperture and is instead determined by the objective lens (NA = 0.6). When the scan is running, the system automatically moves to each position on the stored list, autofocuses, and spectrally scans the sample, collecting an image at each illumination wavelength. The result is a three-dimensional data cube (x, y, λ) identical to that provided by the first-generation system, but with the benefits of being fully automated and requiring much less time. Figure 3 illustrates the process required to complete an HTPWS measurement.

3 Experiments and Results

3.1 HTPWS System Performance

Performance with the new HTPWS microscope is dramatically improved compared with the first-generation system. A single HTPWS measurement can be completed in less than 5 s compared with 3 to 4 min for the first-generation system, an
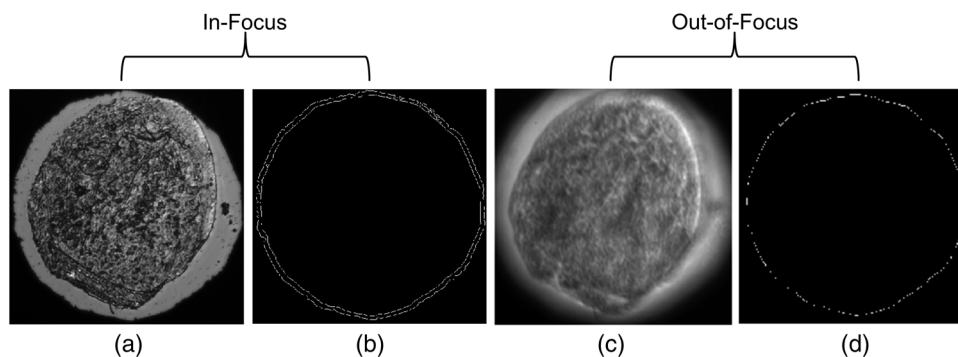
Fig. 2 From left to right: (a) In-focus image of a buccal cell and (b) corresponding edge map for the in-focus image showing the field aperture edges visible outside the border of the mask applied to remove the field. (c) Out-of-focus image of the same buccal cell and (d) corresponding edge map for the out-of-focus image showing no edges detected outside the border of the mask.
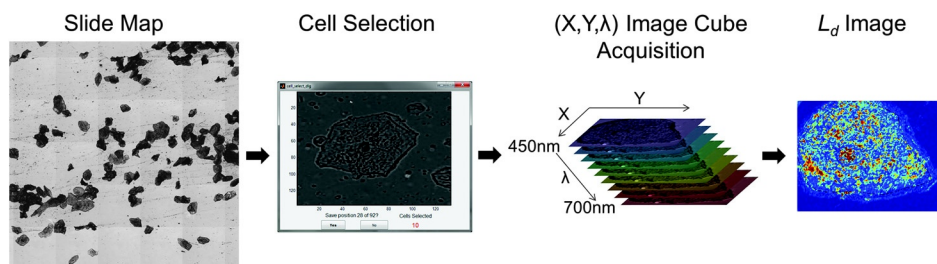
Fig. 3 Flow chart of the HTPWS measurement process including slide mapping, segmenting cell positions, collecting the spectral image cube, and generating an image showing the L_d distribution within a cell.

approximately 42-fold improvement. For a typical measurement of 30 cells, slide mapping is completed in 10 min. Manual cell selection adds another 10 min, while semi-automated cell selection using a segmentation algorithm can be achieved in less than 5 min. Finally, the spectral measurements are completed in 20 min. Thus, for a typical measurement, the entire process on a single slide for approximately 30 cells completes in approximately 40 min compared with 4 to 5 h for the first-generation system. The increases in performance and automation allow approximately seven patients to be measured on the HTPWS system for one on the first-generation system.

3.2 Validation Experiments

In order to characterize the new HTPWS system, a series of experiments were performed to validate that its performance correlated with the previous data collected on the first-generation system. In particular, it was necessary to demonstrate that the new system measured the same spectral information as the first-generation system and that the HTPWS instrument was sensitive to the nanoscale properties of the samples measured. Validation of the spectral data was performed using a uniform 1-μm SiO$_2$ thin-film reference standard (Filmetrics, San Diego, California). Measurements were taken on the same region of the thin film using both the first-generation spectrometer-based PWS system and the new HTPWS system. To account for the use of detectors with different sensitivities on each system, the exposures for each separate detector were set to give the same signal-to-noise ratios for the two systems. The spectra at specific pixels within the field-of-view from both measurements were then compared directly for a match within the same wavelength ranges used for PWS measurements (500 to 700 nm). Further spectral validation was also performed using more complex nonuniform samples. Spectra from the same individual latex microspheres of varying sizes (4.3 and 11 μm) that were allowed to dry out of solution on a glass slide (Thermo Scientific, Waltham, Massachusetts) were measured and compared for a direct match between individual pixels in the same locations. Finally, spectra from identical cells [buccal and HT29 (colon cancer) cell types] measured on both systems were plotted for a match at the same pixels and regions.

Spectral comparison between the HTPWS system and the spectrometer-based first-generation system showed consistently similar results from identical samples. Figure 4 shows the spectra generated from the same location on the SiO$_2$ thin-film reference. These spectra were generated by averaging the spectra over a diffraction-limited area at the same location on the sample for both systems. A sixth-order Butterworth low-pass filter was applied to the signals from both systems to remove noise after normalizing by the spectra collected from a mirror. This result was compared with the spectra provided by the manufacturer of the thin-film reference, and there was an excellent agreement in terms of amplitude, slope, oscillation frequency, and phase. Figure 4 also shows a comparison of spectra averaged from pixels that make up the same diffraction-limited spot in an HT29 colon cancer cell. A match can clearly be observed between the spectra from the two systems at this location in the cell. In this case, perfect matching is much more difficult to achieve due to inhomogeneity of the sample, nonuniform sample topography, and differences in detector pixel sizes between the two systems. With perfectly uniform samples, such as the thin-film reference, the different pixel sizes of the detectors are not an issue, because the signal is the same at every point on the sample. On random samples, such as cells, the signal can vary significantly from pixel to pixel, making the difference in pixel size much more significant when matching spectra. Averaging spectra from the pixels that make up a diffraction-limited spot in each system helps to account for this and allows decent spectral matches to be obtained when the same location in a cell is analyzed.

To demonstrate and compare the nanoscale sensitivities of the first-generation PWS instrument with the new HTPWS instrument, nanoscale phantoms were created and measured. The phantoms were constructed using solutions of polymer
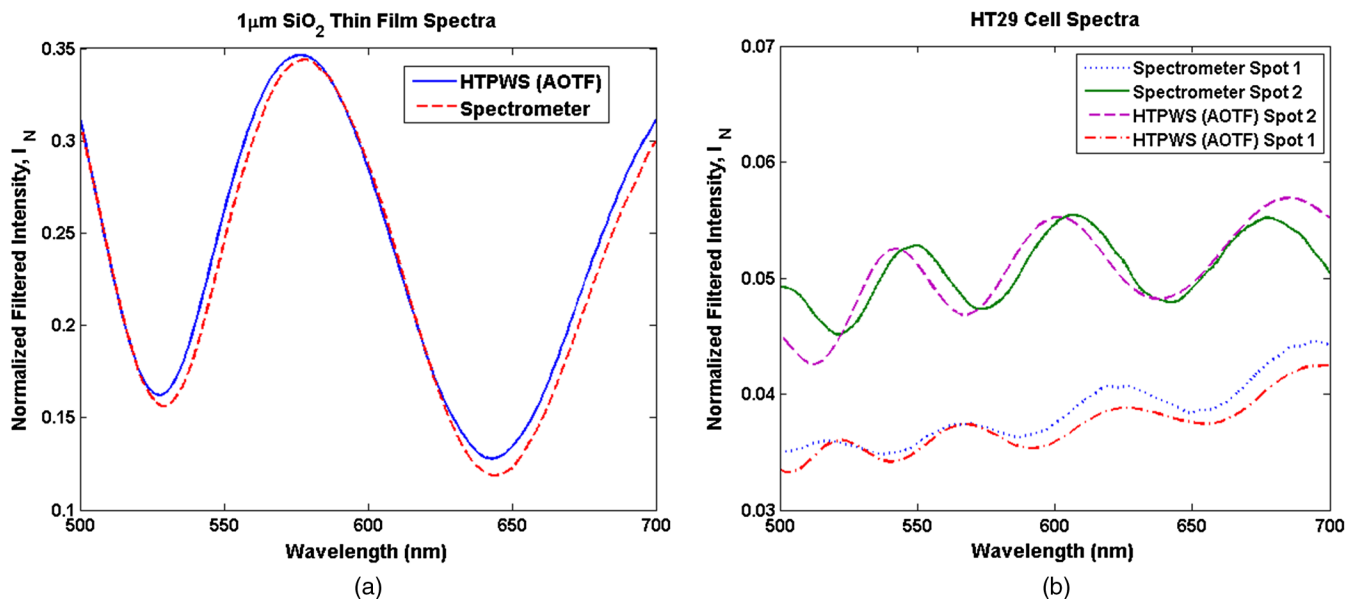

Fig. 4 (a) Normalized and filtered 1-μm thick SiO$_2$ thin-film reference spectra plotted from both the HTPWS system and spectrometer-based first-generation PWS instrument. The spectra are averaged over a diffraction-limited spot in the same location on the sample for both instruments, and a low-pass Butterworth filter is applied to remove noise. Root-mean-square error between the spectra was 0.01. (b) Normalized and filtered spectra comparison averaged at the same diffraction-limited spots in an HT29 colon cancer cell with both the HTPWS and first-generation PWS instrument.

nanospheres (Thermo Scientific). Separate phantoms were created corresponding to different length scales using spheres with diameters of 20, 40, 60, 80, 100, and 125 nm. Each phantom was made by applying a single droplet of the sphere solution to a glass slide and letting the spheres randomly assemble on the slide as the solution dried. This left behind a thin-layered structure of spheres that could be used to represent a random assortment of particles at a specific length scale. HTPWS measurements were taken from each phantom at the same locations on both systems. Twenty-five measurements were acquired from each phantom at different positions to allow for statistical comparison of the data. In order to compare phantoms comprised of spheres of different diameters, measurements were acquired in each phantom from regions of similar thickness based on the number of spectral oscillations (5 to 7 oscillations or 2.5 to 3.5 μm). For each phantom, 25 regions of interest were selected, and L_d analysis was performed on the pixels in these regions. These L_d measurements were then plotted as a function of the phantom nanosphere size to demonstrate sensitivity of L_d to nanoscale length scales.

The measurements of the nanoscale phantoms performed on both systems were analyzed to calculate the L_d value of each phantom. Figure 5 summarizes the results of this analysis, showing the sensitivity of L_d to nanoscale length scales. The length-scale dependence of L_d can clearly be observed as L_d values show a steadily increasing trend with increasing diameter of the nanospheres making up the phantoms. Correlation between the length-scale of phantom spheres and L_d is linear with an R^2 value of 0.93.
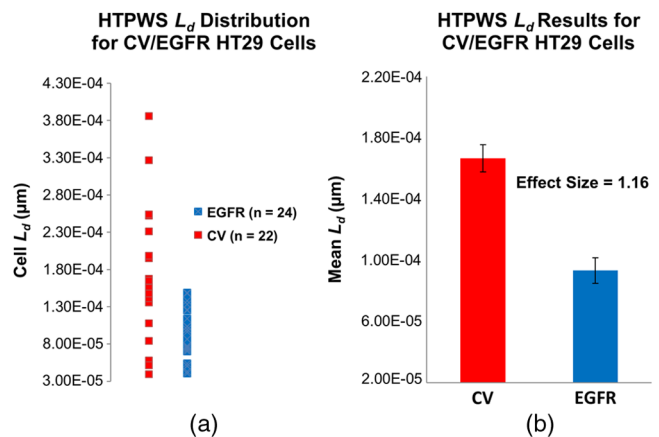
3.3 Diagnostic Performance Experiments

Two experiments were performed to test the diagnostic performance of the new HTPWS system. First, HT29 colon cancer cell lines were used to model a clinical diagnostic test. The experiment consisted of two groups, control vector HT29 (CV) cells and epidermal growth factor receptor (EGFR) knockdown HT29 cells, a less aggressive genetic variant. The HT29 control vector and EGFR knockdown cells were first collected in centrifuge tubes and centrifuged for 5 min at 1000 rpm. The supernatant was then removed, and the cells were plated on a glass chamber slide. The slides were checked to ensure that they contained at least 20,000 cells. Two milliliters of fresh cell culture medium was added to each chamber slide, which was then incubated at
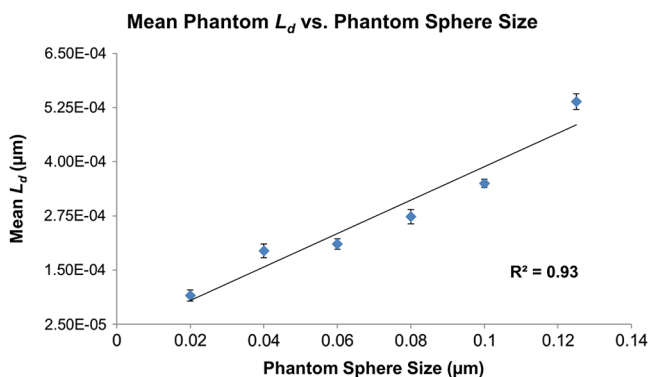
37°C for at least 5 to 6 h. After incubation, the medium was completely removed from the chamber slides, and the slides were washed with 70% ethanol to remove any traces of the medium. Following this, the slides were immediately fixed using 70% ethanol and kept in a 4°C refrigerator until PWS measurements. Using this protocol, one slide each was prepared of control vector HT29 cells and one of EGFR knockdown HT29 cells. The two slides were measured unstained back-to-back on the first-generation spectrometer-based PWS system and the HTPWS instrument. The same 25 cells from each cell line were measured to allow for statistical comparison of the data. Previous PWS measurements of the CV and EGFR knockdown HT29 cells had shown significantly lower L_d values associated with the EGFR knockdown cells compared with CV.[16] Performing a second experiment with these HT29 cell lines on both the HTPWS system and the first-generation system yielded the same result. Figure 6 shows distribution, average L_d values, and the corresponding effect size for the CV and the EGFR knockdown cells. Comparison of the results from both PWS systems shows similar effect sizes for the differences between the mean L_d values for the CV and EGFR cell types, 1.16 for the HTPWS system and 1.23 for the spectrometer, respectively. P values were also comparable with 0.0007 for the HTPWS instrument and 0.0002 for the spectrometer-based PWS instrument.

Clinical lung cancer diagnostic performance with the HTPWS system was also evaluated in a small experiment including 23 patients, consisting of 9 patients with cancer and 14 smokers. This human study was performed in accordance with the Institutional Review Board at NorthShore University HealthSystem. Cells were brushed from each patient's cheek and smeared onto a glass slide before being fixed in 95% ethanol and stained using Papanicolaou stain just prior to measurement. For each patient, approximately 30 cells were measured and used to determine mean L_d values for the individual patients as well as for each diagnostic category. Measurements were also performed on the first-generation system to correlate L_d measurements between the two systems. Diagnostic performance of the system was represented by quantifying the difference in the mean L_d's of the cancer and smoker patient groups using the data collected from all 23

Fig. 5 Dependence of mean phantom L_d as a function of the diameter of spheres making up the phantom. L_d increases linearly with the size of the spheres making up the phantom. Error bars display the standard error of the mean L_d value for each phantom.

Fig. 6 L_d measurements on CV and EGFR knockdown HT29 cell lines. The distribution of the cell L_d values (a) and the mean L_d values are shown (b). The CV cell line measured higher average values for L_d compared with EGFR knockdown cells with $p = 0.0007$ and an effect size of 1.16.
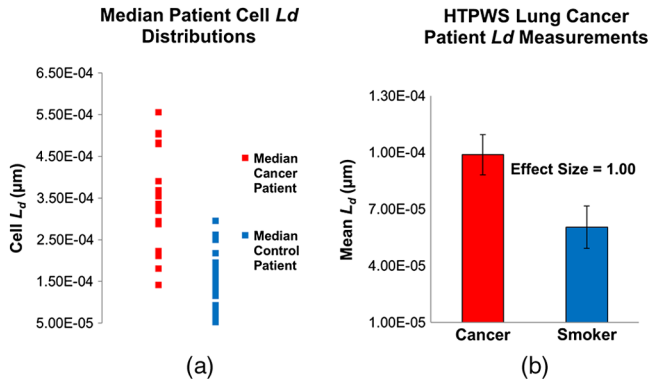
Fig. 7 HTPWS lung cancer patient L_d results. The distribution of the cell L_d values is shown for the median patients in each group (a), and the mean patient L_d values are plotted (b). Cancer patients had significantly higher average L_d values compared with smokers, $p = 0.02$ and effect size $= 1.00$.

patients. Average L_d measurements were computed for each patient and for two groups, patients with cancer and smokers. Figure 7 shows the diagnostic results for the smoker and cancer groups. The cancer group had a significantly higher average L_d compared with the smoker group as measured with the HTPWS instrument, $p = 0.02$ and effect size $= 1.00$.

Similar results to those in Fig. 7 were achieved with the first-generation PWS instrument. Cancer patients had significantly higher L_d values than smokers with $p = 0.03$ and effect size $= 0.90$. To verify consistent results between the two systems, correlation between individual cell L_d values and patient L_d values was plotted for the two systems. Figure 8 shows the correlation between patient L_d values for both systems. The correlation between the patient L_d values for the two systems yielded $R^2 = 0.93$. For individual cell L_d values, the correlation was $R^2 = 0.92$. The greater variance in the L_d values observed for both groups in this study can be attributed to some focus error with this initial version of the automated measurement system and variability in slide quality due to sample collection, smearing, and storage protocols.

4 Discussion

An effective and viable clinical cancer screening technology combines both sensitivity and performance. More specifically,
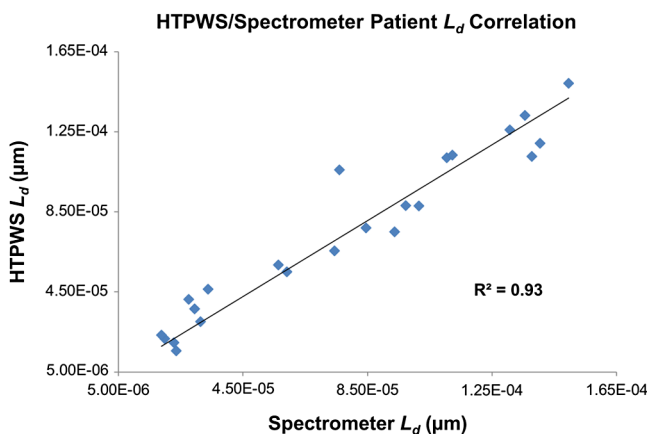
Fig. 8 Correlation of patient L_d values between the HTPWS system and first-generation spectrometer-based PWS system ($R^2 = 0.93$).

the device needs to be sensitive to the earliest known indications of disease such as intracellular nanoarchitectural changes and have the performance to achieve sufficient patient throughput on large at-risk populations in a primary-care setting.[15] Prior to the development of the HTPWS system, PWS had demonstrated nanoscale sensitivity and diagnostic capability in clinical experiments.[15,26–28] However, the throughput required to perform large-scale, multicenter clinical research studies with the PWS technology did not exist. The HTPWS system not only significantly reduces the time required to measure each slide, but also automates much of the process, minimizing the amount of work that a user needs to do to complete a measurement. With the implementation described here, the user is only required to select the cells to be measured and set up the measurement parameters before the software takes over and completes the measurement. While the performance figures for the current implementation cannot yet map, select cells, and complete a measurement in the 4 to 8 min/slide that commercial cytopathology systems achieve, improvement is ongoing as more advanced software algorithms for efficient automation of all aspects of measurement are developed. Future work on the system will seek to test the diagnostic performance of the technique in multicenter clinical research studies, while further software development continues to improve HTPWS measurement times.

In comparison to the first-generation system, the new HTPWS system is much easier to focus as live view of the cell is possible with the detector camera, whereas a separate camera was required for live view of the sample on the original system. Consequently, focusing and sharpness of the final images are both dramatically improved, since the entire image of the cell is visible at the detector at any given time rather than a single 10-μm section of the sample that must be used as part of a reconstruction to get all the spatial information. It is for this reason as well, that spatial resolution of the cell images is improved with the HTPWS system, as this no longer corresponds to the slit width of a spectrometer and is instead set by the pixel size of the detector. While spectral resolution with the AOTF is less than that of the spectrometer (5 nm compared with 4 nm or better, respectively), the HTPWS system still successfully recorded the same spectra in our comparisons shown in Fig. 4, and maintained nanoscale sensitivity as demonstrated in Fig. 5. Diagnostic performance was comparable for both systems, and L_d values correlated well across all measurements performed on both systems.

While PWS has previously completed initial clinical research studies, it was not possible to do high-patient volume, multicenter studies to establish credible clinical performance figures due to extremely low-system throughput. HTPWS has the potential to take on significant clinical research studies and develop into a rapid screening technique that can be used in a primary-care setting. As an example, lung cancer is by far the most deadly cancer, accounting for 29% of cancer deaths among males and 26% among females as compared with the second leading causes 9% (prostate) and 14% (breast), respectively.[29] In the case of lung cancer, many screening techniques have been proposed and tested for early stage detection including chest radiography, sputum cytology, and low-dose computed tomography (LDCT). Until a recent study on LDCT screening showed a 20% reduction in mortality, no screening method had demonstrated a significant reduction in lung cancer mortality.[2,30] Despite this result, LDCT presents several challenges as a screening methodology: high cost, selection of the high-risk screening group,

exposure to ionizing radiation, false-positive results, and sensitivity to early stage lesions.[31] In contrast, HTPWS shows significant potential as the first step in a tiered screening protocol on the entire at-risk population for lung cancer, whereby patients at greatest risk of harboring lesions can be selected for more risky and expensive second-tier tests such as LDCT. Furthermore, HTPWS appears particularly suited for this role in a screening protocol for lung cancer, because it is quantitative, nanoscale sensitive, low cost, and minimally invasive. These facts make it ideally suited to stratify the entire large at-risk population for lung cancer in a primary-care setting via a procedure as simple as a cheek swab without requiring significant additional clinical resources and cost such as interpretation by a specialist.

5 Conclusion

HTPWS demonstrates the potential of spectral nanocytology as a high-throughput quantitative screening tool for cancer. This high-throughput version of the previously developed PWS technology has comparable performance to all previous versions and improves upon those versions in critical areas, specifically sample throughput with a 5× to 14× increase in speed. In contrast to the previous generations of PWS, HTPWS is a potentially clinically relevant technique as it can achieve the patient throughput levels necessary to participate in the large-scale, multicenter clinical trials necessary to clinically demonstrate the screening performance of quantitative spectral nanocytology.

References

1. A. M. Minino et al., "Deaths: final data for 2008," *Natl. Vital Stat. Rep.* **59**, 1–126 (2011).
2. J. Vansteenkiste et al., "Screening and early detection of lung cancer," *Ann. Oncol.* **23**(Suppl 10), x320–x327 (2012).
3. F. S. Miller, L. E. Nagel, and M. B. Kenny-Moynihan, "Implementation of the ThinPrep imaging system in a high-volume metropolitan laboratory," *Diagn. Cytopathol.* **35**(4), 213–217 (2007).
4. A. W. Levi et al., "Implementation of FocalPoint GS location-guided imaging system: experience in a clinical setting," *Cancer Cytopathol.* **120**(2), 126–133 (2012).
5. T. M. Elsheikh et al., "Increasing cytotechnologist workload above 100 slides per day using the ThinPrep imaging system leads to significant reductions in screening accuracy," *Cancer Cytopathol.* **118**(2), 75–82 (2010).
6. A. W. Levi et al., "Increasing cytotechnologist workload above 100 slides per day using the BD FocalPoint GS imaging system negatively affects screening performance," *Am. J. Clin. Pathol.* **138**(6), 811–815 (2012).
7. T. M. Elsheikh et al., "American society of cytopathology workload recommendations for automated pap test screening: developed by the productivity and quality assurance in the era of automated screening task force," *Diagn. Cytopathol.* **41**(2), 174–178 (2013).
8. S. S. Kecheril et al., "Automated lung cancer detection by the analysis of glandular cells in sputum cytology images using scale space features," *Signal Image Video Process.*, 1–13 (2013).
9. S. Doyle et al., "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomed. Eng.* **59**(5), 1205–1218 (2012).
10. P. Filipczuk, M. Kowal, and A. Obuchowicz, "Automatic breast cancer diagnosis based on K-means clustering and adaptive thresholding hybrid segmentation," *Image Process. Commun. Challenges 3* **102**, 295–302 (2011).
11. N. Ikeda et al., "Malignancy associated changes in bronchial epithelial cells and clinical application as a biomarker," *Lung Cancer* **19**(3), 161–166 (1998).
12. B. Susnik et al., "Malignancy-associated changes in the breast—changes in chromatin distribution in epithelial-cells in normal-appearing tissue adjacent to carcinoma," *Anal. Quant. Cytol. Histol.* **17**(1), 62–68 (1995).
13. M. Us-Krasovec et al., "Malignancy associated changes in epithelial cells of buccal mucosa: a potential cancer detection test," *Anal. Quant. Cytol. Histol.* **27**(5), 254–262 (2005).
14. R. A. Kemp et al., "Detection of malignancy associated changes in cervical cell nuclei using feed-forward neural networks," *Anal. Cell. Pathol.* **14**(1), 31–40 (1997).
15. H. K. Roy et al., "Optical detection of buccal epithelial nanoarchitectural alterations in patients harboring lung cancer: implications for screening," *Cancer Res.* **70**(20), 7748–7754 (2010).
16. H. Subramanian et al., "Optical methodology for detecting histologically unapparent nanoscale consequences of genetic alterations in biological cells," *Proc. Natl. Acad. Sci. U. S. A.* **105**(51), 20118–20123 (2008).
17. H. Subramanian et al., "Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy," *Cancer Res.* **69**(13), 5357–5363 (2009).
18. S. Chaitchik et al., "An evaluation of the SCM test for the diagnosis of cancer of the breast," *Eur. J. Cancer Clin. Oncol.* **24**(5), 861–866 (1988).
19. M. Deutsch and A. Weinreb, "Validation of the SCM-test for the diagnosis of cancer," *Eur. J. Cancer Clin. Oncol.* **19**(2), 187–193 (1983).
20. C. MacAulay et al., "Malignancy-associated changes in bronchial epithelial cells in biopsy specimens," *Anal. Quant. Cytol. Histol.* **17**(1), 55–61 (1995).
21. Y. Liu et al., "Elastic backscattering spectroscopic microscopy," *Opt. Lett.* **30**(18), 2445–2447 (2005).
22. J. S. Kim et al., "The influence of chromosome density variations on the increase in nuclear disorder strength in carcinogenesis," *Phys. Biol.* **8**(1), 015004 (2011).
23. Y. Stypula et al., "Understanding biological mechanisms of nuclear disorder strength in early carcinogenesis," *Gastroenterology* **140**, S765–S766 (2011).
24. L. Cherkezyan et al., "Targeted alteration of real and imaginary refractive index of biological cells by histological staining," *Opt. Lett.* **37**(10), 1601–1603 (2012).
25. E. Hughlett and P. Kaiser, "An autofocus technique for imaging microscopy," in *ICASSP-92–1992 Int. Conf. Acoustics, Speech, and Signal Processing*, Vols. 3, pp. C93–C96 (1992).
26. D. Damania et al., "Insights into the field carcinogenesis of ovarian cancer based on the nanocytology of endocervical and endometrial epithelial cells," *Int. J. Cancer* **133**(5), 1143–1152 (2013).
27. D. Damania et al., "Nanocytology of rectal colonocytes to assess risk of colon cancer based on field cancerization," *Cancer Res.* **72**(11), 2720–2727 (2012).
28. H. Subramanian et al., "Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy," *Cancer Res.* **69**(13), 5357–5363 (2009).
29. R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2012," *CA Cancer J. Clin.* **62**(1), 10–29 (2012).
30. D. R. Aberle et al., "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N. Engl. J. Med.* **365**, 395–409 (2011).
31. P. B. Bach et al., "Benefits and harms of CT screening for lung cancer: a systematic review," *JAMA* **307**, 2418–2429 (2012).