High Wavenumber Raman Spectroscopy for in Vivo Detection of Cervical Dysplasia

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2009

Impact Factor: 5.64 · DOI: 10.1021/ac9015159 · Source: PubMed

Anal. Chem. **2009,** *81,* 8908–8915

High Wavenumber Raman Spectroscopy for in Vivo Detection of Cervical Dysplasia

Jianhua Mo,[†] Wei Zheng,[†] Jeffrey J. H. Low,[‡] Joseph Ng,[‡] A. Ilancheran,[‡] and Zhiwei Huang*,[†]

Optical Bioimaging Laboratory, Department of Bioengineering, National University of Singapore, Singapore 117576, and Division of Gynecologic Oncology, Department of Obstetrics & Gynecology, National University Hospital and National University of Singapore, Singapore 119074

Raman spectroscopy is a vibrational spectroscopic technique capable of optically probing the biomolecular changes associated with neoplastic transformation. The purpose of this study was to apply near-infrared (NIR) Raman spectroscopy in the high wavenumber (HW) region (2800−3700 cm^{-1}) for in vivo detection of cervical dysplasia. A rapid-acquisition NIR Raman spectroscopy system associated with a ball-lens fiber-optic Raman probe was developed for in vivo spectroscopic measurements at 785 nm excitation. A total of 92 in vivo HW Raman spectra (46 normal, 46 dysplasia) were acquired from 46 patients with Pap smear abnormalities of the cervix. Significant difference in Raman intensities of prominent Raman bands at 2850 and 2885 cm^{-1} (CH_2 stretching of lipids), 2940 cm^{-1} (CH_3 stretching of proteins), and the broad Raman band of water (peaking at 3400 cm^{-1} in the 3100−3700 cm^{-1} range) were observed in normal and dysplasia cervical tissue. The diagnostic algorithms based on principal components analysis and linear discriminant analysis together with the leave-one-patient-out cross-validation method on in vivo HW Raman spectra yielded a diagnostic sensitivity of 93.5% and specificity of 97.8% for dysplasia tissue identification. This study demonstrates for the first time that HW Raman spectroscopy has the potential for the noninvasive, in vivo diagnosis and detection of precancer of the cervix.

Cervical cancer is the second most common malignancy and also has the fifth highest mortality of different cancer deaths in women worldwide.[1] Early detection and localization of precancer and cancer in the cervix with effective treatment is crucial to improving the survival rates. Screening methods, such as visual inspection after applying acetic acid (VIA) or Lugol's iodine (VILI) and Pap smear, have been commonly practiced for cervical intraepithelial neoplasia (CIN) diagnosis and detection at obstetrics and gynecology (O&G) clinics,[2] and the diagnostic sensitivities and specificities for CIN screening of 61.8% and 86.5% for VIA; of 73.7% and 86.6% for VILI, and of 34.3% and 94.6% for Pap smear, respectively, have been reported in literature.[2] Although the white-light colposcopy has been employed to improve the diagnostic sensitivity (∼96%), the detection specificity still remains poor (∼48%) in identifying cervical tissue changes associated with neoplastic transformation.[2,3] Excisional biopsy currently is the standard diagnostic approach for cervical precancer and cancer diagnosis, but it is invasive and impractical for screening high-risk patients who may have multiple suspicious lesions.

Optical spectroscopic techniques, such as fluorescence spectroscopy, diffuse reflectance, infrared (IR) spectroscopy, and Raman spectroscopy, have been comprehensively investigated for in vitro and in vivo diagnosis of malignancies in various organs, including the cervix.[4−20] IR spectroscopy and Raman spectroscopy

* To whom correspondence should be addressed. Tel: +65- 6516-8856. Fax: +65- 6872-3069. E-mail: biehzw@nus.edu.sg.

[†] National University of Singapore.

[‡] National University Hospital and National University of Singapore.

(1) Parkin, D. M.; Bray, F.; Ferlay, J.; Pisani, P. *Int. J. Cancer* **2001**, *94*, 153–156.

(2) Arbyn, M.; Sankaranarayanan, R.; Muwonge, R.; Keita, N.; Dolo, A.; Mbalawa, C. G.; Nouhou, H.; Sakande, B.; Wesley, R.; Somanathan, T.; Sharma, A.; Shastri, S.; Basu, P. *Int. J. Cancer* **2008**, *123*, 153–160.

(3) Mitchell, M. F.; Schottenfeld, D.; Tortolero-Luna, G.; Cantor, S. B.; Richards-Kortum, R. *Obstet. Gynecol.* **1998**, *91*, 626–631.

(4) Hung, J.; Lam, S.; Leriche, J. C.; Palcic, B. *Lasers Surg. Med.* **1991**, *11*, 99–105.

(5) Huang, Z.; Zheng, W.; Xie, S.; Chen, R.; Zeng, H.; McLean, D. I.; Lui, H. *Int. J. Oncol.* **2004**, *24*, 59–64.

(6) Subhash, N.; Mallia, J. R.; Thomas, S. S.; Mathews, A.; Sebastian, P.; Mahadevan, J. *J. Biomed. Opt.* **2006**, *11*, 014018.

(7) Bensalah, K.; Peswani, D.; Tuncel, A.; Raman, J. D.; Zeltser, I.; Liu, H.; Cadeddu, J. *Urology* **2009**, *73*, 178–181.

(8) Shim, M. G.; Song, L. M.; Marcon, N. E.; Wilson, B. C. *Photochem. Photobiol.* **2000**, *72*, 146–150.

(9) Huang, Z.; McWilliams, A.; Lui, H.; McLean, D. I.; Lam, S.; Zeng, H. *Int. J. Cancer* **2003**, *107*, 1047–1052.

(10) Stone, N.; Kendall, C.; Shepherd, N.; Crow, P.; Barr, H. *J. Raman Spectrosc.* **2002**, *33*, 564–573.

(11) Huang, Z.; Teh, S. K.; Zheng, W.; Mo, J.; Lin, K.; Shao, X.; Ho, K. Y.; Teh, M.; Yeoh, K. G. *Opt. Lett.* **2009**, *34*, 758–760.

(12) Georgakoudi, I.; Jacobson, B. C.; Müller, M. G.; Sheets, E. E.; Badizadegan, K.; Carr-Locke, D. L.; Crum, C. P.; Boone, C. W.; Dasari, R. R.; Van, D. J.; Feld, M. S. *Cancer Res.* **2002**, *62*, 682–687.

(13) Huang, Z.; Lui, H.; Chen, X. K.; Alajlan, A.; McLean, D. I.; Zeng, H. *J. Biomed. Opt.* **2004**, *9*, 1198–1205.

(14) Mahadevan-Jansen, A.; Mitchell, M. F.; Ramanujamf, N.; Malpica, A.; Thomsen, S.; Utzinger, U.; Richards-Kortum, R. *Photochem. Photobiol.* **1998**, *68*, 123–132.

(15) Huang, Z.; Lui, H.; McLean, D. I.; Korbelik, M.; Zeng, H. *Photochem. Photobiol.* **2005**, *81*, 1219–1226.

(16) Teh, S. K.; Zheng, W.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; Huang, Z. *J. Biomed. Opt.* **2008**, *13*, 034013.

(17) Teh, S. K.; Zheng, W.; Lau, P. D.; Huang, Z. *Analyst* **2009**, *134*, 1232–1239.

(18) Wong, P. T. T.; Lacelle, S.; Fung, K. F. M.; Sentermann, M.; Mikhael, N. Z. *Biospectroscopy* **1995**, *1*, 357–364.

10.1021/ac9015159 CCC: $40.75
Published on Web 10/09/2009

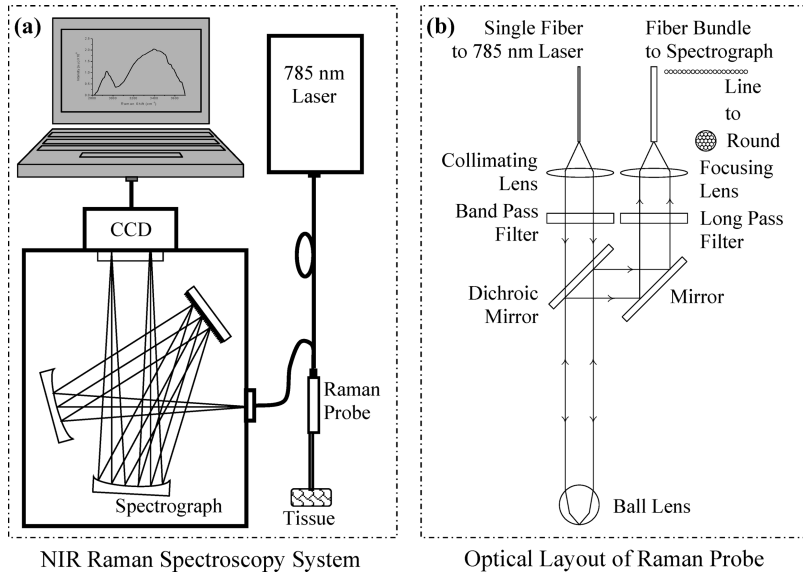

Figure 1. (a) Block diagram of the portable NIR Raman spectroscopy system developed for in vivo Raman measurements. (b) Optical layout of the fiber-optic Raman probe coupled with a ball lens.

are complementary vibrational spectroscopic techniques that are capable of probing the changes of biochemical and biomolecular structures and compositions of tissue associated with malignant transformation.[9,10,15−22] Fourier-transformed IR (FT-IR) spectroscopy has been used for cervical precancer diagnosis and evaluation, but most work is limited to the in vitro tissue measurements due to the difficulties with sample preparation and strong IR absorption interference by tissue water.[18−21] With the use of near-infrared (NIR) excitation light (e.g., 785 nm), NIR Raman spectroscopy can interrogate with tissue in deeper regions, due to less water absorption, and experience far less interference from tissue autofluorescence compared to the ultraviolet (UV) or shorter visible light excitation, thereby enabling in vivo tissue diagnosis and characterization.[9,10,14,15] NIR Raman spectroscopic differences between normal and dysplasia cervical tissues have been observed,[14] demonstrating the ability of NIR Raman spectroscopy for precancer diagnosis of the cervix. To date, most cervical NIR Raman studies were centered on the fingerprint region (i.e., 800−1800 cm^{-1}) owing to the great wealth of biochemical information contained in this spectral region for tissue characterization.[9,10,14] However, the strong fluorescence background and silica Raman signal arising from fiber optic Raman probe also fall into the fingerprint range, which severely interfere with the detection of the inherently weak tissue Raman signal.[14,15] This makes the light filtering modules in Raman probe design complicated, bulky, and unsuitable for in vivo biomedical applications. To tackle this problem, Puppels and co-workers[23] proposed to use Raman spectroscopy in the so-called high wavenumber (HW) region (i.e., 2400−3800 cm^{-1}) for tissue diagnosis and characterization. The main advantages of using HW Raman spectroscopy are the significant reduction of fluorescence/Raman background from optical fibers,[23] more intense tissue Raman signals generated compared to the fingerprint region,[24] as well as the possibility of an unfiltered, simplified single fiber Raman probe design for facilitating in vivo Raman clinical procedures.[25] For instance, HW Raman studies[26] showed that basal cell carcinoma (BCC) lesions could be distinguished with a 100% prediction accuracy, suggesting that diagnostic information extracted from the HW Raman spectral region could compete with those from the fingerprint region. To date, the clinical potential of HW Raman spectroscopy for in vivo identification of cervical dysplasia (i.e., precancer) has yet been reported in literature. The main aims of this work are to develop a rapid fiber-optic HW Raman spectroscopy system and to evaluate the feasibility of utilizing HW Raman spectroscopy for in vivo diagnosis of cervical precancer. Multivariate statistical techniques, including principal components analysis (PCA) and linear discriminant analysis (LDA), are employed to develop diagnostic algorithms for differentiation between normal and precancer cervical tissue. Receiver operating characteristic (ROC) testing is also conducted to further evaluate the performance of PCA−LDA algorithms on in vivo HW Raman spectroscopy for cervical precancer diagnosis.

MATERIALS AND METHODS

Raman Instrumentation. Figure 1a shows the NIR Raman spectroscopy system developed for in vivo tissue measurements of the cervix. The system mainly consists of an external cavity stabilized diode laser at 785 nm (100 mW, B&W TEK Inc., Newark, DE), a high-throughput spectrometer (QE65000, Ocean Optics Inc., FL) equipped with an NIR-enhanced charge-coupled device (CCD) detector (S7031−1006, 1024 × 58 with pixel sizes

(19) Wood, B. R.; Chiriboga, L.; Yee, H.; Quinn, M. A.; McNaughton, D.; Diem, M. *Gynecol. Oncol.* **2004**, *93*, 59–68.
(20) Cohenford, M. A.; Rigas, B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 15327–15332.
(21) Chang, J. I.; Huang, Y. B.; Wu, P. C.; Chen, C. C.; Huang, S. C.; Tsai, Y. H. *Gynecol. Oncol.* **2003**, *91*, 577–583.
(22) Teh, S. K.; Zheng, W.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; Huang, Z. *J. Raman Spectrosc.* **2009**, *40*, 908–914.
(23) Santos, L. F.; Wolthuis, R. S. K.; Almeida, R. M.; Puppels, G. *J. Anal. Chem.* **2005**, *77*, 6747–6752.
(24) Eikje, N. S.; Ozaki, Y.; Aizawa, K.; Arase, S. *J. Biomed. Opt.* **2005**, *10*, 014013.
(25) Nazemi, J. H.; Brennan, J. F., III. *J. Biomed. Opt.* **2009**, *14*, 034009.
(26) Nijssen, A.; Maquelin, K.; Santos, L. F.; Caspers, P. J.; Schut, T. C. B.; Hollander, J. C. D. *J. Biomed. Opt.* **2007**, *12*, 034004.

of 24.6 μm, QE > 90%, Hamamatsu), and a specially designed hand-held fiber-optic Raman probe coupled with a ball lens for maximizing both the tissue excitation and tissue Raman collections. The Raman probe (Figure 1b) comprises two optical arms (one for laser light delivery and one for scattered tissue Raman collection) integrated with optical filtering modules. The 785 nm excitation light is coupled into the excitation arm of the Raman probe through a 200 μm fiber (NA = 0.22) and delivered into the filtering module incorporated with an NIR lens for excitation light collimation and a narrow band-pass (BP) filter (centered at 785 nm, fwhm = ±2.5 nm) for removing fiber fluorescence and laser noise, and then the laser light is focused onto the tissue with a beam size of ~0.2 mm through an NIR-coated sapphire ball lens (5 mm in diameter, refractive index $n = 1.77$) mounted on the tip of the Raman probe. The backscattered tissue Raman photons from the cervical tissue (of up to 700 μm in tissue depth[27−30]) are efficiently collected by the same ball lens (collection cone angles of 55°) and subsequently reflected back into the collection arm through a dichroic mirror and a reflectance mirror. The backscattered light is further filtered by using an edge long pass (LP) filter (cutoff at 800 nm) to block the Rayleigh scattered laser light while allowing the Raman scattered light to pass into the spectrometer through a specially designed round-to-line fiber bundle adapter (28 × 50 μm, NA = 0.22) to improve the signal-to-noise ratio (S/N) of up to 7.6-fold ($\sqrt{58}$) through vertical binning of the entire CCD[31] for maximizing in vivo tissue Raman detection. All the optics of the Raman probe is sealed into the stainless steel sleeve (outer diameter of 8 mm) with a polytetrafluoroethylene (PTFE) gasket for in vivo cervical tissue measurements. The system acquires HW Raman spectra over the wavenumber range of 2800−3700 cm^{-1}, and each spectrum was acquired within 1 s with light irradiance of 1.6 W/cm^2, which is less than the American National Standards Institute (ANSI) maximum permissible skin exposure limit set out for a 785-nm laser beam.[32] Our further calculations based on the finite difference thermal model[33,34] and cervical tissue optical properties[27−30] indicate that the maxima tissue temperature rise without consideration of cooling effects (e.g., perfusion and evaporation in tissue and air) is only about 0.3 °C after 1 min of 785 nm laser radiation with an incident power of 15 mW on a spot diameter of 200 μm during HW Raman measurements. This estimated temperature rise is far below the level to generate cytotoxicity in tissue and cells,[35] suggesting that the laser power density used in this study is safe for in vivo HW Raman measurements. The spectral resolution of the system is about 5 cm^{-1}. All wavelength-calibrated HW Raman spectra were also corrected for the wavelength-dependence of the system using a standard lamp (RS-10, EG&G Gamma Scientific, San Diego, CA).

Patients. A total of 46 women (mean age of 42.6 years old) who underwent colposcopy examinations for abnormal Pap smears were recruited through colposcopy clinics at the National University Hospital (NUH) of Singapore. All patients preoperatively signed an informed consent permitting the in vivo spectroscopic measurements of the cervix before the examinations were carried out. This study was approved by the Ethics Committee of the National Healthcare Group (NHG) of Singapore. Prior to Raman spectroscopic measurements, a complete routine colposcopic examination was performed on the patients by experienced colposcopists. A 5% acetic acid was applied to the cervix for a couple of minutes to allow the inspection of acetic acid whitening changes of epithelium, which was then graded according to the degree of acetic-whiteness, morphologic textures, and the vascular pattern.[36,37] In vivo HW Raman spectra were then acquired from the suspicious lesion sites for each patient through gently placing the ball-lens Raman probe on the cervix. Each abnormal site measured was biopsied and then submitted for histopathologic examination. In addition, in vivo HW Raman spectra were also measured from the surrounding normal sites that appear completely normal in the colposcopists' opinion (i.e., normal tissue does not exhibit colored patterned changes that only accompany dysplastic epithelium),[36] but no biopsies were taken from normal appearing tissue. A total of 92 in vivo HW Raman spectra were acquired from 46 patients with cervical abnormalities, in which 46 were from normal sites, while 46 from cervical dysplasia lesions [i.e., 10 low-grade cervical intraepithelial neoplasia (CIN1) and 36 high-grade cervical intraepithelial neoplasia (9 CIN2 and 27 CIN3)].

Data Preprocessing. The raw HW Raman spectra (2800−3700 cm^{-1}) measured from in vivo cervical tissue represented a composition of Raman signal, autofluorescence background and noise. Thus, the raw spectra were preprocessed by a first-order Savitsky−Golay filter (window width of 3 pixels, which corresponded to the system spectral resolution) to reduce noise.[38] A first-order polynomial was found to be optimal for fitting the autofluorescence background in the noise-smoothed spectrum, and this polynomial was then subtracted from the raw spectrum to yield the tissue HW Raman spectrum alone. Each background-subtracted HW Raman spectrum was also normalized to the integrated area under the curve from 2800 to 3700 cm^{-1}, enabling a better comparison of the spectral shapes and relative Raman band intensities among the different cervical tissues.

Multivariate Statistical Analysis. Principal component analysis was first performed on the in vivo tissue HW Raman data set to reduce the dimension of Raman spectral space (of 138 intensities within 2800−3700 cm^{-1}) while retaining the most diagnostically significant information for effective tissue classification. To eliminate the influence of inter- and/or intra-subject spectral variability on PCA, the entire spectra were standardized so that the mean of the spectra was zero and the standard deviation (SD) of all the spectral intensities was one.

(27) Jaillon, F.; Zheng, W.; Huang, Z. *Phys. Med. Biol.* **2008**, *53*, 937–951.
(28) Jaillon, F.; Zheng, W.; Huang, Z. *Appl. Opt.* **2008**, *47*, 3152–3157.
(29) Abdul-Karim, F.; Fu, Y. S.; Reagan, J. W.; Wentz, W. B. *Obstet. Gynecol.* **1982**, *60*, 210–214.
(30) Kolanjiappan, K.; Manoharan, S.; Kayalvizhi, M. *Clin. Chim. Acta* **2002**, *326*, 143–149.
(31) Huang, Z.; Zeng, H.; Hamzavi, I.; McLean, D. I.; Lui, H. *Opt. Lett.* **2001**, *26*, 1782–1784.
(32) *American National Standard for the Safe Use of Lasers*; ANSI Standard 2136.1−1986; American National Standards Institute: Washington, DC, 1986.
(33) Incropera, F. P.; Witt, D. P. D. *Fundamentals of Heat and Mass Transfer*; John Wiley & Sons: New York, 1990.
(34) Torres, J. H.; Motamedi, M.; Pearce, J. A.; Welch, A. J. *Appl. Opt.* **1993**, *32*, 597–606.
(35) Thomsen, S. *Photochem. Photobiol.* **1991**, *53*, 825–835.
(36) Apgar, B. S.; Brotzman, G. L.; Spitzer, M. *Colposcopy, Principles and Practice: An Integrated Textbook and Atlas*, 2nd ed.; Saunders/Elsevier: Philadelphia, 2008.
(37) Jeronimo, J.; Schiffman, M. *Am. J. Obstet. Gynecol* **2006**, *195*, 349–353.
(38) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627–1639.

Mean centering ensures that the PCs form an orthogonal basis.[39,40] The standardized Raman data sets were assembled into data matrices with wavenumber columns and individual case rows. Thus, PCA was conducted on the standardized spectral data matrices to generate PCs comprising a reduced number of orthogonal variables that accounted for most of the total variance in original spectra. Each loading vector is related to the original spectrum by a variable called the PC score, which represents the weight of that particular component against the basis spectrum. PC scores reflect the differences between different classes. Paired two-sided Student's *t*-test[40] was used to identify the most diagnostically significant PCs ($p < 0.05$). These significant PC scores are last selected as input for the development of linear discriminant analysis algorithms for binary-class classification. LDA determines the discriminant function that maximizes the variances in the data between groups while minimizing the variances between members of the same group. The performance of the diagnostic algorithms rendered by the PCA−LDA models for correctly predicting the tissue groups (e.g., normal vs dysplasia) was estimated in an unbiased manner using the leave-one-patient (i.e., paired spectra associated with one patient)-out cross-validation method[39,41] on all model spectra. In this method, the paired spectra from each patient were held out from the data set and the entire algorithm including PCA and LDA was redeveloped using the remaining tissue spectra. The algorithm was then used to classify the withheld spectra. This process was repeated until all withheld spectra were classified. For the assessment of diagnostic sensitivity and specificity of HW Raman spectroscopy technique, histopathologic examination results serve as the gold standard.

RESULTS

Figure 2a shows the comparison of mean in vivo HW Raman spectra ±1 standard deviation (SD) of normal ($n = 46$) and dysplasia ($n = 46$) cervical tissue. Prominent Raman bands, such as 2850 and 2885 cm^{-1} (CH_2 stretching of lipids), 2940 cm^{-1} (CH_3 stretching of proteins), and the broad Raman band of water (OH stretch vibrations peaking at 3400 cm^{-1} in the 3100−3700 cm^{-1} region),[23,24] are clearly observed in both normal and dysplasia cervical tissue. Overall, precancer cervical tissue shows significantly lower intensities for the Raman bands in the 2800−3000 cm^{-1} region (paired two-sided Student's *t* test, $p < 0.001$, $n = 46$), while being higher for OH bands in the 3100−3700 cm^{-1} region (paired two-sided Student's *t* test, $p < 0.001$, $n = 46$) compared to normal tissue, as shown in the difference spectrum in Figure 2b. Nevertheless, there are significant variations and overlapping intensities of in vivo HW Raman spectra of normal and precancer tissue among intersubjects (Figure 2). Thus, tissue diagnosis simply based on spectral intensities alone is precluded. In addition, it is also difficult to standardize the absolute intensities of spectral measurements in clinical settings. Therefore, in this study we focus on analyzing spectral feature differences following a standardization/normalization data preprocessing procedure as described in the sections Data Preprocessing and Multivariate Statistical Analysis.

(39) Lachenbruch, P. A.; Mickey, M. R. *Technometrics* **1968**, *10*, 1–11.
(40) Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, 7th ed. Thomson/Brooks/Cole: Belmont, CA, 2009.
(41) Deinum, G.; Rodriguez, D.; Romer, T. J.; Fitzmaurice, M.; Kramer, J. R.; Feld, M. S. *Appl. Spectrosc.* **1999**, *53*, 938–942.

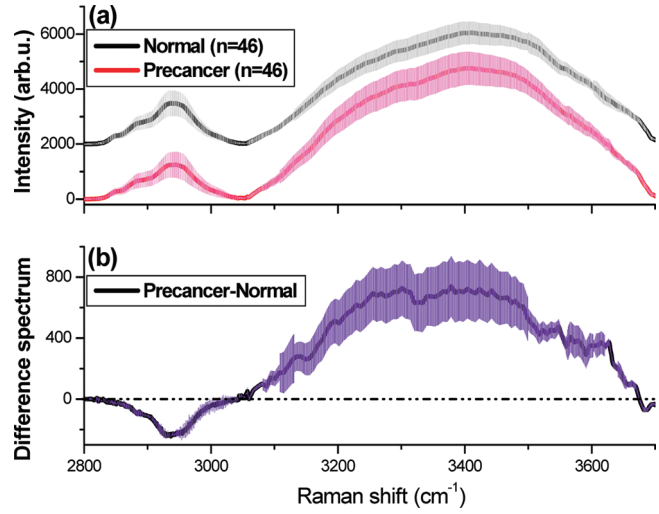

Figure 2. (a) Comparison of mean in vivo HW Raman spectra ±1 standard deviation (SD) of normal ($n = 46$) and precancer ($n = 46$) cervical tissue. (b) Difference spectrum ± 1SD difference between precancer ($n = 46$) and normal cervical tissue ($n = 46$). Note that the mean in vivo HW Raman spectrum of normal tissue was shifted vertically for better visualization (panel a); the shaded areas indicate the respective standard deviations.

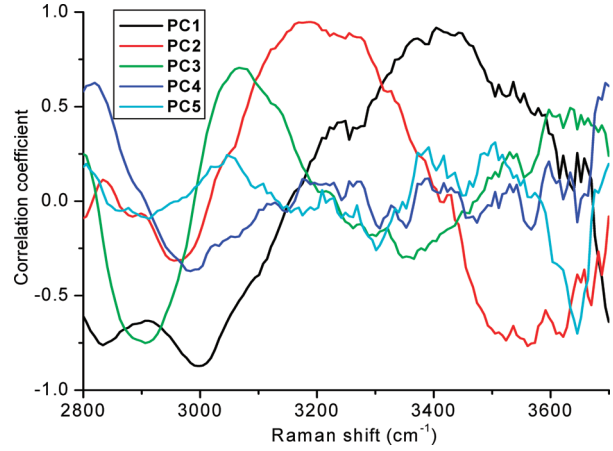

Figure 3. The first five principal components accounting for about 88% of the total variance calculated from in vivo HW Raman spectra of cervical tissue (PC1, 49.6%; PC2, 21.7%; PC3, 10.9%; PC4, 4.7%; and PC5, 1.6%).

We employ the multivariate statistical method (e.g., PCA−LDA) together with paired two-sided Student's *t*-test by incorporating the entire HW Raman spectra to determine the diagnostically significant Raman features for tissue diagnosis and classification. Figure 3 shows the first five principal components (PCs) calculated from PCA on all HW Raman spectra. The first PC accounts for the largest variance (e.g., 49.6% of the total variance), whereas the successive PCs describe the spectral features that contribute progressively smaller variances. Some PC features, such as peaks, troughs, and spectral shapes, are similar to those of tissue HW Raman spectra (Figure 2). The paired two-sided Student's *t*-test on the obtained PCs showed that there were three diagnostically significant PCs (i.e., PC1, PC4, and PC9) ($p < 0.05$) for discriminating dysplasia tissue from normal tissue. Figure 4 shows the correlations between the diagnostically significant PC scores for normal and dysplastic cervical tissue, illustrating the utility of significant PCs for classification of Raman spectra between normal

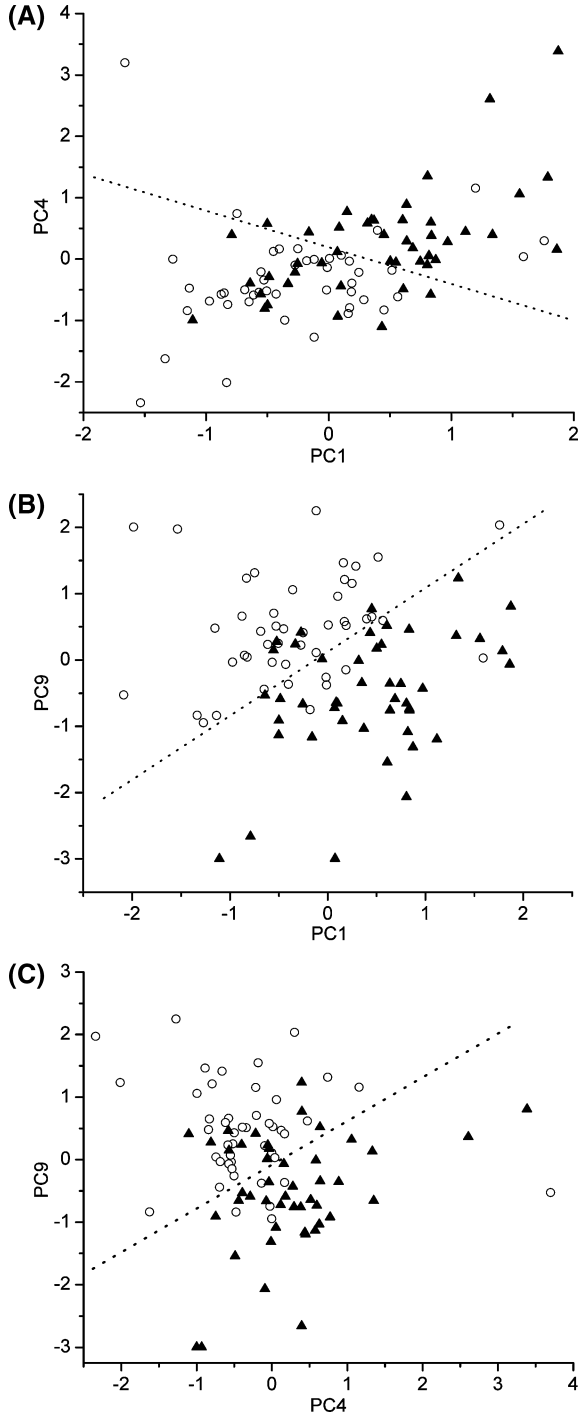

Figure 4. Scatter plots of the diagnostically significant PCs derived from in vivo HW Raman spectra of normal and precancer cervical tissue: (a) PC1 vs PC4, (b) PC1 vs PC9, and (c) PC4 vs PC9. The dotted lines (PC4 = −0.57 PC1 + 0.19; PC9 = 0.96 PC1 + 0.12; PC9 = 0.62 PC4 − 0.08) as diagnostic algorithms classify precancer from normal with sensitivities of 63.0% (29/46), 89.1% (41/46), and 73.9% (34/46) and specificities of 87.0% (40/46), 84.8% (39/46), and 87.0% (40/46), respectively. Key: (○) normal, (▲) precancer.

and precancer cervical tissues. Normal and dysplasia tissues can be largely clustered into two separate groups on the basis of different combinations of significant PCs, and the corresponding separation lines (i.e., diagnostic algorithms) in Figures 4a−c classify dysplasia from normal tissue with the sensitivity of 63.0% (29/46), 89.1% (41/46), and 73.9% (34/46) and the specificity of

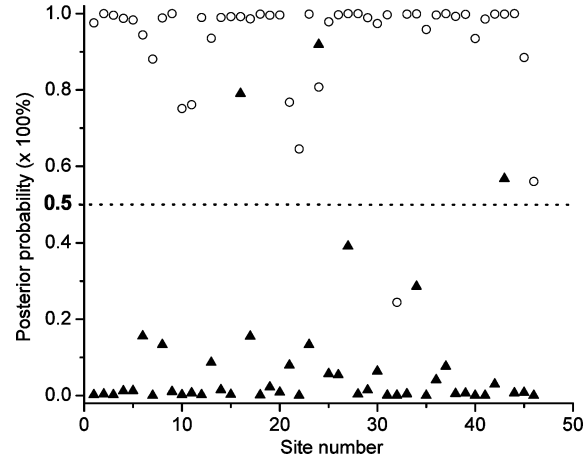

Figure 5. Scatter plot of the posterior probability of belonging to the normal and precancer cervical tissues using the PCA−LDA technique together with the leave-one-patient-out cross-validation method. The separate line yields a diagnostic sensitivity of 93.5% (43/46) and specificity of 97.8% (45/46), for identifying precancer from normal cervical tissue. Key: (○) normal, (▲) precancer.

87.0% (40/46), 84.8% (39/46), and 87.0% (40/46), respectively. These results show that selection of different combinations of significant PCs will lead to different levels of accuracy for tissue classification.

To further improve tissue diagnosis, all the three diagnostically significant PCs were loaded into the LDA model for generating effective diagnostic algorithms for tissue classification. Figure 5 shows the posterior probability of classification results based on the PCA−LDA technique together with the leave-one-patient-out cross-validation method. The PCA−LDA diagnostic algorithms yielded the diagnostic sensitivity of 93.5% (43/46) and specificity of 97.8% (45/46) for distinguishing dysplasia from normal cervical tissue.

To evaluate the performance of the PCA−LDA-based diagnostic algorithms derived from all the significant PCs of tissue HW Raman data set, the receiver operating characteristic (ROC) curve (Figure 6) was generated from the scatter plot in Figure 5 at different threshold levels to determine the correct or incorrect classification of cervical tissues. The integration area under the ROC curve is 0.98, demonstrating the efficacy of PCA−LDA diagnostic algorithms developed that utilize the entire spectral features of HW Raman spectroscopy for in vivo diagnosis of cervical precancer.

DISCUSSION

A Raman spectroscopy technique that is capable of providing a wealth of biochemical structures and compositions of tissues can be a clinically useful tool for precancer and cancer diagnosis in humans; however, an efficient and quick detection of tissue Raman signal that is inherently very weak is critical to realizing in vivo Raman clinical applications.[8,11,14,31] In this work, we have developed a rapid and portable NIR Raman spectroscopy system coupled with a ball-lens fiber-optic Raman probe to provide a high degree of collection efficiency for real-time in vivo cervical tissue measurements. Compared to the conventional fiber-optic Raman probes with a relatively small collection angle (~13°; NA = 0.22),[8] our Raman probe coupled with an NIR-coated sapphire ball lens (diameter of 5 mm, refractive index of 1.77) offers a large

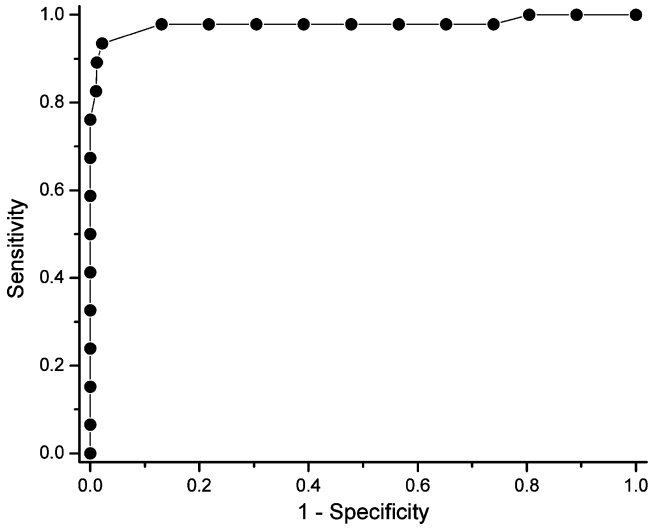

Figure 6. Receiver operating characteristic curve of discrimination results for in vivo HW Raman spectra of cervical tissue using PCA−LDA algorithms together with the leave-one-patient-out cross-validation method. The integration area under the ROC curve is 0.98, illustrating the efficacy of PCA−LDA algorithms for tissue classification.

collection angle (of up to 55°) to maximize the collection of backscattered Raman photons from tissue. In addition, the special round-to-line fiber bundle adapter connected to the ball-lens Raman probe optimizes the coupling of Raman photons collected into the spectrometer for further improving tissue Raman acquisitions.[11,31] As a result, good quality (S/N of up to 30 at 1 s integration time) HW Raman spectra in the range of 2800−3700 cm^{-1} can be acquired from cervical tissue in vivo using the rapid Raman spectroscopy system developed.

We applied this rapid HW Raman spectroscopy for the first time for the in vivo diagnosis and detection of cervical precancer during clinical colposcopic examination. The overall intensity of Raman signals of cervical dysplasia in the region (2800−3050 cm^{-1}) involving proteins (e.g., 2940 cm^{-1}) and lipids (e.g., 2850 and 2885 cm^{-1}) are significantly lower than those in normal cervical tissue. Our Monte Carlo (MC) simulations[27,28] based on the two-layer (i.e., epithelium, stroma) tissue model according to the cervical tissue optical properties[27,29,30] show that the 785 nm laser light from our ball-lens Raman probe can penetrate into the tissue depth of 690 μm in normal cervical tissue, while 630 μm in dysplastic tissue, thereby interrogating with tissue volumes comprising both epithelial and stromal tissue in the cervix. Due to the changes of tissue optical properties (e.g., thickening of epithelium) and morphologies (e.g., higher cellular density resulting from the increased nucleus to cytoplasm ratios) associated with dysplastic transformation,[16,28,42] these effects significantly attenuate the excitation light penetration and also obscure the tissue Raman photons emitted from the underlying stroma in precancer tissue as compared to normal cervical tissue. Further MC simulations indicate that the overall Raman signals in the 2800−3050 cm^{-1} ragion emitted from normal tissue are 1.15-fold stronger than dysplastic cervical tissue, which is in good agreement with the observation in our HW Raman cervical tissue measurements (Figure 2). We have also studied the differences of the two respective tissue layers (i.e., epithelium, and stroma) in contributions to the HW Raman signals between normal and dysplastic cervical tissues. The MC simulation results show that the stromal tissue (i.e., collagen) in dysplasia cervical tissue contribute only about 10% to the overall HW Raman signal in the 2800−3050 cm^{-1} region as compared to the 50% of contributions from normal stromal tissue. This also explains the significant reduction of collagen Raman signal (at 2940 cm^{-1}) observed in dysplasia tissue (Figure 2). Similar results in the finding of much reduced collagen signals associated with cervical dysplastic transformation have also been reported in the literature using fluorescence spectroscopy and Raman spectroscopy in the fingerprint region.[12,30] On the other hand, we have also observed that the water vibrational signal in the 3100−3700 cm^{-1} region is significant higher in dysplastic cervical tissue (Figure 2), indicating the increase of water contents in percentage relative to all the Raman-active components in dysplastic cervical tissue. The increase of water concentrations associated with tumor alterations in other organs had also been reported using FT-IR spectroscopy techniques.[43−45] This phenomenon is probably due to the increase of certain types of aquaporins (AQPs) at the dysplastic cervical cells, which could increase the plasma membrane osmotic water permeability of up to 10-fold, thereby facilitating influx of water molecules into the intracellular space in dysplasia tissue.[46,47] The increased DNA levels or hydration of DNA due to unfolding during cell divisions may also be responsible for the change of water Raman bands observed in cervical dysplastic lesions.[19−21] Therefore, the distinctive spectral differences observed in the in vivo HW Raman spectra of normal and dysplastic cervical tissue in this study suggest the diagnostic potential of HW Raman spectroscopy for noninvasively distinguishing dysplastic tissue from normal cervical mucosa.

We have also evaluated the diagnostic ability of the intensity ratio of protein Raman band intensity at 2940 cm^{-1} (CH_3 stretching of proteins) to the lipid band at 2850 cm^{-1} (CH_2 stretching of lipids) for cervical precancer identification. The Raman intensity ratio value of I_{2940}/I_{2850} for dysplasia tissue is 5.05 ± 0.23 (n = 46), which is significantly higher than the ratio value of 4.16 ± 0.35 (n = 46) for normal cervical tissue (paired two-sided Student's t test, $p < 0.0001$, n = 46). The higher ratio of I_{2940}/I_{2850} (CH_3/CH_2) is probably ascribed to the decrease of the membrane lipid contents and the increased formation of shorter fatty acid chains, as well as the increased of the nucleic hyperchromatism and density associated with dysplastic changes.[18,21] Similar results on the increase of CH_3/CH_2 ratio have also been observed using a FT-IR technique on malignant cervical tissues in vitro.[18−20] With the use of the intensity ratio I_{2940}/I_{2850} as a diagnostic algorithm, a sensitivity of 89.1% (41/46) and specificity of 87% (40/46) can be reached

(42) Teh, S. K.; Zheng, W.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; Huang, Z. *Br. J. Cancer* **2008**, *98*, 457–465.
(43) Hornung, R.; Pham, T. H.; Keefe, K. A.; Berns, M. W.; Tadir, Y.; Tromberg, B. J. *Reprod.* **1999**, *14*, 2908–2916.
(44) Kondepati, V.; Heise, H.; Backhaus, J. *Anal. Bioanal. Chem.* **2008**, *390*, 125–139.
(45) McIntosh, L. M.; Summers, R.; Jackson, M.; Mantsch, H. H.; Mansfield, J. R.; Howlett, M.; Crowson, A. N.; Toole, J. W. P. *J. Invest. Dermatol.* **2001**, *116*, 175–181.
(46) Verkman, A.; Hara-Chikuma, M.; Papadopoulos, M. *J. Mol. Med.* **2008**, *86*, 523–529.
(47) Wolthuis, R.; Aken, M. V.; Fountas, K.; Robinson, J. S.; Bruining, H. A.; Puppels, G. *J. Anal. Chem.* **2001**, *73*, 3915–3920.

for cervical dysplasia identification. Further investigation on employing other intensity ratios such as the Raman band intensity for lipids at 2885 cm^{-1} and water at 3400 cm^{-1} to the protein Raman band intensity at 2940 cm^{-1} also provide good differentiation between normal and dysplasia tissue (data not shown). Hence, the Raman intensity ratios reported in this study could also potentially be used as effective diagnostic algorithms for classifying dysplasia from normal gastric tissue. It should be pointed out that empirical analysis based on intensity ratios only employs a limited number of Raman peaks for tissue diagnosis. This means that only a very tiny fraction of biochemical and biomolecular information contained in Raman spectra is used for analysis; most of Raman features contained in the entire Raman spectra are discarded, limiting the diagnostic accuracy.

Since biological tissue is complex, it is likely that there are many biochemical species influencing diseases concurrently.[42,43] In this work, the multivariate statistical analysis (e.g., PCA and LDA) was also implemented by utilizing the entire HW Raman spectrum ranging from 2800 to 3700 cm^{-1} to determine the diagnostically significant spectral features for improving dysplastic cervical tissue diagnosis and classification. With the limited HW Raman spectral data sets we acquired to date, PCA−LDA together with the leave-one-patient-out cross-validation technique was applied to the HW Raman spectra for dysplastic tissue identification in an unbiased manner. The cross-validated diagnostic sensitivity of 93.5%, specificity of 97.8%, and accuracy of 95.7% achieved for distinguishing dysplasia from normal cervical tissue confirms the diagnostic potential of HW NIR Raman spectroscopy for cervical precancer diagnosis. To further verify if the utilization of the entire HW spectral region (i.e., 2800−3700 cm^{-1}) is robust for providing good tissue classification accuracy, we have also compared the diagnostic abilities based on the two Raman subregions, i.e., 2800−3050 cm^{-1} (containing proteins and lipids biomolecular constituents) and 3100−3700 cm^{-1} (representing water molecules) for cervical precancer diagnosis. PCA−LDA analysis yielded a diagnostic sensitivity of 89.1%, specificity of 89.1%, and accuracy of 89.1% in the spectral region 2800−3050 cm^{-1} and a sensitivity of 84.8%, specificity of 89.1%, and accuracy of 87% in the region 3100−3700 cm^{-1}, respectively, for cervix dysplasia identification. The highest diagnostic accuracy (~96%) achieved by employing the entire HW spectral region from 2800−3700 cm^{-1} compared to the truncated HW Raman subregions (i.e., 2800−3050 and 3100−3700 cm^{-1}) further reinforces that the unique combination of different Raman biomolecular signals ranging from water to lipids to collagen (i.e., 2800−3700 cm^{-1}) can provide better differentiation between normal and dysplastic cervical tissue. The overall diagnostic accuracy rate of 96% achieved in the 2800−3700 cm^{-1} region in this study is, in fact, comparable with those achieved via Raman spectroscopy in the fingerprint spectral region (accuracy of 90−100% in 800−1800 cm^{-1}) for in vivo detection of cervical dysplasia.[6,14] On the other hand, accurate distinction between different stages of cervical dysplasia (CIN1, CIN2/CIN3) plays an important role in a good prognosis for the patient at risk of cervical malignancies. This is because the low-grade squamous intraepithelial lesions (LGSILs) (CIN1) have a high rate to regress to normal without requiring any treatment;[36] while the high-grade squamous intraepithelial lesions (HGSILs) (CIN2/CIN3) are known more or less to progress to invasive cervical cancer.[36,37] Current colposcopic examination is prone to over-diagnosing the LGSILs as HGSILs. Meanwhile, accurate discrimination between normal/benign and LGSILs is even more challenging, because some benign changes in the cervix may also show some features of LGSILs.[36] For instance, HPV infection or inflammation may also exhibit vascular atypia, which is the hallmark of higher grade lesion, and metaplasia may exhibit aceto-white due to the cellular changes.[37] In view of these clinical challenges, we have also evaluated the feasibility of applying HW Raman spectroscopy for discrimination among normal, CIN1, and CIN2/CIN3 in the cervix. Our further PCA−LDA analysis together with leave-one-patient-out cross-validation on the limited HW Raman data set shows that the diagnostic accuracies of 90.0% (9/10) for CIN1, 97.2% (35/36) for CIN2/CIN3, 80.0% (8/10) for CIN1, and 97.8% (45/46) for the normal tissue can be achieved. The high degree of diagnostic accuracy indicates that the CIN2/CIN3 cases are biochemically distinguishable from CIN1 and also that CIN1 is identifiable from the normal cases by using in vivo HW Raman spectroscopy technique. Therefore, HW Raman spectroscopy (2800−3700 cm^{-1}) together with PCA−LDA can be used to yield high diagnostic accuracy for in vivo detection and staging of cervical precancer. One notes that PCA is primarily for data reduction rather than for identification of biochemical or biomolecular components of tissue. It is usually difficult to interpret the physical meanings of the component spectra. More powerful diagnostic algorithms (e.g., genetic algorithms, classification and regression tree (CART) algorithms)[16,17,48] that take into consideration the interactions among different spectral features could also be applied to choose the best subset of Raman features for further improving tissue classification. Distinctive spectral regions that are optimal for tissue differentiation may be identified and related to particular biochemical and biomolecular changes (e.g., proteins, lipids, nucleic acid, carbohydrates) associated with neoplastic transformation.[17,22] These techniques need a much larger cervical tissue Raman data set for robust diagnostic algorithms development. Moreover, to further understand the relationship between the dysplasia-related morphologic/biochemical changes and tissue Raman spectra that is crucial to establishing confidence in clinicians with the application of rapid fiber-optic HW Raman spectroscopy for cervical precancer detection, confocal Raman microspectroscopy should be explored for cervical tissue in vivo or in vitro by measuring the complete Raman spectra of specific tissue microstructures or by mapping the distribution of some specific Raman peaks or principal components within the tissue or, alternatively, by mapping the biochemical distribution at different tissue depths in association with cervical tissue histopathology.[49,50] Work in these areas warrants further investigation.

(48) Mountford, C. E.; Somorjai, R. L.; Malycha, P.; Gluch, L.; Lean, C.; Russell, P.; Barraclough, B.; Gillett, D.; Himmelreich, U.; Dolenko, B.; Nikulin, A. E.; Smith, I. C. *Br. J. Surg.* **2001**, *88*, 1234–1240.
(49) Caspers, P. J.; Lucassen, G. W.; Puppels, G. J. *Biophys. J.* **2003**, *85*, 572–580.

In conclusion, in vivo HW Raman spectra in the range 2800−3700 cm^{-1} from normal and precancer cervical tissue can be acquired in 1 s using the rapid Raman system integrated with a ball-lens fiber-optic Raman probe developed. Distinctive Raman spectral differences are observed for the first time between normal and dysplastic cervical tissue in vivo in the spectral range of 2800−3700 cm^{-1}, which contains water, lipids, and collagen signals. Good differentiation between normal and different stages of dysplastic cervical tissue can be achieved by using HW Raman spectroscopy and PCA−LDA techniques. Currently, HW Raman measurements on a larger series of patients are in progress at NUH to further validate the clinical merits of HW Raman spectroscopy for the early diagnosis and detection of cervical precancer and cancer in vivo during clinical colposcopic examinations.

(50) Shetty, G.; Kendall, C.; Shepherd, N.; Stone, N.; Barr, H. *Br. J. Cancer* **2006**, *94*, 1460–1464.

ACKNOWLEDGMENT

This work was supported by the Biomedical Research Council, the National Medical Research Council, the Academic Research Fund from the Ministry of Education, and the Faculty Research Fund from the National University of Singapore.

Received for review July 8, 2009. Accepted October 2, 2009.

AC9015159