RAMAN SPECTROSCOPY FOR *IN VIVO,* NON-INVASIVE DETECTION OF DYSPLASIA OF THE CERVIX

By

Amy Robichaux Viehoever

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements for
the degree of
DOCTOR OF PHILOSOPHY
in
Biomedical Engineering
May, 2004
Nashville, Tennessee

Approved:
Professor Dean Billheimer
Professor Howard Jones III
Professor Anita Mahadevan-Jansen
Professor Lynn Matrisian
Professor Robert Roselli

ACKNOWLEDGEMENTS

I would first like to thank my advisor on this project Anita Mahadevan-Jansen. She took me on as a graduate student despite my lack of any knowledge of biomedical optics. She has taught me not only about the field of biomedical optics, but also how to do research and write scientifically. I especially appreciate her continuing support for me as a student. She always expressed concern regarding my academic development both as an engineer as a physician, and her efforts to help me graduate in a timely manner will not be forgotten. She often has put my individual needs above the needs of the research project which I greatly appreciate. Lastly, as a wife and an aspiring mother and physician-scientist, I greatly value her example to me as a full time mother, wife and researcher; her example gives me hope that one day I may be able balance the three as well.

I would also like to thank the members of my thesis committee: Robert Roselli, Dean Billheimer, Howard Jones, and Lynn Matrisian. They have been extremely supportive and helpful from before the proposal until the last minute emails before the thesis was due. Their suggestions have changed this project for the better. I especially appreciate the fact that their help has been nothing but constructive. I have never felt that anyone was not behind me 100%.

I will forever be in debt to the members of the optics group who suffered through my ignorance of all things optical and mechanical and helped me gain the knowledge necessary to complete my project. Duco Jansen has provided tremendous insight and ideas for my project; Chad Lieber showed me the in and outs of Raman spectroscopy and wrote many of the Matlab scripts which have made my life easier; Wei-Chiang Lin has not only provided information but also resources for many of the details of my project; Josh Beckham and Shannon Faley have been a tremendous help during my cell culture work. Last but not least, the honorary members of the lab, Kiana and Arvin Jansen and Andrew Faley, have provided much comical relief – their smiling faces always brightened my day.

I also especially appreciate the help and resources provided by Drs. Todd Giorgio and Rick Haselton who have graciously allowed to use their cell culture facilities. The advice and help from the members of the Intracellular Engineering group has also made my life much easier.

The nurses on 4S deserve a special thanks for their willingness to go out of their way to help “the girl with the cart”. Their good natured complaints of “Oh no, Amy’s here!” made me

feel especially loved. I also would like to thank the Gynecology residents who had to take time out of their busy schedules to help me with my study. Beth Huff provided a tremendous amount of help and information to me, especially early in the study. I could have not acquired my data without the help and guidance of the gynecology attending physicians, especially Drs. Lynn Parker, John Currie, Marta Crispens, and Deborah Webster-Claire.

A special thanks goes to Marcella Woods, Mark Mackanos, and Shannon Faley for always being there to chat or listen when a vent was necessary. The comraderieship they have provided has made my time as a graduate student much more enjoyable. I also appreciate their willingness to listen to my endless chatter regarding my obsession with all things wedding. Dr. Mark Bray also has been a tremendous support, both as a friend in Christ and also an sounding board for all things graduate school related.

I would also like to thank the Medical Scientist Training Program for making my education possible. David Robertson and Terry Dermody have supported me when I wanted to take the non-traditional route. Ellen Carter, Jean Briggs, and Michelle Grundy have put up with me and helped smooth the bumps in the road that come with taking the path not traveled.

Robin Ryther deserves a special mention here. My colleague and partner in crime in medical school, she supported and encouraged me to continue to pursue the M.D./Ph.D. even when I was doubting myself. I look forward to working with her again when we return to the wards.

Of course, I would not be here without the love and support of my parents. Even though they dreaded the idea of their little girl moving far away and never coming back, they have supported me and my goals without fail. I owe so much to them, more than I can ever thank or repay them for. My Aunt Jacki and Uncle Ken have also been a source of tremendous support. They have pushed me to pursue my goals and dreams even when no one else would; they can relate to me in ways no one else can, and I know that they will always be there chats about life in academia.

Never did I dream that during this process I would find and marry the man of my dreams. The support of my wonderful husband has helped me tremendously during this stressful period in my life. The thesis would have not come to fruition without his encouragement during these last few months and technical support in writing this thesis. I am a better person because of you, Sterling, thank you.

Lastly, my faith in God has pulled me through the times of self-doubt and questioning. I thank and praise my LORD my God with all of my heart, soul, and strength.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES AND CHARTS

CHAPTER I

INTRODUCTION

Amy Robichaux Viehoever[1]

[1]Department of Biomedical Engineering
Vanderbilt University
Nashville, Tennessee 37235

1.1 Introduction and Motivation

Dysplastic lesions of the cervix, or Squamous Intraepithelial Lesions (SIL), are precursor lesions in the development of cervical cancer. If these precursor lesions can be detected early in the progression of disease, early treatment can prevent the progression to invasive cervical carcinoma. Although the widespread application of the Papanicolaou (Pap) smear as a screening tool has greatly decreased the incidence of cervical cancer [1], sampling and reading errors lead to high false positive and negative rates. The most recent meta-analysis of the accuracy showed that in low prevalence populations, the mean sensitivity and specificity of the Pap smear was 48% and 95%, respectively [2].

Raman spectroscopy has the potential to be a more accurate screening and diagnostic device. The sensitivity and specificity of Raman spectroscopy in differentiating dysplasia from all other tissue categories (including normal, squamous metaplasia and inflammation) *in vitro* was 82% and 92% [3], respectively. A pilot *in vivo* study found that using a simple algorithm, based on two Raman intensity ratios, Raman spectroscopy can separate high grade lesions from all others [6].

Since spectral measurements and analysis can be done in real-time (i.e. without pathologist intervention), Raman spectroscopy also has the potential to be a more cost effective means of screening and diagnosis. A study in *Obstetrics and Gynecology* in 1998 [4] concluded that managing cervical SILs with a "see and treat" strategy in which real-time diagnosis with spectroscopy was immediately followed by treatment with LEEP (Loop Electrocautery Excision Procedure) was more cost effective than the traditional colposcopy directed biopsy, followed by pathologic examination of the biopsy with treatment by LEEP at a later date.

This see and treat strategy also has implications for diagnosing and treating high-risk populations. Patients who do not have access to regular medical care, such as low-income populations or patients in developing countries, often do not receive proper screening and/or follow-up treatment. Lifestyle changes in these populations have led to an increased transmission of Human Papillomavirus (HPV) and a consequent rise in squamous cell carcinoma, for which the contraction of HPV is a known risk factor [5]. Additionally, in many Latin American countries, the incidence of cervical cancer has remained constant despite large screening efforts due to a lack of infrastructure for quality cytologic diagnosis and the absence of follow-up programs for treatment [5]. A see and treat method would reduce these problems by allowing

screening and certain treatments to be conducted simultaneously, therefore moderating the necessity of follow-up patient visits. The see and treat system would also save time and expenses for those patients needing to travel long distances for medical care.

In this thesis, the potential of using Raman spectroscopy for the detection of cervical dysplasia is examined with the long term goal of using Raman spectroscopy as part of a see and treat strategy for the screening of cervical cancer.

1.2 Specific Aims

To evaluate the potential of Raman spectroscopy to distinguish different pathologies in the cervix, the following specific aims were developed:

1.) Characterize the Raman spectral differences of different cervical tissue types *in vivo* for differential diagnosis of cervical lesions.

Raman spectra of the cervix were measured *in vivo* prior to the removal of the tissue during excisional procedures of the cervix. These spectra were examined to determine the spectral features unique to each class of cervix pathology.

2.) Characterize the sources of variation with a given pathology category.

A second clinical study was designed to measure Raman spectra from areas of normal cervix in patients without evidence of cervical dysplasia. Multiple spectra of normal appearing areas were measured in each patient prior to hysterectomy procedures allowing for determination of the optimal measurement time, effects from acetic acid, and components of variance.

3.) Develop algorithms based on multivariate statistics to automatically distinguish the spectra of normal cervix, inflammation, metaplasia, low grade SIL and high grade SIL.

The Raman spectra were grouped into categories according to the histologic diagnosis. A model using two logistic regression algorithms was developed to separate high grade dysplasia from benign areas of the cervix.

4.) Determine the biochemical and cellular basis for the differences seen in spectra of different tissue types.

To better understand the biochemical and cellular changes seen in specific aim1, organotypic raft tissue cultures that mimic the biochemical and cellular composition of the normal and dysplastic cervix were constructed, and Raman spectra of the intact tissue cultures and separate tissue

layers were measured individually. An understanding of the cellular contributions to the Raman spectra enhanced the characterization of the spectral differences between tissue categories.

1.3 References

1. Myers, E.R., *et al.*, Setting the target for a better cervical screening test: characteristics of a cost-effective test for cervical neoplasia screening. *Obstet Gynecol*, 2000. **96**(5 Pt 1): p. 645-52.

2. Nanda, K., *et al.*, Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*, 2000. **132**(10): p. 810-9.

3. Mahadevan-Jansen, A., *et al.*, Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers. *Photochem Photobiol*, 1998. **68**(1): p. 123-32.

4. Cantor, S.B., *et al.*, Cost-effectiveness analysis of diagnosis and management of cervical squamous intraepithelial lesions. *Obstet Gynecol*, 1998. **91**(2): p. 270-7.

5. Vizcaino, A.P., *et al.*, International trends in incidence of cervical cancer: II. Squamous-cell carcinoma. *Int J Cancer*, 2000. **86**(3): p. 429-35.

6. Utzinger, U., *et al.*, Near-Infrared Raman Spectroscopy for *In vivo* Detection of Cervical Precancers. *Applied Spectroscopy*, 2001. **55**(8): p. 955-9.

CHAPTER II

BACKGROUND

Amy Robichaux Viehoever[1] and Anita Mahadevan-Jansen[1]

[1]Department of Biomedical Engineering
Vanderbilt University
Nashville, Tennessee 37235

2.1 Anatomy and Histology of the Cervix

The Cervix is most inferior portion of the uterus and measures 2.5-3 cm in length in the nulliparous adult woman [1]. A small orifice in the center of the cervix, termed the external os, serves as the connection and passageway between the uterine corpus and vagina.

The three zones of the cervix are defined by the type of epithelium: the ectocervix on the periphery of the cervix with stratified squamous epithelium, the endocervix lining the external os with columnar epithelium, and the transformation zone which marks the border between the columnar and stratified squamous epithelial zones. The transformation zone is significant because most cervical cancers arise from this area. Underlying the epithelium in all three zones is the stromal layer primarily consisting of collagen devoid of any endometrial glands, characteristic of the stroma of the uterine corpus. The stromal layer also contains fibroblasts, inflammatory cells, and capillary beds.

The stratified squamous epithelium is 15 to 20 cells thick with a thickness of 200-400 microns. The maturation process begins in the layer of basal cells. Basal cells are elliptical cells with large oval nuclei and high mitotic activity. As they mature, these cells migrate to the surface, accumulate glycogen in the cytoplasm, and acquire a flattened shape. Eventually, the basal cells are shed in the desquamation process and the entire epithelium is replenished, which typically occurs over the course of 4-5 days [2].

2.2 Cervical Dysplasia

Dysplasia is defined as “disordered growth” and is generally thought of as a step in the development of invasive cancer [3]. Although not all dysplastic lesions will progress to invasive cancer, early detection and treatment of dysplastic lesions in the cervix is an important and effective method for the prevention of invasive cancer [4].

The majority of dysplastic lesions arise in the transformation zone, where the acidic environment of the vagina induces areas of columnar epithelium to transform into the more protective stratified squamous epithelium. This process is termed metaplasia. This metaplastic transformation provides an environment ripe for the initiation of dysplastic changes. Overwhelming evidence implicates HPV, a sexually transmitted virus infecting the squamous cells of the cervix, as the most important factor in the initiation of the pathogenesis of cervical cancer [5]. Figure 2.1 shows the theoretical process by which cervical cancers develop [3].

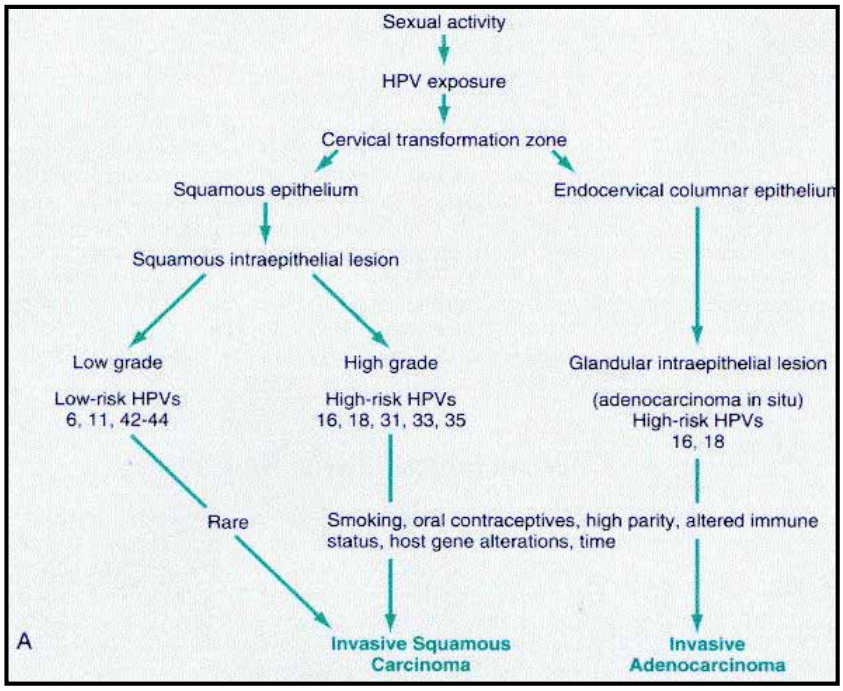

Figure 2.1: Diagram illustrating the stages of development of cervical cancer [3]

Different strains of HPV are known to confer different risks of developing invasive carcinoma. The strains conferring the greatest risk of inducing neoplastic changes ("high risk HPVs") are those that integrate their viral DNA into the host cell DNA. This integration disrupts the open reading frame of the viral gene E2, which is known to regulate expression of viral genes E6 and E7. Overexpression of E6 and E7 confers immortality to the host cell via inhibition of host cell tumor suppressor gene products p53 and Rb, respectively. Conversely, "low risk HPV" strains typically do not integrate into the host cell genome, and are consequently more likely to produce warts or hyperkeratosis than dysplastic changes[3].

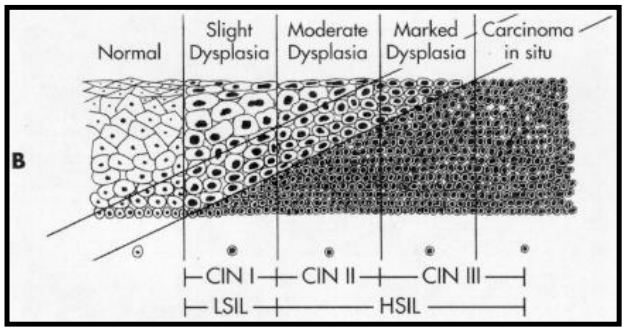

Figure 2.2: Progression from normal epithelium to carcinoma *in situ* (CIS).
Below the illustration, the corresponding classification is denoted. The top classification row (CIN) refers to the histological diagnosis; the bottom row refers to the cytological diagnosis as defined by the Bethesda system [2].

The stages of tissue progression from normal to dysplasia are shown in figure 2 with the corresponding diagnostic classifications. Cervical Intraepithelial Neoplasia (CIN) is a general term for dysplastic changes limited to the epithelium of the cervix. Dysplasia is characterized by a decrease in cytoplasm and glycogen levels, and a rise in mitotic activity. Similarly characteristic of dysplastic lesions are the regional increases in nuclear size and DNA quantity associated with the intensified mitotic activity. Morphologically, dysplastic cells resemble undifferentiated basal cells, but unlike normal basal cells, dysplastic cells exhibit large hyperchromatic nuclei, chromatin clumping, irregular nuclear borders and scant cytoplasm [3]. As the degree of atypia and number of dysplastic cells increase, the grade of the dysplasia is increased, with CIN III representing the highest grade of dysplasia including both severe dysplasia and carcinoma *in situ*. The Bethesda system was developed in 1988 as a method of grouping the pathological categories based on the course of treatment [6]. In this system, changes associated with HPV infection and CIN I are grouped together as low grade SIL whereas CIN II/III are grouped together as high grade SIL.

The development of cancer of the cervix is typically a slow process, taking several years for the transformation from normal epithelium to dysplasia to invasive carcinoma. This slow progression makes it possible for screening programs to detect neoplastic processes before invasive carcinoma develops. Early detection and treatment remains critical for successful eradication of the disease. Screening programs are designed to detect dysplasia in an effort to curb the incidence rate of carcinoma.

2.3 Current Methods of Screening/Diagnosis

2.3.1 Pap Smear

The Pap Smear (named after George Papanicolaou who first developed the technique) is a method of screening for cervical neoplasia by sampling exofoliated cells from the surface of the endocervical canal and ectocervix. A small spatula is used to first scrape cells from the ectocervix. Following this, the endocervical canal is sampled by inserting a cytological brush through the external os. The collected material is traditionally placed on a glass slide and fixed with a spray fixative. Since the widespread implementation of the Pap smear as a screening tool, the incidence of and mortality from cervical cancer has decreased dramatically in developed nations [7]. Yet, research into improving or replacing the Pap smear has increased due to concerns about its inefficiency and low sensitivity (30%-87%) [8, 9]. The low sensitivity is a product of sampling error, laboratory error, and interobserver error [10]. Sampling Error is estimated to contribute 66% of the overall false negative rate and typically occurs when abnormal cells are not acquired, collected cells are inadequately transferred to the slide, or cell clumping/artifacts on the slide obscure proper assessment of collected cells. The remaining 33% of the false negative rate can be attributed to a combination of laboratory error (errors in processing of the slides) and interobserver error (variation due to the subjective nature of screening Pap smear slides) [10, 11].

2.3.2 Colposcopy

Typically, if a Pap smear shows abnormal cells, a woman will be referred for a colposcopic examination in which the cervix is visualized by a magnifying colposcope (4-40X). The surface of the cervix is swabbed with a solution of 4% acetic acid which causes areas of abnormal epithelium to turn white. The acetowhitening of the epithelium slowly decays over several minutes but can be maintained by re-application of acetic acid. The degree of acetowhitening can be quantified (0, 1/2+, 1+, 2+), but such quantification is highly subjective. An iodine solution, which stains normal squamous epithelium but not normal columnar epithelium or abnormal epithelium, is often used to enhance the contrast of the transitional zone. Colposcopy, in expert hands, has been shown to be a very sensitive technique of identifying areas of abnormal epithelium with a sensitivity of 96%. Yet colposcopy often mistakes benign

changes of epithelium, such as squamous metaplasia and inflammation, for dysplasia, resulting in a low specificity (average specificity: 48%) [12]. Thus, to obtain a definitive diagnosis, the colposcopic diagnosis often must be confirmed with one or more biopsies. The need to confirm diagnosis with biopsy and histology increases patient pain and inconvenience, cost of disease management, and patients lost to follow-up as adherence with repeat colposcopy is 7-50% depending on patient population [13].

2.3.3 HPV testing

Studies have shown that women infected with HPV are more likely to develop low grade SIL than those without evidence of infection [5]. Infections with high risk types of HPV (16, 18) have higher rates of progression to high grade SIL than infection with low risk types (6, 11) [8]. Thus, HPV testing has been proposed as screening method to determine which women are at high risk for developing cervical neoplasia. The main benefit to using HPV testing as a screening tool lies in the extremely low false negative rate and high sensitivity (83%-100%) for detecting HSIL and carcinoma. Because not all women infected with HPV will have lesions that progress to SIL, the specificity of HPV testing hovers around 78%, and thus has not gained widespread acceptance as a primary screening tool [14-16].

HPV testing has been shown to be a more effective tool in triaging women when the results from a Pap Smear show Atypical Squamous Cells of Undetermined Significance (ASCUS) [17]. Since only a minority of these lesions will progress to higher grade lesions, HPV testing can aid in determining which women should be evaluated with colposcopy due to the increased likelihood that legions in women infected with HPV will progress to high grade SIL [8]. Since the vast majority (85%) of LSIL will test positive during HPV screening, HPV testing is no longer recommended as a method for triaging patients with a LSIL Pap smear for colposcopy [17].

2.3.4 A flawed gold standard

It should be noted that the calculations of accuracy for all of these screening and diagnostic tests are based on comparisons with histopathologic diagnosis on biopsy or other cervical tissue specimen, which is considered the gold standard for diagnosis of cervical dysplasia and invasive carcinoma. Yet, because histopathologic diagnoses are dependent on the

interpretation of one or more pathologists, the gold standard is not perfect and subject to variability. Analysis from the Atypical Squamous Cells of Undetermined Significance-Low Grade Squamous Intraepthelial Lesion (ALTS) Triage Study has shown that the histopathologic interpretations of punch biopsy (k=0.46) and LEEP (k= 0.49) (with k=1 representing exact agreement and k=0 representing poor agreement) can be extremely variable, even between pathologists at academic medical institutions [18]. Most of the variability in interpretation comes in distinguishing intermediate grades of pathology such as low grade dysplasia. Only 43% of low grade dysplasia interpretations were reproduced while 90% of normal and 76% of high grade dysplasia interpretations were corroborated [18]. Thus it should be noted that one of the inherent difficulties in assessing the accuracy of any new technology is comparison with a flawed gold standard.

2.4 New Technologies in Cervical Cancer Screening

Several new technologies have emerged in recent years aimed at improving the accuracy and accessibility of cervical cancer screening. Four of these have been implemented as possible alternatives to the current standard of treatment and diagnosis: liquid-based thin-layer cytology, computer-assisted cytologic screening and rescreening, visual inspection, and cervicography. Thin Prep® is a liquid-based cytologic preparation designed to reduce the number of errors due to sampling loss and preparation issues such as drying artifacts on traditional slides of cervical specimens. PAPNET® and AutoPAP® are computerized screening applications intended to reduce the number of false negative results attributable to reading errors. With visual inspection, a technique used in regions lacking the economic resources to fund more expensive alternatives, the cervix is examined with the unaided eye after the application of an acetic acid wash. Finally, cervicography is the use of high quality photographs of the cervix taken by trained technicians which then can be read in a manner similar to colposcopy, by an expert either locally or elsewhere. Of all of these, only Thin Prep® has been widely implemented [19].

With the exception of visual inspection, all of these new technologies add to the cost of cervical cancer screening, and none of them address the primary cause of diagnosis error in Pap smears, namely the failure to sample the abnormal cells from the cervix. In addition, all of the aforementioned technologies still require that a specimen (or picture) be sent off for evaluation, resulting in a lapse between the time a patient is evaluated and to the time a diagnosis can be

made. A technology that could eliminate this time lapse would enhance the quality of care and reduce the level of patient anxiety, as treatment decisions could be made at the time of initial evaluation. A real-time diagnosis would enhance the cost and time efficiency of the screening process by eliminating the additional time and cost of pathological diagnosis and reducing the number of visits to health care providers. Thus, methods of optical diagnosis, specifically optical spectroscopy, are considered here as a means for providing non-invasive, real-time diagnosis.

2.5 Use of Spectroscopy in Diagnosis of Cervical Lesions

Optical techniques, particularly optical spectroscopy, increasingly are being applied to detect pathology due to their ability to provide real-time, non-intrusive, and automated information.

2.5.1 Fluorescence Spectroscopy

Laser induced fluorescence spectroscopy is the most extensively investigated type of spectroscopy for the diagnosis of cervical dysplasia. In the most complete study to date, Richards-Kortum and co-workers measured 359 fluorescence spectra of cervical tissue *in vivo* from 95 patients prior to cervical biopsy using three (337, 380, and 460 nm) excitation wavelengths [20]. Classification was performed using a multivariate algorithm that employed principal components analysis for data reduction and an algorithm based on Bayes' theorem for discrimination. This multivariate technique was able to separate high grade dysplasia from all other cervical pathologies with a sensitivity of 79% and specificity of 78%. A separate (but similarly derived) algorithm separated all dysplastic lesions from all benign pathologies with a sensitivity of 82% and a specificity of 68%. Prospective classification of the data using a radial basis function (neural) network yielded analogous results [21].

Further analysis of the fluorescence data shows that the low accuracy is due primarily to the similarity of the spectra of squamous metaplasia and dysplasia leading to difficulties in distinguishing squamous metaplasia from dysplasia spectra, a problem also seen with colposcopy [20]. In attempt to improve the accuracy of fluorescence in separating metaplasia from dysplasia, there have been several studies investigating the use of fluorescence spectroscopy in combination with diffuse reflectance and/or light scattering spectroscopy. Diffuse reflectance spectroscopy measures the elastic scattering of light over a range of wavelengths, and using the light diffusion

theory, information regarding the optical properties of tissue can be extracted from the spectrum to provide information on the morphology and biochemistry of the stroma [22]. Light scattering spectroscopy measures the reflected light from only the superficial epithelial layer using polarized spectroscopy allowing for estimation of amount and sizes of epithelial cell nuclei [23]. Nordstrom *et.al*. [24] developed a system to scan the cervix, collecting a fluorescence and diffuse reflectance spectrum for 120 distinct areas. Although the system performed reasonably well when distinguishing both normal ectocervix from CIN I, achieving an 86% sensitivity and 87% specificity, and normal ectocervix from CINII/III, achieving a 91% sensitivity and 93% specificity, it was only able to distinguish squamous metaplasia from high grade dysplasia with a 77% sensitivity and 76% specificity. Georgakoudi *et. al.* added light scattering spectroscopy to fluorescence and diffuse reflectance spectroscopy (termed trimodal spectroscopy) to measure spectra in a study of 44 patients [22]. Using logistic regression and leave-one-out validation, trimodal spectroscopy was able to classify SIL versus non-SIL (benign) with a sensitivity of 92% and specificity of 71%. Development of trimodal spectroscopy for use as a fully functional, real-time imaging system is ongoing [25].

2.5.2 Near Infrared Absorption Spectroscopy

Although near-infrared spectroscopy (NIR) has been investigated for its potential uses in *in vitro* diagnosis, NIR has had limited *in vivo* applications due to the inability to use the same fiber optic probe for incident and collection light. Tromberg and co-workers [26] employed a technique termed frequency domain photon migration to overcome this limitation. In their study of ten patients being treated for high grade dysplasia with photodynamic therapy, absorption and scattering coefficients at 956 nm were shown to be significantly different between high grade dysplasia and normal cervix, yet no data was collected on low grade dysplasia, squamous metaplasia, or inflammation, three pathologies important for diagnosis of the cervix.

2.5.3 Confocal Imaging

Instead of analyzing reflected light spectrally, the reflected light can be collected to form an image which can be visually inspected and analyzed to give diagnostic information. In confocal imaging, light is collected from specific depths within the tissue, producing depth resolved images similar to histology sections of tissue. In an *in vitro* study, Collier *et. al.*

collected confocal images from biopsy specimens of colposcopically normal and abnormal areas of cervix from 19 patients. The nuclear-cytoplasmic ratio, calculated using image processing software, was able to distinguish CIN II/III from all others with a sensitivity and specificity of 100% and 91% respectively [27]. A second study by Collier *et. al.* found that application of acetic acid prior to collection of confocal images enhanced the nuclear contrast in the image in both normal and abnormal biopsy specimens; the effect served to increase the contrast between normal and dysplastic images [28]. Although the results of these studies are promising, the images were collected using a system designed in house to be a benchtop system. In order to evaluate the potential of this technology *in vivo*, a portable system, capable of taking measurements in a clinical setting, must be designed and implemented.

2.5.4 Cost Effectiveness of Spectroscopy

In an analysis of the cost effectiveness of different strategies for diagnosis and treatment, Cantor *et. al.* [29] found that the most cost effective strategy for management of abnormal Pap smears was a "see and treat" strategy incorporating fluorescence spectroscopy. Colposcopically directed biopsy, the current standard of care, was the most expensive but also the most accurate. Thus, if a method of spectroscopy could be developed, with the accuracy of colposcopically-directed biopsy, its cost effectiveness could make spectroscopy a viable alternative to the current standard of care.

2.6 Raman Spectroscopy for Diagnosis of Cervical Lesions

2.6.1 What is Raman Spectroscopy?

Raman spectroscopy is a molecular specific spectroscopy in which individual bonds are probed. When a photon is incident on a molecule, the light can be absorbed or scattered. Fluorescence and near infrared spectroscopy rely on the absorption of light (and subsequent emission of energy in the case of fluorescence). The Raman effect results from inelastic scattering. During scattering, the photon can reflect off the incident molecule with or without a transfer of energy. Elastic scattering denotes scattering which occurs without a transfer of energy such that the reflected photon leaves the interaction with the same amount of energy (and thus at the same wavelength) as the incident photon. Inelastic scattering occurs when energy is

transferred between the incident photon and the molecule causing a change in the vibrational energy of the bonds within the molecule. The energy transfer causes a shift in the wavelength of the scattered photons. Raman spectroscopy is a measure of this inelastic scattering in which the intensity of the scattering is plotted against the frequency corresponding to the energy lost or gained by the photon as it interacts with the substance [30]. Thus, this technique provides specific molecular information which results in the detection of subtle changes in molecular conformation and structure.

Raman spectroscopy traditionally has been used to study the structure and function of various molecules by analytical chemists and biochemists. More recently, researchers have begun to investigate the potential of Raman spectroscopy as a method of performing non-invasive diagnosis of diseases such as cancer and atherosclerosis [31]. Thus far, *in vitro* studies of the gastrointestinal tract [32], larynx [33], breast [34], skin [35], cervix [36], lung [37], lymph nodes [38] prostate [39] brain [40] and *in vivo* studies of cervix [54], esophagus [41], skin [42], colon [43] and rat carcinogenesis models [44] have shown that Raman spectroscopy is a promising diagnostic tool for the discrimination of dysplastic and cancerous lesions.

2.6.2 Discrimination Algorithms

In order to use spectral information to provide real time diagnosis, a mathematical algorithm or series of algorithms must be developed to extract clinically relevant information from the spectra and calculate a score or set of scores to classify the given spectrum into a category based on a predetermined set of criteria. The most robust method for developing and validating discrimination algorithms is to divide a data set into independent training and test sets. An algorithm or series of algorithms is fit to the data in the training set using the empirical or statistical method of choice, and the criteria for classification into specific categories is determined. Classification of the spectra in the test set determines the unbiased accuracy of the algorithm. In cases of small data sets (as is often the case in pilot studies), division of the data into training and test sets is not feasible.

The leave-one-out cross-validation method is a popular alternative to independent training and test sets. In the leave-one-out method, one spectrum is removed from the data set and the algorithm is derived using the remaining spectra. The algorithm is then tested using the removed spectrum. This process is repeated for every spectrum in the data set, such that an

estimate of the potential accuracy of future algorithms developed using the method in question can be calculated. In reality, the leave-one-out validation cannot be considered an unbiased estimate of accuracy. Often, it is difficult to leave out the spectrum throughout the entire algorithm development process, and thus the estimation of algorithms accuracy is inherently biased. Also, this technique is not effective for situations in which some spectra are related to one another, as is the case for multiple spectra attained from a single patient.

There have been a number of different empirical and multivariate statistical methods used to develop algorithms to discriminate between Raman spectra taken from cancerous and dysplastic tissue. One empirical method which has shown the most promise for separating spectra in different pathological groups is the calculation of peak ratios. Previous studies have used multiple peak ratios to effectively separate spectra of cervical dysplasia and normal cervix [36, 54]. Yet, empirical analysis techniques are limited in scope and subjective in nature. Spectral differences between the various categories of pathology may be subtle and not immediately recognizable, even to the trained eye. Thus, the need exists to develop tools which can automate the classification process such that the classification is both objective and systematic. Several different multivariate techniques are effective for classification in spectral diagnosis, including: neural networks, [45, 46], hierarchical cluster analysis [47], and linear discrimination analysis (LDA) based on Bayes' theorem [36].

Linear discrimination algorithms, such as Fisher's Discriminate Algorithm (FDA) or Quadratic Discrimination Score (QDS) are often used to group or classify measurements into two or more categories [48]. Yet, such algorithms are computationally complex and do not perform well with large data matrices such as those seen in spectroscopy. Also, the number of inputs to a discrimination model must be lower than the number of spectra in the training set to avoid overfitting the model which would decrease the likelihood that the model would perform well on future data sets. Thus, one or more data reduction techniques must be employed before discrimination algorithms can be used for classification.

The most basic data reduction technique is empirical data reduction. An examination of the data can yield a subjective assessment to determine which aspects of the spectrum are likely to be most useful for discrimination. If the biological basis of spectral differences seen between pathology categories is known, those spectral features with biological relevance can be empirically chosen as inputs for the model rather than choosing spectral features which are only

statistically different in the current data set. This strategy would increase the likelihood that the model inputs will perform equally well with future data sets. An objective method of identifying the variance within a data set is Principal Components Analysis (PCA). PCA is a method that uses eigenvector decomposition of the covariance matrix to map the data onto a new set of orthogonal axes such that the variance within the data is maximized and concentrated into a few components. Each principal component describes a decreasing percentage of the total variance (as described by the eigenvalues) [48]. This allows the majority of the variance within a data set to be represented by a minimum number of variables. Thus, data reduction with PCA allows for a minimum number of variables necessary for evaluation in discrimination algorithms without compromising the data variance.

The use of PCA for data reduction, followed by linear discrimination algorithms (LDA) for classification, has become a well established method for spectroscopic data analysis with a number of research groups employing the method specifically for discrimination of Raman spectra [41, 43, 44, 49]. There are some potential flaws to this system, including the restriction of discrimination to a linear surface and the potential to overfit the model to the data. One such alternative to the use of PCA, which has been used for classification in fluorescence, trimodal, near infrared absorption, and Raman spectroscopy, is logistic regression [22, 28, 50]. Logistic regression is a variation of linear regression in which there are only two outcomes. The results of the linear regression are fit to a sigmoidal curve yielding a function that estimates the probability of an event happening based on the data input [51]. When used for spectral classification, the logistic regression algorithm will yield a score or probability that a given spectrum was measured from a high grade lesion (or category of choice).

2.6.3 In vivo Raman spectroscopy for the detection of dysplasia and cancer

Although there has been extensive work on the characterization of Raman spectra for various tissue types *in vitro*, very few studies have investigated the diagnostic potential of Raman spectroscopy *in vivo*. Bakker Schut ***et. al.*** used a rat carcinogenesis model to map the changes in Raman spectra which occur during neoplastic transformation in the oral cavity [44].Upon visual inspection of the data, increases were seen in the peak ~1006 cm^{-1} and region from 1300-1330 cm^{-1} when comparing spectra from normal tissue to dysplastic tissue. A model using PCA for data reduction and LDA for discrimination with leave-one-out validation, was

able to detect low grade dysplasia with a sensitivity of 78% and specificity of 93% and high grade dysplasia and cancer with 100% sensitivity and specificity. This study was the first to show the ability of Raman spectroscopy to detect dysplasia *in vivo* in a controlled model system. The results indicated that in a controlled model system devoid of significant intersubject variation, Raman spectroscopy is capable of detecting dysplastic changes with a high level of accuracy when compared to current methods.

Shim *et. al.* measured Raman spectra *in vivo* from 20 patients during upper or lower GI endoscopy [41]. Spectra were collected using a 5 second measurement time with signal to noise ratios ranging from 3 in the esophagus to 17 in the colon. Subtle changes in the spectra were noted between 1100-1800 cm^{-1} when comparing normal and pathologic states, but no attempt at spectral classification was made due to the small number of patients. Yet, this study represents a significant milestone as it is the first to record *in vivo* clinical Raman spectra measured with a clinically relevant signal collection time (5 seconds) and sufficient signal to noise ratio. In a subsequent study, Molckovsky *et. al.* measured Raman spectra from 19 adenotamous and hyperplastic polyps *in vivo* from 3 patients. Data reduction using PCA and classification using LDA with leave-one-out cross validation was used to develop a diagnostic algorithm which classified the data with 100% sensitivity and 89% specificity [43]. The authors also note in this study that an algorithm developed from an *ex vivo* study of colon polyps was not able to classify the *in vivo* data accurately, indicating that although *in vitro* and *ex vivo* studies may be able to predict whether Raman spectroscopy will succeed *in vivo*, diagnostic algorithms developed using data from *ex vivo* and *in vitro* data may not be translated for *in vivo* data.

Notably, several studies from Puppels and colleagues have demonstrated the measurement of Raman spectra from skin *in vivo* using a confocal Raman microspectrometer [35, 42, 52]. However, since none of these studies have looked cancerous or dysplastic changes of the skin they will not be detailed here.

2.6.4 Previous Studies using Raman Spectroscopy for Diagnosis in the Cervix

In the first study investigating the potential of using Raman spectroscopy for diagnosis in the cervix, Mahadevan-Jansen *et. al.*[36] measured Raman spectra *in vitro* of 36 biopsies from18 patients. Empirical peak intensities were found that differentiated dysplasia from benign pathologies with an average sensitivity and specificity of 88 % and 92% respectively. An

unbiased multivariate analysis using principal components analysis and linear discriminate algorithms separated dysplasia from all benign pathologies with a sensitivity and specificity of 82% and 92% respectively.

This *in vitro* study was followed by a 13 patient study investigating the feasibility of using Raman spectroscopy for *in vivo* diagnosis [54]. A specialized fiber optic probe was designed to maximize throughput of Raman signal collection and minimize collection of diffuse reflectance and Raman peaks from the probe itself [53]. Measurements were collected prior to cervical biopsy, and a total of 24 measurements were suitable for analysis. A simple ratio of peak intensities was able to separate high grade dysplasia from all other pathology types misclassifying only one sample.

These two proof of principle studies show that Raman spectroscopy can be measured *in vivo* and has the potential to be a more accurate method of spectroscopic diagnosis than fluorescence spectroscopy, while retaining the cost-effectiveness and real-time diagnosis capability shown with fluorescence spectroscopy [29]. Yet, these studies only included small numbers of patients and used empirical algorithms based on subjective analysis of the data. Thus, although this study shows that Raman spectroscopy has the potential to be a useful tool for the detection of cervical dysplasia, further studies are needed to fully evaluate this potential.

2.7 References

1. Sheets, E., The Cervix, in Ryan: Kistner's *Gynecology & Women's Health*. 1999, Mosby Inc. p. 93-119.

2. Ferenczy, A. and B. Winkler, *Anatomy and Histology of the Cervix*, in *Blaustein's Pathology of the Female Genital Tract*, R. Kurman, Editor. 1994, Springer-Verlag: New York.

3. Cotran R, K.V., Collins T., *Robbins Pathologic Basis of Disease*. 6th ed. 1999, Philadelphia: W.B. Saunders Company.

4. *NIH Consensus Statement*. 1996, National Institutes of Health.

5. Walboomers, J.M., *et al.*, Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*, 1999. **189**(1): p. 12-9.

6. Kurman, R.J., *et al.*, From Papanicolaou to Bethesda: the rationale for a new cervical cytologic classification. *Obstet Gynecol*, 1991. **77**(5): p. 779-82.

7. Myers, E.R., *et al.*, Setting the target for a better cervical screening test: characteristics of a cost-effective test for cervical neoplasia screening. *Obstet Gynecol*, 2000. **96**(5 Pt 1): p. 645-52.

8. Spitzer, M., Cervical screening adjuncts: recent advances. *Am J Obstet Gynecol*, 1998. **179**(2): p. 544-56.

9. Nanda, K., *et al.*, Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*, 2000. **132**(10): p. 810-9.

10. Baldwin, P., R. Laskey, and N. Coleman, Translational approaches to improving cervical screening. *Nat Rev Cancer*, 2003. **3**(3): p. 217-26.

11. Cohn, D.E. and T.J. Herzog, *New* innovations in cervical cancer screening. *Clin Obstet Gynecol*, 2001. **44**(3): p. 538-49.

12. Mitchell, M.F., *et al.*, Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol*, 1998. **91**(4): p. 626-31.

13. Basen-Engquist, K., *et al*., *Cervical cancer.* Cancer, 2003. **98**(9 Suppl): p. 2009-14.

14. Snijders, P.J., A.J. van den Brule, and C.J. Meijer, The clinical relevance of human papillomavirus testing: relationship between analytical and clinical sensitivity. *J Pathol*, 2003. **201**(1): p. 1-6.

15. Clavel, C., *et al.*, Human papillomavirus testing in primary screening for the detection of high-grade cervical lesions: a study of 7932 women. *Br J Cancer*, 2001. **84**(12): p. 1616-23.

16. de Cremoux, P., *et al.*, Efficiency of the hybrid capture 2 HPV DNA test in cervical cancer screening. A study by the French Society of Clinical Cytology. *Am J Clin Pathol*, 2003. **120**(4): p. 492-9.

17. Solomon, D., M. Schiffman, and R. Tarone, Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J Natl Cancer Inst*, 2001. **93**(4): p. 293-9.

18. Stoler, M.H. and M. Schiffman, Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*, 2001. **285**(11): p. 1500-5.

19. Soler, M.E. and P.D. Blumenthal, New technologies in cervical cancer precursor detection. *Curr Opin Oncol*, 2000. **12**(5): p. 460-5.

20. Ramanujam N, M.M., Mahadevan-Jansen A, Thomasen S, Staerkel G, Malpica A, Wright T, Atkinson A, Richards-Kortum R, Cervical pre-cancer detection using a multivariate statistical algorithm based on laser induced fluorescence spectra at multiple excitation wavelengths. *Photochem Photobiol*., 1996. **64**: p. 720-735.

21. Tumer, K., *et al.*, Ensenbles of radial basis function networks for spectroscopic detection of cervical precancer. *IEEE Trans Biomed Eng*, 1998. **45**(8): p. 953-61.

22. Georgakoudi, I., *et al.*, Trimodal spectroscopy for the detection and characterization of cervical precancers *in vivo*. *Am J Obstet Gynecol*, 2002. **186**(3): p. 374-82.

23. Gurjar, R.S., *et al.*, Imaging human epithelial properties with polarized light-scattering spectroscopy. *Nat Med*, 2001. **7**(11): p. 1245-8.

24. Nordstrom RJ, B.L., Niloff JM and Myrtle JF, Identification of Cervical Intraepithelial Neoplasia (CIN) Using UV-Excited Fluorescence and Diffuse-Reflectance Tissue Spectroscopy. *Lasers in Surgery and Medicine*, 2001. **29**: p. 118-27.

25. Drezek, R.A., *et al*., Optical imaging of the cervix. *Cancer*, 2003. **98**(9 Suppl): p. 2015-27.

26. Hornung, R., *et al*., Quantitative near-infrared spectroscopy of cervical dysplasia *in vivo*. *Hum Reprod*, 1999. **14**(11): p. 2908-16.

27. Collier, T., *et al*., Near real-time confocal microscopy of amelanotic tissue: detection of dysplasia in ex vivo cervical tissue. *Acad Radiol*, 2002. **9**(5): p. 504-12.

28. Collier, T., *et al*., Near real time confocal microscopy of amelanotic tissue: dynamics of aceto-whitening enable nuclear segmentation. *Optics Express*, 2000. **6**: p. 40-48.

29. Cantor, S.B., *et al.*, Cost-effectiveness analysis of diagnosis and management of cervical squamous intraepithelial lesions. *Obstet Gynecol*, 1998. **91**(2): p. 270-7.

30. Ferraro, J. and K. Nakamoto, *Introductory Raman Spectroscopy*. 1994: Academic Press.

31. Hanlon, E.B., *et al*., *Prospects for in vivo Raman spectroscopy.* Phys Med Biol, 2000. **45**(2): p. R1-59.

32. Stone N, *et al*., Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 564-573.

33. Stone, N., *et al*., Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. *Laryngoscope*, 2000. **110**(10 Pt 1): p. 1756-63.

34. Shafer-Peltier, K., *et al*., Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 552-563.

35. Caspers, P.J., *et al*., *In vitro* and *in vivo* Raman spectroscopy of human skin. *Biospectroscopy*, 1998. **4**(5 Suppl): p. S31-9.

36. Mahadevan-Jansen, A., *et al*., Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers. *Photochem Photobiol*, 1998. **68**(1): p. 123-32.

37. Huang, Z., *et al*., Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int J Cancer*, 2003. **107**(6): p. 1047-52.

38. Smith, J., *et al*., Raman spectral mapping in the assessment of axillary lymph nodes in breast cancer. *Technol Cancer Res Treat*, 2003. **2**(4): p. 327-32.

39. Crow, P., *et al*., The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*. *Br J Cancer*, 2003. **89**(1): p. 106-8.

40. Koljenovic, S., *et al*., Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy. *Lab Invest*, 2002. **82**(10): p. 1265-77.

41. Shim, M.G., *et al*., *In vivo* near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy. *Photochem Photobiol*, 2000. **72**(1): p. 146-50.

42. Caspers, P.J., *et al*., *In vivo* confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. *J Invest Dermatol*, 2001. **116**(3): p. 434-42.

43. Molckovsky, A., *et al*., Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. *Gastrointest Endosc*, 2003. **57**(3): p. 396-402.

44. Bakker Schut, T.C., *et al*., *In vivo* detection of dysplastic tissue by Raman spectroscopy. *Anal Chem*, 2000. **72**(24): p. 6010-8.

45. Rovithakis, G.A., *et al*., Artificial neural networks for discriminating pathologic from normal peripheral vascular tissue. *IEEE Trans Biomed Eng*, 2001. **48**(10): p. 1088-97.

46. Goodacre, R., *et al*., Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology*, 1998. **144**(Pt 5): p. 1157-70.

47. Kirschner, C., *et al*., Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. *J Clin Microbiol*, 2001. **39**(5): p. 1763-70.

48. Wichern, R.J.a.D., *Applied Multivariate Statistical Analysis*. 4 ed. 1998, Englewood Cliffs, NJ: Prentice Hall.

49. Stone, N., *et al*., Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 564-573.

50. Wang, C.Y., *et al*., A probability-based multivariate statistical algorithm for autofluorescence spectroscopic identification of oral carcinogenesis. *Photochem Photobiol*, 1999. **69**(4): p. 471-7.

51. Neter J., *et al*., *Applied Linear Statistical Models*. 4th ed. 1996, Chicago, IL: Irwin.

52. Caspers, P.J., *et al*., Combined *in vivo* confocal Raman spectroscopy and confocal microscopy of human skin. *Biophys J*, 2003. **85**(1): p. 572-80.

53. Mahadevan-Jansen, A., *et al.,* Development of a fiber optic probe to measure NIR Raman spectra of cervical tissue *in vivo*. *Photochem Photobiol*, 1998. **68**(3): p. 427-31.

54. Utzinger, U., *et al*., Near-Infrared Raman Spectroscopy for *In vivo* Detection of Cervical Precancers. *Applied Spectroscopy*, 2001. **55**(8): p. 955-9.

CHAPTER III

CHARACTERIZATION OF RAMAN SPECTRA MEASURED *IN VIVO* FOR THE DETECTION OF CERVICAL DYSPLASIA

Amy Robichaux Viehoever[1], Heidi Shappell[2], Dean Billheimer[3], Howard Jones, III[4],and Anita Mahadevan-Jansen[1]

[1]Department of Biomedical Engineering,
[2]Department of Pathology,
[3]Department of Preventive Medicine,
[4]Department of Obstetrics and Gynecology
Vanderbilt University, Nashville, Tennessee 37235

This manuscript was prepared for submission in the *Journal of Biomedical Optics*.

3.1 Abstract

Raman Spectroscopy has been shown to have the potential for providing differential diagnosis in the cervix with high sensitivity and specificity in previous *in vitro* and *in vivo* studies. A clinical study was designed at Vanderbilt University Medical Center to further evaluate the potential of near infrared Raman spectroscopy to detect cervical dysplasia in a clinical setting.

Using a portable system, Raman spectra were collected from the cervix of 79 patients prior to excisional procedures using clinically feasible integration times (5 seconds). Multiple Raman spectral measurements were taken of colposcopically normal and abnormal areas prior to the excision of tissue. The Raman spectra were then extracted from the measured signals using signal processing, The resulting spectra were correlated with the corresponding histopathologic diagnosis to determine empirical differences between different diagnostic categories. Using histology as the gold standard, logistic regression discrimination algorithms were developed to distinguish between normal ectocervix, squamous metaplasia, and high grade dysplasia using independent training and validation sets of data. An unbiased estimate of the accuracy of the model indicates that Raman spectroscopy was able to distinguish between high grade dysplasia and benign areas of the cervix (normal ectocervix and squamous metaplasia) with sensitivity of 89% and specificity 81% while colposcopy in expert hands was able to discriminate with a sensitivity of 87% and specificity of 72%.

3.2 Introduction

If cervical dysplasia, a precursor lesion in the development of cervical cancer, can be detected early in the progression of disease, treatment can prevent the progression to invasive cervical carcinoma [1]. Although the widespread application of the Papanicolaou (Pap) smear as a screening tool has greatly decreased the incidence of cervical cancer, sampling and reading errors in the Pap smear lead to a high number of false positive results. The most recent meta-analysis of the accuracy showed that in low prevalence populations, the mean sensitivity and specificity of the Pap smear was 48% and 95%, respectively [2]. The standard diagnostic method for women suspected of having dysplasia is colposcopy directed biopsy, a procedure in which the physician applies a 3-5% acetic acid solution to the cervix and examines it under a magnifying (14X) colposcope. The application of acetic acid causes areas of abnormal

epithelium to turn "acetowhite", enabling easier identification of abnormal areas. Because benign changes such as squamous metaplasia and inflammation can mimic the appearance of dysplasia during colposcopy, suspicious areas must be biopsied to provide definitive diagnosis. This is a difficult, painful, and time consuming procedure often not capable of providing a complete diagnosis [3]. Hence, there is considerable interest in developing a non-invasive diagnostic tool, which can accurately identify dysplasia, follow suspicious areas over time, and facilitate complete excision of dysplasia and cancer by accurately delineating the margins prior to and during excisional surgery. If such a screening and diagnostic program could be developed, it would streamline the treatment of cervical dysplasia and thus decrease the overall cost in the management of the disease, while also reducing patient pain and inconvenience [4].

Non-invasive optical spectroscopy has been proposed as a potentially more accurate and cost-effective method of screening for cervical cancer. The majority of the researchers investigating the potential of spectroscopy to perform diagnosis in the cervix thus far has used autoflorescence spectroscopy either alone or in combination with other types of spectroscopy, including light scattering spectroscopy and diffuse reflectance [5-7]. The most extensive studies to date reported that fluorescence alone was able to distinguish dysplasia from non-dysplasia (benign areas of the cervix) with a sensitivity of 82% and specificity of 68% [8]. Trimodal spectroscopy, which included fluorescence, diffuse reflectance, and light scattering spectroscopy, improved the accuracy, separating dysplasia from benign with a sensitivity of 92% and specificity of 71% [5]. Research aimed at improving the accuracy of fluorescence spectroscopy is ongoing and focuses primarily on understanding the differences in spectral intensity due to menopausal status as well as the biochemical and cellular basis behind the differences seen between dysplastic and normal spectra [9-13]. Yet, it remains unclear if these studies will lead to better discrimination algorithms.

Raman spectroscopy is another type of optical spectroscopy which also has the potential to be a more accurate and cost-effective screening device than the current standard of care. Raman spectroscopy measures the inelastic scattering of a photon incident on a substance (the substance in this case is cervical tissue) [14]. Raman spectroscopy also has the potential to be more specific than fluorescence spectroscopy given that every molecular bond will scatter light with a unique shift in energy leading to a spectral fingerprint, with each peak representing a specific molecular bond. Conversely, there are only a small number of molecules within tissue

which will emit fluorescence upon interrogation with a specific wavelength of light. Raman spectroscopy has shown thus far to be a promising diagnostic tool for the discrimination of dysplastic and cancerous lesions in *in vitro* studies of the gastrointestinal tract [15], larynx [16], breast [17], skin [18], cervix [19], lung [20], lymph nodes [16], prostate [21], and brain [22] and *in vivo* studies of cervix [32] esophagus [23], colon [24] skin [25], and rat carcinogenesis models [26]

Utizinger *et. al.* conducted a pilot study to examine the potential of using Raman spectroscopy *in vivo* for the detection of cervical dysplasia [32]. In this pilot study, Raman spectra were measured *in vivo* from 25 patients with a diagnosis of an abnormal Pap smear using signal collection times of 60-120 seconds. Data from 13 of those patients was used for analysis. The results showed specific spectral changes associated with spectra from areas of dysplasia as compared with spectra from areas of normal cervix and benign changes of the cervix (inflammation, squamous metaplasia). A model using empirically derived peak ratios was able to distinguish high grade dysplasia from all others misclassifying only one spectrum.

Our current study was designed to follow-up on the results of the pilot study to further test the potential of using Raman spectroscopy to perform real-time diagnosis *in vivo* in the cervix. Some differences in the current study include: a larger study population (79 versus 25 in the pilot study), more spectra measured per patient (on average of 2-6 spectra versus only 2 spectra in the pilot study), utilization of a more clinically relevant measurement time (5 seconds versus 90 seconds in the original study), inclusion of a subset of "normal" subjects undergoing hysterectomy with no evidence of cervical disease, and utilization of an automated signal processing technique to produce yield more reproducible spectra.

The goals of this current study are the following: to characterize the Raman spectral features for five different pathologic categories in the cervix (normal ectocervix, normal endocervix, squamous metaplasia, low grade dysplasia, high grade dysplasia), develop a classification model using logistic regression algorithms to separate the spectra into benign or dysplasia categories, give an unbiased estimate of the accuracy of spectral classification and map the observed spectral differences to known the biochemical and cellular origins for specific Raman peaks.

3.3 Methods

3.3.1 Study Design

Raman spectra from 79 patients undergoing diagnostic or therapeutic procedures to remove tissue from the cervix were measured *in vivo* prior to excision of tissue as part of two different studies approved by the Vanderbilt Institutional Review Board (IRB). Informed consent was obtained from each patient prior to the study.

The first study, which focused on dysplasia patients, measured spectra from 44 patients with a history of an abnormal Pap smear or known abnormalities of the cervix from previous colposcopic evaluation. Any adult, non-pregnant patient undergoing a Loop Electrocautery Excision Procedure (LEEP), cervical biopsy (including conization), or radical hysterectomy at the Vanderbilt Colposcopy Clinic or Vanderbilt Ingram Cancer Center was eligible to participate. Colposcopic examination of the cervix was performed first to identify normal and abnormal areas of the cervix. Multiple (2-6) Raman spectra of colposcopically normal and abnormal appearing sites were measured *in vivo* after the application of 4% acetic acid, but prior to the excision of tissue. One background measurement was made with the laser off and the fiber optic probe in contact with the tissue to measure the ambient light present at the tissue. Following spectral measurements, the measured sites then were marked with a methylene blue paste. The tissue was then excised according to standard clinical protocol (or in the case of radical hysterectomy, the hysterectomy proceeded as normal) and the histological analysis was performed by the participating gynecological pathologist (HS).

The second study focused on patients with a normal cervix undergoing total abdominal or vaginal hysterectomy for indications other than disease of the cervix. Raman spectra were measured from a total of 35 patients. To be eligible for enrollment in the study, patients were required to be undergoing a hysterectomy for reasons other than cervical disease, be between the ages of 18-75, and have a history of previous normal Pap smear or normal exam of the cervix. The study was performed after the patient was placed under general anesthesia but prior to the start of the hysterectomy. Colposcopic examination of the cervix was performed first to identify all colposcopically normal and abnormal (if any) appearing areas of the cervix. Colposcopy was performed for two reasons: to ensure the measured areas were colposcopically normal and to maintain consistency with the study of dysplasia patients. Three areas of normal cervix were identified for measurement. If the squamocolumnar junction was visible, then columnar

epithelium (endocervix) was chosen as one of the sites to increase the overall numbers of endocervix spectra in the study. Multiple Raman spectra were measured from each chosen site *in vivo*. A background measurement was made, and the sites of measurement were marked as in the study of dysplasia patients. The hysterectomy then proceeded according to standard clinical protocol. After the hysterectomy the histological analysis of the cervical tissue specimen was performed by the participating gynecologist pathologist (HS).

Raman spectra were collected using a portable Raman spectroscopy system consisting of a 785 nm diode laser (Process Instruments, Inc., Salt Lake City, UT), fiber optic probe (Visionex Inc., Atlanta, GA), imaging spectrograph (Kaiser Optical Systems, Inc., Ann Arbor, MI), and back-illuminated, deep-depletion, liquid nitrogen cooled charge coupled device (CCD) camera

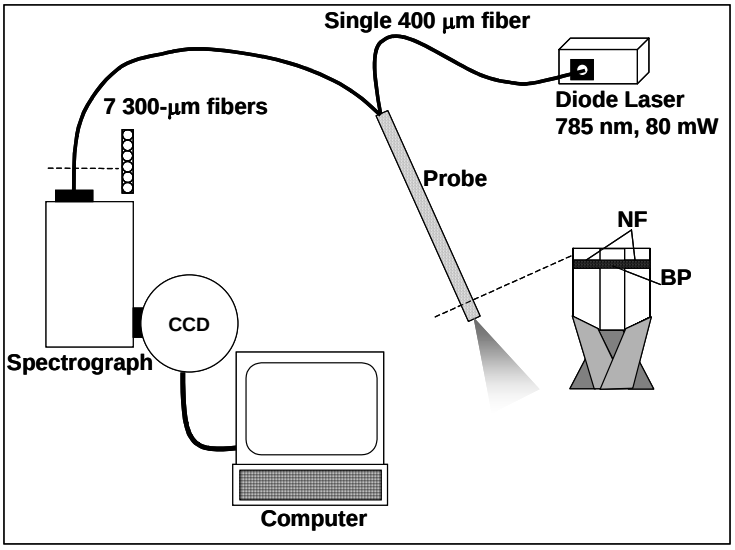

Figure 3.1: Diagram of Raman spectroscopy system used for data collection.
Note: CCD: Charged Coupled Device, NF: notch filterat 785 nm, BP: band pass filter at 785 nm

(Roper Scientific, Inc., Princeton, NJ) all controlled with a laptop computer. A diagram of the system is shown in figure 3.1. The beam-steered fiber optic probe delivers 785 nm incident light onto the tissue and collects the scattered light. The collected light is then filtered with an inline notch filter within the probe itself. The light is then fed into the spectrograph where it is filtered again with a holographic notch filter and dispersed onto the CCD camera where the computer records the signal. For this study, the fiber optic probe delivered 80mW 785 nm incident light

onto the tissue and collected the scattered light for 5-15 seconds. In all cases, the overhead fluorescent lights and colposcope light were turned off during the measurements. Any luminescent lights were left on but turned away from the measurement site.

3.3.2 Data Pre-Processing and Extraction

Prior to the spectral measurements in each patient, a spectral calibration of the system was performed using a neon-argon lamp and naphthalene standard to correct for system wavenumber, laser excitation, and throughput variations. For each Raman spectrum measured, the signal from the CCD camera was binned along the vertical (intensity) axis to create a single spectrum per measurement site. Prior to any signal processing, the spectrum was truncated to only include the region from 950 cm^{-1} to 1850 cm^{-1}. This eliminated the large Raman peaks below 950 cm^{-1} primarily due to the silica present in the fiber optic probe that obscure any tissue Raman peaks and the noise present at the very end of the spectral region. The spectrum was then binned along the horizontal (wavenumber) axis such that each data point represents the intensity from 3.5 cm^{-1}. The spectrum was then filtered using a 2^{nd} order Savitzky-Golay filter (window =17.5 cm^{-1}) for noise smoothing [27]. Fluorescence subtraction was accomplished using a modified polynomial fitting method in which a 5^{th} order polynomial is fit to the fluorescence baseline [28]. Following data pre-processing, each spectrum was normalized to its mean spectral intensity across all Raman bands to account for overall intensity variability. These normalized spectra were used for further comparison and analysis.

3.3.3 Data Analysis

The processed Raman spectra were grouped according to the corresponding histological diagnosis for analysis (see table 3.1).

Table 3.1: Description of the Pathology Categories Used to Group Spectra

PATHOLOGY CATEGORY	DESCRIPTION
Normal ectocervix	Squamous epithelium
Normal endocervix	Columnar epithelium
Squamous metaplasia	Benign change of cervical epithelium from columnar to squamous
Low grade dysplasia	Cervical Intraepithelial Neoplasia I (CIN I), HPV changes
High grade dysplasia	CIN II/III, carcinoma in situ (CIS), invasive squamous cell carcinoma

Data from 11 dysplasia patients were excluded from analysis for one of the following reasons: instrument malfunction, low signal to noise ratio, inability to obtain accurate histological diagnosis, or excessive room light that obscured the Raman spectra. Data from 2 normal patients were excluded due to low signal to noise or camera saturation. Thus, data from a total of 66 patients (33 dysplasia patients, 33 normal patients) were used in this analysis. If multiple spectra were collected from a single location within a given patient, only the first spectra collected was included for data analysis.

The mean and standard deviation of the spectra within each pathology group were calculated to characterize the overall spectral trends for each pathology group. A student's t-test was performed at each wavenumber between individual pairs of pathology groups to help identify regions of spectral distinction between two different pathologies.

The classification model was constructed to automatically classify spectra into one of two categories (high grade dysplasia or benign cervix) using two-tiered logistic regression model [29]. The process to develop the model is diagrammed in chart 3.1. Traditional linear regression fits an equation of a line to model a set of data. Yet linear regression does not work well when attempting to model data with discrete outcomes. Logistic regression overcomes this problem by using a non-linear transformation of the traditional linear regression such that the solution of the algorithm is restricted to values between 0 and 1 (eq. 3.1).

$$\ln\left(\frac{score}{1-score}\right) = a + b \times (inputs) \qquad (eq.\ 3.1)$$

where a = residual constant as determined by model fit

b = array of model coefficients

score = model score/probability for model inputs belonging to high grade dysplasia category

The logistic regression algorithms used in the model described here were trained using the glmfit function and tested using the glmval function in MATLAB® R12. In training of the algorithm, the algorithm coefficients (b) and residual constant (a) are derived by determining a best of the training set inputs and their corresponding assigned scores (assigned by the user). The first algorithm was trained to classify a spectrum as either normal ectocervix (score=0) or high grade dysplasia (score =1) and was developed using independent training and validation sets that were randomly generated by dividing the normal ectocervix and high grade dysplasia data sets into a training set (two-thirds of the patients) and a validation sets (one-third of the patients). Any major peak that showed statistical difference at the level of $p< 0.01$ between normal ectocervix spectra and high grade dysplasia spectra was chosen as an input for the algorithm. Thus, the inputs to the algorithm are the normalized intensity values at 1006, 1058, 1240 1305 1324, 1450, 1550, 1655 cm^{-1} (see figure 3.3) and the logistic regression equation is as shown in equation 3.2.

$$\ln\left(\frac{score}{1-score}\right) = a + b_1 * I_{1006} + b_2 * I_{1055} + b_3 * I_{1240} + b_4 * I_{1305} + b_5 * I_{1324} + b_6 * I_{1450} + b_7 * I_{1550} + b_8 * I_{1655} \qquad (eq\ 3.2)$$

The algorithm outputs a score which represents the probability that the input represents high grade dysplasia. Data from the other pathology categories (normal endocervix, squamous metaplasia, and low grade dysplasia) were also included as part of the test set for the model even though no data from these categories were included in the training set to examine the possibility that a single algorithm could discriminate all spectra of benign pathology from dysplasia spectra (see discussion). Since the specificity of this one algorithm model was less than desired due primarily to misclassifications of squamous metaplasia spectra, a second logistic regression

algorithm developed to separate high grade dysplasia from squamous metaplasia was developed to increase the specificity of the overall model.

The second algorithm was trained with all high grade dysplasia (score=1) and squamous metaplasia (score=0) spectra. The same inputs used in the first algorithm are also used as inputs in the second algorithm, but the output is a value (score) which represents the probability that the spectra was measured from an area of high grade dysplasia as compared with squamous metaplasia. Thus, the algorithm is also represented by equation 3.2, but the coefficients (b1-8) are different than those in the first algorithm. Any spectra from the original test set with a score from the original algorithm of score>0.5 was automatically used to test the second algorithm. Since there were not enough spectra to create a separate training and validation sets, the algorithm was trained with all of the available high grade dysplasia and squamous metaplasia spectra, thus some of the spectra in the test set were also used to train the algorithm.

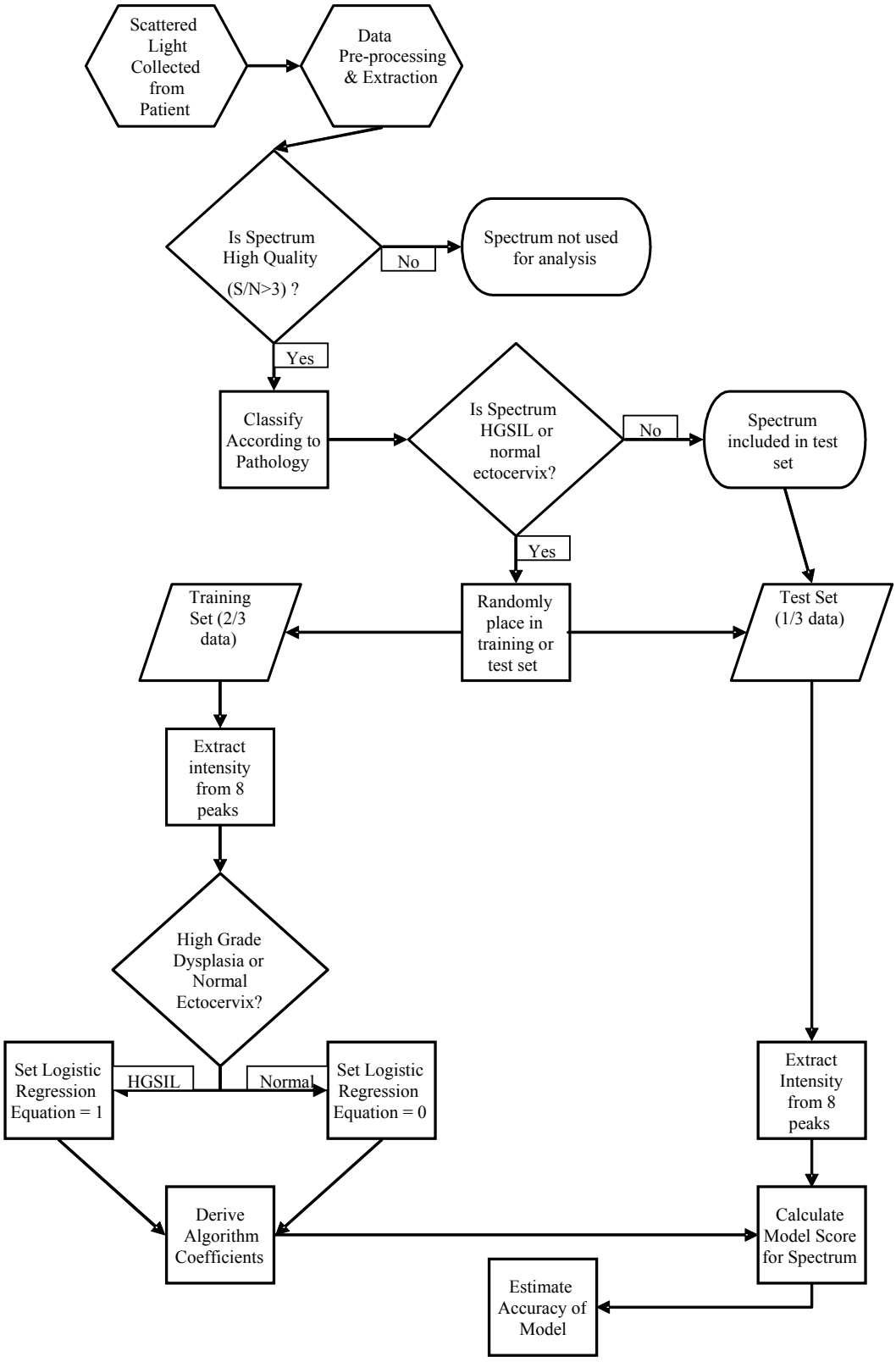

Chart 3.1: From Collection of Spectrum to Generation of First Algorithm

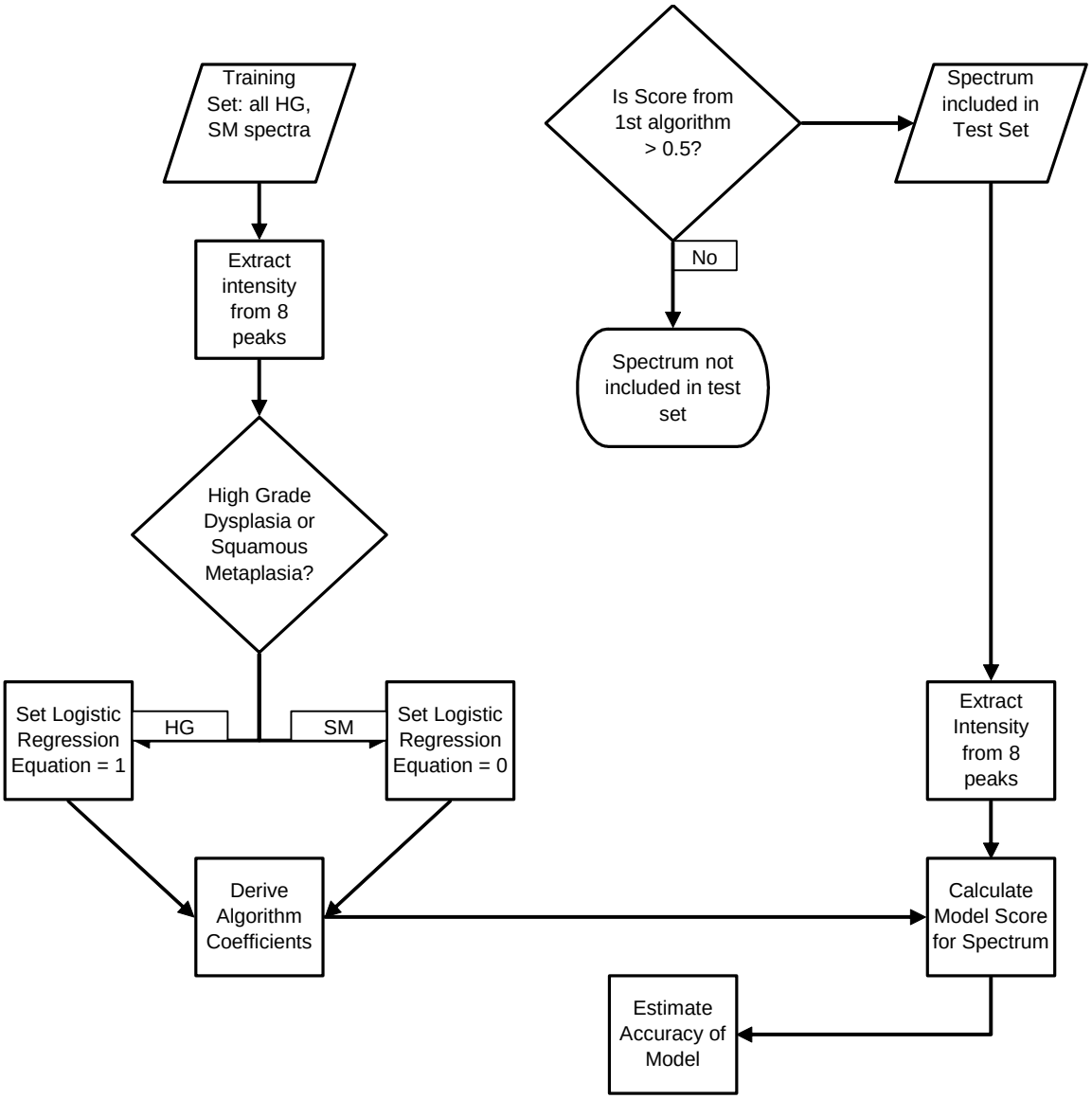

Chart 3.2 Generation of Second Algorithm

3.4 Results

The mean spectra for each pathology category are shown in figure 3.2. The most consistent peaks are labeled and found at 1006, 1058, 1086, 1244, 1270, 1324, 1450, 1550, 1655 cm^{-1}. These peaks are found in all the patient spectra.

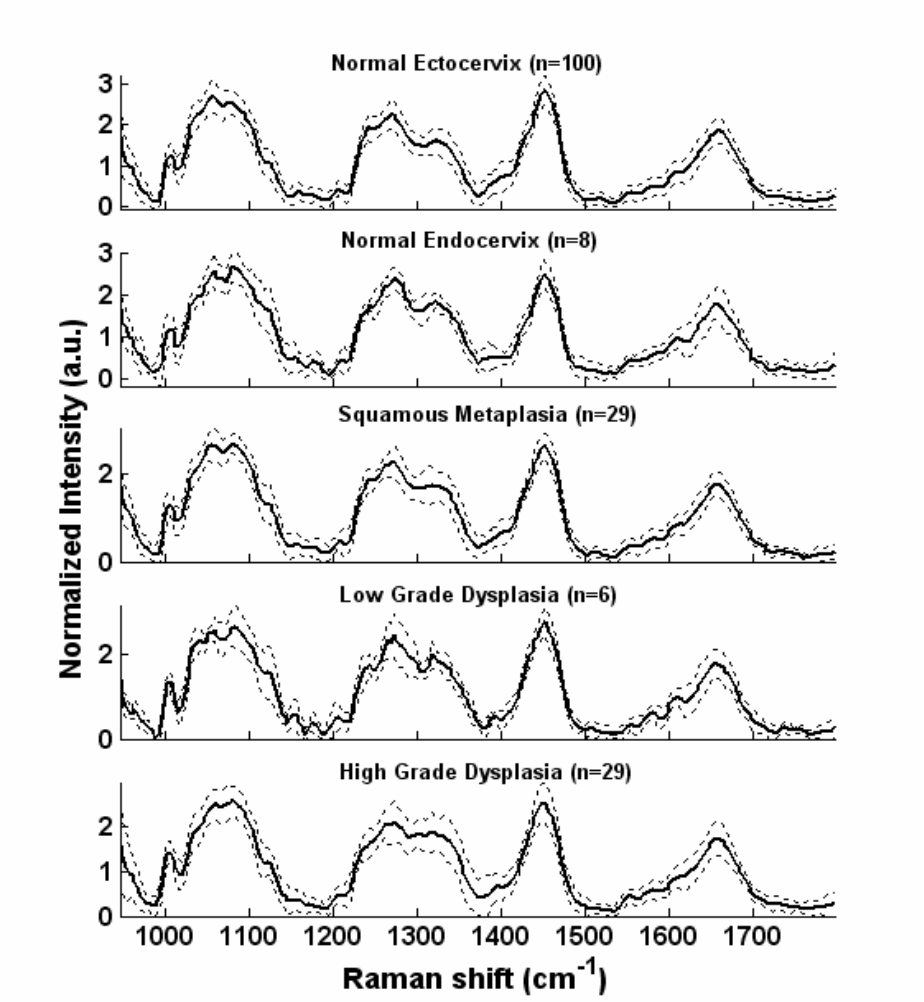

Figure 3.2: Mean Raman spectra of all pathology categories.
The mean ± one standard deviation spectra are plotted for each pathology category. N= number of spectra.

Although the peak shapes and locations show consistency across all pathology classifications, there are small but significant differences in peak intensities between the different

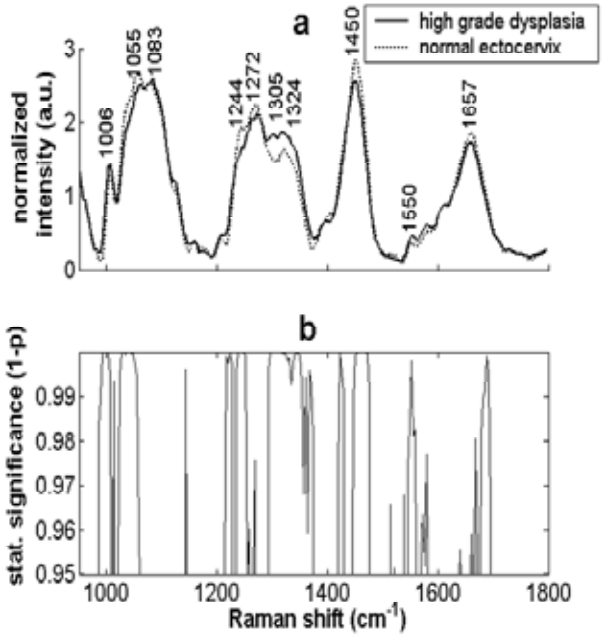

Figure 3.3: Spectra Comparison: High Grade Dysplasia vs. Normal Ectocervix.
(a) Mean spectra overlay of high grade dysplasia (n=29 spectra) and normal ectocervix (n=100 spectra), (b) results of t-test at each wavenumber (Note: y-axis is 1-p).

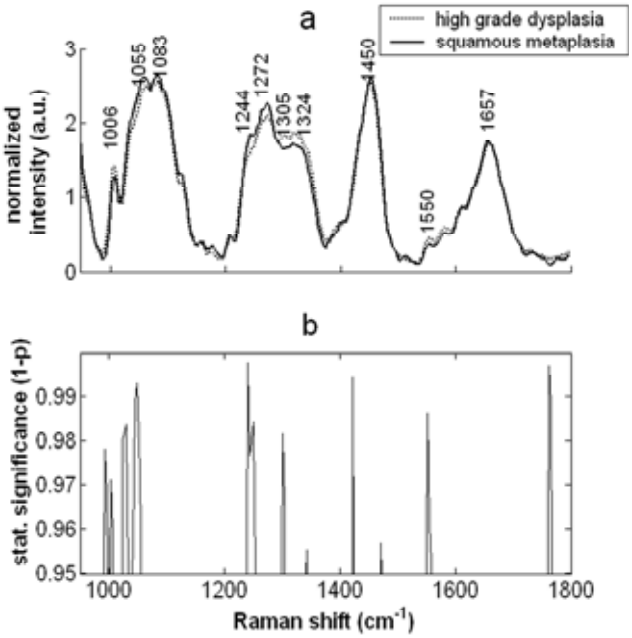

Figure 3.4: Spectra Comparison: High Grade Dysplasia vs. Squamous Metaplasia.
(a) Mean spectra overlay of high grade dysplasia (n=29 spectra) and squamous metaplasia (n=29 spectra), (b) results of t-test at each wavenumber (Note: y-axis is 1-p).

pathology categories as are shown in figures 3.3-3.5 and tables 3.2-3.3. The spectral differences between high grade dysplasia and normal ectocervix spectra (figure 3.3) and high grade dysplasia and squamous metaplasia spectra (figure 3.4) are illustrated with a mean spectral overlay which displays the visual spectral differences and a plot of the corresponding results of t-tests performed at each individual wavenumber. These two spectral comparisons are highlighted to denote that these categories were used in developing the discrimination algorithms.

Looking at the normal ectocervix/high grade dysplasia spectral comparison, several spectral regions, including 1006, 1055, 1305-1330, and 1450 cm^{-1}, show statistical significance at $p< 0.001$ and several more peaks that show significance at $p< 0.01$ including 1550 and 1655 cm^{-1}. The comparison of squamous metaplasia and high grade dysplasia spectra in figure 3.4 shows fewer regions of statistical difference. There are no regions which show significance at

p<0.001, but several peaks show significant differences at the level of p< 0.01, including 1058, 1244, and 1550 cm^{-1}. Although there is a difference at the 1324 cm^{-1} peak, it does not vary significantly. While there is a valley at 1305 cm^{-1} in the normal ectocervix spectra, the same region is flat in the spectra of the squamous metaplasia, and in comparison with the high grade dysplasia spectra, the difference is less significant than that of the normal ectocervix/high grade dysplasia spectral comparison. The 1450 and 1655 cm^{-1} peaks, which were also significantly different between the normal ectocervix and high grade dysplasia spectra, do not show a visual or statistical difference between the squamous metaplasia and high grade dysplasia spectra.

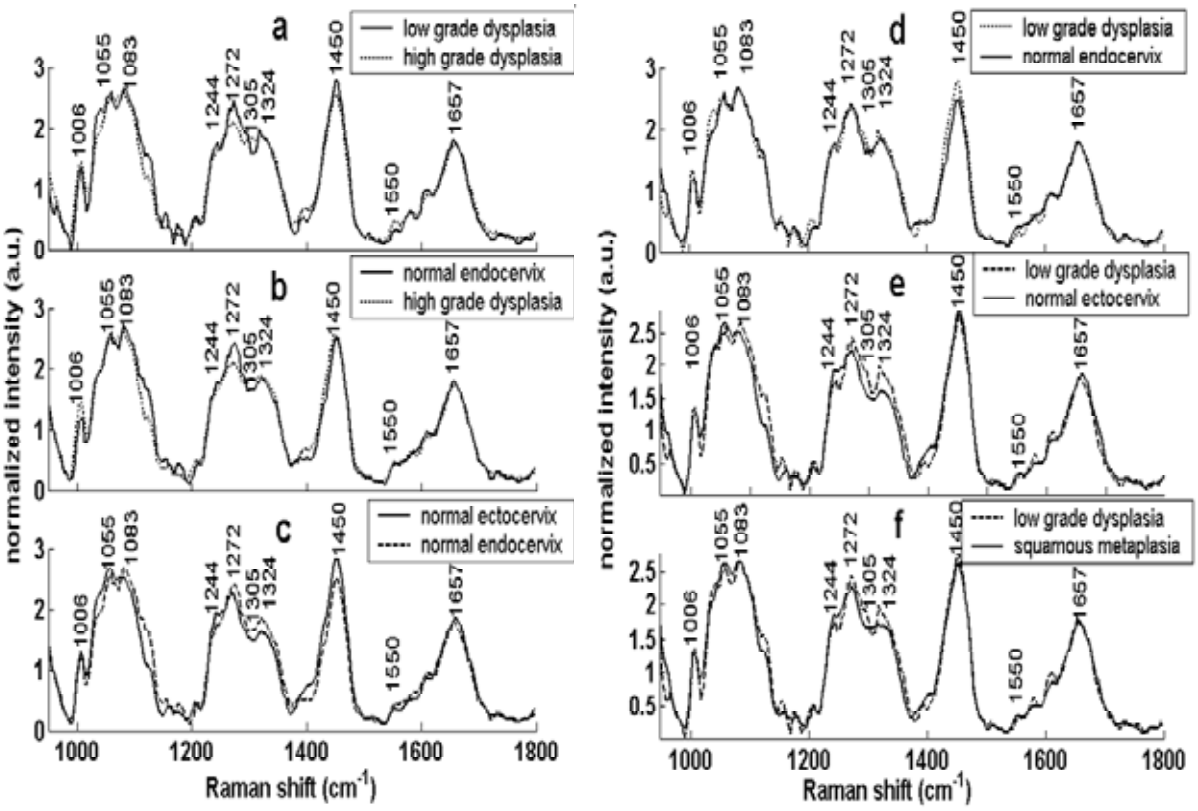

Figure 3.5: Comparisons of Mean Spectra.
Mean spectra overlays for the following categories: (a) high grade dysplasia (n=29 spectra) and low grade dysplasia (n=6 spectra), (b) normal endocervix (n=8 spectra) and high grade dysplasia, (c) normal ectocervix (n=100 spectra) and normal endocervix, (d) low grade dysplasia and normal endocervix, (e) low grade dysplasia and normal ectocervix, (f) low grade dysplasia and squamous metaplasia (29 spectra).

Spectral comparisons of the remaining pathology categories are shown in figure 3.5. Although there are only 6 low grade dysplasia spectra in the analysis, some interesting trends are shown in the mean overlay plots seen in figure 3.5a, d, e and f. In panel e, the low grade dysplasia spectra shows an increase in the 1324 cm^{-1} peak as compared with normal ectocervix which is similar to the high grade dysplasia/normal ectocervix spectral comparison (figure 3.3). Yet the intensity of 1272 cm^{-1} peak in the low grade dysplasia spectra seems to remain similar to that seen in normal ectocervix, unlike in the high grade dysplasia spectra, where the peak is lower when compared with the spectra of normal ectocervix. Similarly, the 1450 cm^{-1} peak in the low grade dysplasia spectra also shows little deviation from that same peak in the normal ectocervix spectra, unlike the significant difference seen at 1450 cm^{-1} in the high grade dysplasia/normal ectocervix spectral comparison. The comparison of the low grade dysplasia with squamous metaplasia spectra shows few differences; the one exception is the area from 1300-1330 cm^{-1}. There is a large valley at 1305 cm^{-1} and subsequent peak at 1324 cm^{-1} in the low grade dysplasia spectra, whereas the squamous metaplasia spectra is relatively flat in that region, showing no real change from 1300- 1330 cm^{-1}. Unlike the comparisons of high grade dysplasia with normal ectocervix and squamous metaplasia spectra, the comparison of high grade dysplasia with normal endocervix spectra in figure 3.5b shows similarities in the intensity at 1324 cm^{-1} between the two categories. Yet, the high grade dysplasia spectra shows a decrease in intensity at the 1272 cm^{-1} peak, a change also seen in the comparisons with normal ectocervix and squamous metaplasia spectra. The normal endocervix spectra also seems to retain the valley seen at 1305 cm^{-1} similar to the normal ectocervix spectra but unlike the squamous metaplasia case. The 1006 cm^{-1} peak shows an increase in the high grade dysplasia spectra, a change also seen when compared with the normal ectocervix spectra. A comparison of normal endocervix and ectocervix spectra shows some striking differences. Most notably, there is an increase in the 1272 and 1324 cm^{-1} peaks in the endocervix spectra as compared with ectocervix spectra while the 1450 cm^{-1} peak shows a decrease in intensity. There is also a significant difference in the shoulder ~1100 cm^{-1} and the valley at 1400 cm^{-1}. All of the aforementioned spectral differences are diagrammed in tables 3.2 and 3.3.

Table 3.2: Spectral Changes Seen Between High Grade Dysplasia and Other Pathologies.
The arrow represents the direction of the change in spectral intensity at that peak (location shown in cm^{-1}). Statistically significant differences ($p<0.01$) are denoted with a larger arrow.

High Grade Dysplasia										
vs.	1006	1058	1083	1244	1272	1305	1324	1450	1550	1657
Normal Ectocervix	↑	↓	↔	↓	↓	↑	↑	↓	↑	↑
Normal Endocervix	↑	↔	↔	↔	↓	↑	↔	↔	↔	↔
Squamous Metaplasia	↑	↓	↔	↓	↓	↑	↑	↔	↑	↔
Low Grade Dysplasia	↔	↔	↔	↔	↓	↑	↔	↑	↔	↔

Table 3.3: Spectral Changes Seen Between Low Grade Dysplasia and Other Pathologies.
The arrow represents the direction of the change in spectral intensity at that peak (location shown in cm^{-1}). Statistically significant differences ($p<0.01$) are denoted with a larger arrow.

Low Grade Dysplasia										
vs.	1006	1058	1083	1244	1272	1305	1324	1450	1550	1657
Normal Ectocervix	↔	↓	↑	↓	↑	↔	↑	↔	↔	↔
Normal Endocervix	↔	↔	↔	↔	↔	↔	↔	↑	↓	↔
Squamous Metaplasia	↔	↔	↔	↔	↑	↔	↑	↔	↔	↔

Figure 3.6a shows the results from the test set as scored by the first algorithm developed to distinguish high grade dysplasia from normal ectocervix.

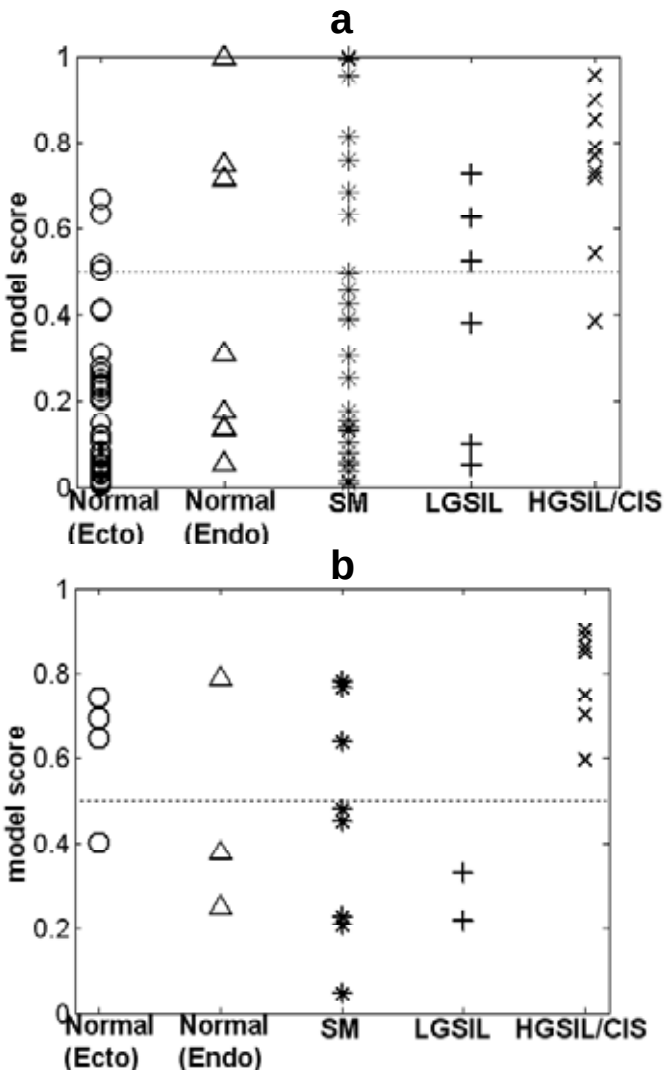

Figure 3.6: Model Classification of Test Set.
(a) model scores for first unbiased algorithm (b) model scores for second algorithm, all spectra with model score of p>0.5 in the first algorithm were used as inputs to second algorithm

Each data point represents an individual spectrum from the independent test set. The spectra are grouped along the x-axis by pathology classification as determined by histological examination, and the corresponding model score is shown on the y-axis. The line drawn at y=0.5 represents an unbiased model separation of high grade dysplasia (y>0.5) from normal ectocervix (y<0.5). The model classifies 8 out of 9 high grade spectra and 31 out of 35 normal ectocervix spectra correctly. The figure also illustrates the results from the model classification of the remaining three pathology types (normal endocervix, squamous metaplasia, low grade dysplasia). Although

the model was not trained with spectra from these pathology categories, an attempt to classify these spectra using this model was made based on the premise that normal ectocervix and high grade dysplasia represent the extremes of pathology within the cervix. The majority of the normal endocervix spectra classify as normal (score<0.5) which is promising for separating normal endocervix from high grade dysplasia. The scores for the majority (4 out of 6) of the low grade dysplasia spectra cluster around score=0.5 which would be expected as low grade dysplasia represents an intermediate from normal to high grade dysplasia. The scores from the squamous metaplasia spectra lack any definitive pattern of classification. Although squamous metaplasia represents a benign change of the cervix, using the initial algorithm, 8 out of 29 spectra classify as high grade dysplasia. Based on these observations, a second algorithm to further distinguish between squamous metaplasia and high grade dysplasia was developed to enhance the specificity of the model when distinguishing high grade dysplasia from all benign areas of the cervix.

The second algorithm is also a logistic regression based algorithm, trained to distinguish between squamous metaplasia and high grade dysplasia. Figure 3.6b shows the results of this model classification. This second algorithm is able to correctly classify 5 out of 8 squamous metaplasia spectra and all of the high grade dysplasia spectra. Again, since all spectra from the original test set with a mode score>0.5 were used as inputs to the second algorithm, some normal ectocervix and endocervix and low grade dysplasia spectra from pathology categories were included in the test despite the fact that those pathologies were not used in the training of the model. Among these, 3 out of 4 of the normal ectocervix spectra, none of the low grade dysplasia spectra, and 1 out of 3 of the normal endocervix classified as high grade dysplasia.

Comparisons of the accuracy of the model classification (with and without the second algorithm) are shown with the corresponding results of the accuracy of the colposcopic diagnosis as recorded in the clinical studies in table 3.4. Results from low grade dysplasia and normal endocervix are not included in the calculations of accuracy due to the fact that these categories were not included in the training algorithms. Thus the computed accuracies shown in table 3.4 compare only normal ectocervix and squamous metaplasia with high grade dysplasia. The colposcopic results were also computed only for diagnoses of normal ectocervix, squamous metaplasia, or high grade dysplasia.

Table 3.4: Accuracy of Classification.
The accuracy of the model classification are compared with the accuracy of the physician colposcopic diagnosis. HG dysplasia=high grade dysplasia. Note: italicized columns represent results from a biased estimates of accuracy.

	HG dysplasia vs. Benign (normal ectocervix + squamous metaplasia			**HG dysplasia vs. Normal ectocervix**		**HG dysplasia vs. squamous metaplasia**	
	Model algorithm 1	Model algorithms 1+ 2	Colposcopy	Model algorithm 1	Colposcopy	Model algorithms 1 + 2	Colposcopy
Sensitivity	89%	89%	87%	89%	89%	89%	97%
Specificity	81%	88%	72%	89%	88%	90%	8%

3.5 Discussion

This study represents the next step in developing Raman spectroscopy as a potential tool in the screening and diagnosis of cervical dysplasia and cancer. Many improvements over the pilot study investigating the potential of using Raman spectroscopy for diagnosis in the cervix have been made in this study [32]. The system used in this study, diagrammed in figure 3.1, allows collection of high quality spectra with a much lower signal collection time (5 seconds versus 60-1200 seconds in the pilot study) [32]. The use of automated signal processing methods yields more reproducible spectra than the pilot study [28]. In the pilot study, peak locations varied with standard deviations ranging from 7.5 – 35 wavenumbers whereas the peak locations in this study vary with a standard deviation of less than 5 wavenumbers . These two advances, coupled with robust discrimination algorithms, make it feasible to develop a system to provide real time diagnosis in the clinic or operating room.

Results from the model classification show that Raman spectroscopy in combination with an logistic regression algorithm can distinguish between high grade dysplasia and benign areas of the cervix (normal ectocervix, squamous metaplasia) with similar sensitivity and higher specificity using unbiased estimates of accuracy when compared to colposcopy. Logistic regression provides a particularly powerful approach to the task of classifying spectra into pathologic categories. Since the output of the model is a value (score) which represents the percentage likelihood that a spectrum falls into one of two categories, the model not only can be used to classify the spectra, but also used to aid health care providers in quantifying and

following the disease of the patient over time. This type of score also enables the provider to use his or her knowledge from the clinical examination in combination with the information given by the model to derive a diagnosis for each individual patient rather than rely solely on a computer model to generate a diagnosis.

The model presented here is a preliminary formulation designed to estimate the ability of Raman spectroscopy to perform pathologic classification of individual spectra on a larger scale. Even though multiple spectra were sometimes collected from the same site in a particular patient, only one spectra was used per location within a given patient for the data analysis. The first spectra collected at each site was used in lieu of average spectra to best predict the ability of Raman spectroscopy to perform diagnosis on a single spectrum. This prevents an overestimation of the accuracy predictions. If only the first algorithm of the model is used, the estimates of model accuracy are completely unbiased and therefore, statistically should perform equally as robust on future data sets. Since the second algorithm of the model contains some of the same data in both the training and test sets, performance estimates are biased, and thus cannot be used to predict future accuracy. Although the model performs well when compared with the standard of care (colposcopy) (see table 3.4), further testing of the second algorithm is necessary in the form of an independent validation set of data to clarify the future performance of the second algorithm. Yet, developing this two tiered model allows the ability to predict the type of algorithm or series of algorithms that are likely to successfully classify the data for future model development.

Intensity values from eight of the major Raman spectral peaks were chosen as inputs for the model based on results of t-tests between spectra of different pathologies. By only using those peaks which showed statistically significant differences, the model inputs represent spectral areas which show potential for future algorithm development. Analysis of the correlation between model coefficients indicated that no two model inputs are highly correlated (i.e. $r>0.75$). This ensures that there are no redundant model inputs. Yet, future development may be able to reduce the number of model inputs.

Though a model could have been constructed with multiple algorithms to classify each pathology category individually using a leave one out method, doing so would not give a true estimate of the future accuracy of such algorithms. Since so few low grade dysplasia and normal endocervix spectra were available for analysis, we chose to not attempt to train a separate

algorithm to specifically classify these categories. Their inclusion in the testing of the model was done as an exploratory process, not for the purposes of assessing accuracy. A multiple algorithm model may prove to be even more accurate at classifying spectra from these pathology categories.

Although the low grade dysplasia and normal endocervix spectra were not included in training of the model, the results from their inclusion in the test set show definite trends of classification (figure 3.6). Using only the first algorithm, the majority (5 out of 8) of the normal endocervix spectra classify with the normal ectocervix spectra, and while promising, it indicates the need for a separate algorithm to distinguish endocervix from dysplasia spectra. When placed in the second algorithm, an additional 2 out of 3 of the normal endocervix spectra classified with the squamous metaplasia spectra (or "benign") bringing the overall classification of endocervix spectra as benign to 7 out of 8. The classification of low grade dysplasia spectra using the first algorithm shows that the model scores for 4 out of 6 low grade dysplasia spectra cluster between 0.4-0.75. This finding could result from the fact that low grade dysplasia is an intermediate stage in the progression of normal ectocervix to high grade dysplasia, and thus the low grade dysplasia would classify in the middle. Thus, it is possible that a single algorithm would be able to separate normal cervix, low grade dysplasia, and high grade dysplasia with the accumulation of a larger training and test set of data. A true estimate of the ability of a model can only occur with a larger data set and an increase in the numbers of low grade dysplasia and normal endocervix spectra.

It is encouraging that the most significant spectral differences of all of the pathology comparisons are seen in the normal ectocervix/high grade dysplasia spectral comparison. Because normal ectocervix and high grade dysplasia represent the two extremes of a continuum of the development of neoplasia, they have more biochemical and morphological difference than any other categories, which would theoretically lead to differences in the Raman spectra. These two groups also have the highest number of spectra of any of the categories, and thus, significant differences are more likely to be detected when present. The results from the model classification of high grade dysplasia and normal ectocervix using the first algorithm also are encouraging (figure 3.6). The discrimination line at p=0.5 was used to ensure that the estimates of algorithm accuracy were completely unbiased. Yet, if the discrimination line is drawn at 0.75, the specificity of the discrimination increases to 100% while only misclassifying one additional high grade dysplasia case, although this would bias the classification on the test set. A future

independent validation set of data could test whether a discrimination line at 0.75 would truly yield better results. It should be noted that 3 of the spectra in the high grade dysplasia test set have a histological diagnosis of invasive squamous cell carcinoma, although no invasive carcinoma spectra were used in the training of the algorithms. These spectra were classified by the model as high grade dysplasia as would be expected. Their inclusion in the test set indicates that although the algorithm was trained to recognize high grade dysplasia spectra, it will also correctly identify more severe pathologies. Of all the categories, squamous metaplasia was the only category not exhibiting any trends of classification using the first algorithm (figure 3.6a). This result is not surprising given that although dysplasia often originates in the context of squamous metaplasia, squamous metaplasia results from a different process of tissue transformation than dysplasia. Although the second algorithm improved the classification of the squamous metaplasia spectra (figure 3.6b), the spectral differences between squamous metaplasia and high grade dysplasia are subtle, and thus generating more robust classification algorithms will likely require large numbers of spectra.

The differences seen between the mean spectra of high grade dysplasia and normal ectocervix correlate well with the differences observed using an organotypic raft culture model (figure 3.3) [30]. In these experiments, *in vitro* tissue cultures that mimic *in vivo* conditions of cervical tissue were grown and analyzed using Raman spectroscopy. The results from raft culture experiments indicate that the 1324 cm^{-1} peak arises primarily from the epithelial layer of tissue [30]. This finding is in agreement with previous studies which assigned the region surrounding 1330 cm^{-1} to the purine bases in the nucleic acid of epithelial cells as evidenced by confocal measurements of skin epithelium and classical Raman spectroscopy peak assignment [15, 18, 31]. In contrast, the 1272 cm^{-1} is an Amide III peak and arises primarily from the type I collagen in the stromal layer of the tissue [30, 31]. Thus, the changes seen in the spectra of high grade dysplasia reflect changes in the structure and biology during the transformation from normal to dysplasia. The increase seen at 1324 cm^{-1} in the high grade dysplasia spectra is most likely due to the increase in the cellular and nucleic acid content in dysplasia while the decrease at 1272 cm^{-1} likely reflects a decrease in the light collected from the stromal layer due to the increase in density (and thus decrease in transparency) of the epithelial layer. The 1450 and 1655 cm^{-1} peaks have contributions from both the epithelial and stromal layers [30]. The 1450 cm^{-1} peak is traditionally thought to arise from the C-H stretch in lipids whereas the 1657 cm^{-1} peak is

an amide I peak [15]. Since the 1657 cm^{-1} peak has origins in the epithelial and stromal layers, it likely arises from a bond common to both collagen and cellular proteins, and likely has contributions from the keratin [18, 30]. The origins of the peaks at 1244 and 1305 cm^{-1} are less clear although 1244 cm^{-1} is often included in the amide III region; both peak intensities are significantly different in the high grade dysplasia/normal ectocervix and high grade dysplasia/squamous metaplasia spectral comparisons (figures 3.3, 3.4). Although the majority of the peak centered at 1070 cm^{-1} is due to the Raman scatter of the silica in the fiber optic probe, the smaller peaks at 1055 and 1083 cm^{-1} likely arise from the tissue [31]. The peak at 1083 cm^{-1} has been shown to arise from acyl backbone in lipids. The 1006 cm^{-1} is known to arise from phenylalanine and the source of 1550 cm^{-1} is thought to be tryptophan [15]. Future studies will establish the origins of these peaks in a more definitive manner.

These results show that Raman spectroscopy, in conjunction with a model developed using logistic regression algorithms, can distinguish high grade dysplasia from normal ectocervix and squamous metaplasia with a similar sensitivity and higher specificity than colposcopy performed by experts. Future clinical studies are needed to increase the numbers of spectra, especially in the low grade dysplasia and normal endocervix categories, such that independent training and test sets can be used to develop more effective algorithms and further discriminate all pathology categories.

3.6 References

1. Myers, E.R., *et al.*, Setting the target for a better cervical screening test: characteristics of a cost-effective test for cervical neoplasia screening. *Obstet Gynecol*, 2000. **96**(5 Pt 1): p. 645-52.

2. Nanda, K., *et al.*, Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*, 2000. **132**(10): p. 810-9.

3. Mitchell, M.F., *et al.*, Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol*, 1998. **91**(4): p. 626-31.

4. Cantor, S.B., *et al.*, Cost-effectiveness analysis of diagnosis and management of cervical squamous intraepithelial lesions. *Obstet Gynecol*, 1998. **91**(2): p. 270-7.

5. Georgakoudi, I., *et al.*, Trimodal spectroscopy for the detection and characterization of cervical precancers *in vivo*. *Am J Obstet Gynecol*, 2002. **186**(3): p. 374-82.

6. Ramanujam, N., Fluorescence spectroscopy of neoplastic and non-neoplastic tissues. *Neoplasia*, 2000. **2**(1-2): p. 89-117.

7. Nordstrom RJ, B.L., Niloff JM and Myrtle JF, Identification of Cervical Intraepithelial Neoplasia (CIN) Using UV-Excited Fluorescence and Diffuse-Reflectance Tissue Spectroscopy. *Lasers Surg Med*, 2001. **29**: p. 118-27.

8. Ramanujam N, M.M., Mahadevan-Jansen A, Thomasen S, Staerkel G, Malpica A, Wright T, Atkinson A, Richards-Kortum R, Cervical pre-cancer detection using a multivariate statistical algorithm based on laser induced fluorescence spectra at multiple excitation wavelengths. *Photochem Photobiol*., 1996. **64**: p. 720-735.

9. Agrawal, A., *et al.*, Fluorescence spectroscopy of the cervix: influence of acetic acid, cervical mucus, and vaginal medications. *Lasers Surg Med*, 1999. **25**(3): p. 237-49.

10. Brookner, C.K., *et al*., Cervical fluorescence of normal women. *Lasers Surg Med*, 1999. **24**(1): p. 29-37.

11. Brookner, C.K., *et al.*, Autofluorescence patterns in short-term cultures of normal cervical tissue. *Photochem Photobiol*, 2000. **71**(6): p. 730-6.

12. Drezek, R., *et al.*, Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: modeling, measurements, and implications. *J Biomed Opt*, 2001. **6**(4): p. 385-96.

13. Drezek, R., *et al.*, Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochemistry with dysplasia. *Photochem Photobiol*, 2001. **73**(6): p. 636-41.

14. Mahadevan-Jansen, A., R. Richards-Kortum, Raman Spectroscopy for the Detection of Cancers and Precancers. *Journal of Biomedical Optics,* 1996. **1**(1): p. 31-70.

15. Stone, N., *et al.*, Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 564-573.

16. Stone, N., *et al.*, Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. *Laryngoscope*, 2000. **110**(10 Pt 1): p. 1756-63.

17. Shafer-Peltier, K., *et al.*, Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 552-563.

18. Caspers, P.J., *et al*., *In* vitro and in vivo Raman spectroscopy of human skin. *Biospectroscopy*, 1998. **4**(5 Suppl): p. S31-9.

19. Mahadevan-Jansen, A., *et al.*, Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers. *Photochem Photobiol*, 1998. **68**(1): p. 123-32.

20. Huang, Z., *et al.*, Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int J Cancer*, 2003. **107**(6): p. 1047-52.

21. Crow, P., *et al.*, The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*. *Br J Cancer*, 2003. **89**(1): p. 106-8.

22. Koljenovic, S., *et al.*, Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy. *Lab Invest*, 2002. **82**(10): p. 1265-77.

23. Shim, M.G., *et al.*, *In vivo* near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy. *Photochem Photobiol*, 2000. **72**(1): p. 146-50.

24. Molckovsky, A., *et al.*, Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. *Gastrointest Endosc*, 2003. **57**(3): p. 396-402.

25. Caspers, P.J., *et al.*, *In vivo* confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. *J Invest Dermatol*, 2001. **116**(3): p. 434-42.

26. Bakker Schut, T.C., *et al.*, *In vivo* detection of dysplastic tissue by Raman spectroscopy. *Anal Chem*, 2000. **72**(24): p. 6010-8.

27. Savitsky, A. and M. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem*, 1964. **36**(8): p. 1627-1639.

28. Lieber, C.A. and A. Mahadevan-Jansen, Automated Method for the Subtraction of Fluorescence from Biological Raman Spectra. *Applied Spectroscopy*, 2003. **57**(11): p. 1363-1367.

29. Hosmer, D.W. and S. Lemeshow, *Applied Logistic Regression*. 1989, New York: John Wiley & Sons.

30. Robichaux Viehoever, A., *et al.*, Organotypic Raft Cultures as an Effective *In vitro* Tool for Understanding Raman Spectral Analysis of Tissue. *Photochem. Photobio.*, 2003. **78**(5): p. 517-24.

31. Ferraro, J. and K. Nakamoto, *Introductory Raman Spectroscopy*. 1994: Academic Press.

32. Utzinger, U., *et al.*, Near-Infrared Raman Spectroscopy for *In vivo* Detection of Cervical Precancers. *Applied Spectroscopy*, 2001. **55**(8): p. 955-9.

CHAPTER IV

CHARACTERIZATION AND QUANTIFICATION OF THE SOURCES OF VARIABILITY PRESENT IN RAMAN SPECTRA OF THE CERVIX.

Amy Robichaux Viehoever[1], Heidi Shappell[2], Dean Billheimer[3], Howard Jones, III[4],and Anita Mahadevan-Jansen[1]

[1]Department of Biomedical Engineering,
[2]Department of Pathology,
[3]Department of Preventive Medicine,
[4]Department of Obstetrics and Gynecology
Vanderbilt University, Nashville, Tennessee 37235

This manuscript was prepared for submission in the *Journal of Biomedical Optics*.

4.1 Abstract

Raman spectroscopy has the potential to provide non-invasive, real-time differential diagnosis of different pathologies in the cervix. Yet the sources and relative contributions of spectral variability for a given pathology must be understood to accurately predict whether Raman spectroscopy can be an effective method for screening for cervical cancer. This study seeks to understand and quantify the sources of variance within spectra of normal ectocervix.

For this study, Raman spectra were collected *in vivo* from 35 patients with no evidence of cervical dysplasia who were undergoing hysterectomy for indications other than cervical disease. Data from the first five patients was used to determine the optimal signal collection time. Data from the next two patients in addition to data from a concurrent study of dysplasia patients examined the effect of the application of acetic acid on the spectra. The remaining patient studies were used to quantify the spectral variance arising from within a site, within a patient, and between different patients. Further analysis of the sources of interpatient variability examined the spectral differences due to menopausal status, smoking history, and overall patient diagnosis (normal from normal patients vs. normal from dysplasia patients).

The optimal signal collection time was determined to be five seconds. The acetic acid analysis showed two small, non-significant spectral changes arising after the application of acetic acid; no time dependant changes were noted. Interpatient variability was found to be the main source of variability within normal ectocervix, contributing 73% of the total variance. Significant spectral differences were found between pre-menopausal and post-menopausal normal spectra as well as between normal spectra from normal patients and normal spectra from dysplasia patients. Spectral differences between smokers and non-smokers was negligible. The analyses presented here improve the understanding of the sources of variation within a given pathology and will aid the future development of discrimination algorithms for the detection of cervical dysplasia.

4.2 Introduction

In recent years, optical spectroscopy has been the focus of several groups as a modality for accurate, real time diagnosis of cancer and precancerous lesions [1, 2]. Among these, Raman spectroscopy is a novel, molecular specific technique that has been shown to have the potential for providing differential diagnosis in the gastrointestinal tract [3], larynx [4], breast [5], skin [6], cervix [7], lung [8], lymph nodes [9], prostate [10], and brain [11]. Previous *in vivo* studies

investigating the potential of using Raman spectroscopy for non-invasive detection of dysplasia (precancer) in the cervix showed that Raman spectroscopy can distinguish high grade dysplasia from all other pathologies using empirically determined peak ratios [21]. Since spectral measurements and analysis can be done in real time, Raman spectroscopy also has the potential to be a more cost effective means of screening, diagnosis, and guidance of therapy.

Preliminary data analysis from an ongoing clinical study examining the ability of Raman spectroscopy to detect cervical dysplasia indicates that a non-trivial amount of spectral variation exists between different spectral measurements of the same tissue pathology [12]. In order to understand and predict whether Raman spectroscopy can be an effective method for screening for cervical cancer, the origins and relative contributions of the sources of this variation must be understood. Such an understanding could lead to better design of clinical trials, rational design of automated algorithms for real time diagnosis, and better prediction of the potential of Raman spectroscopy to accurately perform diagnosis within the general population.

In examining sources of variation of Raman spectra measured from normal skin, Knudsen *et. al.* looked at variations due to time lapse (spectra measured at different times during the day as well as day to day variations), different locations within the body, different measurements at the same location, and different persons in 13 volunteers as well as variations due to skin pigmentation which was studied in 140 volunteers [13]. The variability was quantified by comparing relative intensities of peaks; small differences were found in the overall spectral intensity (affected by differences in skin pigmentation), the amide I peak (intersubject and between subsequent measurements), and the amide III peak (diurnal and intersubject). Numerous studies using fluorescence spectroscopy in the cervix have looked at different sources of intrapatient and interpatient variability including menopausal status, menstrual cycle variations, age, smoking history, the application of acetic acid, and overall diagnosis of the patient (comparing normal spectra from normal patients with normal spectra from dysplasia patients)[14-17]. Yet, none of these studies have quantified the contributions of variance from individual sources to the overall variance.

Thus, this study was designed to examine and quantify the sources of the intrinsic spectral variation observed *in vivo*. Raman spectra were measured from the normal cervix of women undergoing hysterectomy for reasons other than cervical disease. The amount of spectral variation was quantified for the following categories: within the same site (intrasite), between

different locations in the same patient (intrapatient), between different patients (interpatient), and instrinsic measurement variation. The contributions of potential sources of interpatient variation were examined including: menopausal status, smoking history, and overall patient diagnosis. The application of acetic acid was studied as a possible source of intrapatient variation, and the optimal length of time for collecting signal was quantified. By examining the contributions of these sources of variation, a deeper understanding of how these variations affect tissue spectroscopic diagnosis is achieved.

4.3 Methods

4.3.1 Clinical Study Design

Thirty-five patients undergoing total abdominal or vaginal hysterectomies were recruited to participate in the study as approved by the Vanderbilt Institutional Review Board (IRB). To be eligible for enrollment in the study, patients must be undergoing a hysterectomy for reasons other than cervical disease, between the ages of 18-75, and have a history of previous normal Papanicolaou (Pap) smear or normal exam of the cervix. Informed consent was obtained from each patient prior to the procedure. After the patient was placed under anesthesia, but before the hysterectomy procedure began, a colposcopic examination of the cervix under anesthesia was performed and the Raman spectral measurements were collected. Five percent acetic acid was swabbed on the cervix to visually enhance any abnormal areas of epithelium. Colposcopy was performed for two reasons: to ensure the measured areas were colposcopically normal and to maintain consistency with a concurrent study of dysplasia patients [12]. Multiple Raman spectra of colposcopically normal appearing sites then were measured *in vivo*. One background measurement was made with the laser off and the probe in contact with the tissue to measure the ambient light present at the tissue. The measured sites were marked with a methylene blue paste. The hysterectomy procedure then proceeded according to standard clinical protocol. Upon removal of the cervix, histologic analysis was performed by the participating gynecological pathologist (HS).

4.3.2 Data Collection

Spectral measurements were collected using the portable Raman spectroscopy system described in chapter 3. For this study, the fiber optic probe delivered 80mW 785 nm incident light onto the tissue and collected the light. The signal collection time varied between 1-15 seconds depending on the protocol under study (table 4.1). In all cases, the overhead fluorescent and colposcope lights were turned off during the measurements. Any luminescent lights were left on but turned away from the measurement site.

The spectral measurements followed one of three protocols (table 4.1). Studies on the first five patients were designed to determine the optimal signal collection time. In these patients, three sites on the cervix were chosen. Five spectra were acquired at each site, each using a different signal collection time (1, 3, 5, 8 seconds, or 3 accumulations of 5 seconds for a total of 15 seconds). The fiber optic probe was kept in contact with the tissue throughout the five measurements at each individual site. Studies on another 2 patients examined the affect of applying acetic acid to cervix. These patients were a part of a larger analysis ongoing with a concurrent study of dysplasia patients [12]. In this study, three sites were again chosen for measurements. All spectra were collected using a 5 second signal collection time. Spectra were acquired at each site prior to the application of acetic acid and at various time intervals ranging from 0-180 seconds following the application of acetic acid to determine the effect, if any, of acetic acid on the Raman spectra. Studies on the remaining study subjects (subjects 8-35) were designed to examine the sources of variation. In this final study, three sites were chosen for measurements (as in the other protocols); 3 spectra were collected at each site: the first two were collected without moving the fiber optic probe, and the last was acquired after removing the probe and replacing it at the same site. This allowed for determination of the relative variability due to user error of probe placement versus intrinsic variability of the Raman spectra. The measurements at different locations within the normal ectocervix of a given patient allowed for the calculation of variation due to location variability, and the comparison of spectra between different patients allowed for the calculation of variation arising from interpatient variation.

Table 4.1: Details of Different Measurement Protocols

Patient #	Signal Collection Time	# sites measured	# measurements per site
1-5	1,3, 5, 8, 15 seconds	3	5: one for each signal collection time
6-7 (plus 5 patients from dysplasia study)	5 seconds	3	4: relative to application of acetic acid (1) before, (2) 30-60 sec after (3) 60-90 sec after (4) 90+ sec after
8-35	5 seconds	3	3: 2 made without moving the probe, 1 made after removing the probe and placing back on the same site.

4.3.3 Data Pre-Processing and Extraction

Prior to the spectral measurements from each patient, a spectral calibration of the system was performed using a neon-argon lamp and naphthalene standard to correct for system wavenumber, laser excitation, and throughput variations. The spectra were processed as described in chapter 3. Following data processing, each spectrum was normalized to its mean spectral intensity across all Raman bands to account for overall intensity variability. These normalized spectra were categorized according to pathology as determined by histology and used for further comparison and analysis. If multiple spectra were collected from a single location within a given patient, only the first spectra collected was included for data analysis.

4.3.4 Data Analysis

To estimate the optimal integration time, the signal to noise ratio was calculated for each spectra collected at the different measurement times (1, 3, 5, 8, and 15 seconds) for the first five patients. An automated routine was developed to extract the level of signal (intensity at 1450 cm^{-1} peak) and the noise (largest intensity for region surrounding 1700 cm^{-1}) from the unfiltered spectra (each spectrum was binned along horizontal axis and the fluorescence background was subtracted, but the spectrum was not smoothed with any filter).

The effect of acetic acid on the Raman spectra was analyzed using a mean spectral overlay and student's t-test to compare the spectral variability at each wavenumber of spectra collected from areas of dysplasia before and after the application of acetic acid. This effect was

further quantified by calculating the spectral difference between individual spectra measured before and after the application of acetic acid at a specific site. The spectral difference was defined as the difference in intensity between the zero time point (no acetic acid) and various time points after the application of acetic acid at the 10 major peaks in the spectra (1006, 1058, 1083, 1244, 1272, 1305, 1324, 1450, 1550, 1657 cm^{-1}). The mean (across the ten peaks) of these intensity differences was calculated (such that one mean spectral difference was calculated for each spectrum) and used for the analysis.

The variability due to probe placement also was quantified by calculating the mean spectral difference (as described above) between an initial measurement and a subsequent measurement in which the probe was either kept in place or removed and replaced at the initial site. The remaining components of variance were calculated using analysis of variance (ANOVA) and Henderson's method [18]. Henderson's method estimates the variance due to each component by setting its observed mean squared error (MSE) from the ANOVA table equal to its expected value which can be represented as linear combinations of the variance components. The equations used for this analysis from Henderson's method are shown below in equations 4.1- 4.14 and were encoded for automatic calculation into MATLAB® R12.
The following calculations were done for each wavenumber in the spectrum:

First calculate the SS = "SUM OF SQUARES" for each category (of variability) for use in the ANOVA table:

$$\text{ss(patients)} = \sum_i \left(y_{i\,++}^2 / n_{i+} \right) - C \qquad \textit{(eq. 4.1)}$$

where

$$C = \left(y_{+++} \right)^2 / n_{++} \qquad \textit{(eq. 4.2)}$$

$n_{i+} = \sum_j n_{ij}$ = # observations for patient i

$y_{i++} = \sum_{jk} y_{ijk}$ = sum of intensity for patient i

$n_{++} = \sum_{i,j} n_{ij}$ = total number of observations

$y_{+++} = \sum_{i,j,k} y_{ijk}$ = grand sum

$$\text{ss(location)} = \sum_{i} \sum_{j} \left(y_{ij+}^2 / n_{ij} \right) - \sum_{i} \left(y_{i++}^2 / n_{i+} \right) \qquad \textit{(eq. 4.3)}$$

where

n_{ij} = # observations for patient i at location j

$y_{ij+} = \sum_{k} y_{ijk}$ = sum of intensity for patient i at location j

$$\text{ss(meas. error)} = \sum_{i} \sum_{j} \sum_{k} y_{ijk}^2 - \sum_{i} \sum_{j} \left(y_{ij+}^2 / n_{ij} \right) \qquad \textit{(eq. 4.4)}$$

where

y_{ijk} = intensity for the k th measurement for patient i at location j

Now, the ANOVA table can be constructed

Component	Degrees of freedom	SS	Mean Squares (MS)	Expected Vaule of MS (E<MS>)
PATIENTS (interpatient)	$P-1$	ss(patients)	$\frac{\text{ss(patients)}}{P-1}$	$\sigma_\Sigma^2 + C_1\sigma_L^2 + C_2\sigma_P^2$
LOCATION (intrapatient)	$L-P$	ss(location)	$\frac{\text{ss(location)}}{L-P}$	$\sigma_\Sigma^2 + C_3\sigma_L^2$
ERROR (intrinsic spectral variability)	$n_{++}-L$	ss(meas. error)	$\frac{\text{ss(meas. error)}}{n_{++}-L}$	σ_Σ^2

where:

P = Total # of patients

L = Total # of locations measured

σ_Σ^2 = variance due to intrinsic measurement error

σ_L^2 = variance due to location variability

σ_P^2 = variance due to patient variability

$$C_1 = \frac{\sum_{ij}\left(n_{ij}^2 / n_{i+}\right) - \sum_{ij}\left(n_{ij}^2 / n_{++}\right)}{P-1} \qquad \textit{(eq. 4.5)}$$

$$C_2 = \frac{n_{++} - \sum_{i}\left(n_{i+}^2 / n_{++}\right)}{P-1} \qquad \textit{(eq. 4.6)}$$

$$C_3 = \frac{n_{++} - \sum_{i}\sum_{j}\left(n_{ij}^2 / n_{i+}\right)}{P-1} \qquad \textit{(eq. 4.7)}$$

To calculate the variance due to each category, set the MS= E<MS>, leaving 3 equations with 3 unknowns:

$$\sigma_{\Sigma}^2 = \text{MS(MEAS. ERROR)} \qquad \textit{(eq. 4.8)}$$

$$\sigma_L^2 = \frac{MS(LOCATION) - \sigma_{\Sigma}^2}{C_3} \qquad \textit{(eq. 4.9)}$$

$$\sigma_P^2 = \frac{MS(PATIENTS) - \sigma_{\Sigma}^2 - C_1\sigma_L^2}{C_2} \qquad \textit{(eq. 4.10)}$$

For the final analysis, calculate the percentage variance for each category:

$$\text{total variance: } \sigma_T^2 = \sigma_{\Sigma}^2 + \sigma_L^2 + \sigma_P^2 \qquad \textit{(eq. 4.11)}$$

$$\%\text{ total variance for patient variability} = \frac{\sigma_P^2}{\sigma_T^2} \qquad \textit{(eq. 4.12)}$$

$$\%\text{ total variance for location variability} = \frac{\sigma_L^2}{\sigma_T^2} \qquad \textit{(eq. 4.13)}$$

$$\%\text{ total variance for measurement error} = \frac{\sigma_{\Sigma}^2}{\sigma_T^2} \qquad \textit{(eq. 4.14)}$$

4.4 Results

The mean ± one standard deviation spectra for normal ectocervix (panel a) and endocervix (panel b) are shown in figure 4.1. The major peaks are labeled and are conserved between both categories.

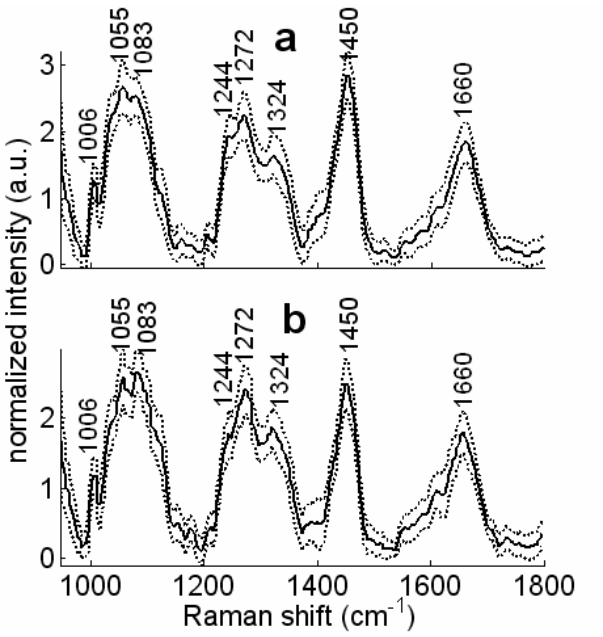

Figure 4.1: Normal Ectocervix vs. Normal Endocervix.
(a) mean spectra (solid) +/- 1 standard deviation (dotted) is shown for all unique normal ectocervix (squamous epithelium) (n=100 spectra, 36 patients) (b) mean spectra (solid) =/- 1 standard deviation (dotted) is shown for all normal endocervix (columnar epithelium) (n=8 spectra, 7 patients).

A comparison of the two mean spectra reveals that the peak intensities at 1006, 1055, 1083, 1550 and 1657 cm^{-1} are similar. Differences emerge at 1272 cm^{-1} and 1324 cm^{-1} where the intensities are higher in the endocervix spectra and at 1450 cm^{-1} where the intensity in the endocervix spectra is decreased. The standard deviation remains fairly constant across the wavenumber axis and varies between 0.18-0.40. It should be noted that only one spectrum per individual location was used in this analysis, such that if three measurements were made at each location, only one of those spectra were used.

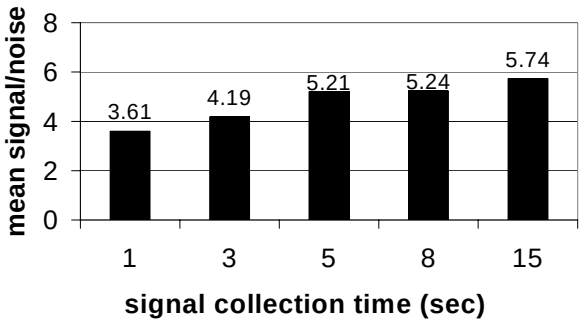

Figure 4.2: Signal to Noise Calculations for Different Measurement Times.
Signal to noise was calculated for 15 spectra from 5 patients collected using different measurement times. Signal was estimated using the Intensity of the 1450 cm-1 peak. Noise was estimated using the region surrounding 1700 cm-1.

Figure 4.2 shows the results from the analysis to determine the optimal signal collection time in a bar chart comparing the average signal to noise ratio for each integration time. The signal to noise ratio increases linearly with the measurement time until 5 seconds. Conversely, the increase in signal to noise at 8 seconds and 15 seconds is not proportional to the increase in signal collection time. Since the signal to noise ratio for 5 seconds was similar to that of the longer integration times, 5 seconds was chosen for the measurement time in subsequent studies.

A comparison of the mean spectra measured before and after the application of acetic acid is shown in figure 4.3a with the corresponding results of the t-test at each wavenumber in figure 4.3b.

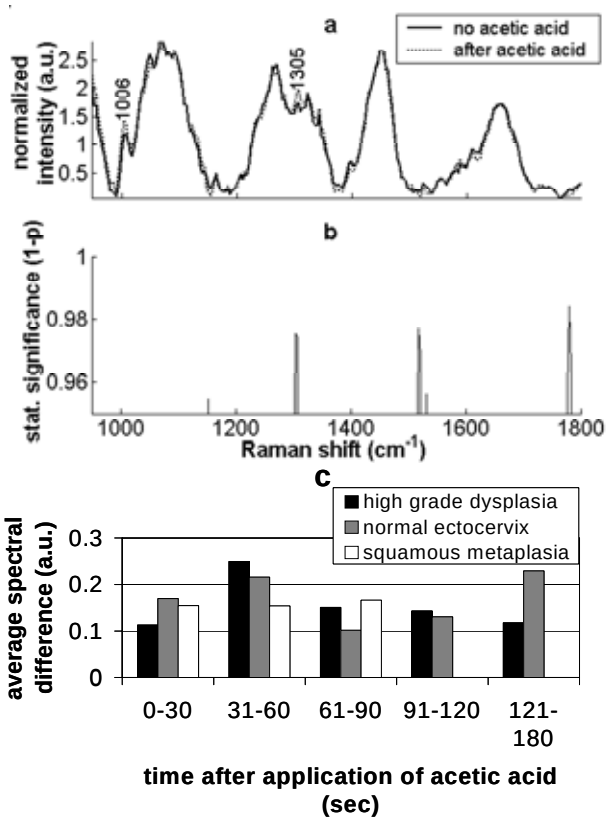

Figure 4.3: Acetic Acid Analysis.
(a) mean spectra of high grade dysplasia before and after acetic acid. (b) corresponding t-test performed at each wavenumber (c) average intensity difference between spectra measured at a given site before and after the application of acetic acid for the 10 major peaks at individual time points (the data are divided into dysplasia (n=3 patients, 6 sites of measurement), normal ectocervix (n=4 patients, 12 sites of measurement), and squamous metaplasia (n=2 patients, 3 sites of measurements).

The mean spectral overlay shows two areas of difference: the intensity at peaks 1006 and 1305 cm^{-1}. The t-test indicates that neither of these differences are significant (at the level of $p<0.01$), although the difference at 1305 cm^{-1} has a significance of $p=0.016$. A bar chart illustrating the spectral differences at different time points following the application of acetic acid is shown in figure 4.3c. The bar chart is stratified according to pathology type. Although there are outlier points in the 31-60 and 121-180 second groups, the spectral differences remain fairly constant across all time points, and the average magnitude of the spectral difference (0.15) is only slightly higher that seen in repeat measurements of the same site with no change in acetic acid application (0.12 – see figure 4.4).

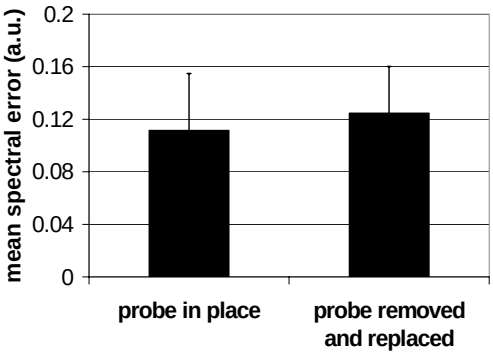

Figure 4.4: Intrasite variation.
The mean difference (or error) between the initial spectrum collected at a site and subsequent spectra measured either with the probe kept in place or removed and replaced to the same site is shown. The error bars represent one standard deviation.

Figures 4.4 and 4.5 show the results from the components of variance analysis in which the variation from between patients, between sites in a single patient, and within a site are quantified. Intrasite variation is illustrated in figure 4.4. Although the average spectral error is slightly higher when the probe was removed and replaced, the difference is not significant. The total variance at each wavenumber is plotted in figure 4.5a.

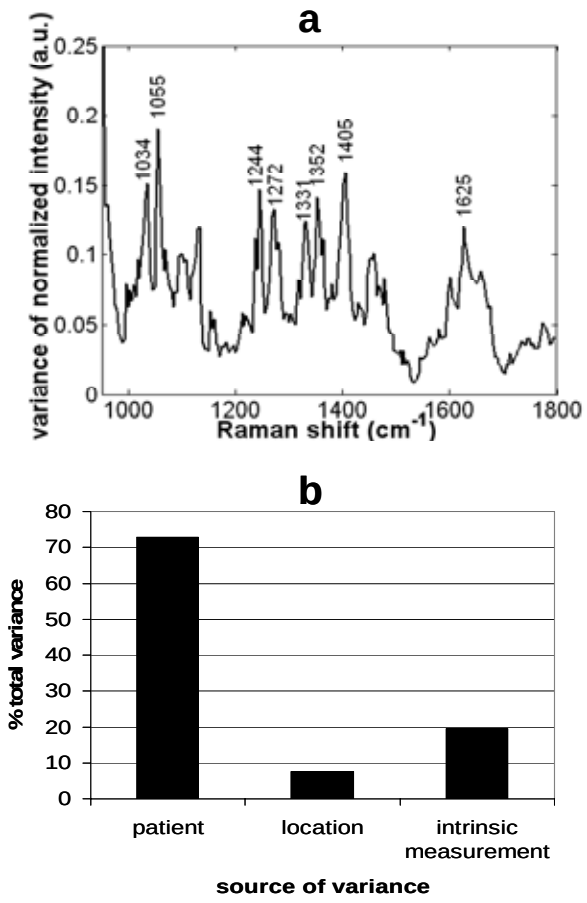

Figure 4.5: Components of variance.
A total of 180 spectra of normal ectocervix from 29 patients were used for this analysis. (a) total variance at each wavenumber, (b) percent of total variance due to interpatient (between patients), intrapatient (between locations), and intrinsic measurement variability.

Although the variance curve does not approximate the spectral curve (figure 4.1), the spikes in variance occur mainly at the sites of major spectral peaks. Figure 4.5b shows the relative contribution of variance from different sources as calculated by Henderson's method [18]. The largest source of total variance is due to interpatient variability. The intrapatient, or variability due to different locations within a patient, is neglible (7%) and contributes less to the overall variability than the error intrinsic to the measurement process (20%).

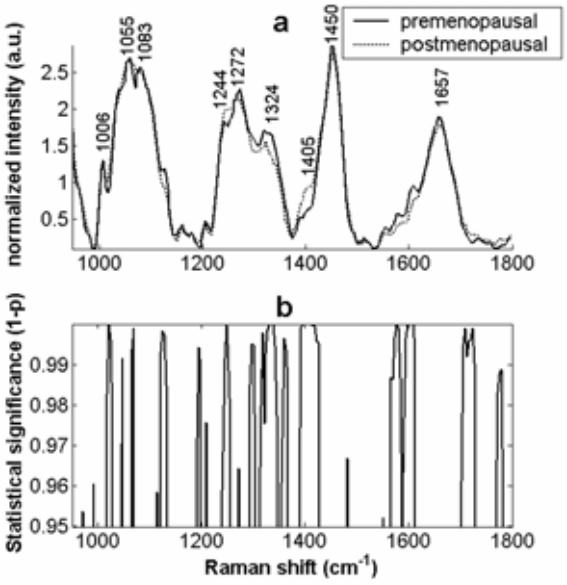

Figure 4.6: Comparison of Spectra from Premenopausal and Postmenopausal Patients.
(a) overlay of mean spectra of normal ectocervix from premenopausal (47 spectra from 17 patients) and postmenopausal (n=43 spectra from 18 patients) (b) results from t-test performed at each wavenumber.

Figures 4.6-4.8 examine three potential sources of interpatient variability. In figure 4.6, the spectra from normal ectocervix are stratified according to menopausal status, and the mean spectra from pre-menopausal and post-menopausal women are compared. The peak at 1324 cm^{-1}, the shoulders at 1100 and 1405 cm^{-1}, and the region around 1600 cm^{-1} are all areas of visual spectral distinction (figure 4.6a) and statistically significant differences (figure 4.6b).

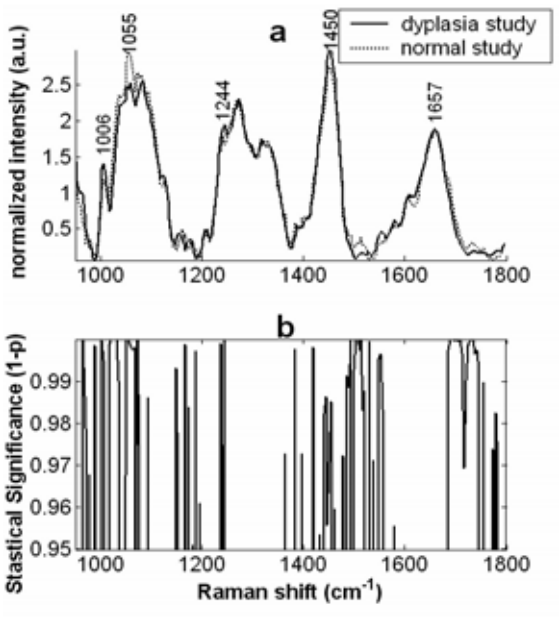

Figure 4.7: Comparison of Normal Ectocervix Spectra from Normal Patients and Dysplasia Patients.
(a) mean spectra from normal ectocervix from the normal study (n=22 spectra from 11 patients) and dysplasia study (n=25 spectra from 15 patients) (note: only spectra from pre-menopausal patients were used). (b) results from t-test performed at each wavenumber.

The mean normal ectocervix spectrum from pre-menopausal patients with normal cervices is compared with that from those with cervical dysplasia in figure 4.7. The analysis was limited to pre-menopausal patients as to not confuse actual spectral differences with those due to menopausal status. Although, histologically, the normal ectocervix from normal and dysplasia patients appear identical, there are spectroscopic differences between the two different sources of normal. The most notable visual distinctions appear in the peaks at 1006, 1055, 1244 and 1450 cm^{-1}. A closer look at the plot of the significance levels in figure 8b shows that there are several regions of statistically significant differences in addition to the ones indicated above.

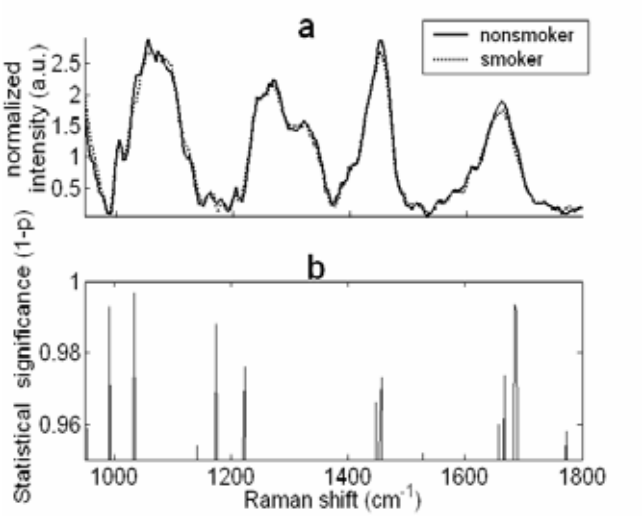

Figure 4.8: Comparison of Spectra from Smokers and Non-smokers.
(a) overlay of mean spectra of normal ectocervix from normal smoking (n=10 spectra from 5 patients) and nonsmoker (n=33 spectra from 11 patients) (b) results from t-test performed at each wavenumber.

A spectral comparison of normal ectocervix spectra from smokers is compared with non-smokers in figure 4.8. The only visual differences in major peaks between the mean spectra are seen at the 1450 and 1655 cm^{-1} peak, neither of which is significant at $p<0.01$. The only significant difference is seen at the shoulder ~1030 cm^{-1}.

4.5 Discussion

Raman spectroscopy has the potential to provide real time, non-invasive diagnosis, yet non-trivial amount of spectral variation exists between different spectral measurements of the same tissue pathology. In an attempt to better understand the origins and relative contributions of the sources of this variation, this study analyzed and quantified the different sources of variance within spectra of normal ectocervix.

Analysis of the signal to noise ratio of spectra collected using different integration times indicated that 5 seconds would be the optimal measurement time for this and future studies in the cervix using the current Raman spectroscopy system. The signal to noise ratio for spectra collected using a 5 second signal collection time is comparable to the signal to noise ratio for 8 or 15 second measurements (figure 4.2) and has the advantage of minimizing any probe movement during the signal collection that could lead to an increase in measurement error. Although the signal to noise ratio might be slightly higher with the 8 or 15 second signal

collection time, user error becomes more of a concern as the inability to hold the fiber optic probe steady for that length of time increases the likelihood of variability due to extraneous movement of the probe on the surface of the tissue.

The acetic acid analysis identified 2 areas within the spectra that show a visual spectra difference before and after the application of acetic acid (figure 4.3). Although these differences are not significant, they do give insight into the changes that occur after the application of acetic acid. These two peaks do show statistical differences between normal ectocervix and high grade dysplasia as well as normal ectocervix and squamous metaplasia (see chapter 3). The application of acetic acid enhances these changes, and thus it would seem advantageous to apply the acetic acid for the detection of dysplasia, both for the visual and spectral distinctions which are enhanced when it is applies. These results are comparable to an analogous study of the effect of acetic acid in fluorescence spectroscopy of the cervix [14]. Figure 4.3c indicates that there are no time dependent changes as a result of the acetic acid, so although acetic acid should be applied prior to spectral measurement for optimal results, it does not seem necessary to standardize the time after the application for taking spectral measurements.

The analysis of the components of variance revealed that the main source of variance between spectra of normal ectocervix lies in patient to patient variation (figure 4.4). Although this finding was expected, the corresponding finding that intrapatient variation (or the variance from one location of normal ectocervix to another in the same patient) is responsible for only 7% of the total variance, less than the intrinsic measurement error, was not expected. This finding indicates that there is little variation in normal tissue biology within a given patient, and thus there is no need to make multiple measurements of normal in a given patient, as the spectra should not be significantly different between different locations. This finding is corroborated by the results from figure 4.5. Figure 4.5 shows that there is no significant difference between repeat measurements keeping the probe in place versus removing and replacing the probe. This indicates that even if the probe is replaced in a slightly different location than the original placement, the spectra will not vary. This is likely due to the fact that the tissue is homogenous, and thus the slight change in location does not affect the spectra. However, this finding cannot be extrapolated to measurements of dysplastic areas. Since dysplastic lesions are by nature heterogeneous, one area of dysplasia may not have the same spectral characteristics as the adjacent area. Thus, probe placement becomes a larger source of variability for measurements of

dysplasia, and multiple measurements would be necessary to ensure an adequate representation of disease status. The spectral variation due to intrinsic measurement error accounts for less than 2% of the overall signal intensity. Several potential sources for this error are possible including motion artifact, throughput variations, and variation due to intrinsic to the interaction of tissue biology and Raman spectroscopy.

Since interpatient variation is the largest source of total variance, it is likely that demographic differences contribute to the overall observed spectral variation. Menopausal status has been shown to a large source of spectral variance in studies of cervical dysplasia using fluorescence spectroscopy [19]. The spectral differences seen between pre-menopausal and post-menopausal patients in figure 4.6 are likely a major source of the quantified interpatient variance. Most notably, there is a significant difference in the intensity at 1324 cm^{-1}. Since this peak has also been shown in a concurrent study to be one of the key spectral differences in distinguishing cervical dysplasia from normal ectocervix, it is likely that future studies will need to stratify the data according to menopausal status and develop separate classification algorithms specific to menopausal status (chapter 3). Due to the patient population (many patients had continuous menstrual bleeding or spotting due to uterine malignancy) and lack of sufficient numbers of pre-menopausal patients, we were unable to stratify the data according to cycle status in this analysis. Yet, future studies should carefully control or stratify data according to menstrual cycle status to fully understand the contribution of hormonal status to spectral variation.

One of the unexpected findings of the study was the significant differences seen between spectra of normal ectocervix measured from patients with normal cervices as compared with spectra of normal ectocervix measured from patients with cervical dysplasia (figure 4.7). This finding indicates that Raman spectroscopy may be able to detect differences in tissue biology which are not evident during histological examination. It also raises the question of whether Raman spectroscopy is detecting changes that are occurring at the subcellular level in normal tissue in the context of dysplasia. Such changes, termed malignancy associated changes, have been documented using high resolution image analysis of cervical smear specimens [20]. In the aforementioned study, the observed malignancy associated changes increased in intensity with the increase in grade of dysplasia. Whether or not the spectral changes observed actually reflect the presence of malignancy associated changes remains unknown and should be the subject of future research.

Overall, the analysis of variance presented in this paper aids in the understanding of the sources of the variation in the Raman spectra of tissue and how this variation affects the ability of Raman spectroscopy to discriminate between different pathologies. Since the majority of the spectral variance from a single pathology emanates from patient to patient variability and not from variability within a given patient, clinical studies can be more efficiently designed to eliminate unnecessary repeat measurements. Similarly, since menopausal status contributes significantly to the interpatient variance, clinical trials and data analysis can now be modified according to menopausal status; this should improve the accuracy of discrimination algorithms that distinguish between normal and dysplasia spectra. Results from the acetic acid analysis indicate that the application of acetic acid enhances some of the spectral changes between normal and dysplasia spectra, and thus it would be advisable to always apply acetic acid before spectral measurement, though regulation of the time after application is not critical. Other differences such as those between normal spectra taken from patients with normal cervices and those with dysplasia will require more research to fully understand how this will affect the ability of Raman spectroscopy to perform diagnosis.

4.6 References

1. Ramanujam, N., Fluorescence spectroscopy of neoplastic and non-neoplastic tissues. *Neoplasia*, 2000. **2**(1-2): p. 89-117.

2. Hanlon, E.B., *et al*., Prospects for in vivo Raman spectroscopy**.** *Phys Med Biol*, 2000. **45**(2): p. R1-59.

3. Molckovsky, A., *et al.*, Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. *Gastrointest Endosc*, 2003. **57**(3): p. 396-402.

4. Stone, N., *et al.*, Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. *Laryngoscope*, 2000. **110**(10 Pt 1): p. 1756-63.

5. Shafer-Peltier, K., *et al.*, Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 552-563.

6. Nijssen, A., *et al.*, Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy. *J Invest Dermatol*, 2002. **119**(1): p. 64-9.

7. Mahadevan-Jansen, A., *et al.*, Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers. *Photochem Photobiol*, 1998. **68**(1): p. 123-32.

8. Huang, Z., *et al.*, Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int J Cancer*, 2003. **107**(6): p. 1047-52.

9. Smith, J., *et al.*, Raman spectral mapping in the assessment of axillary lymph nodes in breast cancer. *Technol Cancer Res Treat*, 2003. **2**(4): p. 327-32.

10. Crow, P., *et al.*, The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*. *Br J Cancer*, 2003. **89**(1): p. 106-8.

11. Koljenovic, S., *et al.*, Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy**.** *Lab Invest*, 2002. **82**(10): p. 1265-77.

12. Robichaux, A., *et al. In vivo* Detection of Cervical Dysplasia with Near Infrared Raman Spectroscopy. in SPIE Photonics West. 2002. San Jose, CA: SPIE Press.

13. Knudsen, L., *et al.*, Natural variations and reproducibility of *in vivo* near-infrared Fourier transform Raman spectroscopy of normal human skin. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 574-9.

14. Agrawal, A., *et al.*, Fluorescence spectroscopy of the cervix: influence of acetic acid, cervical mucus, and vaginal medications. *Lasers Surg Med*, 1999. **25**(3): p. 237-49.

15. Brookner, C.K., *et al*., *Cervical* fluorescence of normal women*. Lasers Surg Med*, 1999. **24**(1): p. 29-37.

16. Cox, D.D., *et al.*, Detecting the signal of the menstrual cycle in fluorescence spectroscopy of the cervix. *Applied Spectroscopy*, 2003. **57**(1): p. 67-72.

17. Drezek, R.A., *et al.*, Laser scanning confocal microscopy of cervical tissue before and after application of acetic acid. *Am J Obstet Gynecol*, 2000. **182**(5): p. 1135-9.

18. Neter J., *et al*., *Applied Linear Statistical Models*. 4th ed. 1996, Chicago, IL: Irwin.

19. Brookner, C.K., *et al.*, Effects of biographical variables on cervical fluorescence *emission spectra. Journal of Biomedical Optics*, 2003. **8**(3): p. 479-83.

20. Guillaud, M., *et al.*, Malignancy associated changes in cervical smears: systematic changes in cytometric features with the grade of dysplasia. *Anal Cell Pathol*, 1995. **9**(3): p. 191-204.

21. Utzinger, U., *et al.*, Near-Infrared Raman Spectroscopy for *In vivo* Detection of Cervical Precancers. *Applied Spectroscopy*, 2001. **55**(8): p. 955-9.

CHAPTER V

ORGANOTYPIC RAFT CULTURES AS AN EFFECTIVE *IN VITRO* TOOL FOR UNDERSTANDING RAMAN SPECTRAL ANALYSIS OF TISSUE

Amy Robichaux Viehoever[1], Douglas Anderson[1], Duco Jansen[1], and Anita Mahadevan-Jansen[1]

[1]Department of Biomedical Engineering
Vanderbilt University, Nashville, TN 37235

Portions of this manuscript were published in
Photochemistry and Photobiology (2003) Volume 78, Issue 5, p. 517-24.

5.1 Abstract

There is a growing body of evidence showing that optical spectroscopy has the potential to be a useful *in vivo* diagnostic tool. Yet, so far there is no definitive cellular and biochemical understanding for the differences seen in the spectra from different tissue categories and disease states. In this study, we examine the use of organotypic raft cultures as an *in vitro* model of *in vivo* tissue conditions in an attempt to overcome some of the limitations of previously used methods. Organotypic raft cultures resembling normal and dysplastic epithelial cervical tissue were constructed and grown at an air–liquid interface for 2 weeks. Raman spectra of normal as well as dysplastic raft cultures were measured and compared with *in vivo* spectra from the corresponding tissue type. Histologic comparisons ensured that the raft cultures had similar structure and morphology to the corresponding intact tissue types. Raman spectra were also acquired from different layers of tissue. Spectral comparisons show that the Raman spectra of the raft cultures are similar to the spectra acquired from the cervix *in vivo* for both normal and dysplastic tissues. These results show that organotypic raft cultures are an effective and useful tool for the cellular and biochemical analysis of tissue spectroscopy.

5.2 Introduction

Optical spectroscopic methods are increasingly being applied to detect tissue pathology due to their ability to provide real-time, non-intrusive, and automated information. Numerous *in vitro* and *in vivo* studies have documented the potential of fluorescence, diffuse reflectance, near-infrared absorption, and Raman spectroscopy to perform automated tissue diagnosis alone or in combination [1]. Among these, Raman spectroscopy is particularly well suited to probe the biochemical changes seen with carcinogenesis since Raman spectroscopy examines many specific chemical bonds present in tissue rather than just a few and provides a highly specific spectral fingerprint. Raman spectroscopy thus far has shown to be a promising diagnostic tool for discrimination of precancerous and cancerous lesions in *in vitro* studies of the gastrointestinal tract [2], larynx [3], breast [4], skin [5], cervix [6] and *in vivo* studies of cervix [7], esophagus [8], colon[9], skin [10], and rat carcinogenesis models [11].

Although these initial studies indicate that optical spectroscopy has the potential to be a useful *in vivo* diagnostic tool, there is limited understanding of how the variability of different tissue components and conditions affects the spectra of tissue. Detailed knowledge of the

morphologic and biochemical basis of the observed spectral features has several foreseeable benefits. First, an understanding of the interactions of tissue biology with spectral changes could influence the design of the fiber optic probe used to deliver and collect the light from the tissue [12]. For example, if such studies were to indicate that diagnostically significant spectral information emanates from a particular tissue region, a fiber optic probe could be designed to maximally collect light only from that region. Second, such a fundamental understanding could lead to more informed *in vitro* and *in vivo* clinical study design. Third, for optical spectroscopy to be useful clinically, automated diagnostic algorithms capable of distinguishing spectra from various normal and non-normal tissue categories must be designed and implemented. Thus far, unbiased statistical algorithms using methods such as linear discrimination based on Bayes' theorem [6] and artificial neural networks [13] have been used to design these algorithms. Although such statistical methods could prove to be sufficient for automated diagnosis once large data sets are collected, rational design of diagnostic algorithms requires an understanding of the biochemical and tissue morphology changes that underlie the spectral changes. This understanding could in turn prove to be critical in the development of more robust algorithms.

In recent years, the primary tool used to investigate the molecular basis for tissue spectra has been frozen tissue microspectroscopy. Several tissue microspectroscopy studies have elucidated the spectral features of various tissue components such as collagen, fat, and epithelial cells and the contributions of these components to tissue spectra using mathematical models [10, 14, 15]. Yet, none of these methods are capable of determining the basis of the variation in the spectral signal due to intrinsic tissue components, in unknown quantities in tissue, nor have they proven to be very useful in determining how these components affect the ability for spectroscopy to perform diagnosis. In this study, we examined the use of organotypic raft cultures as a controlled *in vitro* model of *in vivo* tissue conditions in an attempt to understand how alterations in tissue biology affect tissue spectroscopy.

Organotypic cultures are multilayer, three dimensional cultures designed to reproduce the *in vivo* structure and function of tissue. Organotypic cultures are typically used as a model system when differentiation of cells and cell-cell interaction within a tissue structure is critical to the understanding of the underlying process [16]. Raft cultures represent a specific type of organotypic cultures that reconstruct epithelial tissue types that consist of an avascular epithelium that attaches via a basement membrane to a supporting connective tissue stroma. Raft

cultures have enabled investigation of problems involving cell-cell interactions [17], epidermis-dermis interactions [18], gene therapy [19, 20], viral pathogenesis [21], immunologic regulation [22], and transformation of normal epithelium to dysplastic epithelium [23].

Although organotypic cultures have been used for many years as a model system in many different fields of study, only recently have organotypic cultures been proposed as a model system for tissue spectroscopy [24]. To our knowledge, no studies to date have demonstrated the utility of organotypic raft cultures for analysis of spectroscopic data. We propose that organotypic cultures are a useful model system that could provide insight into the ability and limitations of spectroscopy to provide automated diagnosis; the ability to control and manipulate tissue components and culture conditions will enable experiments that yield information complementary to frozen tissue microspectroscopy studies. By mimicking an *in vivo* environment, the spectroscopic features of tissue components can be examined in their natural environment. Concentrations and types of tissue components can be manipulated, and the resulting spectra can be analyzed to determine the spectral contributions of individual tissue components.

Thus, this study aims to assess the potential of organotypic raft cultures as a novel model system for understanding the biochemical and cellular basis for spectral features and thus the diagnostic potential of tissue Raman spectroscopy. The goal of this paper is to present a robust method for a model system that is useful in tissue spectroscopic analysis. To accomplish this aim, the ability of organotypic raft cultures to accurately model human cervical tissue for Raman spectroscopic analysis was tested. Raft cultures were constructed to resemble normal and dysplastic human cervix. The raft cultures were compared both morphologically (via histology) and spectroscopically (via Raman spectroscopy) to *in vivo* tissue data from an ongoing clinical study using Raman spectroscopy for the detection of cervical precancers [25]. Two different types of squamous cell carcinoma cell lines (see table 1) were used to generate the dysplastic cultures, thus allowing investigation into the spectral variation associated with varying epithelial cells. Additionally, the epithelium was separated from the stromal layer in the raft cultures such that the Raman spectra could be measured from the individual layers to understand the contributions of the two main layers of cervix. These results were then compared to Raman spectra measured from the epithelial and stromal layers separated from intact normal cervical tissue samples *in vitro*. Additional studies examined the effect of the epithelial thickness on the

overall Raman spectra. Thus, the utility of raft cultures as a suitable model system for the preliminary understanding of the effect of tissue morphology and biochemistry on optical signatures was demonstrated.

5.3 Materials And Methods

5.1.1 Construction of Rafts

Four different cell lines were used in the construction of the organotypic raft cultures. Normal human neonatal dermal fibroblasts (NHDF, Cambrex Co., East Rutherford, NJ) were used in the formation of the stromal layer for all raft cultures. The epithelial component was formed using one of three different cells lines, either Normal Human Neonatal Foreskin Keratinocytes (NHK, Cambrex Co., East Rutherford, NJ), Epidermoid Carcinoma from skin (A-431, ATCC, Manassas, VA), or Squamous Cell Carcinoma cells from the cervix (SiHa, ATCC, Manassas, VA) depending on the type of tissue and disease state (normal vs. dysplasia) desired for study. Table 1 diagrams the cell lines and the corresponding growth media used.

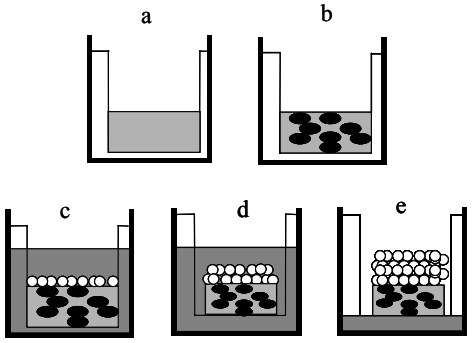

Figure 5.1: Construction of an organotypic culture
(a) Thick layer of collagen, (b) Fibroblast cells mixed with collagen matrix, (c) Epithelial cells seeded on top of matrix with culture media (d) Fibroblasts cause contraction of collagen matrix and (e) Differentiation of epithelial cells after 10-15 days growth at air-liquid interface [17].

Figure 5.1 illustrates the steps taken in the formation of the organotypic raft cultures. The stromal equivalent is formed first by mixing 1 part reconstitution buffer (2.2% $NaHCO_3$, 0.05 M NaOH, 200mM HEPES buffer), 2 parts five-fold concentration High Glucose DMEM (Sigma,

St. Louis, MO), 6.5 parts collagen solution (3.5-4.3 mg/ml rat tail collagen type I in 0.02 N acetic acid from BD Biosciences, Bedford, MA). NaOH (1 N) was then added to the solution until a neutral pH was obtained (indicated by phenol red). To this solution, 0.5 part of cell solution containing NHDF cells suspended in the growth medium (DMEM +10% FBS) at a concentration of 8 X 10^6 cells/ml is added. Two milliliters of this collagen-fibroblast solution is then poured into individual wells of a 24 well plate (Costar, Corning, NY) and allowed to solidify overnight in an atmosphere of 37°C, 5% CO_2. For example, to make the stromal layers for four rafts, 1 ml reconstitution buffer, 2ml 5X DMEM, 6.5 ml of collagen solution, approximately 25 µl 1 N NaOH, and 0.5 ml DMEM containing 4 X 10^6 NHDF cells are mixed and poured into 4 wells. Note: the addition of collagen to the solution causes a contraction of the total volume such that even though 10 ml of total solution is added, the final volume is only 8 ml.

Following the formation of the stromal equivalent, one of three types of epithelial cells (3 X 10^5 cells suspended in 1ml of Keratinocyte Growth Media (KGM-2, Cambrex Co., East Rutherford, NJ) is added to the top of each raft. The epithelial cells are allowed to attach during an overnight incubation, and the following day, the rafts are gently separated from the side walls of the well plate using a sterile spatula. This allows the rafts to contract and form a more dense, tissue-like consistency. After overnight incubation, the rafts are placed onto a 25 mm wire mesh (Sigma, St. Louis, MO) with the epithelial cells facing up. Raft Media (see table 5.1) is then added such that the media reaches the level of the wire mesh but no part of the raft is submerged. The raft cultures are incubated for 10-12 days in an atmosphere of 37°C, 5% CO_2 with a change of media every other day.

Table 5.1: Cell lines used in construction of raft cultures

Cell Line	Cell Line Description	Growth Media
NHDF (Clonetics)	Normal Human Neonatal Dermal Fibroblast	High Glucose Dulbecco's Modified Eagle's Medium (DMEM, Sigma) + 10% Fetal Bovine Serum (FBS, Clonetics)
NHK (Clonetics)	Normal Human Neonatal Dermal Keratinocyte	Keratinocyte Serum Free Media (Invitrogen)
A-431 (ATCC)	Human Epidermoid (Squamous Cell) Carcinoma	High Glucose DMEM +10% FBS
SiHa (ATCC)	Human Cervical Squamous Cell Carcinoma	Modified Eagle's Media (MEM, Sigma) + 10% FBS
Raft Cultures		Raft Media (1 part Keratinocyte Growth Media-2 (Clonetics) 1 part High Glucose DMEM +10% FBS)

Raft cultures grown from 6 different experiments were used for the analysis presented here. An experiment is defined as the formation of a set of raft cultures following the same protocol but utilizing unique flasks of cells, new batches of collagen solution, and begun on different days. In each experiment, an average of 3 distinct rafts of each epithelial cell type (A431: Epidermoid Carcinoma cells, SiHa: Squamous Cell Cervical Carcinoma, NHK: Normal Human Neonatal Dermal Keratinocytes), were grown resulting in an average of 9 raft cultures in each experiment. Raman spectra were acquired from each raft. Spectra from the rafts of each epithelial cell type were averaged from each experiment, thus each experiment was subsequently treated as n=1 for statistical analysis.

5.1.2 Measurement of Raman Spectra

Raman spectra were collected from all raft cultures with a macroscopic Raman spectroscopy system (figure 2) utilizing a fiber optic probe for light delivery and collection.

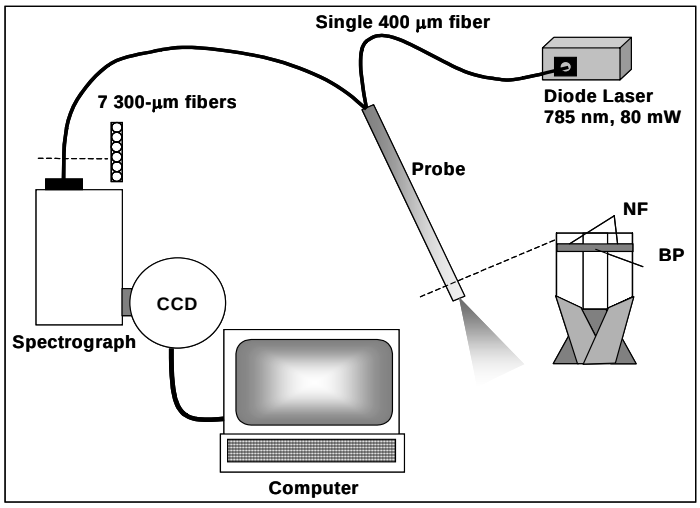

Figure 5.2: Raman spectroscopic system used for data collection.
Note: CCD: Charged Coupled Device, NF: notch filter at 785 nm, BP: band pass filter at 785 nm

The system consists of a 785 nm diode laser (Process Instruments, Inc.) set at 80 mW, fiber-optic probe (Visionex Inc.), a spectrograph (Kaiser Optical Systems, Inc), and a back-illuminated, deep-depletion, liquid nitrogen cooled CCD camera (Roper Scientific Inc.) all controlled with a laptop computer. The fiber-optic probe contains a single, central excitation fiber of 400 μm core diameter surrounded by seven 300 μm beam-steered collection fibers [26]. The collected light is then filtered with an inline notch filter at 785 nm within the probe itself to attenuate light at the laser wavelength. The light is then fed into the spectrograph where it is filtered again with a holographic notch filter at 785 nm and dispersed onto the CCD camera where the computer records the signal.

During spectral measurements from the raft cultures, the rafts remained on the metal grid but were removed from the petri-dish containing the media. The fiber-optic probe was placed in light contact with the raft, and Raman spectrum was recorded using an integration time of 15 seconds. Raman spectra were measured from two different sites on each raft and averaged for final analysis. The probe was rinsed with sterile normal saline between measurements of each raft. After spectral acquisition, one raft from each epithelial cell type was fixed in 10% formalin, sectioned, and stained with hematoxylin & eosin (H & E) for histologic examination. The histology of the raft cultures was compared with typical histological sections of corresponding tissue type and disease category. Raman spectra from the raft cultures were compared to Raman

spectra acquired *in vivo* from the cervix of human patients undergoing excisional procedures as part of an ongoing pre-clinical study assessing the potential of Raman spectroscopy for cervical dysplasia detection [25].

5.1.3 Epithelium and Stroma Measurements: Raft Cultures

After Raman spectra were collected from each intact raft, the epithelial layer was removed from the underlying stroma using either forceps (NHK rafts) or a scapel (A-431 and SiHa rafts) from the remaining two rafts from each epithelial cell type (not including the raft sent for histology). The thin layer of epithelium was placed on the metal grid alongside the remaining stromal layer. Raman spectra of the epithelial and stromal layers were collected separately using the same system (figure 5.2) and protocol used to collect spectra from the intact raft cultures.

5.1.4 Epithelium and Stroma Measurements: Tissue

To further validate raft cultures as a suitable model system, Raman spectra from the epithelial and stromal layers of the raft cultures were compared with spectra from separated epithelial and stromal layers of cervical tissues. Tissue samples of the normal ectocervix were obtained from hysterectomy specimens as approved by the Vanderbilt IRB. The samples were kept frozen at -80°C until analysis. Unlike in the raft cultures where there is no basement membrane attaching the epithelial layer to the stromal layer, the basement membrane present in the tissue necessitates a chemical separation process to break the bonds connecting the epithelium and stromal layers. Thermolysin, a zinc endopeptidase, cleaves bonds found in the lamina lucida (one component of the basement membrane), thereby separating the epithelium from the connective tissue at the level of basement membrane [17].

Tissue specimens were thawed in PBS (phosphate buffered saline) at room temperature at the time of study. Raman spectra were measured from each intact specimen. The ectocervix specimens were then submerged in a 750 μg/ml solution of thermolysin in 10 mM HEPES, 1mM $CaCl_2$, pH 7.2 and incubated at 37°C for 2-3 hours [17]. After incubation, a thin film of epithelium was removed from the top of the tissue specimen and placed on a sheet of aluminum foil or on the metal grids used for the raft cultures alongside the remaining stromal layer. Raman spectra were measured separately from the thin sheet of epithelium and the remaining stroma. All Raman measurements were taken with the previously described system (figure 5.2) and

protocol. Raman spectra acquired from the epithelia and stroma of intact tissues were compared to the corresponding spectra from raft cultures. Additionally the spectra were also compared to the intact tissue and raft culture spectra to assess the effect of each layer on the integrated tissue signature.

5.1.5 Thickness Analysis

To determine the effect of epithelial thickness on the overall Raman spectra, raft cultures were constructed using NHK and SiHa epithelial cells as described above. Five experiments were conducted; in each experiment, three raft cultures of each epithelial cell type were grown. Raman spectra were collected from the raft cultures on days 4, 8 and 14 after the raft cultures were raised to an air-liquid interface. On each day of spectral measurement, one raft culture per epithelial cell type was fixed in 10% formalin, sectioned, and stained with hematoxylin & eosin (H & E) for histologic examination following the collection of Raman spectra.

5.1.6 Data Pre-Processing and Extraction

Prior to each experiment, a spectral calibration of the system was performed using a neon-argon lamp and napthalene standard to correct for system wavenumber, laser excitation, and throughput variations. For each Raman spectrum measured, the signal from the CCD camera was binned along the vertical (intensity) axis to create a single spectrum per measurement site. Only signal between 950 – 1850 cm^{-1} was considered for further analyses due to the interference of silica bands below 950 cm^{-1} from the fiber optics probe. The spectrum was then binned along the horizontal (wavenumber) axis such that each data point represents the intensity from 3.5 cm^{-1}. The spectrum was then filtered using a 2^{nd} order Savitzky-Golay filter (window =17.5 cm^{-1}) for noise smoothing [27]. Fluorescence subtraction was accomplished using a modified polynomial fitting method in which a 5^{th} order polynomial is fit to the fluorescence baseline. Following data processing, each spectrum was normalized to its mean spectral intensity across all Raman bands to account for overall intensity variability. These normalized spectra were used for further comparison and analysis.

5.4 Results

Histopathological examination on a representative raft from each of the epithelial cell types in each experiment confirmed the pathology. Rafts from all experiments showed consistent histopathologic patterns. H&E stained sections from representative raft cultures were compared to that from tissue biopsies of the corresponding pathologic category. Figures 5.3a, b compare histologic sections from an area of normal cervix to the histology from the raft culture made with NHK cells. The raft culture shows all stages of differentiation seen in the normal cervix including an undifferentiated basal layer overlying the stroma, a parabasal layer of cells above the basal layer, an intermediate layer above the parabasal layer, and the terminally differentiated top layer known as the superficial layer. The layer of keratin seen above the normal raft culture arises due to the fact that dermal keratinocytes were used in these cultures in lieu of normal cervical keratinocytes. This layer is not present in the corresponding normal cervix. Figures 5.3e & f compare a histologic section from an area of high grade dysplasia with a raft culture constructed with SiHa cells. The SiHa raft epithelium shows many of the same morphological characteristics seen in high grade cervical dysplasia. For example, there is no differentiation or stratification of the epithelium; instead, cells appear undifferentiated throughout the epithelium and show a lack of organization within the epithelium, with significant nuclear atypia. The main differences between the raft section and the corresponding biopsy are seen primarily in the stromal layer. The stromal layer from intact tissue contains connective fibers (such as collagen, elastin) as well as fibroblasts, inflammatory cells and capillaries. Since the raft cultures were constructed only with collagen, concentrated media, and fibroblasts, the cultures lack the inflammatory cells and capillaries. Thus the stromal layer of the raft section appears less crowded than the stromal layer of the tissue.

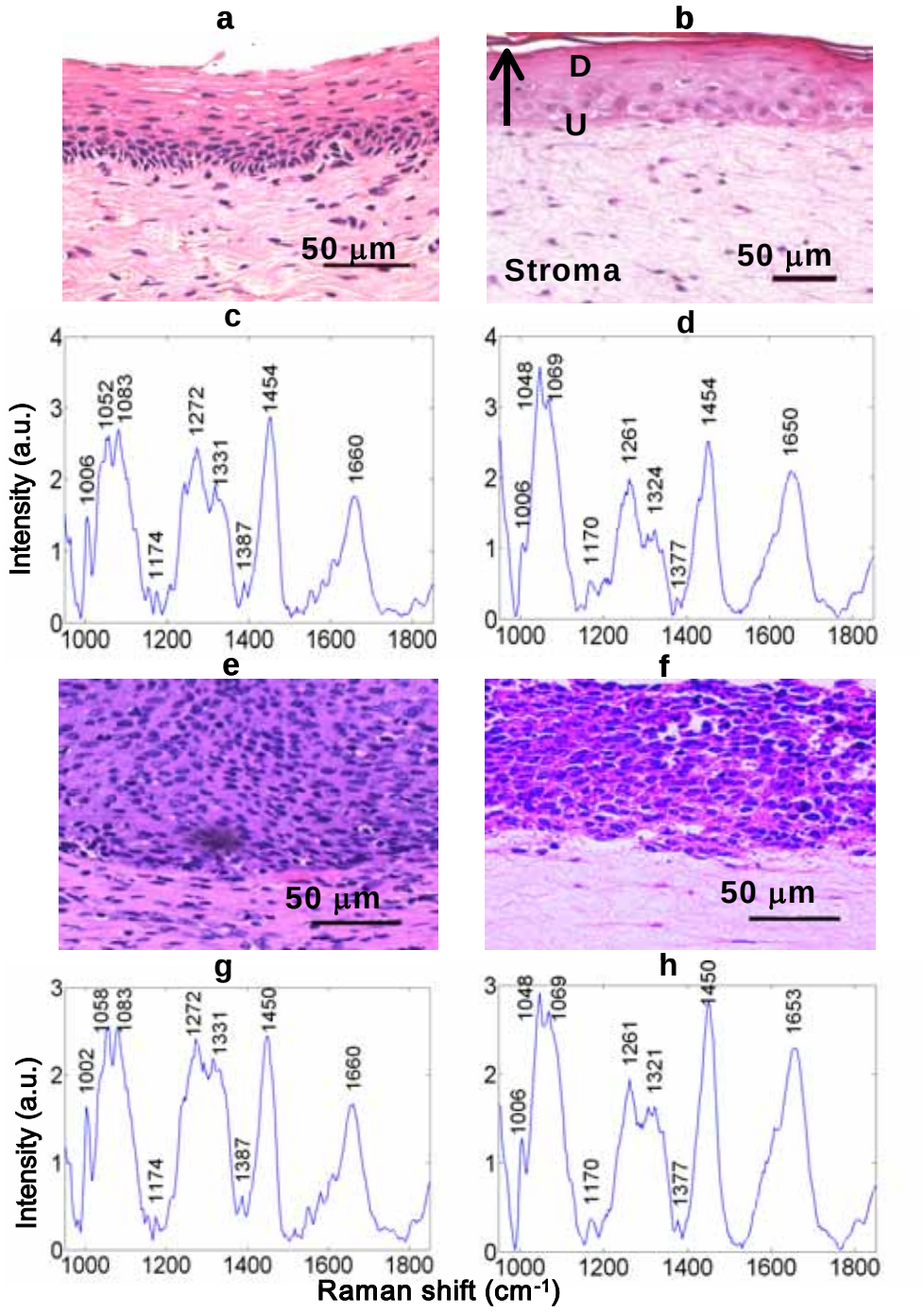

Figure 5.3: *In Vivo* Raft Culture Histology and Spectral Comparison.
Histology of (a) normal squamous epithelium of the cervix and (b) a normal raft culture using human keratinocytes (NHK). Note: the arrow represents the direction of keratinocyte differentiation; beginning in the basal layer as undifferentiated cells (U) and moving upwards to become fully differentiated cells (D). Average Raman spectra of (c) normal cervical tissue using 49 measurements from 24 patients *in vivo* and (d) 18 normal rafts from 6 different experiments. Histology of (e) high grade dysplasia of the cervix and (f) a dysplastic raft culture using cervical squamous cell carcinoma cells (SiHa). Average Raman spectra of (g) dysplastic lesions acquired *in vivo* using 12 measurements from 6 patients and (h) 18 SiHa rafts from 6 different experiments.

These morphological similarities and differences in the histology of the raft cultures described above are consistent with previous reports of raft cultures constructed with normal and cancerous cell lines [23]. The mean Raman spectra from the normal raft culture and normal *in vivo* cervix are shown in figures 5.3c & d. Direct comparison of the spectra shows similarities in the shape of the Raman bands and their relative intensities with some variability in the specific intensities of these bands. The corresponding mean Raman spectra from the SiHa raft and *in vivo* tissue dysplasia lesion show similar trends (figures 5.3g &h).

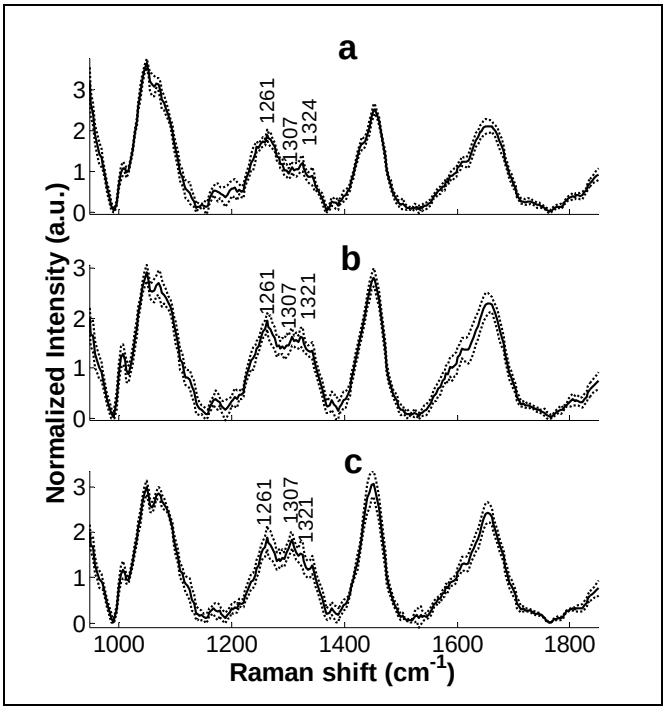

Figure 5.4. Mean and standard deviation spectra for raft cultures.
The mean (solid line) ± one standard deviation (dotted lines) of the Raman spectra from (a) normal (NHK cells) rafts, (b) dysplastic (SiHa cells) rafts and (c) dysplatic (A431 cells) rafts (n=6 experiments, 18 raft cultures in each category).

The Raman spectra of all the raft cultures from six different experiments were averaged by epithelial cell type. The mean (± one standard deviation) Raman spectra from 18 rafts from these six experiments are plotted for each of the three different epithelial types in figure 5.4. All spectra have been normalized to the mean spectral intensity for comparison. To quantify the amount of variation between different experiments, the standard deviation was calculated at each

point along the normalized spectrum. The mean of these standard deviation values (across the spectrum) was then calculated to determine a mean normalized standard deviation for each group of raft cultures. The spectra from raft cultures were observed to be very consistent across the entire fingerprint region for all experiments with a mean normalized standard deviation equaling 0.1326 (A431), 0.15071 (NHK), 0.14654 (SiHa).

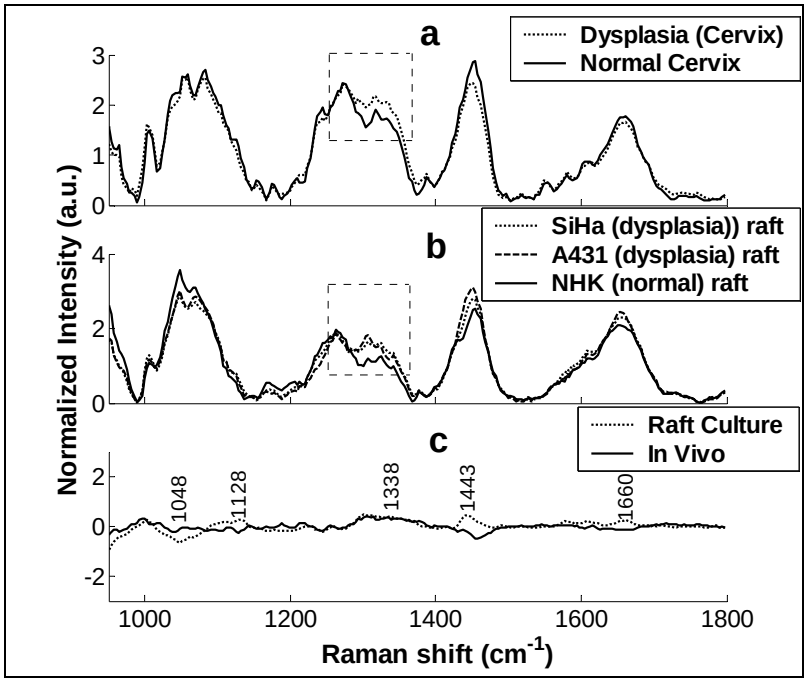

Figure 5.5: *In Vivo* and Raft Culture Dysplastic/Normal Spectral Comparisons.
Raman spectra from (a) normal and dysplastic areas of the cervix acquired *in vivo* [25] and (b) normal (NHK) and two types of dysplastic (SiHa, A431) raft cultures. (c) Difference spectra calculated between normal and dysplastic signals from raft cultures and *in vivo* tissues. The dotted box represents the major spectral differences observed.

Figure 5.5 compares the Raman spectra from normal and dysplastic samples of (a) intact tissue *in vivo* to (b) raft cultures. The spectral differences seen between the SiHa (dysplastic) cultures and NHK (normal) cultures parallel those seen between high grade dysplasia and normal ectocervix *in vivo*. The box enclosing the 1270 cm^{-1} and 1325 cm^{-1} peaks represents the area of greatest visual spectral distinction between normal and dysplastic tissues. The shoulder at 1325 cm^{-1} is noticeably higher in the dysplasia spectra (both raft culture and *in vivo*) when compared with the normal spectra, relative to the peak at 1270 cm^{-1}, which remains constant. To quantify

these spectral differences, ratios of peak intensities were calculated for all major peaks present in the Raman spectra (peaks used: 1058, 1086, 1270, 1325, 1454, 1655 cm^{-1}) with respect to each other. An unpaired Student's t-test was then performed to compare the peak ratios between normal and dysplastic samples. A subset of the ratios compared and their corresponding p-values using spectra from raft cultures as well as *in vivo* samples [25] is presented side-by-side in table 5.2.

Table 5.2: Comparison of peak ratios between normal and dysplastic spectra acquired from raft cultures and *in vivo* tissue.
The corresponding p-values are shown from t-tests comparing normal and dysplastic spectra from an ongoing clinical study, A431 and NHK raft cultures, and SiHa and NHK raft cultures, respectively. Statistically significant differences are shown in bold.

	In vivo	**Raft Culture**	
Peak Ratio (cm^{-1})	**Normal vs. Dysplasia**	**A431 vs. NHK**	**SiHa vs. NHK**
1086/1270	0.307	0.774	0.278
1086/1325	**0.007**	0.013	**.00004**
1086/1454	0.451	**0.00002**	**0.001**
1086/1655	0.957	0.043	0.046
1270/1325	0.014	**0.003**	**.00004**
1270/1454	0.110	0.017	0.241
1270/1655	0.563	0.175	0.592
1325/1454	0.024	0.556	0.005
1325/1655	0.103	0.341	**0.002**
1454/1655	0.299	0.258	0.391

As anticipated from figure 5, multiple peak ratios involving the band at 1325 cm^{-1} show statistically significant differences between normal and dysplasia tissues. Most notably, peak ratios at 1086/1325 and 1275/1325 cm^{-1} that show significant differences *in vivo* are observed to be significant in raft cultures as well. Only one peak ratio (1086/1454 cm^{-1}) is inconsistent between *in vivo* data and both sets of raft culture data. A comparison of the difference spectra calculated between normal and dysplastic spectra from raft cultures and *in vivo* tissues highlight areas of the spectra where raft culture results deviate from that observed *in vivo*. While the

majority of the fingerprint region of the spectra shows good correlation, particularly between 1250 cm^{-1} and 1400 cm^{-1}, spectral differences are inconsistent around 1048, 1128, 1445 and 1660 cm^{-1} between *in vivo* and raft culture spectra.

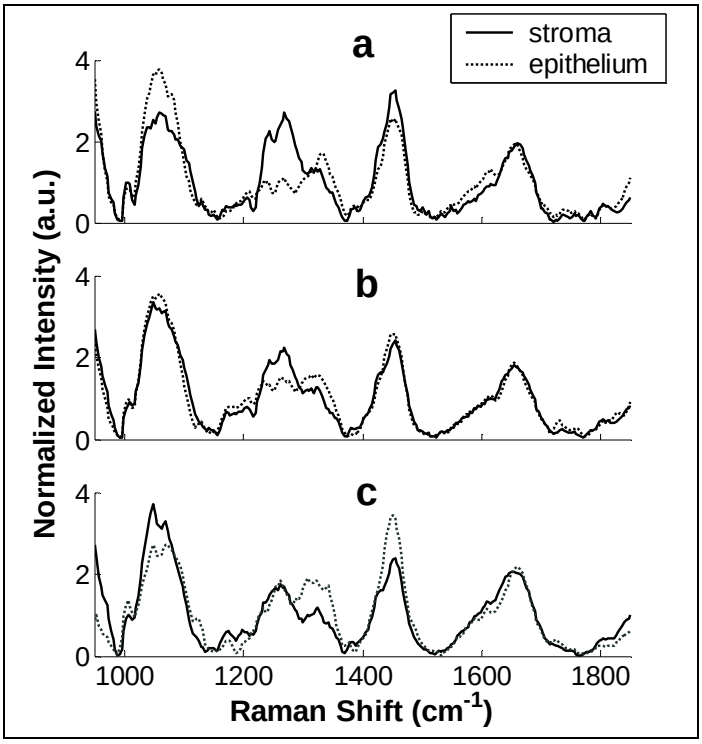

Figure 5.6: Raman spectra measured of epithelium and stroma
Epithelium and stroma separated from (a) a normal cervical biopsy (b) normal raft cultures and (c) dysplastic raft cultures.

Figure 5.6 shows the results from the Raman spectral measurements of the two main layers of tissue, the epithelium and stroma. The mean spectra from the epithelia and stroma of raft cultures measured separately in six experiments are shown for NHK raft (figure 5.6b) and SiHa raft (figure 5.6c). (The Raman spectra from the epithelia and stroma of A431 raft cultures in initial experiments were similar to those of the SiHa rafts and thus are not shown here). In addition, Raman spectra were also measured from separated epithelia and stroma of five normal cervical tissue samples taken from hysterectomy specimens (results shown in figure 5.6a). Comparison of spectra acquired from intact cervical tissue layers with those from the raft cultures indicates that raft cultures show similar variations between the epithelial and stromal

spectra as the spectra measured from the layers of intact tissue. Of note is the striking difference seen in the 1270 cm^{-1} and 1325 cm^{-1} peaks between the stromal and epithelial spectra. The stromal spectrum shows peaks at 1240 cm^{-1} and 1270 cm^{-1} similar to those seen in the *in vivo* data with relatively lower peak intensity at 1325 cm^{-1}. The epithelial spectrum, in contrast, shows a strong peak at 1325 cm^{-1}, and a much smaller peak at 1270 cm^{-1}. In comparing the epithelial spectrum from the SiHa rafts (figure 5.6c) with those from intact normal tissue (figure 5.6a) and the NHK rafts (figure 6c), the SiHa epithelial spectrum appears to have an even greater increase in the 1325 cm^{-1} peak as well as an increase in the 1454 cm^{-1} peak, a change not observed in the intact normal and NHK raft epithelial spectra.

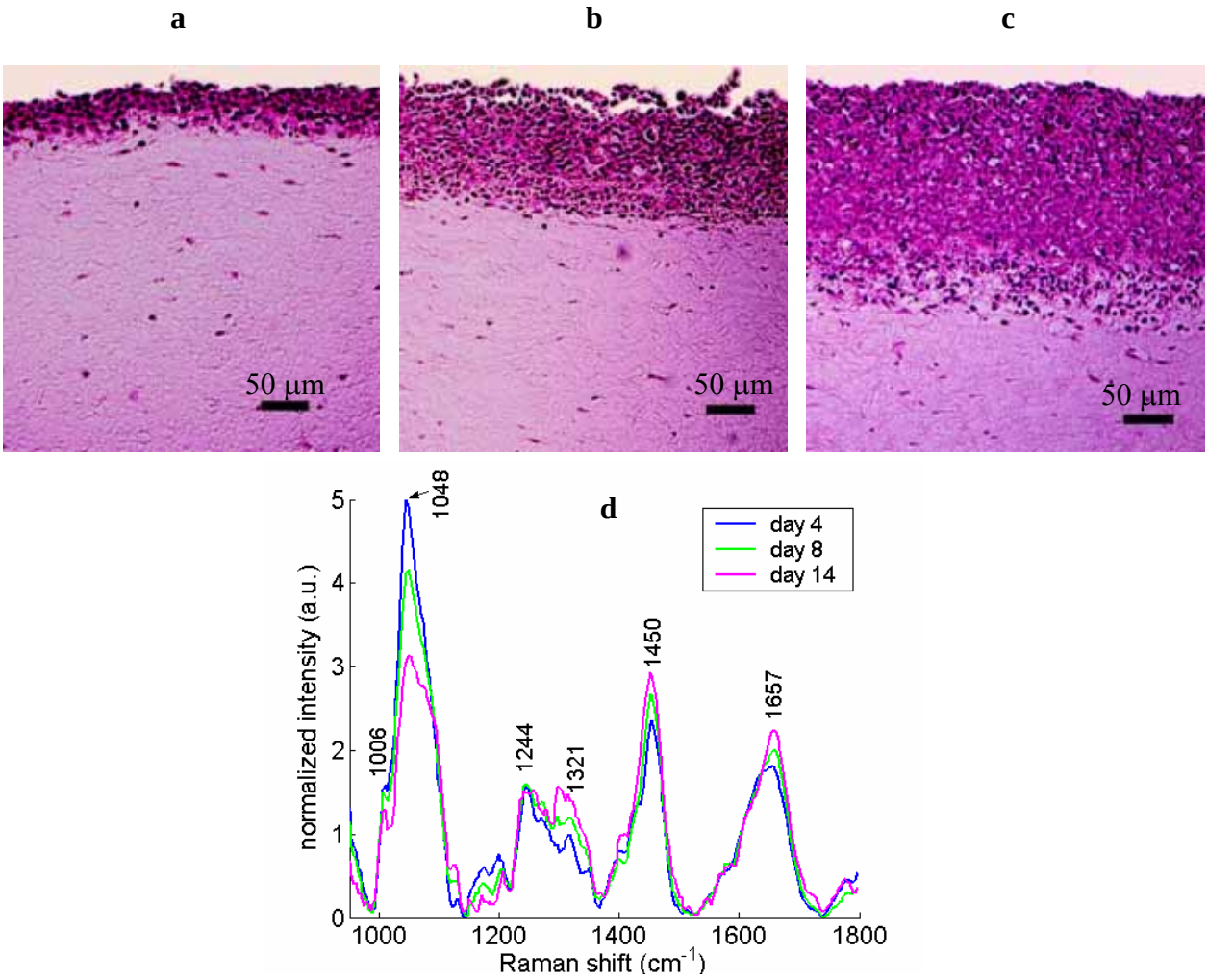

Figure 5.7: Effects of Epithelial Thickness on Raman spectra.
Histology from SiHa (dysplastic) raft cultures is shown for (a) day 4, (b) day 8, (c) day 14 after the raft cultures are raised to the air liquid interface (d) corresponding Raman spectra measured from raft cultures on days 4, 8, and 14.

The results of the experiments looking at the effect of epithelial thickness on the overall Raman spectra are shown in figure 5.7. Histology results from the raft cultures on days 4,8, and 14 are shown in progression in figure 5.7a, b, and c, respectively. The thickness of the epithelium slowly increases from 50 µm at 4 days to 125 µm at 8 days to 200 µm with an additional 60 µm of epithelial cells which appear to be invading the stromal layer at 14 days. The corresponding Raman spectra shown in figure 5.7d show a progressive increase in intensity in the region from 1300-1321 cm^{-1} and the peaks at 1450 and 1657 cm^{-1} from day 4-14. A decrease is seen in the peak at 1048 cm^{-1} while the peak at 1244 cm^{-1} remains constant.

5.5 Discussion

This study indicates that organotypic raft cultures can be used as an accurate, reproducible model system for spectroscopic analysis of tissue. Comparisons of this data from the raft cultures with data collected from an ongoing *in vivo* clinical study [25] indicate that raft cultures are not only an accurate morphologic representation of *in vivo* tissue as shown by histology but also an accurate biochemical representation as shown by comparisons of Raman spectroscopy data. Histology results shown here are consistent with previous reports of raft cultures constructed using normal and cancerous epithelial cells [23]. Analysis of the Raman spectroscopy data shows that all of the major peaks present in spectra from *in vivo* data are also present in similar ratios in spectra from the raft cultures. This suggests that the raft cultures as described here truly represent the tissue biochemistry contributing to the overall Raman spectra (type I collagen, fibroblast cells, and epithelial cells). Other minor components present in the raft culture arise from the media: epidermal growth factor (EGF), epinephrine, insulin, apo-transferrin, hydrocortisone, penicillin/streptomycin/amphotericin B, pituitary hormones, inorganic salts, pyruvate, and amino acids. These minor components are unlikely to contribute to the overall Raman spectra due to their relatively low concentration when compared to the major components. However, future experiments using raft cultures will test this hypothesis and determine the significance of these minor constituents to tissue biochemistry and their affect on the Raman spectra.

As a model system for tissue spectral analysis, there are many foreseeable benefits to using organotypic cultures for analysis of tissue spectroscopy. Organotypic cultures may eliminate some problems associated with human tissue based studies such as difficulty acquiring tissue specimens and inherent diversity between tissue samples due genetics, patient demographics, and state of tissue. Although the ultimate goal is to develop optical spectroscopy to be robust enough to provide accurate diagnosis despite this inherent patient diversity, the ability to provide a reproducible model system will allow more accurate assessment of the ability of optical spectroscopy to distinguish a given pathology independent of patient variation. While the current study presents the benefits of these cultures in the context of Raman spectroscopy, such a model system can be used to understand and enhance other optical methods as well. In fact, organotypic cultures provide a system by which tissue variables can be controlled and manipulated, thus presenting an environment wherein the effect of certain variables (tissue

inflammation, individual biomolecular concentration, enzyme concentrations) on tissue optical spectra can be determined, unlike with actual tissue samples where variables such as molecular concentration are often unknown. Additionally, since the cells within organotypic cultures are maintained viable, the cultures more closely simulate *in vivo* conditions as compared with *in vitro* tissue specimens where the cells begin to die once taken out of the body, thus altering the molecular composition of the tissue.

In a preliminary application of the raft culture system towards understanding the morphological and biochemical basis of cervical tissue Raman spectra, spectral differences between normal and dysplastic tissues grown in culture were compared to those observed *in vivo*. The raft as indicated by figure 5.5. Raman spectra of the raft cultures as well as *in vivo* human cervical tissues indicate that the region surrounding 1325 cm^{-1} shows the most consistent differences between normal and dysplastic tissues. However, inconsistencies between the raft culture and the *in vivo* data are evident at 1048, 1128, 1445, and 1660 cm^{-1}. The discrepancy seen at 1048 cm^{-1} is likely due to differences in the contribution of silica Raman band in that region. Since the raft cultures are placed on a metallic surface, there is an increased reflection of laser light back into the fiber optic probe. In addition, NHK rafts are typically thinner than SiHa rafts resulting in variability in the silica contribution from the two types of rafts. Other variations in the Raman difference spectra around 1128, 1445, and 1660 cm^{-1} are attributed to differences in tissue biology between the rafts and intact tissue. Thus while the raft cultures appear to be an accurate and reproducible model system for cervical tissue, it may not yet include all the constituents of intact cervical tissue. Although the missing constituents have not yet been identified, future experiments will incorporate such changes as using tumor specific fibroblasts in the dysplastic rafts to identify the basis of the discrepancies at the afore-mentioned bands. Such planned studies will help uncover the origins of the spectral differences between normal and dysplastic cervical tissues leading to a better understanding of the morphologic and biochemical basis of tissue spectral signatures.

Figure 5.6 indicates that the differences seen in the 1325 cm^{-1} peak relative to the peaks at 1240 and 1270 cm^{-1} may be associated with changes in the epithelial layer; the band at 1325 cm^{-1} is predominant in the epithelial spectra whereas the stromal spectra is dominated by the peak at 1270 cm^{-1} with virtually no feature at 1325 cm^{-1}. This is likely due to the high concentration of cells (and therefore nucleic acid) within the epithelial layer and is consistent with previous

studies in the skin (5). The 1325-1330 cm^{-1} peak has been attributed to a CH_3CH_2 wagging mode in the purine bases of nucleic acids [2]. Conversely, the 1270 cm^{-1} peak has traditionally been attributed to the Amide III a C-N stretch from alpha helix proteins [2], and thus is likely originating in the stromal region where the bulk of the proteins reside in the form of collagen. The band at 1655 cm^{-1} is primarily attributed to amide I vibrations and is typically associated with the collagen abundantly present in the stroma of cervical tissue [28]. However, the spectra acquired from the epithelial and stromal layers indicate that there may be another source for this peak within the epithelium, since there are few differences between the epithelial and stromal peaks at 1655 cm^{-1}.

Thus, the changes seen in the relative magnitude of the 1325 cm^{-1} peak can be primarily associated with changes seen in the epithelial layer. Comparing the epithelial spectra of the normal and dysplastic raft cultures (figure 5.6b,c), definite differences can be observed. The epithelial spectra from the dysplastic (SiHa) rafts show an relative increase in the 1325 cm^{-1} and 1450 cm^{-1} peaks, two findings consistent with a study by Omberg, *et al* that looked at the changes in the Raman spectra of immortalized cells versus immortalized cells with the capacity to metastasize (via the addition of a T24Ha-ras oncogene) [29]. The corresponding stromal spectra do not show any significant differences; this would be expected since the stromal layer was kept constant. Yet, there are many additional components present in the stromal layer *in vivo* (such as inflammatory cells, glandular structures, capillaries) which were not included in the construction of the raft cultures. To fully understand how the stromal layer contributes to the overall spectra and thus affects diagnosis, these structures should be added to the tissue culture system one by one to determine which spectral features are influenced by these stromal layer components.

In comparison with the Raman spectra of normal keratinocyte rafts, spectra from both the SiHa and the A431 rafts (two different squamous carcinoma cell lines) show similar spectral differences, a relative increase in intensity of the peaks at 1325 cm^{-1} as well as 1450 cm^{-1}. The primary difference between the Raman spectra of the two different dysplasia rafts appears to be an increase in a small peak at 1307 cm^{-1}. This subtle difference seen between the spectra of the A431 and SiHa rafts indicates that different cell lines leading to the same overall pathology can have subtle biochemical differences which can be detected using Raman spectroscopy. Similarly, in human tissues *in vivo*, differences in the epithelial cells can lead to the same end

pathology, even though biochemically these lesions may be completely different. Raman spectroscopy may be used to detect these biochemical variations. On the other hand, these differences may also be a source of variability in clinical Raman spectra as lesions are typically categorized by their pathological diagnosis rather than the biochemical reality, possibly leading to problems of misclassification.

The spectral changes associated with the increase in epithelial thickness (figure 5.7) indicate that the spectral changes seen in spectra of dysplastic raft cultures as compared with normal raft cultures are in part due to the difference in epithelial thickness. On day 4, the spectra of the SiHa raft culture resembles the spectra measured from NHK rafts. The spectral increases seen in figure 5.7d at 1321, 1450 and 1657 cm^{-1} from day 4-14 are similar to differences seen in figure 5.5 between normal and dysplastic spectra. Since the epithelial layer in the raft cultures increases in thickness and density from day 4-14, the observed decrease in intensity of the peak at 1048 cm^{-1} is a similar phenomenon to the one described above in the NHK/SiHa comparison. Because this peak is relatively large, the change in contribution affects the intensity of the remaining peaks through the mean intensity normalization procedure. This makes the interpretation of the remaining changes in peak intensity problematic. Simple subtraction of a background measurement to eliminate the silica peaks is not possible since the tissue optical properties of each raft culture affects the amount of reflection of laser light back into the fiber-optic probe thus affecting the size of the silica Raman peaks. Vector subtraction of the silica Raman peaks is one possible solution employed in another study of tissue Raman spectroscopy using the same fiber-optic probe [11]. Another solution would be to truncate the spectra below 1200 cm^{-1} which would eliminate the spectral regions of silica Raman peaks; this method would also eliminate the peaks at 1006, 1055, and 1086 cm^{-1} which are likely peaks from the tissue. Regardless of artifacts due to spectral normalization, these results indicate that the changes seen in spectra of dysplastic raft cultures are due in part to the increase in epithelial thickness as compared with normal raft cultures. Thus, it is conceivable that Raman spectroscopy may not detect dysplastic changes in cases in which the epithelial layer has been denuded. Yet, results from the Omberg, *et. al*. study indicate that metastatic cells have Raman spectral changes as compared with cells without the ability to metastasize independent of thickness or quantity [29]. Thus, the spectral changes of dysplasia are not solely due to the thickness effect. If spectra of epithelium and stroma could be measured independently, the effects of thickness could be

delineated from the true differences of dysplasia versus normal epithelium. This should be taken into consideration when designing Raman spectroscopy systems for the future clinical detection of cervical dysplasia.

In conclusion, organotypic cultures hold promise as a model system that can be manipulated and controlled to better understand the relationship between tissue components and tissue optical spectra. This study indicates the potential of raft cultures, a type of organotypic culture, as a model system to understand the morphologic and biochemical basis of optical spectroscopy using Raman spectroscopy as one type of tissue spectroscopy. The results presented here indicate the effect of one perturbation of this model system (alteration of epithelial cell content) to show how such a model system could be used to better understand how changes in tissue biology affect tissue spectra. While further studies are planned for a thorough understanding of cervical tissue Raman spectroscopy, this technology could easily be adopted to other organ systems as well as other types of spectroscopic detection.

5.6 References

1. Ramanujam, N., Fluorescence spectroscopy of neoplastic and non-neoplastic tissues. *Neoplasia*, 2000. **2**(1-2): p. 89-117.

2. Stone N, *et al.*, Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 564-573.

3. Stone, N., *et al.*, Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. *Laryngoscope*, 2000. **110**(10 Pt 1): p. 1756-63.

4. Shafer-Peltier, K., *et al.*, Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis *in vivo*. *Journal of Raman Spectroscopy*, 2002. **33**(7): p. 552-563.

5. Caspers, P.J., *et al*., In vitro and in vivo Raman spectroscopy of human skin. *Biospectroscopy*, 1998. **4**(5 Suppl): p. S31-9.

6. Mahadevan-Jansen, A., *et al.*, Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers. *Photochem Photobiol*, 1998. **68**(1): p. 123-32.

7. Utzinger, U., *et al.*, Near-Infrared Raman Spectroscopy for *In vivo* Detection of Cervical Precancers. *Applied Spectroscopy*, 2001. **55**(8): p. 955-9.

8. Shim, M.G., *et al.*, *In vivo* near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy. *Photochem Photobiol*, 2000. **72**(1): p. 146-50.

9. Molckovsky, A., *et al.*, Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. *Gastrointest Endosc*, 2003. **57**(3): p. 396-402.

10. Caspers, P.J., *et al.*, *In vivo* confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. *J Invest Dermatol*, 2001. **116**(3): p. 434-42.

11. Bakker Schut, T.C., *et al.*, *In vivo* detection of dysplastic tissue by Raman spectroscopy. *Anal Chem*, 2000. **72**(24): p. 6010-8.

12. Drezek, R., *et al.*, Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochemistry with dysplasia. *Photochem Photobiol*, 2001. **73**(6): p. 636-41.

13. Tumer, K., *et al.*, Ensenbles of radial basis function networks for spectroscopic detection of cervical precancer. *IEEE Trans Biomed Eng*, 1998. **45**(8): p. 953-61.

14. Brennan, J.F., 3rd, *et al.*, Determination of human coronary artery composition by Raman spectroscopy. Circulation, 1997. **96**(1): p. 99-105.

15. Buschman, H.P., *et al.*, Raman microspectroscopy of human coronary atherosclerosis: biochemical assessment of cellular and extracellular morphologic structures *in situ*. *Cardiovasc Pathol*, 2001. **10**(2): p. 69-82.

16. Parenteau, N., Skin equivalents, in *Keratinocyte Methods*, I. Leigh, Watt, F., Editor. 1994, Cambridge University Press. p. 45-54.

17. Navsaria, H.A., *et al*., Human epidermal keratinocytes, in *Keratinocyte Methods*, I.M. Leigh and F. Watt, Editors. 1994, Cambridge University Press. p. 5-12.

18. Zheng, J. and A. Vaheri, Human skin fibroblasts induce anchorage-independent growth of HPV-16- DNA-immortalized cervical epithelial cells. *Int J Cancer*, 1995. **61**(5): p. 658-65.

19. Eicher, S.A., *et al.*, Evaluation of topical gene therapy for head and neck squamous cell carcinoma in an organotypic model. *Clin Cancer Res*, 1996. **2**(10): p. 1659-64.

20. Pfutzner, W., *et al.*, Selection of keratinocytes transduced with the multidrug resistance gene in an *in vitro* skin model presents a strategy for enhancing gene expression *in vivo*. *Hum Gene Ther*, 1999. **10**(17): p. 2811-21.

21. Chow, L.T. and T.R. Broker, *In vitro* experimental systems for HPV: epithelial raft cultures for investigations of viral reproduction and pathogenesis and for genetic analyses of viral proteins and regulatory sequences. *Clin Dermatol*, 1997. **15**(2): p. 217-27.

22. Delvenne, P., *et al.*, The organotypic culture of HPV-transformed keratinocytes: an effective *in vitro* model for the development of new immunotherapeutic approaches for mucosal (pre)neoplastic lesions. *Vaccine*, 2001. **19**(17-19): p. 2557-64.

23. Rader, J.S., *et al.*, *In vitro* differentiation of epithelial cells from cervical neoplasias resembles *in vivo* lesions. *Oncogene*, 1990. **5**(4): p. 571-6.

24. Sokolov, K., *et al.*, Realistic three-dimensional epithelial tissue phantoms for biomedical optics. *J Biomed Opt*, 2002. **7**(1): p. 148-56.

25. Robichaux, A., *et al. In vivo* Detection of Cervical Dysplasia with Near Infrared Raman Spectroscopy. in SPIE Photonics West. 2002. San Jose, CA: SPIE Press.

26. Shim MG, W.B., Marple E, Wach M, *Study of Fiber-Optic Probes for in vivo Medical Raman Spectroscopy. Applied Spectroscopy*, 1999. **53**(6): p. 619-27.

27. Savitsky, A. and M. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 1964. **36**(8): p. 1627-1639.

28. Mahadevan-Jansen, A., R. Richards-Kortum, Raman Spectroscopy for the Detection of Cancers and Precancers. *Journal of Biomedical Optics*, 1996. **1**(1): p. 31-70.

29. Omberg, K.M., *et al.*, Raman Spectroscopy and Factor Analysis of Tumorigenic and Non-Tumorigenic Cells. *Applied Spectroscopy*, 2002. **56**(7): p. 813-819.

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

Amy Robichaux Viehoever[1]

[1]Department of Biomedical Engineering
Vanderbilt University, Nashville, TN 37235

6.1 Summary

The research presented in this dissertation represents the next step in evaluating the potential of using Raman spectroscopy for differential diagnosis in the cervix. Several key advances were made to the field of tissue Raman spectroscopy. First, the study in chapter 3 represents the largest *in vivo* study using Raman spectroscopy for the detection of dysplastic and cancerous changes to date. Secondly, the analysis of variance in chapter 4 quantified the spectral variance present from different sources using a method not previously used in Raman spectral analysis of tissue. Although the sources of variability examined in chapter 4 were specific to the cervix, the results give insight to the nature of variance of Raman spectra within a given pathology category. Lastly, the development of the organotypic raft culture model as an *in vitro* system for understanding the biochemical and cellular basis of spectral changes seen *in vivo* introduces a method for addressing basic science questions regarding the effect of tissue biology on Raman spectra.

In the first clinical study described in chapter 3, the Raman spectra were characterized for the following cervical tissue pathology categories: normal ectocervix (squamous epithelium), normal endocervix (columnar epithelium), squamous metaplasia, low grade dysplasia, and high grade dysplasia. Peak intensities differences that were statistically significant between normal ectocervix and high grade dysplasia spectra were used to build an unbiased logistic regression algorithm using independent training and test data sets. The algorithm maximally separated the two categories with a sensitivity similar to and specificity superior to colposcopy. The remaining pathology categories were included in the test set, and although only 2/3 of the squamous metaplasia classified with the normal ectocervix (or "benign"), the majority of the normal endocervix were correctly classified as normal. The low grade dysplasia tended to classify in the middle as would be expected as low grade dysplasia is an intermediate step in the progression from normal ectocervix to high grade dysplasia. A second algorithm was developed to improve the specificity of the first algorithm. Using logistic regression, the second algorithm was trained to discriminate between spectra from squamous metaplasia and high grade dysplasia; some of the same spectra used in the training set for the second algorithm were also included in the test set, and thus this algorithm is biased. Even though the cross validation of this second algorithm is not an accurate estimate of the future ability of this algorithm to classify spectra, it does show

how such a multiple algorithm model system could be used to maximally classify spectra into multiple pathology categories.

Thus, my research has shown that Raman spectroscopy can distinguish high grade dysplasia from benign areas of the cervix. Though future studies are needed to further evaluate the accuracy of the classification, the studies presented establish the groundwork to develop Raman spectroscopy as a non-invasive, real-time diagnostic tool for lesions on the cervix. Although there have been numerous reports of Raman spectroscopy sucessfully performing diagnosis *in vitro*, very few studies to date have investigated its potential *in vivo*. Thus, the current work represents a significant advance in the understanding of the power and limitations of Raman spectroscopy for diagnosis in the clinic and operating room settings. Many of the *in vitro* studies in the literature have only used biopsy samples with a single pathology throughout the specimens. This method allows for the collection and delineation of the pure Raman spectral signature for a given pathology, it does not provide an accurate assessment of the capability of the technology to perform diagnosis in a clinical setting in which tissue biology has many spatial variations in pathology. By conducting research in a clinic setting, the variabilities and inaccuracies due to variations in tissue biology, different physician practices, and time constraints become part of the accuracy estimates, thereby yielding a more realistic assessment of the true ability of the technology to be developed into a diagnostic tool.

The second clinical study (chapter 4) examined the sources of variation within the Raman spectra of a single pathology category. The main source of variation was found to originate in patient to patient variability. Menopausal status was shown to be one component of this interpatient variability. The effect of application of acetic acid was also studied. The results show that although there are no time dependent spectral changes due to the application of acetic acid, there are 2 small non-significant differences in peak intensities. Another source of interpatient variation seems to be differences between normal spectra measured from normal patients and normal spectra measured from patients with dysplasia.

Although several other studies have looked at the Raman spectral variance using different methods, the analysis of variance presented here represents the first study which quantifies the contributions of variance from different sources. By understanding the sources of variation, future clinical studies can be optimized and the future performance of the technology can be better evaluated. For example, menopausal status was shown to be a significant source of

interpatient variation, particularly in the spectral region from 1240-1330 cm^{-1}, a region which in chapter 3 was shown to be important for discrimination. Based on this finding, future studies need to be designed such that the data will be stratified according to menopausal status, and separate discrimination algorithms can be developed for each group which will likely improve the accuracy of classification.

The observation that normal spectra measured from normal patients have significantly different spectral features from normal spectra measured from patients with dysplasia is an unparalleled finding which will have implications for future research throughout the field of tissue Raman spectroscopy. Although it remains unclear what if anything these spectral differences represent, it could indicate that Raman spectroscopy can detect subcellular transformations prior to any change visible upon histological examination. It also indicates that the spectra may need to be stratified to reflect overall patient diagnosis. If such a finding holds true in other organ sites in addition to the cervix, it will require a global transformation in the way clinical research are conducted to evaluate tissue spectroscopy. Studies will have to include a subset of normal patients, or patients without evidence of disease, which is considerably more difficult for investigations at organ sites other than the cervix.

The raft culture studies (chapter 5) give insight into the biochemical and cellular basis for the differences seen between normal and dysplastic Raman spectra. Histological and spectral comparisons of raft cultures with *in vivo* cervical tissue illustrates that raft cultures are a good *in vitro* model system for *in vivo* conditions. Difference spectra measuring the divergence of dysplasia from normal spectra showed that the raft culture spectra show similar deviations as are seen *in vivo* with the exception of the 1450 cm^{-1} peak. Spectra measured individually of epithelia and stroma revealed that the 1325 cm^{-1} peak originates primarily in the epithelial layer whereas the 1272 cm^{-1} peak has virtually no contribution from the epithelial layer and is primarily associated with the type I collagen in the stromal layer. All other peaks have equal contributions from both the epithelial and stromal layers.

While raft cultures have been used as a effective model system in fields such as cell biology and virology, the work described in chapter 5 represents the first use of such a novel model system for understanding the biological basis of tissue spectra. If spectral differences can be linked to known biological changes present in the neoplastic transformation, then those spectral differences can be preferentially used in the development of diagnostic algorithms.

Spectral differences which are linked with known biological phenomena are more likely to be present in future data sets, thus increasing the likelihood of robust algorithm development. The experiments detailed in chapter 5 illustrate that spectral differences found in the 1300-1330 region between normal and dysplastic rafts result from the increase in total nucleic acid as a part of changes in the epithelial cell content. Since these biological changes are seen consistently in the transformation from normal to dysplastic epithelium, this region of the spectra will continue to be a critical region for distinguishing the two pathologies. Thus, the inclusion of this region in any discrimination algorithm will likely yield a more robust algorithm as these spectral differences are biologically based rather than based on the statistical differences of the current study population.

This raft culture model system could also be applied in other types of optical spectroscopy and other organ systems. This system represents a completely unique way of addressing the question of how the biology of the tissue affects the spectra. Most other studies aimed at understanding the biology underlying tissue spectra have used frozen tissue. Yet, frozen tissue studies are limited by the numbers of specimens collected and the fact that the tissue is no longer viable. Raft cultures overcome these limitations. First, because the cultures yield reproducible spectra, a conclusion can be drawn within the context of a few repeatable experiments. Since the components of the raft cultures are known ahead of time and can be altered, the contribution and effect of a certain tissue component can be determined, as opposed to frozen tissue spectroscopy in which tissue components are fixed and quantities are unknown. So instead of using the spectroscopy to quantitate the amount of a certain tissue component, the raft culture model system uses changes in the amounts of tissue components to determine the effect of that tissue component on tissue spectroscopy. This represents a shift in perspective and a method of collecting information not possible with the frozen tissue spectroscopy.

Overall, the following conclusions can be made from the data:

(1) There are statistically significant differences between high grade dysplasia and “benign” cervix (normal ectocervix, squamous metaplasia) that can be used to build discrimination models. Visual distinctions are present in the mean spectra of low grade dysplasia and categories of benign cervix indicating potential regions for use in future discrimination algorithms.

(2) An single algorithm logistic regression model can distinguish high grade dysplasia from normal ectocervix with slightly better accuracy than colposcopy using unbiased estimates. A second algorithm added to the model improves the accuracy of the discrimination of squamous metaplasia using biased estimates of accuracy.

(3) With the collection of more spectra in all pathology categories, independent training and test sets can be used to develop algorithms which could distinguish each of the pathology categories individually.

(4) The majority of the variance seen in spectra of normal ectocervix emanates from patient to patient variation. Intrapatient (or location) variation contributes very little to the total variance; thus, repeat measurements of normal sites within a given patient are not necessary. Additional research into the sources of interpatient variation is needed.

(5) Menopausal status is one contributor of interpatient variation. Because there is a significant difference in the peak at 1324 cm^{-1}, a peak associated with nucleic acid found primarily in the epithelial layer. Since this peak is significantly different between spectra from high grade dysplasia, normal ectocervix, and squamous metaplasia, it is likely to continue to be necessary for discrimination, and therefore, it may be necessary to stratify the data according to menopausal status in future studies.

(6) Another contributor of interpatient variation can be found in the differences seen between normal spectra measured from patients with normal cervices and patients with cervical dysplasia. Although sources of these discrepancies are unknown, this finding indicates that Raman spectroscopy may be detecting subcellular changes not apparent upon routine histologic examination.

(7) Results from the acetic acid analysis indicate that the application of acetic acid enhances two of the spectral changes between normal and dysplasia spectra, and thus it would be advisable to always apply acetic acid prior to spectral measurement, though regulation of the time after application is not critical.

(8) The results from the raft culture data provides insight into the origins of some of the spectral peaks:

(a) The 1324 cm^{-1} peak originates primarily in the epithelial layer, likely from the nucleic acid present in epithelial cells.

(b) The 1272 cm^{-1} peak originates primarily in the stromal layer, likely from the type I collagen.

(c) The remainder of the major peaks have origins in both the epithelial and stromal layers.

(d) Changing the epithelial cell content from normal keratinocytes to squamous cell carcinoma cells induces the same spectral changes seen when comparing *in vivo* normal and dysplastic spectra, with the exception of the 1450 cm^{-1} peak. This indicates that the change in the 1450 cm^{-1} peak in dysplasia is likely a result of changes occurring in the stromal layer, since this was not changed in the raft culture model.

(e) The spectral changes seen in the dysplastic rafts as compared with the normal rafts are in part due to the increased epithelial thickness present in dysplasia.

6.2 Limitation of Current Results and Future Directions

This work represents the next step in evaluating the potential of Raman spectroscopy as a tool in the screening and diagnosis of cervical dysplasia. Yet, a significant amount of evaluation of this technology and basic science research to understand the biological basis for the spectral changes observed must occur before this technology can begin to be implemented in a clinical situation.

First, larger clinical studies must be conducted to gain the amount of data necessary to to validate the ability of Raman spectroscopy to distinguish multiple different pathologies in the cervix. Larger data sets will allow for an adequate size training and test set for robust algorithm development and analysis. Specifically, an increase in the numbers of low grade dysplasia and normal endocervix spectra will allow for the development of discrimination algorithms for these categories. Then a multiple algorithm model can be developed to accurately distinguish all pathology categories instead of relying on one or two algorithms as was used in this research. Based on the results from the variability analysis in chapter 5, larger numbers of spectra also are needed to stratify according to menopausal status; acquiring the numbers of patients necessary to accomplish this will be difficult as most patients presenting with cervical dysplasia are pre-menopausal, yet doing so should improve the accuracy of classification. Lastly, studies should be conducted at multiple centers to gain a more diverse population of patients.

The evidence that Raman spectroscopy can distinguish between high grade dysplasia and normal ectocervix is strong, but far from decisive. Issues of whether or not high grade dysplasia can be detected when the epithelium is thin or has been denuded need to be resolved. One possible solution would be to design a fiber-optic probe to collect light separately from the epithelial and stromal layers. Thus the effect of the thickness of the epithelium would have less of an effect on the overall spectrum, and the spectral difference between the dysplastic and normal epithelial cells could be detected regardless of the thickness of the epithelium. Distinguishing squamous metaplasia from dysplasia is not only a challenge for colposcopists, but it also seems to be a challenge with Raman spectroscopy. There are areas of statistical differences between squamous metaplasia and high grade dysplasia spectra, but whether these differences will be sufficient to enable accurate discrimination remains to be seen. Regardless, the collection of a larger data set should increase the accuracy of discrimination since the larger the training set, the more accurate the results will be. Preliminary examination show visual distinctions between spectra of low grade dysplasia and benign cervix, a larger sample size will determine the stastical validity of these variations.

Although the Raman spectroscopy system used in the studies presented here represents a great improvement over the system used in the original pilot study using Raman spectroscopy to detect cervical dysplasia [1], further improvements to the current system are needed to improve the quality of the clinical studies. The main limitations of the current system include: interference of ambient room light with the quality of the spectra and the lack of an accurate system to mark the tissue at the site of spectral measurement. Improvements in the design of the fiber-optic probe could aid in the solution of these problems. The probe could be designed to physically block any ambient light present at the tissue by including material to surround the fiber-optics at the probe tip. An increase in the collection efficiency of the probe also would decrease the amount of ambient light collected by the probe by reducing the length of the integration time required for measurement. A decreased integration time would also improve the clinical utility of the system. Incorporating a marking system which could be intrinsic to the probe itself should also be investigated as a means to improve the current marking system.

Additional research is also needed to develop techniques to aid in the development of robust diagnostic algorithms. Currently, there is no consensus in the field of biomedical Raman spectroscopy on the best methods to use in algorithm development, although there is agreement

on the need for unbiased data analysis and independent validation. A variety of statistical methods typically are used by different research groups to design an algorithm which will maximally separate the data at hand. Logistic Regression was chosen based on the nature of the desired classification and the amount of data available. Yet, if a study were to compare different methods used by different research groups and empirically determine which methods were superior, it could help standardize and improve the process of algorithm development throughout the field. Another method for improving algorithm development lies in understanding the biochemical and cellular basis for the spectral differences between pathology groups. The raft culture experiments presented here give us one example of how understanding the biological basis for the spectral features can aid in algorithm development. A continuation of the raft culture experiments should improve not only the understanding of the effect of biological variables on the tissue spectra but also the development of robust algorithms and accuracy of classification.

The analysis of variance uncovered the basic sources of variability present within normal spectra. Further research should expound upon these findings to examine other potential sources of interpatient variability. There were several potential demographic sources of variation which were unable to be evaluated in this study for various reasons including menstrual cycle status, race, and pregnancy status; these should be evaluated in future studies. Also, based on the results in this study, future studies should stratify the data according to menopausal status and develop separate classification algorithms which should improve the accuracy of classification. The source of the discrepancies seen between normal spectra measured in patients with normal cervices and patients with dysplasia should also be investigated further. It may be necessary to create separate categories for these two groups until a more basic understanding of what is causing these spectral differences is uncovered.

The raft culture model presents a method in which many of these basic science questions can be addressed. Thus far, the only perturbations of the model system have been in the epithelial layer, yet there are indications from studies in fluorescence spectroscopy that there are changes occurring in the stromal layer in dysplasia before invasion occurs [2,3]. Additionally, one major discrepancy at 1450 cm^{-1} was seen in the dysplastic-normal comparison between the *in vivo* and the raft culture data. Since no changes were made to the stromal layer, this discrepancy is likely due to changes in the stromal layer. The first step in investigating these

changes would be through the addition of tumor associated fibroblasts to the stromal layer. The addition of these fibroblasts will allow a more accurate representation of the stromal environment present in dysplasia, and could prove to be the factor behind the discrepancy at 1450 cm^{-1}. The overall effect of fibroblasts should also be examined since the fibroblast concentration was held constant for these studies. If a rigorous comparison of the concentration of fibroblasts *in vivo* and raft cultures could be performed, it may aid to optimize the model system. There are also other structural and cellular components present in the *in vivo* stromal layer which were not included in the initial raft culture experiments, specifically basement membrane, proteoglycans, and laminin [4]. These components should be added one by one to determine their affect on the overall Raman spectra. Inflammatory cells were not included in the composition of the stromal layers. Although inflammatory cells are always present at a basal level in the cervical stroma, the amount and type of inflammatory cells will vary depending on the state of the tissue and external environment (i.e. presence or absence of pathogens). The effect of acute and chronic inflammation on the tissue spectra can be studied by adding differing concentrations of poly-nuclear leukocytes or mono-nuclear leukocytes, respectively. Mature macrophages are often found in tissue as well and could be added to simulate their effect. Although several alterations were made to the epithelial layer in the studies presented in chapter 5, the experiments only looked at the differences between normal and high grade dysplasia, with no investigation of low grade dysplasia. By adding epithelial cells infected with low-risk strains of HPV, the spectral changes seen in low grade dysplasia could be examined. These perturbations of the model will allow for a further evaluation of the biochemical and cellular basis of the Raman spectra.

This dissertation research shows that Raman spectroscopy can distinguish dysplasia from normal cervix *in vivo* and non-invasively. Yet, further research is needed to validate this finding, develop more robust discrimination algorithms, and advance the techonology and design a system which can be used with minimal training. The reproducibility of classification currently is not good enough to warrant replacing the standard of care (colposcopy directed biopsy). Yet, collection of larger data sets should enable the development of more robust algorithms since, theoretically, the larger the training set used to develop the algorithm, the better future classification will be. Also, if more information regarding the biologic basis of spectral features can be gained from raft culture experiments and frozen tissue spectroscopy, that information

should also improve algorithm development and thus classification. Diagnostic accuracy may also improve if Raman spectroscopy is combined with another spectroscopic method. Fluorescence spectroscopy is one possibility. Whereas Raman spectroscopy provides a spectral fingerprint of the chemical bonds present in the tissue and thus is a very specific technique, fluorescence spectroscopy provides measure of the metabolic state of the tissue and has good overall sensitivity. A system could be designed to take advantage of both of these modalities in which fluorescence spectroscopy is used first to scan the cervix and detect suspecious areas, followed by a Raman "optical biopsy" to provide a more specific diagnosis of those suspecious areas. Other potential types of spectroscopy include diffuse reflectance and light scattering spectroscopy which have both been examined in combination with fluorescence for diagnosis in the cervix [5]. Of the two, light scattering spectroscopy improved the performance of fluorescence the most and likely holds the most promise for improving the accuracy of Raman spectroscopy. Light scattering spectroscopy measures the reflected light only from the surface epithelial cells. This may aid in the detection of high grade dysplasia in cases when the epithelium is thin or has been denuded since the thickness of the epithelium does not affect the spectrum in light scattering spectroscopy.

Taking all of these points into consideration, Raman spectroscopy does have a realistic potential for becoming a more accurate and cost-effective method of screening for cervical cancer either alone or in combination with other spectroscopic modalities. Thus, research into developing the technology as a diagnostic tool should continue, and should focus particularly on collecting larger data sets for algorithm training and understanding the biological basis for tissue spectra.

6.3 Protection of Research Subjects

For this technology to be truly evaluated in its ability for noninvasive diagnosis, clinical studies must establish the validity of the technique in the clinical and operating room setting. Human subjects have participated as part of both the *in vivo* clinical dysplasia and normal studies of the cervix (as described in chapters 3-4). Additionally, the *in vitro* raft culture studies aimed at understanding the underlying mechanisms as described in chapter 5 necessitate the use of human cell lines in the preparation of these tissue cultures and a small number (5) of human cervical tissue to validate the model system.

For the *in vivo* clinical studies, the Vanderbilt Institutional Review Board reviewed and approved the two separate protocols entitled "Detection of Cervical Precancers Using Near Infrared Raman Spectroscopy: A Pilot *In Vivo* Study" (IRB# 010245) and "Study of Normal Cervix Using Raman Spectroscopy" (IRB # 02-0345). Informed consent was obtained prior to each study, and patients had the right to refuse to participate or withdraw from the study at any time without compromising their care. Adult patients are included in the study without regard to race, ethnicity, or age. The clinical study added no pain to the procedure, and steps were taken to minimize any additional inconvenience or discomfort the patient might experience due to the increased time required for the study.

The use of organotypic cultures is designed to minimize the need to use human and animal tissues. Yet the use of such cultures was supplemented and validated using tissue from actual patients. The Vanderbilt IRB approved a protocol to obtain cervix samples from hysterectomy specimens. Only tissue not needed for patient care was used, and no patient information was used for the *in vitro* studies. Human cell lines were obtained from the American Tissue Type Collection and the Cambrex Corporation.

6.4 Societal Benefits

The benefits to society have been documented in the background and introduction. In brief, Raman spectroscopy has the potential to be a more accurate and less expensive method of screening and diagnosing cervical dysplasia. The ability of Raman spectroscopy to enable a see and treat protocol would decrease the number of return clinic visits and patient anxiety as the long waits for histopathology diagnoses would be eliminated. Thus, the technology has the potential to increase access to underserved populations where follow-up to screening may not be available. The technology poses no known risks to the patient, and therefore could be a safe alternative to current screening methods.

6.5 References

1. Utzinger, U., *et al.*, Near-Infrared Raman Spectroscopy for *In vivo* Detection of Cervical Precancers. *Applied Spectroscopy*, 2001. **55**(8): p. 955-9.

2. Drezek, R., *et al.*, Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: modeling, measurements, and implications. *J Biomed Opt*, 2001. **6**(4): p. 385-96.

3. Palmer, G.M., *et al.*, Fluorescence Spectroscopy of DMBA Induced Hamster Cheek Pouch Model: Analysis of Fluorescence and Diffuse Reflectance Changes with Disease Progression. *Lasers in Surgery and Medicine*, 2002. **S14**: p. 2.

4. Bosman, F.T. and I. Stamenkovic. *Functional structure and composition of the extracellular matrix*. *J. Path*., 2003. 200(4), 423-8.

5. Georgakoudi, I., *et al.*, Trimodal spectroscopy for the detection and characterization of cervical precancers *in vivo*. *Am J Obstet Gynecol*, 2002. **186**(3): p. 374-82.