Cervical Fluorescence of Normal Women

Carrie K. Brookner, MS,1 Urs Utzinger, PhD,1 Gregg Staerkel, MD,2 Rebecca Richards-Kortum, PhD,1* and Michele Follen Mitchell, MD3

1 Biomedical Engineering Program, University of Texas, Austin, Texas 78712
2 Department of Pathology, UT MD Anderson Cancer Center, Houston, Texas 77030
3 Department of Gynecology, UT MD Anderson Cancer Center, Houston, Texas 77030

Background and Objective: Cervical tissue fluorescence spectra have previously been measured in vivo in women with a recent abnormal Papanicolaou smear. Diagnostic algorithms have been developed to diagnose squamous intraepithelial lesions (SILs) based on these fluorescence emission spectra. However, algorithms have not been tested in women with no history of cervical neoplasia.

Study Design/Materials and Methods: Cervical fluorescence was measured from 54 women with no history of cervical dysplasia, and the spectra were compared to those from colposcopically normal sites in women with suspected dysplasia. Representative spectra from each group were compared and a two-sided, unpaired Student’s t-test was performed to compare mean principal component scores used in previously published diagnostic algorithms. The ability of previously reported diagnostic algorithms to classify these samples as normal tissue was also assessed.

Results: At the 0.05 level of significance, the mean scores of 4 of the 7 important principal components were statistically different for the two populations. However, when the data collected from volunteers in this study were preprocessed in the appropriate manner and the algorithms were applied, more normal samples were correctly classified than in the previous clinical study in which these algorithms were developed.


Key words: cervical intraepithelial neoplasia; diagnosis; spectroscopy; squamous intraepithelial lesion

INTRODUCTION

Fluorescence spectroscopy is a technique that has the potential to improve the accuracy and efficacy of cervical precancer screening and diagnosis [1–8]. Specifically, multivariate statistical algorithms have been developed that can differentially diagnose SILs in vivo based on fluorescence emission spectra collected at three excitation wavelengths (337, 380, and 460 nm) [6–8]. Briefly, the algorithms preprocess the data to reduce inter-patient and intra-patient variation, reduce the data using Principal Component Analysis, select important principal components (PCs), and calculate the posterior probability of the sample belonging to each tissue type [8–11]. The performance of these algorithms has been evaluated retrospectively and prospectively in a group of 95 patients relative to colposcopy in expert hands and to the Pap smear in the referral setting. The sensitivity and specificity of fluorescence-based algorithms (82% and 68%, respectively) compare favorably to those of colposcopy (94% and 48%).

*Correspondence to: Dr. Rebecca Richards-Kortum, Department of Electrical and Computer Engineering, University of Texas, ENS 610, Austin, TX 78712.
E-mail: kortum@mail.utexas.edu
Accepted 27 August 1998

© 1999 Wiley-Liss, Inc.
and those of the Pap smear (62% and 68%) [12,13]. Due to the high sensitivity of colposcopy, this test has a false negative rate of only 6%. Based on these encouraging results, it has been suggested that fluorescence spectroscopy could be used to improve diagnosis and, potentially, screening for cervical dysplasia.

Previously reported data were measured only from patients in a referral setting who had recent abnormal Pap smears. Benedetto et al. [14] have shown that squamous intraepithelial lesions are associated with marked changes in tissue protein thiols and disulfides and that these differences extend to neighboring apparently normal tissue. The goal of this study was to investigate possible spectral differences between colposcopically normal areas in women with and without suspected cervical dysplasia.

We measured the cervical fluorescence emission spectra at three excitation wavelengths from 54 patients who had never had an abnormal Pap smear. These were compared to previously reported data from 95 patients with an abnormal Pap smear. Spectra and diagnostically important principal components were compared using a variety of statistical techniques to identify possible spectral differences. In addition, we evaluated how accurately our previously developed algorithms classified spectra from colposcopically normal areas of squamous and columnar epithelium and the transformation zone in women with no history of an abnormal Pap smear. Statistically significant differences were identified in the fluorescence of normal cervixes from women with and without a history of cervical neoplasia. Given these differences, previously reported algorithms can accurately classify tissue type based on spectra from both groups.

**MATERIALS AND METHODS**

**Patients**

Women at least 18 years of age with no history of abnormal Pap smears, cervical dysplasia, or treatment to the cervix and no current gynecologic symptoms were invited to participate in our study conducted at the University of Texas Women’s Health Center (Houston). A medical history was obtained, noting age, menopausal status, and smoking history. Informed consent was obtained from each subject who participated in the study, and the study was reviewed and approved by the Institutional Review Boards of the University of Texas at Austin and the University of Texas Health Science Center.

**Instrumentation**

The system used to measure cervical fluorescence spectra at 337, 380, and 460 nm excitation has been described in detail previously and was used with two minor modifications [8]. Briefly, the system consists of two dye lasers pumped by a nitrogen laser, an optical fiber probe, and an optical multichannel analyzer. The probe was modified to incorporate a removable outer sheath, which could be disinfected to minimize the amount of time required between consecutive patients. In addition, a new set of longpass filters to block Rayleigh light was incorporated. The 400 nm longpass filter used at 380 nm excitation was replaced with a 420 nm filter.

**Data Collection**

Pap smears were collected with an endocervical brush and an Ayre’s spatula, following routine procedures. All Pap smears were read by a single cytologist at the University of Texas M.D. Anderson Cancer Center (Houston) and were classified using the Bethesda System. After the Pap smear, 6% acetic acid was applied to the cervix for ~2 minutes to duplicate the conditions under which previous data were collected. Acetic acid produces whitening of abnormal areas and is used to identify areas for biopsy during colposcopy [15]. Following the application of acetic acid, a fiber optic probe was placed in contact with the cervix under colposcopic guidance and was used to measure emission spectra at three excitation wavelengths (337, 380, and 460 nm). The probe was used to measure two sites containing squamous epithelium and, where colposcopically visible, two sites containing columnar epithelium and at least one site from the transformation zone (TZ). The transformation zone is the area where the squamous epithelium undergoes metaplasia and begins to cover or grow over areas of columnar epithelium. This process of squamous metaplasia on the cervix is physiological and is considered a normal event [15]. In many of the women, columnar cells were not present on the ectocervix, and the transformation zone was not easily identified.

For each patient studied, the probe sheath was disinfected and background spectra, spectra of a rhodamine standard, and spectra of the cervix were collected as previously described [8], except that 50 spectra were acquired for 50 consecutive laser pulses at 337 nm excitation, 100 spectra
were acquired for 100 consecutive laser pulses at 380 nm excitation, and 200 spectra were acquired for 200 consecutive laser pulses at 460 nm excitation. For each patient, background spectra at each excitation wavelength were averaged. Furthermore, spectra from each cervical site at each excitation wavelength were averaged to give a single spectrum per site. The appropriate average background spectrum was subtracted from each patient measurement. All background-subtracted spectra were corrected for the nonuniform spectral response of the modified detection system using correction factors obtained by recording the spectrum of an N.I.S.T. traceable calibration tungsten ribbon filament lamp. Our daily calibration routine and the use of the new correction factors led to reproducible results, despite the system changes.

Algorithm

Previously described classification rules to separate cervical SILs from tissue without SIL, based on fluorescence emission spectra at three excitation wavelengths, utilize two constituent algorithms to classify samples as: (1) squamous normal (SN) or SIL (SN/SIL algorithm), or (2) columnar normal (CN) or SIL (CN/SIL algorithm) [8]. Both of these constituent algorithms are based on multivariate statistical analysis and have been described in detail previously by Ramanujam et al. [8]. The multivariate statistical analysis of cervical tissue spectra involves four primary steps: (1) preprocess the spectral data to reduce interpatient and intrapatient variation within a tissue type, (2) dimensionally reduce the preprocessed spectra using Principal Component Analysis (PCA), into a set of principal components that describes most of the variance in the original data set [9], (3) select diagnostically important principal components using an unpaired, one-sided Student's t-test [10], and (4) develop a classification algorithm based on logistic discrimination that uses these important principal components [11]. The SN/SIL algorithm requires data that have been preprocessed by normalization, whereas the CN/SIL algorithm requires normalization and mean-scaling [8]. The diagnostically important principal components for the SN/SIL algorithm are PC1, PC3, and PC7 and for the CN/SIL algorithm are PC1, PC2, PC4, and PC5. The classification algorithms use Bayes’ theorem to calculate the posterior probability that an unknown sample belongs to each tissue category. The posterior probability depends upon the prior probability of belonging to each tissue type, the cost of misclassification, and the conditional joint probabilities that, given the particular tissue type, the sample will have a given principal component score. Algorithms were developed and evaluated relative to histopathology using previously reported data from 95 patients with an abnormal Pap smear [8]. The data set was randomly divided into a training set, used to develop and optimize the algorithm, and a validation set used to evaluate its performance.

Data Analysis

Data obtained in this study from women with no history of an abnormal Pap smear were compared to previously reported data from 95 patients with an abnormal Pap smear [8]. As a first method of comparison, representative spectra from colposcopically normal areas of a patient from each study were plotted to allow visual comparison of the line shapes.

Quantitative metrics also were used to compare the entire current data set to our previously reported data. Our previously reported data were divided into a training group and a validation group. The spectra of patients in the training group were assembled into a single data matrix [8]; the eigenvectors of the covariance matrix are used to calculate principal component scores and also can be used to reconstruct any spectrum in the data set. As such, these eigenvectors provide a metric to compare the similarity of spectra from the training set and from another group. Any new data set can be approximated as a linear combination of these eigenvectors; the difference between this approximation and the actual data is a measure of the similarity between the two groups. We used the eigenvectors that account for 95% of the variance in the training data set of the previous study to approximate the spectra of both the validation data set of the previous study and the data set of the current study. The average root mean square (RMS) difference between the approximation and the actual data was calculated at each wavelength. The standard deviation of this RMS difference was also calculated. A two-sided Student’s t-test was used to compare the mean differences of the two data sets. As a reference, the same t-test was used to compare the mean differences between the training and validation sets of the previous study data.

The eigenvectors of the training set from the previously reported data were then used to calculate the principal component scores for the data
set of the current study. The probability density functions of the principal component scores important in the constituent algorithms (SN/SIL algorithm (scores of PC1, PC3, PC7); CN/SIL algorithm (scores of PC1, PC2, PC4, PC5)) were plotted for the data from both studies. Finally, a two-sided, unpaired t-test was performed to compare the means of the important principal component scores of the squamous and columnar normal tissues from these two data sets.

Moreover, the previously developed diagnostic algorithms were applied to the spectra collected from women with no history of dysplasia. Spectra from each cervical site were appropriately preprocessed prior to algorithm implementation. The principal component scores were calculated for each algorithm using the eigenvectors that describe the training set data from the previous clinical study, in which the algorithms were developed. Finally, the posterior probabilities of each site belonging to each tissue type were computed. The conditional probabilities were computed using the probability distributions of the principal component scores of each tissue type from the training set data from the previous clinical study. The costs of misclassifying SIL samples are the same as those in the previous study; the cost associated with misclassification of a SIL was 0.70 for the SN/SIL algorithm and 0.58 for the CN/SIL algorithm. The prior probabilities were determined by calculating the percentage of each appropriate tissue type in the sample population. Specifically, to compute the prior probabilities of being SN or SIL (first algorithm), the SN and TZ sites were considered normal and the CN sites were considered abnormal because previous studies indicate that CN samples are generally misclassified as SILs by the SN/SIL algorithm [6–8]. To calculate prior probabilities of being CN or SIL (second algorithm), the CN sites were considered normal and sites from patients with an abnormal Pap smear were considered SILs to avoid a prior probability of unity. The TZ sites were included only in the computation of the prior probabilities of the SN/SIL algorithm; however, both algorithms were used to classify these sites to determine whether they were better classified as SN or CN.

RESULTS

Spectra were measured from 202 sites in 54 women. The mean age of the participants was 37 ± 9 years. Eight (14.5%) participants were post-menopausal, and two of these eight were on hormone replacement therapy. Six (11%) of the participants were smokers.

The cytology reports for the 54 study participants indicated the following: 10 within normal limits, 4 negative for malignant cells (NMC) with atrophy, 23 NMC with acute inflammation, 6 NMC with Candida or coccobacilli with shift in vaginal flora, 7 NMC with acute inflammation and Candida or coccobacilli, 1 atypical squamous cells of undetermined significance (ASCUS) favor dysplastic process, 1 ASCUS favor human papilloma virus (HPV), 1 suggestive of HPV, and 1 high grade SIL. For the purpose of algorithm implementation, a binary classification scheme was used in which only the four patients with ASCUS, HPV, or SIL were considered abnormal and the remainder were considered normal.

Fluorescence spectra were measured from a total of 186 sites in the 50 women with normal Pap smears; 103 contained squamous normal tissue (SN), 23 contained columnar normal tissue (CN), and 60 were from the transformation zone (TZ). In the four women with abnormal Pap smears, fluorescence spectra were measured from 16 sites; 8 contained squamous tissue, 1 contained columnar tissue, and 7 were from the TZ.

Comparison of Two Sets of Normal Data

Emission spectra at each excitation wavelength from a representative patient from this study are shown in Figure 1a–c. For comparisons, Figure 1d–f shows spectra from a representative patient in the previous clinical study. In the spectra from both studies, at 337 nm excitation (Fig. 1a,d), a red shift is seen for the columnar normal with respect to the squamous normal. For a single patient, the spectrum from a TZ or metaplastic site is found between those for the squamous and columnar normal sites. At 380 nm excitation (Fig. 1b,e), a smaller red shift is seen in the columnar normal spectra. At 460 nm excitation (Fig. 1a,f), columnar normal tissue shows up to 50% additional fluorescence compared to squamous normal at 550–700 nm emission.

Figure 1a–c also shows the average root mean square error between the data measured in this study and its fit to a linear combination of the eigenvectors of the training set data from our previous study. Figure 1d–f shows the similar average RMS error for the validation set data from our previous study. For this validation data set, the average error between the true and reconstructed spectra is <5%. The RMS error between the actual
and reconstructed data from the current study shows a peak of ~10% below 400 nm at 337 nm excitation, below 450 nm at 380 nm excitation, and below 500 nm at 460 nm excitation. No obvious differences between the two studies can be seen above these wavelengths.

The results of the Student’s t-test (for data with unequal standard deviations), to compare the RMS differences of the data sets at each excitation wavelength, are shown graphically in Figure 1a–c for this study and in Figure 1d–f for the previous study. The circles shown at y = 1 indicate the wavelengths at which statistical differences were found between the average RMS differences, and circles at y = 0 indicate the wavelengths at which significant differences were not found. The results in Figure 1d–f show that the average RMS differences between the training and validation sets of the previous study are not statistically different, as expected. In contrast, Figure 1a–c shows that statistically significant differences were found between the average RMS differences of the previous training set and the current data set for most emission wavelengths below 500 nm at all excitation wavelengths. In addition, differences are seen for 337 nm excitation...
tion at 530 and 630 nm emission, for 380 nm excitation at 530 and 600–650 nm emission, and for 460 nm excitation above 600 nm emission.

Figure 2 compares the probability density functions of the important principal component scores—(a) PC1, (b) PC3, (c) PC7—of the SN/SIL algorithm. The scores of PC1 for SN sites from the two data sets are most similar (Fig. 2a). The mean scores of PC1 for the two sets of SN sites were not statistically dif-

Fig. 2. Comparison of probability density functions for the important principal component scores—(a) PC1, (b) PC3, (c) PC7—of the SN/SIL algorithm.
different. However, if the TZ sites from the current study were also included, then the mean scores of PC1 for the two sets were different ($P = .0009$). The mean scores of PC3 and PC7 from the two data sets were found to be statistically different ($P < .005$) regardless of whether the TZ sites were included in the analysis (Fig. 2b,c).

The important principal component scores for the CN/SIL algorithm are compared in Figure 3. The scores of PC4 are the most similar (Fig. 3c). At the 0.05 level of significance, the mean scores of PC1 and PC4 were not statistically different. However, the mean scores of PC2 and PC5 were found to be statistically different ($P = .003$).

**SN/SIL Algorithm**

The SN/SIL algorithm was implemented to calculate the posterior probability that each cervical site was normal squamous tissue or a SIL. The prior probability of being SN (0.876) was calculated as the proportion of the total measurement sites that was either SN or TZ. The prior probability of being a SIL (0.124) was calculated as the proportion of the total that was CN. The algorithm results for SN and TZ samples are shown in Table 1. For comparison, the results of the same algorithm applied to the validation set of the previous clinical study are also given in Table 1. The percentage of SN samples correctly classified exceeds that reported for the validation set SN samples by 17%. The percentage of TZ samples correctly classified exceeds that reported for the validation set SN samples by 8%.

**CN/SIL Algorithm**

The CN/SIL algorithm was implemented to calculate the posterior probabilities that each cervical site was either CN tissue or a SIL. The prior probabilities of being CN and SIL were determined to be 0.59 and 0.41, respectively. The algorithm results for the CN and TZ samples are shown in Table 1. For comparison, the results of the same algorithm applied to the validation set of the previous clinical study are also given in Table 1. The percentage of CN samples correctly classified exceeds that reported for the validation set CN samples by 20%. The classification results for the TZ samples were similar to those obtained using the SN/SIL algorithm. Here, the percentage of TZ sites classified as normal tissue is comparable to that reported for the validation set CN samples.

**DISCUSSION**

Various research groups have developed diagnostic algorithms based on fluorescence spectroscopy, which can be used to discriminate normal from neoplastic tissues in the breast, colon, bladder, lung, and cervix [1–8,16–20]. However, these various diagnostic algorithms are usually developed using patients with suspected disease; the normal sites were measured from these patients as well. This study shows that the fluorescence of normal cervical sites may differ in patients with and without a history of dysplasia, and these potential differences should be considered in the development of screening algorithms.

This is not surprising given the vast literature describing the field effect of cancer and malignancy-associated changes (MACs) found in normal cells in the vicinity of a lesion [21–27].
concept of the field effect is that a cancerous or precancerous lesion can influence the surrounding tissue. The normal-appearing tissue near a lesion may have subvisual biochemical changes occurring that can be detected by measuring certain parameters. Benedetto et al. [14] have shown that SILs are associated with marked changes in tissue protein thiols and disulfides, and that these differences extend to neighboring apparently normal tissue, indicating a field effect. MACs describe subtle changes in the DNA distribution in the nuclei of normal-appearing cells in the healthy tissue surrounding tumors. These subtle changes were first reported 80 years ago [26] and subsequently have been documented in tissues such as peripheral blood, bone marrow, buccal mucosa, uterine cervix, skin, pancreas, liver, and bronchial smears. In a study performed by Hutchinson et al. [25], squamous metaplastic cells were identified in monolayer preparations of cervical cells stained with thionin-Feulgen/orange II stain and classified as normal or dysplastic. A combination of three features (nuclear density, texture, and cytoplasmic density) computed for these metaplastic cells permitted 76% of the cells to be correctly classified as originating from normal or abnormal cell preparations. Quantitative measurements of nuclear texture features also show significant differences between visually normal intermediate cells from cytologically normal and abnormal monolayer cervical smears [23,24]. These factors might influence the fluorescence spectra of colposcopically normal appearing cervix.

However, there were other differences in the populations of women studied with and without a history of abnormal Pap smears that also may influence the spectral data. In particular, comparison of the age and smoking status of the women in these two studies showed some important differences. In the previous clinical study in which the algorithms were developed, the mean age was 30 with a standard deviation of 8.6, and 50% were current smokers. Five women were postmenopausal and none were on hormone replacement therapy. In contrast, the volunteers in the current study had a mean age of 37 with a standard deviation of 9, and only 10% were smokers. Eight women were postmenopausal and two were on hormone replacement therapy. The effect of these factors on the fluorescence data should be explored further.

Despite the differences observed in the spectral data, results of implementation of the previously developed classification algorithms surpassed those achieved in the study in which these algorithms were developed. In the previous clinical study, the SN/SIL algorithm correctly classified 70% of the SN samples and in the present study 87.7% of SN and 78.3% of TZ samples were correctly classified. The CN/SIL algorithm previously classified 76% of normal columnar samples correctly, and here, 95.6% of CN and 76.7% of TZ samples were correctly classified.

In conclusion, this study has shown that spectral differences exist between fluorescence emission spectra collected from colposcopically normal cervical sites in the presence or absence of cervical dysplasia. This is an important finding because diagnostic algorithms generally have been developed using spectra measured from abnormal sites and normal sites from the same patients. If such algorithms are intended for subsequent use in a screening setting, they should be tested in a normal, control population to verify their ability to classify samples correctly in a group with low disease prevalence. In this case, our diagnostic algorithms were extremely robust despite the identified spectral differences.

ACKNOWLEDGMENTS

Financial support from LifeSpex is gratefully acknowledged. The authors acknowledge the contributions of Judith Sandella, RNC, NP, MS, and Glenda Dickerson, HA, in conducting the clinical portion of this study.

REFERENCES

1. Lohmann W, Mussmann J, Lohmann C, Kunzel W. Fluorescence of the cervix uteri as a marker for dysplasia and