# Diagnostic value of nonlinear optical signals from collagen matrix in the detection of epithelial precancer

**Wei Zheng,[1,2] Dong Li,[1] Shuxia Li,[3,4] Yan Zeng,[1] Yanqi Yang,[3,5] and Jianan Y. Qu[1,6]**

[1]*Biomedical Engineering Program, Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China*

[2]*Research Center for Biophotonics, Institute of Biomedical and Health Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*

[3]*Orthodontics, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China*

[4]*Capital Medical University School of Stomatology, Beijing, China*

[5]*e-mail: yangyanq@hku.hk*

[6]*e-mail: eequ@ust.hk*

Received June 28, 2011; accepted August 2, 2011;
posted August 16, 2011 (Doc. ID 150090); published September 13, 2011

During precancer development in epithelium, neoplastic cells remodel the underlying stroma, for example, the basement membrane, capillaries, fibroblasts, and extracellular matrix. The purpose of this study is to investigate the relationship between the nonlinear optical signals from the collagen matrix in stroma and the progression of early epithelial carcinogenesis. Two-photon excited fluorescence (TPEF) and second harmonic generation (SHG) signals were measured from the stroma of hamster oral cheek pouch *in vivo*. We found that three features, including the intensity ratio of TPEF over SHG, the spatial frequency distribution, and the texture feature of SHG images, provide the quantitative identification of epithelial precancer at different pathologic stages. We demonstrated that the combination of all three features by using a support vector machine algorithm can significantly improve the accuracy in the detection of epithelial precancer. 

*OCIS codes:* 170.4580, 170.4730, 180.4315, 170.6930.

Collagen matrix, which forms the basic framework of stroma in tissues, plays an important role in the development of cancer. Accumulated evidences in cell biology and oncology reveal that the progression of epithelial carcinoma depends not only upon the proliferation of neoplastic cells themselves but also upon the interaction with surrounding stromal environment [1,2]. A variety of biochemical and biophysical changes of collagen occur with the epithelial precancer development. Thus, the detection of the signal from collagen could provide us an approach to detect the carcinoma progression. Collagen is an endogenous fluorophore in biological tissues. Its fluorescence is attributed to certain intramolecular and intermolecular cross-links. In addition, owing to its sole noncentrosymmetric structure, collagen can also produce an SHG signal. Therefore, nonlinear optical microscopy, including TPEF and SHG microscopy, could provide complementary information regarding collagen morphology and function without contrast agents. Given the advantages of inherent three-dimensional resolution and superior optical penetration, the nonlinear optical imaging of the collagen matrix has been tested to identify normal and malignant tissues in recent studies [3–7]. However, most of these studies about the stromal modification of carcinoma were conducted in excised tissue samples *ex vivo*. It is unknown whether an *in vivo* quantitative identification of precancer progression can be achieved by the detection of the nonlinear optical signals from the collagen matrix. Moreover, the accuracy of the diagnosis based on the nonlinear optical signals from collagen has not been systematically evaluated. In this study, we assess the feasibility of a nonlinear optical approach for pathological evaluation of epithelial carcinoma progression in a hamster oral carcinogenesis model. The biochemical and morphological alternations of the collagen matrix in different stages of carcinogenesis are quantitatively characterized using different signal features of the collagen matrix in stroma. Furthermore, a support vector machine (SVM) algorithm is utilized to evaluate the performance of the diagnoses based on multiple quantitative features of the collagen matrix.

A combined TPEF and SHG microscope system was instrumented as described in our previous work [8]. The excitation source is a femtosecond Ti:sapphire laser tuning at 745 nm. A linear array of photomultiplier tube connecting with a spectrograph was used as the detection system. The measuring wavelength covers 16 consecutive spectral bands from 300 to 500 nm at 13 nm interval. Therefore, the system provides the capability to detect TPEF and SHG signals and separate them in the wavelength domain simultaneously. An actuator was used to control the imaging depth. Four consecutive images with the interval depth of 3 um below the basal lamina of epithelium were collected for data analysis in each site. The 7,12-dimethylbenz(a)anthracene (DMBA) treated hamster cheek pouch carcinogenesis model was developed following the standard procedure [9]. The right cheek pouch was treated with either 0.5% DMBA in mineral oil or just mineral oil three times per week for about 20 weeks. A total of 34 (11 controlled and 23 treated) male Golden Syrian hamsters were evaluated in this study and 1–2 sites in each animal were chosen for the measurement. The pathologic assessment was performed by a certified pathologist (S.L.). All samples were divided into four categories: normal ($n = 20$), low-grade precancer (mild dysplasia, $n = 30$), mid-grade precancer

 

(moderate dysplasia, $n = 7$), and high-grade precancer (severe dysplasia and carcinoma *in situ*, $n = 7$). All measurements were carried out in anesthetized animals and controlled to be finished in 3 hours for each hamster. The breath and body temperature of the animals were monitored during the whole process of the experiment.

Figures 1(a)–1(c) show the SHG, TPEF, and merged SHG/TPEF images from a normal stroma, respectively. Collagen clusters are interlaced with each other as shown in the SHG image, and single collagen fibers can be identified in the clusters. Albeit fluorescent, the fine structure of collagen fibers cannot be observed in the TPEF image. This may be due to the fact that the collagen fluorescence originates from the inter-cross-link and intra-cross-link between collagen fibers. In the TPEF image, stromal cells could be found, whereas they do not appear in the SHG image. The TPEF and SHG spectra of collagen are displayed in Fig. 1(d). As can be seen, the intensity of SHG is about 2 orders of magnitude greater than that of the TPEF signal. The TPEF spectral peak of collagen has about 10 nm blueshift and the spectra is broad as compared with the TPEF spectra of epithelial cells. The SHG and TPEF signals arise from different mechanisms and the extracellular matrix remodeling associated with the epithelial cancer development has different effects on TPEF and SGH, respectively. Therefore, the intensity ratio of TPEF over SHG could reflect the biochemical property of the collagen matrix in the tissue at different pathological stages and has the potential to classify normal and precancerous tissues. We integrated the TPEF signal from the pixels, which exhibited SHG signals in the same image, and divided the integrated TPEF by total SHG signal. The statistical result is shown in Fig. 1(e). As can be seen, the ratio is about $0.045 \pm 0.009$ in normal tissues, whereas the value increases to $0.078 \pm 0.015$ in high-grade precancerous tissues. This can be partially explained as in the stroma of precancer, the up-regulation of several proteases, such as metalloproteinase, will degrade the collagen fiber into fibril [2]. The decomposition of the noncentrosymmetric structure of collagen yields the decrease of SHG signal.

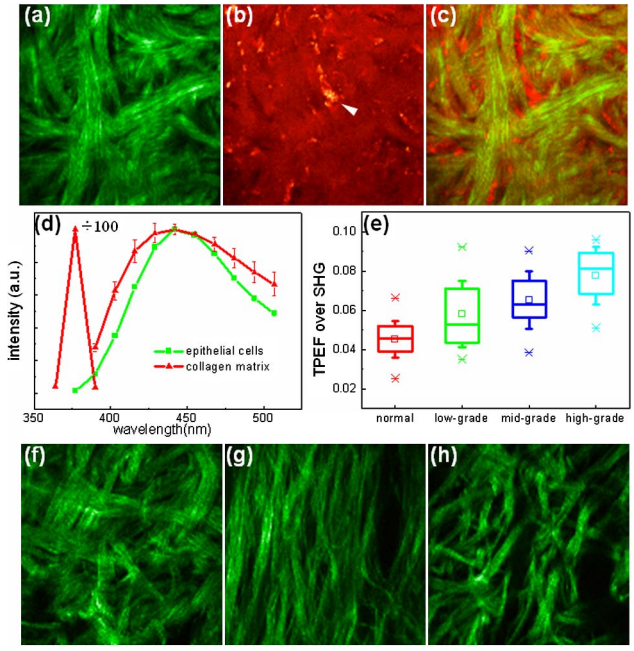


Fig. 1. (Color online) Nonlinear optical signals from collagen matrix. (a)–(c): Nonlinear optical images of normal collagen matrix based on (a) SHG signal, (b) TPEF signal, and (c) merged signal. A stromal cell (white arrowhead) is presented in TPEF image. (d) TPEF and SHG spectra of collagen matrix. (e) Statistical result of the ratio of TPEF over SHG. The line within each notch box represents the median, and the lower and upper boundaries of the box indicate first and third quartiles (25% and 75%), respectively. Error bars (whiskers) represent the standard deviation (SD). The square represents the mean, and the upper and lower asterisks indicate the maximum and the minimum, respectively. (f)–(h): Representative SHG images from (f) low-grade, (g) mid-grade, and (h) high-grade precancers. Sampling area of each image is $100\,\mu m \times 100\,\mu m$.

As SHG intensity was much stronger than TPEF of collagen and the fine structure of the collagen matrix could only be identified in SHG images, the SHG images were utilized for morphological evaluation of carcinogenesis in the following studies. Representative SHG images of the collagen matrix from different stages of cancer growth are presented in Figs. 1(f)–1(h). Morphological alternations in the collagen assembly are observed as compared with the normal stroma. Specifically, during the progression of carcinoma, the collagen clusters are replaced by aggregated taut bundles and more areas appear to be devoid of collagen matrix. The interlaced structure of the collagen matrix in normal tissue becomes more aligned in precancerous tissues. To quantitatively assess these morphological alternations, we implemented two types of methods to analyze the SHG images. One was based on the gray-level co-occurrence matrix (GLCM) texture analysis and the other relied on the fast Fourier transform (FFT) analysis of SHG images.

GLCM texture analysis is based on second-order statistics (i.e., statistics given by pairs of pixels) of gray-level pixel distribution [10]. The joint probability distribution of pairs of pixels is defined as the GLCM and the different texture features are deduced from the GLCM. One of these texture features is the GLCM correlation. It evaluates the linear dependence between each pixel pair in the image. The higher value indicates that there are more similar structures in the image. Therefore, it can provide an estimate of the organization and structure of the collagen matrix. In this study, we utilize the GLCM texture analyzer module (v0.4) in ImageJ software (NIH) to analyze the SHG images. The statistical result is shown in Fig. 2(a). The correlation value of normal tissue is

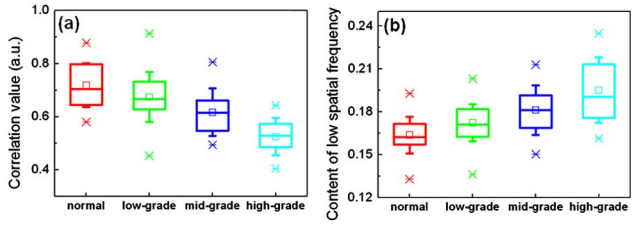


Fig. 2. (Color online) Quantitative assessment of collagen morphological alternations during carcinoma progression. (a) Statistical result of GLCM correlation measured from different groups of epithelial tissues. The GLCM texture analyzer module (v0.4) in ImageJ software (NIH) was used and the pixel distance was set as 1. (b) Statistical result of the content of lower spatial frequency ($<10\,\mu m^{-1}$, excluding the DC component) in image. The legend is the same as Fig. 1(e).

higher than that of precancerous tissues and this value decreases with the progression of carcinoma. The difference between each group is significantly different ($p < 0.05$, $t$-test). This result is consistent with the qualitative changes that the "diffusive" collagen matrix in normal tissue was replaced by the aggregated fiber bundles in precancer.

Unlike the GLCM analysis, which is based on the statistics of pixel intensity distribution, FFT analysis focuses on the spatial frequency of an image. We performed FFT analysis on the SHG images and observed that the lower spatial frequency increases, whereas the higher spatial frequency decreases with the progression of carcinoma. Figure 2(b) shows the statistical result of the content ratio of the lower spatial frequency components ($<10\,\mu m^{-1}$, excluding the DC component). As can be seen, the content of the lower spatial frequency from precancerous tissues is greater than that from the normal tissue. The difference between the normal and each pathologic stage is again significantly different ($p < 0.05$, $t$-test). The high spatial frequency is attributed to the fine structure in the image, such as the single collagen fiber. The content of higher spatial frequency decreases with the progression of carcinoma because collagen fibers degrade in precancerous tissues.

Although the above quantitative assessments can identify the normal from the precancerous tissues, it is expected that the combination of all these assessments may improve the accuracy of the diagnosis because different features of the collagen matrix provide complementary information. In this study, we chose the SVM method to combine the three assessments because the SVM can classify data measured from multiple signal sources (multivariate data) with relatively small sample sizes [11]. In the process of SVM, the multivariate data are represented as sample points in the input space. The SVM maps these points into a feature space with higher dimensions, usually in an infinite-dimensional space, by linear or nonlinear transformation. This mapping makes the distribution of sample points form a wide gap between each group. Then an optimal hyperplane, which has the largest distances to the sample points of different groups, is revealed to describe the criterion of classification. We used a nonlinear SVM algorithm (LIBSVM V3.1 [12], radial basis function kernel) to classify tissues based on the correlation between above three quantitative features and the pathologic states of the tissue. A receiver operating characteristics (ROC) curve is used to visualize the accuracy of each classification and to select different classifiers based on their performance. An exhaustive search based on a fivefold cross-validation was performed to maximize the accuracy for screening purpose, which identifies the abnormal from normal tissues, and for diagnosing propose, which differentiates the high-grade and mid-grade precancer from low-grade and normal tissues, respectively. Although it was reported that the optimal number of folds for cross-validation is 10, we compromised to choose the number of folds as five because the fivefold cross-validation is commonly used in the studies with relatively small sample volumes and should be more precise [13]. The cross-validation results are shown in Fig. 3. Clearly, the combination of three assessments significantly improved the performance of diagnosis. Specifically, the highest accuracy of fivefold cross-validation is 77.2% and 85.4% for screening purpose and diagnostic purposes, respectively.

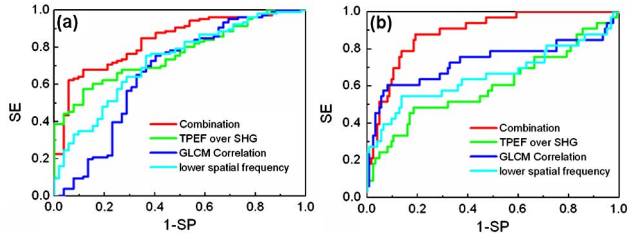


Fig. 3. (Color online) ROC curves of SVM algorithm for (a) screening purpose and (b) diagnostic purpose.

In summary, we found that the nonlinear optical signals from stroma can provide the quantitative identification of epithelial precancer at different pathologic stages. The combination of the intensity ratio of TPEF over SHG, the spatial frequency distribution, and the texture feature of SHG images can significantly improve the accuracy in the detection of epithelial precancer. The capability of quantifying collagen alternation with epithelial carcinoma progression renders nonlinear optical microscopy a promising tool for future noninvasive clinic application.

## References

1. J. A. Joyce and J. W. Pollard, Nat. Rev. Cancer **9**, 239 (2009).
2. M. Egeblad and Z. Werb, Nat. Rev. Cancer **2**, 161 (2002).
3. S. Lin, S. Jee, C. Kuo, R. Wu, W. Lin, J. Chen, Y. Liao, C. Hsu, T. Tsai, Y. Chen, and C. Dong, Opt. Lett. **31**, 2756 (2006).
4. M. W. Conklin, J. C. Eickhoff, K. M. Riching, C. A. Pehlke, K. W. Eliceiri, P. P. Provenzano, A. Friedl, and P. J. Keely, Am. J. Pathol. **178**, 1221 (2011).
5. O. Nadiarnykh, R. B. LaComb, M. A. Brewer, and P. J. Campagnola, BMC Cancer **10**, 94 (2010).
6. N. D. Kirkpatrick, M. A. Brewer, and U. Utzinger, Cancer Epidemiol Biomark. Prev. **16**, 2048 (2007).
7. R. M. Williams, A. Flesken-Nikitin, L. H. Ellenson, D. C. Connolly, T. C. Hamilton, A. Y. Nikitin, and W. R. Zipfel, Translational Oncology **3**, 181 (2010).
8. W. Zheng, Y. Wu, D. Li, and J. Y. Qu, J. Biomed. Opt. **13**, 054010 (2008).
9. S. Andrejevic, J. Savary, C. Fontolliet, P. Monnier, and H. Van Den Bergh, Int. J. Exp. Pathol. **77**, 7 (1996).
10. R. M. Haralick, Proc. IEEE **67**, 786 (1979).
11. C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995).
12. C. Chang and C. Lin, "LIBSVM: a library for support vector machines," http://www.csie.ntu.edu.tw/~cjlin/libsvm.
13. L. Breiman and P. Spector, Int. Stat. Rev. **60**, 291 (1992).