Label-free serum ribonucleic acid analysis for colorectal cancer detection by surface-enhanced Raman spectroscopy and multivariate analysis

Yanping Chen
Gang Chen
Shangyuan Feng
Jianji Pan
Xiongwei Zheng
Ying Su
Yan Chen
Zufang Huang
Xiaoqian Lin
Fenghua Lan
Rong Chen
Haishan Zeng
Label-free serum ribonucleic acid analysis for colorectal cancer detection by surface-enhanced Raman spectroscopy and multivariate analysis

Yanping Chen, a,b* Gang Chen, a Shangyuan Feng, a Jianji Pan, b Xiongwei Zheng, b Ying Su, b Yan Chen, b Zufang Huang, a Xiaoqian Lin, a Fenghua Lan, a Rong Chen, a and Haishan Zeng d

a Fujian Normal University, Key Laboratory of Optoelectronic Science and Technology for Medicine, Ministry of Education and Fujian Provincial Key Laboratory for Photonics Technology, Fuzhou 350007, China
b Teaching Hospital of Fujian Medical University, Fujian Provincial Tumor Hospital, Fuzhou 350014, China
c Research Center for Molecular Diagnosis of Genetic Diseases, Fuzhou General Hospital, Fuzhou 350025, China
d British Columbia Cancer Agency Research Centre, Integrative Oncology Department—Imaging Unit, Vancouver, British Columbia V5Z 1L3, Canada

Abstract. Studies with circulating ribonucleic acid (RNA) not only provide new targets for cancer detection, but also open up the possibility of noninvasive gene expression profiling for cancer. In this paper, we developed a surface-enhanced Raman scattering (SERS) platform for detection and differentiation of serum RNAs of colorectal cancer. A novel three-dimensional (3-D) Ag nanofilm formed by dry MgSO4 aggregated silver nanoparticles, Ag NP, as the SERS-active substrate was presented to effectively enhance the RNA Raman signals. SERS measurements were performed on two groups of serum RNA samples. One group from patients, n = 55 with pathologically diagnosed colorectal cancer and the other group from healthy controls, n = 45. Tentative assignments of the Raman bands in the normalized SERS spectra demonstrated that there are differential expressions of cancer-related RNAs between the two groups. Linear discriminant analysis, based on principal component analysis, generated features can differentiate the colorectal cancer SERS spectra from normal SERS spectra with sensitivity of 89.1 percent and specificity of 95.6 percent. This exploratory study demonstrated great potential for developing serum RNA SERS analysis into a useful clinical tool for label-free, noninvasive screening and detection of colorectal cancers. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.JBO.17.6.067003]

Keywords: surface-enhanced Raman scattering; 3-D Ag nanofilm; colorectal cancer; circulating nucleic acids; cancer screening.

Paper 11799 received Dec. 31, 2011; revised manuscript received Apr. 14, 2012; accepted for publication Apr. 30, 2012; published online Jun. 6, 2012.

1 Introduction

Colorectal cancer (CRC), is the fourth most common cancer in men and the third most common cancer in women, worldwide. Globally, incidence and mortality rates of CRC are on the rise.1 Until more effective prevention measures are instituted, CRC mortality will eventually increase. Pre-symptomatic detection of the earliest stage cancers yields an excellent prognosis, with a five year survival in excess of 97 percent.2 Ideal features of a noninvasive screening test for CRC is highly sensitive, highly specific, fast and widely accessible. However, at present conventional methods such as fecal occult blood testing (FOBT), barium enema, total colonoscopy and conventional tumor markers, cannot meet all of the desired criteria of an ideal screening tool.3 Since the discovery of circulating nucleic acids by Mandel and Metais in 1948, scientists are committed to the early warning application of circulating nucleic acids in the noninvasive diagnosis of clinical disease, including cancer.4 In recent years, many researchers speculated that circulating RNA markers not only provide new targets for cancer detection, but also open up the possibility of noninvasive gene expression profiling for cancer.5,6

Raman spectroscopy has been widely applied to detect cancerous tissues from various body sites such as colon, lung, breast, bladder, prostate, cervix, larynx, naso-pharynx, and skin.7,8 However, the inherently small cross-section, e.g., 10−30 cm2 per molecule, and the auto-fluorescence interference hinder its practical applications. With the development of surface-enhanced Raman scattering (SERS) technique, the Raman signal can be enhanced enormously,10 as high as the order of 1014 to 1015. Moreover, SERS has several advantages over regular Raman spectroscopy such as relatively rapid measurement and only small amounts of material, micrograms to nanograms, with minimum sample preparation are required. Now, detection of cancer is one of the uppermost applications of SERS spectroscopy in biomedicine. Recently, oncology applications are focused on developing SERS based immunoassay,11 which is a powerful tool for multiplex high-throughput screening tumor markers in biochips.12 Another promising way of using SERS is label-free analysis through direct binding onto a metallic, enhancing nanostructure. Numerous examples have been reported on the application of colloidal dispersions for ultrasensitive detection and characterization of micro-biomolecules such as DNA and RNA.13,14 However, none of these studies have been done for application of RNA SERS for cancer detection.15,16

Herein, we demonstrate the facile fabrication of a novel three-dimensional (3-D) silver nanofilm as SERS substrates

*These authors contributed equally to this work.

Address all correspondence to: Shangyuan Feng, Fujian Normal University, Key Laboratory of Optoelectronic Science and Technology for Medicine, Ministry of Education and Fujian Provincial Key Laboratory for Photonics Technology, Fuzhou 350007, China. Tel: (86) 591-87160252; Fax: (86) 591-83465373; E-mail: syleng@fjnu.edu.cn

0091-3286/2012/$25.00 © 2012 SPIE
for serum RNA analysis. Principal component analysis (PCA), combined with linear discriminate analysis (LDA), were used to classify the serum RNA SERS spectra obtained from healthy subjects and cancer patients. To the best of our knowledge, this is the first report on SERS combined with PCA-LDA analyses for the purpose of CRC detection at the genetic level, RNA.

2 Materials and Methods

2.1 Blood Sample Collection

In this study, blood samples were collected from 100 subjects which consisted of 55 CRC patients and 45 healthy volunteers. All patients were from Fujian Provincial Tumor Hospital and had similar ethnic and socioeconomic backgrounds. In addition, all patients were untreated patients with initial pathological diagnosis. The mean age for the CRC group was 57.4 years and the control group 50.5 years. The CRC group consists of 25 cases of stage I-II lesions and 30 cases of stage III-IV lesions. All the study subjects provided written consent and ethics permission was obtained for the study.

2.2 Preparation of the RNA Samples

Serum was separated from the whole blood within 2 h after blood collection and was immediately stored at −80°C.17 Total RNA was extracted from 250 μL of serum using Blood/Liquid Sample Total RNA Rapid Extraction Kit, Spin-column manufactured by BioTeke Corporation, China, according to the manufacturer’s protocol. Lysis Buffer RLS reagent was always added to serum sample based on the volume ratio 3:1. The final elution volume was 30 μL of diethyl pyrocarbonate (DEPC) treated water. The concentrations of all RNA samples were quantified by NanoDrop ND 1000, Nano-drop, Wilmington, Delaware, USA. The concentration of total RNA extracted from serum ranged from 18.8 to 56.4 ng/μL.

2.3 Preparation of SERS-Active Substrate

Stable silver colloid solutions were prepared by the aqueous reduction of silver nitrate with sodium carbonate, according to Lee and Meisel.18 Briefly, 90 mg of silver nitrate was suspended in 500 mL of distilled water, at 90°C, and heated rapidly to the boiling point using stirring, 10 mL of a 1.0% solution of sodium citrate was then immediately added. Boiling was continued for 60 min, the heat source was removed and the solution was cooled naturally. The silver colloids were kept in the dark at room temperature and remained stable for at least one month. The silver colloidal solution was concentrated by centrifugation at 10,000 rpm for 10 min, discarding the supernatant and bringing the final concentrated colloids for use. Absorption spectra of silver colloids were recorded from 200 to 1200 nm, with a resolution of 2 nm, using a PerkinElmer Lambda 950 Spectrophotometer which was manufactured by Waltham, MA, USA. The particles diameter ranged from 70 to 80 nm with an average diameter of 75 nm.

To fabricate a SERS-active substrate for RNA analysis, the silver colloidal solution was initially concentrated by centrifugation at 10,000 rpm for 10 min, discarding the supernatant and keeping the final concentrated colloids for use. 5 μL of concentrated Ag colloids were then dropped on the surface of a rectangular aluminum plate. Next, 2 μL of a 0.1 M MgSO₄ aqueous solution were added into the concentrated Ag colloids, immediately the silver colloid strongly aggregated. Until the MgSO₄-aggregated colloids were dried naturally at room temperature, as shown in Fig. 1(b), 5 μL of RNA sample were then dropped into the dry aggregated colloids with a pipette. The mixture was incubated for 3 min at room temperature before SERS measurements. The scanning electron microscope (SEM), observation of nonaggregated Ag NP and MgSO₄-aggregated Ag NP on the surface of aluminum substrate were performed on a JSM 7500F SEM, manufactured by JOEL, Japan.

2.4 SERS Measurements

The SERS spectra were recorded with a confocal Raman microspectrometer, manufactured by Renishaw, Great Britain. A 785 nm near-infrared (IR) diode laser was used as the excitation source. The laser power was set to 50 percent, where the power at the sample surface was measured to be ~5 mW. The spectra were collected using a Leica 20x objective with a spectral resolution of 2 cm⁻¹, the detection of Raman signal was carried out with a Peltier cooled charge-coupled device, CCD, camera. Extended scan spectra, with a spectral range from 500 to 1500 cm⁻¹, were acquired using an integration time of 10 s. This region is known as the molecular fingerprint region and provides the most information on the biological constituents of a sample including RNA.19,20 Three spectra were collected, from different locations, for each RNA sample to ensure representative sampling and incorporate spot-to-spot variability in the signal. DEPC-treated water was used as a control. The SERS spectra were processed with the software package WIRE 2.0, produced by Renishaw.

2.5 Data Analysis

An automated algorithm for autofluorescence background removal from the original SERS data was applied to the measured raw data to extract pure Raman spectra.21 Then, all of the pure SERS spectra were normalized to the integrated area under the curve. This reduces the spectral intensity variations between different spectra and facilitates more accurate spectral shape analysis. The SPSS software package, produced by SPSS Inc. in Chicago, was used for data analysis. To test the capability of RNA SERS spectra for differentiating cancer from normal, PCA, combined with LDA was performed on the normalized RNA SERS spectra. For the validation of the PCA-LDA model, the leave-one out cross-validation procedure was used.

3 Results

3.1 SEM Image and UV-Visible Absorption Spectra

From the SEM images, we noted that the SEM image of MgSO₄-aggregated Ag NP substrate, Fig. 1(b) shows a very different morphology as compared to that of nonaggregated Ag NP, Fig. 1(a). The aggregated Ag NP formed an obvious 3-D structure, while nonaggregated Ag NP showed tightly packed configuration. The MgSO₄-aggregated Ag NP, that form the 3-D structures, are actually thin Ag nanofilm stacked by closely adjacent nanoparticles. Figure 1(c) shows the UV-visible absorption spectra of RNA, curve a, the pure Ag sol, curve b, and the mixture of Ag colloid and RNA, curve c, as well as the mixture of Ag colloid, MgSO₄ and RNA solution, curve d, respectively. Curve a indicates that the strong absorption band of RNA is at about 269 nm and the intense absorption band of pure silver colloids, curve b, at 417 nm. When RNA was added into Ag sol,
the intense Ag nanosphere absorption band was reduced to some extent, curve c, while a subtle new plasma resonance absorption peak in the longer near IR wavelength region appeared. However, when MgSO₄ was added into the mixture of Ag colloid and RNA, a coupled localized surfaced plasma (LSP) peak emerged at 700 to 900 nm, in addition to that for isolated particles at about 417 nm, curve d. Compared with suspended Ag colloid, the extinction maximum of the dried AgNP is red-shifted from 417 to 480 nm due to inter-particle coupling since they are tightly packed, as shown in Fig. 1(d). This behavior is usually believed to originate from the surface plasma resonance of aggregated silver nanoparticles.

3.2 RNA SERS Spectra

We have recorded the SERS spectra of RNA from patients and healthy subjects, and studied the enhancement effects of the external aggregating agent MgSO₄ to the SERS signals of RNA samples. As shown in Fig. 2(c), only a few RNA SERS peaks can be observed with silver colloidal solution in the absence of MgSO₄. When an RNA sample was dropped into the dry Ag colloid, the intensity of Raman bands were clearly increased and some new Raman peaks appeared, as shown in Fig. 2(b). However, when an RNA sample was dropped into the dry Ag colloids aggregated with MgSO₄, the intensity of many dominant Raman bands were increased dramatically and many new Raman peaks appeared, as shown in Fig. 2(a), indicating that the enhancement effects of dry MgSO₄-aggregated Ag colloid to the RNA Raman signals were the highest. Figure 2(d) shows the background Raman signals of DEPC-treated water mixed with Ag sol and Fig. 2(e) shows the background Raman signal of Ag colloid mixed with MgSO₄. We see almost no interference signals in our interested spectral range.

Figure 3(a) compares the normalized mean SERS spectra of serum RNAs obtained from 55 patients with CRC versus 45 healthy subjects. The shaded areas represent the standard deviations (SD) of the means. The gray shaded area indicates the SD of the normal mean spectra, while the green shaded area represents the SD of the CRC mean spectra. From the low standard deviations, we can see that the SERS signals are relatively stable and very few variations are observed with the change of the measurement location and samples, indicating the good reproducibility of SERS spectra.

In general, the SERS spectra of serum RNA from cancer and normal groups share many similar peaks, except the obvious differences in the peaks at 573, 714, 859, 1008, 1033, 1240, 1277, 1316, and 1391 cm⁻¹. The intensities of SERS peaks at 714,
860, 1008, 1240, and 1391 cm$^{-1}$ are significantly greater for normal than for tumor, while SERS bands at 573, 1033, 1277, and 1316 cm$^{-1}$ are more intense in cancer group. These normalized intensity differences can be viewed more clearly in the difference SERS spectrum between cancerous and normal serum RNA samples, as shown at the bottom in Fig. 3(a). Figure 3(b) shows a comparison of the mean intensities and standard deviations of the selected SERS peaks with significant differences, as evaluated by using the Student’s $t$-test analysis, $p < 0.05$, between the two groups of serum RNAs sample. The difference spectrum reveals the changes of prominent SERS peaks occurring in cancer serum RNA, confirming a potential role of serum RNA SERS for CRC diagnosis.

Moreover, we have confirmed that the different RNA concentrations obtained from the same sample have minimal effect on the SERS spectral shape after normalization. Equal volumes, 5 $\mu$L, instead of equal concentrations of RNA were dropped into the MgSO$_4$-aggregated colloids. Though the intensity of SERS signals was different with the different RNA concentrations, the difference of initial concentrations from the same sample could be eliminated after normalization. As seen in Fig. 4(a), the standard deviation is large without normalization. After normalization, shown in Fig. 4(b), the standard deviation became very small, which indicated that the concentration difference of the same RNA samples could nearly be eliminated.

### 3.3 Statistic Analysis

In this section, we explore the potential of RNA SERS to discriminate between the two groups. Multivariate models, based on PCA and LDA, were performed on all of the normalized SERS spectra. In this analysis, 55 cancerous serum RNA SERS spectra were compared with 45 normal serum RNA SERS spectra. Independent-sample $t$-test on all the PC scores, comparing normal and cancerous groups, displayed that there were three PCs that were most diagnostically significant, $p < 0.05$, for discriminating normal group and cancerous group. Figure 5(a) shows the result of a three dimensional scatter plot of PC scores, PC2, PC6, and PC12 as the three axes, for the healthy and CRC groups. It is clearly seen that they are distributed in separate clusters. Figure 5(b) presents the LDA scores of normal controls compared to CRC groups. A clear distinction can be made between the two groups, with few overlaps, suggesting that the SERS spectra of the CRC group can be differentiated from the healthy group via multivariate analysis methods. This method showed very good prediction accuracy. 49 out of 55 patients with CRC were predicted as cancerous patients, 89.1 percent, and 43 out of 45 healthy controls were predicted as healthy, 95.6 percent, as well as the overall diagnostic accuracy was 92.35 percent. To further assess the performance of the PCA-LDA-based diagnostic algorithm for CRC diagnosis, the relative operating characteristic (ROC) curve, shown in Fig. 6, was generated at different threshold levels. The area under the ROC curve was 0.96. These results demonstrated that the serum RNA SERS spectra could potentially be used for CRC detection and screening with good sensitivity and specificity.

### 4 Discussions

#### 4.1 SERS Substrate

It is interesting to find that after drying naturally at room temperature, the silver colloids aggregated with the addition of MgSO$_4$.
MgSO₄ and formed analogous to 3-D Ag nanofilm on the surface of aluminum substrate, which produced great enhancement effects to the Raman signals of RNA samples. The 3-D Ag nanofilm was formed by repetitive stacking of aggregated Ag NP. Close examination of the SEM image, of Fig. 1(b) reveals that there are ample nano-cavities and nano-gaps formed by the stereotypical stacking of closely adjacent nanoparticles. Recent studies have shown an extremely strong local field enhancement in the gap between two, closely spaced, silver nanoparticles. Therefore, the 3-D SERS substrate may generate a lot of effective electromagnetic hot spots and strong SERS responses over the surface area. When RNA molecules were absorbed and located in the nano-gap of the closely adjacent Ag particles, Raman signal could be largely enhanced under the coupled LSP resonances. Moreover, the absorption area of the stereotypical aggregated particles is much larger than that of nonaggregated Ag nanoparticles, which can improve the adsorption of the biological macromolecules onto the surface of the Ag nanofilm. The presence of nanocavities on the MgSO₄-aggregated Ag colloids was an important factor for SERS enhancement of RNA molecules. All of the above are probably responsible for the observed higher electromagnetic enhancement effect and additional chemisorption. The total electromagnetic enhancement may reach 10¹¹ in hot nanojunctions. Thus, when RNAs were closely adsorbed onto the 3-D Ag nanofilm substrate, the Raman scattering of these molecules would be enhanced enormously, while the SERS signal of RNAs was poor on the nonaggregated Ag colloid.

4.2 SERS Spectra

The results of our preliminary study indicated that there were significant differences in SERS spectra for serum RNA obtained from patients with CRC versus from healthy subject, suggesting great potential for SERS in CRC detection and screening applications. In many cases, Raman bands can be associated with the vibration of a particular chemical bond or a single functional group in the molecule, so distinctive SERS spectral features and intensity differences for cancer and normal groups could reflect molecular changes associated with malignant transformation. To better understand the molecular components and relative content of serum RNA samples, Table 1 lists tentative assignments for the observed SERS bands, cm⁻¹, according to the literature.

The major SERS spectral difference between the RNA of normal and cancer blood serum were in the relative intensities of the bands 573, 714, 859, 1008, 1033, 1240, 1277, 1316, and 1391 cm⁻¹. These changes probably reflected variations on tumor-specific RNA relative contents and conformations in the process of occurrence and development of cancer. The detailed mechanisms for these spectral changes deserve further investigation. According to Silva et al. epithelial tumor RNA,
CK19, and CEA are highly expressed in the plasma of colon cancer patients, but low in healthy plasma. In addition, this molecular event is associated with advanced stages and circulating tumor cells. Our own work about nasopharyngeal and gastric cancer detection based on plasma SERS analysis also reported that plasma samples from cancer patients have higher relative concentrations of nucleic acids bases as compared with control samples.25,31 In addition to differential expression, circulating RNA alterations might be detectable ahead of cancer diagnosis, and increasing tumor-relative RNAs content might correlate with higher histological grade and worse prognosis.30 This put forward the possibility of exploiting RNA specific SERS signals from serum/plasma samples as novel and noninvasive biomarkers for diagnosis, estimating prognosis, predicting therapeutic efficacy, maintaining surveillance following surgery, and forecasting disease recrudescence of cancer.

4.3 Statistical Analysis

From the 3-D scatter plot of PC scores in Fig. 4(a) it is clear to see that the healthy and cancer groups are distributed in separate areas. The normal group forms one cluster and the cancer group forms the other. Also, as can be seen in the figure of LDA scores in Fig. 4(b), a clear distinction can be made between the two groups, with the exception of six outlier values from the cancer group mixed with the normal group, and two outlier values from the normal group mixed with the cancer group, which means we can discriminate between the SERS spectra of the CRC group and the healthy control group. The sensitivity of this technique for identifying CRC from the serum RNA SERS analysis is 89.1 percent, and the specificity is 95.6 percent. Our result demonstrated that the sensitivity of serum RNA detection based on SERS spectroscopy is higher than conventional methods. For instance, FOBT has a large variation in sensitivity ranges from 26 percent to 69 percent. Flexible sigmoidoscopy (FS), identifies far less than half premalignant or malignant lesions. The sensitivity of barium enema is 82.9 percent. Virtual colonoscopy continues to show widely varying results with sensitivities for detecting polyps ≥10 mm ranging from 55 percent to 92 percent.

Meanwhile, from Fig. 4(a), we can see that the regional distribution of the CRC group is wider than for the healthy control group. The most important reason is cancer derived serum RNA contents may be variable from patient to patient due to the heterogeneity of cancer development. Furthermore, recent studies found that the integrity of circulating RNA was significantly reduced in cancerous patients, and the RNA integrity was different with the different stages.32 These reasons are probably responsible for the observed phenomenon that the regional distribution of the CRC group on the PCA plot is much larger than for the healthy control group.

5 Conclusions

In this paper, we report a new methodology, based on SERS technology, to analyze serum RNAs obtained from CRC patients and healthy volunteers. The novel 3-D nanofilms, formed by MgSO₄-aggregated Ag colloid with ample nanocavities, is an excellent SERS substrate for the detection of serum RNA. Using RNA SERS spectroscopy combined with PCA-LDA multivariate analysis, we were able to differentiate CRC patients from healthy subjects with very good diagnostic sensitivity, 89.1 percent, and specificity, 95.6 percent. The results of this preliminary study demonstrated tremendous promise for

<p>| Table 1 SERS peak positions and vibrational mode assignments. |
|---------------------------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Peak (cm⁻¹)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>527</td>
<td>U</td>
</tr>
<tr>
<td>573</td>
<td>C, G</td>
</tr>
<tr>
<td>678</td>
<td>G</td>
</tr>
<tr>
<td>714</td>
<td>A</td>
</tr>
<tr>
<td>736</td>
<td>A</td>
</tr>
<tr>
<td>821</td>
<td>O-P-O stretch DNA</td>
</tr>
<tr>
<td>859</td>
<td>Phosphate group</td>
</tr>
<tr>
<td>902</td>
<td>Phosphodiester, Deoxyribose</td>
</tr>
<tr>
<td>1008</td>
<td>Stretching of C-O and C-C</td>
</tr>
<tr>
<td>1033</td>
<td>Stretching of C-O and C-C</td>
</tr>
<tr>
<td>1078</td>
<td>C-C or PO2 stretch (nucleic acid)</td>
</tr>
<tr>
<td>1130</td>
<td>CC skeletal stretch transconformation</td>
</tr>
<tr>
<td>1240</td>
<td>Stretching of asymmetric phosphate [PO2- [asym.]]</td>
</tr>
<tr>
<td>1277</td>
<td>Nucleic acids and phosphates</td>
</tr>
<tr>
<td>1316</td>
<td>G</td>
</tr>
<tr>
<td>1399</td>
<td>A, G</td>
</tr>
<tr>
<td>1391</td>
<td>CH rocking</td>
</tr>
<tr>
<td>1446</td>
<td>CH2 deformation</td>
</tr>
<tr>
<td>1467</td>
<td>Deoxyribose</td>
</tr>
</tbody>
</table>

developing SERS based serum RNA analysis into a clinical tool for noninvasive detection and screening of CRC. Further work needs to be carried out for large-scale validation of the reliability of this new cancer detection method, including testing on adenoma with various degree of dysplasia and other systematic cancers, such as lung cancer.

Acknowledgments

We thank Longfeng Ke for excellent technical assistance, Zheng Lin, Jing Wang and Yun Yu for statistical analysis, and Minhua Hu for providing samples. This work was supported by the National Natural Science Foundation of China (Nos. 11104030, 61178090, 81101110), the program for Changjiang Scholars and Innovative Research Team in University (IRT1115), the Project of the Educational Office of Fujian Province (No. JA11055) and the Canadian Institutes of Health Research International Scientific Exchange Program.

References