Multivariate analysis of tissues Raman spectra using regression methods

A A Lykina\textsuperscript{1}, D N Artemyev\textsuperscript{1}, V I Kukushkin\textsuperscript{2}, I A Bratchenko\textsuperscript{1}, S T Avraamova\textsuperscript{3} and N S Aleksandrov\textsuperscript{3}

\textsuperscript{1}Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
\textsuperscript{2}Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001
\textsuperscript{3}Sechenov University, Trubetskaya st. 8-2, Moscow, Russia, 119991

e-mail: zuum63@mail.ru

Abstract. This work is devoted to the study of intact and tumor kidney tissues spectral properties. 532 nm, 785 nm, and 1064 nm lasers were used to stimulate Raman spectra and autofluorescence. Visually and morphologically unchanged tissues of human kidneys as well as tumor tissue were examined. Raman spectra have specific spectral markers for each tested subject. The analysis of the spectral characteristics was carried out with using projection on latent structures method to highlight the spectral features between the tissues.

1. Introduction
At the present time, kidney cancer is one of the leading problems in the modern oncology, associated with a high increase of sickness rate and difficulty of its diagnosis [1]. Histology is utilized to investigate pathological processes in biological tissues [2]. Usually, during the surgery, the biopsy of the suspicious tissue is taken to the histopathology laboratory where it is tested using the standard staining techniques and optical microscopy. Unfortunately, this method of cancerous tissue identification is a time consuming procedure and cannot be performed directly during the surgery. Spectral methods have become widely used for tissue analysis [3-6]. They provide an opportunity to measure the optical properties of substances and tissues. One of the most common optical methods is autofluorescence (AF) [3], Raman spectroscopy (RS) [4,5]. The registered spectra of tissues carry features, which differ depending on the component composition of biomolecules in the tissue. Analysis of experimental data and assessment of the possibility of tissues type separation was carried out using regression analysis.

2. Material and methods
2.1. Experimental setup
Registered spectra were collected by three experimental devices: portable spectrometers from EnSpectr (532 nm and 1064 nm) and measuring systems that include: LML-785.0RB-04, commercial Raman probe (Inphotonics RPB785) and spectrograph Shamrock SR-500i-D1-R with deeply cooled digital camera Andor iDus DU416A-LDC-DD (air-cooled up to -70 ° C). The optical scheme of the
Detailed information about the experimental setup presented in paper [7]. The parameters of the experimental setups are presented in Table 1. Experimental devices for Raman scattering registration is presented in Fig. 1.

![Figure 1. Optical scheme of the experimental setup (LML-785.0RB-04 laser) (LPF- long-pass filter, BPF - band-pass filter, DM- dichroic mirror, M- deviating mirror, L1,3- lens).](image)

| Registered signal | Name of the laser     | Excitation wavelength, nm | Laser power, mW | Optical resolution of the spectrograph, cm⁻¹ |
|-------------------|-----------------------|---------------------------|-----------------|------------------------------------------|
| AF                | EnSpectr R532         | 532                       | 30              | 2                                        |
| RS                | LML-785.0RB-04        | 785                       | 200             | 1                                        |
| RS                | EnSpectr R1064        | 1064                      | 300             | 10                                       |

The exposure time was 60 seconds. Three spectra for each studied sample were registered sequentially.

### 2.2. Tested samples

In this work, the tissue sections of the kidneys were studied after tumor resection from patients at the Sechenov University. The standardized collection of 11 samples with pathological neoplasms was performed. 1 sample corresponded to intact tissue of kidney (Intact tissue), 9 samples are clear cell types of the kidney cell cancer adenocarcinoma (CCA), and 1 sample is chromophobe subtype of the renal cell cancer (ChRCC). Each tissue sample had a size of about 10x15 mm. The samples were placed in blocks after resection and fixed in formalin. Human tissue type (intact or tumor) was determined by histological and immunohistochemical methods.

In the experiment, the inelastic scattering of the human kidney was recorded at three points on the sample. Four spectra were registered at each point on the sample.

### 2.3. Data processing methods

Raman spectra of kidney tissues have hidden connections between different bands of the spectrum, the contribution of which are due to the same chemical components. Therefore, projection to latent structures (PLS) method was used for the experimental data analysis [8]. The objective with PLS is to select a model with the appropriate number of components that has good predictive ability.

The registered signal includes AF and Raman scattering, so a raw spectrum preprocessing was performed for the AF background removal. The Savitzky–Golay smoothing filter was used to reduce the noise effect on the registered signals. The background signal was eliminated by baseline correction with asymmetric least squares smoothing. This method uses a smoothed signal as the baseline for raw signal estimation. The method includes the parameters responsible for the signal smoothing procedure, the baseline position for raw signal and number of iterations. The data were normalized using the standard deviation of a random variable with a normal distribution (Standard Normal Variate - snv). Preprocessing methods were implemented in the cloud service TPTcloud (https://tptcloud.com/).
Spectral informative bands of the regression model were determined from the analysis of the variable importance in projection (VIP) [9].

3. Results and discussion

3.1 Inelastic scattering of human kidney tissue by visible and near-infrared radiation

Registration of inelastic scattering of renal tumor tissues morphologically corresponding to clear cell and chromophobic variants of renal cell carcinoma was carried out. Spectra of renal tissues were compared with the registered signal of intact human kidney tissue. Normalized registered signals of human kidney tissues in the visible and near infrared range are presented in Fig. 2.

![Figure 2. Inelastic scattering of human kidney tissues in visible (a-EnSpectr532) and NIR ranges (b-LML-785.0RB-04, c- EnSpectr1064) (blue line- CCA, yellow line- ChRCC, green line- intact tissue).](image)

It can be seen in Fig. 2, that the intensity of the signal of the tissues takes a negative value. The intensity takes a negative value due to the normalization of the data using the standard deviation of a random variable [10]. This method is based on subtracting the averaged signal from the original
signal, then dividing by the standard deviation. Registered signals of intact tissues and tumor tissues of the human kidney with a 1064 nm laser excitation coincide in shape and have a common Raman peak at a wave number of 880 cm\(^{-1}\), which corresponds to hydroxyproline (C\(_5\)H\(_9\)NO\(_3\), collagen) [2]. Registered signals of intact tissue and ChRCC for a 785 nm laser coincide in shape and don’t have Raman peaks, in contrast to the registered signals of CCA (1250 cm\(^{-1}\) of Amide III in lipids and proteins and 1370 cm\(^{-1}\) corresponding for wagging vibrations of CH\(_2\) in collagen and nucleotide acids) [4,5].

Considering the registered inelastic scattering of human kidney tumors by radiation at a wavelength of 532 nm, it can be noted that the signals of CCA and ChRCC have a coincidence in shape and intensity. Registered intact tissue signal has a high ratio of RS to AF intensity relative to the signals of other tumor tissues.

### 3.2 Regression model for Raman spectra analysis

The next step is the construction of a regression model for the Raman spectra of tumor and intact human kidney tissues. Raw spectra of tumor and safe human kidney tissues were processed by methods of AF background removal, signal smoothing, and data normalization to obtain pure Raman spectra. Informative bands of Raman spectra of human kidney tissues were determined by VIP-scores. VIP-variables were used to isolate the clear cell adenocarcinoma among other samples. Raman spectra of tumor and intact human kidney tissues using a 532 nm laser had a high intensity of AF, which overlapped informative Raman peaks. Therefore, Raman spectra of tumor and intact human kidney tissues were used to study VIP-variables only for NIR radiation (Fig. 3).

![VIP variables of PLS models for discrimination of clear cell adenocarcinoma human kidney cancer from other types of tissues for 785 nm (green line) and 1064 nm (red line) lasers excitation (AF- autofluorescence, Phe - phenylalanine, Trp-tryptophan).](image)

The most informative peaks for clear cell cancer determination using 785 nm and 1064 nm lasers are 330-500 cm\(^{-1}\) (AF), 750 cm\(^{-1}\) (tryptophan in lipids and proteins), 880 cm\(^{-1}\) (hydroxyproline, C\(_5\)H\(_9\)NO\(_3\), collagen), 950 cm\(^{-1}\) (CH stretching bonds of phenylalanine in proteins), 1002 cm\(^{-1}\) (phenylalanine in proteins), 1250 cm\(^{-1}\) (Amide III in lipids and proteins), 1300 cm\(^{-1}\) (CH\(_3\), CH\(_2\) vibrations in collagen), 1350 cm\(^{-1}\) (CH\(_3\), CH\(_2\) vibrations in collagen), 1450 cm\(^{-1}\) (deformation vibrations of CH\(_2\) groups in lipids and proteins) and 1600 cm\(^{-1}\) (tryptophan in lipids and proteins), 1650 cm\(^{-1}\) (Amide I in proteins)[11,12]. It should be noted that VIP-variables made it possible to identify the most informative Raman peaks of clear cell adenocarcinoma (for 1064 nm laser excitation) at the wave numbers of 1002 cm\(^{-1}\), 1450 cm\(^{-1}\) and 1650 cm\(^{-1}\). In addition, the VIP-variables of the PLS model (785 nm laser) identified the most informative bands as 1335 cm\(^{-1}\), 1002 cm\(^{-1}\) and 330-500 cm\(^{-1}\) (AF).

In this paper, spectral differences of human kidney tissues could be caused by the presence of high concentrations of mitochondria in the cytoplasm for renal oncocytoma cells [13]. The study [12] demonstrated SERS method with silver nanoparticles for the differentiation of cancer and healthy kidney cells. The most informative peaks of cancer cells were 667 cm\(^{-1}\) (lipoprotein and glutathion),
800 cm\(^{-1}\) (cytosine, proteins, lactic acid and glutathione), 993 cm\(^{-1}\) (phenylalanine and aspartic acid), 1330 cm\(^{-1}\) (proteins) and 1576 cm\(^{-1}\) (proteins) bands. Raman peak at 993 cm\(^{-1}\) is increased for cancer cells relatively to healthy cells. This fact indicates the increase concentration of phospholipids in cancer cells, because of possible increase in production of fatty acids. In cancer cases the cell division is accelerated, so they need to produce new phospholipids to maintain cell membrane integrity. Positions of Raman peaks of kidney tissues in our work and work \[12\] have coincidences on wavenumbers of 1002 cm\(^{-1}\) and 1336 cm\(^{-1}\).

4. Conclusion

The Raman spectra of kidney tissues has the major peaks of 330-500 cm\(^{-1}\) (AF), 980-1130 cm\(^{-1}\) and 1360-1420 cm\(^{-1}\) for 785 nm laser excitation. Accordingly, the informative peaks for the radiation of 1064 nm laser are 1002 cm\(^{-1}\), 1450 cm\(^{-1}\), 1650 cm\(^{-1}\). Spectral differences of human kidney clear-cell adenocarcinoma are probably caused by the presence of mitochondria high concentrations in the cytoplasm of cancer cells.

5. References

[1] Humphrey P A, Dehner L P and Pfeifer J D 2008 The Washington Manual of Surgical Pathology (Philadelphia: Lippincott Williams & Wilkins) p 980
[2] Huang N, Short M, Zhao J, Wang H, Lui H, Korbelik M and Zeng H 2011 Optics Express 19
[3] Liu Y, Du Z, Zhang J and Jiang H 2017 Oncotarget 8
[4] Lorinz A, Haddad D, Naik V, Naik R, Fung A, Cao A, Manda P, Pandya A, Auner G, Rabah R, Langenburg S E and Klein M D 2004 J Pediatr. Surg. 39
[5] Stewart S, Kirshner H, Treado P J, Priore R, Tretiakova M 2014 J. of Raman spectroscopy 45
[6] Bratchenko I A, Alonova M V, Myakinin O O, Muratov A A, Kozlov S V and Zakharov V P 2016 Hyperspectral visualization of skin pathologies in visible region Computer Optics 40(2) 240-248 DOI: 10.18287/2412-6179-2016-40-2-240-248
[7] Lykina A A, Artemyev D N and Bratchenko I A 2017 JBPE 2
[8] Esbensen K Multivariate Data Analysis 2000 (New Jersey: in Practice) p 587
[9] Wold S, Sjostrom M and Eriksson L 2001 Chemometrics and Intelligent Laboratory Systems 58
[10] Li X, Zhang Y and He Y 2016 Scientific Reports 6
[11] Huang Z, Mcwilliams A, Lam S, English J, Melean D I, Lui H and Zeng H 2003 International J. of oncology 23
[12] Mert S, Culha M 2014 Applied spectroscopy 68
[13] Kalasinsky K S, Hadfield T, Shea A A, Kalasinsky V F, Nelson M P, Neiss J, Drauch A J, Vanni G S and Treado P J 2007 Anal. Chem. 79

Acknowledgments

This research was supported by the Ministry of Education and Science of the Russian Federation.