



OPEN FT-Raman data analyzed by multivariate and machine learning as a new methods for detection spectroscopy marker of platinum-resistant women suffering from ovarian cancer

Marta Kluz-Barłowska¹, Tomasz Kluz^{2,3}, Wiesław Paja⁴, Jaromir Sarzyński⁴, Monika Łączyńska-Madera², Adrian Odrzywolski⁵, Paweł Król⁶, Józef Cebulski^{7,9} & Joanna Depciuch^{5,8,9}✉

The phenomenon of platinum resistance is a very serious problem in the treatment of ovarian cancer. Unfortunately, no molecular, genetic marker that could be used in assigning women suffering from ovarian cancer to the platinum-resistant or platinum-sensitive group has been discovered so far. Therefore, in this study, for the first time, we used FT-Raman spectroscopy to determine chemical differences and chemical markers presented in serum, which could be used to differentiate platinum-resistant and platinum-sensitive women. The result obtained showed that in the serum collected from platinum-resistant women, a significant increase of chemical compounds was observed in comparison with the serum collected from platinum-sensitive woman. Moreover, a decrease in the ratio between amides vibrations and shifts of peaks, respectively, corresponding to C–C/C–N stretching vibrations from proteins, amide III, amide II, C=O and CH lipids vibrations suggested that in these compounds, structural changes occurred. The Principal Component Analysis (PCA) showed that using FT-Raman range, where the above-mentioned functional groups were present, it was possible to differentiate the serum collected from both analyzed groups. Moreover, C5.0 decision tree clearly showed that Raman shifts at 1224 cm^{-1} and 2713 cm^{-1} could be used as a marker of platinum resistance. Importantly, machine learning methods showed that the accuracy, sensitivity and specificity of the FT-Raman spectroscopy were from 95 to 100%.

Ovarian malignancies are heterogenous neoplasms of the distinct origin, precursor lesions, clinical course, risk factors, molecular profiles, treatment and outcomes. Histologically, WHO classifies the ovarian malignancies in distinct groups based on their origin, the most frequently encountered are epithelial ovarian cancers (EOC), constituting 90% of malignant ovarian tumors¹.

The data published by GLOBOCAN in 2020 showed that there had been 313 959 new cases and 207 252 deaths due to ovarian cancer, which made it 8th most commonly diagnosed cancer in females and was ranked 8th as far as cause of cancer deaths is concerned². Interestingly, ovarian carcinoma is the leading cause of death

¹Department of Pathology, Fryderyk Chopin University Hospital, F. Szopena 2, 35-055 Rzeszow, Poland. ²Department of Gynecology, Gynecology Oncology and Obstetrics, Fryderyk Chopin University Hospital, F. Szopena 2, 35-055 Rzeszow, Poland. ³Institute of Medical Sciences, Medical College of Rzeszow University, Kopisto 2a, 35-959 Rzeszow, Poland. ⁴Institute of Computer Science, College of Natural Sciences, University of Rzeszow, Rzeszow, Poland. ⁵Department of Biochemistry and Molecular Biology, Medical University of Lublin, 20-093 Lublin, Poland. ⁶College of Medical Sciences, Institute of Physical Culture Studies, University of Rzeszow, 35-959 Rzeszow, Poland. ⁷Institute of Physics, College of Natural Sciences, University of Rzeszow, 35959 Rzeszow, Poland. ⁸Institute of Nuclear Physics Polish Academy of Sciences, 31342 Krakow, Poland. ⁹These authors jointly supervised this work: Józef Cebulski and Joanna Depciuch. ✉email: joanna.depciuch@ifj.edu.pl

in women diagnosed with gynecological cancers^{3,4}. Since there is no reliable screening program and the course of disease is often asymptomatic, many patients have clinically advanced cancer at the time of the diagnosis. The staging is determined by FIGO classification which is based on both surgical and histopathological assessment. In the vast majority of cases the therapy of ovarian cancer cannot be based solely on cytoreductive surgery, and systemic treatment is crucial.

Platinum-based compounds are currently in use not only for ovarian cancer, but also lung cancer, head and neck cancer, breast cancer and many others. Platinum-based compounds target cancer cells by forming adducts/crosslinks with DNA purine bases, with a preference for guanine. These crosslinks result in DNA damage that impedes proper genome replication, transcription, and triggers cell apoptosis⁵. The accumulation of platinum antitumor agents inside the cells is the necessary assurance of cytotoxicity, so decreased influx or increased efflux is responsible for platinum resistance⁶.

Nowadays, the 1st line treatment in ovarian carcinoma consists of the following: either cytoreductive surgery or neoadjuvant chemotherapy, addition of bevacizumab to carboplatin and paclitaxel to the protocol, and in the case of positive response to platinum-based compounds, and the implementation of PARP inhibitors⁷. However, up to 25% of women with ovarian cancer have so-called platinum-refractory disease. Even if patients are sensitive to 1st line platinum therapy, they may develop recurrence and acquire progressive resistance over time⁸. Generally, patients treated for ovarian cancer may be divided into the following categories:

- Platinum refractory—disease progressing during therapy or within 4 weeks after the last dose
- Platinum resistant—disease progressing within 6 months of platinum-based therapy
- Partially platinum sensitive—disease progressing between 6 and 12 months
- Platinum sensitive—disease progressing with an interval of more than 12 months^{9,10}.

Both platinum refractory and platinum resistant groups of patients have poorer prognosis and are perfect candidates for clinical trials. Drugs used in such cases include cisplatin (PLD) and paclitaxel, sometimes in combination with bevacizumab as the first line chemotherapy. As the second line we usually apply topotecan, gemcitabine and doxorubicin. Cisplatin is able to crosslink with the purine bases on the DNA causing the DNA damage. Paclitaxel targets microtubules causing mitotic arrest. On the other hand bevacizumab inhibits VEGF (Vascular endothelial growth factor). Topotecan binds to the topoisomerase I—DNA complex and prevents relegation of these single strand breaks. Gemcytabine interferes with DNA synthesis and targets ribonucleotide reductase. Doxorubicin inhibits the progression of topoisomerase II. Drugs used in such cases include topotecan, PLD, gemcitabine and paclitaxel in combination with bevacizumab. However, cisplatin is used as the first anyway. Therefore, a method is needed to determine whether a given patient is platinum-sensitive or platinum-resistant in order to increase her chance of recovery. The Raman spectroscopy may be a method. This technique provides information about chemical characterization of the sample under the study. Consequently, chemical fingerprint of the analyzed sample could be obtained¹¹. Moreover, the Raman spectroscopy is a fast, non-invasive, non-expensive and non-destructive technique, which means, that in a short time results could be obtained¹². Furthermore, a small amount of sample is needed, which causes that in a lot of medical and biological research the Raman spectroscopy is used, also in oncology research, e.g. to differentiate cancer and non-cancer tissues¹³, to show effectiveness of a treatment¹⁴ or to determine tumor resection margin¹⁵. Consequently, in this study, for the first time, chemical differences occurred in the serum collected from platinum-sensitive and platinum-resistant women suffering from ovarian cancer using the FT-Raman spectroscopy will be determined. To show significance of these differences, a statistical analysis will be performed. In order to show differentiation between the analyzed samples, a multivariate analyses will be done. Finally, to show accuracy, selectivity and sensitivity of the FT-Raman spectroscopy, a machine learning (ML) method will be used. For this purpose, four ML algorithms were applied: decision trees C5.0, Random Forest (RF), k-Nearest Neighbors (kNN) and Support Vector Machine (SVM). The first one was used to obtain information about the most significant Raman shifts, which could be used as a FT-Raman marker of platinum-resistance. In turn, random forest is an ensemble learning algorithm that combines multiple decision trees to create a more robust and accurate model. Next, SVM algorithm was used to find a hyperplane that maximizes the margin between different classes while minimizing classification errors. Finally, The KNN algorithm was used to measure the similarity between data points. For classification tasks, the class labels of the k nearest neighbors are used to predict the class of the query point.

Results

The FT-Raman spectroscopy was used to determine chemical differences in the serum collected from women suffering from ovarian cancer, where one part of patients were platinum-resistant women and the second part—platinum-sensitive ones. In both the FT-Raman spectra vibrations from hydroxyproline ($\sim 890\text{ cm}^{-1}$), C-H in-plane bending mode ($\sim 1000\text{ cm}^{-1}$) and C–C vibrations ($\sim 1090\text{ cm}^{-1}$) was noticed¹⁶. Moreover, CH₃ bonds from lipids and proteins were visible around $\sim 1370\text{ cm}^{-1}$ ¹⁷. Amide III, amide II and amide I vibrations were located at $\sim 1280\text{ cm}^{-1}$, $\sim 1550\text{ cm}^{-1}$ and $\sim 1670\text{ cm}^{-1}$ Raman shifts¹⁸, while lipids C=O and C–H vibrations were visible around $\sim 1770\text{ cm}^{-1}$ and $\sim 2950\text{ cm}^{-1}$ ¹⁹. Moreover, the additional peak around 1467 cm^{-1} corresponding CH₂ bonds from lipids and proteins was visible in the FT-Raman spectra of serum collected from platinum-sensitive women¹⁷, Fig. 1a. The detailed positions of the described peaks for each analyzed diseases were located in Table 1.

Figure 1b showed that in the FT-Raman spectrum of serum collected from platinum-resistant women a significant increase of hydroxyproline (888 cm^{-1}), C–C band (1085 cm^{-1}), C–C/C–N stretching vibrations from proteins (1182 cm^{-1}), three amides vibrations (1270 cm^{-1} , 1590 cm^{-1} , 1660 cm^{-1}), CH₃ and CH₂ groups from lipids and proteins (1467 cm^{-1} , 1367 cm^{-1} , respectively), as well as CH lipids vibrations (2940 cm^{-1}) were observed in comparison with the FT-Raman spectrum of serum collected from platinum-sensitive woman. Moreover,

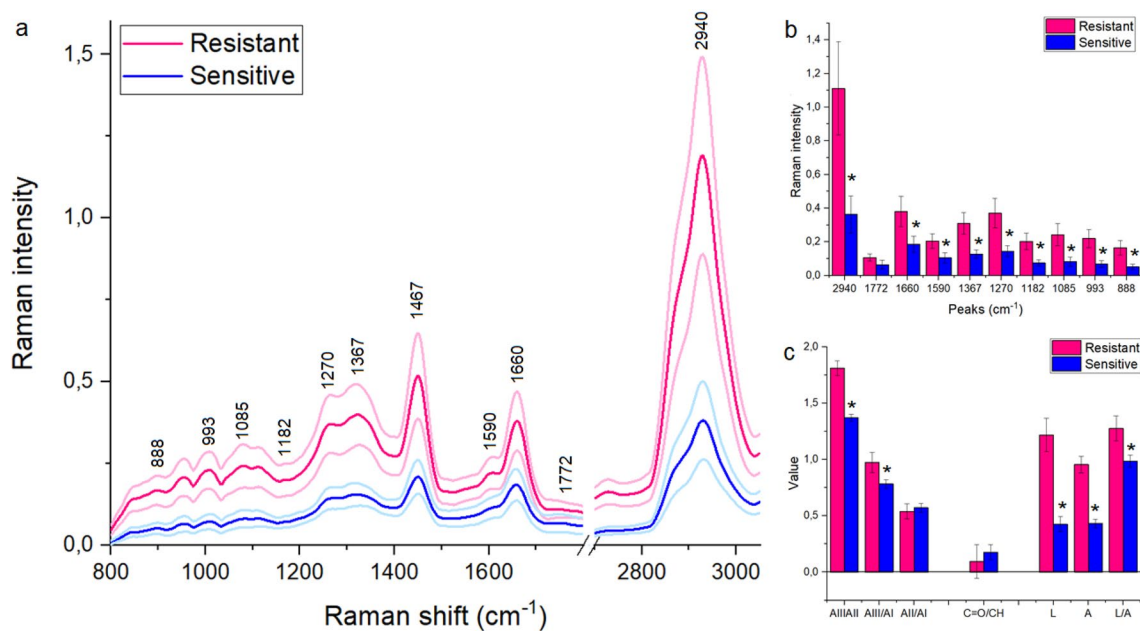


Figure 1. Average of FT-Raman spectra \pm Standard Deviation (SD) of serum collected from platinum-sensitive (blue curve) and platinum-resistant (pink color) women suffering from ovarian cancer (a); average value of peaks intensities \pm SD, where # mean significant differences between platinum-sensitive and platinum-resistant woman. The degree of significance was denoted as $p < 0.05$ (b); the ratio between respectively amides and lipids vibrations, as well as the sum of amides (=A), lipids (=L) and the ratio between lipids and amides (L/A) (c).

Sensitive (S) (cm^{-1})	Resistant (R) (cm^{-1})	S-R (cm^{-1})	Band assigned
888	892	-4	Hydroxyproline
993	996	-3	C-H in-plane bending mode
1085	1085	0	C-C
1182	1197	-15*	C-C/C-N stretching vibrations from proteins
1270	1286	-16*	Amide III
1367	1371	-4	CH ₃ band from proteins and lipids
1467	1452	15*	CH ₂ from lipids and proteins
1590	1590	0	Amide II
1660	1668	-8	Amide I
1772	1783	-11*	C5O from lipids
2940	2964	-24*	Cholesterol and cholesterol ester

Table 1. Peaks analyzed in the FT-Raman spectra of platinum-sensitive and platinum-resistant patients with a band assigned and differences between positions, where “*” means the shift higher than 8 cm^{-1} .

Fig. 1c showed an increase in the ratio between (i) amide III/amide II, (ii) amide III/amide II in platinum-resistant woman in comparison with the platinum-sensitive one. Moreover, higher lipids, as well as amides level was observed in the platinum-resistant women. In the same group an increase in the ratio between lipids and amides vibrations was noticed.

Table 1 showed that in the FT-Raman spectrum of serum collected from platinum-resistant woman, a significant shift of peaks corresponding to C-C/C-N stretching vibrations from proteins, amide III, CH₂ from lipids and proteins, C=O and CH lipids vibrations toward lower Raman shifts were observed. These suggested that in these chemical compounds changes in the structure could occur.

To show if it was possible to differentiate serum collected from platinum-resistant and platinum-sensitive women using the FT-Raman spectroscopy, the PCA analysis was performed, Fig. 2a. This analysis was done for two FT-Raman ranges: $800\text{--}1800 \text{ cm}^{-1}$ and $2800\text{--}3000 \text{ cm}^{-1}$. In the case of the first range, the PCA analysis showed that the samples collected from platinum-sensitive women had the negative value of PC1, and positive as well as the negative value of PC2, Fig. 2a1. Therefore, PC1 played the most important role in distinguishing the samples collected from two analyzed groups of women. Loading plots of PC2 showed positive peaks at 1182 cm^{-1} , 1467 cm^{-1} and 1590 cm^{-1} , 1660 cm^{-1} Raman shifts and negative peaks at 950 cm^{-1} , 1085 cm^{-1} , 1270 cm^{-1} and 1425 cm^{-1} Raman shifts, Fig. 2b1. In the Raman range between 2800 cm^{-1} and 3000 cm^{-1} , the values of PC1 and PC2 for the samples collected from platinum-resistant and platinum-sensitive women were

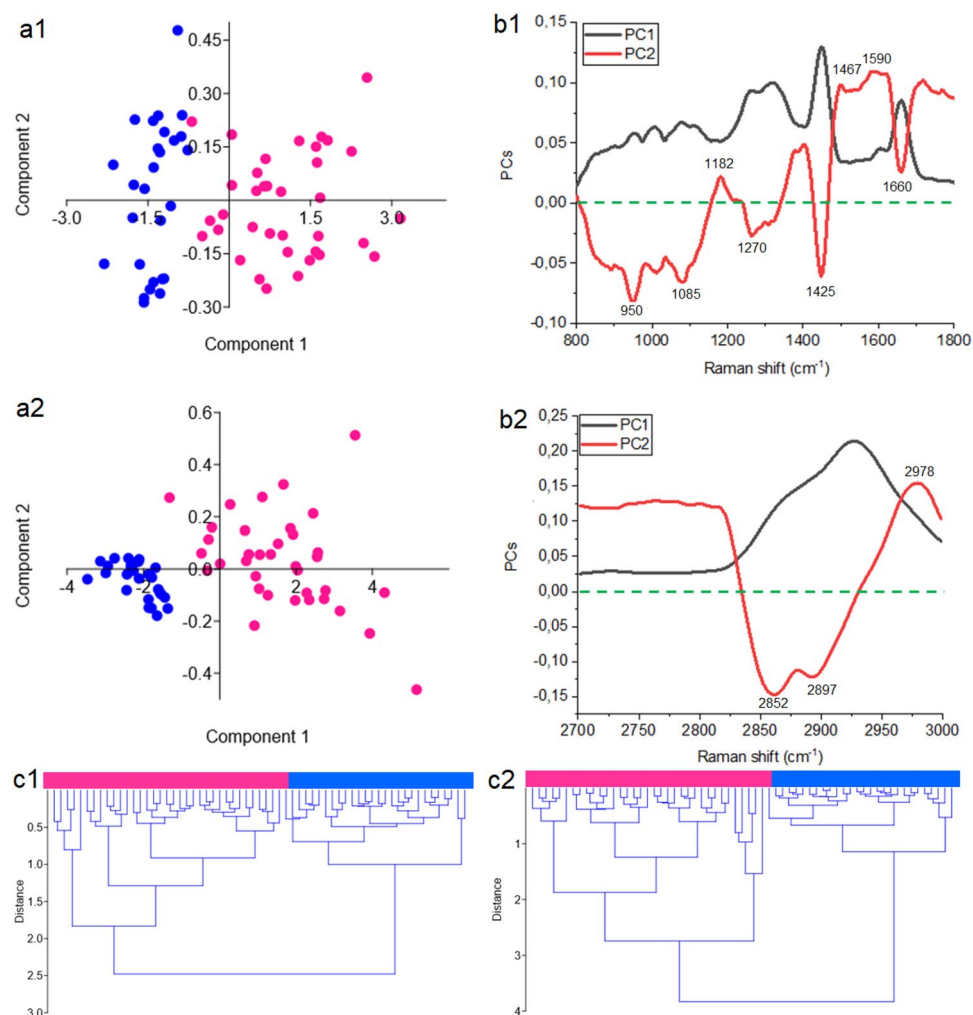


Figure 2. PCA analysis (a) with loading plots (b) and HCA analysis (c) of FT-Raman data of serum collected from platinum-sensitive (blue curve) and platinum-resistant (pink color) women suffering from ovarian cancer. The analyses were done using two spectral ranges: 800–1800 cm^{-1} range (1) and 2800–3000 cm^{-1} (2).

the same as for ranges 800–1800 cm^{-1} , Fig. 2a2. In the loading plots negative peaks at 2852 cm^{-1} and 2897 cm^{-1} and one negative peak at 2978 cm^{-1} were noticed, Fig. 2b2.

Furthermore, HCA analysis showed creation of two similarity groups, where in the first group all the samples collected from platinum-sensitive women were placed, and in the second one—all the samples from other analyzed group. It was visible for both analyzed FT-Raman regions, Fig. 2c1,c2. The results suggested that changes in the proteins and lipids could play an important role in platinum resistance phenomenon.

To sum up, the FT-Raman results showed significant differences in the chemical compositions of serum collected from platinum-resistant and platinum-sensitive woman. Consequently, machine learning methods were used to show accuracy, selectivity and specificity of this technique. The results of the analysis of four created datasets using ML algorithms were shown in Table 2. The calculated values of parameters showed that accuracy was from 95 to 100%, sensitivity—from 89 to 100% and specificity from 97 to 100%.

The analysis of the difference in the spectrum of Raman intensity values (Fig. 3, Table 3) for the categories of positive and negative cases shows that about 6 Raman shifts ranges can be identified, which clearly distinguishes the serum collected from platinum-sensitive and platinum-resistant women.

For the 800–1800 cm^{-1} analyzed FT-Raman range, a strong difference in absorption was seen at the 800–926 cm^{-1} , 969–1016 cm^{-1} , 1035–1074 cm^{-1} , 1081–1441 cm^{-1} , 1455–1479 cm^{-1} and 2700–3000 cm^{-1} ranges, Table 3. Consequently, these ranges could be used as strong markers that differentiate platinum-sensitive and platinum-resistant women.

Also, the C5.0 decision tree learning model generated from these data (Fig. 4) indicates that 1224 cm^{-1} was the most important Raman shift, and was a value that unambiguously distinguishes the two categories of patients. Similarly, for the 2800–3000 cm^{-1} range, a another FT-Raman range under the study, could be used as a marker that differentiates platinum-sensitive and platinum-resistant women.

Dataset	No. of features	Algorithm	Accuracy	Sensitivity	Specificity	Precision	f1	mcc
800–1800 cm^{-1}	260	C5.0	0.95	0.89	1.00	1.00	0.91	0.94
		RF	0.98	1.00	0.97	0.96	0.97	0.98
		kNN (k=1)	0.98	1.00	0.97	0.96	0.97	0.98
		SVM	0.98	1.00	0.97	0.96	0.97	0.98
800–1800 selected wavenumbers	176 selected	C5.0	0.95	0.89	1.00	1.00	0.91	0.94
		RF	0.98	1.00	0.97	0.96	0.97	0.98
		kNN (k=1)	0.98	1.00	0.97	0.96	0.97	0.98
		SVM	0.98	1.00	0.97	0.96	0.97	0.98
2800–3000 cm^{-1}	79	C5.0	0.98	0.96	1.00	1.00	0.97	0.98
		RF	0.98	1.00	0.97	0.96	0.97	0.98
		kNN (k=1)	0.98	1.00	0.97	0.96	0.97	0.98
		SVM	1.00	1.00	1.00	1.00	1.00	1.00
2800–3000 selected wavenumbers	79 selected	C5.0	0.98	0.96	1.00	1.00	0.97	0.98
		RF	0.98	1.00	0.97	0.96	0.97	0.98
		kNN (k=1)	0.98	1.00	0.97	0.96	0.97	0.98
		SVM	1.00	1.00	1.00	1.00	1.00	1.00

Table 2. FT-Raman mean values of the quality assessment parameters of the four obtained learning algorithms, using 4 datasets.

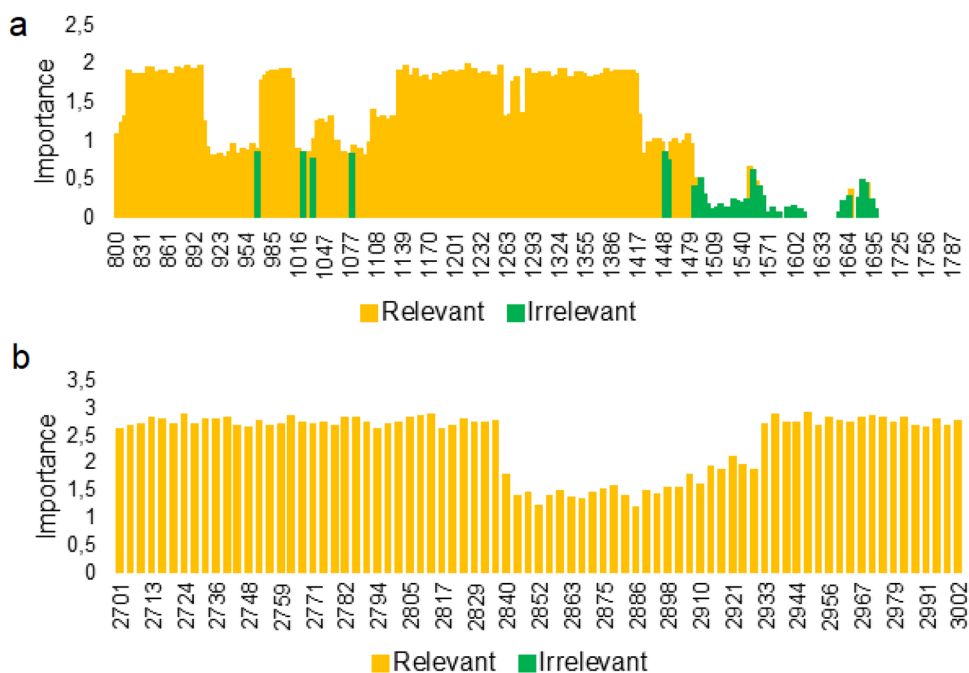


Figure 3. Significant ranges of Raman shifts in the differentiation serum collected from platinum-sensitive and platinum-resistant woman obtained using machine learning algorithms. Significant ranges were calculated for two Raman ranges: 800–1800 cm^{-1} (a) and 2700–3000 cm^{-1} (b).

Raman 800–1800 cm^{-1}	800–962 cm^{-1}
	969–1016 cm^{-1}
	1035–1074 cm^{-1}
	1081–1444 cm^{-1}
	1455–1479 cm^{-1}
Raman 2800–3000 cm^{-1}	2700–3000 cm^{-1}

Table 3. Identified significant ranges of wavenumbers.

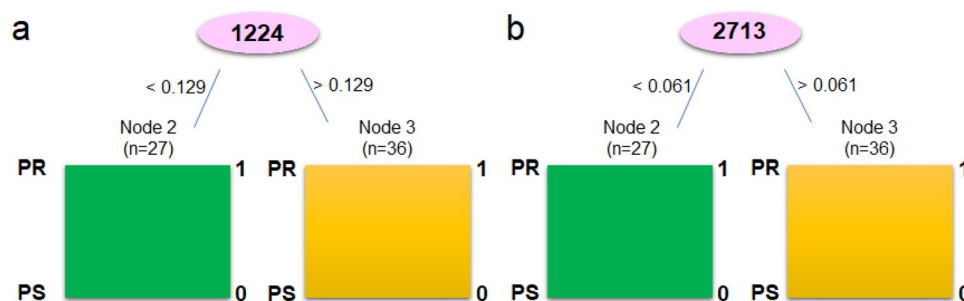


Figure 4. Decision tree for FT-Raman ranges 800–1800 cm^{-1} (a) and 2700–3000 cm^{-1} (b).

Discussion

The survival rates of the ovarian cancer are poor. It is caused by advanced disease stages and disease recurrence, which results of platinum chemotherapy resistance²⁰. Until now, ovarian cancer has been treated by paclitaxel/carboplatin combination chemotherapy, and unfortunately the majority of patients became platinum-resistant with subsequent relapses of the disease⁸. Moreover, platinum resistance can be detected only after 6 months, which is associated with a decrease in patients' chances of recovery²¹. Therefore, it is needed to find platinum-resistance biomarkers, which will show if patient should be treated by platinum compounds or not since in histopathological examination we cannot observe any indicators which will differentiate these two groups. Moreover, also the mechanism of platinum-resistant in ovarian cancer is still not well known, which causes additional problems in chemotherapy of this kind of cancer. Ottevanger et al. showed that the most probable mechanism of platinum resistant is the fact that during chemotherapy it remains, the cancer stem cells remain at rest, consequently drugs have no effect on cancer stem cells²². Nowadays, the most likely marker of platinum resistance is bone morphogenetic protein 2 (BMP2), which is upregulated in ovarian cancer cells, and which is correlated with poor prognosis^{23, 24}. However, although the evidence, that these proteins could be biomarkers of platinum resistance is compelling, prognostic testing or development of targeted treatments are still not accessible. Moreover, only by means of expensive methods, the level of BMP2 in platinum-resistant and platinum-sensitive patients can be investigated. Therefore, in this study, we wanted to investigate if it is possible to find another platinum-resistance marker, which could be found fast and, which will be inexpensive. For this purpose, the FT-Raman spectroscopy in combination with machine learning and multivariate analyses were used. Indeed, the results obtained showed that the greatest differences between platinum-sensitive and platinum-resistant patients occurred in the region of vibration of protein functional groups (Fig. 1a,b), which reflected the proposal of BMP2 protein as a marker of platinum resistance^{23, 24}. Furthermore, in platinum-resistant women suffering from ovarian cancer, higher expression of low density lipoprotein receptor (LDLR) was visible²⁵, which also showed that the FT-Raman marker region obtained could have a potential application. This is even more convincing as the result of C5.0 decision tree method indicated the amide II region as a potential spectroscopic marker region, Table 3. Moreover, in platinum-resistant women, higher amount of amides, as well as changes in the amides balance was observed (Fig. 1c), which suggested, that changes in the proteins expression and structure occurred²⁵. However, analyses of the FT-Raman spectra also showed that C-H lipids functional groups could be used as a platinum-resistant marker (Figs. 2a2, 4b Table 3). Indeed, in chemotherapy, where platinum drugs were used, the fatty acid binding protein 4 (FABP4) plays a very important role as it is responsible for promotion of the uptake of long chain fatty acids into cells^{26, 27}. Consequently, in platinum-sensitive woman a lower amount of free fatty acid should be noted. Our results showed that in platinum-resistant women higher amount of lipids functional groups, as well as global lipids fraction and changes in the lipids balance were observed, Fig. 1. Moreover, PCA analysis of lipids FT-Raman region clearly showed that using this FT-Raman range it was possible to differentiate platinum-resistant and platinum-sensitive women with ovarian cancer. Importantly, in this study we used baseline correction and normalization of all obtained spectra using the same methods. It means that we minimized the influence of sample thickness and homogeneity changes on the results obtained, especially quantitative ones. Moreover, we used the same volume of samples for measurements. Taking into account results obtained by other Authors, which used molecular biology methods^{23–27} and these which we showed, the correlation between the data obtained by different techniques was visible and, what is the most important, has a sense with medical data about platinum-resistant mechanism.

Summarizing, we showed, that using the FT-Raman spectroscopy, functional groups distinguishing samples obtained for two analyzed groups of women suffering from ovarian cancer could be found. These functional groups build protein and lipid structures. In the case of the protein structures, the results are consistent with protein molecular platinum-resistant markers^{23, 28}. The indication of lipids as markers of platinum resistance is a novelty, however, the correlation with molecular studies also exists²⁶. Additionally, taking into account the fact that PCA, HCA and machine learning methods confirm that lipids may have an impact on platinum resistance, we believe that the result is valid and noteworthy. However, further research in this direction should be conducted.

Conclusions

In the presented study, the FT-Raman spectroscopy was used to determine chemical differences in serum collected from platinum-resistant and platinum-sensitive women. Moreover, using multivariate and machine learning methods the spectroscopy data was analyzed to show specific chemical features which could be used to differentiate both studied groups of serum. The results obtained clearly showed that quantitative and qualitative differences between serum collected from both groups of women suffering from ovarian cancer existed. In the FT-Raman spectrum of serum collected from platinum-resistant women significant increase of functional groups building hydroxyproline, three amides vibrations, CH₃ and CH₂ groups from lipids and proteins, as well as CH lipids vibrations was observed in comparison with the FT-Raman spectrum of serum collected from platinum-sensitive woman. Furthermore, structural differences were visible as changes in the value ratio between, respectively, amides vibrations and shift of peak at 1182 cm⁻¹, 1270 cm⁻¹, 1590 cm⁻¹, 1772 cm⁻¹ and 2940 cm⁻¹. Using these differences, it is possible to differentiate serum collected from both groups of women, which was presented in PCA and C5.0 decision tree methods. Importantly, calculated values of parameters showed that accuracy, sensitivity and specificity were around 95%. Consequently, the results showed a possibility of using the FT-Raman spectroscopy to determine if women were platinum resistant or not. Of course, larger studies with more cases from each category have to be required to confirm these results. Moreover, these are the first studies of this type.

Materials and methods

Participants

All the patients who participated in this study signed informed consent. Moreover, the study was approved by the Bioethics Committee of the Regional Medical Chamber in Rzeszów–24 November 2016 (Resolution No. 90/B/2016). Furthermore, all research was performed in accordance with relevant guidelines/regulations, and from all participants (patients) the consent was obtained.

The study group of patients consisted of 21 women with histopathological diagnosis of high-grade ovarian adenocarcinoma who were treated at the Fryderyk Chopin University Hospital in Rzeszów between 2017 and 2020. Among them, 12 were platinum-resistant and 9 were platinum-sensitive. All the samples were collected during the primary surgery before any adjuvant therapy (i.e. chemotherapy, immunotherapy, hormonal therapy and radiotherapy). All of the patients were ranked as stage III according to FIGO (International Federation of Gynecology and Obstetrics) classification and the histopathological type was endometrioid Adenocarcinoma. None of the patients had any type of malignancy ever diagnosed in medical history. The average age of patients was 52,3. In the study group 63% of patients were suffering from arterial hypertension and 27% from obesity. As far as genetic studies are concerned, based on next-generation sequencing (NGS) no mutation, neither in BRCA1 and BRCA 2 gene, nor homologous recombination deficiency (HRD) were found.

Methods

All obtained blood samples were centrifuged for 15 min at 3000 rpm to obtain pure serum. Next, the serum samples were stored at –80 °C. The samples were dispersed just before taking measurements on the FT-Raman spectrometer.

FT-Raman measurements

FT-Raman spectra were recorded using a Nicolet NXR 9650 FT-Raman spectrometer (Thermo Fisher Scientific, USA). In this spectrometer an Nd:YAG laser (1064 nm) and a germanium detector was used. The measurements were performed in the range from 150 to 3.700 cm⁻¹ with a laser power of 1 W. Unfocused laser beam was used with a diameter of approximately 100 μm and a spectral resolution of 8 cm⁻¹. Raman spectra were processed by the Omnic/Thermo Scientific software based on 64 scans. The spectra were normalized using vector normalization in OPUS 7.0 software (Bruker Optik GmbH, Ettlingen, Germany). Moreover, each spectrum was smoothed using 21 points by Savitzky–Golay algorithm.

Analyses of FT-Raman spectra

To determine significant differences in the Raman intensity of peaks presented in the FT-Raman spectra of serum collected from platinum-resistant and platinum-sensitive women, one-way ANOVA with Tukey's post hoc test was done using Past 4.0. software. Furthermore, to show differences in the ratio between amides and lipids vibrations, respectively, the ratio between Raman intensity of peaks originating from amide III, amide II, amide I, as well as C=O and C-H lipids vibrations were used. Finally, to determine global amount of amides (= A) and lipids (= L), the sum of the intensities of peaks at 1270 cm⁻¹, 1590 cm⁻¹ and 1660 cm⁻¹ (amides) and the sum of intensities of peak at 1772 cm⁻¹ and 2940 cm⁻¹ (lipids) was calculated. Furthermore, using results of these sums, the ratio between lipids and amides (L/A) was shown. Significance was defined as less than a 0.05 p-value. Moreover, to obtain information about differentiation as well as similarity between the samples collected from platinum-resistant and platinum-sensitive women, the spectroscopic data was analyzed using PCA and HCA with Euclidean distance and paired group (UPGMA) algorithm analyses. For this purpose, two different FT-Raman range were analyzed: 800–1800 cm⁻¹, where 260 points was taken and between 2700 cm⁻¹ and 3000 cm⁻¹ (52 points). Both analyses were performed using in Past software (developed by Oyvind Hammer). Next, in order to apply ML algorithms, the data obtained from FT-Raman spectroscopy experiments were transformed into the form of a so-called decision system. Such the system takes the form of a decision table where the columns denote features describing the object (in this case, Raman shifts), while the rows are specific learning objects (in this case, disease cases). The last column in the table is a decision column containing the value of the category class assigned to each learning object (in this case, the platinum-sensitive or platinum-resistant class). This form of data was the input for the ML algorithms. To sum up, the columns represent the individual: 260 Raman shifts for the 800–1800 cm⁻¹

range and 79 Raman shifts for the 2700–3000 cm^{-1} range. The rows represent the Raman intensity values of these Raman shifts for specific patients, the last column is the patient category: resistant (positive) and non-resistant (negative) to cisplatin. Four ML algorithms were used for the analysis: decision trees C5.0, Random Forest (RF), k-Nearest Neighbors (kNN) and Support Vector Machine (SVM). C5.0 is a popular decision tree algorithm for data mining and machine learning. It is primarily used for classification tasks, where the goal is to assign a label or a category to a set of input features. C5.0 works by creating a decision tree that recursively splits the data into subsets based on the most informative attributes, ultimately leading to a classification for each instance. In turn, random forest is an ensemble learning algorithm that combines multiple decision trees to create a more robust and accurate model. It was introduced as an improvement over individual decision trees, which can suffer from overfitting and instability. Multiple decision trees are built independently using the bootstrapped datasets and the randomly selected features. Then each tree "votes" for a class, and the class with the most votes becomes the predicted class. The primary goal of SVM algorithm is to find a hyperplane that maximizes the margin between different classes while minimizing classification errors. The margin is the perpendicular distance between the hyperplane and the nearest data point of each class. Support vectors are the data points closest to the hyperplane, and they have an influence on the position and orientation of the hyperplane. These support vectors are used to calculate the margin and determine the optimal hyperplane. The KNN algorithm makes predictions based on the similarity between a query point and its k nearest neighbors in the training dataset. KNN relies on a distance metric (e.g., Euclidean distance, Manhattan distance, or others) to measure the similarity between data points. For classification tasks, the class labels of the k nearest neighbors are used to predict the class of the query point. These models were tested on test cases using a leave-one-out cross validation (LOOCV) approach. Additionally, to determine the most significant differences (specific Raman shifts), the process of relevant feature selection will be carried out, known in machine learning. For this purpose, an algorithm will be used to assess the relevance of features using a random forest. Feature selection using a Random Forest algorithm is a technique that leverages the importance scores assigned to features by a Random Forest model to select the most relevant features for a given task. Random Forests are well-suited for this purpose because they can assess feature importance based on the impact of individual features on the model's performance. Common measures of feature importance in Random Forest include Gini impurity, mean decrease in impurity, or mean decrease in accuracy. The features with higher scores are considered more important, while those with lower scores are less important. In addition, the Boruta algorithm will be used to distinguish between relevant and irrelevant features. Boruta is based on the concept of "shadow features" that are essentially duplicates of original features from dataset. The shadow features make it possible to delineate relevant features from irrelevant features in such a way that all original features having a significance higher than the shadow features are considered essential features.

Data availability

The data supporting this study's findings are available from the corresponding author upon reasonable request.

Received: 4 October 2023; Accepted: 22 November 2023

Published online: 26 November 2023

References

- Höhn, A. K. *et al.* 2020 WHO classification of female genital tumors. *Geburtshilfe Frauenheilkd* **81**, 1145–1153 (2021).
- Ferlay, J. *et al.* Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **149**, 778–789 (2021).
- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
- Arora, T., Mullangi, S. & Lekkala, M. R. Ovarian cancer. In *StatPearls* (eds Arora, T. *et al.*) (StatPearls Publishing, 2023).
- Dasari, S. & Bernard Tchounwou, P. Cisplatin in cancer therapy: Molecular mechanisms of action. *Eur. J. Pharmacol.* **740**, 364–378 (2014).
- Zhou, J. *et al.* The drug-resistance mechanisms of five platinum-based antitumor agents. *Front. Pharmacol.* **11**, 343 (2020).
- Mahmood, R. D., Morgan, R. D., Edmondson, R. J., Clamp, A. R. & Jayson, G. C. First-line management of advanced high-grade serous ovarian cancer. *Curr. Oncol. Rep.* **22**, 64 (2020).
- Davis, A., Tinker, A. V. & Friedlander, M. "Platinum resistant" ovarian cancer: What is it, who to treat and how to measure benefit?. *Gynecol. Oncol.* **133**, 624–631 (2014).
- Friedlander, M. *et al.* Clinical trials in recurrent ovarian cancer. *Int. J. Gynecol. Cancer* **21**, 771–775 (2011).
- Stuart, G. C. E. *et al.* 2010 Gynecologic cancer InterGroup (GFIG) consensus statement on clinical trials in ovarian cancer: Report from the fourth ovarian cancer consensus conference. *Int. J. Gynecol. Cancer* **21**, 750–755 (2011).
- Kizil, R. & Irudayaraj, J. Spectroscopic technique: Fourier Transform Raman spectroscopy. In *Modern Techniques for Food Authentication* 2nd edn (ed. Sun, D.-W.) 193–217 (Academic Press, 2018). <https://doi.org/10.1016/B978-0-12-814264-6.00006-2>.
- Lin, Y. K., Leong, H. Y., Ling, T. C., Lin, D.-Q. & Yao, S.-J. Raman spectroscopy as process analytical tool in downstream processing of biotechnology. *Chin. J. Chem. Eng.* **30**, 204–211 (2021).
- Hanna, K. *et al.* Raman spectroscopy: Current applications in breast cancer diagnosis, challenges and future prospects. *Br. J. Cancer* **126**, 1125–1139 (2022).
- Fenn, M. B. *et al.* Raman spectroscopy for clinical oncology. *Adv. Opt. Technol.* **2011**, 1–20 (2011).
- Aguiar, R. P. *et al.* Discriminating neoplastic and normal brain tissues *in vitro* through Raman spectroscopy: A principal components analysis classification model. *Photomed. Laser Surg.* **31**, 595–604 (2013).
- Shetty, G., Kendall, C., Shepherd, N., Stone, N. & Barr, H. Raman spectroscopy: Elucidation of biochemical changes in carcinogenesis of oesophagus. *Br. J. Cancer* **94**, 1460–1464 (2006).
- Huang, Z. *et al.* Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int. J. Cancer* **107**, 1047–1052 (2003).
- Malini, R. *et al.* Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: A Raman spectroscopy study. *Biopolymers* **81**, 179–193 (2006).
- Notingher, I. *et al.* Discrimination between ricin and sulphur mustard toxicity *in vitro* using Raman spectroscopy. *J. R. Soc. Interface* **1**, 79–90 (2004).
- Van Zyl, B., Tang, D. & Bowden, N. A. Biomarkers of platinum resistance in ovarian cancer: What can we use to improve treatment. *Endocrine-Relat. Cancer* **25**, R303–R318 (2018).

21. Colombo, N. *et al.* ESMO-ESGO consensus conference recommendations on ovarian cancer: Pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Ann. Oncol.* **30**, 672–705 (2019).
22. Ottevanger, P. B. Ovarian cancer stem cells more questions than answers. *Semin. Cancer Biol.* **44**, 67–71 (2017).
23. Le Page, C. *et al.* Gene expression profiling of primary cultures of ovarian epithelial cells identifies novel molecular classifiers of ovarian cancer. *Br. J. Cancer* **94**, 436–445 (2006).
24. Le Page, C. *et al.* BMP-2 signaling in ovarian cancer and its association with poor prognosis. *J. Ovarian Res.* **2**, 4 (2009).
25. Malissa, A., Cappa, F., Schreiner, M. & Marchetti-Deschmann, M. Spectral features differentiate aging-induced changes in parchment—A combined approach of UV/VIS, μ -ATR/FTIR and μ -Raman spectroscopy with multivariate data analysis. *Molecules* **28**, 4584 (2023).
26. Gharpure, K. M. *et al.* FABP4 as a key determinant of metastatic potential of ovarian cancer. *Nat. Commun.* **9**, 2923 (2018).
27. Mukherjee, A. *et al.* Adipocyte-induced FABP4 expression in ovarian cancer cells promotes metastasis and mediates carboplatin resistance. *Cancer Res.* **80**, 1748–1761 (2020).
28. Zheng, L., Li, L., Lu, Y., Jiang, F. & Yang, X.-A. SREBP2 contributes to cisplatin resistance in ovarian cancer cells. *Exp. Biol. Med. (Maywood)* **243**, 655–662 (2018).

Author contributions

M.K.-B.—conceptualization, sample collected, sample description, writing original draft; Tomasz Kluz—conceptualization, sample collected; W.P., J.S.—machine learning analyzed, visualization; M.L.-M.—samples collection and preparation; A.O., P.K.—review and edition; J.C.—supervisor, review and edited; J.D.—supervisor, measurement, visualization, writing original draft.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023