



OPEN

## Spectroscopic evaluation of carcinogenesis in endometrial cancer

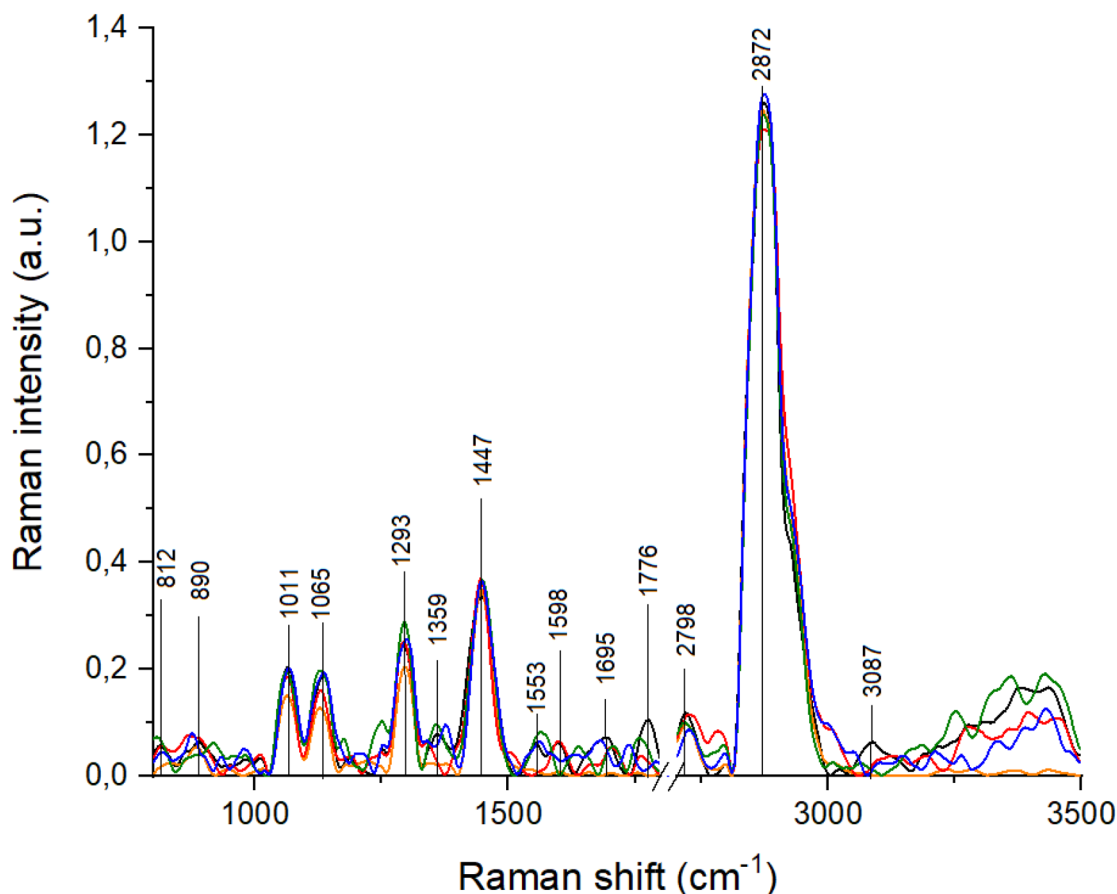
Joanna Depciuch<sup>1,7</sup>✉, Edyta Barnas<sup>2,7</sup>, Joanna Skręt-Magierło<sup>3,7</sup>, Andrzej Skręt<sup>3</sup>, Ewa Kaznowska<sup>4</sup>, Kornelia Łach<sup>5</sup>, Paweł Jakubczyk<sup>6</sup> & Jozef Cebulski<sup>6</sup>

Carcinogenesis is a multifaceted process of cancer formation. The transformation of normal cells into cancerous ones may be difficult to determine at a very early stage. Therefore, methods enabling identification of initial changes caused by cancer require novel approaches. Although physical spectroscopic methods such as FT-Raman and Fourier Transform InfraRed (FTIR) are used to detect chemical changes in cancer tissues, their potential has not been investigated with respect to carcinogenesis. The study aimed to evaluate the usefulness of FT-Raman and FTIR spectroscopy as diagnostic methods of endometrial cancer carcinogenesis. The results indicated development of endometrial cancer was accompanied with chemical changes in nucleic acid, amide I and lipids in Raman spectra. FTIR spectra showed that tissues with development of carcinogenesis were characterized by changes in carbohydrates and amides vibrations. Principal component analysis and hierarchical cluster analysis of Raman spectra demonstrated similarity of tissues with cancer cells and lesions considered precursor of cancer (complex atypical hyperplasia), however they differed from the control samples. Pearson correlation test showed correlation between cancer and complex atypical hyperplasia tissues and between non-cancerous tissue samples. The results of the study indicate that Raman spectroscopy is more effective in assessing the development of carcinogenesis in endometrial cancer than FTIR.

Endometrial cancer is commonly divided into oestrogen-dependent (type I) and oestrogen-independent (type II). Type I constituting 80% of all cases usually represents low-grade tumours. While type II including the remaining 20% of all cases has poorer clinical prognosis<sup>1</sup>. Carcinogenesis is a multistage process preceding cancer observed in type II endometrial cancer. Endometrial hyperplasia without and with histological atypia are generally accepted as stages of type I carcinogenesis. These stages have only a potential role in type I endometrial cancer carcinogenesis<sup>2</sup>. Hyperplasia without atypia transforms into cancer only in 1–3% of cases, while hyperplasia with atypia in 29% of cases<sup>3</sup>. Currently, based on microscopic evaluation it is impossible to predict which cases transfer to cancer. Therefore, new methods are required to enable this prediction. Spectroscopy candidates for this role. Specimens for both pathological spectroscopic analysis of stages of endometrial cancer carcinogenesis and tissues adjacent cancer are easily available in curettage<sup>4</sup>. Another aim of spectroscopic analysis of stages preceding endometrial cancer may be to determine a new mechanism operating behind the initiation of carcinogenesis.

Currently, histopathological methods are still the gold standard for cancer cells diagnosis in biopsy material. However, these methods are used in high stage of diseases. These methods are not only invasive and time consuming, but also the sensitivity of histopathological methods depends on the subjective judgement of a pathologist. Therefore, very high false negative (1.0 to 1.5 per 1000 women) and false positive (121.2 per 1000 women) rates, were observed<sup>5,6</sup>. Moreover, hematoxylin and eosin (H&E) dyes were used for histochemical staining of the tissue samples. Unfortunately, these dyes are non-specific for cancer cells<sup>7</sup>. Biomarkers, defined as disease-related molecular changes in tissues or body fluids, are analysed to decrease the percent of false positive rates<sup>8</sup>. However,

<sup>1</sup>Institute of Nuclear Physics, Polish Academy of Science, 31-342 Krakow, Poland. <sup>2</sup>Institute of Health Sciences, Medical College, University of Rzeszow, Kopisto 2a, 35-959 Rzeszow, Poland. <sup>3</sup>Institute of Medical Sciences, Medical College, University of Rzeszow, Kopisto 2a, 35-959 Rzeszow, Poland. <sup>4</sup>Chair of Morphological Sciences, Department of Pathomorphology, Medical College, University of Rzeszow, Kopisto 2a, 35-959 Rzeszow, Poland. <sup>5</sup>Department of Pediatrics, Institute of Medical Sciences, Medical College, University of Rzeszow, Warzywna 1A, 35-310 Rzeszow, Poland. <sup>6</sup>Institute of Physics, College of Natural Sciences, University of Rzeszow, Pigonia 1, 35-310 Rzeszow, Poland. <sup>7</sup>These authors contributed equally: Joanna Depciuch, Edyta Barnas and Joanna Skręt-Magierło. ✉email: joanna.depciuch@ifj.edu.pl



**Figure 1.** Raman spectra of samples: control (black spectrum); atrophic endometrium (orange spectrum); complex atypical hyperplasia (red spectrum); endometrial polyp (green spectrum); endometrioid adenocarcinoma (blue spectrum).

the methods used for biomarkers marking are expensive. Moreover, lots of biomarkers are nonspecific<sup>9–11</sup>. Therefore, it is very important to find non-invasive and cost-effective tests, which can be used as novel diagnostic methods. Consequently, it is of prime importance to conduct interdisciplinary studies e.g. in physics to develop methods which can be used in diagnostic of cancer diseases<sup>4,12–14</sup>. Some of these methods are Raman and FTIR spectroscopy. They allow chemical characterization of studied samples. Moreover, Raman and FTIR spectroscopy are non-invasive, simple, accurate and rapid. These non-destructive testing methods do not cause any damage to the evaluated materials. Spectroscopy could be used as a comparative for gold standard methods facilitating clinical decision-making and patient outcomes by detecting biochemical changes in tissues before the changes can be observed under microscope. Nowadays, Raman and FTIR spectroscopy methods are used to detect chemical changes caused by lung<sup>12,15</sup>, bones<sup>16,17</sup>, thyroid<sup>13,18</sup>, breast<sup>19</sup>, prostate<sup>20</sup>, endometrial cancers<sup>4</sup>. However, recognizing changes visible during neoplastic process is important in the cancer diagnostics. Changes caused in the subsequent stages of cancers should be investigated to understand the neoplastic disease development process. For this purpose, Raman and FTIR spectroscopy can be used to obtain information about chemical changes. Importantly, our previous study demonstrated the potential of FT-Raman and FTIR spectroscopy for the identification of chemical changes in atypical complex hyperplasia and endometrioid adenocarcinoma cancer<sup>4</sup>. To the best of our knowledge, this is the first study applying physical methods such as FT-Raman and FTIR spectroscopy with multidimensional analysis and correlation test to evaluate carcinogenesis in endometrial cancer. The paper also attempts to investigate which spectroscopy method is more effective in recognizing the different stages of carcinogenesis.

## Results

Raman spectra of all analyzed groups of samples are presented in Fig. 1. Raman spectrum vibrations from nucleic acid (812  $\text{cm}^{-1}$ , 1065  $\text{cm}^{-1}$ , 1293  $\text{cm}^{-1}$ ), proline and hydroxyproline (890  $\text{cm}^{-1}$ ), tryptophan (1359  $\text{cm}^{-1}$ ), protein (1553  $\text{cm}^{-1}$ , 1598  $\text{cm}^{-1}$ ), amide I (1695  $\text{cm}^{-1}$ ), proteins and lipids (1447  $\text{cm}^{-1}$ ) and only lipids (1776  $\text{cm}^{-1}$ , 2798  $\text{cm}^{-1}$ , 2872  $\text{cm}^{-1}$ ) and water (3087  $\text{cm}^{-1}$ ) were visible in control. The differences in the presence of Raman shift were noticed in the tissues from other study groups. The absence of OH vibrations were observed in all study groups compared to the control group. Moreover, peaks corresponding to proline, hydroxyproline, tyrosine,  $\text{PO}_2^-$  stretching of nucleic acids and C=C, tryptophan (protein assignment) disappeared in Raman spectra of tissues from atrophic endometrium and complex atypical hyperplasia groups. Furthermore, lack of peaks at 1011  $\text{cm}^{-1}$  and 1598  $\text{cm}^{-1}$  was observed also for Raman spectrum of atrophic endometrium and endometrial

No.	Control	Atrophic endometrium	Atypical complex hyperplasia	Endometrial polyp	Endometrioid adenocarcinoma	Vibrations
1	812			806	<b>821</b>	Proline, hydroxyproline, tyrosine, PO <sub>2</sub> <sup>-</sup> stretching from nucleic acids
2	890	884	<b>867</b>		<b>876</b>	C–C stretching from proline and hydroxyproline
3	1011		1010			Stretching vibrations of CO, CC, OCH from ring of polysaccharides and pectin
4	1065	1064	1064	1061	1067	Proline
5	1293	1296	<b>1300</b>	1295	<b>1302</b>	Amide III (collagen assignment)
6	1359			1358	<b>1376</b>	Tryptophan
7	1447	1447	1446	1448	1447	CH <sub>2</sub> bending from lipids and proteins
8	1553			1564	1561	C=C, tryptophan (protein assignment)
9	1598		1599			C=N and C=C stretching from protein
10	1695		<b>1705</b>	<b>1638</b>	<b>1685</b>	Amide I
11	1776		1763	<b>1757</b>	<b>1792</b>	C=O stretching from lipids
12	2798	<b>2720</b>	<b>2782</b>	2794	2795	CH <sub>3</sub> stretching from lipids
13	2872	2869	2872	2871	2873	CH <sub>2</sub> stretching from lipids
14	3087					OH vibrations from water

**Table 1.** Raman shift with corresponding vibrations described in the Raman spectra from Fig. 1<sup>20–24</sup>. Bold means statistically significant shift.

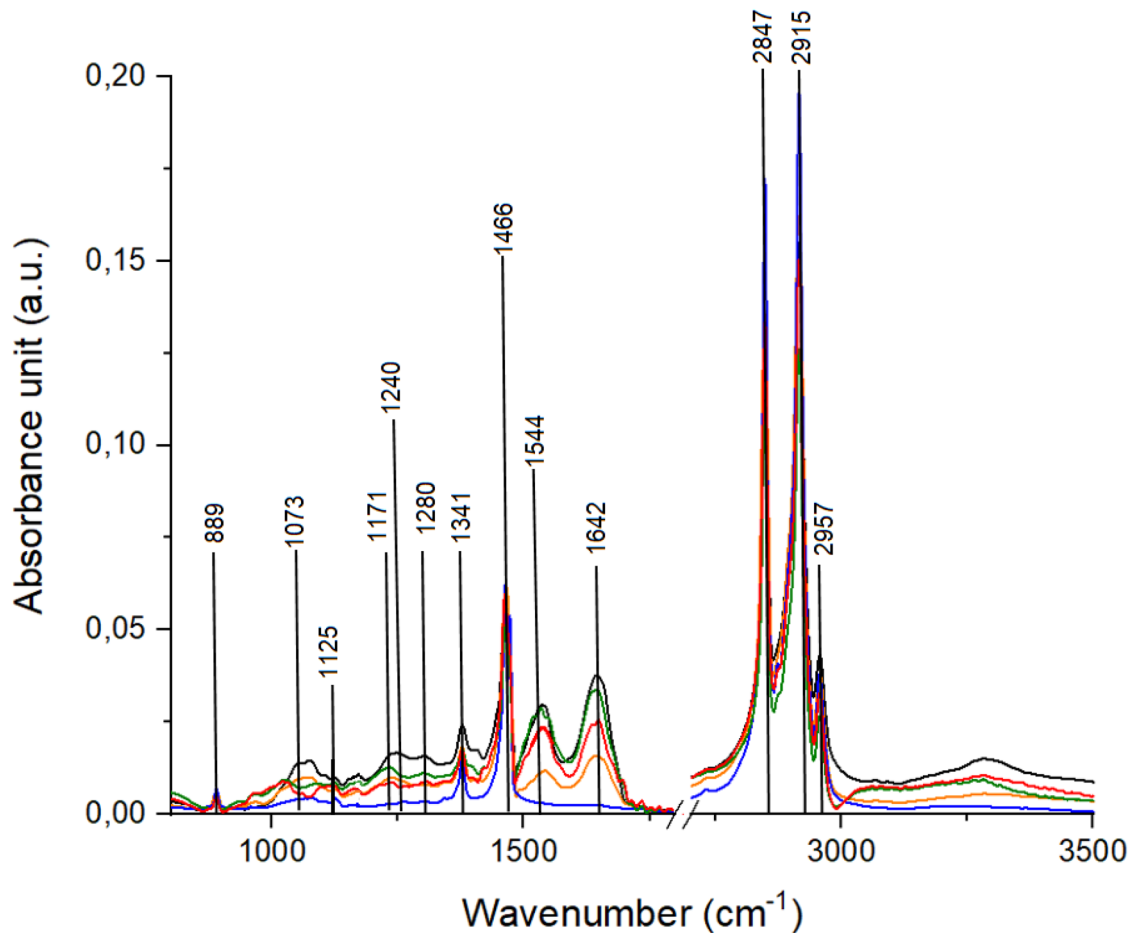
polyp. Raman shift originating for C–C stretching from proline and hydroxyproline disappeared in the spectrum of endometrial polyp.

Comparison of the study groups and control one revealed the disappearance of peaks and peak shifts. Significant shift of peak corresponding to CH<sub>3</sub> stretching from lipids was visible in the Raman spectrum of atrophic endometrium. Moreover, the shift of peaks originating from C–C stretching from proline and hydroxyproline, amide I and CH<sub>3</sub> stretching from lipids, were noticed in the Raman spectrum of complex atypical hyperplasia. In the spectrum of endometrial polyp, a significant shift of peaks corresponding to amide I and C=O stretching vibrations from lipids were observed. The most significant shift of the largest number of peaks was visible in the Raman spectrum of endometrioid adenocarcinoma when compared to the control group. Significant shift of peaks originating from: proline, hydroxyproline, tyrosine, PO<sub>2</sub><sup>-</sup> stretching of nucleic acids, C–C stretching from proline and hydroxyproline, phosphodiester groups in nucleic acids, tryptophan, amide I and C=O stretching from lipids, were visible in the Raman spectrum of endometrioid adenocarcinoma. All analyzed Raman peaks and their shifts are presented in Table 1.

FTIR spectra of all analyzed groups of samples are presented in Fig. 2 which demonstrated that all study group samples differed from the control one. In FTIR spectrum of atrophic endometrium (orange plot), significant shift of peaks corresponding to amide II, amide I and CH<sub>2</sub> as well as CH<sub>3</sub> groups from lipids vibrations was noticed. Moreover, lack of amide III vibrations and presence of C–O carbohydrates group was observed in comparison with the control group samples. FTIR spectrum of complex atypical hyperplasia (red plot) was characterized by shifting of peaks originating from C–O stretching mode of serine, threonine, and tyrosine, C–O deformation vibrations of proteins, glycogen, carbohydrates, all three amides modes and N–H vibrations of cytosine. The absence of CH<sub>2</sub> wagging for proline (amino acids and collagen) was also noticed. Comparison of FTIR spectrum of endometrial polyp (red plot) with FTIR spectrum of control group samples (black plot) revealed the presence of C–O group from carbohydrates and absence of C–O stretching mode from serine, threonine, and tyrosine of protein and N–H group of cytosine. Moreover, significant shift of peaks at 1168 cm<sup>-1</sup>, 1245 cm<sup>-1</sup>, 1560 cm<sup>-1</sup>, 2842 cm<sup>-1</sup> was observed. Finally, in FTIR spectrum of endometrioid adenocarcinoma (blue plot), shift of peaks corresponding to C–O stretching mode of C–OH groups from serine, threonine, tyrosine, amide III, amide II, amide I and absence of deformation N–H cytosine group was visible in comparison to FTIR spectrum of the control group. Description of analyzed peaks present on the FTIR spectra was presented in Table 2.

Differences in the values of absorbance and Raman intensity were visible in the Figs. 1 and 2. Therefore, average values of these parameters were calculated for all analyzed group of tissue samples, Fig. 3.

Figure 3 presents mean ± SEM intensity values of peaks from the Raman, Fig. 3a, and FTIR, Fig. 3b, spectra. Only statistically significant differences were considered. Figure 3a demonstrated differences between control and atrophic endometrium samples in the 1293 cm<sup>-1</sup> peak, while differences between control and complex atypical hyperplasia samples were visible at 812 cm<sup>-1</sup>, 890 cm<sup>-1</sup>, 1011 cm<sup>-1</sup>, 1065 cm<sup>-1</sup>, 1293 cm<sup>-1</sup>, 1359 cm<sup>-1</sup>, 1447 cm<sup>-1</sup>, 1695 cm<sup>-1</sup>, 1776 cm<sup>-1</sup>, 2872 cm<sup>-1</sup> and 3088 cm<sup>-1</sup> Raman shifts. Differences between control and polyp tissue samples were visible at 890 cm<sup>-1</sup>, 1065 cm<sup>-1</sup>, 129 cm<sup>-1</sup> 3, 1359 cm<sup>-1</sup> and 1447 cm<sup>-1</sup>. Moreover, control samples differed from cancer ones in all analyzed Raman spectra. Comparison of atrophic endometrium with complex atypical hyperplasia samples showed no differences at the 1598 cm<sup>-1</sup>, 1776 cm<sup>-1</sup>, 2789 cm<sup>-1</sup>. Atrophic endometrium sample differed from polyp one at the 890 cm<sup>-1</sup>, 1065 cm<sup>-1</sup>, 1293 cm<sup>-1</sup>, 1359 cm<sup>-1</sup>, 1447 cm<sup>-1</sup>, 1598 cm<sup>-1</sup> Raman shifts, while comparison of this sample with cancer one showed differences in all analyzed spectra. Furthermore, complex atypical hyperplasia tissue samples differed from endometrial polyp ones at the 2872 cm<sup>-1</sup>, 1293 cm<sup>-1</sup> and 1447 cm<sup>-1</sup>, while with the endometrioid adenocarcinoma, complex atypical hyperplasia tissues differed at 890 cm<sup>-1</sup>, 1065 cm<sup>-1</sup>, 1293 cm<sup>-1</sup>, 1359 cm<sup>-1</sup>, 1447 cm<sup>-1</sup>, 2872 cm<sup>-1</sup>, 2789. When compared to cancer



**Figure 2.** FTIR spectra of samples: control (black spectrum); atrophic endometrium (orange spectrum); complex atypical hyperplasia (red spectrum); endometrial polyp (green spectrum); endometrioid adenocarcinoma (blue spectrum).

samples, polyp tissue ones did not show any differences at the  $1011\text{ cm}^{-1}$  and  $1598\text{ cm}^{-1}$ . Figure 3b demonstrated, that statistically significant differences between groups were observed only in the case of two analyzed peaks:  $1073\text{ cm}^{-1}$  and  $1125\text{ cm}^{-1}$ . These differences were visible when we compared control and atrophic endometrium, endometrial polyp and control and complex atypical hyperplasia, endometrial polyp. Moreover, statistically significant differences between atrophic endometrium and complex atypical hyperplasia were noticed for the peak at  $1073\text{ cm}^{-1}$ . In case of the peak at  $1125\text{ cm}^{-1}$ , statistically significant differences were observed between endometrial polyp, complex atypical hyperplasia samples and the control ones.

FT-Raman and FTIR spectra variation among the types of samples and their similarity were analysed with principal component analysis (PCA) and hierarchical cluster analysis (HCA), Fig. 4. All analyzed Raman spectra peaks were statistically significant, while in the case of FTIR spectra, only two peaks turned out to be statistically significant. Therefore, PCA analyses were performed only for statistically significant data, which can be further used to differentiate various types of endometrial tissue samples.

PCA analysis of average spectra (Fig. 4a) showed, that Raman spectra of atrophic endometrium, atypical complex hyperplasia, endometrial polyp, endometrioid adenocarcinoma groups differed from the control group. However, similarities in the Raman spectra of endometrial polyp and endometrioid adenocarcinoma groups, were visible. These observations conformed with HCA analysis (Fig. 4b), which showed similarity between these two types of samples. Moreover, HCA analysis also showed that the Raman spectrum of atrophic endometrium group was similar to the one of complex atypical hyperplasia. PCA analysis of average FTIR spectra (Fig. 4c) demonstrated that FTIR spectra of complex atypical hyperplasia and endometrioid adenocarcinoma were placed in the same quarter of the coordinate system. This analysis revealed also similarity between control and endometrial polyp group. Furthermore, HCA analysis of average FTIR spectra (Fig. 4d) showed that control, atrophic endometrium and endometrioid adenocarcinoma create one group which is similar to FTIR spectrum of endometrial polyp tissues. Moreover, PCA of Raman data of all samples, Fig. 4e, revealed that the largest points in the same quarter of the coordinate system referred to endometrioid adenocarcinoma tissue group. However, the highest dispersion was found for atrophic endometrium tissue group. Similar case occurred in the PCA plot obtained for one selected FTIR region, Fig. 4f. Also in this case, significant separation between analyzed group of samples was not observed. Therefore, we marked vectors of least-squared lines representing different groups in

No.	Control	Atrophic endometrium	Atypical complex hyperplasia	Endometrial polyp	Endometrioid adenocarcinoma	Vibrations
1	889	888	889	887	889	C–C, C–O deoxyribose, fatty acid, saccharide
2	1073	1078	<b>1080</b>	*	<b>1066</b>	C–O stretching mode of C–OH groups of serine, threonine, and tyrosine of protein
3	*	<b>1125</b>	<b>1124</b>	<b>1124</b>	*	vC–O Carbohydrates
4	1171	1184	<b>1167</b>	<b>1168</b>	1169	v(C–O), v(C–C), def. C–O–H (proteins, glycogen, carbohydrates)
5	1240	*	<b>1237</b>	<b>1245</b>	1241	Amide III (N–H bending, C–N stretch, C–C stretch) (proteins, DNA, phospholipids)
6	1280	*	<b>1304</b>	*	*	Deformation N–H cytosine
7	1341	1341	*	1341	1340	CH <sub>2</sub> wagging for proline (amino acids and collagen)
8	1466	1466	<b>1462</b>	1465	<b>1462</b>	CH <sub>2</sub> group scissoring modes
9	1544	<b>1558</b>	<b>1535</b>	<b>1560</b>	<b>1535</b>	Amide II due to N–H bending and C–N stretching of proteins
10	1642	<b>1645</b>	<b>1646</b>	1642	<b>1649</b>	Amide I (v(C=O), v(CN), γ(CCN), δ(NH)) (proteins)
11	2847	<b>2843</b>	2847	<b>2842</b>	2848	Symmetric stretching of the CH <sub>2</sub> group due to mainly lipids, with little contribution from proteins, carbohydrates and nucleic acids
12	2915	<b>2912</b>	2915	2915	2915	CH <sub>2</sub> asymmetric stretch: mainly lipids, with little contribution from proteins, carbohydrates, nucleic acids
13	2957	<b>2953</b>	2956	2957	2956	CH <sub>3</sub> asymmetric stretch: mainly lipids

**Table 2.** FTIR wavenumbers with corresponding vibrations described in the FTIR spectra presented in Fig. 2<sup>25–30</sup>. Bold means statistically significant shift.

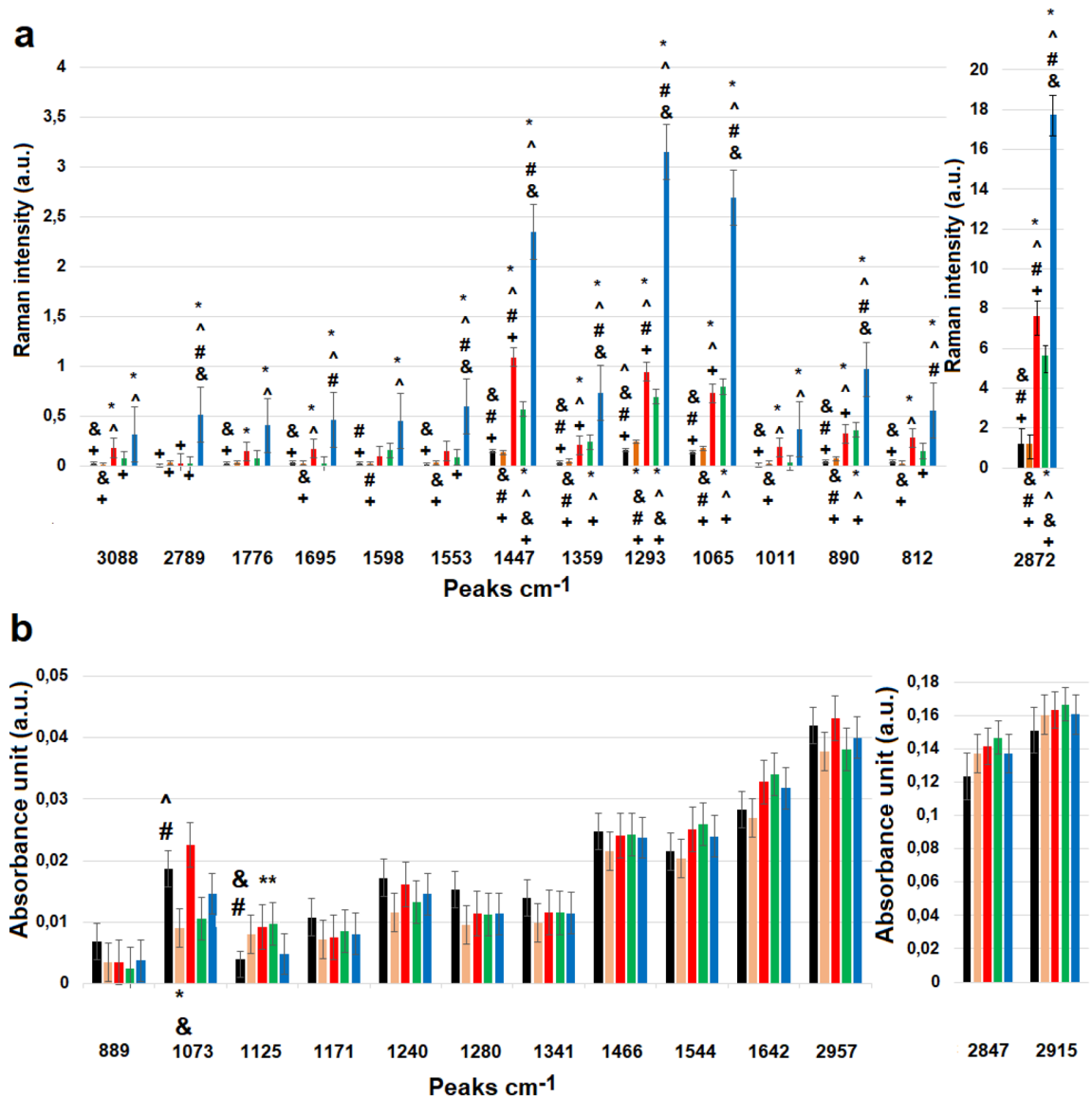
Fig. 4e,f. Consequently, PCA obtained from Raman data showed, that atrophic endometrium, complex atypical hyperplasia and endometrioid adenocarcinoma, were separated from control and polyp endometrial tissues. PCA obtained from FTIR data, showed separation between control and polyp, as well as cancer tissue groups. Therefore, Partial Least Squares analysis (PLS) with variables importance in projection (VIP) was performed, Fig. 5.

The PLS results presented as plots of predictor showed, that in the Raman spectra, the region which could be used to separate endometrial tissues with different carcinogenesis stages was between 500 and 1500 cm<sup>-1</sup> and around 2800–2900 cm<sup>-1</sup>, Fig. 5a. Moreover, the region between 1500 and 1700 cm<sup>-1</sup>, which is very important in description of changes in proteins, can give false positive or false negative results. Obtained prediction correlation was characterized by good predicted response for training data, which was visible in linear fit presented in Fig. 5b. FTIR data analysed with PLS showed good linear fit, Fig. 5e for data enabling separation between each analysed sample in the range between 500 and 1500 cm<sup>-1</sup> and in the wavenumbers corresponding to CH<sub>2</sub> and CH<sub>3</sub> vibrations of lipids, Fig. 5d. VIP values were generated from the model as shown in Fig. 5c,f for Raman and FTIR data, respectively. The threshold level was established at 0.8. Note that the VIP values were not associated for all the measured Raman and FTIR spectra. For the Raman data, VIP values showed that separation of the control and atrophic endometrium samples was impossible, while VIP values obtained for FTIR spectra showed, that polyp samples could not be separated from others. Furthermore, Random Forest as well as second learning machine method C5.0 classification model were calculated to obtain information about accuracy of Raman and FTIR spectroscopy in distinguishing analyzed samples. These two analyses were done for fingerprint FTIR and Raman region (800–1800 cm<sup>-1</sup>), and for a selected region (only peaks which differences were statistically significant—data taken from Fig. 3). The results were presented in Table 3.

The results obtained using Random forest and C5.0 algorithms indicate that both Raman and FTIR approaches can effectively identify groups of cases. The classification accuracy is in the range of 62.71 to 96.61%. It can be seen that much better results were obtained using Random forest algorithm which is an ensemble machine learning method. At the same time, the obtained results provide a basis for further extensive research on a larger number of learning instances.

The correlation between all measured samples was investigated using Pearson correlation test, Table 4.

The Pearson correlation test obtained from Raman spectra and presented in Table 3 showed, that it was possible to determine the correlation between control, atrophic endometrium and endometrial polyp tissues. Moreover, correlation between atrophic endometrium, endometrial polyp and endometrioid adenocarcinoma was also observed in Raman spectra. Furthermore, correlation in Raman spectra between polyp, control, atypical hyperplasia and endometrioid adenocarcinoma was noticed. Interestingly, lack of correlation between endometrioid adenocarcinoma and other analyzed tissues was visible in Raman spectra. Pearson correlation test obtained from FTIR spectra (Table 3) showed, that each analyzed sample correlated with all others, e.g. correlation between control, atrophic endometrium, atypical hyperplasia, endometrial polyp and endometrioid adenocarcinoma was observed.



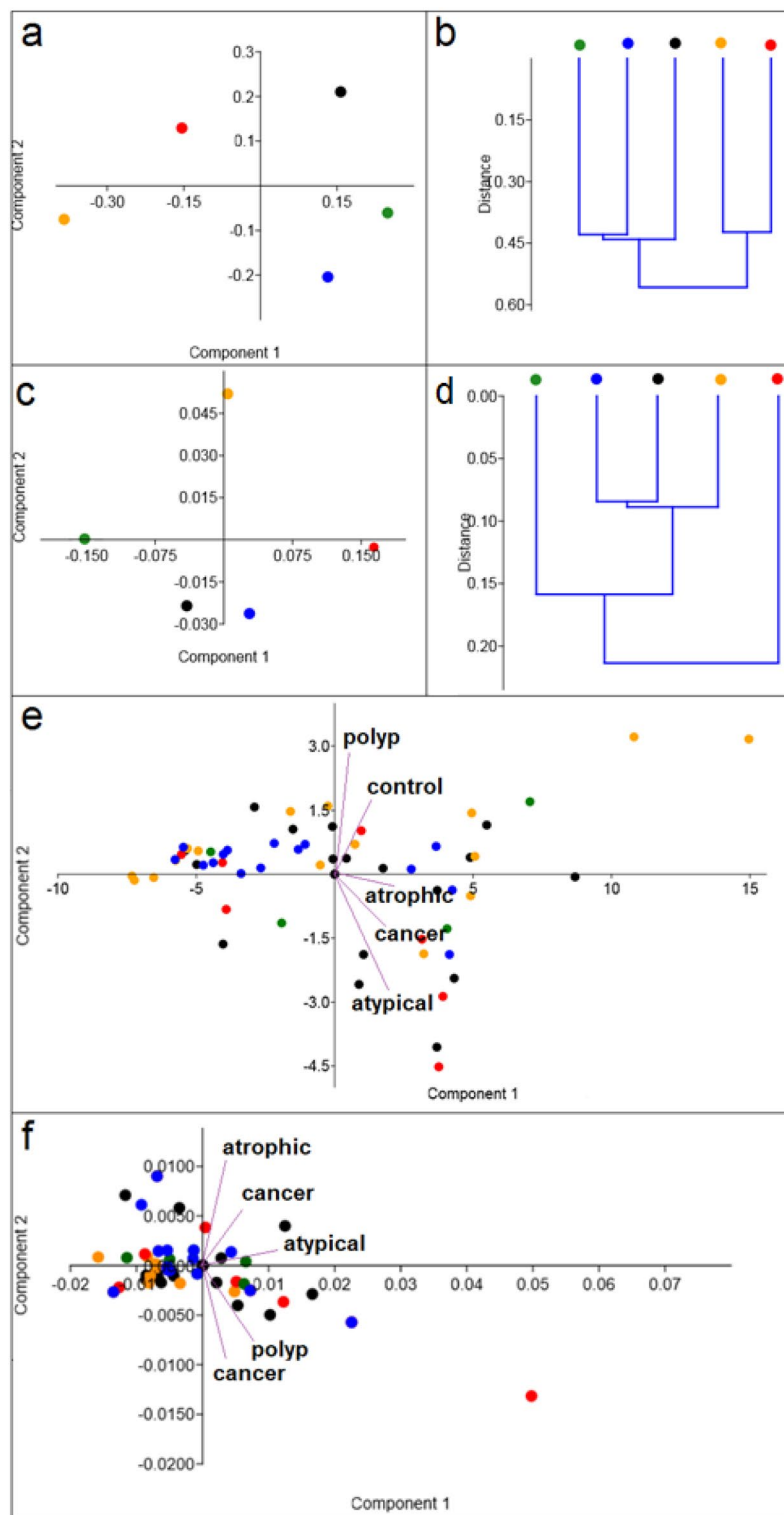
**Figure 3.** Mean  $\pm$  standard error of the mean (SEM) intensity values of the individual peaks measured with Raman (a) and FTIR (b) for all groups of tissue samples, where color bars stand for: control (black spectrum); atrophic endometrium (orange spectrum); complex atypical hyperplasia (red spectrum); endometrial polyp (green spectrum); endometrioid adenocarcinoma (blue spectrum). Data was analyzed using one-way ANOVA followed by Tukey's post hoc test. Statistical significance was adopted at \* $p < 0.05$  versus Control; ^  $p < 0.05$  versus atrophic endometrium; &  $p < 0.05$  versus complex atypical hyperplasia; #  $p < 0.05$  versus endometrial polyp; +  $p < 0.05$  versus endometrioid adenocarcinoma.

## Discussion

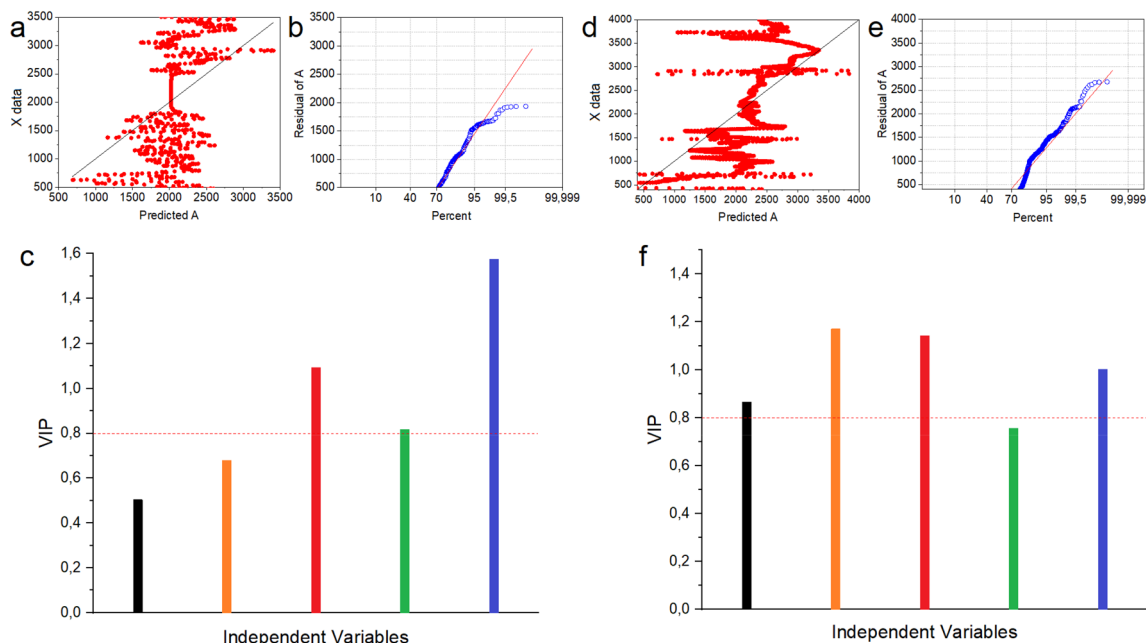
Observation of lesions characteristic for cancer cells without significant architectural changes in tissues is very important to understand the carcinogenesis of endometrial cancer. However, the diagnostic gold standard used currently does not allow for it. Therefore, a new diagnostic technique is required to facilitate this application in future. Consequently, this study showed chemical changes that occur during carcinogenesis process in endometrial tissues using FT-Raman and FTIR spectroscopy.

Recent studies show, that all spectroscopic techniques can be used to observe chemical differences between healthy and non-healthy endometrial tissues<sup>4,31–36</sup>. Patel et al. showed, that amides and prolines vibrations can be used as spectroscopic marker in endometrial cancer<sup>31</sup>. Our Raman results also showed, that quantitative and qualitative changes in endometrial cancer tissues are observed at 1695  $\text{cm}^{-1}$  and 812  $\text{cm}^{-1}$  wavenumbers, which corresponded to amide I and proline vibrations, respectively, Fig. 1. Moreover, we also noticed structural changes in amides vibrations in FTIR spectra, Fig. 2. Similar results were obtained by Taylor et al.<sup>37</sup>. They used FTIR spectroscopy to identify chemical changes that occur at different stages of endometrial cancer. Furthermore, the changes in amides vibrations are more significant, when the cancer stage is higher. As shown in Table 2 shifts were observed in amides vibrations. Interestingly, in comparison with control samples, these shifts were more





**Figure 4.** PCA (a, c, e, f) and HCA (b, d) analysis of: control (black dot); atrophic endometrium (orange dot); complex atypical hyperplasia (red dot); endometrial polyp (green dot); endometrioid adenocarcinoma (blue dot) Raman (a, b, e) and FTIR (c, d, f) spectra. Two-dimensional (2D) scores plot of samples with differences in chemical compositions presented through the selected spectral regions: all spectral region for Raman spectra and points corresponding to C–O stretching mode of C–OH groups of serine, threonine, and tyrosine of protein and  $\nu$ C–O vibration from carbohydrates for FTIR data. The results for average spectra were shown in Figure (a–d) and for all analyzed samples—e, f with marker with vectors of least-squared lines representing the various groups.



**Figure 5.** Plots of predicted for Raman (a) and FTIR (d) region, which could be used to separate endometrial tissues with different carcinogenesis stages. Normal Probability Plot of residual collected from predicted plots for Raman (b) and FTIR (e) spectra. VIP values plot for the all analyzed Raman (c) and FTIR (f) range with 0.8 VIP thresholds, where colors stands for the following samples: control (black); atrophic endometrium (orange); complex atypical hyperplasia (red); endometrial polyp (green); endometrioid adenocarcinoma (blue).

	Random forest		C5.0	
	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
Raman selected	92.73	7.27	76.36	23.64
Raman fingerprint	92.73	7.27	83.64	16.36
FTIR selected	96.61	3.39	62.71	37.29
FTIR fingerprint	96.61	3.39	84.75	15.25

**Table 3.** Percentage accuracy with error value of selected and for fingerprint region (800–1800 cm<sup>-1</sup>) for obtained Raman and FTIR spectra.

	Control	Atrophic endometrium	Atypical hyperplasia	Endometrial polyp	Endometrioid adenocarcinoma
<b>FT-Raman spectroscopy</b>					
Control		0.80*		0.86*	
Atrophic endometrium	0.80*		0.75*		
Atypical hyperplasia				0.95*	
Endometrial polyp	0.86*	0.97*	0.95*		
Endometrioid adenocarcinoma		0.93*	0.97*	0.93*	
<b>FTIR spectroscopy</b>					
Control		0.95*	0.93*	0.95*	0.98*
Atrophic endometrium	0.95*		0.97*	0.84*	0.98*
Atypical hyperplasia	0.93*	0.97*		0.81*	0.96*
Endometrial polyp	0.95*	0.84*	0.81*		0.90*
Endometrioid adenocarcinoma	0.98*	0.98*	0.96*	0.90*	

**Table 4.** Pearson correlation ( $p < 0.05$ ) test for Raman and FTIR spectra. \*Statistically significant, when  $p < 0.05$ .



significant for samples with more developed carcinogenesis process. It is important, especially, when attempting to differentiate complex atypical hyperplasia and endometrial cancer, because differential diagnosis between these two types of endometrial tissues still a problem<sup>38</sup>. Some research showed, that p53 protein can be a factor, which could be used in diagnostics of these two endometrial changes<sup>39</sup>. Therefore, results obtained from the Raman and FTIR range corresponding to amides vibrations confirmed the molecular biology hypothesis. Moreover, the differences in the proteins region in spectra can suggest, that expression of genes and, consequently, proteins conformation could be used as a biomarker in the different stage of endometrial carcinogenesis.

Paraskevaïdi et al. used blood serum collected from women suffering from complex atypical hyperplasia and endometrial cancer stage I and II<sup>36</sup>. We observed in FTIR spectra of analysed samples that chemical changes visible in tissue samples corresponded with these noticed in blood serum, especially in amides IR range. However, Paraskevaïdi et al. showed also that CH<sub>2</sub> wagging vibrations from collagen played important role in carcinogenesis process detected from blood serum. In our study significant shift of collagen vibrations was visible in Raman spectra of the samples characterized by the most advanced carcinogenesis process, Fig. 1, Table 2. Moreover, intensity of peaks originating from functional groups building collagen structure decrease together with carcinogenesis process. It could be caused by collagen linearization and tissue stiffness and such changes increase the tumor incidence and progression<sup>40</sup>. Collagen constitutes the scaffold of tumor microenvironment and regulating its extracellular matrix remodeled by collagen degradation and re-deposition, and promoting tumor infiltration, angiogenesis, invasion and migration<sup>41</sup>. Therefore, we think, that also vibrations of collagen functional groups can be used as spectroscopic marker of carcinogenesis.

Moreover, PCA analysis of average spectra of different types of tissues, showed, that only in the case of Raman data, differentiation between cancer and complex atypical hyperplasia was possible, Fig. 4a. These two kinds of endometrial changes differed enough to allow differentiation<sup>38</sup>. The histopathology image of complex atypical hyperplasia is very similar to the one of cancer. Therefore, new methods are required that enable differentiation of these two kinds of endometrial changes. However, PCA analysis of all analyzed samples measured with Raman (Fig. 4a) and FTIR (Fig. 4b) did not show separation between each stage of carcinogenesis. The similarity between control and polyp endometrial tissues, was detected from the Raman data, while PCA analysis of FTIR spectra showed non-separation between control, atrophic endometrium and complex atypical hyperplasia. Polyp and atrophic endometrium are classified as normal endometrial tissues without any signs of neoplastic changes<sup>42</sup>. Complex atypical hyperplasia is very similar to the image of cancer<sup>38</sup>. Therefore, FTIR may not be used to separate very similar carcinogenesis stages. However, in this study we used sample collected from women at different age and it is known that age greatly influences FTIR and Raman spectra obtained<sup>43</sup>.

As shown in Figs. 1, 2 and 3 endometrial tissues in different carcinogenesis differs fundamentally from normal tissue in terms of structure, genetics, and cellular activity. Therefore, obtained spectra were analyzed by multivariate analysis methods<sup>44,45</sup>. In this study we performed PCA analysis and Person correlation test, Fig. 3, Table 3, respectively. PCA analysis, as well as Person correlation test demonstrated that Raman spectroscopy offers higher possibility to distinguish between analysed types of tissues. The reason could be the difference in the physical principles of both techniques<sup>46</sup>, e.g. FTIR spectroscopy is sensitive especially for OH stretching in water, while Raman spectroscopy to—C—C, C=C, and C≡C bonds<sup>47</sup>. A high number of carbon functional groups was observed in biological materials, therefore Raman spectroscopy can give a more precise chemical composition of a sample. Particularly, in Fig. 3 the differences in the peak area were observed for almost all analyzed Raman peaks, and only for two FTIR peaks, which corresponded with functional groups building carbohydrates and protein amino acid. These suggest, that changes observed by us using two complementary methods, which have the greatest impact on endometrial carcinogenesis occur in carbohydrates and amino acids, thus consequently in protein. Indeed, lectins, which are carbohydrate-binding proteins permit glycoproteins to be expressed on cancer cells<sup>48</sup>. Furthermore, other research showed, that in the cancer cells there are changes in the expression of genes that code glycoproteins<sup>49</sup>. Importantly, the changes in the carbohydrates and amino acids metabolism and expression, were visible not only in the case of endometrial carcinogenesis, but also in other types of cancer, e.g. colon<sup>50</sup>, ovarian<sup>51</sup> ones etc.”

## Conclusions

In this study, Raman and FTIR spectroscopy were used to evaluate a carcinogenesis in endometrial cancer. Moreover, the methods were compared in terms of efficiency in detection of chemical changes during carcinogenesis of endometrial cancer. Obtained Raman spectra showed, that together with endometrial cancer development, chemical structure and compositions of tissues differs from the ones observed in control samples. Moreover, these changes are statistically significant in all analyzed Raman ranges, while only two peaks of values of absorbance in the FTIR spectra were statistically significant (1073 cm<sup>-1</sup> and 1125 cm<sup>-1</sup>). Furthermore, PCA and HCA analysis of Raman spectra showed that tissues with cancer cells or changes which occurred before cancer (complex atypical hyperplasia) are similar to each other, but they are different from the control samples or changes which can be observed in non-cancerous endometrial tissues. Correlation was found between non-cancerous tissues, as well as between cancer and complex atypical hyperplasia tissues in the correlation test. In contrast, PCA and HCA results as well as correlation obtained from FTIR data showed, that samples which are not histologically similar, correlated and were characterized by similar chemical changes. Overall, our results indicate that Raman spectroscopy is more effective than FTIR in assessing the development of carcinogenesis in endometrial cancer. Using Raman spectroscopy helps to investigate information about atrophic endometrium, complex atypical hyperplasia and endometrioid adenocarcinoma development, while FTIR spectroscopy facilitates differentiation between healthy, cancer and polyp endometrial tissues. Consequently, FTIR spectroscopy methods can be used to detect more visible changes in pathomorphologic images, but not in the case of very similar carcinogenesis phase. Importantly, the amount of tissues collected from the women at different age investigated in this study

was not extensive. Therefore, further tests should be performed to confirm the obtained results. Moreover, the results need to be verified by other researchers. Moreover, application of Raman and FTIR spectroscopies as diagnostics tools in carcinogenesis process should be studied regarding other organs.

## Materials and methods

**Materials limitation.** The study was conducted under Institutional Review Board (No. 9/05/2019) at Rzeszow University. The experimental protocols used in this study were approved by the institutional ethics committees (IECs) of the University of Rzeszow, and were carried out in accordance with the approved guidelines. Informed consents were obtained from all subjects. All studied material (59 tissues samples) was obtained from the women treated in Podkarpackie Gynecology and Obstetrics Ward, Oncology Center in Brzozow, Gynecology and Obstetrics Clinic of the Provincial Clinical Hospital No. 1 in Rzeszow and the Municipal Hospital in Rzeszow. All samples were divided into 5 groups: control (17 tissues); atrophic endometrium (12 tissues); complex atypical hyperplasia (8 tissues); endometrial polyp (6 tissues); endometrioid adenocarcinoma (16 tissues). The medical characteristics of patients, from whom we obtained samples, is presented in Table S1. Moreover, particularly important is the issue that samples were collected from 16 patients over several years, where each next sample were characterized by other endometrial changes. Furthermore, noteworthy is relatively small number of tissue samples in overall group and distribution of their subgroups. It is connected with character of the Clinic and does not reflect their frequency in population.

**Materials preparation.** All obtained materials were prepared like in the paper Depciuch et al.<sup>52</sup>. First, tissues were placed for 12 h in a liquid fixative. Secondly, ethanol within the tissue was gradually replaced with xylene. A paraffin infiltration of the tissue was performed at a temperature of 52 °C. When infiltration sections were embedded in paraffin block, they were prepared by pouring liquefied paraffin into a metal mold and thin piece of tissue was inserted with the appropriate spatial orientation. Each section was flattened on a hot water surface. For the spectroscopic measurements, the obtained samples were placed on CaF<sub>2</sub> slides. Moreover, an immunohistochemical diagnostics was performed for each obtained sample.

## Methods

All methods were performed in accordance with the relevant guidelines and regulations.

**FT-Raman measurement.** For the FT-Raman spectra Nicolet NXR 9650 FT-Raman Spectrometer was used. The spectrometer has an Nd:YAG laser (1064 nm) and a germanium detector. The measurement range was between 150 and 3.700 cm<sup>-1</sup>. The laser power was 0.5 W. Moreover, each sample was measured using 64 scans with 8 cm<sup>-1</sup> resolution. All obtained spectra were analyzed by OPUS software using baseline correction, smoothing (7 points) and normalization using vector normalization functions.

**FTIR measurements.** Vertex 70v spectrometer by Bruker was used to obtain FTIR spectra of all analyzed samples. Moreover, Attenual Total Reflectance (ATR) technique was used with diamond crystal. All samples were measured in the IR range between 400 and 4000 cm<sup>-1</sup> using 32 scans with 2 cm<sup>-1</sup> resolutions. Moreover, all measurements were made in triplicate. The obtained spectra were normalized, smoothed and baseline corrections were made in the OPUS software.

**Statistics—multivariate analysis, Pearson correlation test.** Principal component analysis (PCA) was performed to obtain information about the spectra variation among the types of samples. PCA reduced the dimensionality, the number of variables of the data, while maintaining as much variance as possible. PCA was performed based on the selected spectral regions, which were determined after counting the average values of Raman intensities and absorbance of analyzed peaks. The statistical significance of the calculated values of Raman intensities and absorbance were analyzed by one-way ANOVA followed by the Tukey's test (Statistica 10). The obtained experimental results were represented as the means ± SEM (the standard error of the mean). Furthermore, HCA analysis was performed to determine similarity between each group of samples. Moreover, to obtain information about correlation between obtained Raman and FTIR spectra of measured samples, the Pearson's test was performed with  $p < 0.05$  and significance level of 95%. All analyses were performed using Past 3.0 software. In addition, considering low number of samples in this study, Partial Least Square (PLS) analysis was performed. It was used in case of multicollinearity problems associated with complex biological data when the number of predictors was much larger than the number of samples in tall or wide data sets, like in this study. Moreover, Variables importance in projection (VIP) was calculated to define the most important vibrational band associated with the separation between each type of measured samples. The PLS analysis and VIP factor were calculated using Origin 2019 software. Random forest as well as second learning machine method C5.0 classification model were calculated to obtain information about accuracy of Raman and FTIR spectroscopy in distinguishing of analyzed samples. We used the suggested Random forest algorithm<sup>53</sup> and additionally applied the traditional C5.0 single decision tree algorithm well known in the literature<sup>54</sup>. The R environment and the Random Forest and C5.0 software packages were used to conduct the experiments.

Received: 11 December 2020; Accepted: 8 April 2021

Published online: 27 April 2021

## References

- Rodriguez, A. C., Blanchard, Z., Maurer, K. A. & Gertz, J. Estrogen signaling in endometrial cancer: a key oncogenic pathway with several open questions. *Horm. Cancer* **10**, 51–63 (2019).
- Cornel, K. M. C. *et al.* Gene promoter methylation in endometrial carcinogenesis. *Pathol. Oncol. Res.* **25**, 659–667 (2019).
- Reed, S. D. *et al.* Complex hyperplasia with and without atypia: clinical outcomes and implications of progestin therapy. *Obstet. Gynecol.* **116**, 365–373 (2010).
- Barnas, E. *et al.* Simultaneous FTIR and Raman spectroscopy in endometrial atypical hyperplasia and cancer. *Int. J. Mol. Sci.* **21**, 4828 (2020).
- Epstein, J. B., Zhang, L. & Rosin, M. Advances in the diagnosis of oral premalignant and malignant lesions. *J. Can. Dent. Assoc.* **68**, 617–621 (2002).
- Khanmohammadi, M., Bagheri, G. A., Samani, S., Ghasemi, K. & Asahuri, A. Application of linear discriminant analysis and attenuated total reflectance Fourier transform infrared microspectroscopy for diagnosis of colon cancer. *Pathol. Oncol. Res.* **17**, 435–441 (2011).
- Bergner, N. *et al.* Tumor margin identification and prediction of the primary tumor from brain metastases using FTIR imaging and support vector machines. *Analyst* **138**, 3983–3990 (2013).
- Poste, G. Bring on the biomarkers. *Nature* **469**, 156–157 (2011).
- Holmstrom, B. *et al.* Prostate specific antigen for early detection of prostate cancer. Longitudinal study. *BMJ* **339**, b3537 (2009).
- Seregini, E., Coli, A. & Mazzucca, N. Circulating tumour markers in breast cancer. *Eur. J. Nucl. Med. Mol. Imaging* **31**, S15–S22 (2004).
- Soletormos, G., Nielsen, D., Schioler, V., Mouridsen, H. & Dombernowsky, P. Monitoring different stages of breast cancer using tumour markers CA 15-3, CEA and TPA. *Eur. J. Cancer* **40**, 481–486 (2004).
- Kaznowska, E. *et al.* Application of infrared spectroscopy for the identification of squamous cell carcinoma (lung cancer). Preliminary study. *Infrared Phys. Technol.* **89**, 282–290 (2018).
- Depciuch, J. *et al.* Spectroscopic analysis of normal and neoplastic (WI-FTC) thyroid tissue. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **204**, 18–24 (2018).
- Depciuch, J., Kaznowska, E., Kozirowska, A. & Cebulski, J. Verification of the effectiveness of the Fourier transform infrared spectroscopy computational model for colorectal cancer. *J. Pharm. Biomed. Anal.* **145**, 611–615 (2017).
- Zhiwei, H. *et al.* Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int. J. Cancer* **107**(6), 1047–1052 (2003).
- Chaber, R. *et al.* Application of infrared spectroscopy in the identification of Ewing sarcoma: a preliminary report. *Infrared Phys. Technol.* **83**, 200–205 (2017).
- Morris, M. D. & Mandair, G. S. Raman assessment of bone quality. *Clin. Orthop. Relat. Res.* **469**(8), 2160–2169 (2011).
- Sbroscia, M. *et al.* Thyroid cancer diagnosis by Raman spectroscopy. *Sci. Rep.* **10**, 13342 (2020).
- Zúñiga, W. C. *et al.* Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy. *Sci. Rep.* **9**, 14639 (2019).
- Medipally, D. K. R. *et al.* Monitoring radiotherapeutic response in prostate cancer patients using high throughput FTIR spectroscopy of liquid biopsies. *Cancer (Basel)* **11**, 925 (2019).
- Laska, J. & Widlarz, J. Spectroscopic and structural characterization of low molecular weight fractions of polyaniline. *Polymer* **46**, 1485–1495 (2005).
- Seballos, L., Zhang, J. Z. & Sutphen, R. Surface-enhanced Raman scattering detection of lysophosphatidic acid. *Anal. Bioanal. Chem.* **383**, 763–767 (2005).
- 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).
- Movasaghi, Z., Rehman, S. & Rehman, I. U. Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **42**, 493–541 (2007).
- Shaikh, R. *et al.* A comparative evaluation of diffuse reflectance and Raman spectroscopy in the detection of cervical cancer. *J. Biophotonics* **10**, 242–252 (2017).
- Argov, S. *et al.* Inflammatory bowel diseases as an intermediate stage between normal and cancer: a FTIR-microspectroscopy approach. *Biopolymers* **75**, 384–392 (2004).
- Richter, T. *et al.* Identification of tumor tissue by FTIR spectroscopy in combination with positron emission tomography. *Vibr. Spectrosc.* **28**, 103–110 (2002).
- Naumann, D. FT-infrared and FT-Raman spectroscopy in biomedical research. In *Infrared and Raman Spectroscopy of Biological Materials* (eds Griemlich, H. U. & Yan, B.) 323–377 (Marcel Dekker, 2001).
- Wong, P. T., Wong, R. K., Caputo, T. A., Godwin, T. A. & Rigas, B. Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10988–10992 (1991).
- Fujioka, N., Morimoto, Y., Arai, T. & Kikuchi, M. Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy. *Cancer Detect. Prev.* **28**, 32–36 (2008).
- Conti, C. *et al.* FT-IR microimaging spectroscopy: a comparison between healthy and neoplastic human colon tissues. *J. Mol. Struct.* **881**, 46–51 (2008).
- Patel, I. I. *et al.* High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares. *Analyst* **136**, 4950–4959 (2011).
- Notarstefano, V. *et al.* Vibrational characterization of granulosa cells from patients affected by unilateral ovarian endometriosis: new insight from infrared and Raman microspectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **212**, 206–214 (2019).
- Parlatan, U. *et al.* Raman spectroscopy as a non-invasive diagnostic techniques for endometriosis. *Sci. Rep.* **9**, 19795 (2019).
- Gajjar, K. *et al.* Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer. *Analyst* **138**, 3917–3926 (2013).
- Paraskevaidi, M. *et al.* Detecting endometrial cancer by blood spectroscopy: a diagnostic cross-sectional study. *Cancer (Basel)* **12**, 1256 (2020).
- Taylor, S. E. *et al.* Infrared spectroscopy with multivariate analysis to interrogate endometrial tissue: a novel and objective diagnostic approach. *Br. J. Cancer* **104**, 790–797 (2011).
- Silverberg, S. Problems in the differential diagnosis of endometrial hyperplasia and carcinoma. *Mod. Pathol.* **13**, 309–327 (2000).
- Sherman, M. E. Theories of endometrial carcinogenesis: a multidisciplinary approach. *Mod. Pathol.* **13**, 295–308 (2000).
- Wei, S. C. *et al.* Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nat. Cell Biol.* **17**(5), 678–688 (2015).
- Fang, M., Yuan, J., Peng, C. & Li, Y. Collagen as a double-edged sword in tumor progression. *Tumour Biol.* **35**, 2871–2882 (2014).
- Nijkang, N. P., Anderson, L., Markham, R. & Manconi, F. Endometrial polyps: Pathogenesis, sequelae and treatment. *SAGE Open Med.* **7**, 2050 (2019).
- Depciuch, J. *et al.* Phospholipid-protein balance in affective disorders: analysis of human blood serum using Raman and FTIR spectroscopy. A pilot study. *J. Pharmaceut. Biomed. Anal.* **131**, 287–296 (2016).
- Walsh, M. J. *et al.* ATR microspectroscopy with multivariate analysis segregates grades of exfoliative cervical cytology. *Biochem. Biophys. Res. Commun.* **352**, 213–219 (2007).

45. Martin, F. L. *et al.* Identifying variables responsive for clustering in discriminant analysis of data from infrared microspectroscopy of a biological sample. *J. Comput. Biol.* **14**, 1176–1184 (2007).
46. Tuschel, D. Practical group theory and Raman spectroscopy, part I: normal vibrational modes. *Spectroscopy* **29**, 14–23 (2014).
47. Wartewig, S. *IR and Raman Spectroscopy: Fundamental Processing*, XXX (Wiley-VCH Verlag GbmH & Co., 2003).
48. Zarogoulidis, P. *et al.* Use of proteins as biomarkers and their role in carcinogenesis. *J. Cancer* **6**(1), 9–18 (2015).
49. Siegsmun, M. *et al.* Association of the P-glycoprotein transporter MDR1(C3435T) polymorphism with the susceptibility to renal epithelial tumors. *J. Am. Soc. Nephrol.* **13**, 1847–1854 (2002).
50. Konno, A. *et al.* Carbohydrate expression profile of colorectal cancer cells is relevant to metastatic pattern and prognosis. *Clin. Exp. Metastasis* **19**(1), 61–70 (2002).
51. Qu, W. *et al.* B19, a novel monocarbonyl analogue of curcumin, induces human ovarian cancer cell apoptosis via activation of endoplasmic reticulum stress and the autophagy signaling pathway. *Int. J. Biol. Sci.* **9**, 766–777 (2013).
52. Depciuch, J. *et al.* Comparing paraffined and deparaffinized breast cancer tissue samples and an analysis of Raman spectroscopy and infrared methods. *Infrared Phys. Technol.* **76**, 217–226 (2016).
53. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
54. Quinlan, J. R. *C4.5: Programs for Machine Learning* (Morgan Kaufmann Publishers Inc., 1993).

### Author contributions

Conceptualization, E.B. and A.S. Methodology, E.K., K.Ł., J.S.-M. and J.D.; Validation, J.D.; Formal analysis, J.D. and K.Ł.; Writing—original draft preparation, J.D. and E.B.; Writing—review and editing, J.C. and A.S.; Visualization, J.D.; Supervision, J.C. and A.S. All authors have read and agreed to the published version of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-88640-7>.

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

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2021