

Article

Potential of mid-infrared spectroscopy as a non-invasive diagnostic test in urine for endometrial or ovarian cancer

Paraskevaidi, Maria, Medeiros-De-morais, Camilo De lelis, Lima, Kassio M.G., Ashton, Katherine, Stringfellow, Helen F., Martin-Hirsch, Pierre L. and Martin, Francis L

Available at <http://clock.uclan.ac.uk/23115/>

Paraskevaidi, Maria, Medeiros-De-morais, Camilo De lelis ORCID: 0000-0003-2573-787X, Lima, Kassio M.G., Ashton, Katherine, Stringfellow, Helen F., Martin-Hirsch, Pierre L. and Martin, Francis L ORCID: 0000-0001-8562-4944 (2018) Potential of mid-infrared spectroscopy as a non-invasive diagnostic test in urine for endometrial or ovarian cancer. Analyst, 143 . pp. 3156-3163. ISSN 0003-2654

It is advisable to refer to the publisher's version if you intend to cite from the work.

<http://dx.doi.org/10.1039/C8AN00027A>

For more information about UCLan's research in this area go to <http://www.uclan.ac.uk/researchgroups/> and search for <name of research Group>.

For information about Research generally at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

1 **Potential of mid-infrared spectroscopy as a non-invasive**
2 **diagnostic test in urine for endometrial or ovarian cancer**

3 Maria Paraskevaïdi^{a,1}, Camilo L.M. Morais^{a,b}, Kássio M.G. Lima^b, Katherine M. Ashton^c,
4 Helen F. Stringfellow^c, Pierre L. Martin-Hirsch^d and Francis L. Martin^{a,1}

5 ^a*School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1*
6 ²*HE, UK; ^bInstitute of Chemistry, Biological Chemistry and Chemometrics, Federal University*
7 <sup>of Rio Grande do Norte, Natal 59072-970, Brazil; ^cPathology Department, Lancashire
8 <sup>Teaching Hospitals NHS Foundation, Preston PR2 9HT, UK; ^dDepartment of Obstetrics and
9 ^{Gynaecology, Lancashire Teaching Hospitals NHS Foundation, Preston PR2 9HT, UK}</sup></sup>

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24 ¹To whom correspondence should be addressed. Email: mparaskevaïdi@uclan.ac.uk or
25 flmartin@uclan.ac.uk; Tel: +44 (0) 1772 89 6482

26 **Abstract**

27

28 The current lack of an accurate, cost-effective and non-invasive test that would allow for
29 screening and diagnosis of gynaecological carcinomas, such as endometrial and ovarian cancer,
30 signals the necessity for alternative approaches. The potential of spectroscopic techniques in
31 disease investigation and diagnosis has been previously demonstrated. Here, we used
32 attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy to analyse
33 urine samples from women with endometrial (n=10) and ovarian cancer (n=10), as well as from
34 healthy individuals (n=10). After applying multivariate analysis and classification algorithms,
35 biomarkers of disease were pointed out and high levels of accuracy were achieved for both
36 endometrial (95% sensitivity, 100% specificity; accuracy: 95%) and ovarian cancer (100%
37 sensitivity, 96.3% specificity; accuracy 100%). The efficacy of this approach, in combination
38 with the non-invasive method for urine collection, suggest a potential diagnostic tool for
39 endometrial and ovarian cancers.

40

41

42

43

44

45

46

47 **Keywords:** infrared spectroscopy; chemometrics; non-invasive; ovarian cancer; endometrial
48 cancer; diagnosis

49 1. Introduction

50

51 Endometrial and ovarian cancers are the commonest cancers in post-menopausal
52 women worldwide. In the UK alone, 9,300 women develop endometrial cancer, of whom 2,100
53 die annually ¹; ovarian cancer affects 7,300 women resulting in 4,100 deaths per year (*i.e.*,
54 ~40% overall survival) ². The epidemiology of endometrial and ovarian cancer is closely
55 entwined, histological subtypes of endometrial cancer mirror subtypes found in ovarian cancer
56 and the same risk factors seem to influence both diseases ³. Endometrial cancer is often
57 symptomatic at an early stage (stage I) when there is still time for treatment ⁴. In the case of
58 ovarian carcinoma, however, symptoms present late in most cases and after the cancer has
59 already metastasized within the abdomen, resulting in late-stage disease and poor prognoses ⁵.
60 An accurate and early diagnosis of both diseases, and especially ovarian cancer, is of major
61 need as it would permit an early intervention and potentially early-stage diagnosis and
62 consequently improved prognosis.

63 The gold standard for diagnosis of endometrial cancer is biopsy performed either in an
64 outpatient or inpatient setting after a patient presents with symptomatic bleeding. For ovarian
65 cancer diagnosis in patients with symptoms, women initially undergo a pelvic examination,
66 followed by measurement of serum cancer antigen (CA-125); if symptoms persist in the
67 absence of raised CA-125 levels, an abdominal and transvaginal ultrasound follow ⁵. In
68 asymptomatic women for ovarian cancer screening, a combination of these biomarkers is used.
69 In the future, with the escalating incidence of endometrial cancer secondary to obesity, there
70 might be a role for screening for disease. All of the above-mentioned diagnostic approaches
71 have drawbacks, either being invasive (*e.g.*, biopsy) or expensive (*e.g.*, ultrasound). Even
72 though a blood biomarker would be an ideal diagnostic approach, CA-125 has now been found
73 to be unsuitable for early-stage diagnosis as it is only elevated in 50% of the individuals ⁶. A

74 number of research groups are actively investigating the utility of multiple biomarkers for a
75 more accurate diagnosis of endometrial and ovarian cancers ^{6,7}. However, current methods of
76 biomarker identification are heavily dependent on multiplex assays and molecular techniques
77 which are costly.

78 Vibrational spectroscopy has gained increasing attention in the recent years due to its
79 potential as a diagnostic tool for various diseases, by providing chemical and structural
80 information of the sample in use ^{8,9}. Both infrared (IR) and Raman spectroscopic techniques
81 have been extensively used for cancer diagnostics using tissue, cells or biofluids, such as blood
82 plasma/serum, urine, bile, ascitic fluid and cerebrospinal fluid ^{10,11}. A screening or diagnostic
83 test should be non-invasive to facilitate compliance and, as such, venepuncture and urine
84 analysis are ideal. Blood- and urine-based spectroscopy have already been applied successfully
85 in studies including brain ¹², breast ¹³, gynaecological ^{14,15} and other types of cancer. In the
86 present study, attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy
87 was used to analyse urine samples from women with endometrial and ovarian cancer and
88 facilitate towards their segregation from healthy controls. The classification models that were
89 employed to distinguish between these groups were partial least squares discriminant analysis
90 (PLS-DA), principal component analysis with support vector machines (PCA-SVM) and
91 genetic algorithm with linear discriminant analysis (GA-LDA).

92 2. Materials and Methods

93

94 2.1 Study population and sample collection

95

96 All samples were collected at Royal Preston Hospital UK after obtaining ethical
97 approval (16/EE/0010). All experiments were performed in accordance with relevant laws and
98 guidelines, and approved by the ethics committee at University of Central Lancashire (UCLan).
99 Informed consent was obtained from all human subjects. Urine samples were collected from

100 30 individuals: 10 healthy women with no symptoms of cancer who were used as controls, 10
101 women with endometrial cancer and 10 women with ovarian cancer. All urine specimens were
102 obtained after patients were administered a general anaesthetic prior to hysterectomy for benign
103 or malignant indications. All patients had undergone a period of at least 6 hours fasting; the
104 suggested preoperative fasting time is 6-8 hours for light meals and 2 hours for fluids ^{16, 17}.
105 Prolonged fasting (12-16 h) should be avoided as it triggers gluconeogenesis precipitation and
106 increases the organic response to trauma ¹⁸. Specimens were obtained after aseptic preparation
107 of the urethra and after catheterisation, thus avoiding any contamination. All cases were staged
108 according to the guidelines by the International Federation of Gynecologic Oncology (FIGO).
109 Samples were kept at -80°C until the time of spectroscopic analysis. Before analysis, all
110 samples were left to thaw at room temperature and 50 µl were deposited on low-emissivity
111 (low-E) slides (MirrIR Low-E slides, Kevley Technologies, USA); they were then left to air-
112 dry for approximately 45 minutes. All urines were taken pre-operatively on the day of surgery
113 in both controls and disease patients and patients had not received any treatment for the disease
114 prior to surgery. All endometrial and ovarian cancers were high grade cancers. Sub-group
115 analysis for other incidental diseases or factors such as type 2 diabetes, hypertension or
116 medication was not performed in this study.

117 2.2 Spectroscopic analysis

118

119 ATR-FTIR spectroscopy was employed for the analysis of the urine samples. A Tensor
120 27 FTIR spectrometer with a Helios ATR attachment (Bruker Optics Ltd, Coventry, UK) which
121 contained a diamond crystal, was used for the data collection. The OPUS 7.2 software was used
122 for spectral acquisition. Spectral resolution was set at 8 cm⁻¹ and mirror velocity at 2.2 kHz; 32
123 scans were acquired for each spectrum for optimal signal-to-noise ratio. The diamond crystal
124 was cleaned with distilled water after the use of each sample and a background spectrum was

125 collected to eliminate atmospheric changes. A CCTV camera was used for visualisation and
126 navigation across the sample's surface; ten spectra were acquired from different locations of
127 each sample to minimize bias.

128 2.3 Data analysis

129 After collection, the raw spectra need to be pre-processed in order to account for
130 inconsistencies relating to the experimental procedure and spectral acquisition. Pre-processing
131 and computational analysis of the data was performed using PLS Toolbox version 7.9.3
132 (Eigenvector Research, Inc., Manson, USA) and an in-house developed IRootLab toolbox
133 (<http://trevisanj.github.io/irootlab/>)¹⁹. For the purposes of this study, raw spectra were initially
134 pre-processed as following: cut to the bio-fingerprint region (1800-900 cm⁻¹), rubberband
135 baseline corrected and vector normalised. Rubberband baseline correction is used to correct
136 underlying oscillations on the baseline of the spectra which can be caused by scattering effects,
137 reflection, temperature, concentration, among other instrumental anomalies that render
138 wavenumbers, known for having no absorption, with absorbance values different from zero²⁰.

139 For further classification spectra were divided into training (60%, $n = 18$ patients) and
140 test (40%, $n = 12$ patients) sets using the Kennard-Stone sample selection algorithm²¹. Partial
141 least squares discriminant analysis (PLS-DA), principal component analysis with support
142 vector machines (PCA-SVM) and genetic algorithm with linear discriminant analysis (GA-
143 LDA) were used as classification methods. PLS-DA is a linear classification technique that
144 uses partial least squares (PLS) to find a straight line that divide the classes spaces²². PCA-
145 SVM makes use of principal component analysis (PCA)²³ for data compression; the PCA
146 scores are then used as input variables for a support vector machine (SVM) classifier²⁴. The
147 SVM classifier was based on a radial basis function (RBF) kernel which is used to transfer the
148 data to a feature space by means of a non-linear discriminant criterion; a linear decision surface

149 is then constructed in this feature space to separate the classes analysed ²⁵. Both PLS-DA and
150 PCA-SVM were optimized using cross-validation venetian blinds in a “leave one patient out”
151 fashion (10 splits with 1 sample per split). GA-LDA was optimized using an external validation
152 data set having half of the samples of the test set. The algorithm was applied three times, using
153 100 generations with 200 chromosomes, and the best model was selected. Crossover and
154 mutation probabilities were set to 60% and 10%, respectively.

155 In order to study the differences at specific wavenumbers, we implemented a simple
156 approach, namely difference-between-means (D-B-M) spectra, which subtracts the mean
157 spectra from a reference class (*i.e.*, healthy controls). A peak-detecting algorithm was then used
158 to denote six of the most differentiating peaks.

159 2.4 Availability of data

160

161 All data (raw and pre-processed spectra) along with appropriate code identifiers have
162 been uploaded onto the publicly accessible data repository Figshare
163 ([https://figshare.com/articles/Potential_of_mid-infrared_spectroscopy_as_a_non-](https://figshare.com/articles/Potential_of_mid-infrared_spectroscopy_as_a_non-invasive_diagnostic_test_for_endometrial_or_ovarian_cancer_in_urine/5929516)
164 [invasive_diagnostic_test_for_endometrial_or_ovarian_cancer_in_urine/5929516](https://figshare.com/articles/Potential_of_mid-infrared_spectroscopy_as_a_non-invasive_diagnostic_test_for_endometrial_or_ovarian_cancer_in_urine/5929516)).

165 2.5 Statistical analysis

166

167 The classification performance of the chemometric algorithms was evaluated according
168 to the accuracy, sensitivity and specificity on the test set. The accuracy (AC) represents the
169 number of samples correctly classified considering true and false negatives; sensitivity (SENS)
170 and specificity (SPEC) measure the proportion of positives and negatives that are correctly
171 identified, respectively ²⁶. These parameters are calculated as follows:

$$172 \text{AC}(\%) = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \right) \times 100 \quad (3)$$

173
$$\text{SENS}(\%) = \left(\frac{\text{TP}}{\text{TP} + \text{FN}} \right) \times 100 \quad (4)$$

174
$$\text{SPEC}(\%) = \left(\frac{\text{TN}}{\text{TN} + \text{FP}} \right) \times 100 \quad (5)$$

175 where TP stands for true positive, TN for true negative, FP for false positive and FN for false
176 negative.

177 The peaks that were responsible for the differentiation after the D-B-M spectra
178 approach, were imported into GraphPad Prism 7 (GraphPad Software, Inc., La Jolla, CA,
179 92037, USA) to conduct statistical analyses and calculate the *P*-values. Differences between
180 the two groups (*i.e.*, healthy vs cancer) were assessed using a Student's t-test (two-tailed, non-
181 parametric, Mann-Whitney test, 95% confidence interval). The data were expressed as the
182 mean \pm standard deviation (SD). A *P*-value of 0.05 or less was considered significant.

183 3. Results

184 3.1 Segregation between cancer patients and controls

185

186 All spectra were pre-processed before comparison of the cancer patients with the
187 healthy controls. Supplementary Fig. 1A shows the average pre-processed spectra for each
188 class. The most discriminatory peaks for the comparison between endometrial cancer and
189 healthy women were: 1593 cm^{-1} ($P < 0.0001$, 95% CI = -0.0179 to -0.012), 1508 cm^{-1} (P
190 < 0.0001 , 95% CI = 0.0038 to 0.0103), 1462 cm^{-1} ($P < 0.0001$, 95% CI = -0.0115 to -0.0058),
191 1400 cm^{-1} ($P < 0.0001$, 95% CI = 0.0107 to 0.0161), 1335 cm^{-1} ($P < 0.0001$, 95% CI = 0.0053
192 to 0.0093), 1041 cm^{-1} ($P < 0.0001$, 95% CI = 0.0063 to 0.0118) (Supplementary Fig. 1B). Fig.
193 1 shows the differences in the absorbance of the above-mentioned peaks; a general increase
194 was denoted in the endometrial cancer patients with the exception of the peaks at 1593 cm^{-1}
195 and 1462 cm^{-1} which showed decreased levels. The means and SD values for these peaks were:
196 1593 cm^{-1} (mean/SD for healthy: 0.18/0.011; mean/SD for cancer: 0.164/0.0098), 1508 cm^{-1}
197 (mean/SD for healthy: 0.0445/0.0115; mean/SD for cancer: 0.0506/0.0104), 1462 cm^{-1}

198 (mean/SD for healthy: 0.0958/0.0151; mean/SD for cancer: 0.0833/0.0064), 1400 cm^{-1}
199 (mean/SD for healthy: 0.0509/0.007; mean/SD for cancer: 0.0645/0.0095), 1335 cm^{-1}
200 (mean/SD for healthy: 0.0237/0.0053; mean/SD for cancer: 0.0315/0.007), 1041 cm^{-1}
201 (mean/SD for healthy: 0.0286/0.0083; mean/SD for cancer: 0.0383/0.0109).

202 The peaks responsible for differentiation between healthy and ovarian cancer patients
203 were: 1597 cm^{-1} ($P < 0.0001$, 95% CI = -0.0173 to -0.0114), 1508 cm^{-1} ($P < 0.0001$, 95% CI =
204 0.0038 to 0.0103), 1408 cm^{-1} ($P < 0.0001$, 95% CI = 0.0102 to 0.0149), 1373 cm^{-1} ($P < 0.0001$,
205 95% CI = 0.0076 to 0.0122), 1231 cm^{-1} ($P < 0.0001$, 95% CI = 0.0042 to 0.0074), 1041 cm^{-1} (P
206 < 0.0001 , 95% CI = 0.0063 to 0.0118) (Supplementary Fig. 1B). Similarly, to endometrial
207 cancer, the majority of the peaks showed increased absorbance when cancer was present, apart
208 from the peak at 1597 cm^{-1} (Fig. 2). Means and SD values for each of the abovementioned
209 peaks were: 1597 cm^{-1} (mean/SD for healthy: 0.18/0.0104; mean/SD for cancer: 0.166/0.0098),
210 1508 cm^{-1} (mean/SD for healthy: 0.0445/0.0115; mean/SD for cancer: 0.0506/0.0104), 1408
211 cm^{-1} (mean/SD for healthy: 0.0548/0.0067; mean/SD for cancer: 0.0676/0.0085), 1373 cm^{-1}
212 (mean/SD for healthy: 0.0326/0.0075; mean/SD for cancer: 0.0427/0.0085), 1231 cm^{-1}
213 (mean/SD for healthy: 0.0198/0.0079; mean/SD for cancer: 0.0256/0.0079), 1041 cm^{-1}
214 (mean/SD for healthy: 0.0286/0.0083; mean/SD for cancer: 0.0383/0.0109).

215 3.1 Classification algorithms to calculate diagnostic accuracy

216 All classification algorithms (PLS-DA, PCA-SVM and GA-LDA) were applied to the
217 data after the same pre-processing (cut to the bio-fingerprint region [1800-900 cm^{-1}],
218 rubberband baseline correction and vector normalisation).

219 PLS-DA was employed to differentiate healthy, endometrial and ovarian cancer
220 samples using 10 latent variables (LVs), accounting for 96.02% of cumulative variance (Fig.
221 3A). The number of LVs was selected according to the lowest error of cross-validation (15.4%

222 for healthy; 16.2% for endometrial cancer; and 12.5% for ovarian cancer) and maximum
223 explained variance (96.02%). PLS-DA loadings are shown in Fig. 3B, where the first three LVs
224 have higher coefficients at $\sim 1041\text{ cm}^{-1}$, $\sim 1082\text{ cm}^{-1}$, $\sim 1462\text{ cm}^{-1}$, $\sim 1547\text{ cm}^{-1}$, $\sim 1589\text{ cm}^{-1}$ and
225 $\sim 1670\text{ cm}^{-1}$. The predicted classes for each sample spectrum analysed by PLS-DA are shown
226 in Fig. 3C (healthy), 3D (endometrial cancer) and 3E (ovarian cancer), where a degree of
227 superposition is observed among the three classes. Only the ovarian cancer dataset showed
228 clearer separation from the other two classes.

229 PCA-SVM was performed using 10 principal components (PCs), accounting for
230 97.33% of cumulative variance (Fig. 4A). The PCA loadings (Fig. 4B) had higher coefficients
231 in regions very similar to PLS-DA: $\sim 1042\text{ cm}^{-1}$, $\sim 1090\text{ cm}^{-1}$, $\sim 1130\text{ cm}^{-1}$, $\sim 1462\text{ cm}^{-1}$, ~ 1508
232 cm^{-1} , $\sim 1543\text{ cm}^{-1}$, $\sim 1589\text{ cm}^{-1}$ and $\sim 1667\text{ cm}^{-1}$. The predicted classes for each sample are shown
233 in Fig. 4C (healthy), 4D (endometrial cancer) and 4E (ovarian cancer). In comparison to PLS-
234 DA, Fig. 4 C-E shows a clearer separation among the classes, with only a few samples being
235 misclassified.

236 GA-LDA classified healthy, endometrial and ovarian cancer with a fitness of 1.53 (Fig.
237 5A). The GA-LDA discriminant function (DF) plot for the three classes is shown in Fig. 5B
238 with clear segregation between the three classes. A total of 20 variables showed differences
239 between the three classes: 922 cm^{-1} , 972 cm^{-1} , 1007 cm^{-1} , 1011 cm^{-1} , 1018 cm^{-1} , 1045 cm^{-1} ,
240 1049 cm^{-1} , 1061 cm^{-1} , 1084 cm^{-1} , 1265 cm^{-1} , 1362 cm^{-1} , 1366 cm^{-1} , 1400 cm^{-1} , 1412 cm^{-1} , 1500
241 cm^{-1} , 1535 cm^{-1} , 1562 cm^{-1} , 1566 cm^{-1} , 1682 cm^{-1} and 1716 cm^{-1} (Fig. 5B).

242 Table 1 shows the classification rates achieved by the three algorithms, with PCA-SVM
243 being superior. Both accuracy and sensitivity values for PCA-SVM model ranged from 92.5%
244 (healthy) to 100% (ovarian cancer); and specificity ranged from 96.3% (ovarian cancer) to
245 100% (endometrial cancer). GA-LDA had accuracy ranging from 90.0% (endometrial cancer)

246 to 98.3% (ovarian cancer); sensitivity ranging from 70.0% (endometrial cancer) to 100%
247 (healthy/ovarian cancer); and specificity ranging from 87.5% (healthy) to 100% (endometrial
248 cancer). PLS-DA was the worst model as only the ovarian cancer data set had quality
249 parameters as good as the other algorithms. Accuracy ranged from 57.5% (endometrial cancer)
250 to 92.5% (ovarian cancer); sensitivity ranged from 62.5% (endometrial cancer) to 87.5%
251 (ovarian cancer); and specificity ranged from 86.3% (healthy) to 90% (ovarian cancer). The
252 classification rates for the training and test sets using all three algorithms are shown in
253 Supplementary Table 1.

254 4. Discussion

255

256 The wavenumbers that were mostly responsible for segregation between the different
257 classes could facilitate as potential diagnostic biomarkers. In endometrial cancer patients the
258 majority of the IR bands, associated with proteins and nucleic acids, were increased in
259 comparison to healthy individuals. This could potentially be due to an elevated concentration
260 of biomolecules, previously suggested as biomarkers for endometrial cancer, such as human
261 epididymis protein 4 (HE4), CA-125 or carcinoembryonic antigen (CEA)^{7, 27, 28}; the increased
262 level of nucleic acid may be caused by the unconstrained proliferation of cells. Only two out
263 of the six discriminatory peaks were lower in the cancer cases; these were attributed to C-C
264 vibrations of phenyl rings of proteins ($\sim 1593\text{ cm}^{-1}$, Amide II) and CH_2 vibrations of lipids
265 ($\sim 1462\text{ cm}^{-1}$). These results could be potentially explained by a number of possible reasons.
266 For instance, preceding research has demonstrated a simultaneous increased degradation of
267 proteins as well as a decreased protein synthesis during cancer cachexia²⁹. Another study
268 demonstrated decreased expression of follicle-stimulating hormone (FSH) in endometrial
269 cancer patients when these were compared to healthy controls³⁰. The same study also showed
270 decreased levels of matrix metalloproteinases (MMP), which is a family of enzymes implicated

271 in normal and pathological processes, and previously suggested as novel biomarkers and/or
272 therapeutic targets in human cancer ³¹. Also, apolipoprotein-1 (ApoA-1), prealbumin (TTR)
273 have also been shown to be decreased in endometrial cancer patients ⁷.

274 With regards to the decrease in the lipid region of endometrial cancer cases, previous
275 work may again justify the results of the current study. After studying a number of lipids in
276 urine, Skotland et. al revealed that increased and/or decreased levels of molecular lipids could
277 be used as non-invasive biomarkers for prostate cancer with high levels of diagnostic accuracy.
278 Therefore, similar conclusions could possibly be extrapolated to endometrial cancer ³².
279 Previous research has also suggested that lipids, and specifically cholesterol, were lower in
280 blood samples of endometrial cancer than controls ³³; this might explain the lower absorbance
281 in the lipid region ($\sim 1462\text{ cm}^{-1}$) of endometrial cancer patients. More recent studies have further
282 confirmed the increased risk of low cholesterol concentration in other types of cancer as well,
283 such as lung, prostate or colon ^{34,35}.

284 When we compared ovarian cancer patients with healthy controls, the discriminatory
285 peaks were mainly attributed to proteins and nucleic acids. Increased levels of these
286 biomolecules were observed in cancerous samples with an exemption of a peak 1597 cm^{-1}
287 which was assigned to C-C phenyl ring of proteins. Continuous research has previously shown
288 that a cancer biomarker can be either upregulated or downregulated. After reviewing several
289 biomarkers, a total of 111 were found significantly altered between ovarian cancer and controls,
290 with $\sim 60\%$ of them being elevated in cancer and $\sim 40\%$ decreased ³⁶. Some of the biomarkers
291 showing lower levels in cancerous state are, for instance, ApoA-1, FSH, microtubule-
292 associated protein 1 light chain 3 (LC3) and epidermal growth factor receptor (EGFR) ^{6,37,38}.
293 Therefore, this may explain the observed decrease in the Amide II region. On the contrary,
294 increased peaks could potentially be attributed to other established biomarkers such as HE4,
295 previously found to be increased in urine samples of ovarian cancer patients, CA-125, cancer

296 antigen 15-3 (CA15-3) and others^{6, 39}. A relatively recent study, also demonstrated increased
297 levels of the anti-apoptotic protein B-cell lymphoma 2 (Bcl-2) when urine samples from
298 ovarian cancer patients were analysed⁴⁰.

299 Three classification algorithms were employed to calculate the diagnostic accuracy
300 with which spectroscopy identified endometrial and ovarian cancer. The optimal approach was
301 PCA-SVM which identified endometrial cancer with 95% sensitivity and 100% specificity
302 (95% accuracy) and ovarian cancer with 100% sensitivity and 96.3% specificity (100%
303 accuracy), which are exceptionally high in comparison to conventional molecular and imaging
304 methods.

305 Previously, numerous studies have investigated blood biomarkers as a relatively non-
306 invasive approach towards diagnosis of endometrial cancer. A study using serum HE4 yielded
307 sensitivity of 45.4% and 95% sensitivity⁴¹; another study developing a multimarker panel for
308 the early detection of endometrial cancer suggested that prolactin could be used as an accurate
309 biomarker with sensitivity and specificity of ~98%³⁰. Combination of three different
310 biomarkers (ApoA-1, prealbumin and transferrin) distinguished normal samples from early-
311 stage endometrial cancers with 71% sensitivity and 88% specificity, as well as normal samples
312 from late-stage cancer with 82% sensitivity and 86% specificity⁷. After reviewing 13 studies,
313 Timmermans *et. al.*, showed that ultrasonography achieved sensitivity and specificity of 90-
314 98% and 35-54%, respectively⁴². Magnetic resonance imaging (MRI) has been shown to detect
315 early and advanced endometrial cancer with high sensitivity (87-100%) and specificity (90-
316 99%) but stage Ic and stage II disease had significantly reduced sensitivity (19-56%) whereas
317 specificity remained high (86-96%)⁴³.

318 Currently, molecular tests measuring serum CA-125 for ovarian cancer, achieve
319 sensitivity of only 50-60% for early-stage disease and specificity of >95%^{6, 44}. Moreover,

320 transvaginal ultrasound (TVS), computed tomography (CT), MRI and power Doppler are of
321 high-cost and achieve sensitivity <90% for early ovarian cases and relatively high false positive
322 results which render them less useful for screening ⁶. Combination of different biomarkers has
323 been shown to achieve higher sensitivity and specificity values. For example, two combinations
324 of serum biomarkers for ovarian cancer are CA-125, CA 72-4, CA 15-3 and macrophage
325 colony-stimulating factor (M-CSF) ⁴⁵, as well as CA-125, ApoA-1, a truncated form of
326 transthyretin and a cleavage fragment of inter-alpha-trypsin inhibitor heavy chain H4 ⁴⁶; the
327 above-mentioned combinations of biomarkers improved sensitivity and specificity to 70-73%
328 and 97-98%, respectively. Even though the improved accuracy is acceptable, there is still room
329 for improvement. A different study found that a blood-based assay of 11 analytes could
330 distinguish ovarian cancer from benign case with sensitivity and specificity of 90% ³⁷.
331 However, an important drawback of molecular methods is their expense and laborious sample
332 preparation and analysis, in contrast to spectroscopic methods which are rapid and label-free.

333 5. Conclusion

334

335 This pilot study demonstrates the efficacy of ATR-FTIR spectroscopy in detecting
336 endometrial and ovarian cancers in urine samples, with high levels of accuracy. Being rapid,
337 non-destructive and at the same time cost-effective, spectroscopy is introduced as an ideal
338 method for studying these types of cancer and could potentially be translated into clinical
339 practise in the future as either a screening or diagnostic test. An adequately powered study will
340 be required to demonstrate the true diagnostic accuracy and validate these preliminary results.
341 Furthermore, the quick and non-invasive nature of urine collection and subsequent analysis has
342 the potential of a preferable vehicle for repeated measurements, thus facilitating monitoring of
343 disease progression/regression/recurrence or even therapeutic response.

344 Conflict of interest

345 There are no conflicts of interest to declare.

346

347 Acknowledgements

348 MP was supported by the Rosemere Cancer Foundation (RCF). CLMM would like to
349 acknowledge CAPES for his research grant. KMGL acknowledges CNPq (grant 305962/2014-
350 0) for financial support.

351

352 References

- 353 1. Cancer Reserach UK, [http://www.cancerresearchuk.org/health-professional/cancer-](http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer/incidence)
354 [statistics/statistics-by-cancer-type/uterine-cancer/incidence](http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer/incidence) (Accessed December 2017).
- 355 2. Cancer Research UK, [http://www.cancerresearchuk.org/health-professional/cancer-](http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Six)
356 [statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Six](http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Six) (Accessed October
357 2017).
- 358 3. D. W. Cramer, *Hematol Oncol Clin North Am*, 2012, **26**, 1-12.
- 359 4. P. Morice, A. Leary, C. Creutzberg, N. Abu-Rustum and E. Darai, *Lancet*, 2016, **387**, 1094-
360 1108.
- 361 5. G. C. Jayson, E. C. Kohn, H. C. Kitchener and J. A. Ledermann, *Lancet*, 2014, **384**, 1376-1388.
- 362 6. Z. Yurkovetsky, S. Skates, A. Lomakin, B. Nolen, T. Pulsipher, F. Modugno, J. Marks, A.
363 Godwin, E. Gorelik and I. Jacobs, *J Clin Oncol*, 2010, **28**, 2159-2166.
- 364 7. G. Farias-Eisner, F. Su, T. Robbins, J. Kotlerman, S. Reddy and R. Farias-Eisner, *Am J Obstet*
365 *Gynecol*, 2010, **202**, 73.e71-73.e75.
- 366 8. M. Paraskevaidi, C. L. M. Morais, K. M. G. Lima, J. S. Snowden, J. A. Saxon, A. M. T.
367 Richardson, M. Jones, D. M. A. Mann, D. Allsop, P. L. Martin-Hirsch and F. L. Martin, *Proc*
368 *Natl Acad Sci USA*, 2017, **114**, E7929-e7938.
- 369 9. K. L. Andrew Chan and S. G. Kazarian, *Chem Soc Rev*, 2016, **45**, 1850-1864.
- 370 10. M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszewski, G.
371 Thieffin and G. D. Sockalingum, *Chem Soc Rev*, 2016, **45**, 1803-1818.
- 372 11. F. L. Martin, J. G. Kelly, V. Llabjani, P. L. Martin-Hirsch, I. I. Patel, J. Trevisan, N. J. Fullwood
373 and M. J. Walsh, *Nat Protocols*, 2010, **5**, 1748-1760.
- 374 12. J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D.
375 Jenkinson, R. W. Lea, C. Walker and M. J. Baker, *J Biophotonics*, 2014, **7**, 189-199.
- 376 13. J. Backhaus, R. Mueller, N. Formanski, N. Szlama, H.-G. Meerpohl, M. Eidt and P. Bugert,
377 *Vib Spectrosc*, 2010, **52**, 173-177.
- 378 14. K. Gajjar, J. Trevisan, G. Owens, P. J. Keating, N. J. Wood, H. F. Stringfellow, P. L. Martin-
379 Hirsch and F. L. Martin, *Analyst*, 2013, **138**, 3917-3926.
- 380 15. S. E. Taylor, K. T. Cheung, I. I. Patel, J. Trevisan, H. F. Stringfellow, K. M. Ashton, N. J.
381 Wood, P. J. Keating, P. L. Martin-Hirsch and F. L. Martin, *Br J Cancer*, 2011, **104**, 790-797.
- 382 16. E. Søreide, L. Eriksson, G. Hirlekar, H. Eriksson, S. Henneberg, R. Sandin and J. Raeder, *Acta*
383 *Anaesthesiol Scand*, 2005, **49**, 1041-1047.
- 384 17. O. Ljungqvist and E. Søreide, *Br J Surg*, 2003, **90**, 400-406.
- 385 18. J. E. de Aguiar-Nascimento and D. B. Dock-Nascimento, *World J Gastrointest Surg*, 2010, **2**,
386 57.
- 387 19. J. Trevisan, P. P. Angelov, A. D. Scott, P. L. Carmichael and F. L. Martin, *Bioinformatics*,
388 2013, **29**, 1095-1097.

- 389 20. J. G. Kelly, J. Trevisan, A. D. Scott, P. L. Carmichael, H. M. Pollock, P. L. Martin-Hirsch and
390 F. L. Martin, *J Proteome Res*, 2011, **10**, 1437-1448.
- 391 21. R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137-148.
- 392 22. R. G. Brereton and G. R. Lloyd, *J Chemometrics*, 2014, **28**, 213-225.
- 393 23. R. Bro and A. K. Smilde, *Anal Methods*, 2014, **6**, 2812-2831.
- 394 24. C. L. Morais, F. S. Costa and K. M. Lima, *Anal Methods*, 2017, **9**, 2964-2970.
- 395 25. C. Cortes and V. Vapnik, *Mach Learn*, 1995, **20**, 273-297.
- 396 26. C. L. Morais and K. M. Lima, *Chemometrics Intellig Lab Syst*, 2017.
- 397 27. L. Zanotti, E. Bignotti, S. Calza, E. Bandiera, G. Ruggeri, C. Galli, G. Tognon, M. Ragnoli, C.
398 Romani and R. A. Tassi, *Clin Chem Lab Med*, 2012, **50**, 2189-2198.
- 399 28. P. B. Panici, G. Scambia, G. Baiocchi, L. Perrone, S. Greggi, F. Battaglia and S. Mancuso,
400 *Gynecol Obstet Invest*, 1989, **27**, 208-212.
- 401 29. K. Smith and M. Tisdale, *Br J Cancer*, 1993, **67**, 680.
- 402 30. Z. Yurkovetsky, S. Ta'asan, S. Skates, A. Rand, A. Lomakin, F. Linkov, A. Marrangoni, L.
403 Velikokhatnaya, M. Winans and E. Gorelik, *Gynecol Oncol*, 2007, **107**, 58-65.
- 404 31. R. Roy, J. Yang and M. A. Moses, *J Clin Oncol*, 2009, **27**, 5287-5297.
- 405 32. T. Skotland, K. Ekroos, D. Kauhanen, H. Simolin, T. Seierstad, V. Berge, K. Sandvig and A.
406 Llorente, *Eur J Cancer*, 2017, **70**, 122-132.
- 407 33. C. A. Swanson, N. Potischman, R. J. Barrett, M. L. Berman, R. Mortel, L. B. Twiggs, G. D.
408 Wilbanks, R. N. Hoover and L. A. Brinton, *Cancer Epidemiol Biomarkers Prev*, 1994, **3**, 575-
409 581.
- 410 34. H. Kikuchi, A. Nanri, A. Hori, M. Sato, K. Kawai, H. Kasai and T. Mizoue, *Nutr Metab*, 2013,
411 **10**, 59.
- 412 35. M. Eichholzer, H. B. Stähelin, F. Gutzwiller, E. Lüdlin and F. Bernasconi, *Am J Clin Nutr*, 2000,
413 **71**, 569-574.
- 414 36. B. M. Nolen and A. E. Lokshin, *Future Oncol*, 2012, **8**, 55-71.
- 415 37. S. D. Amonkar, G. P. Bertenshaw, T. H. Chen, K. J. Bergstrom, J. Zhao, P. Sessaiah, P. Yip
416 and B. C. Mansfield, *PLoS One*, 2009, **4**, e4599.
- 417 38. Y. Shen, D.-D. Li, L.-L. Wang, R. Deng and X.-F. Zhu, *Autophagy*, 2008, **4**, 1067-1068.
- 418 39. J. B. Liao, Y. Y. Yip, E. M. Swisher, K. Agnew, K. E. Hellstrom and I. Hellstrom, *Gynecol
419 Oncol*, 2015, **137**, 430-435.
- 420 40. N. S. Anderson, Y. Bermudez, D. Badgwell, R. Chen, S. V. Nicosia, R. C. Bast and P. A. Kruk,
421 *Gynecol Oncol*, 2009, **112**, 60-67.
- 422 41. R. G. Moore, A. K. Brown, M. C. Miller, D. Badgwell, Z. Lu, W. J. Allard, C. Granai, R. C.
423 Bast and K. Lu, *Gynecol Oncol*, 2008, **110**, 196-201.
- 424 42. A. Timmermans, B. C. Opmeer, K. S. Khan, L. M. Bachmann, E. Epstein, T. J. Clark, J. K.
425 Gupta, S. H. Bakour, T. van den Bosch, H. C. van Doorn, S. T. Cameron, M. G. Giusa, S.
426 Dessole, F. P. Dijkhuizen, G. Ter Riet and B. W. Mol, *Obstet Gynecol*, 2010, **116**, 160-167.
- 427 43. O. Ortashi, S. Jain, O. Emmanuel, R. Henry, A. Wood and J. Evans, *Eur J Obstet Gynecol
428 Reprod Biol*, 2008, **137**, 232-235.
- 429 44. V. Nossov, M. Amneus, F. Su, J. Lang, J. M. T. Janco, S. T. Reddy and R. Farias-Eisner, *Am J
430 Obstet Gynecol*, 2008, **199**, 215-223.
- 431 45. S. J. Skates, N. Horick, Y. Yu, F. J. Xu, A. Berchuck, L. J. Havrilesky, H. W. de Bruijn, A. G.
432 van der Zee, R. P. Woolas, I. J. Jacobs, Z. Zhang and R. C. Bast, Jr., *J Clin Oncol*, 2004, **22**,
433 4059-4066.
- 434 46. Z. Zhang, R. C. Bast, Jr., Y. Yu, J. Li, L. J. Sokoll, A. J. Rai, J. M. Rosenzweig, B. Cameron,
435 Y. Y. Wang, X. Y. Meng, A. Berchuck, C. Van Haaften-Day, N. F. Hacker, H. W. de Bruijn,
436 A. G. van der Zee, I. J. Jacobs, E. T. Fung and D. W. Chan, *Cancer Res*, 2004, **64**, 5882-5890.

437

438

439

440

441

442 Figure Legends

443

444 **Figure 1:** Analysis of the top six discriminatory peaks between healthy controls and
445 endometrial cancer patients.

446 **Figure 2:** Analysis of the top six discriminatory peaks between healthy controls and ovarian
447 cancer patients.

448 **Figure 3:** Cumulative explained variance using PLS-DA (A); PLS-DA loadings on LV1, 2 and
449 3 (B); predicted healthy class *versus* endometrial and ovarian cancer (C); predicted endometrial
450 cancer class *versus* healthy and ovarian cancer (D); predicted ovarian cancer class *versus*
451 healthy and endometrial cancer (E). Class measured 1 = healthy control; 2 = endometrial
452 cancer; 3 = ovarian cancer. LV: Latent Variable.

453 **Figure 4:** Cumulative explained variance using PCA (A); PCA loadings on PC1, 2 and 3 (B);
454 predicted probability of healthy class *versus* endometrial and ovarian cancer (C); predicted
455 probability of endometrial cancer class *versus* healthy and ovarian cancer (D); predicted
456 probability of ovarian cancer class *versus* healthy and endometrial cancer (E). Class measured
457 1 = healthy control; 2 = endometrial cancer; 3 = ovarian cancer. PC: Principal Component.

458 **Figure 5:** Fitness function (A); Discriminant Function (DF) plot (B); and selected variables by
459 GA-LDA (C).