Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer

Ketan Gajjar[1,2], Júlio Trevisan[1,3], Gemma Owens[2], Patrick J. Keating[2], Nicholas J. Wood[2], Helen Stringfellow[2], Pierre L. Martin-Hirsch[2], Francis L. Martin[1]
[1]Centre for Biophotonics, Lancaster Environment Centre, Lancaster University, Lancaster, UK; [2]Department of Obstetrics and Gynaecology, Central Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK; [3]School of Computing and Communications, Lancaster University, Lancaster, UK

List of Figures

List of Tables

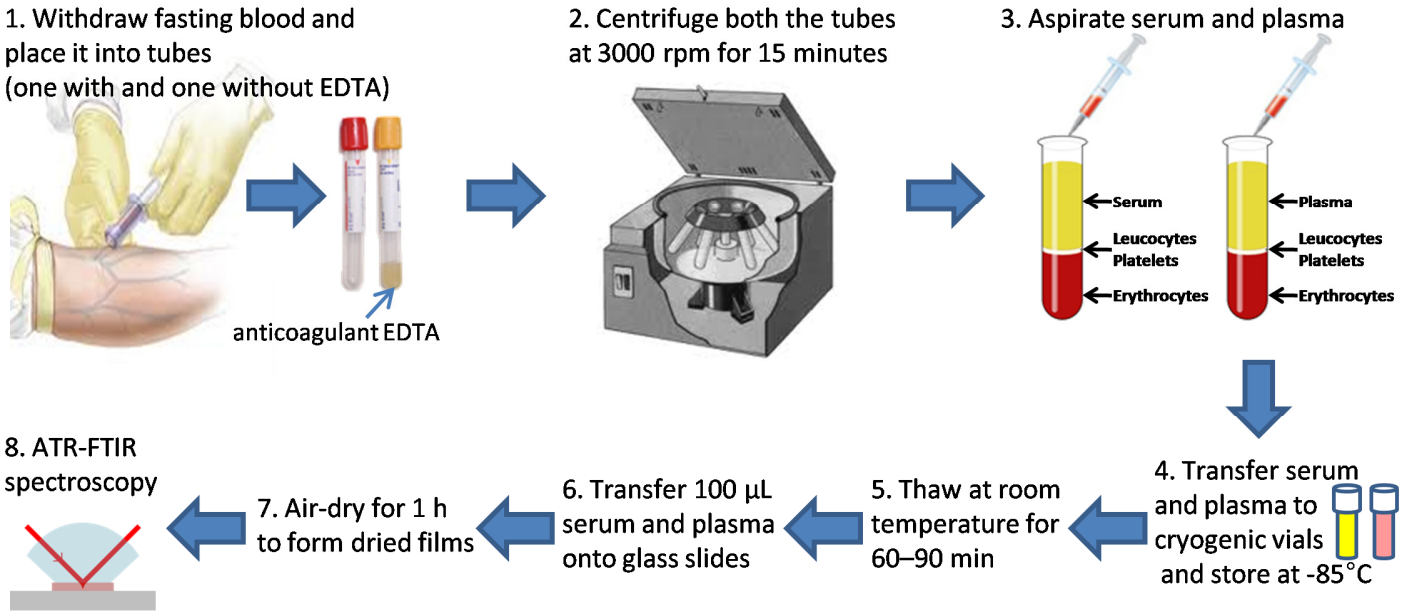

Figure S1 - Flow diagram of sample collection, storage and preparation for spectroscopic analysis.

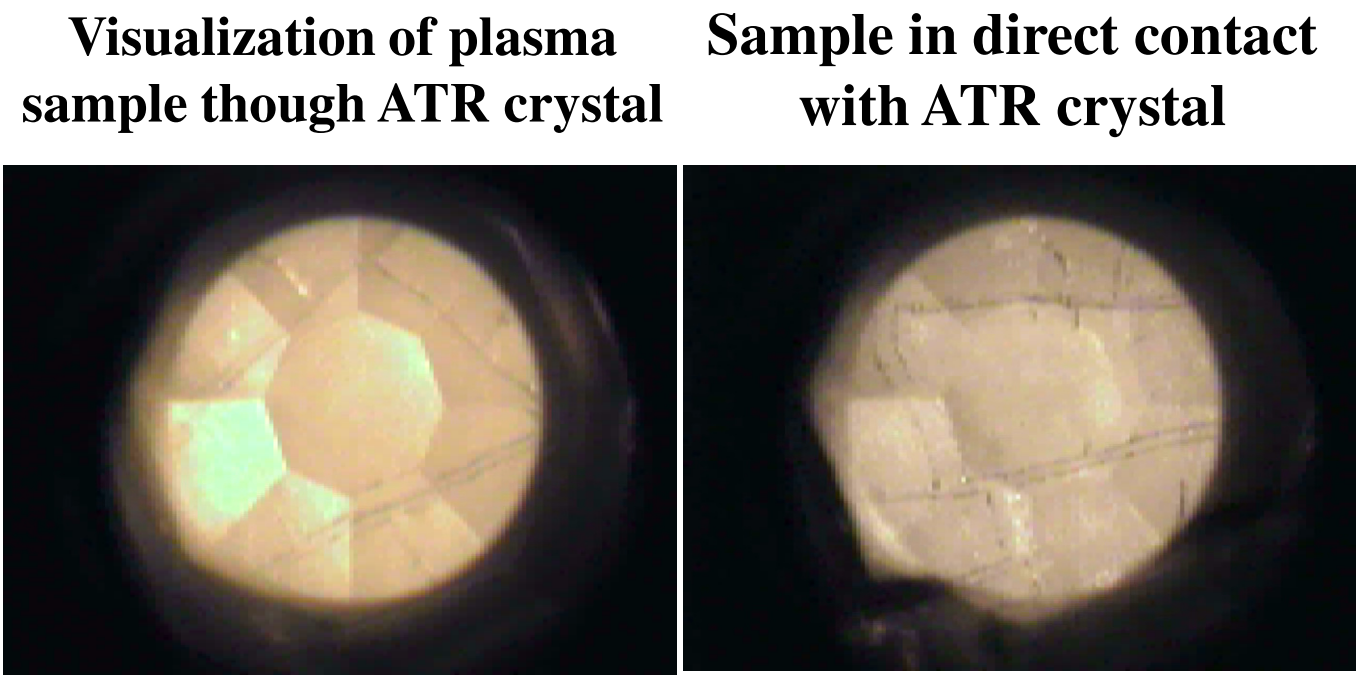

Figure S2 - Homogeneous nature of samples.
A sample of blood plasma on a Low-E glass slide viewed through or in direct contact with an ATR crystal. The homogeneity of the sample is clearly evident in the images taken through the ATR crystal. All the analyzed samples had a similar homogeneous appearance.

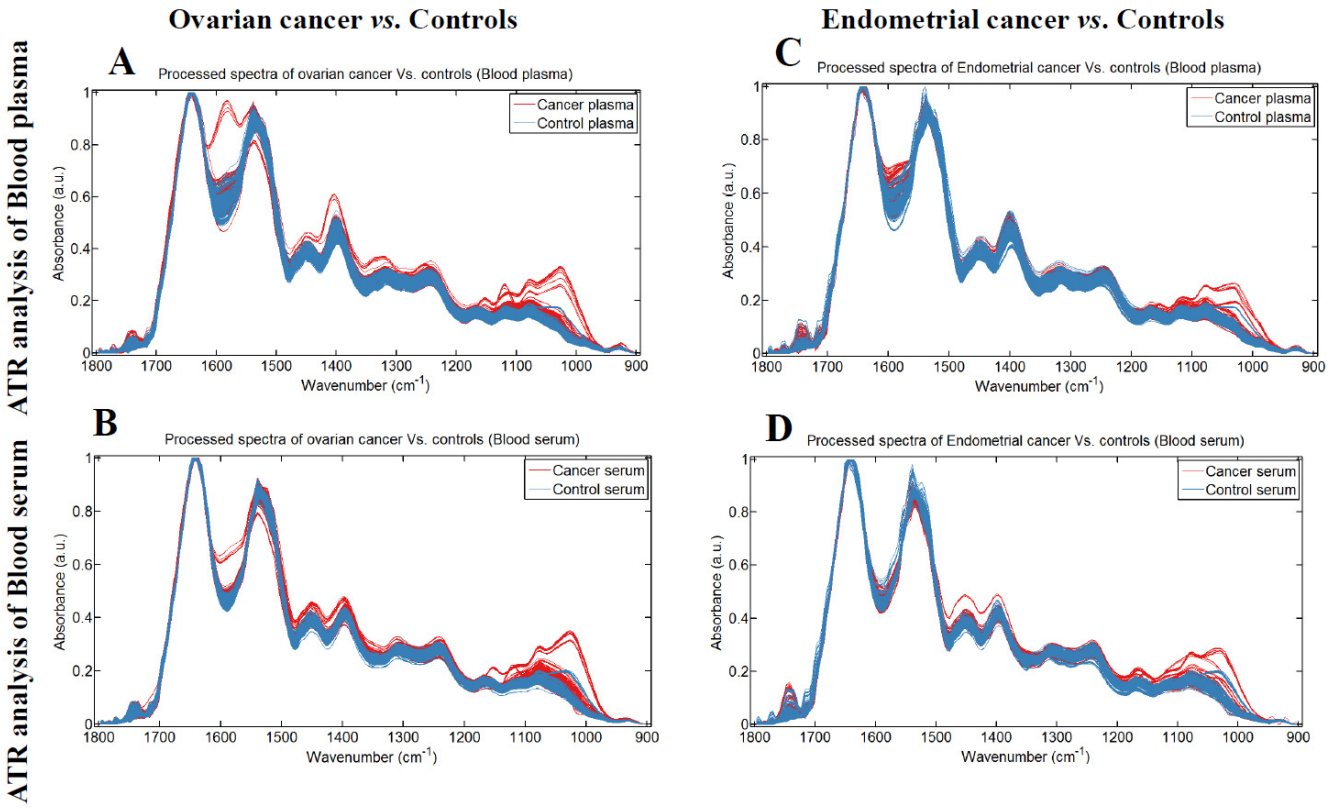

Figure S3 - Spectra for all datasets, pre-processed for visualization.
Spectra were cut to the bio-fingerprint region (1800 cm^{-1} - 900 cm^{-1}), followed by rubberband baseline correction and normalization to the Amide I peak (1650 cm^{-1}). (**A**) shows the pre-processed infrared (IR) spectra obtained from blood plasma of ovarian cancer *vs.* controls; and, (**B**) shows the IR spectra obtained from blood serum of ovarian cancer *vs.* controls. Similarly (**C**) shows the IR spectra from blood plasma after pre-processing for endometrial cancer *vs.* controls; and, (**D**) shows the IR spectra of blood serum of endometrial cancer *vs*. controls.

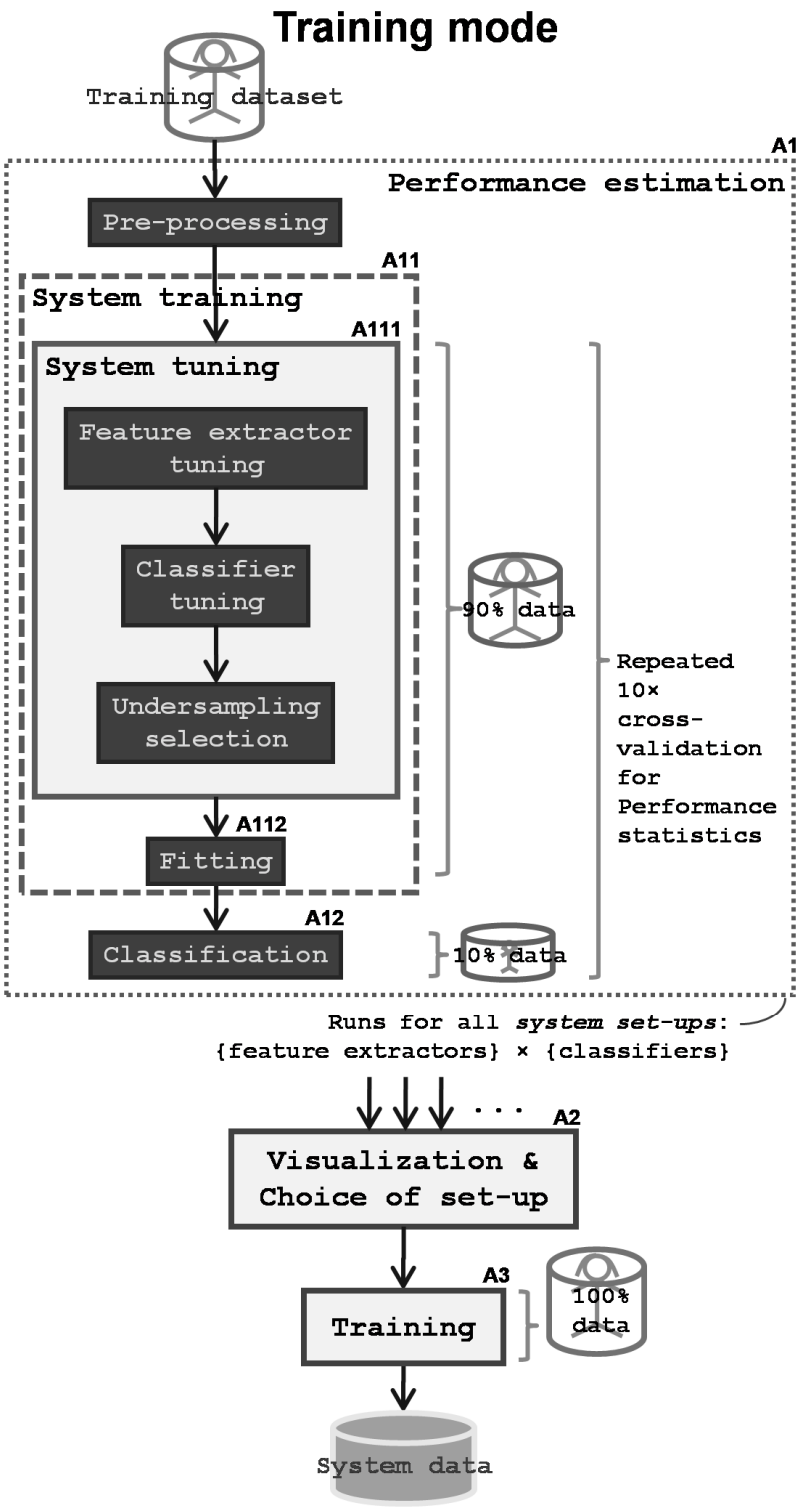

Figure S4 - Classification machine in training mode.
Patient data is first pre-processed prior to being used for training. System training consists of tuning followed by fitting. After the system is tuned, the data is used again to fit the system, which is assessed through classification performed using a spared 10% of data. This process is repeated 10 times in a cross-validated fashion to obtain 10 performances estimations that are used to calculate statistics. These statistics are used to compare system set-ups. Once chosen, the final system undergoes the training process depicted in A11, however using 100% of available data this time. The final system data represents all data-dependent model parameters, structures, and matrices etc, which were determined during the training process.

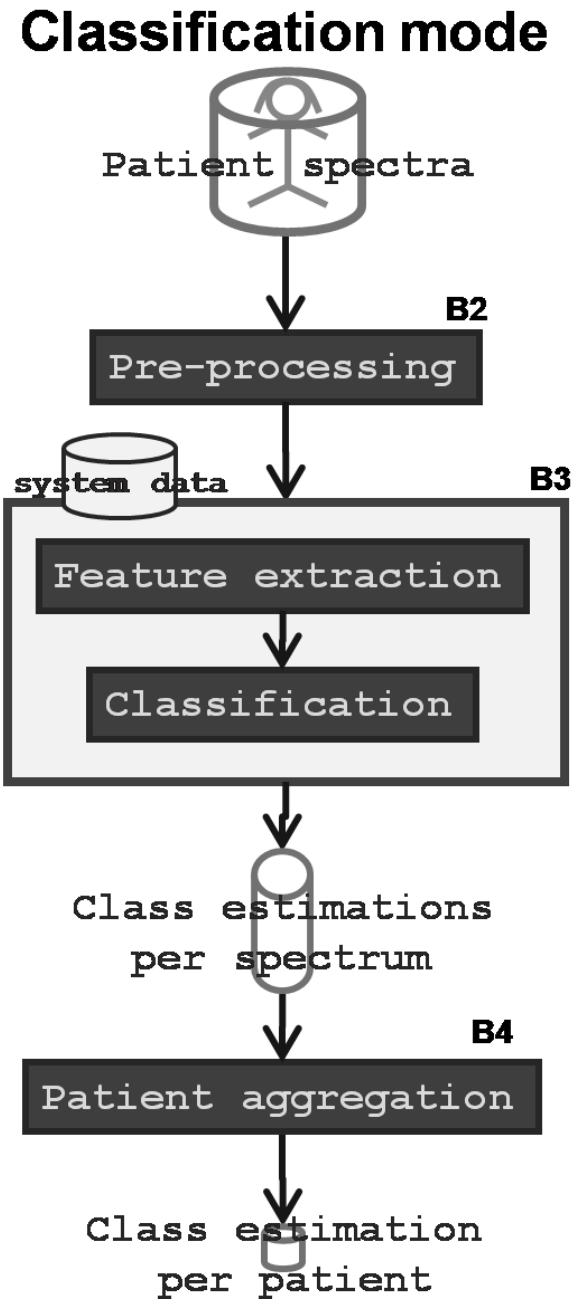

Figure S5 - Machine in classification mode.
Feature extraction and classification are performed using system knowledge gained during training mode (system data). Class estimations per spectrum are used in a pooled decision (*e.g.*, major voting) to generate a final estimation per patient.

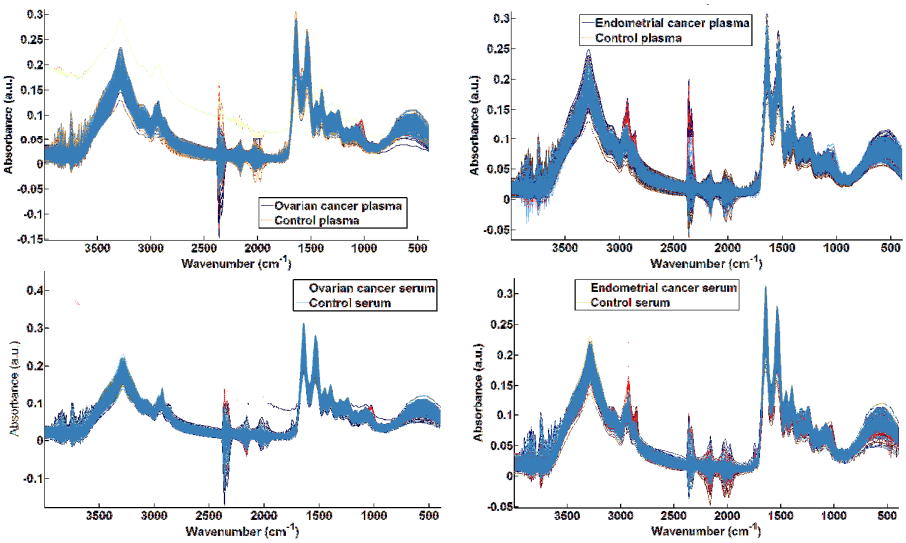

Figure S6 - Absence of artefacts in raw ATR-FTIR spectra.
Examples of raw ATR-FTIR spectra derived from ovarian cancer (plasma or serum) samples. There is an absence of any apparent artefacts, such as Mie scattering, in raw spectra acquired through the ≈250 μm × 250 μm ATR crystal.

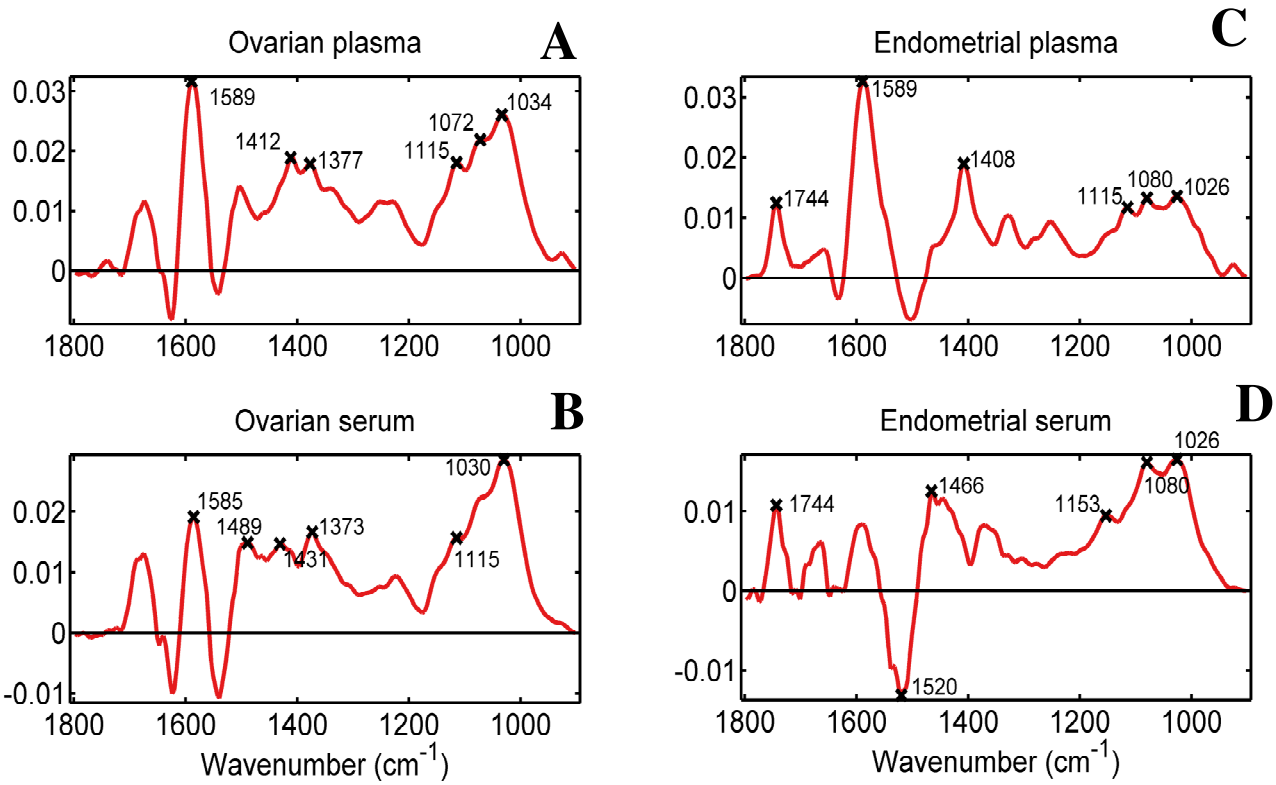

Figure S7 - Top six discriminating wavenumbers for cancer *versus* control.
The curves represent differences in absorbance (Cancer minus Control). For ovarian cancer plasma *vs.* control plasma (Fig. 4A), the discriminating wavenumbers (biomolecular assignments) are 1034 cm^{-1} (collagen), 1072 cm^{-1} (nucleic acid band), 1115 cm^{-1} (RNA), 1377 cm^{-1} (oligonucleotide base and sugar), 1412 cm^{-1} (stretching C-N, deformation N-H, deformation C-H) and 1589 cm^{-1} (ring C-C stretch of phenyl). In case of ovarian cancer serum *vs.* control serum (Fig. 4B), the discriminating wavenumbers (biomolecular assignments) are 1030 cm^{-1} (stretching C-O ribose), 1115 cm^{-1} (RNA), 1373 cm^{-1} (stretching C-N cytosine, guanine), 1431 cm^{-1} (proteins), 1489 cm^{-1} (in-plane CH bending vibration) and 1585 cm^{-1} (Amide I). The top six discriminating wavenumbers for endometrial cancer plasma *vs.* control plasma (Fig. 4C) are 1026 cm^{-1} (glycogen), 1080 cm^{-1} ($\nu_s PO_2^-$), (RNA), 1408 cm^{-1} (COO^- symmetric stretching vibrations of fatty acids and amino acid), 1589 cm^{-1} (ring C-C stretch of phenyl) and 1744 cm^{-1} (lipids). For endometrial cancer serum *vs.* control serum (Fig. 4D), the discriminating wavenumbers with their biomolecular assignments are 1026 cm^{-1} (glycogen), 1080 cm^{-1} ($\nu_s PO_2^-$), 1153 cm^{-1}, 1466 cm^{-1}, 1520 cm^{-1} and 1744 cm^{-1} (lipids).

Table S1 - Patients characteristics in the control group.

GUTB code	Age (years)	Menopausal status	Histology	Endometrial thickness
E1	46	Premenopausal	Uterine adenomyosis	5 mm
E3	44	Premenopausal	Leiomyoma	7 mm
E4	35	Premenopausal	Normal histology	12 mm
E5	47	Premenopausal	Normal histology	6 mm
E6	31	Premenopausal	Endometriosis	3 mm
E7	49	Premenopausal	Leiomyoma	8 mm
E8	42	Premenopausal	Normal histology	Not available
E9	47	Premenopausal	Endometriosis	Not available
E10	71	Postmenopausal	Normal histology (Prolapse)	Not available
E11	43	Premenopausal	Normal histology	6 mm
E12	47	Premenopausal	Uterine adenomyosis	4 mm
E13	44	Premenopausal	Normal histology	8 mm
E14	46	Premenopausal	Normal histology	2 mm
E15	77	Postmenopausal	Normal histology (Prolapse)	Atrophic
E16	43	Premenopausal	Normal histology	6 mm
E19	67	Postmenopausal	Normal histology (Prolapse)	Atrophic
E20	38	Premenopausal	Normal histology	6 mm
E21	50	Premenopausal	Endometriosis	3 mm
E23	46	Premenopausal	Normal histology	5.9 mm
E24	35	Premenopausal	Normal histology	Not available
E25	39	Premenopausal	Uterine adenomyosis	6 mm
E26	42	Premenopausal	Normal histology	14 mm
E27	42	Premenopausal	Leiomyoma	12 mm
E28	39	Premenopausal	Normal histology	Not available
E30	76	Postmenopausal	Normal histology (prolapse)	Atrophic
E31	47	Premenopausal	Normal histology	Not available
E32	38	Premenopausal	Leiomyoma	Very thin
E33	47	Premenopausal	Leiomyoma	10 mm
E35	32	Premenopausal	Normal histology	6 mm
E36	44	Premenopausal	Normal histology	Not available
Average	**46.47**			

Table S2 - Patients characteristics in the ovarian cancer group.

GUTB code	Age (years)	Menopausal status	Serum CA125	Histology	FIGO stage
OC1	77	Postmenopausal	<35	Serous/endometrioid adenocarcinoma	3B
OC2	60	Postmenopausal	>35	Metastatic adenocarcinoma	4
OC4	69	Postmenopausal	>35	Serous/mucinous carcinoma	1A
OC5	60	Postmenopausal	>35	Mucinous carcinoma	1A
OC6	40	Premenopausal	>35	Serous adenocarcinoma	3A
OC7	86	Postmenopausal	>35	Clear cell carcinoma	2A
OC8	74	Postmenopausal	>35	Serous adenocarcinoma	1C
OC9	74	Postmenopausal	>35	Serous adenocarcinoma	3C
OC10	65	Postmenopausal	>35	Serous adenocarcinoma	1C
OC11	64	Postmenopausal	>35	Serous adenocarcinoma	1C
OC12	65	Postmenopausal	>35	Mucinous carcinoma	1A
OC13	62	Postmenopausal	>35	Serous adenocarcinoma	1A
OC14	81	Postmenopausal	>35	Mucinous carcinoma	1A
OC15	60	Postmenopausal	>35	Clear cell carcinoma	1C
OC16	65	Postmenopausal	>35	Serous adenocarcinoma	1C
OC17	56	Postmenopausal	>35	Serous adenocarcinoma	3C
OC19	62	Postmenopausal	>35	Endometrioid adenocarcinoma	1C
OC21	64	Postmenopausal	<35	Serous adenocarcinoma	3B
OC22	78	Postmenopausal	>35	Metastatic carcinosarcoma	4
OC24	75	Postmenopausal	>35	Serous adenocarcinoma	1A
OC25	60	Postmenopausal	>35	Clear cell carcinoma	3C
OC26	53	Postmenopausal	>35	Serous adenocarcinoma	3C
OC27	36	Premenopausal	>35	Serous adenocarcinoma	1A
OC29	56	Postmenopausal	>35	Serous adenocarcinoma	3A
OC30	70	Postmenopausal	<35	Mucinous carcinoma	2A
OC32	65	Postmenopausal	>35	Mucinous carcinoma	1A
OC33	59	Postmenopausal	<35	Metastatic carcinosarcoma	3C
OC34	85	Postmenopausal	>35	Serous adenocarcinoma	3B
OC35	63	Postmenopausal	>35	Mucinous carcinoma	1C
OC36	73	Postmenopausal	>35	Serous adenocarcinoma	1C
Average	**65.23**				

Table S3 - Patients characteristics in the endometrial cancer group.

GUTB code	Age (years)	Menopausal status	Histology: grade (G1,G2,G3) and subtype	Endometrial thickness
EC1	77	Postmenopausal	G1, Endometrioid type	> 5 mm
EC2	87	Postmenopausal	G3, Endometrioid type	> 5 mm
EC3	64	Postmenopausal	G1, Endometrioid type	> 5 mm
EC4	78	Postmenopausal	G3, Carcinosarcoma	> 5 mm
EC6	66	Postmenopausal	G3, Endometrioid/Serous type	> 5 mm
EC7	73	Postmenopausal	G3, Carcinosarcoma	> 5 mm
EC12	72	Postmenopausal	G3, Clear Cell carcinoma	> 5 mm
EC14	78	Postmenopausal	G1, Endometrioid type	> 5 mm
EC18	71	Postmenopausal	G3, Clear Cell carcinoma	> 5 mm
EC20	82	Postmenopausal	G3, Uterine Serous cancer	> 5 mm
EC21	72	Postmenopausal	G3, Clear Cell carcinoma	> 5 mm
EC22	68	Postmenopausal	G2, Endometrioid type	> 5 mm
EC23	82	Postmenopausal	G3, Carcinosarcoma	> 5 mm
EC26	68	Postmenopausal	G3, Carcinosarcoma	> 5 mm
EC27	57	Postmenopausal	G3, Endometrioid type	> 5 mm
EC28	64	Postmenopausal	G3, Mixed endometrioid/clear cell type	> 5 mm
EC29	66	Postmenopausal	G1, Endometrioid type	> 5 mm
EC30	57	Postmenopausal	G1, Endometrioid type	> 5 mm
EC32	60	Postmenopausal	G1, Endometrioid type	> 5 mm
EC33	63	Postmenopausal	G3, Carcinosarcoma	> 5 mm
EC34	62	Postmenopausal	G3, Clear Cell carcinoma	> 5 mm
EC35	56	Postmenopausal	G1, Endometrioid type	> 5 mm
EC37	76	Postmenopausal	G3, Uterine serous cancer	> 5 mm
EC38	57	Postmenopausal	G1, Endometrioid type	> 5 mm
EC41	73	Postmenopausal	G1, Endometrioid type	> 5 mm
EC43	61	Postmenopausal	G3, Endometrioid type	> 5 mm
EC44	48	Premenopausal	G1, Endometrioid type	> 5 mm
EC45	59	Postmenopausal	G1, Endometrioid type	> 5 mm
EC47	67	Postmenopausal	G1, Endometrioid type	> 5 mm
EC48	82	Postmenopausal	G1, Endometrioid type	> 5 mm
Average	**68.2**			

Table S4 - Spectral acquisition settings.
These parameters are found in the "Advanced measurement" box in OPUS 6.5 software (Bruker Optik GmbH).

Parameter name	Value
Basic	
Imaging device	Helios
Objective	HELIOS ATR
Advanced	
Resolution	8 cm^{-1}
Sample scan time	32 Scans
Background scan time	32 Scans
Save data from	4000 cm^{-1} to 400 cm^{-1}
Result spectrum	ATR Spectrum
Data blocks to be saved	- ATR Spectrum - Single Channel - Background
Optic	
External synchronization	Off
Source setting	MIR
Beamsplitter	KBr
Optical Filter setting	Open
Aperture setting	6 mm
Measurement channel	Sample Compartment
Background meas. Channel	Sample Compartment
Detector setting	RT-DLaTGS [Internal]
Scanner velocity	2.2 KHz
Sample signal gain	Automatic
Background signal gain	Automatic
Delay after device change	0
Delay before measurement	0
Optical bench ready	OFF
Acquisition	
Wanted high frequency limit	8000
Wanter low frequency limit	0
High Pass filter	Open
Low Pass filter	10 KHz
Acquisition mode	Double Sided, Forward-Backward
Correlation mode	OFF
Fourier transformation (FT)	
Phase resolution	32
Phase Correction mode	Mertz / No Peak Search
Apodization function	Blackman-Harris 3-Term
Zerofilling factor	2

Table S5 - Classification results: ovarian plasma data - all system set-ups.
Each system set-up is a combination of a feature extraction method (left column) with a classification method (top row). The numbers shown represent classification rates (%) ± standard deviations (%).
(This table is identical to Figure 4A in the main manuscript and repeated here for convenience).

	eClass1	*k*-NN	LDC	QDC	SVM
PCA	85±12.3	80±18.92	95±8.05	85±16.57	86.67±10.54
PLS	86.67±10.54	71.67±8.05	91.67±8.78	86.67±10.54	81.67±16.57
Spline	91.67±11.79	76.67±16.1	86.67±15.32	90±8.61	83.33±11.11
FS-LASSO	96.67±7.03	75±21.15	93.33±8.61	90±11.65	93.33±8.61
FS-Fisher	76.67±14.05	81.67±12.3	75±14.16	75±11.79	76.67±14.05
FFS (*classifier*)	91.67±11.79	88.33±13.72	93.33±8.61	93.33±8.61	85±14.59
FFS (MANOVA)	88.33±8.05	86.67±17.21	91.67±8.78	90±14.05	85±12.3
Identity	93.33±8.61	76.67±17.92	91.67±8.78	19.69±20.5	81.67±9.46

Table S6 - Classification results: ovarian plasma data - 10 best system set-ups from Table S5.
These setups are listed here to show, apart from the classification rates, the computational time, and the *P*-values. The latter represent the probability of the observed classification rates of adjacent rankings, given that they are equal. The null hypothesis may be formulated as $H_0 : r_i - r_{i+1} = 0$, where r_i and r_{i+1} are respectively the classification rates for ranking i and i+1. The *P*-values were calculated using a *T*-test. For example, the *P*-value of 0.296 results from the hypothesis test involving the setups Identity→eClass1 and FFS (MANOVA)→LDC.

	Ranking	System set-up	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**FS-LASSO→eClass1**	**96.67 ± 7.03**	**1.27 ± 0.11**	**0.172**
	2	**PCA→LDC**	95.00 ± 8.05	0.15 ± 0.01	0.172
	3	**FFS→QDC**	93.33 ± 8.61	0.07 ± 0.01	0.500
	4	**FS-LASSO→LDC**	93.33 ± 8.61	0.08 ± 0.01	0.500
	5	**FFS→LDC**	93.33 ± 8.61	0.08 ± 0.01	0.500
	6	**FS-LASSO→SVM**	93.33 ± 8.61	0.28 ± 0.04	0.500
	7	**Identity→eClass1**	93.33 ± 8.61	3.17 ± 0.37	0.296
	8	**FFS (MANOVA)→LDC**	91.67 ± 8.78	0.08 ± 0.00	0.500
	9	**Identity→LDC**	91.67 ± 8.78	0.12 ± 0.01	0.500
	10	**PLS→LDC**	91.67 ± 8.78	0.23 ± 0.07	0.500

Table S7 - Classification results: ovarian plasma data - ensemble systems.
The *P*-values calculated as in **Table S6**.

	Ranking	Ensemble of …	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**1 system**	**96.67 ± 7.03**	**2.18 ± 0.63**	**0.172**
	2	**2 systems**	95.00 ± 8.05	2.64 ± 1.04	0.500
	3	**3 systems**	95.00 ± 8.05	11.50 ± 1.75	0.500
	4	**6 systems**	95.00 ± 8.05	12.57 ± 1.64	0.500
	5	**7 systems**	95.00 ± 8.05	12.41 ± 0.95	0.500
	6	**8 systems**	95.00 ± 8.05	15.65 ± 1.79	0.500
	7	**9 systems**	95.00 ± 8.05	15.55 ± 1.80	0.500
	8	**10 systems**	95.00 ± 8.05	17.16 ± 1.69	0.500
	9	**11 systems**	95.00 ± 8.05	17.41 ± 1.28	0.500
	10	**12 systems**	95.00 ± 8.05	20.42 ± 2.70	0.500

Table S8 - Classification results: ovarian serum data - all system set-ups.
Each system set-up is a combination of a feature extraction method (left column) with a classification method (top row). The numbers shown represent classification rates (%) ± standard deviations (%).
(This table is identical to Figure 4B in the main manuscript and repeated here for convenience).

	eClass1	*k*-NN	LDC	QDC	SVM
PCA	78.33±13.72	78.33±11.25	81.67±14.59	75±16.2	76.67±14.05
PLS	78.33±13.72	81.67±12.3	83.33±15.71	73.33±17.92	75±16.2
Spline	81.67±16.57	88.33±13.72	76.67±21.08	73.33±16.1	86.67±13.15
FS-LASSO	88.33±13.72	78.33±13.72	90±11.65	83.33±13.61	85±9.46
FS-Fisher	85±16.57	88.33±11.25	85±16.57	86.67±15.32	80±20.49
FFS (*classifier*)	86.67±10.54	95±8.05	83.33±11.11	85±16.57	91.67±11.79
FFS (MANOVA)	86.67±13.15	81.67±21.44	88.33±13.72	86.67±13.15	83.33±13.61
Identity	78.33±15.81	83.33±15.71	80±17.21	7.77±18.29	86.67±13.15

Table S9 - Classification results: ovarian serum data - 10 best system set-ups from Table S8.
The *P*-values calculated as in **Table S6**.

	Ranking	System set-up	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**FFS→*k*-NN**	**95.00 ± 8.05**	**0.36 ± 0.12**	**0.172**
	2	**FFS→SVM**	91.67 ± 11.79	0.71 ± 0.54	0.339
	3	**FS-LASSO→LDC**	90.00 ± 11.65	0.32 ± 0.23	0.339
	4	**FS-Fisher→*k*-NN**	88.33 ± 11.25	0.35 ± 0.11	0.500
	5	**FFS (MANOVA)→LDC**	88.33 ± 13.72	0.26 ± 0.19	0.500
	6	**Spline→*k*-NN**	88.33 ± 13.72	0.39 ± 0.11	0.500
	7	**FS-LASSO→eClass1**	88.33 ± 13.72	2.51 ± 0.91	0.379
	8	**FFS→eClass1**	86.67 ± 10.54	2.46 ± 1.22	0.500
	9	**FFS (MANOVA)→QDC**	86.67 ± 13.15	0.12 ± 0.08	0.500
	10	**Spline→SVM**	86.67 ± 13.15	1.11 ± 0.47	0.500

Table S10 - Classification results: ovarian serum data - ensemble systems.
The *P*-values calculated as in **Table S6**.

	Ranking	Ensemble of …	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**1 system**	**95.00 ± 8.05**	**0.25 ± 0.14**	**0.500**
	2	**2 systems**	95.00 ± 8.05	0.88 ± 0.57	0.500
	3	**5 systems**	95.00 ± 8.05	3.95 ± 1.15	0.500
	4	**6 systems**	95.00 ± 8.05	4.58 ± 1.52	0.500
	5	**7 systems**	95.00 ± 8.05	4.32 ± 1.60	0.500
	6	**8 systems**	95.00 ± 8.05	7.94 ± 0.96	0.500
	7	**17 systems**	95.00 ± 8.05	14.46 ± 2.64	0.500
	8	**3 systems**	95.00 ± 11.25	1.10 ± 0.61	0.500
	9	**4 systems**	95.00 ± 11.25	1.62 ± 1.04	0.339
	10	**18 systems**	93.33 ± 8.61	15.51 ± 2.44	0.500

Table S11 - Classification results: endometrial plasma data - all system set-ups.
Each system set-up is a combination of a feature extraction method (left column) with a classification method (top row). The numbers shown represent classification rates (%) ± standard deviations (%).
(This table is identical to Figure 4C in the main manuscript and repeated here for convenience).

	eClass1	*k*-NN	LDC	QDC	SVM
PCA	63.18±13.71	60.57±12.32	63.01±13.96	69.74±16.99	68.69±10.52
PLS	62.13±11.62	55.71±17.91	63.84±13.38	65.34±13.63	67.77±16.63
Spline	63.83±12.17	57.7±12.59	61.94±11.92	68.59±18.3	71.78±13.32
FS-LASSO	68.83±15.73	62.01±16.53	63.98±13.9	71.43±15.7	68.33±14.92
FS-Fisher	73.26±19.27	71.69±15.47	72.9±18.76	72.57±20.53	69.34±20.43
FFS	70.98±16.31	58.14±21.65	66.59±13.49	69.48±17.33	67.41±16.62
FFS (MANOVA)	66.61±13.54	65.93±16.54	67.03±12.93	66.56±15.98	66.19±17.37
Identity	62.3±13.76	63.84±15.92	62.95±16.34	24.26±27.93	74.85±16.01

Table S12 - Classification results: endometrial plasma data - 10 best set-ups from Table S11.
The *P*-values calculated as in **Table S6**.

	Ranking	System set-up	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**Identity→SVM**	**80.00 ± 18.92**	**1.50 ± 0.37**	**0.453**
	2	**FS-Fisher→*k*-NN**	79.17 ± 18.84	0.39 ± 0.08	0.249
	3	**Spline→QDC**	73.75 ± 18.22	0.49 ± 0.26	0.500
	4	**FS-Fisher→QDC**	73.75 ± 19.45	0.25 ± 0.07	0.500
	5	**FS-LASSO→QDC**	73.75 ± 19.45	0.31 ± 0.18	0.475
	6	**Spline→SVM**	73.33 ± 16.10	1.14 ± 0.72	0.500
	7	**FS-Fisher→LDC**	73.33 ± 22.50	0.16 ± 0.06	0.500
	8	**FS-Fisher→eClass1**	73.33 ± 22.50	2.54 ± 0.73	0.399
	9	**FFS→eClass1**	71.67 ± 17.66	3.58 ± 1.56	0.457
	10	**FFS→QDC**	70.83 ± 18.43	0.19 ± 0.07	0.500

Table S13 - Classification results: endometrial plasma data - ensemble systems.
The *P*-values calculated as in **Table S6**.

	#	Ensemble of …	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**3 systems**	**81.67 ± 18.34**	**2.25 ± 0.81**	**0.379**
	2	**1 system**	80.00 ± 18.92	1.71 ± 0.57	0.339
	3	**4 systems**	78.33 ± 19.33	2.54 ± 0.87	0.500
	4	**5 systems**	78.33 ± 19.33	2.55 ± 0.87	0.296
	5	**2 systems**	76.67 ± 17.92	1.99 ± 0.78	0.500
	6	**7 systems**	76.67 ± 21.08	5.74 ± 1.81	0.500
	7	**8 systems**	76.67 ± 21.08	5.86 ± 1.70	0.296
	8	**11 systems**	75.00 ± 19.64	8.93 ± 2.57	0.500
	9	**6 systems**	75.00 ± 21.15	3.95 ± 1.08	0.500
	10	**9 systems**	75.00 ± 19.64	8.39 ± 1.57	0.500

Table S14 - Classification results: endometrial serum data - all system set-ups.
Each system set-up is a combination of a feature extraction method (left column) with a classification method (top row). The numbers shown represent classification rates (%) ± standard deviations (%).
(This table is identical to Figure 4D in the main manuscript and repeated here for convenience).

	eClass1	*k*-NN	LDC	QDC	SVM
PCA	60.86±12.39	57.79±9.32	59.61±13.95	56.88±13.67	58.02±24.52
PLS	62.36±16.52	58.64±14.62	60.19±12.07	61.46±20.15	67.75±20.51
Spline	64.4±14.91	51.9±16.96	66.51±13.41	63.49±17.62	70.47±16.96
FS-LASSO	64.09±19.2	56.39±19.51	61.26±12.61	58.88±22.06	74.05±17.84
FS-Fisher	67.32±14.34	60.66±17.17	67.2±14.43	66.68±11.75	72.9±16.07
FFS	68.1±15.24	64.63±18.41	64.3±21.44	63.74±13.14	53.78±10.6
FFS (MANOVA)	62.36±16.48	63.87±14.95	63.67±16.92	63.74±13.29	71.09±17.34
Identity	59.24±13.32	58.26±14.56	61.2±13.6	21.33±31.86	71.09±22.3

Table S15 - Classification results: endometrial serum data - 10 best set-ups from Table S14.
The *P*-values calculated as in **Table S6**.

	Ranking	System set-up	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**FS-Fisher→SVM**	**76.67 ± 14.05**	**1.14 ± 0.63**	**0.373**
	2	**FS-LASSO→SVM**	75.42 ± 21.01	2.03 ± 1.89	0.461
	3	**FFS (MANOVA)→SVM**	75.00 ± 21.15	1.52 ± 0.94	0.276
	4	**Identity→SVM**	72.08 ± 22.23	4.85 ± 2.58	0.390
	5	**FS-Fisher→QDC**	70.42 ± 10.48	0.14 ± 0.05	0.500
	6	**Spline→SVM**	70.42 ± 17.17	1.68 ± 0.71	0.500
	7	**FFS→eClass1**	70.42 ± 21.56	2.41 ± 0.67	0.379
	8	**Spline→eClass1**	68.33 ± 14.59	3.11 ± 1.16	0.500
	9	**Spline→LDC**	68.33 ± 16.57	0.13 ± 0.06	0.457
	10	**FFS→*k*-NN**	67.50 ± 24.98	0.33 ± 0.13	0.477

Table S16 - Classification results: endometrial serum data - ensemble systems.
The *P*-values were calculated as in **Table S6**.

	Ranking	Ensemble of …	Classification rate (%)	Training+test time (seconds)	*P*-values
»	**1**	**2 systems**	**77.08 ± 15.87**	**1.85 ± 0.97**	**0.427**
	2	**1 system**	76.67 ± 14.05	0.65 ± 0.40	0.500
	3	**3 systems**	76.67 ± 17.92	2.83 ± 1.15	0.500
	4	**4 systems**	76.67 ± 17.92	5.34 ± 1.98	0.172
	5	**5 systems**	75.00 ± 19.64	5.29 ± 1.88	0.154
	6	**6 systems**	70.42 ± 15.27	6.44 ± 1.97	0.500
	7	**7 systems**	70.42 ± 17.17	7.28 ± 1.99	0.500
	8	**10 systems**	70.42 ± 17.17	8.65 ± 1.92	0.500
	9	**11 systems**	70.42 ± 17.17	9.95 ± 1.94	0.500
	10	**12 systems**	70.42 ± 15.27	10.11 ± 2.03	0.500