Supplementary Information

Mid-infrared spectral classification of endometrial cancer compared to benign controls in serum or plasma samples.

David Mabwa, Ketankumar Gajjar, David Furniss, Roberta Schiemer, Richard Crane, Christopher Fallaize, Pierre L Martin-Hirsch, Francis L Martin, Theordore Kypraios, Angela B Seddon, Sendy Phang

No. of Pages = 11

No. of Figures = 8

No. of Tables = 2

Endometrial Plasma

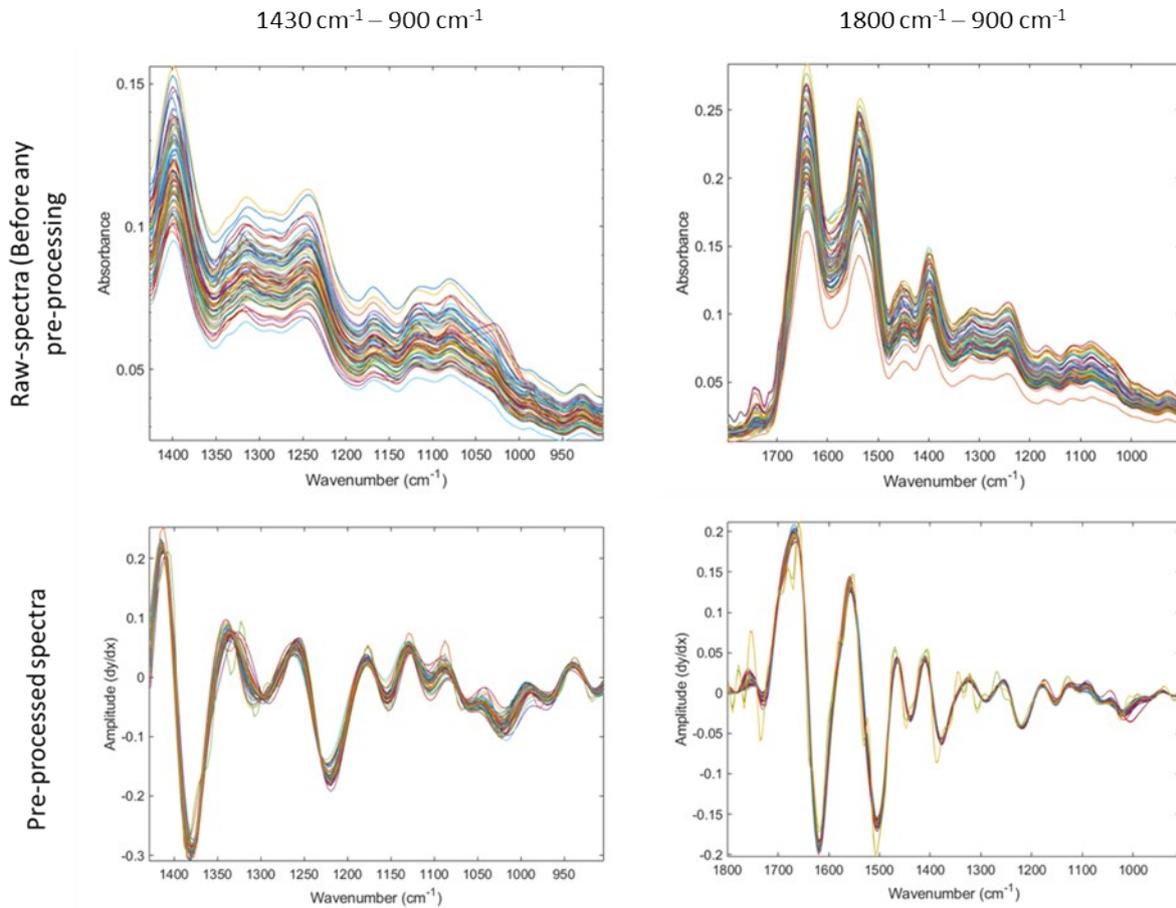1430 cm^{-1} – 900 cm^{-1} 1800 cm^{-1} – 900 cm^{-1}

Figure S1. Spectral data collected from endometrial plasma cancer and control patients. Top row shows the spectra before any pre-processing was performed and the bottom row shows the spectra after the following pre-processing steps were performed: 1. Cut spectra (1430 cm^{-1} – 900 cm^{-1} – LEFT COLUMN, or 1800 cm^{-1} – 900 cm^{-1} – RIGHT COLUMN), SG-filtering (polynomial order of 5) and spectral derivation (1st Order) and vector normalisation. These spectra are only a representation of the full sample, therefore only 80 random individual spectra are shown.
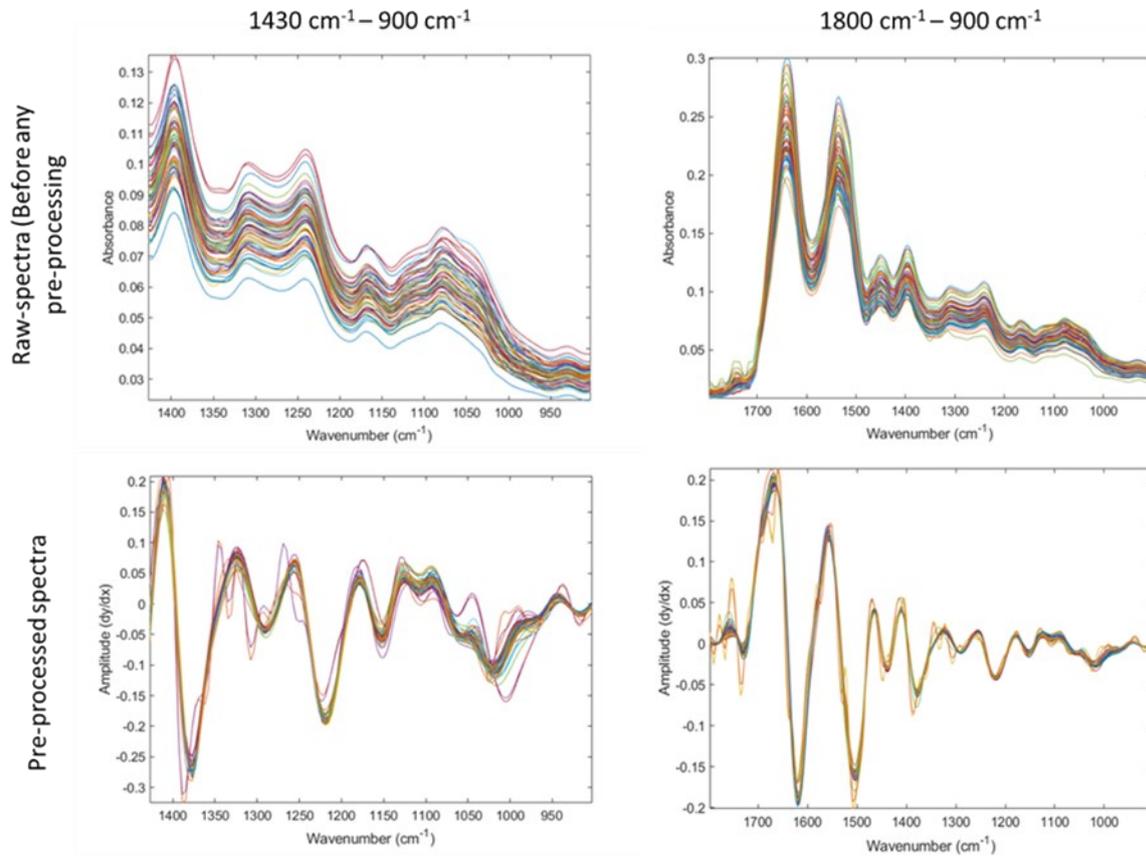
Endometrial Serum

Figure S2. Spectral data collected from endometrial serum cancer and control patients. Top row shows the spectra before any pre-processing was performed and the bottom row shows the spectra after the following pre-processing steps were performed: 1. Cut spectra (1430 cm⁻¹ – 900 cm⁻¹ – LEFT COLUMN, or 1800 cm⁻¹ – 900 cm⁻¹ – RIGHT COLUMN), SG-filtering (polynomial order of 5) and spectral derivation (1ˢᵗ Order) and vector normalisation. These spectra are only a representation of the full sample, therefore only 80 random individual spectra are shown.
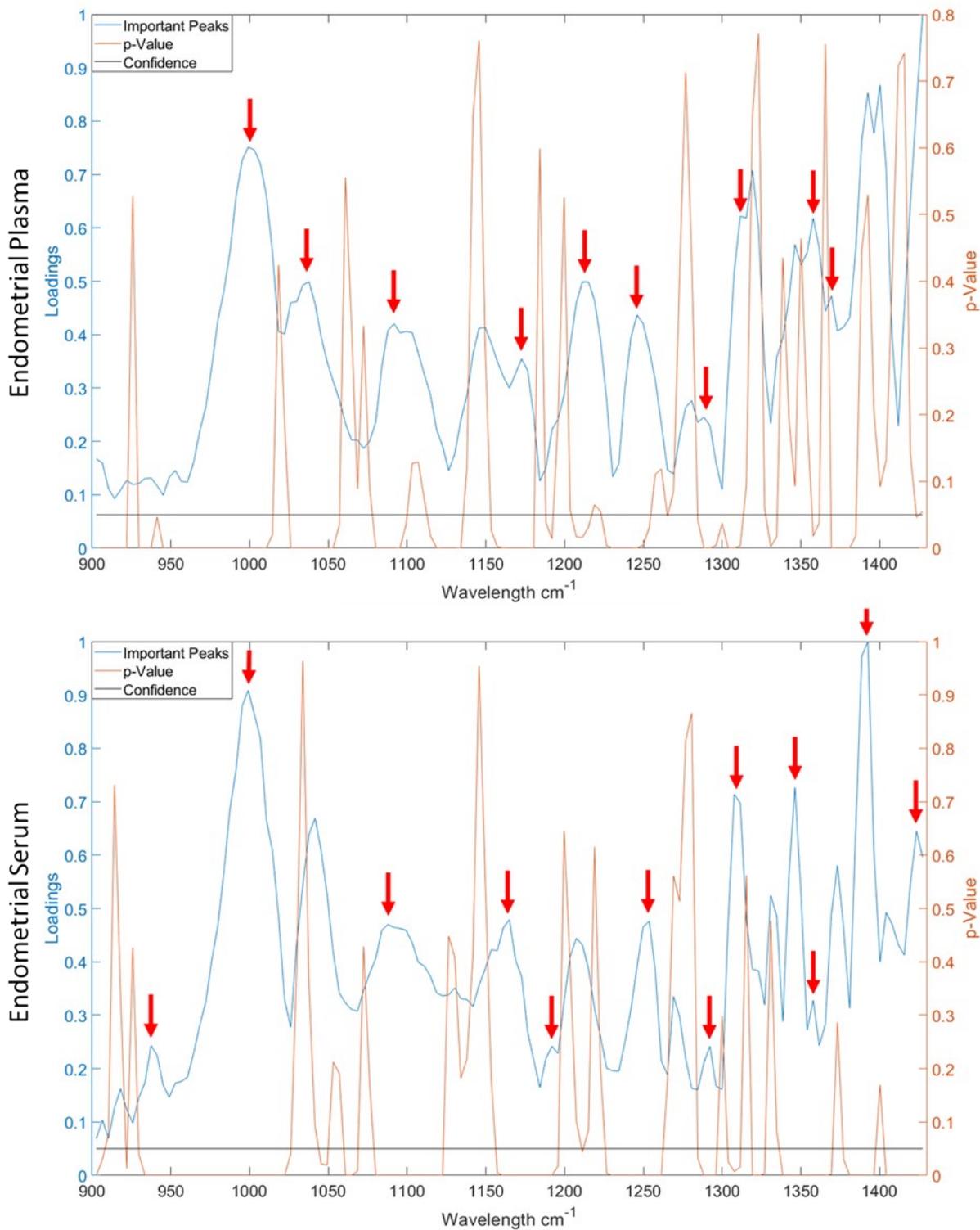
Figure S3. *Important features found in endometrial plasma and serum in the 1430 cm^{-1} to 900 cm^{-1} region. Blue line plot shows the PCA loadings/weightings obtained by performing PCA on derivatised spectra. These loadings, which are validated by a two-sample t-test, indicate the importance of each wavenumber. Orange line plot (P-value) shows the result of the independent samples t-test, indicating the significance of each wavenumber. If the P-value exceeds 0.05 he corresponding wavenumber is deemed insignificant. Significant features are indicated by red arrows.*

Figure S4. *Important features found in endometrial plasma and serum in the 1800 cm^{-1} to 900 cm^{-1} region. Blue line plot shows the PCA loadings/weightings obtained by performing PCA on derivatised spectra. These loadings, which are validated by a two-sample t-test, indicate the importance of each wavenumber. Orange line plot (P-value) shows the result of the independent samples t-test, indicating the significance of each wavenumber. If the P-value exceeds 0.05 the corresponding wavenumber is deemed insignificant. Significant features are indicated by red arrows.*
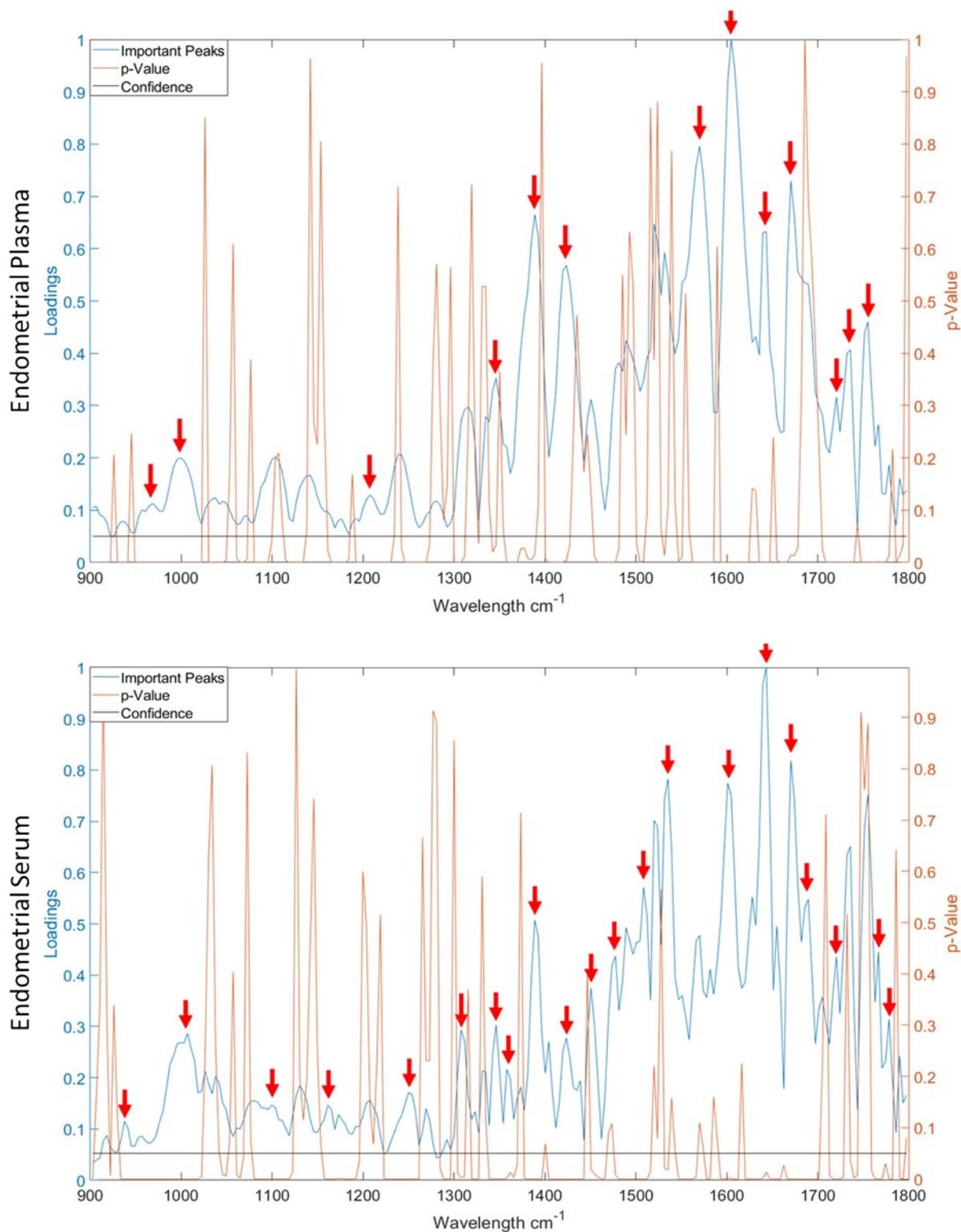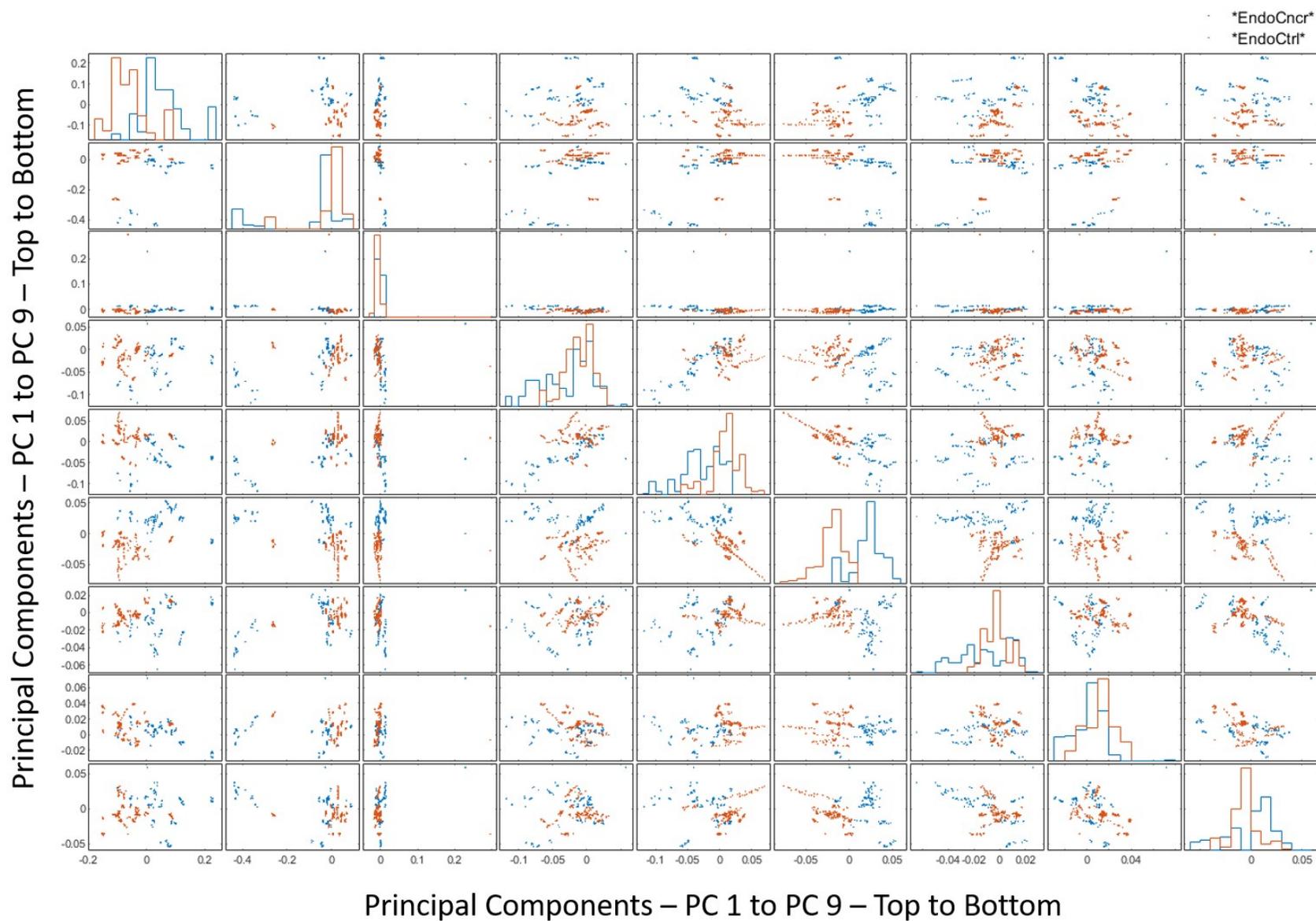
Figure S5. *The PCA outcome from endometrial plasma (1430 cm^{-1} to 900 cm^{-1}). Nine PCs accounting for 95% variance in the dataset were found. On the x-axis, PC 1 to PC 9 are shown from left to right, while on the y-axis, they are shown from top to bottom. Blue dots refer to cancer (*EndoCncr*). Orange dots refer to control (*EndoCtrl*).*

Figure S6. *The PCA outcome from endometrial plasma (1800 cm^{-1} to 900 cm^{-1}). Eight PCs accounting for 95% variance in the dataset were found. On the x-axis, PC 1 to PC 8 are shown from left to right, while on the y-axis, they are shown from top to bottom. Blue dots refer to cancer (*EndoCncr*). Orange dots refer to control (*EndoCtrl*).*
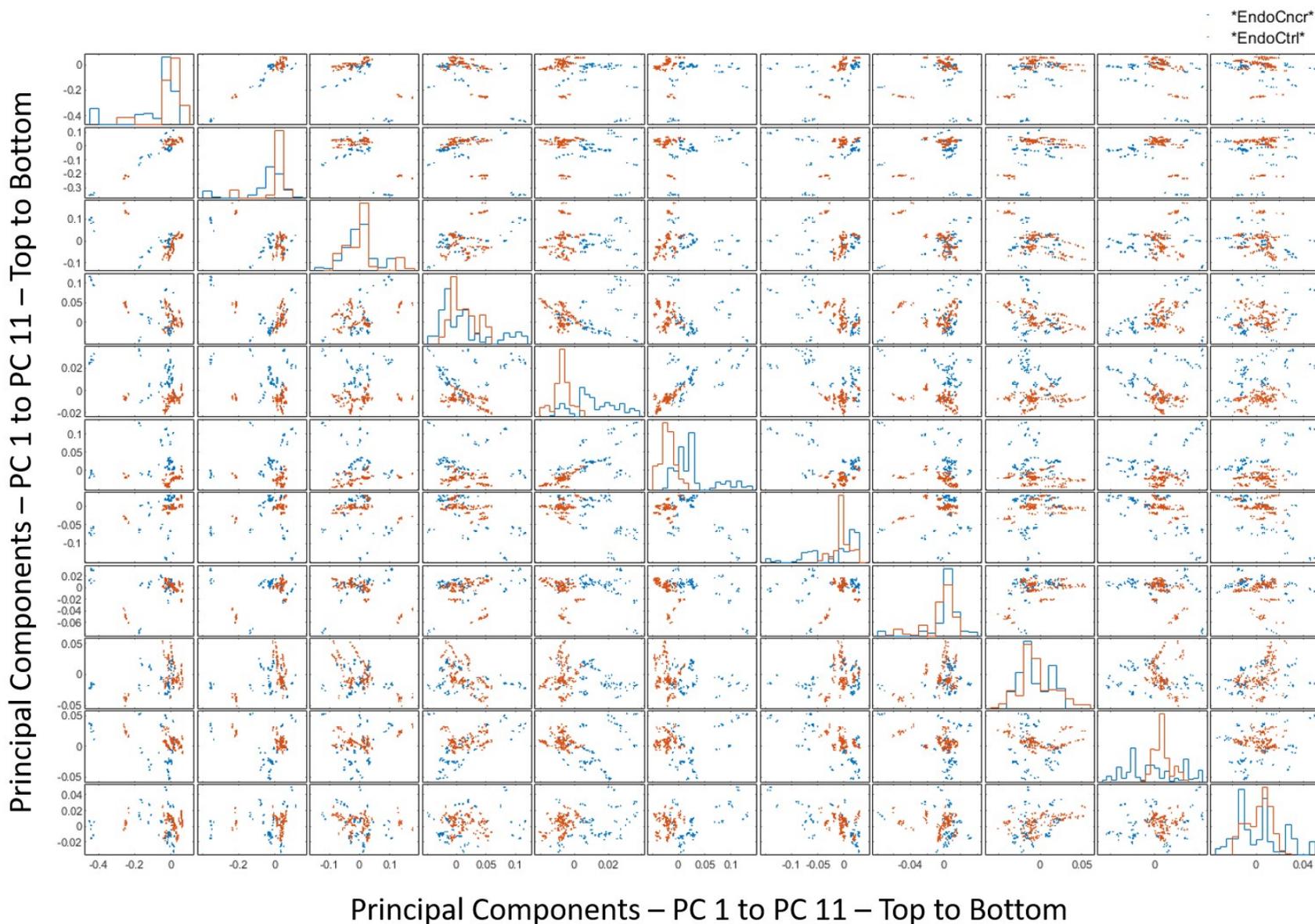
Figure S7. *The PCA outcome from endometrial serum (1430 cm⁻¹ to 900 cm⁻¹). Eleven PCs accounting for 95% variance in the dataset were found. On the x-axis, PC 1 to PC 11 are shown from left to right, while on the y-axis, they are shown from top to bottom. Blue dots refer to cancer (*EndoCncr*). Orange dots refer to control (*EndoCtrl*).*
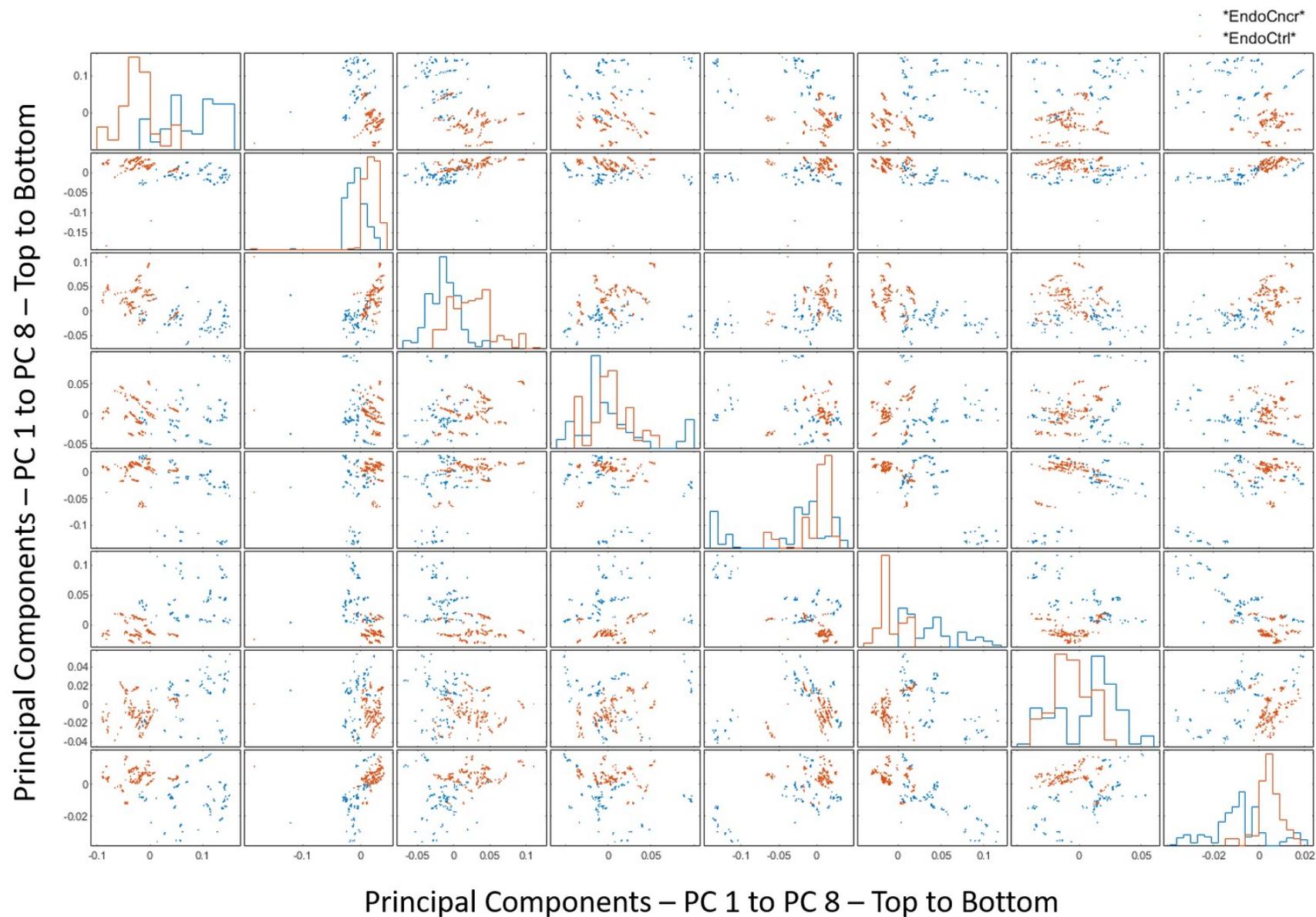
Figure S8. *The PCA outcome from endometrial serum (1800 cm⁻¹ to 900 cm⁻¹). Ten PCs accounting for 95% variance in the dataset were found. On the x-axis, PC 1 to PC 11 are shown from left to right, while on the y-axis, they are shown from top to bottom. Blue dots refer to cancer (*EndoCncr*). Orange dots refer to control (*EndoCtrl*).*
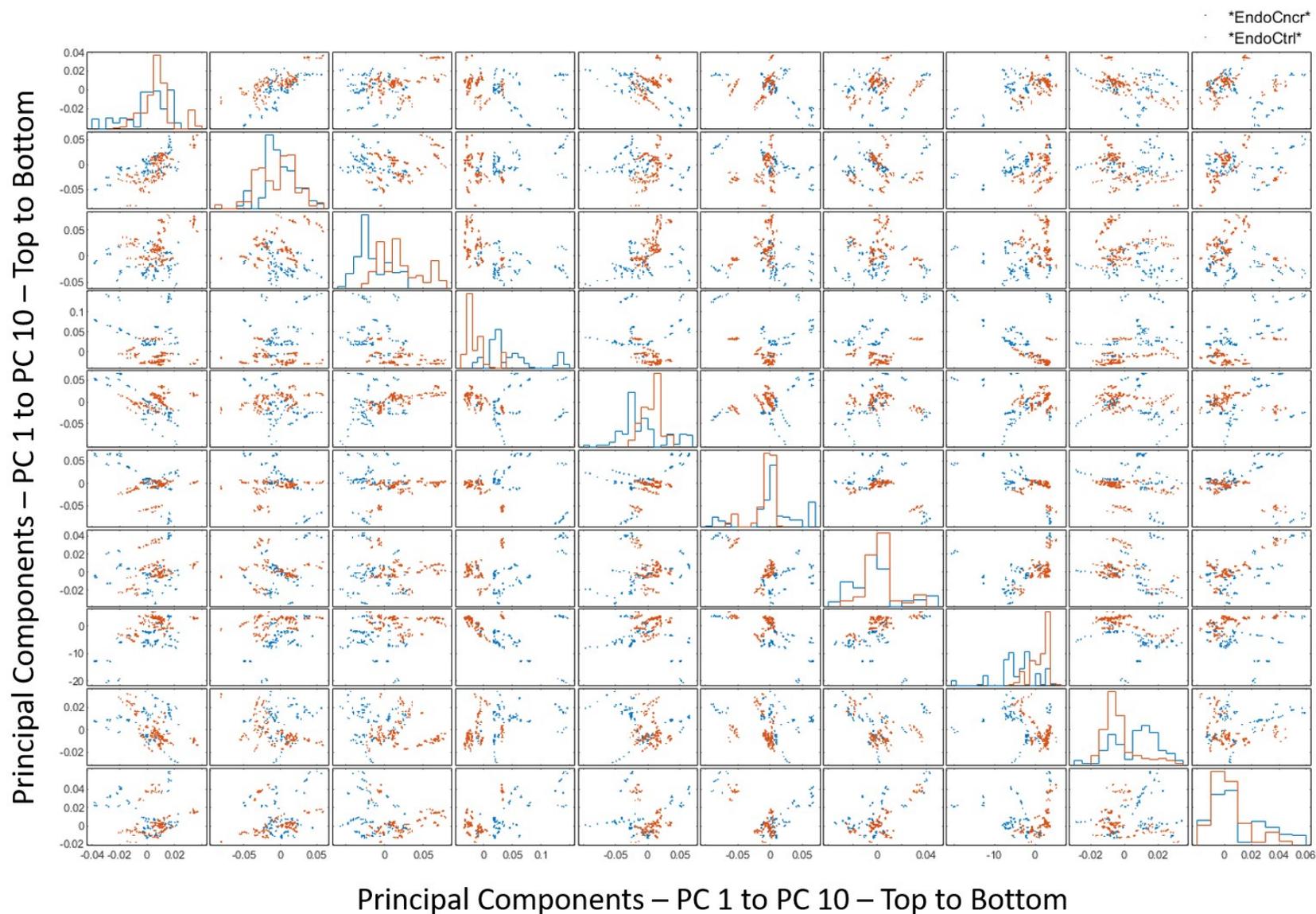
Table S1. Important features found in endometrial plasma and serum in the 1430 cm^{-1} to 900 cm^{-1} region, with vibrational modes and tentative assignments [1].

Wavenumber / cm^{-1} - Plasma	Wavenumber / cm^{-1} - Serum	Tentative Assignments	Vibrational Modes
-	1423	Fatty acids, amino acid side chains	C=O stretching of –COO$^-$
-	1393	Proteins, lipids	Symmetric C-H deformation of CH$_3$
1369	-	Tyrosine, guanine	Stretching C-N
1358	1358	Fatty acids, amino acid side chains	C=O stretching of –COO$^-$, CH$_2$ wagging for proline
1312, 1288	1346, 1308, 1292	Proteins	C-N/N-H deformation (Amide III)
1246, 1215	1254	Nucleic acids (cfDNA), phospholipids	Asymmetric P=O stretching in PO$_2^-$
1173	1192, 1165	Polysaccharides, carbohydrates	C-O-C and C-O-P stretching and ring vibrations, Symmetric C-O stretching coupled to C-O-H bending
1092	1088	Nucleic acids (cfDNA), phospholipids	Symmetric stretching in PO$_2^-$, CO-O-C symmetric stretching vibration
1038	-	Carbohydrates	Symmetric C-O-C stretching
999	999	Carbohydrates	Symmetric C-O stretching
-	937	Carbohydrates or fatty acids	C-O or C-C stretching

Table S2. Important features found in endometrial plasma and serum in the 1800 cm^{-1} to 900 cm^{-1} region, with vibrational modes and tentative assignments [1, 2].

Wavenumber / cm^{-1} – Plasma	Wavenumber / cm^{-1} - Serum	Tentative Assignments	Vibrational Modes
1755, 1736 1720	1778, 1766, 1720	Phospholipids, thymine, uracyl	C=O stretching vibration of esters
1670	1690, 1670	Proteins	Anti-parallel β-sheet, turns and loops
1643, 1605	1643, 1601	α-helix, proteins (fibrinogen)	C=O stretch of Amide I group
1570	1535, 1508	Proteins (fibrinogen)	N-H bending, C-H stretching, C-O bending, C-C and N-C stretching (Amide II band)
-	1467, 1450	Lipids, proteins, nucleic acids	Symmetric and asymmetric C-H scissoring of –CH$_2$
1423	1423	Fatty acids, amino acid side chains	C=O stretching of –COO$^-$
1389	1389	Proteins (fibrinogen), lipids	Symmetric C-H deformation of CH$_3$
-	1358 (falling edge of 1400)	Fatty acids, amino acid side chains	C=O stretching of –COO$^-$, CH$_2$ wagging for proline
1346	1346, 1308	Proteins	C-N/N-H deformation (Amide III)
1207 (edges of 1235)	1254 (edges of 1235)	Nucleic acids (cfDNA), phospholipids	Asymmetric P=O stretching in PO$_2^-$
-	1161	Polysaccharides, carbohydrates	C-O-C and C-O-P stretching and ring vibrations, Symmetric C-O stretching coupled to C-O-H bending
-	1099	Nucleic acids (cfDNA), phospholipids	Symmetric stretching in PO$_2^-$, CO-O-C symmetric stretching vibration
999 (descending edge of 1030)	1007 (descending edge of 1030)	Carbohydrates	Symmetric C-O-C/C-O stretching
968	937	Carbohydrates or fatty acids	C-O, C-C stretching

References

[1] L. Shi and R. R. Alfano, Deep imaging in tissue and biomedical materials: Chapter 8: Biomedical

   applications in probing deep tissue using mid-infrared supercontinuum optical biopsy, Singapore: Pan

Standord Publishing Pte. Ltd., 2017.

[2] M. Boix, S. Eslava, G. C. Machado, E. Gosselin, N. Ni, E. Saiz and J. D. Coninck, "ATR-FTIR

measurements of albumin and fibrinogen adsorption: Inert verses calcium phosphate ceramics," *Journal*

of Biomedical Materials Research Part A, vol. 103, no. 11, pp. 3493-3502, 2015.