Electronic Supplementary Material (ESI) for Analyst.

Characterization of Ovarian Cancer-Derived Extracellular Vesicles by Surface-Enhanced Raman Spectroscopy – Supplementary Information

Nina M. Ćulum,[1] Tyler T. Cooper, [2] Gillles A. Lajoie,[2] Thamara Dayarathna,[3] Stephen H. Pasternak,[3] Jiahui Liu,[4] Yangxin Fu,[4] Lynne-Marie Postovit,[4,5] François Lagugné-Labarthet[1]*

[1] University of Western Ontario (Western University), Department of Chemistry, 1151 Richmond St., London, Ontario, Canada, N6A 5B7

[2] University of Western Ontario (Western University), Department of Biochemistry, 1151 Richmond St., London, Ontario, Canada, N6A 5B7

[3] University of Western Ontario (Western University), Robarts Research Institute, 1151 Richmond St., London, Ontario, Canada, N6A 5B7

[4] University of Alberta, Department of Oncology, 116 St. & 85 Ave., Edmonton, Alberta, Canada, T6G 2R3

[5] Queen's University, Department of Biomedical & Molecular Sciences, 99 University Ave., Kingston, Ontario, Canada, K7L 3N6

Table S1. Parameters for Q Exactive Plus.

Parameter	**Parameter Setting**
Orbitrap Resolution (MS1)	7×10^4
Mass Range	400 – 1500 m/z
MS1 Injection Time	200 ms
MS1 Automatic Gain Control (AGC) Target	3×10^6 ions/cycle
Lock Mass	445.120025 m/z
MS2 Detection	Fourier Transform
MS2 Resolution	1.75×10^4
MS2 AGC Target	2×10^5 ions/cycle
MS2 Injection Time	50 ms
Loop Count	12
Isolation Width	1.2 m/z
Isolation Offset	0.5 m/z
MS2 Activation	Higher-energy C-trap dissociation
Normalized Collision Energy	25 %
Dynamic Exclusion	12
Minimum AGC Target	2×10^3 ions/cycle
MS2 Intensity Threshold	8×10^4
Exclusion Duration	30 s
Charge Exclusion	Unassigned, 1, 8, >8
Polarity	Positive

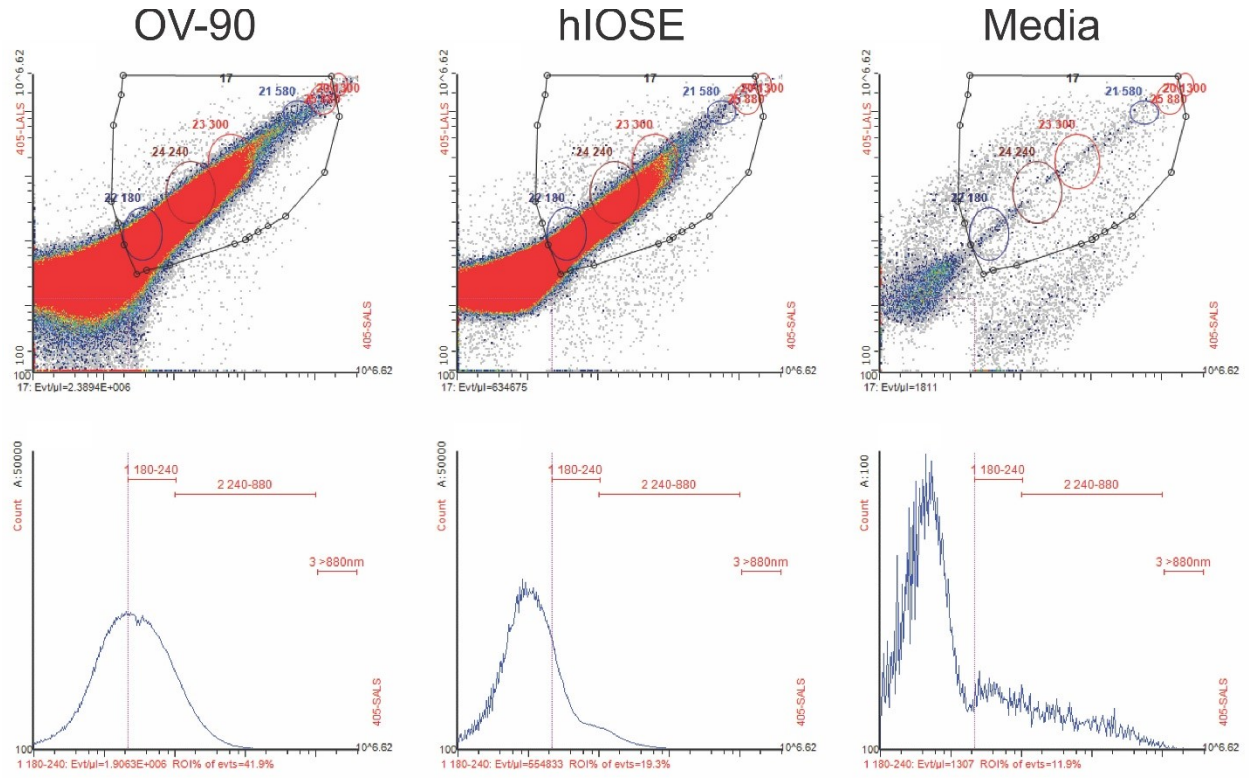

Figure S1. Representative flow cytometry plots (top) and corresponding size distributions (bottoms) of EV samples from OV-90 (left) and hIOSE (middle) cell lines, as well as conditioned media (right).

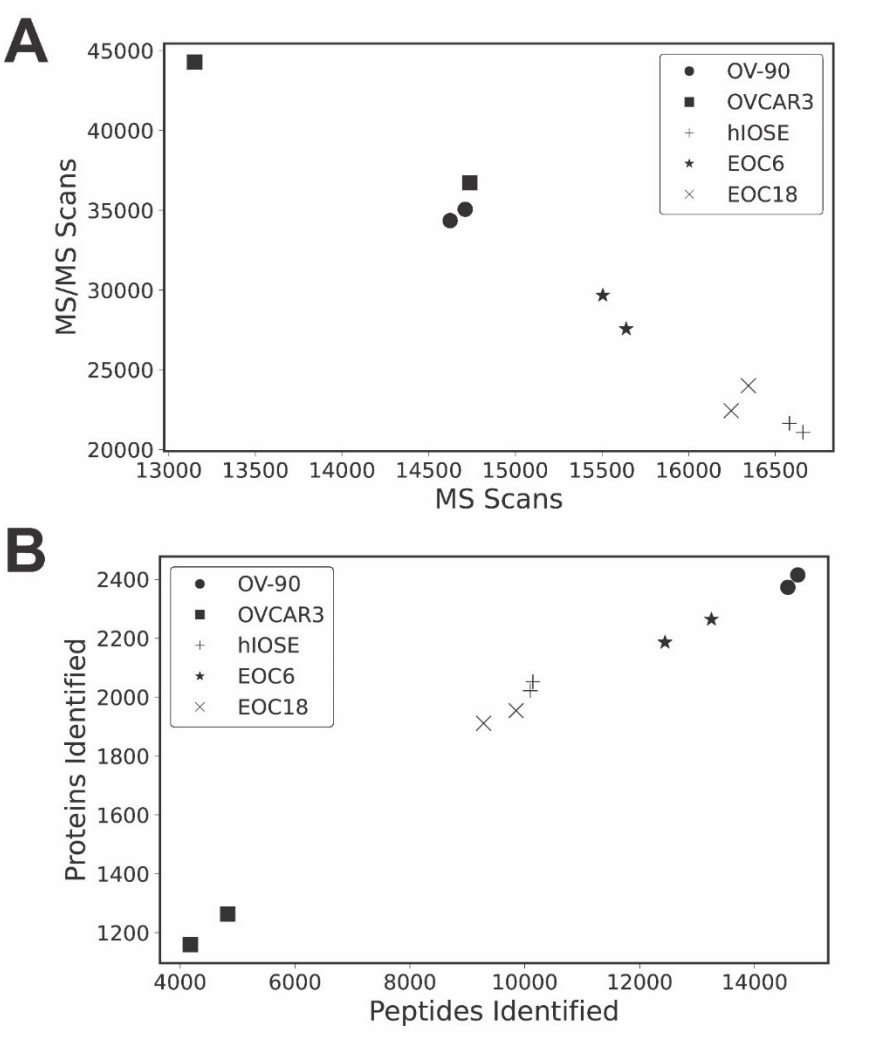

Figure S2. (A) Number of MS1 and MS2 scans for duplicate injections of each EV sample. **(B)** Number of peptides identified by de novo sequencing, which led to confident protein identification with false discovery rate ($p < 0.01$).

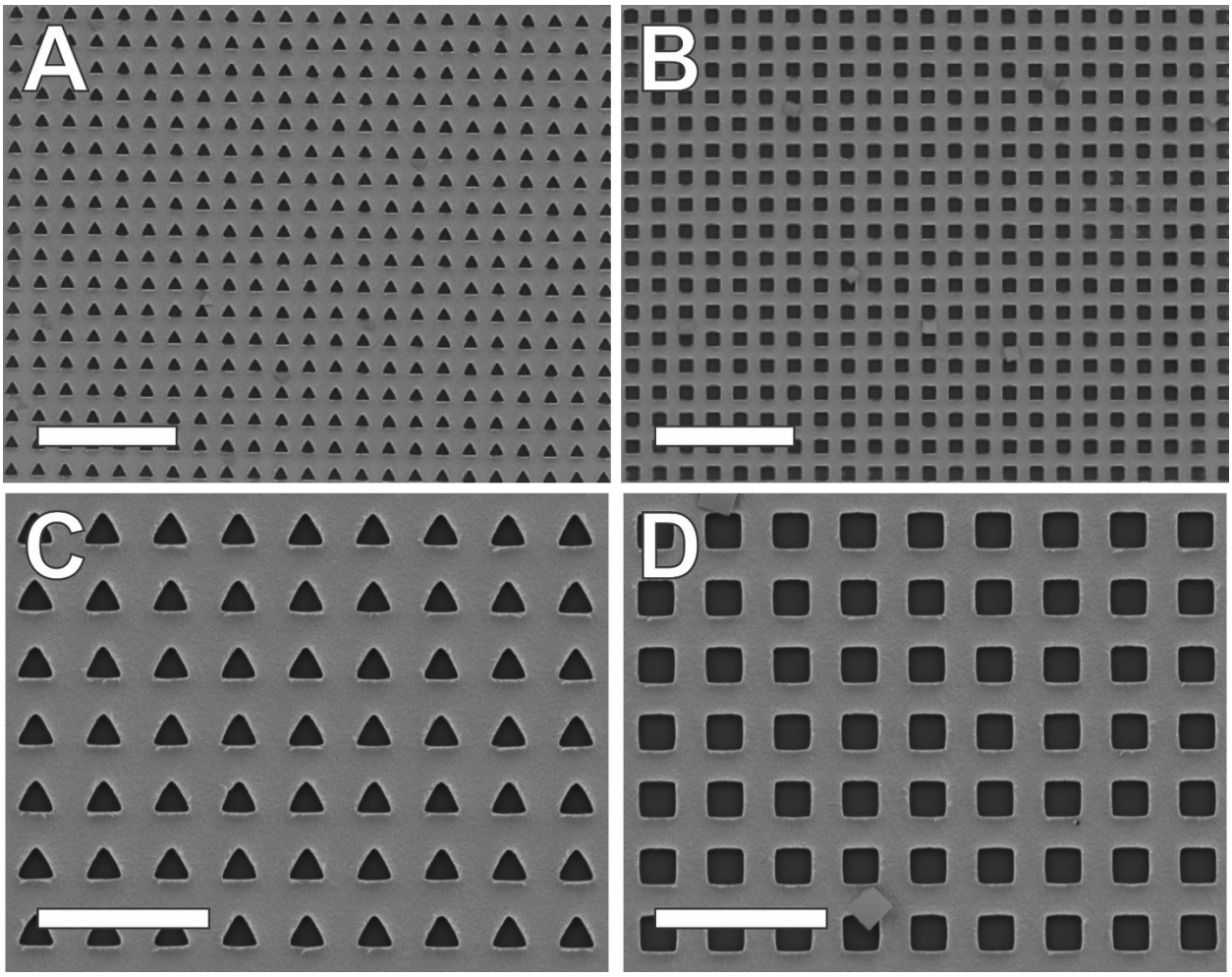

Figure S3. SEM images of typical nanohole arrays (edge length = 1.0 μm) used for EV capture and analysis, including **(A)** triangular arrays and **(B)** square arrays (scale bars = 10 μm). **(C)** and **(D)** represent magnified images of **(A)** and **(B)**, respectively (scale bars = 5 μm).

Table S2. Summary of peak assignments of hIOSE EVs, as highlighted in **Fig. 3A**.

Raman Shift (cm^{-1})	**Presumed Assignment**	**Ref. Peak (cm^{-1})**
630	Glycerol	630[1,2]
692	Ring deformation	686[2]
724	Ring breathing mode of adenine	725[1,2]
755	Symmetric breathing of tryptophan	755[2]
787	Ring breathing mode of cytosine, uracil, thymine	786[1,2]
818	C-C stretching in collagen	817[1,2]
837	Deformative vibrations of amine groups	838[1,2]
865	C-C stretching or C-O-C skeletal mode in carbohydrates	868[1,2]
935	C-C stretching mode of proline, valine, and protein backbone (α-helix); glycogen	935[1,2]
961	Unassigned in protein assignments	963[1,2]
1029	Phenylalanine of collagen	1030[2]
1090	Symmetric phosphate stretching vibrations	1090[1,2]
1185	Cytosine, guanine, adenine	1180 – 1184[1,2]
1254	Lipids	1255[1,2]
1303	CH_2/CH_3 twisting, wagging, or bending mode of lipid/collagen; amide III	1302[1,2]
1356	Guanine	1355, 1357[1,2]
1404	C-H deformation	1404[1]
1455	Deoxyribose; CH_2 scissoring of proteins and lipids	1455[2]
1482	Ring breathing mode of guanine, adenine	1485[1,2]
1545	Amide II	1544[1]

Table S3. Summary of peak assignments of OV-90 EVs, as highlighted in **Fig. 3B**.

Raman Shift (cm^{-1})	**Presumed Assignment**	**Ref. Peak (cm^{-1})**
632	Glycerol	630[1,2] *
675	Ring breathing mode in guanine	678[1,2]
714	C-N (membrane phospholipid head); adenine	717[2]
719	C-N (membrane phospholipid head), symmetric stretch vibration of choline group $N^+(CH_3)_3$; nucleotide	719[1,2]
755	Symmetric breathing of tryptophan	755[2] *
782	Ring breathing in thymine, cytosine, uracil	782[1,2]
822	Phosphodiester	822[1,2]
908	Tyrosine	906[1]
943	Skeletal modes in polysaccharides	941[1,2]
1008	Phenylalanine	1008[1,2]
1036	C-H in-plane bending mode of phenylalanine	1036[2]
1118	Glucose	1117[1,2]
1151	C-N stretching in proteins	1152[1,2]
1186	Cytosine, guanine, adenine	1180 – 1184[1,2] *
1201	Nucleic acids and phosphates; aromatic C-O and C-N	1200[1,2]
1220	C=N=C stretching	1220[1,2]
1274	Amide III	1275[1,2]
1335	CH_2/CH_3 twisting and wagging in collagen and nucleic acids; C-N stretching in amide III	1335[1,2]
1370	Saccharide band	1370[1,2]
1403	C-H deformation	1404[1] *
1467	Lipids	1465[1,2]
1533	Amide II	1542[1]

* Also present in control (hIOSE) spectrum.

Table S4. Summary of peak assignments of OVCAR3 EVs, as highlighted in **Fig. 3B**.

Raman Shift (cm^{-1})	**Presumed Assignment**	**Ref. Peak (cm^{-1})**
675	Ring breathing mode in guanine	678[1,2]
687	Ring deformation	686[2] *
724	Ring breathing mode of adenine	725[1,2] *
741	DNA, tryptophan	742[1]
818	C-C stretching in collagen	817[1,2] *
848	C-O-C skeletal mode in carbohydrates	847[1,2]
929	Carbohydrates	931[1,2]
935	C-C stretching mode of proline, valine, and protein backbone (α-helix); glycogen	935[1,2] *
956	CH_3 stretching in proteins (α-helix)	951[1,2]
994	C-O ribose, C-C	996[1,2]
1055	C-O stretching, C-N stretching in proteins	1053[1,2]
1176	C-H bending in tyrosine	1176[1]
1197	Tryptophan ring breathing	1199[1]
1226	Amide III (β-sheet)	1224[1,2]
1299	Acyl chains, fatty acids	1298[1,2]
1324	CH_2/CH_3 wagging mode in collagen and purine bases	1324[1,2]
1346	Adenine and guanine; C-H deformation of proteins	1344[2]
1376	Ring breathing mode of adenine	1376[2]
1455	Deoxyribose; CH_2 scissoring of proteins and lipids	1455[2] *
1483	Ring breathing mode of guanine, adenine	1485[1,2] *
1529	-C=C- in-plane vibrations	1525[1,2]
1584	C=C bending mode of phenylalanine	1583[1,2]

* Also present in control (hIOSE) spectrum.

Table S5. Summary of peak assignments of EOC6 EVs, as highlighted in **Fig. 3C**.

Raman Shift (cm^{-1})	**Presumed Assignment**	**Ref. Peak (cm^{-1})**
639	Tyrosine ring breathing	639[1]
694	Ring deformation	686[2] *
736	Phosphatidylserine	733[1,2]
775	Phosphatidylinositol	776[1,2]
797	Ring breathing mode in uracil	802[1,2]
845	C-O-C skeletal mode in carbohydrates	847[1,2]
886	Ring deformation and symmetric C-N-C stretching	886[2]
939	C-C skeletal stretching in proteins	939[2]
1003	C-C skeletal mode, phenylalanine	1003[1,2]
1023	Glycogen	1023[1,2]
1096	Phosphodioxy group (PO_2^- in nucleic acids)	1096[1,2]
1159	C-C/C-N stretching in proteins	1158[1]
1162	Tyrosine	1163[1]
1225	Amide III (β-sheet)	1224[1,2]
1265	Amide III of collagen; C-C_6H_5 stretching in phenylalanine	1265[2]
1332	C-C stretching in phenyls, C-O stretching, C-H in-plane bending	1332[1,2]
1367	CH_3 stretching in phospholipids	1367[1,2]
1404	C-H deformation	1404[1] *
1439	CH_2/CH_3 deformation in collagen	1439[1,2]
1520	-C=C- in-plane vibrations	1525[1,2]
1558	Tryptophan, tyrosine, amide II	1558[1,2]

* Also present in control (hIOSE) spectrum.

Table S6. Summary of peak assignments of EOC18 EVs, as highlighted in **Fig. 3C**.

Raman Shift (cm^{-1})	**Presumed Assignment**	**Ref. Peak (cm^{-1})**
612	Cholesterol ester	614[1,2]
648	Ring, cyclic deformation	649[2]
677	Ring breathing in guanine	678[1,2]
756	Symmetric breathing of tryptophan	755[2] *
835	Deformative vibrations of amine groups	838[1,2] *
852	Proline, hydroxyproline, tyrosine	852[1,2]
935	C-C stretching mode of proline, valine, and protein backbone (α-helix); glycogen	935[1,2] *
987	Phenylalanine	991[1]
1032	CH_2/CH_3 bending modes of phenylalanine and proline of collagen, phospholipids	1032[1,2]
1076	Symmetric stretching of PO_4^{3-}	1076[1]
1166	Lipids	1168[1,2]
1209	$C\text{-}C_6H_5$ stretching mode in tryptophan and phenylalanine	1209[1,2]
1248	Amide III	1248[1]
1262	Ring breathing mode in thymine, adenine; =C-H bending in proteins	1263[1,2]
1300	CH_2 twisting in lipids, fatty acids	1300[1,2]
1338	Amide III	1338[2]
1362	Tryptophan	1360[1,2]
1386	CH_3 band	1386[1,2]
1425	Deoxyribose	1424[1,2]
1466	Lipids	1465[1,2]
1473	C=N stretching	1470[1,2]
1577	Guanine, adenine	1578[1]

* Also present in control (hIOSE) spectrum.

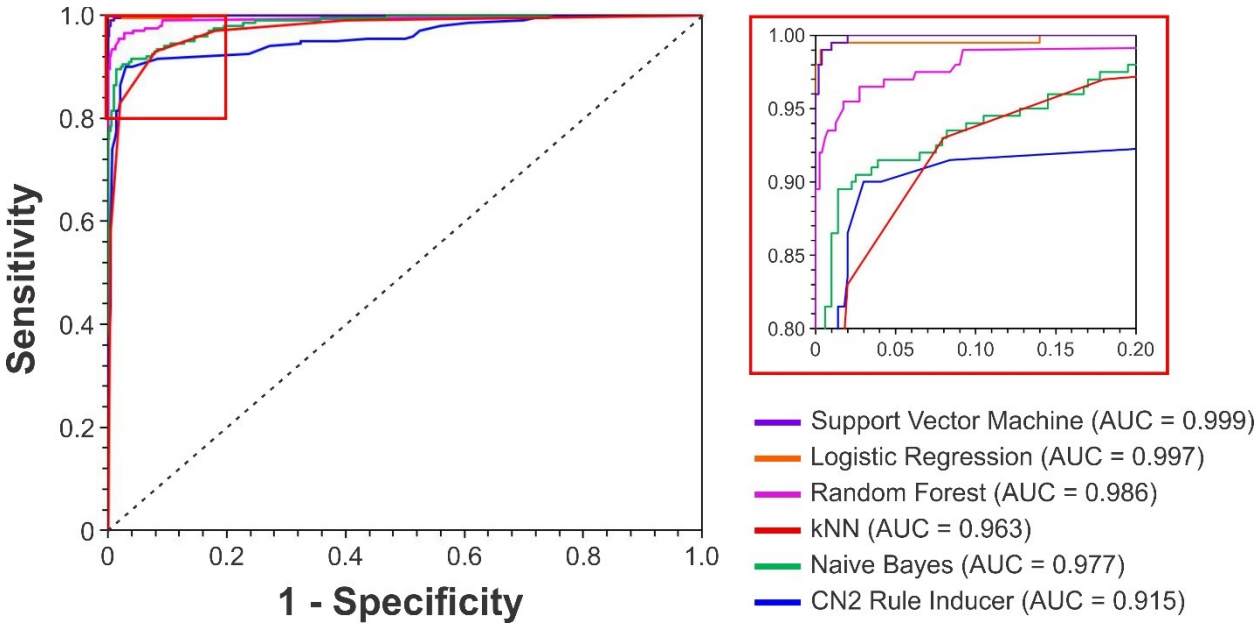

Figure S4. ROC curves comparing six different machine learning algorithms: support vector machine (SVM), logistic regression, random forest, kNN, Naïve Bayes, and CN2 rule inducer. The upper left-most portion is zoomed in and highlighted in red (right).

Table S7. Comparison of accuracies, sensitivities, and specificities achieved with the six learning machine algorithms shown in Fig. S3. Although SVM has a slightly higher AUC (0.999) than logistic regression (0.997), logistic regression is able to classify each EV type with higher accuracy, precision, and recall.

Model	**Accuracy**	**Precision**	**Recall**
Logistic Regression	98.6 %	98.6 %	98.6 %
SVM	97.3 %	97.4 %	97.3 %
Random Forest	91.0 %	91.1 %	91.0 %
Naïve Bayes	86.5 %	88.4 %	86.5 %
kNN	86.5 %	87.1 %	86.5 %
CN2 Rule Inducer	77.4 %	77.4 %	77.4 %

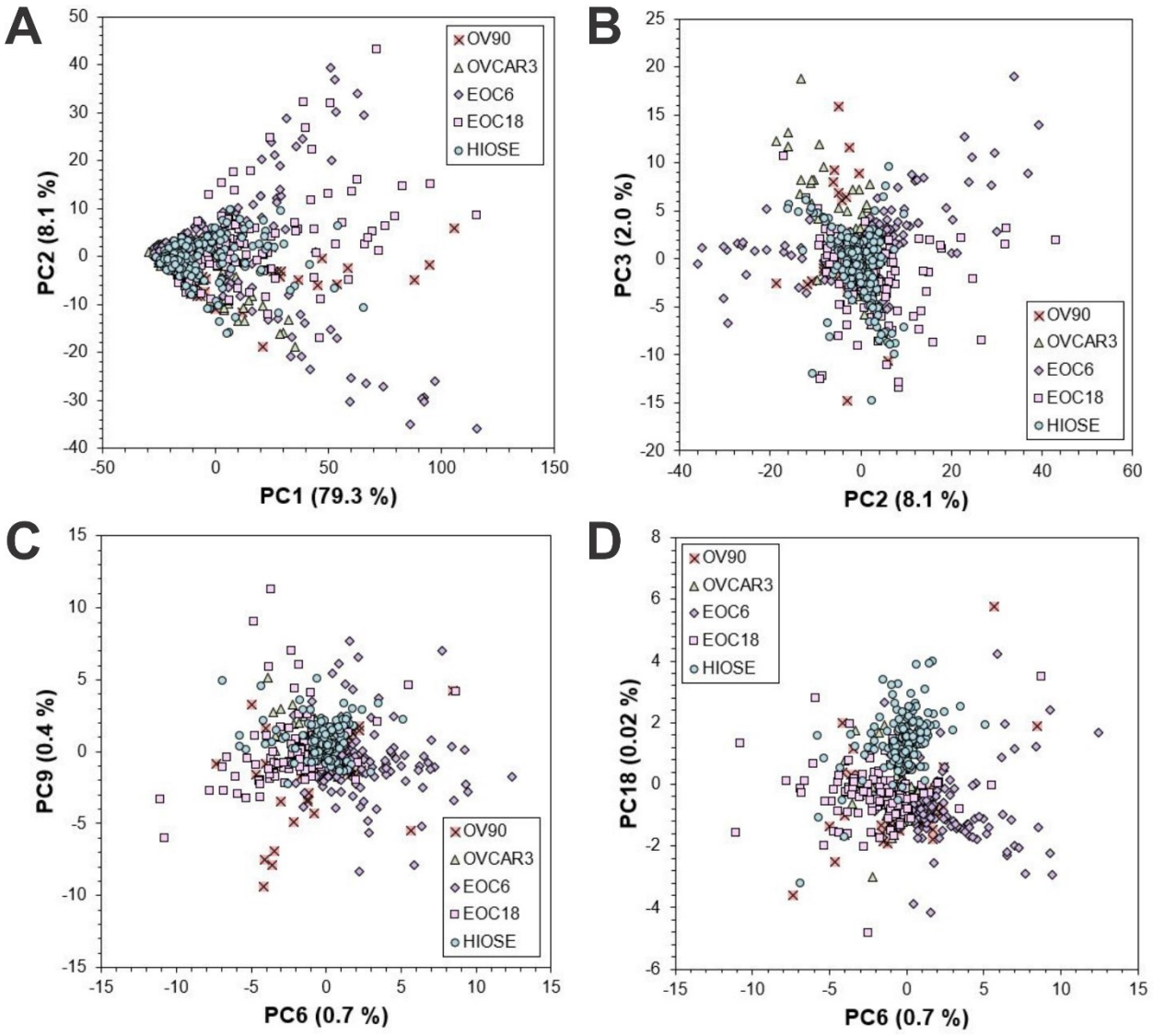

Figure S5. Score plots containing (A) first two PCs (i.e., a traditional score plot), (B) PC2 and PC3 (informative projection for the first 5 PCs), (C) PC6 and PC9 (informative projection for both the first 10 and 15 PCs), and (D) PC6 and PC18 (informative projection for the first 20 PCs).

Table S8. A comparison of the descriptive statistics when varying amounts of PCs are retained for machine learning.

Number of PCs Retained	Total Explained Variance	Accuracy	Precision	Recall
5	91.5 %	46.3 %	50.6 %	46.3 %
10	94.0 %	69.0 %	69.6 %	69.0 %
15	95.5 %	69.0 %	69.6 %	69.0 %
20	96.5 %	94.6 %	94.8 %	94.6 %
25	97.2 %	98.6 %	98.6 %	98.6 %

Table S9. Confusion matrix generated when the first 5 PCs are retained for machine learning.

		Predicted				
		EOC6	**EOC18**	**OV-90**	**OVCAR3**	**hIOSE**
Actual	**EOC6**	70	12	2	16	57
	EOC18	13	48	9	6	60
	OV-90	2	3	45	0	50
	OVCAR3	12	0	4	38	44
	hIOSE	21	22	8	30	119

Table S10. Confusion matrix generated when the first 15 PCs are retained for machine learning.

		Predicted				
		EOC6	**EOC18**	**OV-90**	**OVCAR3**	**hIOSE**
Actual	**EOC6**	133	3	4	4	13
	EOC18	4	79	4	6	43
	OV-90	1	3	74	4	18
	OVCAR3	5	2	12	58	21
	hIOSE	20	18	12	17	133

Table S11. Confusion matrix generated when the first 20 PCs are retained for machine learning.

		Predicted				
		EOC6	**EOC18**	**OV-90**	**OVCAR3**	**hIOSE**
Actual	**EOC6**	144	7	5	1	0
	EOC18	4	127	2	1	2
	OV-90	4	1	95	0	0
	OVCAR3	0	6	0	92	0
	hIOSE	1	3	0	0	196

Table S12. Confusion matrix generated when the first 25 PCs are retained for machine learning.

		Predicted				
		EOC6	**EOC18**	**OV-90**	**OVCAR3**	**hIOSE**
Actual	**EOC6**	153	1	0	1	2
	EOC18	0	134	1	1	0
	OV-90	0	0	99	0	1
	OVCAR3	0	1	0	97	0
	hIOSE	1	1	0	0	198

References

1. A. C. S. Talari, Z. Movasaghi, S. Rehman and I. U. Rehman, *Appl. Spectrosc. Rev.* 2015**, 50**, 46-111.

2. I. U. Rehman, Z. Movasaghi, S. Rehman, in *Vibrational Spectroscopy for Tissue Analysis*, CRC Press, Boca Raton, 2012, vol. 1, pp 213-294.