ELECTRONIC SUPPORTING INFROMATION

Serum Raman spectroscopy as a diagnostic tool in patients with Huntington's disease

Anna Huefner[♪§#], *Wei-Li Kuan*[§#], *Sarah L. Mason*[§], *Sumeet Mahajan*[‡*], *Roger A. Barker*[§*]

♪ Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1WE, United Kingdom

§ John van Geest Centre for Brain Repair and WT-MRC Cambridge Stem Cell Institute, University of Cambridge, Forvie Site, Robinson Way, Cambridge, CB2 0PY, United Kingdom

‡ Institute for Life Sciences and Department of Chemistry, University of Southampton, Highfield Campus, SO17 1BJ, Southampton, United Kingdom

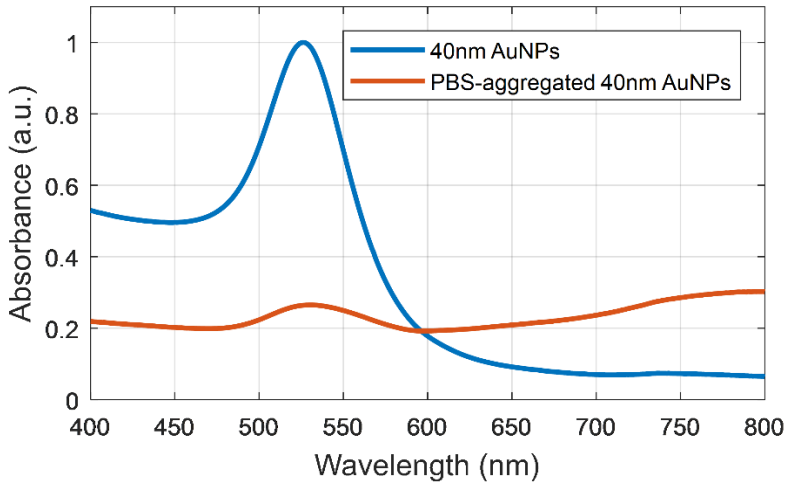
[#]*AH and WLK contributed equally to the work.*

*Corresponding authors: s.mahajan@soton.ac.uk, rab46@cam.ac.uk

SERS spectroscopy on cortical homogenates and serum samples from female R6/2 mice – a well-known transgenic murine model of HD

For SERS experiments, a solution of concentrated gold nanoparticles (AuNPs) was aggregated using phosphate buffered solution before mixing with mice cortical homogenates or serum (for details see Experimental section). After 30min of rest, aggregated AuNPs were spun down and the mixture was pipetted onto a glass coverslip and covered by another glass coverslip. As expected, the UV-Vis data shows a red-shift of the absorption maximum as the gold nanoparticle size increases after PBS aggregation (ESI Fig. S1). Areas on larger AuNP aggregates were chosen and measurements were performed in StreamLine© mode for map scans generating 200-500 individual spectra. This was repeated for all the mouse samples (ESI Table S1).

ESI Fig. S1 Representative Ultraviolent-visible absorption spectroscopy analysis indicates a red-shift of the absorption spectra following incubation with 10x PBS, confirming the aggregation of AuNPs for SERS analysis.

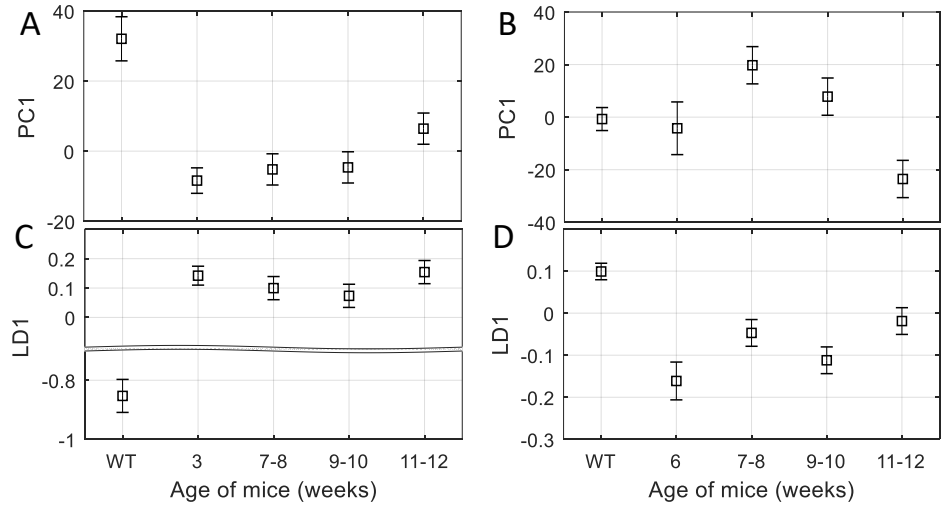


ESI Table S1 Details for all mice cortical homogenates or serum samples including strain, age and number PC1 loadings for analysis.

	Cortical homogenates					Serum				
Mice strain	WT	R6/2	R6/2	R6/2	R6/2	WT	R6/2	R6/2	R6/2	R6/2
Age of mice (weeks)	6	3	7-8	9-10	11-12	6-12	6	7-8	9-10	11-12
Number of PC1 loadings	6	18	12	12	12	21	4	8	8	8

PCA analysis was used to generate the Principal Component loadings for further analysis. The SERS data comprising of PC1 loadings from the mouse cortical homogenates or serum, was grouped into WT and HD mice samples (3 to 12 weeks of age) and analysed by principal component analysis (PCA) and linear discriminant analysis (LDA, grouped into WT and HD). In brief, PCA reduces the dimensions of the data sets while retaining key features of it, and LDA transforms the spectra into LD space allowing one to segregate the assigned groups. Their distribution was analysed using ANOVA with *post-hoc* adjustment (ESI Fig. S2, ESI Table S2). While SERS of cortical homogenates showed a significant distinction between WT and HD in both the PC and the LD space, a significant segregation between WT and HD samples was only observed in the LD space for serum samples.

ESI Fig. S2 PC1 (A-B) and LD1 (C-D) scores generated from SERS spectra of cortical homogenates of the R6/2 mice (A and C) and serum (B and D) at different disease stages. Box and whiskers indicate estimated mean and standard deviation following PCA-LDA analysis.



ESI Table S2 Difference between the estimated group means (Δ), the lower (CI_-) and upper limits (CI_+) for 95% confidence intervals for the true mean difference and the *p*-value after statistical analysis (post-hoc ANOVA) for the different comparison groups.

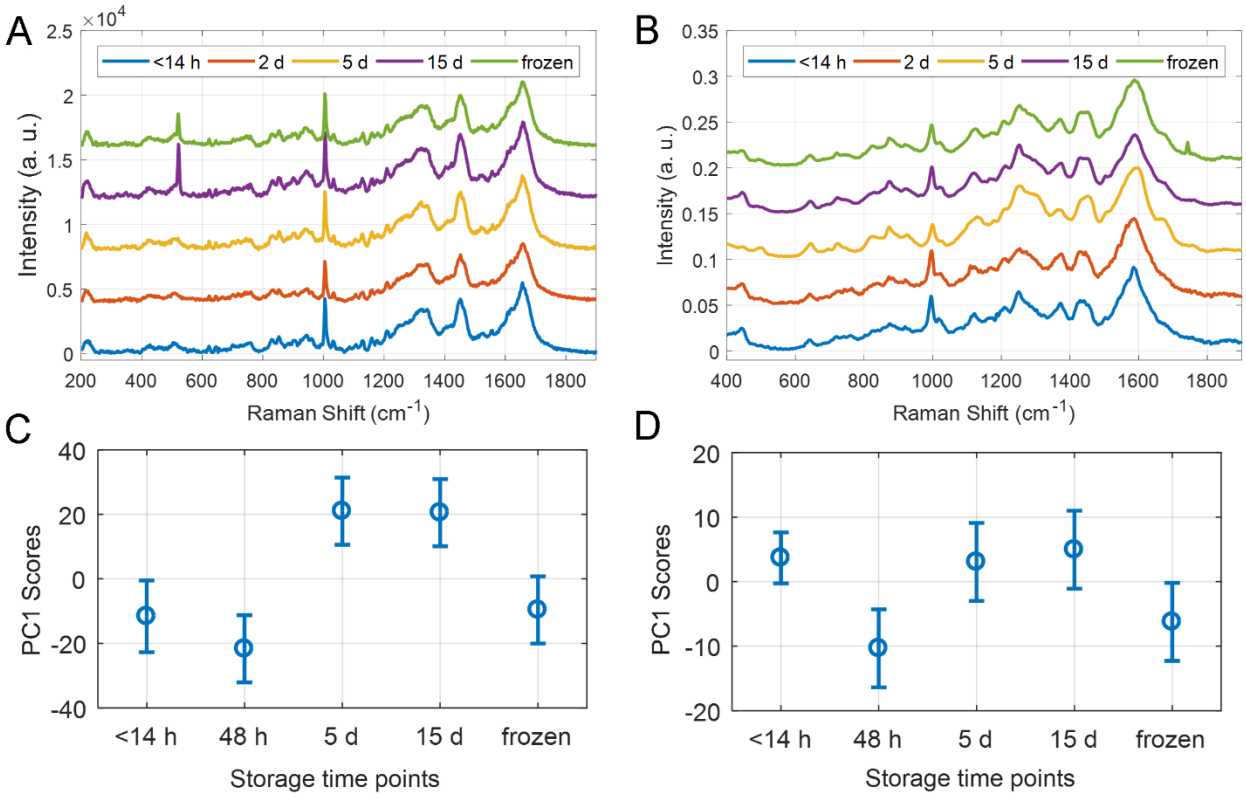
	Comparison Group		CI_-	Δ	CI_+	*p*-value	CI_-	Δ	CI_+	*p*-value
			PC1				LD1			
Cortical homogenates	WT	3 weeks	19.95	40.46	60.96	<0.0001	-1.18	-1.00	-0.81	<0.0001
	WT	7-8 weeks	15.52	37.26	59.01	<0.0001	-1.15	-0.95	-0.76	<0.0001
	WT	9-10 weeks	14.94	36.69	58.44	<0.0001	-1.12	-0.93	-0.73	<0.0001
	WT	11-12 weeks	3.89	25.63	47.38	0.01	-1.20	-1.01	-0.81	<0.0001
	3 weeks	7-8 weeks	-19.40	-3.19	13.02	0.98	-0.10	0.04	0.19	0.92
	3 weeks	9-10 weeks	-19.97	-3.77	12.44	0.96	-0.07	0.07	0.21	0.66
	3 weeks	11-12 weeks	-31.03	-14.82	1.39	0.09	-0.16	-0.01	0.13	1.00
	7-8 weeks	9-10 weeks	-18.33	-0.57	17.18	1.00	-0.13	0.03	0.18	0.99
	7-8 weeks	11-12 weeks	-29.39	-11.63	6.13	0.36	-0.21	-0.05	0.10	0.86
	9-10 weeks	11-12 weeks	-28.81	-11.06	6.70	0.41	-0.24	-0.08	0.08	0.60
			PC1				LD1			
Serum	WT	6 weeks	-27.59	3.51	34.61	1.0	0.12	0.26	0.40	<0.0001
	WT	7-8 weeks	-44.15	-20.47	3.22	0.12	0.04	0.15	0.25	<0.0001
	WT	9-10 weeks	-32.22	-8.53	15.15	0.84	0.10	0.21	0.32	<0.0001
	WT	11-12 weeks	-0.88	22.80	46.49	0.06	0.01	0.12	0.22	0.02
	6 weeks	7-8 weeks	-58.89	-23.98	10.93	0.30	-0.27	-0.11	0.04	0.24
	6 weeks	9-10 weeks	-46.95	-12.04	22.87	0.86	-0.21	-0.05	0.11	0.90
	6 weeks	11-12 weeks	-15.62	19.29	54.20	0.52	-0.30	-0.14	0.01	0.09
	7-8 weeks	9-10 weeks	-16.57	11.94	40.44	0.76	-0.06	0.07	0.19	0.60
	7-8 weeks	11-12 weeks	14.77	43.27	71.78	<0.001	-0.16	-0.03	0.10	0.97
	9-10 weeks	11-12 weeks	2.83	31.34	59.84	0.02	-0.22	-0.09	0.03	0.25

Dependency between serum Raman/SERS data and sample storage time

RS/SERS signatures of whole blood samples have been shown to change due to storage time of the sample, which is the time between taking the sample from the patient and conducting the RS/SERS measurements. These changes have primarily been linked to the deterioration of blood cellular components (red blood cells, white blood cells, and/or platelets) by Premasiri and co-

workers.[1] These changes appear to be attributed to hypoxanthine, a product of purine degradation, which are evident in SERS spectra of whole blood and plasma (vibrational band of the hypoxanthine found at 724 cm^{-1}), but not evident in RS spectra.[1]

ESI Fig. S3 Serum Raman and SERS spectra depending on the storage time. Representative raw spectra for RS (A), as well as the PC1 loading of SERS (B), at different time points. The PC1 loadings, rather than the raw SERS data, were shown due to the heterogeneity associated with local enhancement of the signal. For Raman (C) and SERS (D) data, freezing as well as storage up to 48 h did not result in significant deterioration of the samples. Data were analyzed using ANOVA with *post-hoc* Bonferroni correction.



The use of serum for the conducted experiments eliminates any effects caused by degradation of blood cellular components. In order to evaluate the effect of storage time on serum RS/SERS

spectra, measurements were carried out following different storage times (< 14 h , 48 h , 5 d, and 15 d, all at 4˚C) as well as samples immediately frozen at -20˚C and thawed (ESI Fig. S3). The groups contained data (3 repeats per patient per time point) from four male HD patients with CAG repeat size of 43 and aged 54 to 57 years (median 56 years). For RS, storage time points of < 14 h and 48 h as well as frozen samples could not be distinguished following PCA analysis, while samples stored for 5 d and 15 d were significantly different from 48 h stored samples (ESI Fig. S3A, ESI Table S3). For SERS, all samples groups were not statistically different (ESI Fig. S3A, ESI Table S3).

ESI Table S3 Multi-comparison of storage times. Difference between the estimated group means (Δ), the upper limits (CI+) for 95% confidence intervals for the true mean difference and the p-value after statistical analysis (ANOVA with post-hoc Bonferroni correction) for the different storage times for Raman and SERS data.

Comparison Group			Raman				SERS			
			CI_-	Δ	CI_+	*p*-value	CI_-	Δ	CI_+	*p*-value
Storage time of serum samples	<14 h	48 h	-33.80	10.06	53.92	0.963	-6.21	14.02	34.24	0.306
	<14 h	5 d	-76.46	-32.60	11.26	0.227	-19.59	0.63	20.86	1
	<14 h	15 d	-75.99	-32.13	11.73	0.240	-21.49	-1.26	18.96	1
	<14 h	frozen	-45.85	-1.99	41.87	1	-10.31	9.92	30.14	0.647
	48 h	5 d	-85.03	-42.66	-0.28	0.048	-37.31	-13.38	10.55	0.524
	48 h	15 d	-84.56	-42.19	0.19	0.051	-39.21	-15.28	8.65	0.389
	48 h	frozen	-54.43	-12.05	30.32	0.923	-28.03	-4.10	19.83	0.989
	5 d	15 d	-41.91	0.47	42.84	1	-25.83	-1.89	22.04	0.999
	5 d	frozen	-11.77	30.61	72.98	0.252	-14.64	9.29	33.22	0.813
	15 d	frozen	-12.24	30.14	72.51	0.266	-12.75	11.18	35.11	0.687

Correlation of Raman/SERS results to clinical assessment parameters

The demographic information of our donors is summarized in ESI Table S4. Due to the complexity of the RS/SERS data as well as patient data, a methodology was developed to identify

correlations between clinical assessment parameters and the PC1 scores, which were calculated for spectral intervals of $50cm^{-1}$. For each subject, PC1 scores were generated and plotted against the respective clinical assessment parameters. Healthy and HD subjects were considered separately for all clinical assessment parameters, except for age. Clinical assessment parameters were grouped into four bins covering the full range of the rating scales and patient scores (ESI Table S5). ANOVA with *post-hoc* Bonferroni correction were generated as were estimated means and standard deviations for each of the five groups.

ESI Table S4 Demographic table showing participant details. Results are expressed as the median (interquartile range), minimum-maximum values and were analyzed using non-parametric Kruskal-Wallis test.

	HD male (n = 28)	**HD female** (n = 21)	**Healthy male** (n = 5)	**Healthy female** (n = 10)	**P**
Age at data collection	57(18), 33-84	44 (25), 29-74	72±6.5 (67-82)	70±8.7 (55-75)	0.001
CAG repeat size	43 (3), 39-47	45 (5), 42-57	n/a	n/a	0.002
Disease burden	390 (93.8), 248-580	461 (194.5), 276-672	n/a	n/a	0.229
UHDRS motor	21 (27), 0-59	34 (29), 0-67	n/a	n/a	0.511
Total functional capacity	7 (8), 2-13	7 (7), 1-13	n/a	n/a	0.826
Functional assessments	19 (11), 10-25	17 (11), 1-25	n/a	n/a	0.466
Independence	70 (13), 55-100	70 (15), 45-100	n/a	n/a	0.669

In the first instance, a linear correlation was sought between the PC1 scores and the grouped clinical assessment parameters, excluding healthy subjects. Following the linear fit of the estimated means of all HD groups, the adjusted R^2 (aR^2) values, which measures how well the

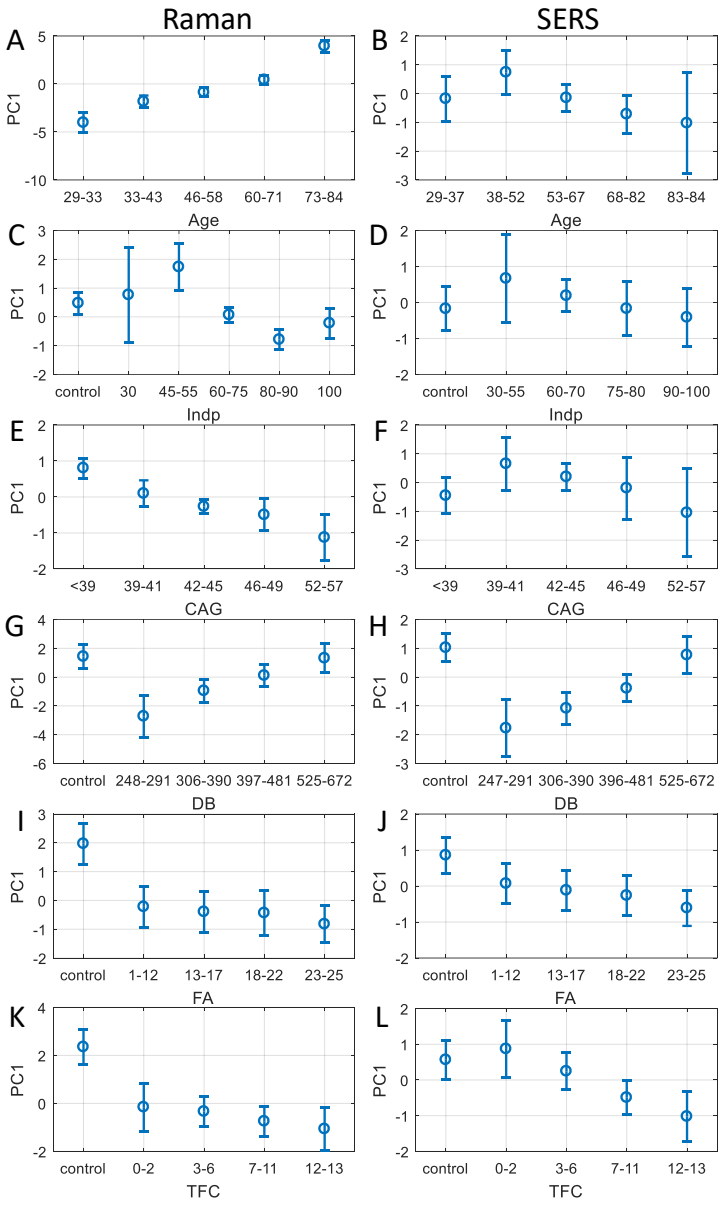
fitted regression approximates to the grouped data, were used which allows for comparison of spectral regions to clinical parameters.

EFI Table S5: Spectral segments (50cm^{-1} width) showing linearity between PC1 scores on RS/SERS and clinical assessment parameters as measured by the adjusted R^2 (aR^2). The best spectral region is characterised by the highest achieved aR^2. For corresponding boxplots, see **Error! Reference source not found.**. Colouration: aR^2 <0.9 (orange), 0.9 <aR^2< 0.95 (yellow), aR^2 > 0.95 (green).

			Clinical Assessment						
			Age	**Indp**	**CAG**	**DB**	**FA**	**TFC**	**UHDRS**
RS	**Regions (cm^{-1}) with aR^2 > 0.9**		650-700, 1750-1900	-	300-350* 1350-1450 1750-1850*	1600-1700	-	750-800 1200-1300	-
	Best region	**cm^{-1}**	1750-1800	550-600	300-350	1600-165	850-900	1200-125	750-800
		aR^2	0.98	0.79	1.00	0.93	0.88	0.98	0.80
SERS	**Regions (cm^{-1}) with aR^2 > 0.9**		-	1300-1350* 1550-1600	550-600 850-900* 1000-1100* 1550-1650* 1850-1900	700-750* 1000-1100* 1800-1900*	1550-1600 1800-1900*	700-750 1800-1900	1200-1300* 1600-1700
	Best region	**cm^{-1}**	1350-1400	1300-1350	1600-1650	700-750	1850-1900	1800-1850	1200-1250
		aR^2	0.55	0.98	1.00	1.00	0.96	1.00	0.91

On this basis, spectral segments (50 cm^{-1} width of each segment) were identified showing a good overall linear correlation between PC1 scores and the respective clinical assessment parameter (ESI Fig. S4 and ESI Table S5). Using either RS (ESI Fig. S4A,C,E,G,I,K) or SERS (ESI Fig. S4B,D,F,H,J,L), for each clinical assessment parameter a spectral segment exists showing an aR^2 greater than 0.91 (highlighted in green, ESI Table S5**Error! Reference source not found.**). Peaks within the identified relevant spectral segments showing high correlations with clinical assessment parameters were assigned (ESI Table S6) and used for the molecular interpretation of the results.

ESI Fig. S4 Grouped clinical assessment parameters, including age (A and B), Indp (C and D), CAG repeat size (E and F), DB score (G and H), FA (I and J) and TFC (K and L), vs PC1 scores from Raman (left) and SERS (right). PC1 scores were computed for the spectral segments ($50cm^{-1}$ width) showing the best linear correlation (highest aR^2) to the respective clinical assessment parameters as found in ESI Table S5.



ESI Table S6 Assignments of RS/SERS peaks linked to spectral intervals, which show a linear correlation to clinical assessments with an adjusted R^2 value that is greater than 0.9. (* indicates adjusted $R^2 > 0.95$). Abbreviations: Cysteine (Cys), Phenylalanine (Phe) Tryptophan (Trp), Tyrosine (Tyr).

Clinical Assessment	Spectral region (cm^{-1})		Peak	Assignment
	RS	SERS		
Age	650-700, 1750-1900	-	655 (r), 692 (r)	Tyr[2-5], C-S stretching mode[3], Cys, C-C twist[4,5] C-S twist[4], cholesterol[5]
Indp	-	1300-1350*, 1550-1600	1350 (s), 1565 (s), 1585 (s), 1628 (s)	Trp, α helix, phospholipid[2] Trp[2] Tyr[2], C=C bending[6], Phe[5,6], riboflavin[6] Tyr, Trp, C=C str[2], Amide I[7,8]
CAG	300-350*, 1350-1450, 1750-1850*	550-600, 850-900*, 1000-1100*, 1550-1650*, 1850-1900	350 (r), 1344 (r), 1408 (r) 640 (s), 875 (s), 996 (s), 1022 (s), 1120 (s), 1565 (s), 1585 (s), 1628 (s), 1855 (s)	Protein, Tyr[2] CH_3CH_2 wagging, Trp, adenine, guanine[4,6,9] Glutathione[2], symmetric CO_2^- stretch[7], uric acid[10] Tyr[2-4,7], C-S stretching[11], C-C twist, skeletal ring def in uric acid[12,13] Trp[2,3], lipids[5] Phe C-H stretch in phe[2,4,6], uric acid[12,13] Protein, lipids[2], C-C stretch[7], C-N in uric acid[12,13] Trp[2] Tyr[2], C=C bending[6,11], Phe[3,6,11], riboflavin[6,11] Tyr, Trp, C=C str[2], Amide I[7,8] C-C stretch of proline ring, ring breathing of Tyr[4]
DB	1600-1700	700-750*, 1000-1100*, 1800-1900*	1658 (r), 1667(r), 720 (s), 745 (s), 996 (s), 1022 (s), 1120 (s),	Protein, amide I, α helix, phospholipids[6] Amide I, β-sheet[7,8] C-H bending adenine[3,4,6,11], coenzyme[4,6,11], adenine[3] Phospholipid[2], Tyr[2], C-C stretch of proline ring[6] Phe C-H stretch in phe[2,4,6], uric acid[12,13] Protein, lipids[2], C-C stretch[7], C-N in uric acid[12,13] C-C stretch of proline ring, ring breathing of Tyr[4]
FA	-	1550-1600, 1800-1900*	1565 (s), 1585 (s), 1628 (s), 1855 (s)	Trp[2] Tyr[2], C=C bending[6,11], Phe[5,6,11], riboflavin[6,11] Tyr, Trp, C=C str[2] C-C stretch of proline ring, ring breathing of Tyr[4]
TFC	750-800, 1200-1300	700-750, 1800-1900	760 (r), 810 (r), 1210 (r), 1245 (r), 1274 (r), 720(s),	Trp[2] Trp[2], C-O-C and C-C backbone[5] Trp[2,4] Tyr[3,7], Phe[11], ring vibration[3,11] Lipids and amide III[5,9], β-sheet[7,8] C-H bending adenine, coenzyme[6,11]

			745(s), 1855(s)	C-H bending adenine[3,4,6,11], coenzyme[4,6,11] Phospholipid[2], Tyr[2], C-C stretch of proline ring[6] C-C stretch of proline ring, ring breathing of Tyr[4]
UHDRS	-	1200-1300*, 1600-1700	1205 (s), 1250 (s), 1632 (s) 1667 (s)	Trp[2,4], Tyr[3,7], Phe[11], ring vibration[11], Lipids and amide III[5,9], β-sheet Tyr, Trp, C=C str[2], Amide I[7,8] Amide I, β-sheet[7,8]

REFERENCE

1. W.R. Premasiri, J.C. Lee, L.D. Ziegler, Surface-Enhanced Raman Scattering of Whole Human Blood, Blood Plasma, and Red Blood Cells: Cellular Processes and Bioanalytical Sensing. *J. Phys. Chem. B*, 2012, **116**, 9376-9386.

2. J.L. Gonzalez-Solis, J.C. Martinez-Espinosa, L.A. Torres-Gonzalez, A. Aguilar-Lemarroy, L.F. Jave-Suarez, P. Palomares-Anda, Cervical Cancer Detection Based on Serum Sample Raman Spectroscopy. *Lasers Med. Sci.,* 2014, **29**, 979-985.

3. D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng, R. Chen, Colorectal Cancer Detection by Gold Nanoparticle Based Surface-Enhanced Raman Spectroscopy of Blood Serum and Statistical Analysis. *Opt. Express*, 2011, **19**, 13565-13577.

4. S. Li, Y. Zhang, J. Xu, L. Li, Q. Zeng, L. Lin, Z. Guo, Z. Liu, H. Xiong, S. Liu, Noninvasive Prostate Cancer Screening Based on Serum Surface-Enhanced Raman Spectroscopy and Support Vector Machine. *Appl. Phys. Lett.*, 2014, **105**, 091104.

5. M. Muratore, Raman Spectroscopy and Partial Least Squares Analysis in Discrimination of Peripheral Cells Affected by Huntington's Disease. *Anal. Chim. Acta*, 2013, **793**, 1-10.

6. S.X. Li, Y.J. Zhang, Q.Y. Zeng, L.F. Li, Z.Y. Guo, Z.M. Liu, H.L. Xiong, S.H. Liu, Potential of Cancer Screening with Serum Surface-Enhanced Raman Spectroscopy and a Support Vector Machine. *Laser Phys. Lett.,* 2014, **11**, 065603.

7. N.C. Maiti, M.M. Apetri, M.G. Zagorski, P.R. Carey, V.E. Anderson, Raman Spectroscopic Characterization of Secondary Structure in Natively Unfolded Proteins: A-Synuclein. *J. Am. Chem. Soc.,* 2004, **126**, 2399-2408.

8. N.M. Perney, L. Braddick, M. Jurna, E.T. Garbacik, H.L. Offerhaus, L.C. Serpell, E. Blanch, L. Holden-Dye, W.S. Brocklesby, T. Melvin, Polyglutamine Aggregate Structure in Vitro and in Vivo; New Avenues for Coherent Anti-Stokes Raman Scattering Microscopy. *PLoS ONE*, 2012, **7**, e40536.

9. A. Sahu, S. Sawant, H. Mamgain, C.M. Krishna, Raman Spectroscopy of Serum: An Exploratory Study for Detection of Oral Cancers. *Analyst*, 2013, **138**, 4161-4174.

10. C. Westley, Y. Xu, B. Thilaganathan, A.J. Carnell, N.J. Turner, R. Goodacre, Absolute Quantification of Uric Acid in Human Urine Using Surface Enhanced Raman Scattering with the Standard Addition Method. *Anal. Chem.,* 2017, **89**, 2472-2477.

11. S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, H. Zeng, Nasopharyngeal Cancer Detection Based on Blood Plasma Surface-Enhanced Raman Spectroscopy and Multivariate Analysis. *Biosens. Bioelectron.,* 2010, **25**, 2414-2419.

12. L. Zhao, J. Blackburn, C.L. Brosseau, Quantitative Detection of Uric Acid by Electrochemical-Surface Enhanced Raman Spectroscopy Using a Multilayered Au/Ag Substrate. *Anal. Chem.,* 2015, **87**, 441-447.

13. A. Bonifacio, S. Dalla Marta, R. Spizzo, S. Cervo, A. Steffan, A. Colombatti, V. Sergo, Surface-Enhanced Raman Spectroscopy of Blood Plasma and Serum Using Ag and Au Nanoparticles: A Systematic Study. *Anal. Bioanal. Chem.,* 2014, **406**, 2355-2365.