Central Lancashire Online Knowledge (CLoK)

Title	Attenuated total reflection Fourier-transform infrared (ATR - FTIR) spectroscopy to diagnose osteoarthritis in equine serum
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/28004/
DOI	https://doi.org/10.1111/evj.13115
Date	2020
Citation	Paraskevaidi, Maria, Hook, P. D., Medeiros-De-morais, Camilo De lelis orcid iconORCID: 0000-0003-2573-787X, Anderson, J. R., White, R., Martin-Hirsch, P. L., Peffers, M. J. and Martin, Francis L (2020) Attenuated total reflection Fourier-transform infrared (ATR - FTIR) spectroscopy to diagnose osteoarthritis in equine serum. Equine Veterinary Journal, 52 (1). pp. 46-51. ISSN 0425-1644
Creators	Paraskevaidi, Maria, Hook, P. D., Medeiros-De-morais, Camilo De lelis, Anderson, J. R., White, R., Martin-Hirsch, P. L., Peffers, M. J. and Martin, Francis L

MISS MARIA PARASKEVAIDI (Orcid ID : 0000-0002-1353-7849)

Attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy to diagnose osteoarthritis in equine serum

M. Paraskevaidi*[1], P. D. Hook[2], C. L. M. Morais[1], J. R. Anderson[3], R. White[4], P. L. Martin-Hirsch[5], M. J. Peffers[3] and F. L. Martin*[1]

[1]School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, UK;

[2]Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK;

[3]Institute of Ageing and Chronic Disease, University of Liverpool, L69 3BX, UK;

[4]Myerscough College, Bilsborrow, Preston PR3 0RY, UK and

[5]Sharoe Green Unit, Lancashire Teaching Hospitals NHS Foundation, Preston PR2 9HT, UK.

*Corresponding author email: mparaskevaidi@uclan.ac.uk or flmartin@uclan.ac.uk

Running head: Spectroscopy for osteoarthritis diagnosis

List of abbreviations

ATR-FTIR: attenuated total reflection Fourier-transform infrared

CI: confidence interval

OA: osteoarthritis

PCA-LDA/QDA: principal component analysis - linear discriminant analysis/ quadratic discriminant analysis

Summary

Background: Reliable and validated biomarkers for osteoarthritis (OA) are currently lacking.

Objective: To develop an accurate and minimally invasive method to assess OA-affected horses and provide potential spectral markers indicative of disease.

Study design: Observational, cross-sectional study.

Methods: Our cohort consisted of 15 horses with OA and 48 without clinical signs of the disease, which were used as controls. Attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy was used to investigate serum samples (50 μL) collected from these horses. Spectral processing and multivariate analysis revealed differences and similarities, allowing for detection of spectral biomarkers that discriminated between the two cohorts. A supervised classification algorithm, namely principal component analysis coupled with quadratic discriminant analysis (PCA-QDA), was applied to evaluate the diagnostic accuracy.

Results: Segregation between the two different cohorts, OA-affected and controls, was achieved with 100% sensitivity and specificity. The six most discriminatory peaks were attributed to proteins and lipids. Four of the spectral peaks were elevated in OA horses, which could be potentially due to an increase in lipids, protein expression levels and collagen, all of which have been previously reported in OA. Two peaks were found decreased and were tentatively assigned to the reduction of proteoglycan content that is observed during OA.

Main limitations: The control group had a wide range of ages and breeds. Pre-symptomatic OA cases were not included. Therefore, it remains unknown whether this test could be also used as an early diagnostic tool.

Conclusions: This spectrochemical approach could provide an accurate and cost-effective blood test, facilitating point-of-care diagnosis of equine OA.

Introduction

Osteoarthritis (OA) accounts for an estimated 60% of all observed lameness being in horses [1]. Current diagnosis of OA is based on clinical signs, imaging techniques or molecular biomarkers using tissue or biofluids, such as synovial fluid. However, there is yet no single biomarker that is considered accurate enough in OA detection and panels of different markers are preferred instead [1,2]. An ideal marker would be minimally invasive and require a cost-effective and less laborious analytical approach.

Indications of clinical equine OA can include swelling of the joint or pain during locomotion. Currently, a primary diagnosis is based on gait analysis/lameness assessment, diagnostic imaging and arthroscopy [3]. The results of a recent symposium have also been published, aiming to identify diagnostic/prognostic biomarkers and promising analytic techniques towards detection of the condition [1]. Analytical approaches included metabolomics, proteomics and spectroscopy. One study used a panel of biomarkers to improve the diagnostic capability in OA-affected horses using synovial fluid and serum; however, this required the use of expensive and laborious assays [2].

Vibrational spectroscopy is a promising method for diagnostics of a great variety of disorders. Minimally- or non-invasive biofluids, such as blood, urine, tears or saliva, have been investigated as important sources of biological information [4]. Spectrochemical

methods generate vibrational movements within the chemical bonds of a sample's constituents, producing a "spectral fingerprint". The spectra provide information about the chemical bonds, their structure and function; this allows one to make a reference range of what represents normal and then classify, using sophisticated computational algorithms, samples outside this range into emerging or developed disease.

Attenuated total refection Fourier-transform infrared (ATR-FTIR) spectroscopy uses an internal reflection element to direct the IR beam to the sample, thus creating an evanescent wave. The latter penetrates the sample a few microns in order to derive its chemical information [5]. The use of ATR-FTIR spectroscopy is advantageous in comparison to previous FTIR studies (in transmission or transflection mode), as it allows easier sample preparation and reduced light scattering which could cause spectral distortion. The main objective of this study was to investigate the feasibility of ATR-FTIR spectroscopy to detect OA using serum samples recovered from OA-affected horses. We hypothesised that circulating biomarkers of disease are reflected in the spectral profile. Using a classification algorithm, segregation between the OA and control groups was undertaken and subsequent diagnostic accuracy assessed. The levels of the discriminatory spectral markers, indicative of disease, were also calculated and correlated with certain bonds within biological molecules.

Materials and Methods

ATR-FTIR spectra were acquired after blood serum collection from 48 healthy horses (average age of 10.1 years [s.d. ± 7.1; range: 1 day-26 years]) of a range of breeds (Thoroughbreds and Thoroughbred cross, Cob, Warmblood, Pony, Sporthorses, Irish Draught Cross, Shire, Shetland, Part Bred Arab, Welsh, and Trotter) and 15 Thoroughbred racehorses diagnosed with OA (average age of 6.4 years [s.d. ± 1.8; range: 3-10 years]); age was

unknown for five controls. The minimum number of animals for a power of 80% was 44 considering two classes, based on a power test using a two-tailed *t*-test (data input as mean and s.d. in absorbance units of the pre-processed spectra for each class). Horses were assigned to the healthy group when no clinical diagnosis of OA was present in their clinical records. Horses with a clinical (e.g. gait analysis, diagnostic imaging techniques) or post-mortem (e.g. gross macroscopic or Mankin/OARSI scoring) diagnosis of OA were assigned into the OA group, with OA identified in one or more of the metacarpophalangeal, metatarsophalangeal or carpus joints.

Serum samples were collected and aliquoted into typical 1 ml microtubes. Centrifugation followed after blood clotting to discard clot and cellular material. Samples were allowed to stand at room temperature for 30 minutes to promote clotting. They were then centrifuged at 3500 rpm for 10 mins and the supernatant taken off and aliquoted. Samples were stored at -80°C until further use. For spectroscopic analysis, a volume of 50 μL serum was used from each horse after thawing, deposited on an IR-reflective slide and left to air-dry. A Bruker Tensor 27 FTIR spectrometer[a] with a Helios ATR attachment[b], containing a diamond crystal, was used. Spectral resolution was set at 8 cm^{-1} and 32 scans were acquired for each spectrum for optimal signal-to-noise ratio; two times zero-filling was used, giving a final 4 cm^{-1} data-spacing. The diamond crystal was cleaned with distilled water after each sample and a background spectrum was collected to eliminate atmospheric changes. Five spectra were acquired per blood spot for each donor (Supplementary Item 1).

Pre-processing and analysis were performed using an in-house developed toolbox within MATLAB (IRootLab[c]) while further classification was performed using lab-made routines. Spectral pre-processing (truncation at 1800-900 cm^{-1}, rubberband baseline correction, vector normalisation) was first carried out to correct for any non-biological differences (e.g. sample thickness) during measurements (Fig 1). Cross-validated principal

component analysis followed by linear discriminant analysis (PCA-LDA) and quadratic discriminant analysis (PCA-QDA) were applied independently to the spectral dataset. PCA reduces the original spectral dataset into a few number of principal components (PCs) accounting to the majority of the original explained variance, where the first PC accounts to the largest variance, followed by the second, third etc. [6]. Each PC is composed of scores and loadings, the first one representing the variance on sample direction, being used to assess similarities/dissimilarities between samples and formation of clusters. The loadings represent the variance on wavenumber direction, thus showing the weight of each wavenumber to the discrimination pattern observed on the PCA scores. Therefore, the loadings profiles can be used to extract spectral markers based on a peak detection algorithm, where the wavenumbers with highest absolute coefficients are the most important ones towards classification. The PCA decomposition takes the form of:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{01}$$

where $\mathbf{X}$ is a matrix containing the spectral dataset; $\mathbf{T}$ is a matrix containing the PCA scores; $\mathbf{P}$ is a matrix containing the PCA loadings; and $\mathbf{E}$ is a matrix containing the residuals. T superscript represents the matrix-transpose operation.

In PCA-LDA or PCA-QDA, the PCA scores $\mathbf{T}$ are used as input variables to a linear or quadratic discriminant classifier. PCA-LDA (L_{ik}) and PCA-QDA (Q_{ik}) classification scores are calculated in a non-Bayesian form as follows [7; 8]:

$$L_{ik} = (\mathbf{t}_i - \bar{\mathbf{t}}_k)^{\mathrm{T}} \mathbf{C}_{\mathrm{pooled}}^{-1} (\mathbf{t}_i - \bar{\mathbf{t}}_k) \tag{02}$$

$$Q_{ik} = (\mathbf{t}_i - \bar{\mathbf{t}}_k)^{\mathrm{T}} \mathbf{C}_k^{-1} (\mathbf{t}_i - \bar{\mathbf{t}}_k) \tag{03}$$

where $\mathbf{t}_i$ represents the PCA scores for sample i; $\bar{\mathbf{t}}_k$ the mean PCA scores array for class k; $\mathbf{C}_{\mathrm{pooled}}$ the pooled covariance matrix between the classes; and $\mathbf{C}_k$ the variance-covariance matrix for class k.

Both PCA-LDA and PCA-QDA are classification algorithms based on a Mahalanobis distance calculation. However, QDA accounts for the variance structure of each class individually, whereas LDA assumes classes having similar variances and then uses a single pooled covariance matrix among the classes; this usually makes QDA superior [8]. For classification, the pre-processed spectra (5 replicas) were averaged for each sample and the classes were divided into training (70%; 43 samples: 33 healthy controls, 10 OA) and test (30%; 20 samples: 15 healthy controls; 5 OA) sets using the Kennard-Stone sample selection algorithm [9]. The training set was used for model construction while the test set for final model evaluation. Since the classification was performed on a sample basis, the model output was a single value (e.g. -1 or 1) for each sample, thus each horse was classified as either healthy or OA; horses with other conditions were not included. Age and breed were not tested as a discrimination factor on the algorithms. Sensitivity and specificity values were calculated based on the external test set as follows [10]:

$$\text{Sensitivity } (\%) = \frac{\text{TP}}{\text{TP+FN}} \times 100 \tag{1}$$

$$\text{Specificity } (\%) = \frac{\text{TN}}{\text{TN+FP}} \times 100 \tag{2}$$

where TP stands for true positives; TN for true negatives; FP for false positives; and FN for false negatives.

The two groups were compared with a two-tailed Mann–Whitney U test in GraphPad Prism 7.0$^{\text{d}}$. Normality was tested by means of an Anderson-Darling test, where the null hypothesis (the data follow a normal distribution) was rejected at a 95% confidence level. A *P*-value ≤0.05 was considered significant.

Results

Multidimensional PCA scores plot and loadings are depicted in Figure 2A and 2B, respectively. Spectral data were initially processed by PCA-LDA in order to discriminate control from OA samples. Leave-one-out cross-validated PCA-LDA model was built using 10 PCs, accounting to 99.97% of cumulative explained variance and achieved an accuracy of only 35% (80% sensitivity and 20% specificity). Because of the poor predictive performance of PCA-LDA, PCA-QDA model was subsequently applied to classify the groups. PCA-QDA (10 PCs, 99.97% of explained variance) achieved the best classification performance, with 100% accuracy, sensitivity and specificity in the test set. Figure 3A depicts the PCA-QDA discriminant function (DF) where the separation of both groups of samples can be seen. The few samples overlapping in DF plot are from the training set, which achieved an accuracy of 85%. A Receiver Operating Characteristic (ROC) curve is shown for the test set in Figure 3B [area under the curve (AUC) = 1] and a graphical confusion matrix with the correct classification rate for each class in the test set is shown in Figure 3C.

A PCA loadings plot (Fig 2B) was generated to depict the spectral regions that were mostly responsible for the differentiation: 1539 cm^{-1} (bending vibration of (N-H) of proteins, Amide II band; $P = 0.01$; 95% CI = 0.00044 to 0.00345), 1474 cm^{-1} (bending vibration of (CH$_2$) of proteins/lipids; $P = 0.005$; 95% CI = -0.00305 to -0.00059), 1400 cm^{-1} (stretching vibration of (COO$^-$) of amino acids and/or bending vibration of (CH$_3$) of proteins; $P = 0.2$; 95% CI = -0.00042 to 0.00201), 1304 cm^{-1} (Amide III band, components of proteins; P <0.001; 95% CI = -0.00535 to -0.00232), 1161 cm^{-1} (stretching vibration of C-OH groups of proteins and/or (C-O) and (C-O-C) groups of carbohydrates; P =0.06; 95% CI = -0.00117 to -0.00002) and 1034 cm^{-1} (collagen; P<0.001; 95% CI = -0.00267 to -0.00146) (Fig 4) (Supplementary Items 2 and 3)[12].

Discussion

In this study we accurately differentiated between horses with and without OA using serum biomarkers. The spectral regions that were responsible for the segregation were mainly representative of lipids and proteins. A general increase of absorbance was noticed for the majority of the peaks in the OA group, including lipid and protein bands (1474 cm^{-1}, 1304 cm^{-1}) as well as the collagen region (1034 cm^{-1}) (Fig 4). The increase in lipid and protein levels may be attributed to the accumulation of lipids and total fatty acids (such as hypercholesterolemia), and increased protein expression levels, which have been previously reported in OA [13,14]. Other studies have shown that inflammatory mediators are increased in OA in comparison to healthy subjects, which can be reflected outside the joint tissues in blood plasma and peripheral blood leukocytes [15-17]. Similar to a number of preceding studies demonstrating increased levels of collagen in OA subjects after analysing cartilage or blood samples, our results also showed a statistically significant increase of the collagen band (1034 cm^{-1}) in OA (Fig 4) [18-22]. For instance, the observed increase may be explained by the overexpression of collagen type X, which has been previously found elevated in serum samples of OA patients and has been associated with cartilage degradation and inflammation [18]. It is also worth noting that Thoroughbred racehorses tend to be young and collagen expression can vary according to maturity. In contrast, one peak was decreased in horses with OA, namely the Amide II band (1539 cm^{-1}). Two peaks (1400 cm^{-1} and 1161 cm^{-1}) were not significantly different even though they were picked out by the peak selection algorithm in IRootLab^{c}.

Although PCA-LDA has been previously reported as a useful technique to detect canine OA [11], this model achieved a low diagnostic accuracy (80% sensitivity, 20% specificity) in the current study, indicating a poor classification performance for this dataset. However, after classification of the two populations with PCA-QDA, the diagnostic accuracy

was exceptionally high, reaching 100% sensitivity and specificity. This could potentially lead to a clinical test that would be beneficial for the diagnosis and management for equine OA. Blood sampling is minimally invasive thus facilitating clinical monitoring and assessment of OA progression.

In contrast to traditional molecular approaches, spectroscopic techniques have a great advantage in their low cost. Biospectroscopy generates a spectrum indicative of a range of biological molecules (proteins, lipids, carbohydrates and nucleic acids) thus generating something similar to a multi-marker assay, allowing for more robust results than isolated biomarkers. A number of previous studies have employed vibrational spectroscopy using synovial fluid or tissue samples to investigate OA [23-26]. Specifically, transmission IR and Raman spectroscopies have been successfully employed to interrogate synovial fluid for the diagnosis of OA in horses as well as in humans [23,25,26]. A further study used Raman spectroscopy to examine the molecular changes accompanying OA, using the articular surface area in the knees of rats [24].

A limitation of the current study was the lack of pre-symptomatic OA cases which should be addressed in future studies to determine whether spectroscopy could be utilised as an early diagnostic tool. A subgroup analysis showing any potential connection between the severity of OA and the spectral peaks would be of definite value, however, outside the scope of this preliminary study. Our study was also limited by the wide range of ages and breeds in the control group. Larger numbers of animals would be needed for more detailed exploration and would also allow to control for differences in the breed or age.

Conclusions

The current study presents novel findings on OA spectroscopic detection using serum. Along with the high sensitivity and specificity with which the condition was diagnosed, spectral markers of joint degenerative changes were also detected. Blood spectroscopy is a tool with great promise in the field of OA research and potentially clinical translation.

Authors' declaration of interests

No competing interests have been declared.

Ethical animal research

Samples were collected under the regulations of the Hong Kong Jockey Club (HKJC), under the University of Central Lancashire (UCLan) Home Office licence and with Institutional Review Board approval.

Owner informed consent

Horse owners gave informed consent.

Source of funding

Manufacturers' addresses

[a]Bruker Tensor

[b]Helios

[c]IRootLab http://trevisanj.github.io/irootlab/

[d]GraphPad Software, La Jolla, California, USA.

Figure Legends

Fig 1: (A) Raw spectra (solid line) with s.d. (coloured area); **(B)** Pre-processed spectra (solid line) with s.d. (coloured area). HC = healthy controls and OA = osteoarthritis cases.

Fig 2: ATR-FTIR spectroscopic analysis. **(A)** Multi-dimensional PCA scores plot (PC1 x PC2 x PC3). **(B)** PCA loadings along PC1 with the top six discriminating peaks accounting for the segregation between osteoarthritis cases (OA) and controls; the dotted-line spectrum at the background indicates the mean spectra before PCA.

Fig 3: PCA-QDA. **(A)** Normalised discriminant function (DF) for training and test samples;

(B) Receiver Operating Characteristic (ROC) curve for the test set.

HC = healthy controls; OA = osteoarthritis cases; AUC = area under the curve.

Fig 4: The absorbance differences of the top six discriminatory wavenumbers, along with the *P*-values after statistical analysis (Mann Whitney U test): 1539 cm^{-1} (bending vibration of (N-H) of proteins, Amide II band), 1474 cm^{-1} (bending vibration of (CH_2) of proteins/lipids), 1400 cm^{-1} (stretching vibration of (COO^-) of amino acids and/or bending vibration of (CH_3) of proteins), 1304 cm^{-1} (Amide III band, components of proteins), 1161 cm^{-1} (stretching vibration of C-OH groups of proteins and/or (C-O) and (C-O-C) groups of carbohydrates) and 1034 cm^{-1} (collagen). Blue and red dots are expressed as real data points. Mean ± s.d. are expressed as horizontal bars. A *P*-value ≤0.05 was considered significant. OA = osteoarthritis cases.

References

[1] McIlwraith, C.W., Kawcak, C.E., Frisbie, D.D., Little, C.B., Clegg, P.D., Peffers, M.J., Karsdal, M.A., Ekman, S., Laverty, S. and Slayden, R.A. (2017) Biomarkers for equine joint injury and osteoarthritis. *J. Orth. Res.* **36**, 823-831.

[2] Frisbie, D., Al-Sobayil, F., Billinghurst, R., Kawcak, C. and McIlwraith, C. (2008) Changes in synovial fluid and serum biomarkers with exercise and early osteoarthritis in horses. *Osteoarthr. Cartil.* **16**, 1196-1204.

[3] Orlando Júnior, N., Leão, M.G.d.S. and Oliveira, N.H.C.d. (2015) Diagnosis of knee injuries: comparison of the physical examination and magnetic resonance imaging with the findings from arthroscopy. *Rev. Bras. Ortop.* **50**, 712-719.

[4] Baker, M.J., Hussain, S.R., Lovergne, L., Untereiner, V., Hughes, C., Lukaszewski, R.A., Thiefin, G. and Sockalingum, G.D. (2016) Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.* **45**, 1803-1818.

[5] Martin, F.L., Kelly, J.G., Llabjani, V., Martin-Hirsch, P.L., Patel, I.I., Trevisan, J., Fullwood, N.J. and Walsh, M.J. (2010) Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat. Protoc.* **5**, 1748-1760.

[6] Bro, R. and Smilde, A.K. (2014) Principal component analysis. *Anal. Methods* **6**, 2812-2831.

[7] Dixon, S.J. and Brereton, R.G. (2009) Comparison of performance of five common classifiers represented as boundary methods: Euclidean Distance to Centroids, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Learning Vector Quantization and Support Vector Machines, as dependent on data structure. *Chemom. Intellig. Lab. Sys.* **95**, 1-17.

[8] Morais, C.L. and Lima, K.M. (2018) Principal Component Analysis with Linear and Quadratic Discriminant Analysis for Identification of Cancer Samples Based on Mass Spectrometry. *J. Braz. Chem. Soc.* **29**, 472-481.

[9] Kennard, R.W. and Stone, L.A. (1969) Computer Aided Design of Experiments. *Technometrics* **11**, 137-148.

[10] Morais, C.L. and Lima, K.M. (2017) Comparing unfolded and two-dimensional discriminant analysis and support vector machines for classification of EEM data. *Chemom. Intell. Lab. Syst.* **170**, 1-12.

[11] Hou, S. (2016) Development of diagnostic models for canine osteoarthritis based on serum and joint fluid mid-infrared spectral data using five different discrimination and classification methods. *J. Chemom.* **30**, 663-681.

[12] Movasaghi, Z., Rehman, S. and ur Rehman, D.I. (2008) Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* **43**, 134-179.

[13] Villalvilla, A., Gómez, R., Largo, R. and Herrero-Beaumont, G. (2013) Lipid transport and metabolism in healthy and osteoarthritic cartilage. *Int. J. Mol. Sci.* **14**, 20793-20808.

[14] Kostopoulou, F., Gkretsi, V., Malizos, K.N., Iliopoulos, D., Oikonomou, P., Poultsides, L. and Tsezou, A. (2012) Central role of SREBP-2 in the pathogenesis of osteoarthritis. *PloS One* **7**, e35753.

[15] Berenbaum, F. (2013) Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). *Osteoarthr. Cartil.* **21**, 16-21.

[16] Sokolove, J. and Lepus, C.M. (2013) Role of inflammation in the pathogenesis of osteoarthritis: latest findings and interpretations. *Ther. Adv. Musculoskelet. Dis.* **5**, 77-94.

[17] Daghestani, H.N. and Kraus, V.B. (2015) Inflammatory biomarkers in osteoarthritis. *Osteoarthr. Cartil.* **23**, 1890-1896.

[18] He, Y., Siebuhr, A.S., Brandt-Hansen, N.U., Wang, J., Su, D., Zheng, Q., Simonsen, O., Petersen, K.K., Arendt-Nielsen, L. and Eskehave, T. (2014) Type X collagen levels are elevated in serum from human osteoarthritis patients and associated with biomarkers of cartilage degradation and inflammation. *BMC Musculoskel. Disord.* **15**, 309.

[19] Kumar, R., Grønhaug, K.M., Afseth, N.K., Isaksen, V., de Lange Davies, C., Drogset, J.O. and Lilledahl, M.B. (2015) Optical investigation of osteoarthritic human cartilage (ICRS grade) by confocal Raman spectroscopy: a pilot study. *Anal. Bioanal. Chem.* **407**, 8067-8077.

[20] Rieppo, L., Töyräs, J. and Saarakkala, S. (2017) Vibrational spectroscopy of articular cartilage. *Appl. Spectrosc. Rev.* **52**, 249-266.

[21] Grimmer, C., Balbus, N., Lang, U., Aigner, T., Cramer, T., Müller, L., Swoboda, B. and Pfander, D. (2006) Regulation of type II collagen synthesis during osteoarthritis by prolyl-4-hydroxylases: possible influence of low oxygen levels. *Am. J. Pathol.* **169**, 491-502.

[22] Von der Mark, K., Kirsch, T., Nerlich, A., Kuss, A., Weseloh, G., Glückert, K. and Stöss, H. (1992) Type X collagen synthesis in human osteoarthritic cartilage. Indication of chondrocyte hypertrophy. *Arthritis Rheum.* **35**, 806-811.

[23] Esmonde-White, K.A., Mandair, G.S., Raaii, F., Jacobson, J.A., Miller, B.S., Urquhart, A.G., Roessler, B.J. and Morris, M.D. (2009) Raman spectroscopy of synovial fluid as a tool for diagnosing osteoarthritis. *J. Biomed. Opt.* **14**, 034013-034013-034018.

[24] de Souza, R.A., Xavier, M., Mangueira, N.M., Santos, A.P., Pinheiro, A.L.B., Villaverde, A.B. and Silveira, L. (2014) Raman spectroscopy detection of molecular changes associated with two experimental models of osteoarthritis in rats. *Lasers Med. Sci.* **29**, 797-804.

[25]   Vijarnsorn, M., Riley, C.B., Shaw, R.A., McIlwraith, C.W., Ryan, D.A., Rose, P.L. and Spangler,
          E. (2006) Use of infrared spectroscopy for diagnosis of traumatic arthritis in horses. *Am. J.
          Vet. Res.* **67**, 1286-1292.

[26]   Eysel, H., Jackson, M., Nikulin, A., Somorjai, R., Thomson, G. and Mantsch, H. (1997) A novel
          diagnostic test for arthritis: multivariate analysis of infrared spectra of synovial fluid.
          Biospectroscopy **3**, 161-167

Supporting Information

Supplementary Item 1: (A) Mean raw spectra; (B) mean pre-processed spectra.

Supplementary Item 2: PCA scores for the 10 first principal components (PCs).

Supplementary Item 3: Absorbance value at the top six discriminatory wavenumbers of the average
pre-processed spectrum per sample.

A

B

A

B