SCIENTIFIC REP🅞RTS

Imaging of Osteoarthritic Human Articular Cartilage using Fourier Transform Infrared Microspectroscopy Combined with Multivariate and Univariate Analysis

J. Oinas[1,2], L. Rieppo[1,2], M. A. J. Finnilä[1,2,3], M. Valkealahti[2,4], P. Lehenkari[2,4,5] & S. Saarakkala[1,2,6]

The changes in chemical composition of human articular cartilage (AC) caused by osteoarthritis (OA) were investigated using Fourier transform infrared microspectroscopy (FTIR-MS). We demonstrate the sensitivity of FTIR-MS for monitoring compositional changes that occur with OA progression. Twenty-eight AC samples from tibial plateaus were imaged with FTIR-MS. Hyperspectral images of all samples were combined for K-means clustering. Partial least squares regression (PLSR) analysis was used to compare the spectra with the OARSI grade (histopathological grading of OA). Furthermore, the amide I and the carbohydrate regions were used to estimate collagen and proteoglycan contents, respectively. Spectral peak at 1338 cm^{-1} was used to estimate the integrity of the collagen network. The layered structure of AC was revealed using the carbohydrate region for clustering. Statistically significant correlation was observed between the OARSI grade and the collagen integrity in the superficial ($r = -0.55$) and the deep ($r = -0.41$) zones. Furthermore, PLSR models predicted the OARSI grade from the superficial ($r = 0.94$) and the deep ($r = 0.77$) regions of the AC with high accuracy. Obtained results suggest that quantitative and qualitative changes occur in the AC composition during OA progression, and these can be monitored by the use of FTIR-MS.

Articular cartilage (AC) is a specialized connective tissue that covers the ends of long bones. AC redistributes the loads directed on bones and, together with synovial fluid, provides almost frictionless articulation between the bones[1,2]. The main constituents of AC are type II collagen, proteoglycans (PGs) and interstitial water[1]. Degradation of the extracellular matrix (ECM) of AC is a hallmark of the initiation of osteoarthritis (OA)[3]. In the early stages of OA, depletion of PGs, especially in the superficial layer, and alterations in the orientation of the collagen fibril network have been reported[4–7]. Wear and degradation of the entire AC tissue and the possible formation of reparative tissue including fibrocartilage are typical signs of late stage OA[8–10].

Different non-invasive or minimally invasive imaging methods, such as magnetic resonance imaging (MRI) and computed tomography (CT), have been used clinically to investigate the degenerative stage of AC in OA[11–13]. Typically, only the morphology of AC, *i.e.*, wear of the tissue, can be detected with routine MRI and CT protocols in clinical practice. There is an increased interest in developing new imaging protocols to evaluate compositional changes in AC, which precede morphological changes. For instance, it is possible to detect the depletion of PG

[1]Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulu, Finland. [2]Medical Research Center, University of Oulu and Oulu University Hospital, Finland. [3]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland. [4]Department of Surgery, Oulu University Hospital, Finland. [5]Research Group of Cancer and Translational Medicine, Faculty of Medicine, University of Oulu, Finland. [6]Department of Diagnostic Radiology, Oulu University Hospital, Finland. Correspondence and requests for materials should be addressed to J.O. (email: joonas.oinas@oulu.fi)

and alterations in the collagen network organization with quantitative MRI using $T_{1\rho}$ weighting and T_2 mapping[14,15]. Contrast-enhanced CT can also be used to detect PG content in AC[16]. However, these non-invasive or minimally invasive methods that are capable of evaluating tissue-level changes in AC are still in the developmental stage.

When considering very early OA compositional changes in AC, neither MRI nor CT provides adequate sensitivity. Thus, more direct approaches that are applicable during arthroscopic examination have been introduced for OA diagnostics. These include, *e.g.*, optical coherence tomography (OCT), ultrasound (US) imaging[17] and near infrared (NIR) spectroscopy, all of which are still in the investigational stage and are not in routine clinical use[18,19]. Spectroscopic methods are especially promising, as they allow the direct assessment of the composition of hyaline cartilage[20], thus, providing the possibility to detect early OA changes[21]. However, NIR spectroscopy does not provide similar high-level molecular sensitivity that is achievable using the mid-infrared region[22]. High molecular sensitivity is an essential requirement not only to improve OA diagnostics but also to provide further understanding of the biochemical changes during the development and progression of OA.

Fourier Transform Infrared (FTIR) spectroscopy is a convenient method for investigating the macromolecular composition of biological samples[23,24]. FTIR spectroscopy is based on the absorption of infrared light by molecules at characteristic frequencies. FTIR microspectroscopy (FTIR-MS), which enables chemical imaging, is performed by combining a FTIR spectrometer with a conventional light microscope. Consequently, every pixel in an acquired image contains an FTIR absorption spectrum. FTIR images can be analyzed pixel-by-pixel to resolve the spatial distribution and structure of multiple biochemical components, *e.g.*, collagen and PGs in the ECM, from an unstained histological section within a single imaging session[25–27].

Univariate spectral analysis methods form the basis for analysis of FTIR-MS data[27]. Camacho *et al.*[28] demonstrated in their pioneering work that collagen and PG contents in AC could be estimated using univariate analysis[28]. Furthermore, the ratio of the integrated areas under CH_2 side-chains (1300–1360 cm^{-1}) and the amide II (1485–1585 cm^{-1}) regions was shown to decrease when the collagen network is degraded. The ratio parameter, which is considered a measure of collagen integrity, was later shown to decrease as the OA grade increases[4,27,29]. It is notable that there are no published studies in which these univariate parameters have been quantitatively correlated with the histopathological OA grade in human samples.

Univariate analysis methods are simple to use, but they discard vast amounts of spectral information. More sophisticated multivariate analysis methods, which derive information from the entire spectrum, could provide more detailed information about the biochemical composition of AC. Partial least squares regression (PLSR) is a multivariate regression method that constructs new variables to explain the co-variance between the reference data and the spectra[30]. PLSR has been used to predict the biomechanical properties[31] and composition[32,33] of AC from FTIR-MS data, and the status of OA in human and rabbit AC from infrared fiber optic probe data[34–36]. However, unlike FTIR-MS, infrared fiber optic probe measurements are limited to the superficial cartilage. To the best of our knowledge, PLSR has not been used to predict the histopathological state (*i.e.*, OARSI grade[10]) of human AC tissue from FTIR-MS data.

Cluster analysis belongs to another class of multivariate analysis methods. Clustering algorithms classify spectra into groups by minimizing differences within groups and maximizing differences between groups. This classification can be used for segmentation of FTIR images[37–43]. It is well known that AC has a layered structure and it is typically divided into three zones from the articulating surface to subchondral bone based on the orientation of collagen fibrils and the biochemical composition of the ECM[2,44,45]. Earlier, clustering algorithms have been used to reveal zonal structure in healthy bovine and rabbit AC[40]. Furthermore, cluster analysis has also been used to differentiate intact and repaired AC from patellar grooves in rabbit femurs[41]. However, cluster analysis has not yet been used to reveal zonal structure in human AC.

The first aim of the present study was to investigate the quantitative and qualitative changes in the FTIR spectra of human AC as a function of the histological OA grade (OARSI grading). We hypothesized that the zonal structure of the AC and changes in the macromolecular structures during OA progression could be detected when cluster analysis is conducted simultaneously on the spectra of all samples. We expected that particular clusters would concentrate on samples with a certain OARSI grade. The second aim of the study was to construct multivariate PLSR models to predict the histopathological OARSI grade of the human AC tissue. We hypothesized that a more accurate prediction would be obtained using the spectral information from the surface layer than from the deep layer of the AC. In addition, we hypothesized, as suggested by earlier literature, that negative correlations between the OARSI grade and every univariate parameter (amide I, carbohydrate region, collagen integrity and second derivative peaks 1202 cm^{-1} and 1064 cm^{-1}) would be observed[4,21,29,34].

Results

Cluster analysis. After careful preliminary analyses of the large data matrix (including all the samples), five clusters were found to be optimal for revealing the layered structure of AC. Different spectral regions were evaluated by clustering to expose the layered structure of AC tissue and to explore the changes in the resulting cluster images during OA progression. The layered structure of AC in each sample was observed in the cluster images using the carbohydrate region for clustering (Fig. 1A). The lowest intensity was observed in the average raw spectra of the green cluster in the carbohydrate region, which indicates the lowest relative PG content (Fig. 1B). Major spectral differences between the clusters were observed in the spectral region from 1058 cm^{-1} to 1080 cm^{-1} in the second derivative spectra when the carbohydrate region was used for cluster analysis (Fig. 1C). In Fig. 2, Safranin O-stained sections, cluster images that were obtained using the carbohydrate region and PLM images of three representative samples with different OA grades are presented (Fig. 2). The corresponding images for all 28 samples can be found in the Supplementary Material. Similar features can be observed in the images. For example, the superficial, the middle and the deep layers of the AC are visible in PLM and cluster images. Furthermore, the green cluster in the cluster images complies with the Safranin-O image. On the surface of the Safranin-O images,
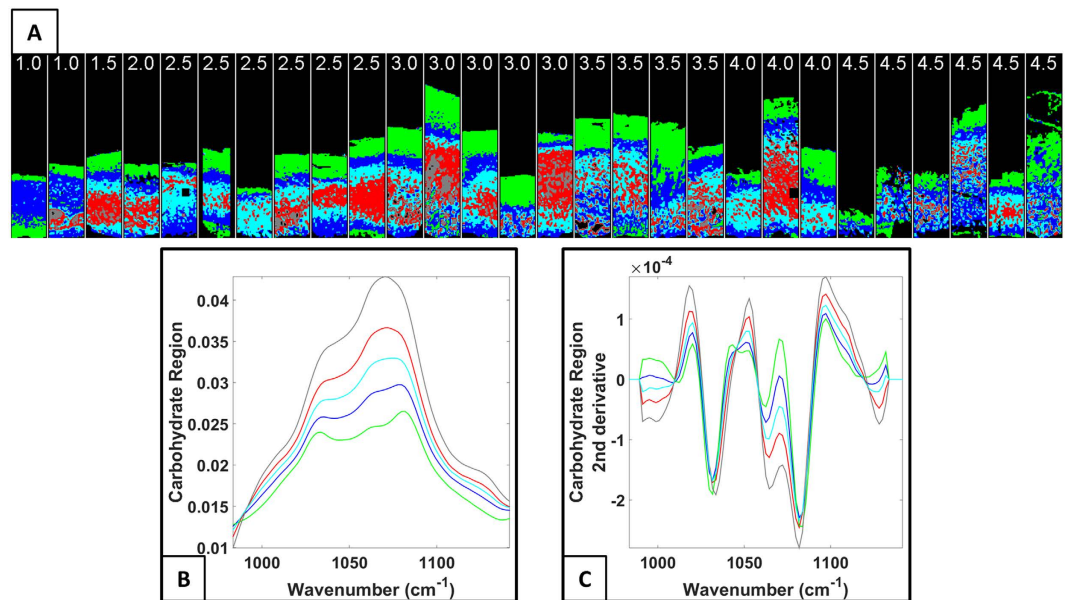
Figure 1. (**A**) Cluster images obtained using the carbohydrate region for clustering. The characteristic layered structure of AC can be observed in the cluster images. The OARSI grade is indicated above each sample. (**B**) Average spectra of each cluster. (**C**) The second derivative spectra of each cluster. Major differences between the second derivative spectra can be seen around the peak at $1060\,\mathrm{cm}^{-1}$. The colors of the spectra correspond to the cluster colors.

the amount of staining is low. Correspondingly, the green cluster in the cluster images is located in the superficial layer of the AC (Fig. 2). A visual correlation between the OARSI grade and the cluster image obtained from the carbohydrate region was not observed. Similarly, when using the extended carbohydrate region, the combination of the amide I and amide II regions, or the combination of the amide I, amide II and carbohydrate regions, there was no clear correspondence between the cluster images and the OARSI grade.

Partial least squares regression. PLSR predicted the OARSI grade more accurately from the second derivative spectra (Fig. 3A,B) ($r_{Surface} = 0.94$, $p < 0.0001$, $r_{Deep} = 0.77$, $p < 0.001$) than from the raw spectra (Fig. 3C,D) ($r_{Surface} = 0.78$, $p < 0.0001$, $r_{Deep} = 0.66$, $p < 0.001$) in both the surface and the deep layers of AC (Fig. 3). The percent errors for the second derivative spectrum models were also lower ($E_{surface} = 11.51\%$, $E_{deep} = 17.66\%$) than for the raw spectrum models ($E_{surface} = 18.30\%$, $E_{deep} = 21.25\%$). Overall, the OARSI grade was more accurately predicted from the surface layer than from the deep layer of AC. The number of latent variables (LVs) varied from two to six between the models. With the spectra from the surface, six and five LVs were used in the models built from the second derivative and the raw spectra, respectively. Correspondingly, five and two LVs were used in the models in the deep layer. The CARS algorithm selected eight optimal variables (wavenumbers) for the PLSR models in both the surface and the deep layer when the raw spectra were used. When the second derivative spectra were used, 21 and 20 variables were selected to the models for the superficial and the deep layer, respectively. The variables are marked with black dots in the spectra of corresponding model (Fig. 3A–D).

Univariate analysis. Statistically significant correlations were not observed between the OARSI grade and the amide I region, the carbohydrate region, the carbohydrate region normalized with the amide I region or the second derivative peak $1064\,\mathrm{cm}^{-1}$ (PG) in either of the analyzed zones (surface and deep). However, statistically significant negative correlations were observed between the OARSI grade and the collagen integrity parameter in both the surface and the deep zones (surface: $r = -0.55$, $p < 0.003$, deep: $r = -0.41$, $p = 0.03$) (Fig. 4A,B). A statistically significant negative correlation ($r = -0.43$, $p = 0.02$) was also found between the OARSI grade and the second derivative peak at $1202\,\mathrm{cm}^{-1}$ (collagen) in the surface layer of AC (Fig. 4C).

Statistically significant differences were found between the OA groups (group 1 = early OA, group 2 = intermediate OA and group 3 = severe OA) in all univariate parameters (amide I region, carbohydrate region, collagen integrity and the second derivative peak $1202\,\mathrm{cm}^{-1}$) (Fig. 5). In the amide I and carbohydrate regions, statistically significant differences were observed between intermediate OA and severe OA in the surface and middle layers (Fig. 5A,B). Collagen integrity was significantly different between early OA and severe OA in the surface layer and between early OA and intermediate OA in the deep layer (Fig. 5C). The peak at $1202\,\mathrm{cm}^{-1}$ exhibited significant differences between intermediate OA and severe OA in the surface layer and between early OA and severe OA in the middle layer (Fig. 5D).

Discussion

In the present study, human AC samples with various histological grades of OA were studied using FTIR-MS. The aims of the study were to detect quantitative and qualitative osteoarthritic changes in the FTIR spectra of AC. We
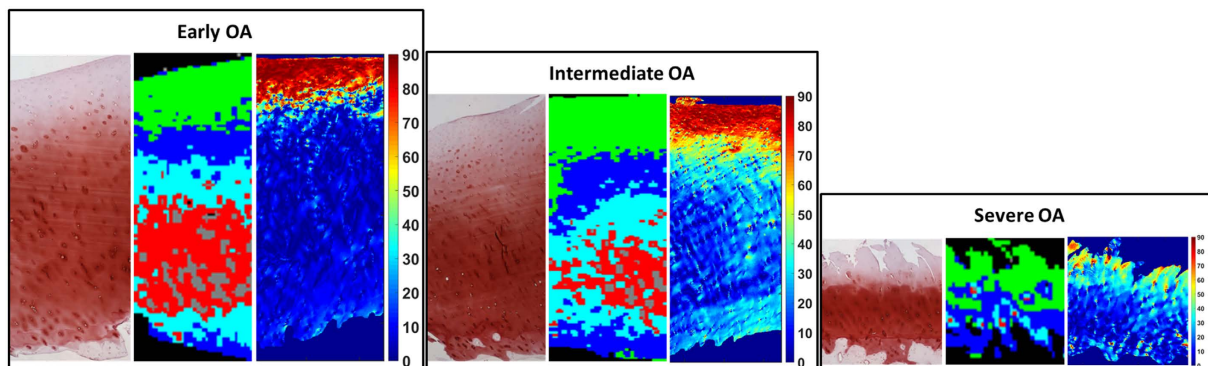
Figure 2. Safranin O, cluster and PLM images from three samples with different severity of OA (Early OA = OARSI grade 1.5, Intermediate OA = OARSI grade 3.5 and Severe OA = OARSI grade 4.5). The corresponding cluster images from the carbohydrate region are extracted from the combined hyperspectral data. Black pixels in the cluster image are due to the removal of poor quality spectra.
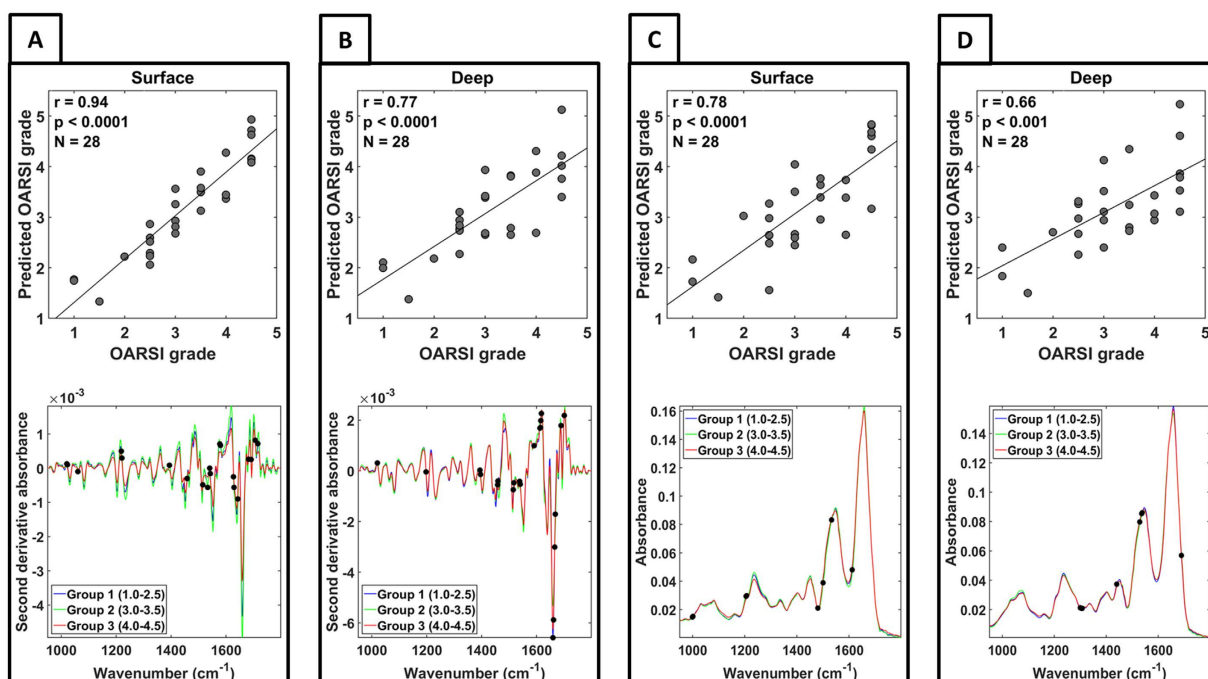

Figure 3. Scatter plots between the reference values and the predicted values of the OARSI grade and the average spectra of each pooled OA group. The black dots in the spectra indicate the wavenumbers that were selected to the PLSR models by the CARS algorithm. (**A,B**) the plots from the models that used second derivative spectra. (**C,D**) the plots from the model that used raw vector normalized spectra.

hypothesized that the layered structure of AC could be observed from the cluster analysis images and that the particular cluster will concentrate on a certain OARSI grade. The results obtained indicate that the structural and compositional changes caused by OA can be observed with both the uni- and multivariate analyses of spectral data. The present study is the first to conduct cluster analysis on FTIR images of multiple human AC samples simultaneously, *i.e.*, by combining hyperspectral data from several individual samples into one large data matrix. This is advantageous because it allows direct comparison of the clusters between the samples. Unfortunately, compared to our original hypothesis, cluster analysis did not reveal any OA-related clusters. However, the PLSR models demonstrated that OA-related changes occur in both the surface and the deep layers of AC.

The layered structure of AC in cluster images is similar to the structure observed in PLM orientation images. Based on the visual evaluation of cluster images from all the spectral regions, the best correspondence with PLM images was observed when the carbohydrate region was used. It can be assumed that the relative amounts of collagen and PGs are the main differences between the clusters in the carbohydrate region. For example, the green cluster is found in the superficial layer of AC in the cluster images (Fig. 1A). The AC surface contains low amounts of PGs. Furthermore, the mean spectra of clusters (Fig. 1B) demonstrate that the overall absorbance in the carbohydrate region is at its lowest in the green cluster. Therefore, it can be assumed that the green cluster indicated the
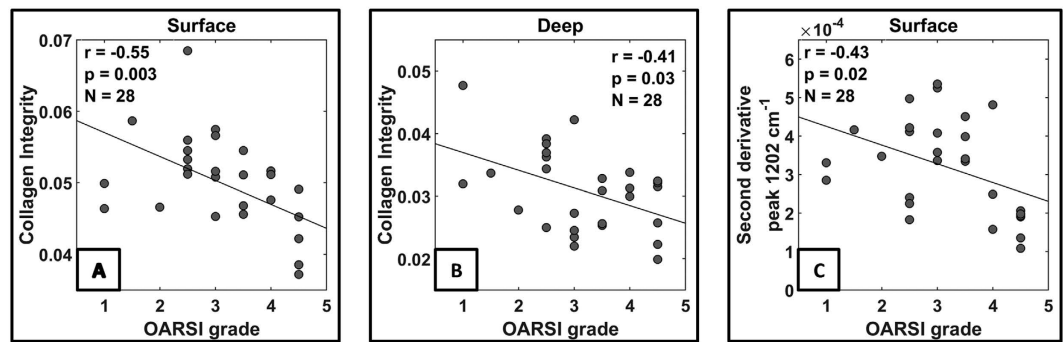
Figure 4. Statistically significant negative correlations were observed between the OARSI grade and the collagen integrity parameter in both the surface (**A**) and the deep layer (**B**), and the second derivative peak 1202 cm^{-1} in the superficial layer (**C**).

regions with the lowest relative amounts of PGs in the sample sections. In contrast, by examining the overall absorbance of the mean spectra in the carbohydrate region (Fig. 1B), the relative PG content is at its highest in the grey cluster that is found in the deep zone. The other three clusters lie between the grey and green clusters in terms of their relative PG content (Fig. 1B). In the second derivative spectra, the main differences are observed in the spectral region around the peak at 1060 cm^{-1} (Fig. 1C). The vibrations from this region have been shown to arise from the C–O stretching vibrations of carbohydrate residues in collagens and in PGs and from SO$_3^-$ stretching vibration in sulfated glycosaminoglycans[46–48]. Because Safranin-O is a cationic dye, it binds to the negatively charged PGs, and therefore, it can be used for observing the distribution of PGs[49]. When the cluster image is compared with the Safranin-O stained sections, the green cluster can be observed to concentrate in the regions where the amount of the Safranin-O stain is low. This supports the idea that the clustering result follows the relative amount of the PGs.

Because the layered structure of AC is traditionally defined by the organization of the collagen network, the amide region was presumed to reveal the layered structure more clearly than the other spectral regions that were investigated. The characteristic layered structure of healthy rabbit and bovine AC could be revealed with clustering algorithms using the amide I, amide II or carbohydrate region[40]. In earlier studies, the clustering results followed the concentration of AC constituents and the organization of the collagen network. In our study with human AC, the layered structure was not evident in all samples when using the amide I or amide II regions alone. One factor that affects the amide I and amide II regions is the traces of water vapor in the measurement chamber. Water vapor displays several sharp bands that overlap the amide I and amide II regions[51]. A second factor that may explain the difference is the different species investigated (rabbit vs. human). Finally, in the study by Kobrina et al.[40], the clustering was performed on one sample at a time, whereas in this study clustering was performed simultaneously on all samples.

Compared to our our original hypothesis, cluster analysis did not detect a cluster that is directly related to osteoarthritic changes in AC. K-means clustering is an unsupervised algorithm, *i.e.*, it was not specifically taught to find osteoarthritic changes, but it groups the spectra according to their similarities based on Euclidean distance. The clusters that are detected seem to be related to variations in the relative PG and collagen contents in the AC. It is possible that clusters associated with OA would be found if the number of clusters would be significantly increased. However, the interpretation of all clusters would become more difficult as their number increases; therefore, based on our preliminary analyses it was not seen feasible in this study.

Although our clustering results did not correlate strongly with the OARSI grade, the PLSR analysis demonstrated that the OARSI grade can be predicted from the FTIR spectra. It is often suggested that the first signs of OA in AC are observed in the superficial layer[5,6]. A competing view suggests that OA could be initiated in the subchondral bone, as during OA changes also occur in the subchondral bone[52]. In our study, the OARSI grade was predicted more accurately from the superficial layer than the deep layer (as indicated by both the correlation and the percentage error analyses). This supports the idea that the most significant biochemical changes occurred in the surface of the AC in this relatively heterogeneous sample set. The OARSI grading system does not directly take into account the biochemical changes that occur in AC. However, as the PLSR satisfactorily predicted the OARSI grade from the deep AC layer, biochemical changes may also occur in the deep layers of AC during OA progression. Furthermore, as expected, the results of the second derivative spectra, which enhance the separation of the overlapping peaks[53], resulted in a more accurate prediction of the OARSI grade.

With regard to univariate analysis of the FTIR spectra, no significant correlations were observed between the OARSI grade and the integrated absorbances of the amide I and carbohydrate regions, even though an earlier study suggested that quantitative correlations may exist[54]. However, a significant correlation was observed in the superficial layer between the OARSI grade and the second derivative peak at 1202 cm^{-1}, which can be used to monitor the collagen content[53]. The difference in the results between the amide I and the second derivative peak might be due to the better separation of overlapping peaks obtained using second derivative spectroscopy[53]. However, the lack of a correlation between the collagen content and the OARSI grade would not be that surprising, as some earlier studies did not detect a difference in the collagen content between healthy and OA human AC[55]. This suggests that the biochemical changes in AC are not linearly related to the OARSI grade. Group comparisons revealed that changes are observed primarily in the first 40% of the thickness from the cartilage surface:
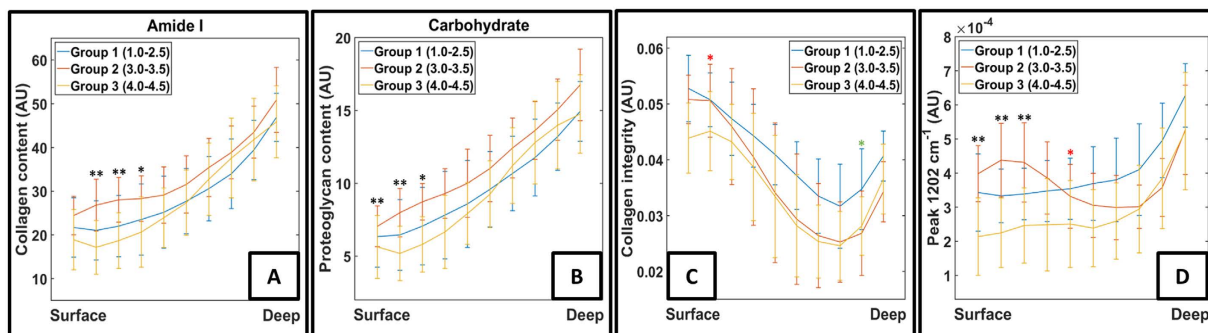

addition, subgrades of 0.5 were also used in the grading. Samples with an OARSI grade greater than 4 were not used in this study because the AC is already completely worn off at grade 5.

Fourier Transform Infrared Microspectroscopy (FTIR-MS). FTIR-MS measurements were conducted using an FTIR spectrometer (Tensor 27, Bruker Inc., Billerica, MA, USA) coupled with a Bruker Hyperion 3000 microscope (Bruker Inc., Billerica, MA, USA) equipped with a focal plane array (FPA) 64×64 detector. The system was controlled with the manufacturer's Opus 7.2 software. First, panoramic images were collected in visual imaging mode using the same microscope. From the visual images, $500\,\mu m$ wide areas from the cartilage surface to subchondral bone ($500–2000\,\mu m$) were selected for FTIR-MS. All measurements were conducted in transmission mode, where incident infrared light is beamed through the sample section, and an FTIR absorption spectrum is obtained for every pixel. The spectral region from $900\,cm^{-1}$ to $3800\,cm^{-1}$ (corresponding to wavelengths from $2.6\,\mu m$ to $11.1\,\mu m$) was collected. The spectral resolution was set to $4\,cm^{-1}$, and each spectrum was averaged 16 times. Spatial resolution was set to $21.6\,\mu m$ using pixel binning. The sample chamber was purged with dry air to stabilize measurement conditions.

Data preprocessing. All preprocessing and analysis of spectral data were performed with custom-made MATLAB (MathWorks Inc., MA, USA) scripts and the commercial MATLAB toolbox Cytospec 2.00.01 (Berlin, Germany). A Resonant Mie Scattering Correction (RMieSC) algorithm was used to remove the scattering effects from the spectra[60]. Predefined thresholds for peak intensity and signal-to-noise ratios (SNR) were used as quality criteria, and poor quality spectra were removed from the dataset using the quality test functions in Cytospec. Principal component analysis (PCA) was used to detect the cartilage-bone interface. PCA is a method that creates a new set of linearly uncorrelated variables named principal components (PCs)[61]. PCs are classified so that the first PC explains most of the variance in the data and the second PC the second most and so on. PCA was performed using the carbohydrate region of the data matrix that included all the samples. A gray scale image for discriminating the bone from the AC was created using score values of the first PC. The score image was visually inspected and the bone spectra were manually removed from the dataset. After the removal of bone spectra, the spectra were truncated to a fingerprint region ($950–1800\,cm^{-1}$) that contains detailed information about proteins and carbohydrates. When calculating the collagen integrity parameter, a linear two-point baseline correction was performed on the CH_2 side-chain and the amide II regions individually. Furthermore, for the cluster analysis, the spectra were vector normalized after RMieSC. Thus, the differences in the amounts of material, *i.e.*, the total absorbance, are removed, and clustering is based only on qualitative properties, *i.e.*, the shape of the sample's spectra. The second derivatives spectra for the univariate and PLSR analyses were calculated using the Savitzky-Golay algorithm with nine smoothing points[62].

Cluster analysis. Cluster analysis methods are unsupervised, *i.e.*, no *a priori* information about class memberships is given. The K-means clustering (KMC) algorithm was used in this study. The basic mathematical principle of the KMC algorithm has been presented by MacQueen[63]. The KMC algorithm is a "coarse" method for classifying the spectra because a single spectrum can be a member of only one class (cluster). The algorithm classifies the spectra so that the spectral differences within clusters are minimized and the spectral differences between clusters are maximized. The Euclidean distance is used to define the differences between the spectra[38]. First, the algorithm selects a predefined number (K) of random spectra from the data set as the centroid spectra. Then, the algorithm calculates the distance between every spectrum to each centroid spectrum and classifies the spectra in K classes according to the distances. New centroid spectra are created by calculating the average spectrum for each class. Again, the distances between each spectrum to each centroid spectrum are calculated and new K classes are created. In this study, the number of these iterations was set to 20, which was visually observed to be an adequate number (no notable differences were observed in the resulting cluster images when the number of iterations was increased). As a result, the algorithm generates an image in which one color represents the particular class (cluster)[38,39]. Black pixels in the cluster images are due to the removal of poor quality spectra.

The 28 individual hyperspectral data images that were measured were combined into one large data matrix for cluster analysis. Different spectral regions can be used in KMC classifications. In this study, the following spectral regions were evaluated for cluster analysis: the carbohydrate region ($985–1140\,cm^{-1}$), the extended carbohydrate region ($950–1400\,cm^{-1}$), the amide I region ($1585–1720\,cm^{-1}$), the amide II region ($1485–1585\,cm^{-1}$), and the combination of the carbohydrate, amide I and amide II regions. Spectral differences between the clusters were investigated by observing the average spectra of each cluster. In addition, second derivative spectra of the cluster averages were also calculated.

First, the zonal structure of AC in the clustering images that were obtained was visually compared with the collagen fibril orientation images obtained using polarized light microscopy (PLM). Second, the obtained cluster images were compared with the Safranin-O stained sections to obtain information about the PG contents of different clusters.

Partial least squares regression (PLSR). PLSR is a powerful multivariate analysis method, which is widely used in chemometric analysis. The spectral peaks from different molecular vibrations tend to overlap in FTIR spectra of biological tissues. PLSR is insensitive to such collinear variables and tolerates a large number of variables. PLSR projects the spectral data onto orthogonal latent variables (LVs)[30]. In the present study, the PLSR was used to predict the OARSI grade of each sample using the FTIR spectra as predictors. The models were created separately for the spectral data from the surface layer (0–15% of the thickness from the surface towards the tidemark) and the deep layer (65–100% of the thickness from the surface toward the tidemark). Using the second derivative and the raw spectra, four models were created. Additionally, the competitive adaptive reweighted sampling (CARS) algorithm[64] was used to select an optimal combination of variables (wavenumbers) for the PLSR

model. The maximum number of LVs was limited to 20. Leave-one-out cross-validation was used for validation of the multivariate models. Spearman's correlation was used to measure the agreement between the reference values and the predicted OARSI grades. In addition, the mean percent error was calculated for each model.

Univariate analysis. The integrated areas under the amide I and carbohydrate regions as well as the ratio of the integrated areas under the region around the $1338\,cm^{-1}$ peak and the amide II region were calculated for each spectrum for univariate analysis. Additionally, to avoid sample thickness variations, the integrated area of the carbohydrate region was also normalized by the integrated area of the amide I region. Consequently, depth-wise average profiles were calculated for every parameter. The regions of interest for the superficial and the deep layers were selected similarly to those used in the PLSR analysis. From the second derivative spectra, the peaks at $1064\,cm^{-1}$ and $1202\,cm^{-1}$ were used to analyze PG and collagen contents, respectively[53]. The results were compared with the OARSI grade using correlation analysis.

Polarized light microscopy (PLM). PLM was used as a reference method when studying the layered structure of AC[65]. All 28 samples were imaged using an Abrio PLM system (CRi, Inc., Woburn, MA, USA) mounted on a conventional light microscope (Nikon Diaphot TMD, Nikon, Inc., Shinagawa, Tokyo, Japan). The Abrio system consists of a green bandpass filter, a circular polarizer, and a computer-controlled analyzer composed of two liquid crystal polarizers and a CCD camera. In a PLM orientation image, each pixel consists of information about the average orientation of the collagen fibrils.

Statistical analysis. The Spearman's correlation test was used for all correlation analyses. For the group comparisons, the OARSI grades were pooled into three groups. Group 1 (early OA) included the grades from 1.0 to 2.5 ($n = 10$), group 2 (intermediate OA) included the grades from 3.0 to 3.5 ($n = 9$), and group 3 (severe OA) included the grades from 4.0 to 4.5 ($n = 9$). The Kruskall-Wallis test with pairwise comparison was used to analyze the statistical significance between the groups. The group comparisons were performed for the amide I region, carbohydrate region, collagen integrity and second derivative peak at $1202\,cm^{-1}$. The statistical analyses for the univariate parameters were performed using SPSS 22 software (IBM SPSS Statistics, SPSS Inc., Chicago, IL, USA).

References

1. Aigner, T., Sachse, A., Gebhard, P. M. & Roach, H. I. Osteoarthritis: Pathobiology—targets and ways for therapeutic intervention. *Advanced Drug Delivery Reviews* **58,** 128–149 (2006).
2. Huber, M., Trattnig, S. & Lintner, F. Anatomy, biochemistry, and physiology of articular cartilage. *Invest. Radiol.* **35,** 573–580 (2000).
3. Aigner, T. & McKenna, L. Molecular pathology and pathobiology of osteoarthritic cartilage. *Cellular and Molecular Life Sciences* **59,** 5–18 (2002).
4. Bi, X., Yang, X., Bostrom, M. P. G. & Camacho, N. P. Fourier transform infrared imaging spectroscopy investigations in the pathogenesis and repair of cartilage. *Biochimica et Biophysica Acta-Biomembranes* **1758,** 934–941 (2006).
5. Buckwalter, J. A. & Mankin, H. J. Articular cartilage. Part II: Degeneration and osteoarthrosis, repair, regeneration, and transplantation. *Journal of Bone and Joint Surgery - Series A* **79,** 612–632 (1997).
6. Guilak, F., Ratcliffe, A., Lane, N., Rosenwasser, M. P. & Mow, V. C. Mechanical and biochemical changes in the superficial zone of articular cartilage in canine experimental osteoarthritis. *J. Orthop. Res.* **12,** 474–484 (1994).
7. Panula, H. E. *et al.* Articular cartilage superficial zone collagen birefringence reduced and cartilage thickness increased before surface fibrillation in experimental osteoarthritis. *Ann. Rheum. Dis.* **57,** 237–245 (1998).
8. Aigner, T. & Schmitz, N. In *Rheumatology*, 5th edition (eds Hetherington, P. & Gunzel, L.) 1741–1759 (Elsevier Ltd., Philadelphia, 2011).
9. Buckwalter, J. A. & Martin, J. A. Osteoarthritis. *Adv. Drug Deliv. Rev.* **58,** 150–167 (2006).
10. Pritzker, K. P. H. *et al.* Osteoarthritis cartilage histopathology: Grading and staging. *Osteoarthritis and Cartilage* **14,** 13–29 (2006).
11. Cockman, M. D. *et al.* Quantitative imaging of proteoglycan in cartilage using a gadolinium probe and microCT. *Osteoarthritis and Cartilage* **14,** 210–214 (2006).
12. Hirvasniemi, J. *et al. In vivo* comparison of delayed gadolinium-enhanced MRI of cartilage and delayed quantitative CT arthrography in imaging of articular cartilage. *Osteoarthritis and Cartilage* **21,** 434–442 (2013).
13. Matzat, S. J., Kogan, F., Fong, G. W. & Gold, G. E. Imaging Strategies for Assessing Cartilage Composition in Osteoarthritis. *Curr. Rheumatol. Rep.* **16,** 11, doi: 10.1007/s11926-014-0462-3 (2014).
14. Shen, S., Wang, H., Zhang, J., Wang, F. & Chen, M. T1ρ magnetic resonance imaging quantification of early articular cartilage degeneration in a rabbit model. *BMC Musculoskelet. Disord.* **16,** 1, doi: 10.1186/s12891-015-0810-0 (2015).
15. Garnov, N. *et al. In vivo* MRI analysis of depth-dependent ultrastructure in human knee cartilage at 7T. *NMR Biomed.* **26,** 1412–1419 (2013).
16. Bansal, P. N., Joshi, N. S., Entezari, V., Grinstaff, M. W. & Snyder, B. D. Contrast enhanced computed tomography can predict the glycosaminoglycan content and biomechanical properties of articular cartilage. *Osteoarthritis Cartilage* **18,** 184–191 (2010).
17. O'Malley, M. J. & Chu, C. R. Arthroscopic optical coherence tomography in diagnosis of early arthritis. *Minimally Invasive Surg.* **2011,** doi: 10.1155/2011/671308 (2011).
18. Afara, I. O., Moody, H., Singh, S., Prasadam, I. & Oloyede, A. Spatial mapping of proteoglycan content in articular cartilage using near-infrared (NIR) spectroscopy. *Biomedical Optics Express* **6,** 144–154 (2015).
19. Marticke, J. K. *et al.* How do visual, spectroscopic and biomechanical changes of cartilage correlate in osteoarthritic knee joints? *Clin. Biomech.* **25,** 332–340 (2010).
20. Palukuru, U. P., McGoverin, C. M. & Pleshko, N. Assessment of hyaline cartilage matrix composition using near infrared spectroscopy. *Matrix Biol.* **38,** 3–11 (2014).
21. Hanifi, A. *et al.* Infrared fiber optic probe evaluation of degenerative cartilage correlates to histological grading. *Am. J. Sports Med.* **40,** 2853–2861 (2012).
22. Mizaikoff, B. Waveguide-enhanced mid-infrared chem/bio sensors. *Chem. Soc. Rev.* **42,** 8683–8699 (2013).
23. Baker, M. J. *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols* **9,** 1771–1791 (2014).
24. Komal Kumar, J. & Devi Prasad, A. G. Fourier transform infrared spectroscopy an advanced technique for identification of biomolecules. *Drug Invention Today* **4,** 616–618 (2012).
25. Rieppo, L., Rieppo, J., Jurvelin, J. S. & Saarakkala, S. Fourier transform infrared spectroscopic imaging and multivariate regression for prediction of proteoglycan content of articular cartilage. *PLoS One* **7,** 2, doi: 10.1371/journal.pone.0032344 (2012).

26. Faillace, M. E., Phipps, R. J. & Miller, L. M. Fourier transform infrared imaging as a tool to chemically and spatially characterize matrix-mineral deposition in osteoblasts. *Calcif. Tissue Int.* **92,** 50–58 (2013).

27. Boskey, A. & Pleshko Camacho, N. FT-IR imaging of native and tissue-engineered bone and cartilage. *Biomaterials* **28,** 2465–2478 (2007).

28. Camacho, N. P., West, P., Torzilli, P. A. & Mendelsohn, R. FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage. *Biopolymers–Biospectroscopy Section* **62,** 1–8 (2001).

29. West, P. A., Bostrom, M. P. G., Torzilli, P. A. & Camacho, N. P. Fourier transform infrared spectral analysis of degenerative cartilage: An infrared fiber optic probe and imaging study. *Appl. Spectrosc.* **58,** 376–381 (2004).

30. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab. Syst.* **58,** 109–130 (2001).

31. Rieppo, L., Saarakkala, S., Jurvelin, J. S. & Rieppo, J. Prediction of compressive stiffness of articular cartilage using Fourier transform infrared spectroscopy. *J. Biomech.* **46,** 1269–1275 (2013).

32. Yin, J., Xia, Y. & Lu, M. Concentration profiles of collagen and proteoglycan in articular cartilage by Fourier transform infrared imaging and principal component regression. *Spectrochimica Acta-Part A: Molecular and Biomolecular Spectroscopy* **88,** 90–96 (2012).

33. Hanifi, A., McCarthy, H., Roberts, S. & Pleshko, N. Fourier Transform Infrared Imaging and Infrared Fiber Optic Probe Spectroscopy Identify Collagen Type in Connective Tissues. *PLoS One* **8,** 5, doi: 10.1371/journal.pone.0064822 (2013).

34. Bi, X. *et al.* Fourier transform infrared imaging and MR microscopy studies detect compositional and structural changes in cartilage in a rabbit model of osteoarthritis. *Analytical and Bioanalytical Chemistry* **387,** 1601–1612 (2007).

35. Hanifi, A. *et al.* Infrared fiber optic probe evaluation of degenerative cartilage correlates to histological grading. *Am. J. Sports Med.* **40,** 2853–2861 (2012).

36. Li, G. *et al.* A chemometric analysis for evaluation of early-stage cartilage degradation by infrared fiber-optic probe spectroscopy. *Appl. Spectrosc.* **59,** 1527–1533 (2005).

37. Bonifacio, A. *et al.* Chemical imaging of articular cartilage sections with Raman mapping, employing uni- and multi-variate methods for data analysis. *Analyst* **135,** 3193–3204 (2010).

38. Hedegaard, M. *et al.* Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging. *Theoretical Chemistry Accounts* **130,** 1249–1260 (2011).

39. Kobrina, Y. *et al.* Cluster analysis of infrared spectra of rabbit cortical bone samples during maturation and growth. *Analyst* **135,** 3147–3155 (2010).

40. Kobrina, Y., Rieppo, L., Saarakkala, S., Jurvelin, J. S. & Isaksson, H. Clustering of infrared spectra reveals histological zones in intact articular cartilage. *Osteoarthritis and Cartilage* **20,** 460–468 (2012).

41. Kobrina, Y. *et al.* Cluster analysis of infrared spectra can differentiate intact and repaired articular cartilage. *Osteoarthritis and Cartilage* **21,** 462–469 (2013).

42. Croxford, A. M. *et al.* Specific antibody protection of the extracellular cartilage matrix against collagen antibody-induced damage. *Arthritis Rheum.* **62,** 3374–3384 (2010).

43. Croxford, A. M. *et al.* Chemical changes demonstrated in cartilage by synchrotron infrared microspectroscopy in an antibody-induced murine model of rheumatoid arthritis. *J. Biomed. Opt.* **16,** 6, doi: 10.1117/1.3585680 (2011).

44. Sophia Fox, A. J., Bedi, A. & Rodeo, S. A. The basic science of articular cartilage: Structure, composition, and function. *Sports Health* **1,** 461–468 (2009).

45. Pearle, A. D., Warren, R. F. & Rodeo, S. A. Basic science of articular cartilage and osteoarthritis. *Clin. Sports Med.* **24,** 1–12 (2005).

46. Kohler, A. *et al.* Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. *Analytical and Bioanalytical Chemistry* **389,** 1143–1153 (2007).

47. Jackson, M., Choo, L. P., Watson, P. H., Halliday, W. C. & Mantsch, H. H. Beware of connective tissue proteins: Assignment and implications of collagen absorptions in infrared spectra of human tissues. *Biochimica et Biophysica Acta - Molecular Basis of Disease* **1270,** 1–6 (1995).

48. Servaty, R., Schiller, J., Binder, H. & Arnold, K. Hydration of polymeric components of cartilage - An infrared spectroscopic study on hyaluronic acid and chondroitin sulfate. *Int. J. Biol. Macromol.* **28,** 121–127 (2001).

49. Király, K. *et al.* Application of selected cationic dyes for the semiquantitative estimation of glycosaminoglycans in histological sections of articular cartilage by microspectrophotometry. *Histochem. J.* **28,** 577–590 (1996).

50. Dousseau, F. & Pézolet, M. Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods. *Biochemistry* (*N. Y.*) **29,** 8771–8779 (1990).

51. Martens, H., Bruun, S. W., Adt, I., Sockalingum, G. D. & Kohler, A. Pre-processing in biochemometrics: Correction for path-length and temperature effects of water in FTIR bio-spectroscopy by EMSC. *J. Chemometrics* **20,** 402–417 (2006).

52. Ding, M., Odgaard, A. & Hvid, I. Changes in the three-dimensional microstructure of human tibial cancellous bone in early osteoarthritis. *J. Bone Jt. Surg. Ser. B* **85,** 906–912 (2003).

53. Rieppo, L. *et al.* Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage. *Osteoarthritis Cartilage* **20,** 451–459 (2012).

54. Saarakkala, S. *et al.* Depth-wise progression of osteoarthritis in human articular cartilage: investigation of composition, structure and biomechanics. *Osteoarthritis and Cartilage* **18,** 73–81 (2010).

55. Lippiello, L., Hall, D. & Mankin, H. J. Collagen synthesis in normal and osteoarthritic human cartilage. *J. Clin. Invest.* **59,** 593–600 (1977).

56. Squires, G. R., Okouneff, S., Ionescu, M. & Poole, A. R. The pathobiology of focal lesion development in aging human articular cartilage and molecular matrix changes characteristic of osteoarthritis. *Arthritis Rheum.* **48,** 1261–1270 (2003).

57. Sandell, L. J. & Aigner, T. Articular cartilage and changes in arthritis An introduction: Cell biology of osteoarthritis. *Arthritis Res.* **3,** 107–113 (2001).

58. Jackson, M. & Mantsch, H. H. The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30,** 95–120 (1995).

59. West, P. A., Torzilli, P. A., Chen, C., Lin, P. & Camacho, N. P. Fourier transform infrared imaging spectroscopy analysis of collagenase-induced cartilage degradation. *J. Biomed. Opt.* **10,** 1, doi: 10.1117/1.1854131 (2005).

60. Bassan, P. *et al.* Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst* **135,** 268–277 (2010).

61. Wang, L. & Mizaikoff, B. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Anal. Bioanal. Chem.* **391,** 1641–1654 (2008).

62. Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36,** 1627–1639 (1964).

63. MacQueen, J. K-means some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. Math. Stat. Probab.* **1,** 287–297 (1967).

64. Li, H., Liang, Y., Xu, Q. & Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling methlod for multivariate calibration. *Anal. Chim. Acta* **648,** 77–84 (2009).

65. Rieppo, J. *et al.* Practical considerations in the use of polarized light microscopy in the analysis of the collagen network in articular cartilage. *Microsc. Res. Tech.* **71,** 279–287 (2008).

Acknowledgements

Author Contributions

Design and conception of the study: J.O., L.R., M.A.J.F. and S.S. Data acquisition, analysis and interpretation of the data: J.O. Collection of the human tissue samples in the Oulu University Hospital: M.V. and P.L. Writing of the manuscript: J.O., L.R., M.A.J.F., M.V., P.L. and S.S. All authors approved the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Oinas, J. *et al.* Imaging of Osteoarthritic Human Articular Cartilage using Fourier Transform Infrared Microspectroscopy Combined with Multivariate and Univariate Analysis. *Sci. Rep.* **6**, 30008; doi: 10.1038/srep30008 (2016).