## Title

Ensemble multivariate analysis to improve identification of articular cartilage disease in noisy Raman spectra



## Journal

Journal of Biophotonics, 8(7)

## ISSN

1864-063X

## Authors

Richardson, Wade
Wilkinson, Dan
Wu, Ling
et al.

## Publication Date

2015-07-01

## DOI

10.1002/jbio.201300200

Peer reviewed

HHS Public Access

Author manuscript

*J Biophotonics*. Author manuscript; available in PMC 2015 July 03.

Published in final edited form as:

*J Biophotonics*. 2015 July ; 8(7): 555–566. doi:10.1002/jbio.201300200.

# Ensemble multivariate analysis to improve identification of articular cartilage disease in noisy Raman spectra

**Wade Richardson**[1], **Dan Wilkinson**[1], **Ling Wu**[2], **Frank Petrigliano**[2], **Bruce Dunn**[*,1], and **Denis Evseenko**[*,2]

[1]Department of Materials Science and Engineering, University of California, Los Angeles

[2]Department of Orthopaedic Surgery, University of California, Los Angeles

## Abstract

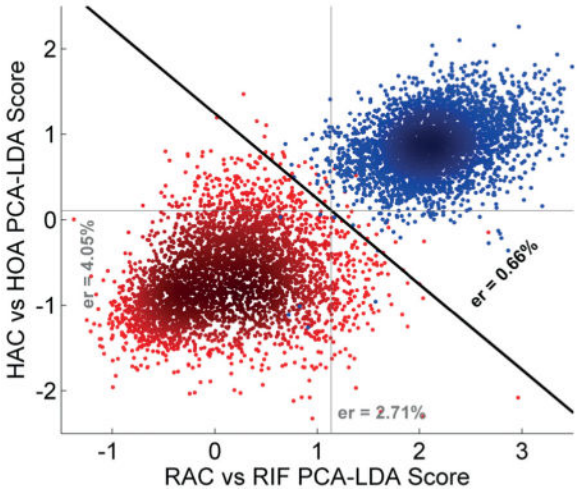
The development of new methods for the early diagnosis of cartilage disease could offer significant improvement in patient care. Raman spectroscopy is an emerging biomedical technology with unique potential to recognize disease tissues, though difficulty in obtaining the samples needed to train a diagnostic and excessive signal noise could slow its development into a clinical tool. In the current report we detail the use of principal component analysis – linear discriminant analysis (PCA-LDA) on spectra from pairs of materials modeling cartilage disease to create multiple spectral scoring metrics, which could limit the reliance on primary training data for identifying disease in low signal-to-noise-ratio (SNR) Raman spectra. Our proof-of-concept experiments show that combinations of these model-metrics has the potential to improve the classification of low-SNR Raman spectra from human normal and osteoarthritic (OA) cartilage over a single metric trained with spectra from the same healthy and OA tissues.

## Abstract

[*]Corresponding authors: bdunn@ucla.edu, or DEvseenko@mednet.ucla.edu.

**Author biographies** Please see Supporting Information online.

Supporting Information: Additional supporting information can be found in the online version of this article at the publisher's website: doi: http://dx.doi.org/10.1002/jbio.201300200.




Scatter plot showing the PCA-LDA derived human-disease-metric scores versus rat-model-metric scores for 7656 low signal-to-noise spectra from healthy (blue) and osteoarthritic (red) cartilage. Light vertical and horizontal lines represent the optimized single metric classification boundary. Dark diagonal line represents the classification of boundary resulting from the optimized combination of the two metrics. Abbreviations: er (error rate), PCA-LDA (principal component analysis – linear discriminant analysis), HOA (human osteoarthritis), HAC (human articular cartilage), RIF (rat injury fibrocartilage), RAC (rat articular cartilage).

**Keywords**

Raman spectroscopy; multivariate analysis; articular cartilage; osteoarthritis

## 1. Introduction

Degenerative diseases of articular cartilage, such as osteoarthritis (OA), represent a widespread clinical challenge with deficient early diagnostic and treatment options available [1]. Current diagnostic methods often rely on direct observations by clinicians of morphological tissue abnormality or on changes in bulk properties, such as density, detected by radiology (i.e. X-ray or MRI) [2, 3]. Alternatively, biomarkers have been sought that can provide a simple blood test [4], but will likely be less sensitive than interrogation of the affected tissue – articular cartilage. Invasive and often counterproductive extraction of tissue samples for histological analysis is currently required to detect and measure the primary biochemical changes that occur in the extracellular matrix of pathological cartilage, such as reduced levels of glycosaminoglycans (GAGs). There is clinical need for a minimally-invasive cartilage evaluation tool that can examine these molecular changes, with the sensitivity to detect disease in pre-clinical patients.

Though conventionally utilized in materials chemistry, Raman spectroscopy has recently gained significant attention as a tool for biomedical characterization as well as medical diagnosis [5–7]. Inelastic Raman scattering reveals the distinctive molecular vibrational states of a material, providing a label-free characterization approach that is sensitive to small

molecular changes. Recent investigations using multivariate analysis to decode the large amount of vibrational information in a spectrum, demonstrate that Raman spectroscopy has enormous potential as a diagnostic tool [8–12]. Particularly encouraging is work that has shown that Raman can be used to distinguish the samples of synovial fluid [13] as well as bone [14] taken from healthy and OA patients. However, direct endoscopic examination of articular cartilage *in vivo* probably remains the most promising means to create an effective clinical Raman diagnostic method for cartilage disease [15].

However, there remain several potential impediments to the realization of such a diagnostic method. Obtaining the data to train this type of method might prove difficult, especially data from pre-clinical patients aimed at detecting disease earlier than conventional methods. In addition, maintaining the a minimally-invasive test with an endoscopic probe necessitates short spectral acquisition times. Because Raman scattering produces an inherently weak signal, shorter acquisition times can result in spectra with a low signal-to-noise ratio (SNR), which could degrade spectral information. While a number of researchers has demonstrated techniques of combining several multivariate methods [16–18], training a classification system with spectra from several related materials that model the disease to some extent has not often been considered. The potential advantages of this approach include the ability of utilize easily obtained training samples and possible improvement to the classification of low-SNR spectra.

In the research presented here, we performed *in vitro* pilot experiments exploring the potential use of multiple model materials in a Raman method to identify cartilage disease. We used a multivariate method trained with Raman spectra from three different pairs of model training materials (Figure 1). One member of each pair represents healthy cartilage and the other pathological cartilage, and the three pairs vary greatly in how closely they relate to the OA and human cartilage samples used as the control pair. By training the multivariate analysis with spectra from each pair, we created three diagnostic metrics on which a spectrum's score indicates its similarity to one training material or the other. The three metric were prepared, respectively, by comparing purified collagen type II (COL2) to collagen type I (COL1), rat articular cartilage (RAC) to rat injury-induced fibrocartilage (RIF), and human articular cartilage (HAC) to human osteoarthritic cartilage (HOA). We tested these metrics for the capacity to correctly classify noisy spectra obtained as spectral maps of either healthy human or osteoarthritic human cartilage. We then tested all possible combinations of the metrics for the classification of these same healthy and osteoarthritic spectral maps. These combined classifications were evaluated based on the rate of spectra misclassified by a linearly optimized decision boundary.

We have found that each of the metrics is able to correctly classify healthy and OA cartilage to some extent, even the highly simplified collagen model metric. However, when the ensemble classification was measured, the rate of misclassified low-SNR spectra was reduced for all possible combinations. The improvement in classification may relate to the similarity of the coefficients belonging to the metrics used together, and appears to stem from the fact that each of the metrics will misclassify different individual spectra.

## 2. Experimental

### 2.1 Materials

Four specimens of human normal and Four specimens of osteoarthritis cartilage (International Cartilage Repair Society (ICRS) grade 2–3, indicating lesions up the the full depth of the cartilage) were obtained from either National Desease Research Interchange (NDRI) or Dr Marcel Karperien (University of Twente, Netherlands). Sprague-Dawley 11-week old rats (all male, Harlan Laboratories, Inc. Indiana, USA) were used for animal studies. Collagen type I from rat tail was purchased from Sigma Aldrich Corporation (St. Louis, MO, USA) and rat lathyritic collagen type II was purchased from MD Bioproducts (St. Paul, MN, USA). Both were used as received.

### 2.2 Rat injury model

The surface joint defects were generated in the femoral intra-condylar region of a knee joint in rats using methods previously described [19]. Briefly, three rats were anesthetized and medial para-patellar arthrotomy were carried out on one hind leg under a dissection microscope (Olympus, USA) using procedure previously described. Osteochondral defects were made with special metal tools. At weeks 2, 3 and 4, rats were sacrificed and dissected for examination. The 4 week post-injury rat was found to have the most substantial tissue repair, so this sample was used for Raman analysis. The injured joint was used to obtain fibrocartilage spectra, while the uninjured contralateral joint of the same rat was used to obtain cartilage spectra.

### 2.3 Raman sample preparation

Tissue samples for Raman analysis were prepared by fixing samples in 2% (vol/vol) paraformaldehyde then embedding in paraffin wax for sectioning. Sections 10 microns thick were prepared on quartz microscope slides and wet-mounted under 0.25 mm quartz coverglass to avoid fluorescence from standard soda-lime glass.

### 2.4 Raman spectroscopy

Raman spectra were acquired with a Renishaw In-Via micro-Raman spectrometer (Renishaw PLC, Wotton-under-Edge, Gloucestershire, UK) equipped with a 1800 lines $mm^{-1}$ grating, Rayleigh line rejection edge filter, Peltier-cooled deep depletion CCD array detector (576 × 384 pixels). The excitation source was the 514 nm line from argon ion laser at 25 mW. A 63×/1.2NA water-immersion objective was used to produce a spot size of approximately 1 μm on the sample. Wavenumber measurements were calibrated using the peak at 520 $cm^{-1}$ taken from a silicon wafer. Spectra used to create training sets were acquired by collecting data for 300 seconds following 60 seconds of photo-bleaching. In the case of cartilage these were obtained from matrix in the interterritorial regions of the middle zone. Spectral maps were prepared using an XYZ automated mapping sample stage with 0.1 μm step size, and an individual spectral acquisition time of 30 seconds with 30 seconds of photobleaching. All acquisitions were made in static scan mode.

Spectral training sets were derived from 48 rat articular cartilage (RAC) spectra and 59 rat injury-formed fibrocartilage spectra (RIF) taken from a matched pair of samples from

contralateral injured and uninjured joints; 80 human adult articular cartilage (HAC) spectra were taken from one specimen and 80 osteoarthritis-effected cartilage (HOA) spectra were taken from one specimen; and 16 purified collagen type II (COL2) spectra and 16 purified collagen type I (COL1) spectra taken from saturated solutions of each protein under quartz coverglass.

### 2.5 Multivariate analysis

Spectra were prepared for analysis by first truncating the data to include wavenumbers from 650 to 1800 $cm^{-1}$. Sample spectra, as well as spectra taken from blank areas on the same samples (i.e. water and quartz only) were then both subjected to a modified range-independent algorithm to remove background fluorescence not eliminated through photo-bleaching [20]. An average of the cleaned blank spectra was then subtracted from the cleaned tissue spectra. Each modified tissue spectrum was then leveled and normalized to its standard deviation.

Using built-in functions in Matlab (Mathworks INC, Natick, MA, USA), PCA was first performed on the combination of two training sets to obtain PCs representing the variance of the combined data set. LDA was then performed using these PCs as the input variables to determine the spectral features that best differentiate between the two training sets. To improve the quality of our classification, training sets were subjected to an iterative leave-one-out (LOO) method. The leave-one-out procedure is often used as a means of validating such PCA-LDA analyses [21]. In this procedure, each observation (spectrum) is tested by removing it from the set prior to analysis, running the analysis without it, then applying the resulting classification metric to the removed observation and examining if it was classified correctly. In our method each incorrectly classified spectra was removed from the data set and the process was repeated. Because removing all misclassified spectra at once will affect the overall analysis, new misclassifications can be produced. This iteration was repeated until no further misclassifications were produced. Misclassification was defined as a PCA-LDA score of the opposite arithmetic sign from what was designated for the spectrum's class.

The number PCs input into the LDA was varied between 2 and 20 and the analysis, including the iterative LOO, was repeated for each increment. The number of PCs used in the final analysis was determined by measuring the spread of the training set PCA-LDA scores, defined as the difference of the mean score of each set, divided by the sum of the standard deviation of two sets. The highest number of PCs after which no further increases in this value was observed was the number of PCs used in the analysis. In this way, over-fitting [22, 23] is avoided. For further details on PCA-LDA scoring see Supporting Information.

### 2.6 Immunohistochemistry

5 μm thick paraffin sections of human and rat tissues were used for immunohistochemistry. Rabbit antibodies against COL I and COL II (both Abcam, Cambrige, MA) were used, followed by incubation of HRP conjugated secondary antibodies against rabbit IgG (Vector Laboratories, Burlingame, CA). Antibodies were then visualized by Peroxidase substrate Kit

DAB (Vector Laboratories, Burlingame, CA). Images were acquired using the Zeiss Axiovision software version 4.8 Carl Zeiss Microscope (Carl Zeiss, Germany) equipped with ApoTome.2: Modules for Axio Imager.2 and Axio Observer with 10, 20 and 40× (1.3 numerical aperture (NA)) and 63× (1.4 NA) oil-immersion objectives.

### 2.7 RNA isolation and quantitative PCR

RNA samples were extracted with the RNeasy Micro Kit (Qiagen, Germantown, MD) from single cells suspension of collagenase digested cartilage from three healthy and three OA specimens. For quantitative polmerase chain reaction (qPCR), one microgram of total RNA was reverse-transcribed into cDNA using the Omniscript RT kit (Qiagen, Hilden, Germany). qPCR was performed on cDNA samples by using the Maxima SYBR Green/ROX qPCR Master Mix (Applied Biosystems, Austin, TX). Reactions were carried out on ViiA™ 7 Real-Time PCR System (Life Technologies, Grand Island, NY). For each reaction a melting curve was generated to test primer dimer formation and non-specific priming. The sequences of primers are available on request. Relative Expression was calculated using the double delta Ct method (24). A geometric average of B2M ($\beta$2 microglobulin), TBP (TATA-box binding protein) and RPL7 (ribosomal protein L7) was used for gene normalization.

## 3. Results

### 3.1 Raman spectra from training set pairs

Raman spectra for training the three metrics were obtained from three pairs of materials, each modeling healthy and pathological cartilage in a different way. The average spectrum for each training material, after background and fluorescence removal, is represented in Figure 2a. In the first pair, training spectra were obtained from the same human normal and osteoarthritic cartilage samples that provided the low SNR spectral maps used to test the metrics derived from all of the material pairs. The non-independent classification with this metric – classification of the same material used for training – was designed to provide a positive control representing the best possible single-metric classification. Given that the OA sample used represents a mid-to-late stage of the disease, the human metric only serves in our study as a generalized direct model of cartilage disease, rather than a model of early stage OA.

For the collagen metric, comparison of spectra from two purified collagen isoforms provide a highly simplified model of a very limited compositional change. COL2 is the primary component of hyaline cartilage, while COL1 is almost entirely absent from middle zone articular cartilage. Though the primary pathomorphological changes in OA do not include the shift in collagen isoforms, an increase in COL1 has been observed in OA cartilage, particularly in more advanced stages [24, 25]. We have verified this overexpression immunostaining and qPCR in healthy and OA samples of the same ICRS grade as the Raman sample (Suppl. Figure 1). Additionally, the collagen metric is particularly well suited for supplementing the human metric. As seen in Figure 2a, the strong resemblance between the tissue spectra and purified collagen spectra suggests that the Raman scattering from the collagen component of cartilage tissues may dominate due to a combination of concentration and scattering efficiency.

Finally, we have employed a modified microfracture surgery on rat joints – most aspects of human articular cartilage are well-conserved in rats. Micro-fracture surgery is an articular cartilage repair technique performed by boring tiny fractures in the cartilage and underlying bone to induce healing of the injury with fibrocartilage, which does not exhibit the same biomechanical functionality as native cartilage [26, 27]. Figure 3a shows a cross-sectional micrograph of the rat femoral head 4 weeks post-surgery with a clearly distinguishable morphological boundary between the injury site and adjacent native cartilage. While the rat injury model is not the best rat model for OA – several others more closely mimic the disease in humans [28, 29] – it does represents a good intermediate between the human control metric and the simplified collagen metric. The COL2-to-COL1 shift is much more dramatic in fibrocartilage than OA, but the rat metric will also reflect the wider compositional and structural differences between cartilage and healed fibrocartilage. Thus both the rat and human metrics, in a more generalized sense, capture the breakdown of the highly organized structure of cartilage.

The average spectra for tissue and collagen samples (Figure 2a) are each similarly characterized by several prominent Raman bands – the complex amide I band (1665 $cm^{-1}$), the C–H stretch band (1451 $cm^{-1}$) and the amide III band (1245 $cm^{-1}$) dominate the higher wavenumbers, while the sharp peak associated with the phenylalanine ring-breathing mode (1004 $cm^{-1}$) as well as several C–C bands assigned to the protein backbone (940 $cm^{-1}$, 816 $cm^{-1}$), proline ring (922 $cm^{-1}$, 857 $cm^{-1}$) and hydroxyproline (922 $cm^{-1}$) are prominent in the lower wavenumbers. The complete list of peak assignments is compiled in Table 1. All of these features are consistent with published spectra from other sources [9, 15, 30, 31]. The similarity of all of the tissue-derived training spectra highlights the benefit of employing a multivariate analysis that can rely on many small variations.

### 3.2 PCA-LDA metrics

Iterated PCA-LDA of the Raman spectra from the three pairs of training materials (HAC vs. HOA, RAC vs. RIF, and COL2 vs. COL1) resulted in separated distributions of PCA-LDA scores for the two classes in each training set (Figure 4a). Separation of the distributions in the human, rat, and collagen metrics was maximized by using 17, 12, and 12 PCs and removing 1, 5, and 3 misclassified training spectra, respectively.

The metric coefficients (Figure 3b), which represent the relative influence of spectral features in the calculation of the PCA-LDA scores, show significant variations between the three metrics. While several bands (phenylalanine ring breathing mode at 1004 $cm^{-1}$, and C–C protein backbone modes at 940 $cm^{-1}$ and 1130 $cm^{-1}$) are consistently represented, and the overall shapes for all three coefficient spectra appear to follow similar trends in certain regions, the sharper local features often fail to line up with one another. The Pearson correlation values were calculated to be 0.318 between the human and rat metric coefficients, 0.424 between the rat and collagen coefficients, and 0.182 between the human and collagen coefficients. These observations suggest that, while not completely uncorrelated, the metrics produced by the different training materials are based on substantially dissimilar features in the training spectra.

### 3.3 Classification of low-SNR spectral maps

We constructed spectral maps from low SNR spectra taken from equally spaced points to test the classification of our metrics. Spectral maps serve as a means for visualizing the spatial distribution and local variation of PCA-LDA scores, while providing a large number of individual spectra. Mapping the injury site region in the rat injury model sample, shown in Figure 3b provided a convenient way of validating our PCA-LDA method in the rat metric. In this map we observe the accurate reproduction of the tissue features, with the red and pink region of the injury site clearly distinguished from the blue of the surrounding articular cartilage. Classification was also found to be accurate in two independent pair-matched rat injury samples (Suppl. Figure 2). The effectiveness of the PCA-LDA is further supported by applying the collagen metric scoring to this same injury site map (Figure 3c). This COL2 vs. COL1 map closely mirrors the features of the RAC vs. RIF map with spectra from the uninjured cartilage scoring as very similar to collagen type II and the fibro-cartilage in the injury site scoring as more similar to collagen type I. This is in agreement with the histological staining for collagen type II and type I (Suppl. Figure 3), as well as the established matrix characteristics of tissue formed after microfracture surgery [26].

The performance of the three metrics was tested by the classification of low-SNR spectra taken from the same healthy and osteoarthritic cartilage samples used to train the human metric. We acquired six low SNR spectral maps each from the normal human articular cartilage and human osteoarthritic cartilage samples (Figure 4b) to provide the large number of spectra required to adequately estimate the statistical distribution of scores, as well as the ability to include local variation within the mapped region and variation across the sample. HAC map 1 and HOA map 1 were each taken from a region including the edge of a chondrocyte to provide some internal contrast (see Suppl. Figure 4).

A purely unguided classification – the classification boundary fixed at zero on each PCA-LDA metric – results in only the human metric producing a reasonably small misclassification rate of 5.3%. The metrics were calibrated by shifting the threshold to minimize the error rate (total fraction of misclassified spectra), in order to measure the maximum capacity for classification in each metric. As shown in Figure 4, once the boundary is optimized, misclassified spectra make up 4.1% of the 7656 total spectra from the 6 healthy cartilage and 6 osteoarthritic maps when scored with the human metric and 2.7% of the total when scored with the rat metric. Meanwhile, despite the exceedingly good classification of its training sets, the collagen metric did not perform as well as a diagnostic for osteoarthritic cartilage, misclassifying 14.7% of the map spectra. The distribution HOA and HAC scores are relatively symmetric, though in all metrics the HOA distribution is slightly broader and less normal in shape reflecting the inherent variability of OA disease. As a result of this symmetry that that the calibrated metrics have roughly equal sensitivity and specificity (see Supporting Information). Figure 4c shows that while the human and rat metric produce total-class score distributions (dark red and dark blue curves) with roughly the same separation of means and overlap, the single-map score sub-distributions (light red and light blue curves) do not maintain the same ordering of map-mean scores between two metrics. In the human metric OA map 6 has its mean most near to the classification boundary and map 3 has its mean near the overall mean for OA spectra. In the rat metric

scoring, the reverse is true. This suggests a somewhat distinct basis for classification between these metrics.

### 3.4 Combination of PCA-LDA metrics

To test our hypothesis that multiple PCA-LDA metrics, applied in concert could more accurately classify noisy cartilage and OA spectra than a single metric, we calculated the linear combination of the metric coefficients that best reduced the misclassification rate for the low SNR healthy and osteoarthritic cartilage map spectra. This relatively simple optimization was done by performing a second LDA, using combinations of the three different metric scores from the maps as input variables. The scores of one metric are plotted against those of another in Figure 5a. The diagonal line that best separates the two classes of spectra represents the optimal combination of metrics as calculated by the LDA. In these 2-metric plots it can also be observed that distribution of scores for each class form a rounded more or less circular shape, indicating that the relative scores of individual spectra within a class are not strongly correlated between any of the metrics.

Reduced misclassification was achieved for all metric combinations. Although the greatest reduction was obtained by combining all three metrics, this value was statistically the same as the 0.66% error rate observed for the combination based on the human and rat metrics. Other metric combinations also lead to a decrease in misclassification rate as shown in the Table in Figure 5d. Taken together, these results suggest that linear combinations of PCA-LDA scoring metrics base on model materials have the potential to improve classification of Raman spectra from healthy and disease cartilage, but that the extent of the advantage will be determined by the specific models used.

## 4. Discussion

We have examined the use of different materials that model cartilage disease to create spectroscopic metrics on which a test spectrum is scored as being normal or pathological. Because each metric performs its diagnostic classification based on different spectroscopic characteristics gleaned from the model it was derived from, combining them seems to prevent misclassifications made by individual metrics. We have found that some of these combinations eliminate nearly all misclassification of noisy, low-acquisition-time normal and OA cartilage spectra.

As expected, the non-independently trained human metric was able to correctly classify spectra from our healthy and OA human cartilage samples, despite low SNR. Alternatively, the ability of the rat and collagen metrics to separate these spectra demonstrates the potential of using independent model materials to train the classification in a cartilage diagnostic. While the accuracy of the collagen metric is too low to provide acceptable classification, our results indicate that very simple models based on purified components might be effective when used to supplement other metrics. For this type of training set, alternative multivariate methods, such as partial least squares regressions [33], might prove able to produce a more effective metric for future clinical investigations. Conversely, the rat metric has a remarkably low error rate, even ignoring any species-specific variation and considering only the significant dissimilarity between OA cartilage and fibrocartilage. Though neither of the

exogenous metrics is as inherently effective as the human metric due to the required shift in classification boundary, they offer a unique opportunity for testing the use of metrics in combination.

Applying the metrics in combination achieves improved classification, which appears to relate to the similarity of the metrics combined. One rationale behind employing model materials to train several classification metrics was that their varied spectroscopic basis would reduce overlap of misclassified spectra. This is analogous to other engineered redundancy systems where the mode of failure for each layer must be different for the combination to be effective [34]. The spectroscopic variation of the metrics is reflected in how closely their coefficients correlate. The combining the most highly correlated rat and collagen metric coefficients ($R^2 = 0.424$) only reduces the misclassification rate of the more effective rat metric from 2.7% to 2.5%. Conversely, combination of the collagen metric with the weakly correlated human metric ($R^2 = 0.182$) reduces misclassification by over half from 4.1% to 1.9%. Both of these combinations may be limited by the relatively poor performance of the collagen metric, since combining the two most effective metrics (rat and human) has the greatest reduction in misclassification.

In our results, the tendency observed of each metric to each separate the two classes of low-SNR spectra by score, while varying substantially between metrics in the scoring of single maps and individual spectra within a class suggests the mechanism by which improved classification is achieved. The circular class score distributions found in Figure 5 show that some spectra that would be misclassified by either single metric are not by the other. If arrangement of individual spectra within the distributions of each class were similar, we would expect to the scatter points cluster closely around a positive diagonal line. In which case there would be little potential to improve the classification by using a 2-dimensional decision boundary over a value on a single metric. The rearrangement of maps in Figure 4c suggests that some of this tendency is based on local variations in tissue, though a comparison of the human metrics scoring of its training data and the low-SNR maps shows that the score distribution of low-SNR scores is broadened even in a single localized map. Collectively, this suggests that misclassification due to local variation or due to spectral noise both might be addressed in the future by the use of multiple metrics in concert. Local variation in this proof-of-concept study may be considered a stand-in for patient-to-patient variation in the clinic. In which case these results are encouraging for the possibility of successfully supplementing limiting clinical training data with related model material data, but must considered preliminary.

The seemingly weak within-class scoring correlation between the metrics supports that they classify based on different biochemical bases. While interpretation becomes more difficult with the more layers of multivariate analysis applied [16], several notable features of the metrics we have derived can be extracted from examining the PCA-LDA coefficients. For instance, several features in the amide III region between 1225 $cm^{-1}$ and 1300 $cm^{-1}$, including the polar and non-polar peak positions [30] at 1270 $cm^{-1}$ and 1245 $cm^{-1}$, match well between the rat and human coefficients, but poorly with the collagen coefficient. Since the amide III mode is particularly sensitive to side chain effects [35], this indicates that variation in protein expression which is important for classification cannot be modelled by

the collagen metric. Lack of feature alignment in the majority of the coefficients suggests once again a largely unrelated molecular underpinning for the rat and human metrics. Particularly notable is the lack of shared features in the amide I band. This band is sensitive to protein secondary structure [21], suggesting that conformational changes between the training set pairs may be an important difference in the three metrics. Though the feature at 940 $cm^{-1}$ common to all three metrics indicates that the PCA-LDA analysis consistently tracks disruption in the helical conformation [36] of collagen, which may begin early in OA [37].

The rat and human coefficients are conspicuously missing features that would indicate the strong influence of GAGs on the classification mechanism. Change in the GAGs content of cartilage is a well-established characteristic of OA, particularly in early stages [22]. The prominent O-S-O stretching mode peak at 1062 $cm^{-1}$ from chondroitin sulfate (CS) has been employed in multivariate analyses to represent the presence of GAGs in cartilage matrix [8, 10, 23, 24]. We observe this peak in the coefficients of some of the lower order PCs from both the rat and human metrics (Suppl. Figure 5), but not in the coefficients representing the full PCA-LDA. Given the similarity between spectra from cartilage tissues and pure collagen, the scattering from GAGs may be too weak to register in the PCA-LDA It is also possible that changes in GAG content are accounted for in the metrics through more subtle spectral features that are difficult to interpret directly from the coefficients. A gradient in rat metric scores is observed in depth-maps of OA and normal human cartilage (Suppl. Figure 6), which corresponds to the gradient in GAG concentration known to exist between the superficial and middle zones in articular cartilage [38]. Sensitivity to changes in GAGs might explain the rat metric's accurate classification of human samples relative to the human metric, which produces no depth-profile scoring gradient.

Ultimately, effective early disease diagnosis in cartilage by Raman will likely be done endoscopically in a clinical setting. Our work suggests a significant advantage in using multiple metrics to increase confidence in a diagnosis, but the spectra used for training should reflect the clinical methods as closely as possible. The results presented here offer a proof-of-concept that can be leveraged in selection source materials for the clinical training sets. The good performance of the rat model in this report suggests that spectra easily obtained during routine arthroscopic follow ups of microfracture surgery patients might also be a useful basis for a supplementary scoring metric. Similarly, the collagen metric has shown that simple pure molecule training sets can improve classification when combined with a disease specific metric. Optimization of the included pure materials would be simplified with a clinical data set. Still, molecular variation during OA or other cartilage disease progression [41] suggests that an effort should be made to acquire early stage disease-specific training set for preclinical diagnosis.

## 5. Conclusion

In conclusion, we have found evidence that two or more multivariate scoring metrics can be considered in concert to improve Raman classification of healthy and OA cartilage. It appears that spectra from both model tissue and purified component bio-molecules might be useful as supplementary data for the classification of low-SNR spectra from healthy and

diseased human tissues. While this evidence is preliminary, it presents intriguing possibilities. From clinical prospective, the most valuable application of Raman spectroscopy is expected in patients with pre-clinical and pre-pathomorphological stages of OA not detectable with MRI or other currently used techniques. The ability to employ multiple classification metrics in a future clinical diagnostic may help to overcome the challenges of correctly classifying noisy spectra with a limited preclinical training data set.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work has been partially supported by the National Institutes of Health through Grant P01 GM081621.

## References

1. Freedman M, Hootman J, Helmick C. MMWR Morb Mortal Wkly Rep. 2007; 56
2. Moskowitz, RW. Osteoarthritis: Diagnosis and Medical/surgical Management. Lippincott Williams & Wilkins; 2007.
3. Potter K, Butler JJ, Horton WE, Spencer RG. Arthritis Rheum. 2000; 43:1580–1590. [PubMed: 10902763]
4. Attur M, Krasnokutsky-Samuels S. Curr Opin Rheumatol. 2013; 25:136–144. [PubMed: 23169101]
5. Krafft C, Sergo V. J Spectrosc. 2006; 20:195–218.
6. Ellis DI, Cowcher DP, Ashton L, O'Hagan S, Goodacre R. Analyst. 2013; 138:3871–84. [PubMed: 23722248]
7. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Shenk R, Wang N, Dasari RR, Fitzmaurice M, Feld MS. J Biomed Opt. 2009; 14:054023. [PubMed: 19895125]
8. Bonifacio A, Beleites C, Vittur F, Marsich E, Semeraro S, Paoletti S, Sergo V. Analyst. 2010; 135:3193–204. [PubMed: 20967391]
9. Lim NSJ, Hamed Z, Yeow CH, Chan C, Huang Z. J Biomed Opt. 2011; 16:017003. [PubMed: 21280924]
10. Pudlas M, Brauchle E, Klein TJ, Hutmacher DW, Schenke-Layland K. J Biophotonics. 2013; 6:205–211. [PubMed: 22678997]
11. Mansfield J, Moger J, Green E, Moger C, Winlove CP. J Biophotonics. 2013; 6:803–814. [PubMed: 23303610]
12. Krafft C, Steiner G, Beleites C, Salzer R. J Biophotonics. 2009; 2:13–28. [PubMed: 19343682]
13. Esmonde-White KA, Mandair GS, Raaii F, Jacobson JA, Miller BS, Urquhart AG, Roessler BJ, Morris MD. J Biomed Opt. 2009; 14:034013. [PubMed: 19566306]
14. Buchwald T, Niciejewski K, Kozielski M, Szybowicz M, Siatkowski M, Krauss H. J Biomed Opt. 2012; 17:017007. [PubMed: 22352673]
15. Esmonde-White KA, Esmonde-White FWL, Morris MD, Roessler BJ. Analyst. 2011; 136:1675–1685. [PubMed: 21359366]
16. Yang P, Hwa Yang Y, Zhou BB, Zomaya AY. Curr Bioinform. 2010; 5:296–308.
17. Menze BH, Petrich W, Hamprecht FA. Anal Bioanal Chem. 2007; 387:1801–1807. [PubMed: 17237926]
18. Beleites C, Salzer R. Anal Bioanal Chem. 2008; 390:1261–1271. [PubMed: 18228011]
19. Eltawil NM, De Bari C, Achan P, Pitzalis C, Dell'accio F. Osteoarthritis Cartilage. 2009; 17:695–704. [PubMed: 19070514]
20. Krishna H, Majumder SK, Gupta PK. J Raman Spectrosc. 2012; 43:1884–1894.
21. Chan J, Lieu D. J Biophotonics. 2009; 668:656–668. [PubMed: 19653219]

22. Grimbergen MCM, van Swol CFP, Kendall C, Verdaasdonk RM, Stone N, Bosch JLHR. Appl Spectrosc. 2010; 64:8–14. [PubMed: 20132590]
23. Wang, M.; Perera, A.; Gutierrez-Osuna, R. Proc IEEE Sensors 2004. Vienna, Austia: Oct. 2004 p. 591-594.
24. Miosge N, Hartmann M, Maelicke C, Herken R. Histochem Cell Biol. 2004; 122:229–236. [PubMed: 15316793]
25. Lahm A, Kasch R, Mrosek E, Spank H, Erggelet C, Esser J, Merk H. Histol Histopathol. 2012; 27:609–615. [PubMed: 22419025]
26. Chen H, Chevrier A, Hoemann CD, Sun J, Pi-card G, Buschmann MD. J Orthop Res. 2013; 31:1757–1764. [PubMed: 23843172]
27. Kalson NS, Gikas PD, Briggs TWR. Int J Clin Pract. 2010; 64:1444–1452. [PubMed: 20716151]
28. Hayami T, Pickarski M, Zhuo Y, Wesolowski GA, Rodan GA, Duong LT. Bone. 2006; 38:234–243. [PubMed: 16185945]
29. de Souza RA, Xavier M, Mangueira NM, Santos AP, Pinheiro ALB, Villaverde AB, Silveira L. Lasers Med Sci. 2014; 29:797–804. [PubMed: 23979802]
30. Frushour B, Koenig J. Biopolymers. 1975; 14:379–391. [PubMed: 1174668]
31. Notingher I, Bisson I, Bishop AE, Randle WL, Polak JMP, Hench LL. Anal Chem. 2004; 76:3185–3193. [PubMed: 15167800]
32. Bonifacio A, Beleites C, Vittur F, Marsich E, Semeraro S, Paoletti S, Sergo V. Analyst. 2010; 135:3193–3204. [PubMed: 20967391]
33. Wold S, Sjöström M, Eriksson L. Chemom Intell Lab Syst. 2001; 58:109–130.
34. Mitra S, McCluskey EJ. Proc 19th IEEE VLSI Test Symp VTS 2001, IEEE Comput Soc. :190–195. n.d.
35. Barth A, Zscherp C. Q Rev Biophys. 2002; 4:369–430. [PubMed: 12621861]
36. Pezolet M, Pigeon-Gosselin M, Caille JP. Biochim Biophys Acta – Protein Struct. 1978; 533:263–269.
37. Huebner JL, Williams JM, Deberg M, Henrotin Y, Kraus VB. Osteoarthritis Cartilage. 2010; 18:397–405. [PubMed: 19825496]
38. Mow, VC.; Hung, CT. Basic Biomech Musculoskelet Syst. Nordin, M.; Frankel, V., editors. Lea and Febiger; London: 1989. p. 60-101.
39. Koljenović S, Bakker Schut TC, van Meerbeeck JP, Maat APWM, Burgers SA, Zondervan PE, Kros JM, Puppels GJ. J Biomed Opt. 2004; 9:1187–1197. [PubMed: 15568939]
40. Kunstar A, Leijten J, van Leuveren S, Hilderink J, Otto C, van Blitterswijk CA, Karperien M, van Apeldoorn AA. J Biomed Opt. 2012; 17:116012. [PubMed: 23117807]
41. Pearle AD, Warren RF, Rodeo SA. Clin Sports Med. 2005; 24:1–12. [PubMed: 15636773]

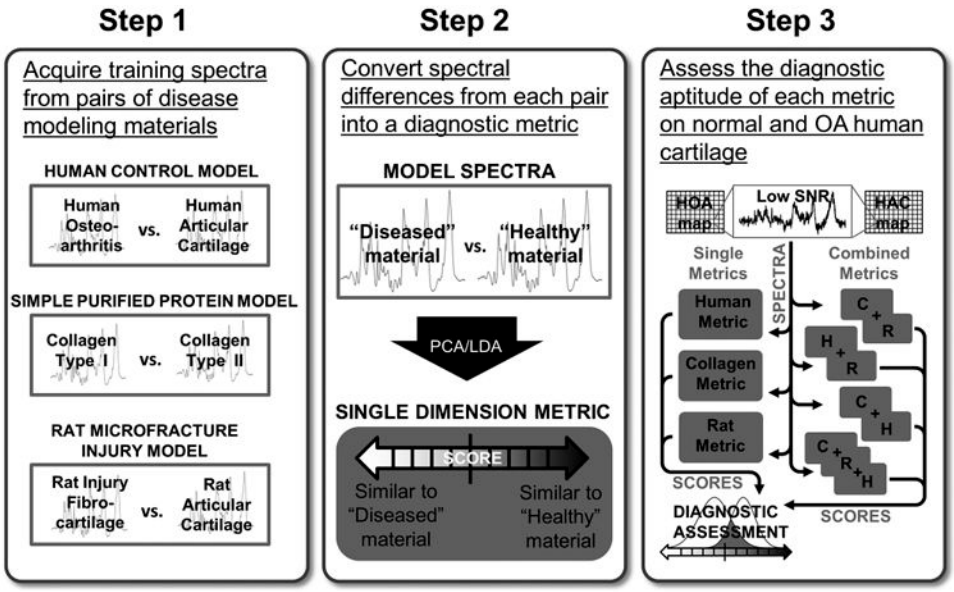



**Figure 1.**

A schematic representing the experimental process presented in this report. High-quality training spectra were first obtained from three sets of materials to model biochemical changes in osteoarthritis. The spectra from each set of materials was then subjected to multivariate analysis to create a scoring metric. Finally, the metrics were applied singly or in combination to maps of noisy spectra from healthy and osteoarthritic cartilage, the degree to which this produced separable groups of scores for the two tissues was measured. Abbreviations: OA (osteoarthritis), HOA (human osteoarthritis), HAC (human articular cartilage), SNR (signal-to-noise ratio), C (collagen metric), R (rat metric), H (human metric).

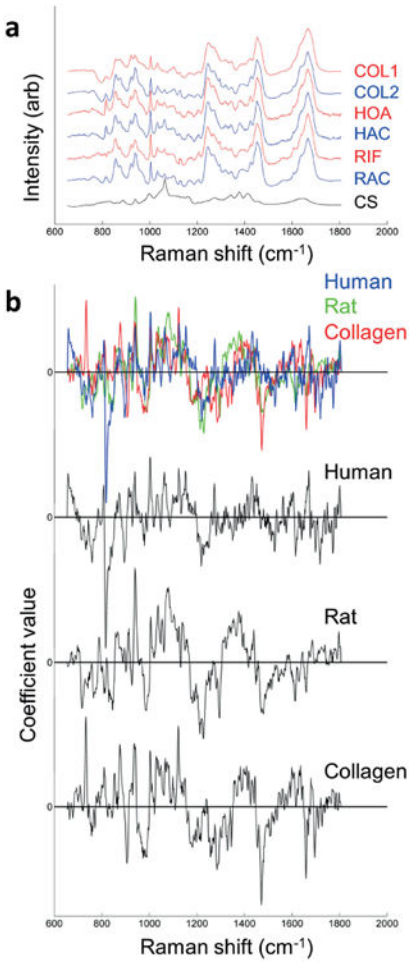



**Figure 2.**

(**a**) The average spectra for the training sets acquired from rat articular cartilage (RAC), rat injury-induced fibrocartilage (RIF), human articular cartilage (HAC), human osteoarthritic cartilage (HOA), purified collagen type II (COL2), and purified collagen type I (COL1). The similarity of the tissue derived spectra to those from the purified collagens suggests that this component of the tissue dominates the spectral signature. Also included for reference is the average spectrum from 15 spectra taken from purified chondroitin sulfate (CS), which spectral has features distinct from the cartilage tissue spectra. (**b**) PCA-LDA coefficients for the metrics from the HAC vs. HOA (Human), RAC vs. RIF (Rat), and COL2 vs. COL1 (Collagen) analyses.

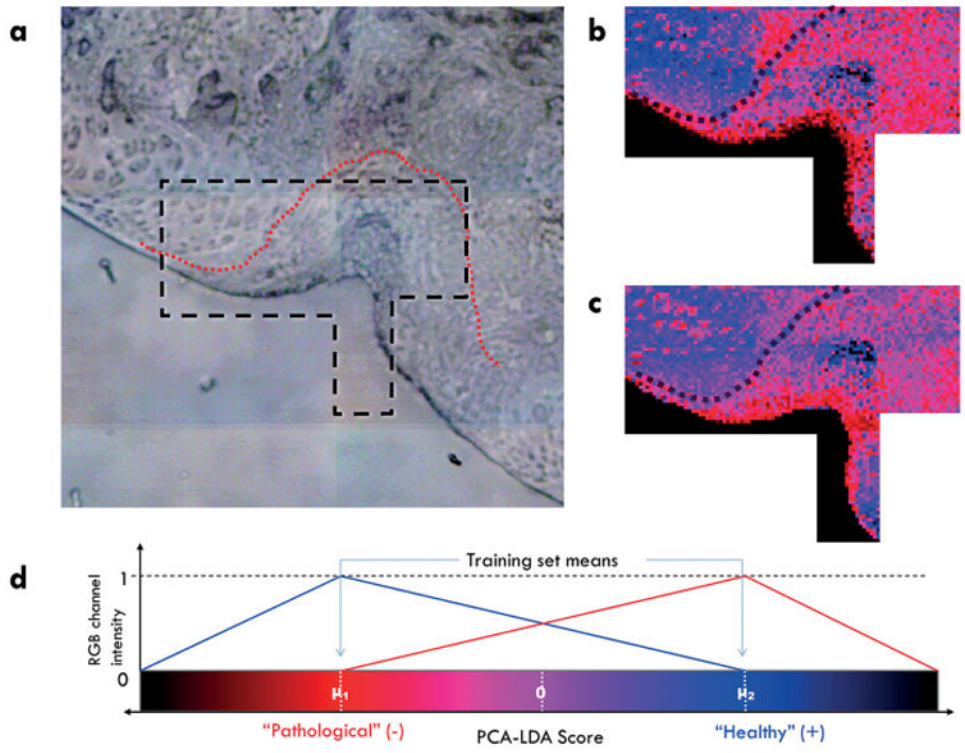



**Figure 3.**

(**a**) A phase contrast image of a 10 μm-thick cross-section of osteochondral defect on rat femur condyle, 4 weeks post-injury. The red dotted line indicates the approximate injury site. The black dashed line marks the region covered by spectral Raman mapping. (**b**) Raman spectral map of the injury site, for which the coloration of each pixel (7 μm × 7 μm) represents the rat metric PCA-LDA score of the low signal-to-noise spectrum taken at that location. In this map the distinctive scoring of the cartilage and fibrocartilage region is visualized. (**c**) Raman spectral map scored with the collagen metric. The similarity between the two maps likely reflects the similarity between the model materials from which the metrics were derived. (**d**) Schematic of the relationship between a spectrum's score and its representative color.

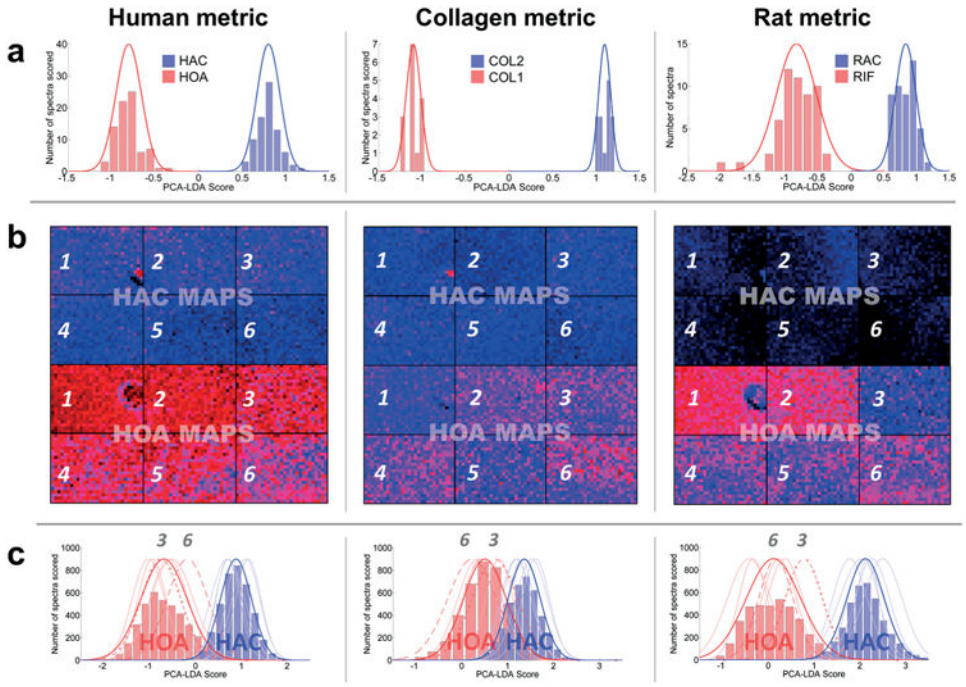
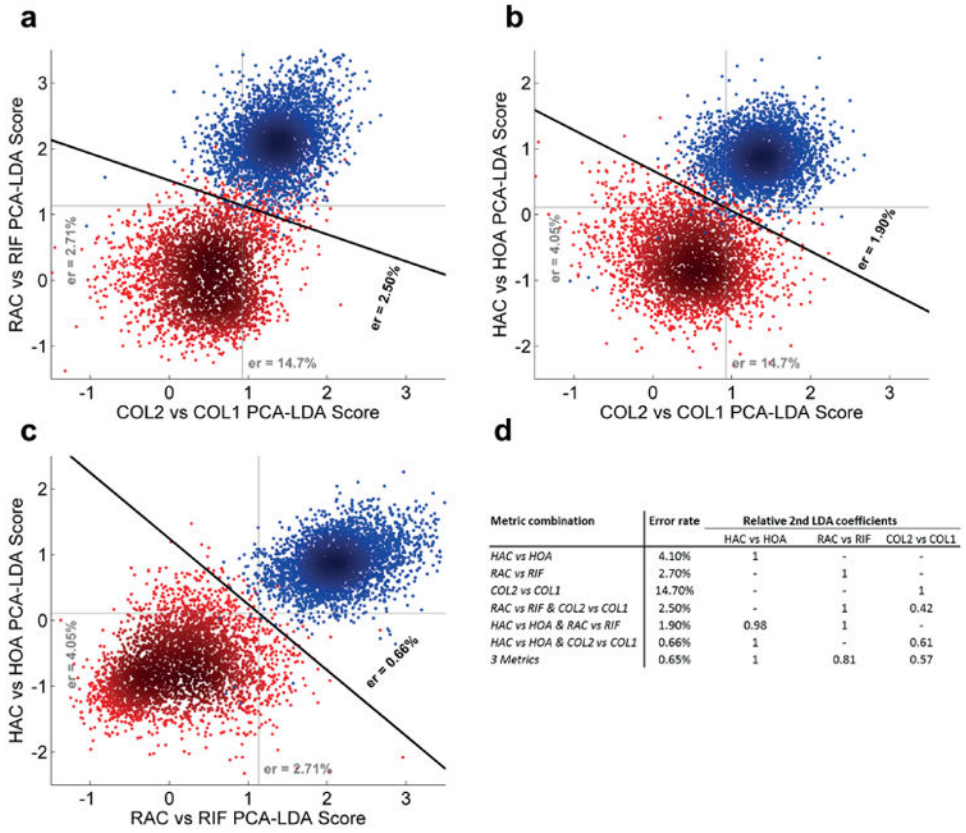



**Figure 4.**

(**a**) Histograms (red and blue bars) and fitted Gaussians (red and blue curves) of the PCA-LDA score distribution of the training sets for the HAC vs. HOA (Human, left), RAC vs. RIF (Collagen, middle), and COL2 vs. COL1 (Rat, right) metrics. (**b**) Low signal-to-noise Raman spectral maps taken from HAC (top) and HOA (bottom) samples, and scored by human (left), rat (middle), and collagen (right) metrics. (**c**) Distributions of PCA-LDA scores from the maps in (**b**) as scored by human (left), collagen (middle), and rat (right) metrics. The histograms and dark fitted Gaussians curves represent the combined distributions for the HAC maps (red) and HOA maps (blue). The light curves represent the distributions of individual HAC maps (red) and HOA maps (blue). The distributions of human metric scores from low-SNR human maps were broadened compared to scoring of the human training sets, highlighting the challenge of noisy spectra. Score distributions created by the rat metric scoring of the maps show a similar separation of the classes when compared to the human metric, but shifted towards the cartilage-like scores. The collagen metric does not perform as well in creating well-separated distributions for the two classes of spectra. Differential scoring of HOA maps 3 and 6 are highlighted by the darker dotted and dashed lines respectively, which indicates a tendency of the different metrics to vary in scoring of maps within a class, while still scoring the two classes as distinct groups.




| Metric combination | Error rate | Relative 2nd LDA coefficients | | |
|---|---|---|---|---|
| | | HAC vs HOA | RAC vs RIF | COL2 vs COL1 |
| *HAC vs HOA* | 4.10% | 1 | - | - |
| *RAC vs RIF* | 2.70% | - | 1 | - |
| *COL2 vs COL1* | 14.70% | - | - | 1 |
| *RAC vs RIF & COL2 vs COL1* | 2.50% | - | 1 | 0.42 |
| *HAC vs HOA & RAC vs RIF* | 1.90% | 0.98 | 1 | - |
| *HAC vs HOA & COL2 vs COL1* | 0.66% | 1 | - | 0.61 |
| *3 Metrics* | 0.65% | 1 | 0.81 | 0.57 |

**Figure 5.**

Scatter plots of the human metric score versus collagen metric score (**a**), human metric score versus rat metric score (**b**), and rat metric score versus collagen metric score (**c**) of each spectrum from the HAC (blue) and HOA (red) low signal-to-noise maps. Light vertical and horizontal lines represent present the optimized single metric classification boundary. Dark diagonal lines represent the optimized classification of boundary resulting from the combination of two metrics through a second LDA (er = error rate). The relatively isotropic spread of the within-class scatter – as opposed to following a positive diagonal – shows the weak correlation between to scores of individual spectra produced by different metrics. In each of the plots the scatter points present that represent spectra misclassified by one or the other metric, but correctly classified by the combined boundary directly illustrate the improvement possible with ensemble classification. (**d**) Table summarizing the optimized classification of low signal-to-noise map spectra by single metric and ensemble methods.

## Table 1

Raman peaks and their assignments.

| Peak ($cm^{-1}$) | Ref. | | Assignment |
|---|---|---|---|
| 1665 | [30, 32] | s | Amide I |
| 1637 | [30] | sh | Amide I |
| 1606 | [30, 31] | m | Phe, Tyr |
| 1560 | [32] | m | Amide II |
| 1460 | [30] | sh | $CH_2$ def. |
| 1452 | [15, 30, 32] | s | $CH_2$ def. |
| 1442 | [15] | sh | $CH_2$ def. |
| 1425 | [30] | m | $COO^-$ (GAGs) |
| 1390 | [30] | m | unassigned |
| 1343 | [30, 31] | w | $CH_2$ wag. |
| 1319 | [30, 31] | w | $CH_2$ twist |
| 1295 | [15, 31] | sh | $CH_2$ def |
| 1270 | [30] | s | Amide III |
| 1245 | [30] | s | Amide III |
| 1207 | [30–32] | m | Hypro, TYr |
| 1165 | [30] | w | C–C/C-N str. |
| 1130 | [30] | w | C–N str. |
| 1070–1100 | [30] | w | $PO^{2-}$ str. (DNA/RNA) |
| 1062 | [15, 30, 32] | m | $OSO^{3-}$ symm. str. (GAGs) |
| 1031 | [31] | m | Phe C–H in-plane |
| 1004 | [15, 30–32] | s | Phe ring breathing |
| 965 | [30] | w | Amide III′ |
| 940 | [30–32] | s | C–C protein backbone |
| 922 | [15, 30] | sh | C–C Pro |
| 874 | [15, 30, 32] | s | C–C Hypro |
| 857 | [15, 30–32] | sh | C–C aromatic amino acid |
| 817 | [30] | s | C–C protein backbone |

*Abbreviations: s (strong), m (medium), w (weak), sh (shoulder), def (deformation), str (stretching), wag (wagging), Phe (phenylalanine), Tyr (tyrosine), Pro (proline), Hypro (hydroxyproline).