



NRC Publications Archive Archives des publications du CNRC

Collagen morphology and texture analysis : from statistics to classification

Mostaço-Guidolin, Leila B.; Ko, Alex C.-T.; Wang, Fei; Xiang, Bo; Hewko, Mark; Tian, Ganghong; Major, Arkady; Shiomi, Masashi; Sowa, Michael G.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1038/srep02190>

Scientific Reports, 3, 2013-07-12

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=58e777b9-73a7-484e-918d-49d5e07f0b9b>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=58e777b9-73a7-484e-918d-49d5e07f0b9b>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.





OPEN

Collagen morphology and texture analysis: from statistics to classification

SUBJECT AREAS:

COMPUTATIONAL
BIOPHYSICS
BIOPHOTONICS
ATHEROSCLEROSIS
IMAGING TECHNIQUESLeila B. Mostaçõ-Guidolin^{1,2}, Alex C.-T. Ko^{1,2}, Fei Wang¹, Bo Xiang¹, Mark Hewko¹, Ganghong Tian^{1,3}, Arkady Major², Masashi Shiomi⁴ & Michael G. Sowa¹

¹National Research Council Canada, Medical Devices Portfolio 435 Ellice Avenue, Winnipeg, MB, Canada R3B 1Y6, ²Department of Electrical and Computer Engineering, University of Manitoba 75A Chancellor's Circle, Winnipeg, Manitoba, Canada R3T 5V6, ³Department of Physiology, School of Medicine, University of Manitoba 745 Bannatyne Ave., Winnipeg, Manitoba, Canada R3E 0J9, ⁴Institute for Experimental Animals, Kobe University, School of Medicine Kobe 650-0017, Japan.

Received
6 March 2013Accepted
7 June 2013Published
12 July 2013Correspondence and
requests for materials
should be addressed to
A.C.T.K. (alex.ko@nrc-
cnrc.gc.ca)

In this study we present an image analysis methodology capable of quantifying morphological changes in tissue collagen fibril organization caused by pathological conditions. Texture analysis based on first-order statistics (FOS) and second-order statistics such as gray level co-occurrence matrix (GLCM) was explored to extract second-harmonic generation (SHG) image features that are associated with the structural and biochemical changes of tissue collagen networks. Based on these extracted quantitative parameters, multi-group classification of SHG images was performed. With combined FOS and GLCM texture values, we achieved reliable classification of SHG collagen images acquired from atherosclerosis arteries with >90% accuracy, sensitivity and specificity. The proposed methodology can be applied to a wide range of conditions involving collagen re-modeling, such as in skin disorders, different types of fibrosis and muscular-skeletal diseases affecting ligaments and cartilage.

A variety of pathological conditions in humans directly or indirectly involve remodeling or regenerating the collagenous framework in tissue. Some of these conditions are characterized by excessive collagen deposition while others present altered collagen organization (e.g., cirrhosis, scleroderma, keloid, pulmonary fibrosis, diabetes, etc.)^{1–5}. Abnormal deposition of collagen may impair vital functions, and changes in the architecture of the focal collagen network may also lead to disabling conditions^{6,7}. The ability to accurately characterize collagen morphology is therefore an essential component in the pursuit of ultimate understanding of these pathologies.

Traditionally, tissue collagen organization is inspected using histochemistry, immunohistochemistry as well as *in situ* hybridization. These standard methods require multiple steps of tissue processing and such sample preparation can lead to un-desirable morphological alterations in the extracellular matrix. Several imaging methods, such as MRI^{8–10}, small angle X-rays^{11–13}, and electron microscopy¹⁴ were tools developed in the past for direct imaging of collagen without the need for tissue processing. However, these imaging modalities often suffer from low chemical specificity and low spatial resolution. Special experimental conditions are often required for some procedures that can damage tissue structure permanently.

Emerging over the last two decades, second-harmonic generation (SHG) microscopy has become a viable tool for direct visualization of extracellular collagen in bulk tissue without invasive tissue staining^{15–19}. It is a coherent, elastic optical process where two excitation photons are effectively combined in an optically nonlinear medium, to create a new energy-doubled photon at a wavelength exactly half of the excitation wavelength. Because of non-zero second-order generation susceptibility, molecules possessing non-centrosymmetric structures are particularly strong SHG emitters. SHG signal magnitude has a quadratic dependence on incident laser intensity thus allowing for highly localized optical excitation. This results in high-axial and high-lateral resolution comparable to confocal microscopy with added biochemical specificity. In biology, SHG has been extensively used for label-free imaging of membranes and protein fibrils^{20–24}. Collagen I fibril is one such structure and by far the most well-documented source of tissue SHG^{25–27}. For tissue imaging, unlike two-photon excited fluorescence (TPEF), SHG does not suffer from phototoxicity nor photobleaching because there is no net energy deposition in the sample²⁸. Another advantage of using SHG for collagen imaging arises from the use of near-infrared (NIR) excitation. Common NIR wavelengths used for SHG and TPEF imaging range between 780–900 nm leading to extended imaging depths in weakly absorbing but highly-scattering tissue structures while at the same time minimizing thermal effects on the sample.



Despite the success of SHG in biomedical research, most published work relied on SHG to describe collagen organization without using quantitative measures. In most studies, collagen SHG images were presented to describe empirical observations that were linked to a particular pathological condition. While understanding these empirical associations between collagen SHG images and pathology is important, it is equally important to be able to track such correlation using quantifiable measures for objective comparison. To date, quantitative collagen analysis methods have largely relied on image pixel-counting applied to histological images of tissue^{29,30}. Others have also used X-ray diffraction, MRI and electron microscopy images for collagen quantification but with less success^{31,32}. Despite these efforts, none of these methods was able to reliably define distinct collagen patterns based on pre-defined quantitative parameters and there are no reports in the literature to develop quantitative collagen pattern classification. Quantitative SHG imaging has not received much attention among researchers until recently. Several SHG collagen studies have recently proposed new methodologies for quantifying imagery features^{33–40}. These studies however either used simple image pixel intensity-based approaches^{33–37}, or only analyzed histological tissue sections, not intact bulk tissues^{38–40}.

In this study we present a methodology capable of quantifying changes in collagen networks caused by various pathological conditions. The method described fills an existing gap in the literature where empirical interpretation of nonlinear optical microscopy (NLOM) images is used to classify tissue biochemical morphology. We explore the use of texture analysis tools to extract SHG image features that are related to the structural and biochemical changes associated with collagen network pathologies in both sectioned and intact tissues. We also performed multi-group classification of SHG images based on these extracted quantitative parameters.

Two types of collagenous tissue, infarcted myocardium (heart muscle) of rats and atherosclerotic arteries of rabbits, were used to develop the methodology. In the infarcted myocardium, collagen remodeling occurs to repair the damage caused by myocardial infarction⁴¹. A recent study indicated that treatment of infarcted hearts using adipose-derived stem cells (ASCs) influenced the collagen remodeling process by reducing collagen deposition in the infarcted zone⁴². In this study, SHG images of collagen type I fibrils of tissue sections from the stem cells-treated hearts and un-treated hearts were compared using texture analysis. A binary classification based on images acquired from ASCs-treated and un-treated hearts, was subsequently performed. The second type of tissue that was investigated was atherosclerotic plaque obtained from myocardial infarction (MI) prone Watanabe heritable hyperlipidemic (WHHL) rabbits. Atherosclerotic plaques contain mostly collagen types I and III⁴³, with type I collagen comprising approximately two-thirds of the total collagen⁴⁴. Thick type I collagen deposits are particularly abundant in the fibrous cap region of the plaque^{44–46}, overlaying a pool of lipid-rich structures. Type I collagen images collected from bulk, un-sectioned atherosclerotic arteries were analyzed using texture measurements. A multi-group classification was tested subsequently to determine accuracy of the proposed method. The texture-analysis method reported in this study is not limited to the conditions highlighted in this work. In fact, this method can be easily applied to a wide range of conditions involving collagen re-modeling. While others have reported imaging of stem-cell treated infarcted rat hearts with TPEF and SHG microscopy⁴⁷, our study is the first to characterize focal collagen organizational features in an infarcted myocardium.

NLOM, such as SHG microscopy, holds great promises for disease diagnosis because it can provide very specific tissue biochemistry while bypassing conventional histopathological procedures^{48,49}. Such capability would eventually allow for *in vivo* pathology to be realized in the near future. Although clinical NLOM may still be years away from reality, recent technological advances have already

pushed for its miniaturization for endoscopic imaging⁵⁰, and commercialization (*MPTflex*, JenLab, Germany). In order to complement ongoing hardware development, this study is aimed at developing a set of standardized SHG collagen imaging descriptors based on texture parameters for objective analyses of collagen SHG images with minimal subjective interpretation.

Results

Texture analysis of SHG images of collagen remodeling occurred in infarcted rat hearts. A rat model of myocardial infarction was originally developed to study therapeutic potential of adipose-derived stem cells (ASCs) for treatment of heart failure⁴². Histological and MRI studies have both confirmed that ASCs reduced infarct size and improved cardiac contractile function of the infarct rat hearts. In this study, we examined the utility of texture analysis of nonlinear optical images (mainly SHG) for quantitative tracking of the changes related to collagen fibril remodeling in the ASCs-treated and un-treated infarcted rat hearts.

Representative SHG collagen images from un-treated, ASCs-treated infarcted rat hearts and non-infarcted heart are illustrated in Figure 1a, 1b and 1c, respectively. Significant reduction of collagen deposition in the ASCs-treated infarcted heart is clearly seen while the lack of collagen is also evident in non-infarcted heart. This observation is consistent with previous findings⁴² in which ASCs demonstrated therapeutic effect in improving cardiac function via enhancement of cardiomyocyte regeneration and suppression of collagen fibrillogenesis. The collagen density in images Fig 1a and 1b was calculated using a simple threshold pixel counting method. The result shows approximately a 50% drop in collagen density for the ASCs-treated rat hearts compared to the un-treated hearts, which is also consistent with previous data⁴². More importantly, the SHG microscopic images reveal a highly-directional and organized collagen fibril morphology in the ASCs-treated infarct myocardium compared to a less-organized collagen structure characterized by shorter but denser fibril bundles in the un-treated infarcted myocardium. Additionally, representative images of histopathological sections obtained from ASCs-treated, un-treated and control heart are illustrated in Fig. 1d–1f. Compared to un-treated infarcted heart, ASCs-treated infarcted hearts contain a larger percentage of viable myocardium in the infarct zone.

Changes in collagen fibril morphology such as fibril length, fibril thickness, fibril alignment are tracked by texture analysis of SHG images using first-order statistics (FOS) and second-order statistics such as gray-level co-occurrence matrix (GLCM).

FOS textures are directly related to the gray tone distribution of pixels intensity and ignore inter-pixel correlations. In contrast, image second-order statistics depend on the spatial arrangements of pixels intensities present in the region of interest (ROI). It is a measure of the probability of a pair of pixel values occurring at selected distances apart in the image, providing textural information for that region. This probability function is known as co-occurrence matrix. In other words, FOS are parameters extracted directly from the original image while second-order statistics are derived from a matrix (e.g. gray-level co-occurrence matrix) that is built upon inter-pixel correlation of the original image⁵¹.

Figure 2 shows the distribution of values extracted from 90 images for each texture parameter selected from a set of FOS and GLCM parameters. These texture parameters provide a means of capturing and quantifying the morphological changes on these images. The mathematical expression and the interpretation of each texture feature are summarized in a supplemental table (Table S1).

From Fig. 2, it is clear that little statistical difference exists in the FOS parameters between the un-treated and ASCs-treated infarcted rat hearts, according to the Kruskal-Wallis test. GLCM parameters (i.e. IDM, entropy and inertia), on the other hand, show distinction between these two groups. For example, the

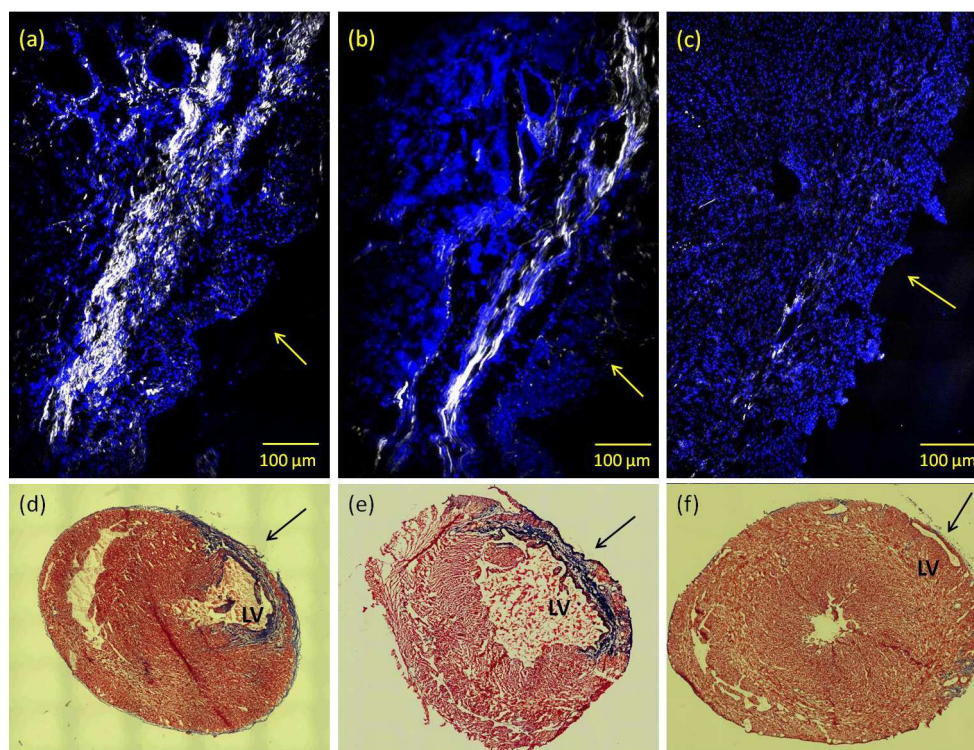


Figure 1 | Co-localized SHG (shown in white, un-labeled) and TPEF (shown in blue, labeled with DAPI) images visualize collagen fibril organization and cardiac muscle cell nuclei, respectively, in the histological section of infarcted myocardium of (a) an untreated infarcted rat heart; (b) an ASCs-treated infarcted rat heart; (c) an image obtained from a histological section of a non-MI rat heart. Images were acquired using $10\times$ 0.45NA dry objective lens. Excitation wavelength is at 800 nm. Collagen SHG signal was collected using a 400 ± 5 nm band-pass filter in the forward direction while the DAPI -TPEF signal was collected in the backscattered (epi) direction through a 505 ± 50 nm filter. Arrows are pointing to the epicardium region. (d), (e) and (f) show representative short-axis histopathological sections of untreated, ASCs-treated infarct rat heart and non-MI heart, respectively. Heart tissue sections were stained with Masson's Trichrome to delineate the infarct region, and images were acquired using $5\times$ objective lens. LV: left-ventricle.

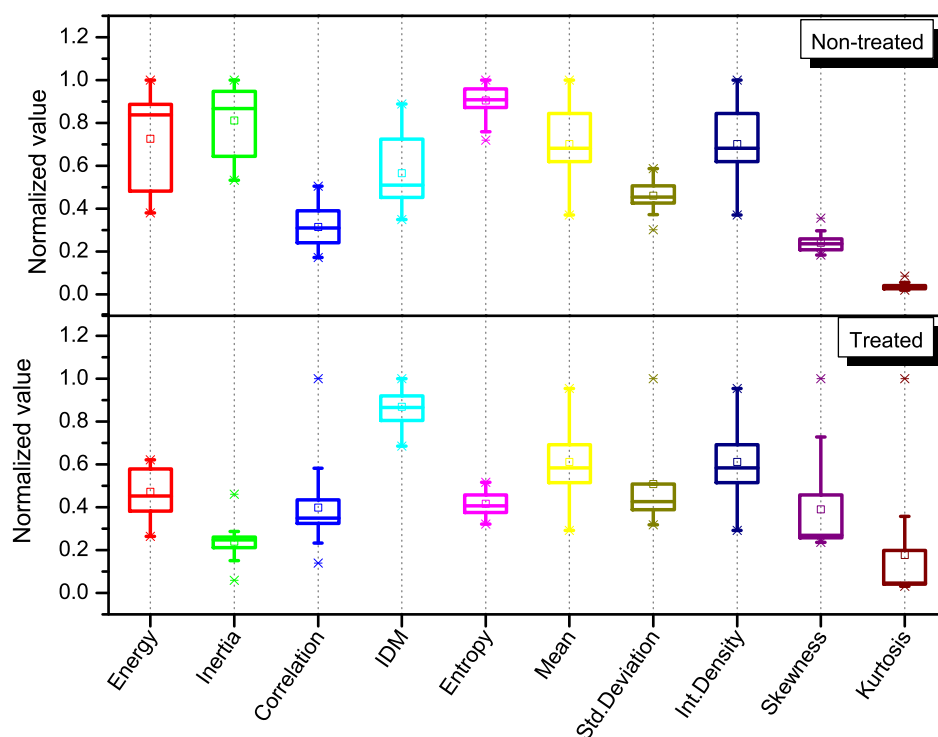


Figure 2 | FOS (mean, standard deviation, integrated density, skewness and kurtosis) and GLCM (energy, inertia, correlation, IDM and entropy) texture parameters extracted from SHG images of the ASCs-treated and un-treated infarcted rat hearts. The top and bottom of each rectangular box denote the 75th and 25th percentiles, respectively, with the median shown inside the boxes. Vertical bars extending from each box represent the 90th and 10th percentiles.

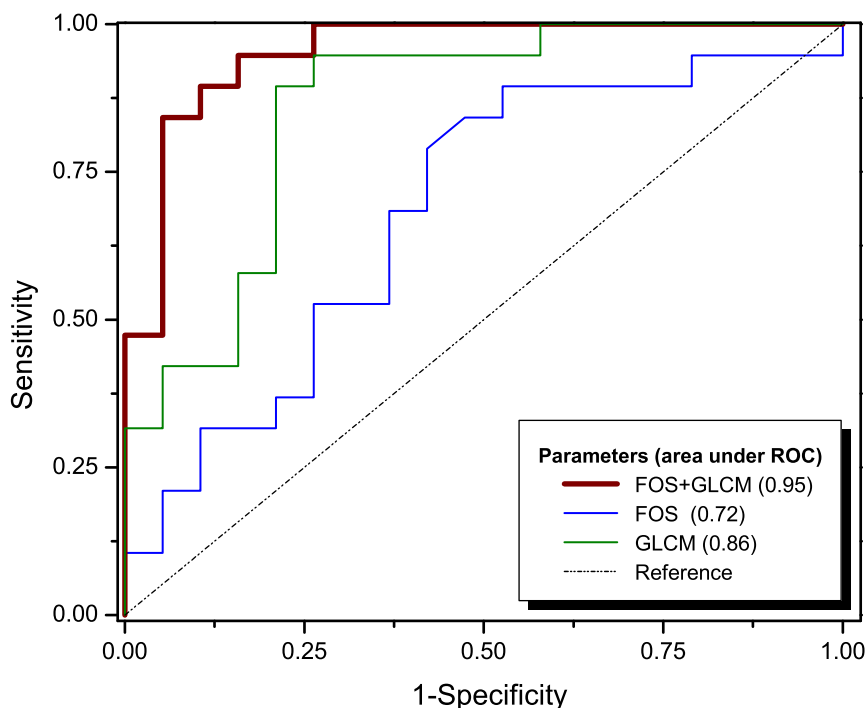


Figure 3 | ROC curves for all three texture sets tested: FOS, GLCM and FOS + GLCM. Values suggest that the classification based on texture analysis has a good predictive value, as the area under the ROC curve was 0.95 for all texture parameters (FOS + GLCM), 0.72 for FOS parameters and 0.86 for GLCM parameters.

group of the ASCs-treated hearts has a statistically higher IDM, lower entropy and lower inertia values compared to its non-treated counterpart. The Shapiro-Wilk test was applied for checking the normality of the data distribution.

Classification of images of the infarcted rat hearts: ASCs-treated v.s. un-treated. Data classification is a supervised learning strategy that is used to analyze the organization and categorization of data into distinct classes. A widely used classifier, support vector machine (SVM), was chosen to test the strength of using FOS and GLCM parameters to differentiate between ASCs-treated from un-treated infarcted rat hearts. SVM is a classification method commonly used in bioinformatics and it is known for its ability to deal with high-dimensional data and flexibility for modeling diverse sources of data^{52–53}. SVM performs classification by constructing an N -dimensional hyperplane that optimally separates the data into different categories. Using a nonlinear SVM classifier, we tested the ability of the texture parameters to classify distinct groups of collagen fibers, based on the first (FOS) and second (GLCM) order features extracted from SHG images. Figure 3 shows the receiver operating characteristic (ROC) curve obtained by combining all texture features (both FOS and GLCM), and those obtained from individual group of features (FOS or GLCM).

The accuracy of the classifier, as represented by the area under the ROC curve was determined to be 0.95 when combined GLCM and FOS parameters were used for classification. This value dropped to, 0.86 and 0.72 when only the GLCM parameters or only the FOS parameters were used, respectively, to classify the infarcted tissue. These values suggest that classification based on combined GLCM and FOS parameters has an improved predictive power in differentiating ASCs-treated from un-treated rat hearts compared using either GLCM or FOS alone.

Texture analysis of collagen fibrils accumulated in atherosclerotic plaques. In order to evaluate the strength and the limitations of the proposed methodology to solve more complex problems, we

performed texture analysis on collagen SHG images acquired from bulk atherosclerotic artery tissues. In previous studies, it has been documented that nonlinear optical imaging microscopy employing TPEF,SHG and coherent anti-Stokes Raman scattering (CARS) could be used to visualize extracellular morphology characteristic of atherosclerotic plaques^{54–60}. Collagen remodeling throughout the progression of atherosclerosis is dynamic, complex and multifactorial. The complexity of this process makes it a good candidate model to assess the power of texture analysis in recognizing various morphological features found in collagen networks.

The images used in this study were acquired directly from the lumen of atherosclerotic arterial segments (un-sectioned bulk) harvested from WHHL rabbits. Figure 4 shows examples of several collagen patterns captured with SHG images along the aorta affected by atherosclerotic plaques. Different patterns in the collagen distribution during plaque development are clearly evident.

A total of 414 SHG images acquired from 14 rabbits were used in the analysis. The age of the rabbits ranged from 2 to 24 months. Based on fibril characteristics such as shape, size and organization, the images were first visually examined and then manually divided into 5 distinct groups by two individuals, independently. The origin of the images was blind to the examiners during this manual classification process. Only those images that received the same categorization from both examiners were used for subsequent texture analysis. Images were segregated into groups A,B,C,D and E, composed of 87, 72, 68, 79 and 108 images, respectively. Representative images from each of the five groups are illustrated in Fig. 5.

Group A images are characterized by thinner, well-defined curled fibrils with a certain degree of fibril orientation. Group B images show a higher focal collagen density and more uniform fibril orientation compared to group A. In contrast, group C and D images both display a more randomly oriented network with shorter fibril length. Group E images are marked by the presence of long, straight and uniformly orientated fibrils. Some level of correlation between distribution of the age groups and that of the 5 image groups were observed. While group A images could be largely correlated with

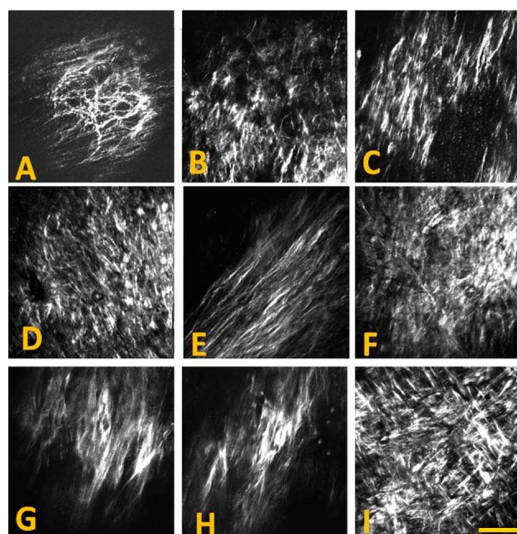


Figure 4 | Epi-SHG images acquired from atherosclerotic plaques on the aorta of WHHL rabbits, showing examples of different collagen fibril morphology detected on atherosclerotic artery. SHG images were acquired using 20×0.75 NA dry objective lens (Olympus) and 800 nm laser excitation. A $2 \times$ digital zoom was used for imaging. Each image shown has 512×512 pixels or approx. $200 \times 200 \mu\text{m}$. Scale bar: $50 \mu\text{m}$.

younger rabbits at an age between 0 and 4 months, collagen features represented in group B–E images were mostly found in rabbits older than 6 months. Furthermore, features shown in group B and C were mainly associated with older rabbits at an age > 16 months.

Five FOS parameters (mean, standard deviation, integrated density, kurtosis, and skewness) were calculated for all 414 images. The distributions of these FOS values corresponding to each group are illustrated in Fig. 6. In 3 FOS parameters, mean, standard deviation, and integrated density, images in group A receive low values compared to the other 4 groups, as a result of the thin and sparse fibrillar organization in this type of plaque. The distribution of skewness and the kurtosis values of the gray level intensity histograms are also illustrated in Figure 6. The skewness indicates the imbalance between

the extent of areas (or number of pixels) that are darker or brighter than the mean. Images containing thicker fibrils, typically from older plaques, present a more asymmetrical pattern left-tailed than that captured in early plaque images (e.g. group A images). Therefore group A images display higher skewness than the other groups.

While skewness is a measure of asymmetry of a distribution relative to the mean value, the kurtosis describes whether distribution of gray tones is more spread-out (flat) or it is more concentrated around the mean (peaked). In advanced plaques, most collagenous networks are better developed, generally thicker and cover larger area therefore generating stronger SHG signals. This usually leads to a more spread-out distribution of pixel intensities, thus lower kurtosis values. Complementary to the five FOS parameters, five second-order GLCM parameters, inverse difference moment (IDM), energy, inertia, entropy and correlation were extracted from the images as well. Their values are presented in Fig. 7. IDM (also called local homogeneity) quantifies the local similarities inside the computational window. It is expected to be higher for GLCMs with elements concentrated near the diagonal. These GLCMs correspond to textures of organized and poorly contrasted features, with only a few gray levels at the same distance d from one another. This parameter quantifies the degree of homogeneity in the region of interest.

The energy feature is sometimes referred to as the second angular moment or uniformity of the GLCM. The lowest value of energy is attained when all the probability density functions $P_{d,0}(i,j)$ are equal, and there are no dominant gray levels. Most gray levels are equally probable.

The inertia texture feature (also called second difference moment) is very sensitive to large differences occurring inside the co-occurrence matrix. Highly contrasted regions will have a high inertia, whereas more homogeneous regions will have a low inertia.

The correlation quantifies the dependence of gray levels between two pixels separated by distance d . Low correlation means that the gray levels are generally independent from one another, i.e., there is no regular structure in the image. However, if correlation is high, there is a high probability that one or several patterns repeat themselves inside the computational window.

The entropy measures the lack of spatial organization inside the computational window. Entropy is high when all $P_{d,0}(i,j)$ are equal,

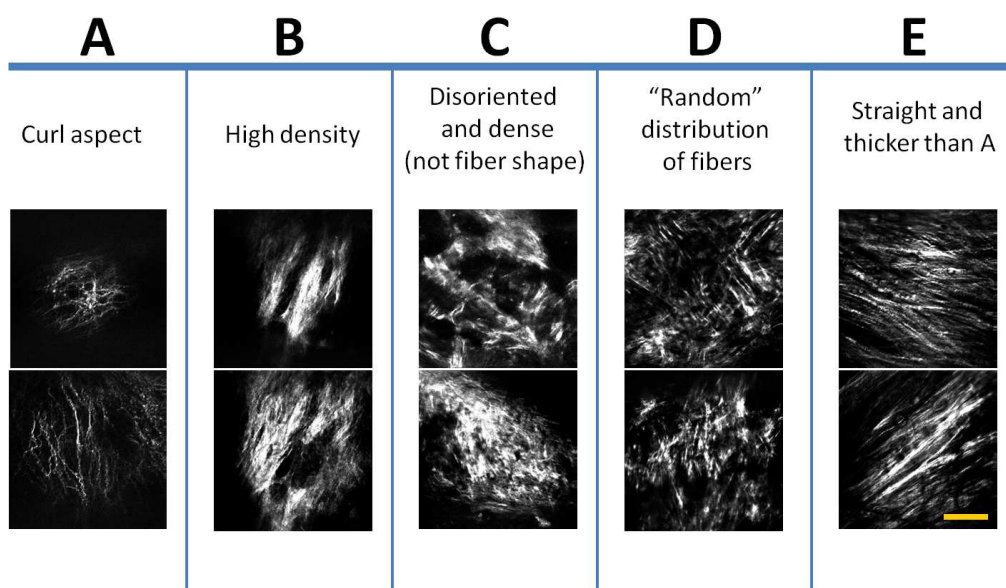


Figure 5 | All collagen SHG images acquired from the arteriosclerotic aortic segments of the WHHL rabbits were classified into five groups A–E. Each group of the images has its own characteristic morphological features such as the fibril’s shape, size and organization. Images are showing the fibrous cap, accumulated closer to the intima layer. Representative images from each group (A–E) are shown. Scale bar: $50 \mu\text{m}$.

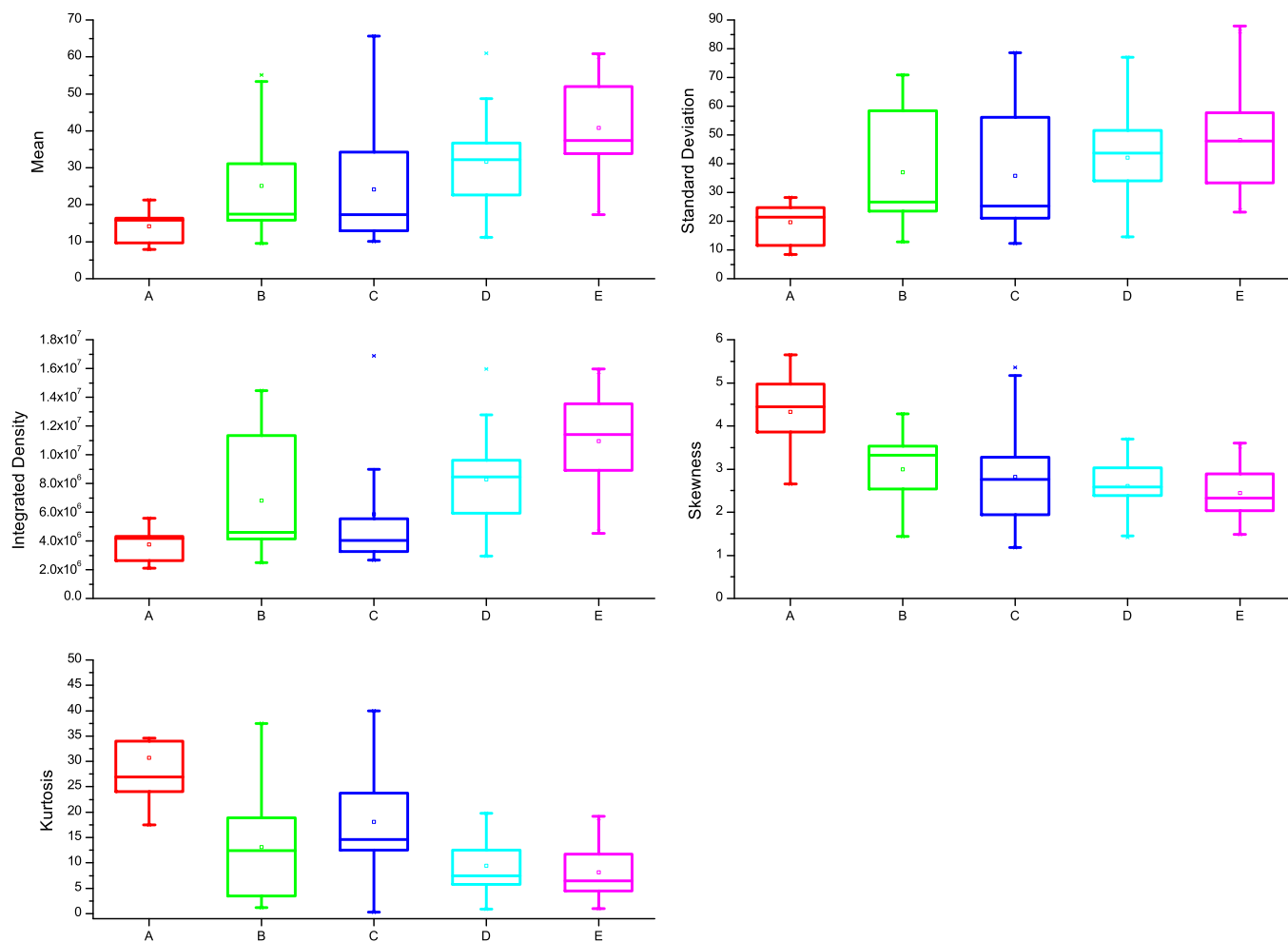


Figure 6 | Comparison of texture parameters calculated for each group defined in the text. The top and bottom of each rectangular box denote the 75th and 25th percentiles, respectively, with the median shown inside the box. Vertical bars extending from each box represent the 90th and 10th percentiles.

which corresponds to a rough texture, and low when the texture is more homogeneous or smoother.

Kruskal-Wallis test was performed to test for statistical differences between any two groups and all p values are shown in the supplemental Table S2. These results indicate insufficient differentiating power for most tested FOS and GLCM parameters in classifying atherosclerotic collagen fibril development, when used individually. One exception is with group A, which stands out as the only group with distinct texture characteristics that can be easily differentiated from the other groups using any of the GLCM parameters, or some of the FOS parameters.

Multi-group classification of SHG images. A nonlinear SVM classifier was trained and applied to classify collagen SHG images of bulk atherosclerotic arterial tissues using FOS and GLCM texture parameter descriptors of the images. This time, the performance of SVM classifier was tested for a 5-group classification. Results are shown in Table I.

The nonlinear SVM classifier has an overall accuracy of 90% when 10 combined FOS and GLCM parameters are included in the classification model. Accuracy drops to 84% and 87% (supplemental Table S3) when only FOS or GLCM parameters were used, respectively. Group A was distinguished from other groups as it shows the highest accuracy, sensitivity and specificity, with little dependence upon the selection of textural parameter (compare Table S2 and Table S3 in the supplementary material). Classification of other

groups, however, is more sensitive to the selection of classification parameters.

Additional tests were also performed to evaluate the effectiveness of the proposed methodology for classification across data sets (atherosclerosis v.s. infarcted heart). All images acquired from the arteries (all five groups) and from infarcted hearts (ASCs-treated and un-treated) were included for analysis. Collagen images collected from artery could be confidently differentiated from those collected from infarcted hearts with 99% accuracy when all texture parameters (FOS + GLCM) were used for classification. Accuracy dropped to 81% when considering only FOS parameters, and 96% when considering only GLCM parameters. Figure S4 (supplemental material) compares each texture parameters calculated for images acquired from atherosclerotic arteries and infarcted myocardium. The obtained ROC curve is presented in Fig. S5 (supplemental material) for each test set.

Overall, classifiers using the full set of calculated texture parameters (combined FOS and GLCM) give the best classification accuracy. When comparing GLCM with FOS parameters, GLCM textures typically give a more accurate classification than classifiers based on FOS textures.

Discussion

In the binary classification of the ASCs-treated and un-treated infarcted rat hearts, our results showed that GLCM parameters were able to provide effective classification with very high accuracy,

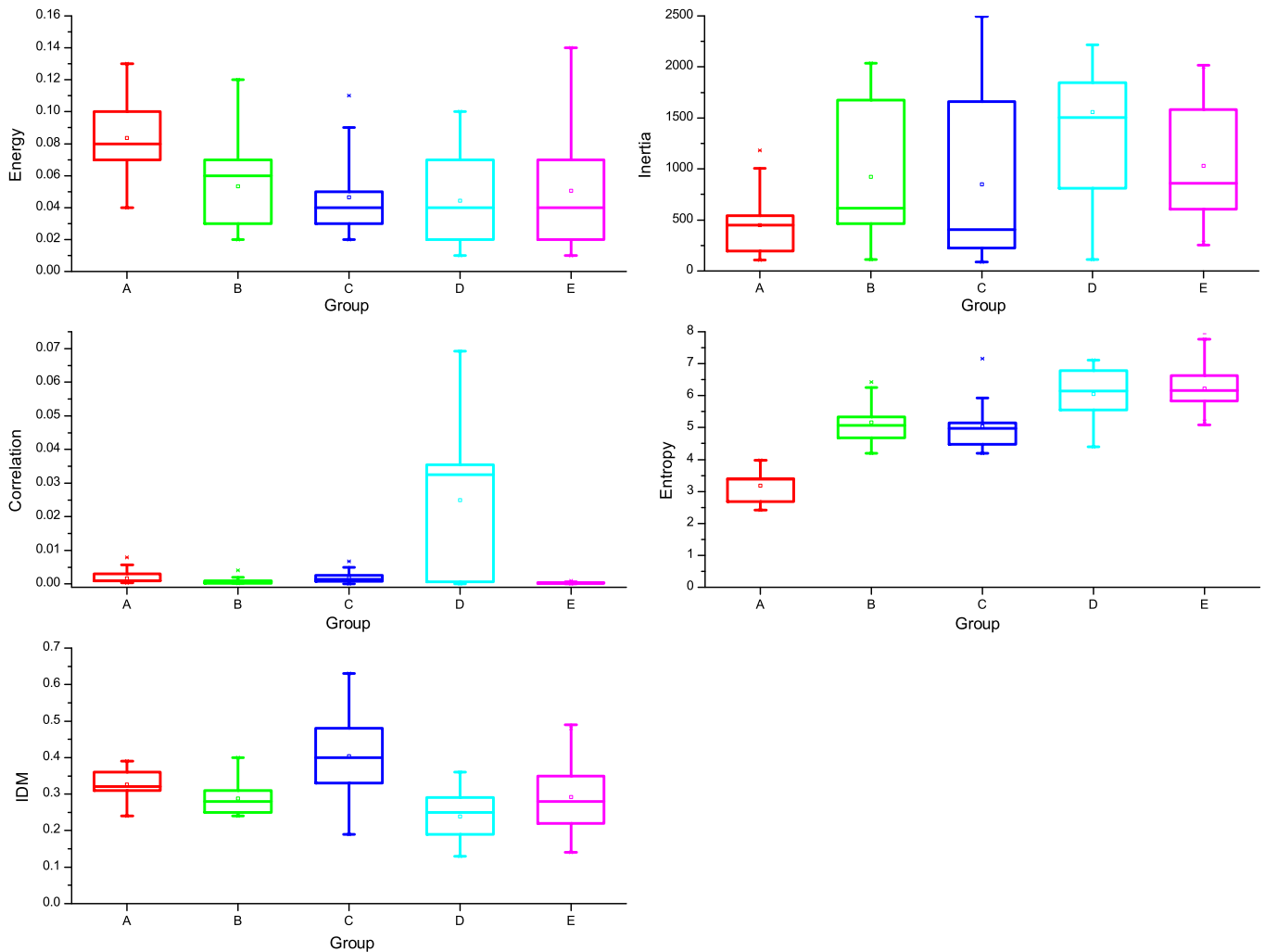


Figure 7 | Values of GLCM texture parameters calculated from each group. The top and bottom of each rectangular box denote the 75th and 25th percentiles, respectively, with the median shown inside the box. Vertical bars extending from each box represent the 90th and 10th percentiles.

conveying both high sensitivity and specificity. From Fig. 2, it is very clear that none of the FOS variables were able to reveal significant differences between the ASCs-treated and un-treated infarcted hearts, suggesting that the intensity-based image features were not sensitive enough to capture the variability of properties of the collagen deposited in infarcted myocardium. GLCM parameters, on the other hand, consider every pixel and its neighborhood, thus generating a map that is able to account for more complex structures in the images.

Three GLCM parameters, IDM, entropy and inertia, all exhibit certain differentiating power. For IDM, a higher value can be associated with a denser or thicker collagen fibril structure in the ASCs-treated hearts. As for entropy, a lower value typically means that the

ASCs-treated hearts have a more homogeneous local collagen morphology than the non-treated hearts. This interpretation is supported by the presence of thicker and better defined fibers in the ASCs-treated hearts (see Fig. 1). Inertia is another texture parameter that quantifies the heterogeneous distribution of collagen in the un-treated hearts. A higher inertia value is often associated with a greater variability of gray levels inside a computational window. In other words, structures and/or particles are distributed within a certain area without a preferential alignment nor orientation, as one would expect in the case of fibrils.

In contrast to data collected from infarcted hearts, the interpretation for data collected from atherosclerotic plaques is not so straightforward due to the more complex and diverse nature of the collagen networks. While it was still possible to link some variations in the FOS parameters directly to certain visual differences between the images, we found it much harder to do the same kind of comparison with the GLCM parameters. As a result, we did not attempt to interpret the direct correlation between the GLCM parameters and the visual presentation of the original images.

Our results showed that GLCM-based classification provided an accuracy at 87%. Classification accuracy was improved modestly to 90% when dimension of the feature space was expanded to include FOS textures. Sensitivity and specificity were however statistically unaffected. This indicates that FOS and GLCM measures are somewhat complimentary to each other in terms of tracking pattern difference in these SHG images.

Table 1 | Summary of the nonlinear SVM classification for collagen SHG images acquired from atherosclerotic arteries based on a 5-group model and combined FOS and GLCM parameters

| Group | Test Set | Accuracy | Sensitivity | Specificity |
|----------------|--------------|------------|-------------|-------------|
| A | 6160 | 100% | 100% | 98% |
| B | 6520 | 73% | 79% | 100% |
| C | 6640 | 88% | 65% | 88% |
| D | 6360 | 90% | 86% | 94% |
| E | 4480 | 93% | 95% | 90% |
| Overall | 30160 | 90% | 91% | 92% |



Based on data presented in Table I, it is quite evident that group A is the most distinct group and can be easily distinguished from the other groups. The collagen textures detected in group A predominantly arise from younger rabbits exhibiting early stage atherosclerosis. The collagen fibers visualized in these images can be characterized by curlier aspect and a sparse distribution without a preferential direction. A closer look at table S2 (supplemental material) also reveals that from a pattern-recognition point of view, group A is the only group that can be separated from the other 4 groups with high confidence. Based on these observations, we can conclude that the collagen features presented in early stage atherosclerotic plaques (rabbits younger than 4 months) are distinctly different from those found in rabbits older than 6 months. This conclusion is consistent with the observation that the progression of atherosclerosis in this rabbit model proceeds faster in the younger animal and then slowly levels off after a certain age.

As for groups B and C, they are mainly characterized by a lack of collagen organization. Interestingly, GLCM textures could not identify groups B as accurately as FOS textures (Table S3, supplementary material). This was one of the reasons why 5-group classification was more accurate when both FOS and GLCM parameters were included in the classification analysis. This demonstrated the benefits of using complimentary texture parameters for evaluating more complex collagen morphologies.

Overall, nonlinear SVM classification shows lower accuracy when only FOS texture parameters are used (73%, 81%, and 92%). This lower accuracy is not surprising given that FOS parameters are directly related to individual pixel intensities and do not consider neighboring values; therefore, different texture patterns can sometimes result in the same FOS value. GLCM parameters, on the other hand, account for certain inter-pixel relationships.

Finally, the methodology was tested for its potential to classify collagen fibril formations in atherosclerotic plaques against those in myocardial infarct hearts, a test across data-sets. Although it was cast as a binary classification problem, individual groups in this test in fact contained much more diversity compared to those presented in the case of the infarcted hearts. With such high level of image inhomogeneity, we were able to obtain a very high classification accuracy at 99% or 96% using combined FOS/GLCM parameters, or GLCM alone, respectively. The FOS texture parameters alone delivered a modest result, at only 81%. Once again we have shown that our methodology is robust in distinguishing structural variations present in collagen network, whether it is for intra- or inter-data set comparison. Secondly, we also proved that this type of collagen feature classification benefits from using a combined FOS/GLCM model.

In summary, using two disease models we have demonstrated the feasibility of performing classification of collagen fibril morphology based on first-order and second-order texture statistical parameters derived from SHG images. Using a nonlinear SMV classifier, it is shown that in more complex cases, the classification accuracy can be improved with combined FOS and GLCM texture variables, compared to the case when either one is used. On the other hand, in a binary classification of ASCs-treated and un-treated infarcted hearts, one group of texture parameters was sufficient to generate classification accuracy of better than 90%. However, when comparing collagen morphology observed in two different data sets (artery v.s heart), the combined FOS/GLCM model is superior to the model that use individual group textures (FOS or GLCM). Although our study is only a proof-of-concept with a limited sample size, its implication is that non-subjective texture based classification of SHG images could have practical clinical applications in distinguishing collagen pathologies. In the clinical world where digital pathology is becoming more popular, automatic classification could be a very helpful tool for the pathologist to increase sample throughput and to help minimize interpretation errors.

Methods

All animal experiments conformed to the guidelines set out by the Canadian Council on Animal Care regarding the care and use of experimental animals and were approved by the local Animal Care Committee of the National Research Council of Canada.

Infarcted hearts: adipose derived stem cells treated and un-treated. Cryo-sections of post-MI rat hearts were obtained from an earlier study which assessed the potential therapeutic efficacy of adipose-derived stem cells (ASCs) on infarcted hearts⁴². In short, myocardial infarction was induced in inbred female Lewis rats by occlusion of the left anterior descending artery (LAD). One week after LAD occlusion, the rats were divided into three groups and subjected to transplantation of ASCs or transplantation of cell culture medium (CCM) or remained untreated. At the end of the 4-wk recovery period, the animals were euthanized and the hearts were excised. ASC-treated, CCM-treated, and untreated hearts were transversely cryosectioned into 6- μ m thick slices from the apex to the base. Cell nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI; Sigma). Technical details on the animal model, cryo-sectioning and histological protocols can be found elsewhere⁴².

Arterial samples. The myocardial infarction prone Watanabe heritable hyperlipidemic (WHHL-MI) rabbits spontaneously develop atherosclerotic plaques due to a hereditary defect in LDL (low-density lipoprotein) processing^{61,62}. This animal model was used previously to study plaque development with nonlinear optical imaging microscopy⁵⁸. The excised aorta was dissected from the ascending aorta to the external iliac artery and then rinsed in heparinized saline. The exterior aorta was subdivided into ~60–80 mm sections that were cut open longitudinally exposing the luminal surface. The samples were placed in petri dishes with the luminal surface facing up on a moist surface and hydration was maintained throughout the measurements by applying PBS solution periodically. Regions of interest were identified prior to SHG imaging measurements. More procedural details can be found in previous work^{58,59}.

Multiphoton microscopy. An in-house, custom-built multiphoton microscope was used for tissue imaging and was previously described⁵⁸. A Ti:Sapphire oscillator (Spectra-Physics, Tsunami) with a center wavelength at 800 nm and a pulse width of 100 fs was used as the laser source for generating SHG and TPEF signals. The laser pulses were first passed through a Faraday isolator (Newport) and pre-compressed using a pair of chirped mirrors (Layertec GmbH, Germany) to compensate for the positive pulse chirping introduced by the microscopy optics. After passing through the various lenses and polarizing optics, the pulses were sent into the microscope assembly where a non-descanned modular type PMT (photomultiplier tube) detector (Hamamatsu) was used for signal detection in either the epi- or the forward direction. The laser pulses were focused onto a sample through a 20 \times , 0.75 NA infinity corrected air objective lens (Olympus) with the SHG signal being collected through the same objective lens. Typically 25 mW of pump and 8 mW of Stokes (measured after the 20 \times air objective lens) were used for imaging. ScanImage software (Cold Spring Harbour Laboratory, NY) was used for laser scanning control and image acquisition.

Image processing and data analysis. Ten texture parameters were determined using the histogram and the gray level co-occurrence matrix (GLCM) of the image. The co-occurrence matrix represents the probability of occurrence of a pixel pair, with a given gray-tone difference, separated by predefined distance taken in a predefined direction, while the image histogram is the frequency of occurrence of a gray tone in an investigated region.

Co-occurrence matrices describe the second-order statistics of the images. In this case statistics depend on the spatial arrangement of the gray levels present in the region of interest and provide textural information for that region. This method is based on the estimation of the second-order joint conditional probability density functions $P_{d,\theta}(i,j)$. Each $P_{d,\theta}(i,j)$ is the probability of going from a gray level i to a gray level j in a given direction θ at a given intersample spacing d .

The co-occurrence matrix $P_{d,\theta}(i,j)$ is a representation of the estimated values. It is a square matrix of dimension N_g (N_g is the number of gray levels in the image). Based on a co-occurrence matrix, a number of texture features can be defined. A set of 14 features, defined by Haralick *et al.*⁶³, were studied for their value in relating SHG images to collagen morphology. These were the angular second moment, inertia, correlation, variance, inverse difference moment (IDM), sum average, sum variance, difference variance, sum entropy, entropy, difference entropy, information measures of correlation one, information measures of correlation two and maximum probability. However, after performing a feature selection test based on Fisher score⁶⁴, only five FOS and five GLCM features were found to be useful in the context of defining structures in collagen fibers from SHG images⁶⁵. Others did not show the ability to distinguish between the various collagen morphologies detected using SHG imaging. The results derived from the energy, inertia, correlation, IDM and entropy texture features are reported.

The co-occurrence matrix was calculated in four orientations: horizontal, vertical and the two diagonals (directions defined by four angles: 0°, 45°, 90° and 135°), and an average value was obtained for characterization purposes. A computational window size of 8 pixels was adopted to extract features from 16 bit images.

Post image processing was performed in ImageJ software. Image background correction, intensity normalization and calculation of various image texture parameters were carried out using Matlab7.5, according to the procedure outlined



previously^{59,66}. A custom-built texture analysis toolkit based on some of the texture analysis functions available in the MATLAB image processing toolbox was used to calculate GLCM parameters. FOS parameters were calculated using ImageJ's histogram analysis toolbox⁶⁷.

Statistical analysis. After all texture features were extracted, the Shapiro-Wilk test was performed to examine the normality of the data distribution. Shapiro-Wilk is one of the most widely used tests for its ability to work with a broad range of data sets. It is also the most powerful normality test available for detecting small variations from normality. The Kruskal-Wallis test was applied to evaluate statistical differences between different groups of data (e.g. ASCs-treated and un-treated hearts). We chose the Kruskal-Wallis test because it is a non-parametric test, not making assumptions about normality and can be applied to examine groups of unequal size.

Nonlinear support vector machine (SVM) classifier. Support Vector Machine (SVM)^{52,53} methods have been shown to be powerful tools for supervised classification. A nonlinear classification SVM (*ksvm* in the R package *kernelab*)⁶⁸ was used to classify SHG images obtained from arteries and hearts. Using the built-in *rbfdot* kernel function in the package, a training set and a validation set were randomly created from the set of SHG images. The training set consisted of 2/3 of the whole data set, while the test set consisted of the 1/3 of the remaining images. The SVM nonlinear classifier was trained and validated using k-fold cross-validation. For experiments involving heart images, k was equal to 3 and for arterial images the used k was 5. A grid search was performed aiming to best define the cost and gamma parameters. The grid search was done on k-fold cross-validation of the training set, and after reaching the maximum possible accuracy, these parameters were then applied to classify the test set. The procedure of defining the cost and gamma parameters was independently repeated for each data set (arteries and hearts).

- Razzaque, M. S. & Taguchi, T. Pulmonary fibrosis: Cellular and molecular events. *Path. Intern.* **53**, 133–145 (2003).
- Friedman, S. L. Molecular Regulation of Hepatic Fibrosis, an Integrated Cellular Response to Tissue Injury. *J. Biol. Chem.* **275**, 2247–2250 (2000).
- Marra, F., Aleffi, S., Bertolani, C., Petrai, I. & Vizzutti, F. Review article: the pathogenesis of fibrosis in non-alcoholic steatohepatitis. *Aliment. Pharm. & Therap.* **22**, 44–47 (2005).
- Varga, J. & Abraham, D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. *J Clin Invest.* **117**, 557–567 (2007).
- Krieg, T., Perlish, J. S., Mauch, C. & Fleischmajer, R. Collagen Synthesis by Scleroderma Fibroblasts. *An. of the New York Acad. of Sci.* **460**, 375–386 (1985).
- Zeichen, J., van Griensven, M., Albers, I., Lobenhoffer, P. & Bosch U. Immunohistochemical localization of collagen VI in arthrofibrosis. *Arch. of Orthop. and Tr. Surg.* **119**, 315–318 (1999).
- Riley, G. The pathogenesis of tendinopathy. A molecular perspective. *Rheumat.* **43**, 131–142 (2004).
- Nissi, M. J. *et al.* Proteoglycan and collagen sensitive MRI evaluation of normal and degenerated articular cartilage. *J. Orthop. Res.* **22**, 557–564 (2004).
- Mahmoud-Ghoneim, D., Amin, A. & Corr, P. MRI-based texture analysis: a potential technique to assess protectors against induced-liver fibrosis in rats. *Radiol Oncol* **43**, 30–40 (2009).
- Chen, J. *et al.* Remodeling of cardiac fiber structure after infarction in rats quantified with diffusion tensor MRI. *Am J Physiol Heart Circ Physiol* **285**, H946–H954 (2003).
- Oi, G. J. *et al.* Fourier transform infrared imaging and small angle x-ray scattering as a combined biomolecular approach to diagnosis of breast cancer. *Med. Phys.* **35**, 2151–2161 (2008).
- Suhonen, H., Fernandez, M., Serimaa, R. & Suortti, P. Simulation of small angle x-ray scattering from collagen fibrils and comparison with experimental patterns. *Phys.Med. Biol.* **50**, 5401–5416 (2005).
- Meek, K. M. & Quantock, A. J. The Use of X-ray Scattering Techniques to Determine Corneal Ultrastructure. *Progr. in Ret.and Eye Res.* **20**, 95–137 (2001).
- Takeda, U. *et al.* Disruption of Dermatopontin Causes Abnormal Collagen Fibrillogenesis. *J. of Invest. Dermat.* **119**, 678–683 (2002).
- Williams, R. M., Warren, R. Z. & Watt, W. W. Interpreting second-harmonic generation images of collagen I fibrils. *Biophysical journal* **88**, 1377 (2005).
- Strupler, M. *et al.* Second harmonic microscopy to quantify renal interstitial fibrosis and arterial remodeling. *J. Biomed. Opt.* **1**, 54041–54041 (2008).
- Pena, A.-M. *et al.* Three-dimensional investigation and scoring of extracellular matrix remodeling during lung fibrosis using multiphoton microscopy. *Microsc. Res. Tech.* **70**, 162–170 (2007).
- Gailhouste, L. *et al.* Fibrillar collagen scoring by second harmonic microscopy: A new tool in the assessment of liver fibrosis. *Journal of Hepatology* **52**, 398–406 (2010).
- Campagnola, P. J. & Loew, L. M. Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nature biotechnology* **21**, 1356–1360 (2003).
- Kissick, D. J. *et al.* Selective Detection of Protein Crystals by Second Harmonic Microscopy. *J. Am. Chem. Soc.* **130**, 14076–14077 (2008).
- Kissick, D. J., Gualtieri, E. J., Simpson, G. J. & Cherezov, V. Nonlinear Optical Imaging of Integral Membrane Protein Crystals in Lipidic Mesophases. *Anal. Chem.* **82**, 491–497 (2002).
- Campagnola, P. J., Wei, M. D., Lewis, A. & Loew, L. M. High-resolution nonlinear optical imaging of live cells by second harmonic generation. *Biophys J.* **77**, 3341–3349 (1999).
- Cohen, B. E. Biological imaging: Beyond fluorescence. *Nature* **467**, 407–408 (2010).
- Prent, N. *et al.* Applications of nonlinear microscopy for studying the structure and dynamics in biological systems. *Proc. SPIE 5971, Phot. Appl. in Nonlinear Opt. Nanophot. and Microwave Phot.* **597106** (2005).
- Georgiou, E. *et al.* Second and third optical harmonic generation in type I collagen, by nanosecond laser irradiation, over a broad spectral region. *Opt. Commun.* **176**, 253–260 (2000).
- Roth, S. & Freund, I. Second harmonic generation and orientational order in connective tissue: a mosaic model for fibril orientational ordering in rat-tail tendon. *J. Appl. Cryst.* **15**, 72–78 (1992).
- Campagnola, P. J., Wei, M. D., Lewis, A. & Loew, L. M. High-resolution nonlinear optical imaging of live cells by second harmonic generation. *Biophys J.* **77**, 3341–3349 (1999).
- Mohler, W., Millard, A. C. & Campagnola, P. J. Second harmonic generation imaging of endogenous structural proteins. *Methods* **29**, 97–109 (2003).
- Rawlins, J. M., Lam, W. L., Karoo, R. O., Naylor, I. L. & Sharpe, D. T. Quantifying Collagen Type in Mature Burn Scars: A Novel Approach Using Histology and Digital Image Analysis. *J. of Burn Care & Res.* **27**, 60–65 (2006).
- George, M. H. *et al.* Injection-site lesions: incidence, tissue histology, collagen concentration, and muscle tenderness in beef rounds. *J. of animal Sci.* **73**, 3510–3518 (1995).
- Han, X., Burke, R. M., Zettel, M. L., Tang, P. & Brown, E. B. Second harmonic properties of tumor collagen: determining the structural relationship between reactive stroma and healthy stroma. *Opt. exp.* **16**, 1846–1859 (2008).
- Chen, X. *et al.* Quantitative analysis of collagen change between normal and cancerous thyroid tissues based on SHG method. In *Phot. and Optoelect. Meetings* **83290H–83290H** (2012).
- Lin, S. J. *et al.* Monitoring the thermally induced structural transitions of collagen by use of second-harmonic generation microscopy. *Opt. lett.* **30**, 622–624 (2005).
- Su, P. J. *et al.* The discrimination of type I and type II collagen and the label-free imaging of engineered cartilage tissue. *Biomater.* **31**, 9415–9421 (2010).
- Lu, K. *et al.* Multiphoton laser scanning microscopy of localized scleroderma. *Skin Res. and Tech.* **15**, 489–495 (2009).
- Lee, P. F., Yeh, A. T. & Bayless, K. J. Nonlinear optical microscopy reveals invading endothelial cells anisotropically alter three-dimensional collagen matrices. *Exp. cell res.* **315**, 396–410 (2009).
- Chen, W. L. *et al.* Multiphoton imaging and quantitative analysis of collagen production by chondrogenic human mesenchymal stem cells cultured in chitosan scaffold. *Tissue Engineering Part C: Methods* **16**, 913–920 (2010).
- Mega, Y. *et al.* Quantification of lamellar orientation in corneal collagen using second harmonic generation images. *Opt. Lett.* **37**, 3312–3314 (2012).
- Guilbert, T. *et al.* A robust collagen scoring method for human liver fibrosis by second harmonic microscopy. *Opt. Exp.* **18**, 25794–25807 (2010).
- Ambekar, R., Lau, T. Y., Walsh, M., Bhargava, R. & Toussaint, K. C. Quantifying collagen structure in breast biopsies using second-harmonic generation imaging. *Biomed. Opt. Exp.* **3**, 2021–2035 (2012).
- Cleutjens, J. P., Verluyten, M. J., Smiths, J. F. & Daemen, M. J. Collagen remodeling after myocardial infarction in the rat heart. *The American journal of pathology* **2**, 325 (1995).
- Wang, L. *et al.* Adipose-derived stem cells are an effective cell candidate for treatment of heart failure: an MR imaging study of rat hearts. *Am. J. of Phys.-Heart and Circ. Physiol.* **297**, H1020–H1031 (2009).
- Shekhonin, B. V., Domogatsky, S. P., Idelson, G. L., Koteliensky, V. E. & Rukosuev, V. S. Relative distribution of fibronectin and type I, III, IV, V collagens in normal and atherosclerotic intima of human arteries. *Atheroscl.* **67**, 9–16 (1987).
- Murata, K., Motayama, T. & Kotake, C. Collagen types in various layers of the human aorta and their changes with the atherosclerotic process. *Atheroscl.* **60**, 251–262 (1986).
- Ooshima, A. & Muragaki, Y. Collagen metabolism in atherogenesis. *Ann. NY Acad. Sci.* **598**, 582–584 (1990).
- Katsuda, S. *et al.* Collagens in human atherosclerosis. Immunohistochemical analysis using collagen type-specific antibodies. *Arterioscler Thromb.* **12**, 494–502 (1992).
- Wallenburg, M. A., Wu, J., Li, R. K. & Vitkin, I. A. Two-photon microscopy of healthy, infarcted and stem-cell treated regenerating heart. *J Biophotonics* **4**, 297–304 (2011).
- Pavlova, I. *et al.* Multiphoton microscopy as a diagnostic imaging modality for lung cancer. In *Proc Soc Photo Opt Instrum Eng* **7569**, 756918–1 (2010).
- Wang, C. C. *et al.* Differentiation of normal and cancerous lung tissues by multiphoton imaging. *Journal of Biomedical optics* **14**, 044034–044034 (2009).
- Bao, H., Boussiotas, A., Jeremy, R., Russell, S. & Gu, M. Second harmonic generation imaging via nonlinear endomicroscopy. *Opt. express* **18**, 1255–1260 (2010).
- Bevk, M. & Kononenko, I. A statistical approach to texture description of medical images: a preliminary study. In *Computer-Based Medical Systems. Proceedings of the 15th IEEE Symposium on* **64**, 239–244, IEEE (2002).
- Saeyns, Y. *et al.* A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507–2517 (2007).



53. Yang, Z. R. Biological applications of support vector machines. *Brief. in Bioinf.* **5**, 328–338 (2004).
54. Zoumi, A., Lu, X., Kassab, G. S. & Tromberg, B. J. Imaging coronary artery microstructure using second-harmonic and two-photon fluorescence microscopy. *Biophys. J.* **87**, 2778 (2004).
55. Lilledahl, M. B., Haugen, O. A., de Lange Davies, C. & Svaasand, L. O. Characterization of vulnerable plaques by multiphoton microscopy. *J. Biomed Opt.* **12**, 044005–044005 (2007).
56. Wang, H. W., Le, T. T. & Cheng, J. X. Label-free imaging of arterial cells and extracellular matrix using a multimodal CARS microscope. *Opt. comm.* **281**, 1813–1822 (2008).
57. Le, T. T., Langohr, I. M., Locker, M. J., Sturek, M. & Cheng, J. X. Label-free molecular imaging of atherosclerotic lesions using multimodal nonlinear optical microscopy. *J. Biomed Opt.* **12**, 054007–054007 (2007).
58. Ko, A. C.-T. et al. Multimodal nonlinear optical imaging of atherosclerotic plaque development in myocardial infarction-prone rabbits. *J. Biomed Opt.* **15**, 020501–020501 (2010).
59. Mostaço-Guidolin, L. B. et al. Differentiating atherosclerotic plaque burden in arterial tissues using femtosecond CARS-based multimodal nonlinear optical imaging. *Biomed. Opt. Express.* **1**, 59–73 (2010).
60. Ko, A. C.-T. et al. Nonlinear optical microscopy in decoding arterial diseases. *Biophysical Reviews* **4**, 323–334 (2012).
61. Shiomi, M., Ito, T., Yamada, S., Kawashima, S. & Fan, J. Development of an animal model for spontaneous myocardial infarction (WHHLMI rabbit). *Arterioscl., thromb., and vasc. biol.* **23**, 1239–1244 (2003).
62. Shiomi, M., Ito, T., Yamada, S., Kawashima, S. & Fan, J. Correlation of Vulnerable Coronary Plaques to Sudden Cardiac Events Lessons from a Myocardial Infarction-prone Animal Model (the WHHLMI Rabbit). *J. Atheroscler. and Thromb.* **11**, 184–189 (2004).
63. Haralick, R. M., Shanmugam, K. & Dinstein, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions* **6**, 610–621 (1973).
64. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *The Journal of Machine Learning Research* **3**, 1157–1182 (2003).
65. Mostaço-Guidolin, L. B. et al. Evaluation of texture parameters for the quantitative description of multimodal nonlinear optical images from atherosclerotic rabbit arteries. *Phys. in Med. & Biol.* **56**, 5319–5323 (2011).
66. Strupler, M. et al. Second harmonic microscopy to quantify renal interstitial fibrosis and arterial remodeling. *J. Biomed. Opt.* **13**, 054041–054041 (2008).
67. Rasband, W. S. "WS 1997–2008." ImageJ. *US National Institutes of Health* (2008).
68. Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. Kernlab-an S4 package for kernel methods in R. *J. Stat. Software* **11**, 1–19 (2004).

Acknowledgments

We acknowledge E.K. Kohlenberg for her assistance in preparing the arterial samples. L.B. M-G also acknowledges financial support from Edward R. Toporeck Graduate Fellowship, Berdie, Irvin Cohen Fellowship, Elizabeth Anne Hogan Memorial Scholarship and University of Manitoba Graduate Fellowship. This work is partially supported by National Research Council Canada, Genomics and Health initiative, and Natural Science and Engineering Research Council (NSERC) Discovery Grant in a form of student financial support. Partial support by an emerging team grant for regenerative medicine and nanomedicine from Canadian Institutes of Health Research (CIHR) is also acknowledged.

Author contributions

L.B.M.-G. conceived of the method, performed imaging measurements and data analyses. A.C.-T.K. designed the experiments, performed imaging measurements and supervised the project. F.W. and B.X. prepared rat heart samples, also conducted histological sectioning and immunohistochemistry study of sectioned heart tissue. M.H. established the rabbit colony, also co-ordinated animal service and imaging works. G.T. supervised the rat study and provided histological interpretations of sectioned rat heart tissue. A.M. co-supervised the project and also provided assistance in imaging protocols. M.S. developed the rabbit model for atherosclerosis and helped M.H. to establish the rabbit colony at National Research Council Canada. M.G.S. performed SVM analyses, and provided guidance on texture analysis of images. L.B.M.-G., A.C.-T.K. and M.G.S. co-wrote the manuscript with input from all other authors.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

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

How to cite this article: Mostaço-Guidolin, L.B. et al. Collagen morphology and texture analysis: from statistics to classification. *Sci. Rep.* **3**, 2190; DOI:10.1038/srep02190 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>