Data in Brief 37 (2021) 107129

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# A comprehensive dataset of histopathology images, grades and patient demographics for human Osteoarthritis Cartilage ☆

Venkata P. Mantripragada [a,*], Nicolas S. Piuzzi [b], George F. Muschler [a,b], Ahmet Erdemir [a], Ronald J. Midura [a]

[a] Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA
[b] Department of Orthopedic Surgery, Cleveland Clinic, Cleveland, OH 44195, USA

ARTICLE INFO

*Article history:*
Received 25 February 2021
Revised 20 April 2021
Accepted 30 April 2021
Available online 14 May 2021

*Keywords:*
Cartilage
Osteoarthritis
Histology
Brightfield imaging
Polarized light images
Image-dataset
Early-osteoarthritis

ABSTRACT

Osteoarthritis (OA) is a leading cause of disability in older adults and takes substantial toll at personal, economic and societal levels. There is inadequate comprehension of OA disease progression specifically during the early phases of OA. This knowledge is critical to understanding the heterogeneity in OA progression as well as enable development of targeted therapeutics at the start of the disease rather than end-stage. Histopathology of cartilage is a common method used to assess in situ state of cartilage tissue. The data presented in this article assesses the histopathological status of human cartilage specimens collected from 90 patients ($n$ = 180). Each specimen was processed for histology and stained with hematoxylin and eosin (HE) and safranin O fast-green (SafO) for acquiring brightfield images to visualize changes in cartilage structure, cells, gycosaminoglycan content and tidemark integrity. The unstained sections were imaged using polarized light microscopy (PLM) to visualize changes in collagen organization and composition within the cartilage specimen. All the specimens were systematically graded by three scorers using established primary OA cartilage grading systems including Histological–Histochemical Grading System (HHGS), advanced Osteoarthritis Research Society

☆ All the work was performed at Lerner Research Institute, Cleveland Clinic

* Corresponding author.
*E-mail address:* mantriv@ccf.org (V.P. Mantripragada).

https://doi.org/10.1016/j.dib.2021.107129

International (OARSI) system and Polarized Light Microscopy (PLM) scoring system. These data can be used by the OA community as an educational resource to train new reviewers (scorers), it serves as a comprehensive image database for experienced OA community to review the wide spectrum of histopathological features presented by these mild to moderate OA specimens, to define different OA-subtypes, and to generate hypothesis on OA progression mechanisms. Finally, the high quality images can be used to develop machine learning algorithms for classification of OA, automated detection and segmentation of existing or new OA features that can serve as early OA histopathological indicators.



## Specifications Table

| | |
|---|---|
| Subject | Health and Medical Sciences (Orthopaedics, Sports Medicine and Rehabilitation) |
| Specific subject area | Histopathological Assessment of Primary Human Osteoarthritis |
| Type of data | Table<br>Image<br>Figures |
| How data were acquired | Leica DM6000 microscope for brightfield images<br>Leica DM4000 B LED for polarized light microscopy images<br>Histological–Histochemical Grading System (HHGS) for histopathology<br>Advanced Osteoarthritis Research Society International (OARSI) system for histopathology<br>Polarized Light Microscopy (PLM) scoring system for histopathology<br>Age, gender and surgery site (left or right total knee arthroplasty) from consented patients at the time of discard tissue collection |
| Data format | Raw images (tif)<br>Histological scores (csv, xml)<br>Patient demographic information (csv, xml) |
| Parameters for data collection | 10x magnification at a resolution of 1.71 μm/pixel for brightfield montage images of Hematoxylin and Eosin (HE) and Safranin O Fast Green (SafO) stained cartilage sections.<br>1.25x magnification at a resolution of 5.16 μm/pixel for polarized microscopy images of unstained cartilage sections.<br>Three independent reviewers for histopathological grading of the cartilage sections using the HHGS, OARSI and PLM systems. |
| Description of data collection | Two osteochondral specimens (4 × 4 × 8 mm) were systematically obtained from the weight bearing portion of the lateral femoral condyle (LFC); one from the medial (CM) and one from the lateral (CL) side to the LFC midline from a total of 90 patients undergoing total knee arthroplasty. Specimens were fixed, decalcified, paraffin embedded, sectioned and stained with HE and SafO for brightfield imaging and unstained section were used for polarized light imaging. Two adjacent sections per stain were obtained for imaging and were referred to when scoring using HHGS, OARSI and PLM systems. |
| Data source location | Cleveland Clinic, Cleveland, Ohio, USA |
| Data accessibility | Data publicly accessible at: https://doi.org/10.18735/77ye-yh24<br>Note: The data is currently hosted on SimTK. The complete database is 27GB. The DOI directs to the page with the Study Title and Description. In order to access the image database, click on the Study Title “Human Knee Cartilage Histopathology Assessment”. At the bottom of this page, user can download the complete image database by clicking on the “Download Archive (27GB)”. The user can choose to login or not to SimTK to successfully download. |

(*continued on next page*)

| | |
|---|---|
| | Another method to selectively download the dataset of interest is by clicking on "Query Data". We have identified 6 variables in the dataset that the user can query on: Patient's age, gender and total knee arthroplasty (TKA) surgery side (left or right), Cartilage histopathology scores including mean total HHGS score, mean total OARIS score and PLM score. To query the data, under "Set Query Rules" dropdown options, choose the variable and the corresponding value. Additional query conditions can be added using "Add rule" or "Add group". Once query conditions are finalized, click on "Search" to identify data subgroups meeting the query requirement. Click on "Get Data" to download selective dataset meeting the query requirements. When downloading queried data, it is advised to create an account on SimTK. Registration is free for any new users and provides additional information on statistics of total downloads. An email is sent to the user, using the user's registered email address, when the packaged data is ready for download. The user can then download the package using the link from the email. |
| Related research article | Mantripragada VP, Gao W, Piuzzi NS, Hoemann CD, Muschler GF, Midura RJ. "Comparative assessment of primary osteoarthritis progression using conventional histopathology, polarized light microscopy and immunohistochemistry", Cartilage, 2020, Online ahead of print. PMID: 32,659,115. DOI: 10.1177/1,947,603,520,938,455 |

## Value of the Data

- To our knowledge, this is the most comprehensive dataset of primary human osteoarthritis (OA) cartilage specimens with about 95% of specimens exhibiting early-mild-moderate OA features rather than late-stage OA features. There is insufficient knowledge and understanding of OA disease progression, particularly during the early phases of OA (due to lack of specimens for observation and study), which are critical to understand and enable development of therapeutics and treatment procedures at the first-stages of the disease rather than end-stage.
- The datasets can be used as an educational resource to train new histopathology reviewers (scorers). In addition, they serve as a foundational image database for experienced OA community to review the wide spectrum of histopathological features presented by these specimens, to define different OA-subtypes, and to generate hypothesis on OA progression mechanisms.
- The high quality images can help with the development of automated image analysis or machine learning algorithms to either identify and score existing OA features or identify new OA features that can serve as early OA histopathological indicators.

## 1. Data Description

Ninety patients (female = 45) with mean age of 62.4 years (range, 37–84) scheduled for total knee arthroplasty (TKA) were recruited over a period of 5 years (2014 to 2018). The relatively preserved lateral femoral condyle (LFC) was collected from these varus knee TKA patients after making the distal femoral cut during TKA, and the anterior-posterior orientation was noted. Two osteochondral specimens ($4 \times 4 \times 8$ mm) were systematically obtained from the weight bearing portion of the LFC; one from the medial ($C_M$) and one from the lateral ($C_L$) side to the LFC midline, by placing the condyle in an in-house fabricated miter box in anterior-posterior orientation [1,2,6]. The two osteochondral specimens were processed for histological staining and assessment using the established HHGS, advanced OARSI and PLM scoring systems [3,4,6]. Table 1 provides a summary of the four scoring systems used and the respective criterias. The summary of overall scores distribution across age groups, gender and surgery side are shown in Figs. 1, 2, and 3. The median HHGS and OARSI scores for the cartilage specimens in this dataset is 5.0,

**Table 1**
Histopathological grading for the primary human osteoarthritis cartilage specimens was performed using four scoring systems: (1) Histological–Histochemical Grading System (HHGS), (2) advanced Osteoarthritis Research Society International (OARSI), (3) Polarized light microscopy (PLM) scoring system developed by Mantripragada et al. [6], (4) Polarized light microscopy (PLM) scoring system developed by Changoor et al. [5]. Details of the various sub-score parameters along with the images required for each scoring system are described below.

| Scoring System | Score Interpretation | Total Score Calculation | Total score Range | Sub-Score Parameters | Sub-Score Ranges | Images Required | Image Resolution in the dataset |
|---|---|---|---|---|---|---|---|
| HHGS | Higher score => More severe disease | Sum of sub-score parameters | 0–14 | 1. Structure | 0–6 | HE | 1.71 μm/pixel |
| | | | | 2. Cells | 0–3 | HE | 1.71 μm/pixel |
| | | | | 3. Safranin O staining | 0–4 | SafO | 1.71 μm/pixel |
| | | | | 4. Tidemark | 0–1 | HE | 1.71 μm/pixel |
| OARSI | Higher score => More severe disease | Product of sub-score parameters | 0–24 | 1. Grade | 0–6 | HE and SafO | 1.71 μm/pixel |
| | | | | 2. Stage | 0–4 | HE and SafO | 1.71 μm/pixel |
| PLM-Mantripragada et al. | Higher score => More severe disease | Sum of sub-score parameters | 0–8 | 1. PLM-SZ collagen organization | 0–4 | PLM | 5.06 μm/pixel |
| | | | | 2. PLM-DZ collagen organization | 0–4 | PLM | 5.06 μm/pixel |
| PLM-Changoor et al. | Higher score => Less severe disease | Only one sub-score parameter | 0–5 | 1. Collagen Organization | 0–5 | PLM | 5.06 μm/pixel |

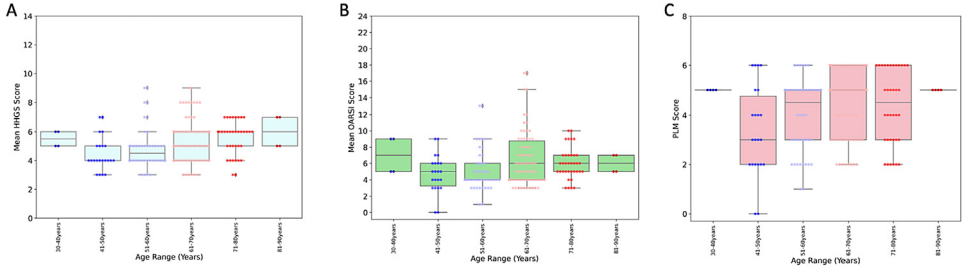


**Fig. 1.** Distribution of histopathological scores for human cartilage as determined by (A) Histological–Histochemical Grading System (HHGS), (B) advanced Osteoarthritis Research Society International (OARSI) system and (C) Polarized Light Microscopy (PLM) system by Mantripragada et al. [6] across different age groups in the 90 patient study cohort.

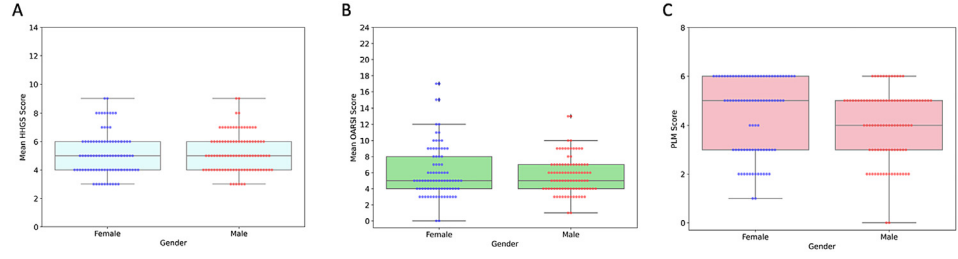


**Fig. 2.** Distribution of histopathological scores for human cartilage as determined by (A) Histological–Histochemical Grading System (HHGS), (B) advanced Osteoarthritis Research Society International (OARSI) system and (C) Polarized Light Microscopy (PLM) system by Mantripragada et al. [6] across the two gender groups in the 90 patient study cohort.

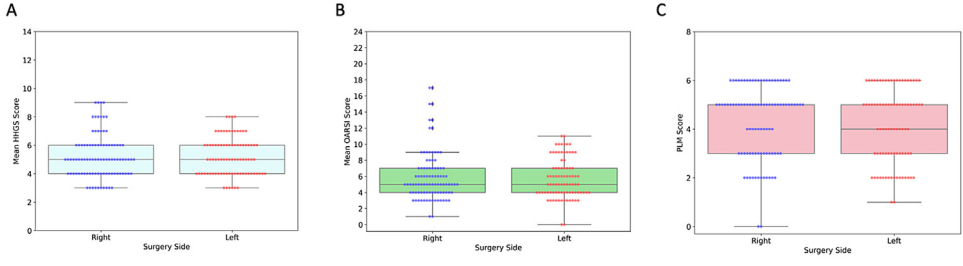


**Fig. 3.** Distribution of histopathological scores for human cartilage as determined by (A) Histological–Histochemical Grading System (HHGS), (B) advanced Osteoarthritis Research Society International (OARSI) system and (C) Polarized Light Microscopy (PLM) system by Mantripragada et al. [6] across the total knee arthroplasty surgery side in the 90 patient study cohort.

suggesting that 50% of specimens exhibit mild OA stage features. There were no specimens with HHGS score of 10 or above and only 2 specimens with an OARSI score above 15.4, suggesting that <3% of specimens with severe or late-stage OA features [2].

The repository contains data organized for each patient (a total of 90 folders). Each patient folder comprises a metadata xml file with information on patient gender, age, surgery side, mean HHGS score, mean OARSI score and PLM score. Each patient folder consists of two more folders, one for Lateral cartilage specimen and another for Medial cartilage specimen. Each of these specimen folders comprise of HE (Hematoxylin & Eosin stained images), SafO (Safranin O Fast Green stained images) folders. Each of the HE and SafO folders comprises of brightfield montaged images of two serial cartilage sections stained with respective histological stains. Additionally, the Medial cartilage specimen folder contains the PLM (polarized light microscopy images) folder

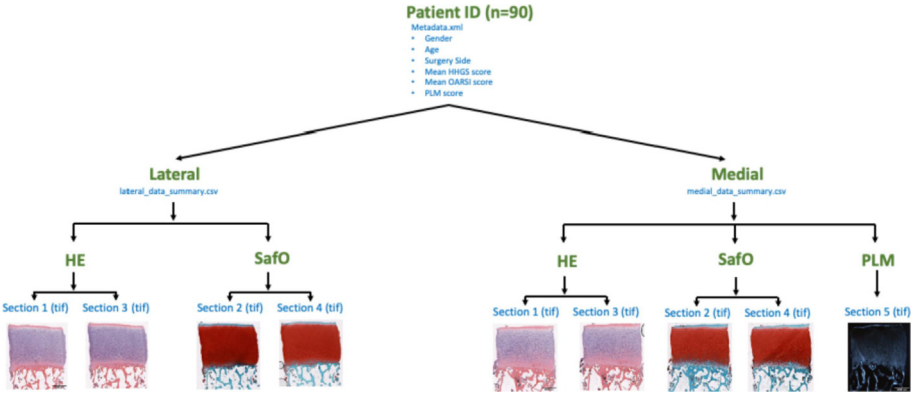


**Fig. 4.** Folder structure hierarchy highlighted in green along with the data files included under each folder highlighted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with a polarized light image of the unstained cartilage section. Fig. 4 illustrates the folder structure, names and data files organization.

Using the images, primary OA cartilage specimens were histopathologically graded using four different scoring systems.

To summarize,

1. Total of 90 patient folders.
2. One metadata.xml file per patient folder. This xml file has patient gender, age, surgery side, mean HHGS score, mean OARSI score and PLM score.
3. Two csv files per patient folder: “lateral_data_summary.csv” and “medial_data_summary.csv”. These csv files have raw data on all the histopathological scores measured in the study by three independent scorers on three occasions (read 1,2,3).
4. Nine image (tif) files per patient folder: Lateral-HE (#2), Lateral-SafO (#2), Medial-HE (#2), Medial -SafO (#2), Medial-PLM (#1).

## 2. Experimental Design, Materials and Methods

### *2.1. Subjects overview*

Patient inclusion criteria included idiopathic OA (primarily medial compartment and/or patellofemoral disease) exhibiting a relatively preserved lateral compartment (Mean Joint space width (JSW): 5.8 mm) based on pre-operative weight-bearing, anterior-posterior (AP) radiographs taken in full extension and 30° of fixed flexion. Patient exclusion criteria included secondary arthritis related to systemic inflammatory arthritis, history of autoimmune disorders, gout or pseudogout, previous surgery to the index knee, current or previous treatment with systemic glucocorticoids or osteotropic medication, cancer within previous 2 years, known or suspected infection; and osteonecrosis.

### *2.2. Human cartilage procurement and processing*

Immediately after surgical retrieval of the LFC in the operating room, the osteochondral specimens were cut and collected in 10% neutral buffered formalin containing 0.5% cetylpyridium chloride (preserves proteoglycan and hyaluronan contents in cartilage tissues) and were fixed for

48 h at 4 °C. Fixation was followed by decalcification for 5 weeks at 4 °C using Cal-rite (ThermoScientific, MA). Decalcification was followed by dehydration in alcohol series into a xylene-substitute and embedding in paraffin with a consistent spatial orientation [1]. Five-micron thick paraffin sections were obtained by cutting the embedded tissue in the plane perpendicular to the surface of the cartilage. For preparing unstained sections for polarized light microscopy (PLM) imaging, sections were deparaffinized using Clear-rite (xylene-substitute), rehydrated in alcohol series followed with water and cover slipped with Cytoseal XYL mounting media. For preparing stained section for brightfield (BF) imaging, sections were stained with hematoxylin and eosin (HE) and safraninO and fast green (SafO) using established protocols. Two adjacent sections per stain were obtained for imaging and were referred to when scoring. Quality of all the sections was confirmed by the senior investigator (scorer 1) before beginning the process of imaging and scoring.

### 2.3. Image acquisition

BF images of the HE and SafO stained sections were obtained on a routinely calibrated Leica DM6000 microscope (Leica Microsystems, Wetzlar, Germany) using color QICam camera (QImaging, Surrey BC, CA). Since the images were acquired over a period of time, every time the microscope was used, white balance and brightfield shading corrections were performed. Using set exposure (1 ms), gain (1x) settings and predictive autofocus, montage (tiled) images were obtained at 10x magnification using a bin factor of 2. Tif images were saved at a resolution of 1.71 µm/pixel. Unstained sections were imaged on a conventional light microscope (Leica DM4000 B LED) with polarizer (positioned in the light path between the specimen and camera) and analyzer (positioned in the light path between the light source and specimen) and a CCD camera (Leica DFC 7000T on C-mount). The analyzer was inserted into the optical path and the polarizer was rotated to 270° with respect to the transmission azimuth, so as to maximize the birefringence signal from collagen fibrils, while keeping the background at minimum light signal intensity. Using a set exposure (15 ms) and gain (2.3x) settings, the image of the complete osteochondral specimen was captured at 1.25x magnification and tif images were saved at a resolution of 5.06 µm/pixel.

### 2.4. Histology scoring

Using HHGS and advanced OARSI scoring systems, HE and SafO stained cartilage sections were graded by three blinded scorers (1, 2, 3): senior investigator (scorer 1), post-doctoral fellow (scorer 2), orthopedic surgery fellow (scorer 3). Prior to initiating the review, scorer 1 was experienced in using HHGS, however scorers 2 and 3 were not. All scorers were inexperienced using OARSI and underwent a training period using self-education and discussions with colleagues in this field to become familiar with the OARSI scoring system. For technical validation, the first fifty patients (100 cartilage specimens) were blindly graded by each scorer (scorer 1,2,3) on three occasions (read1, read2, read3), with each occasion separated by at least a month. After gaining significant experience and observing minimal inter and intra reader variability, each reader scored the remaining 40 patients (80 cartilage specimens) once [1,2]. Using the PLM-Changoor and the new PLM scoring system, the unstained-PLM images were graded once by scorer 2.

Total HHGS score was determined as sum of four sub-score categories: HHGS structure score (score range: 0–6), HHGS cell score (score range: 0–3), HHGS Safranin O staining score (score range: 0–4) and HHGS tidemark score (score range: 0–1) [3]. For scoring structure, cells and tidemark, HE sections were primarily used. Based on the evaluation of both the HE sections (Sections 1, 3), the highest score for each of the sub-scores for a given cartilage specimen was recorded. For instance, if we observed tidemark breach in only one section, a score of 1 was

recorded. For Safranin O staining score, SafO sections were used. Similar to HE, based on evaluating both the SafO sections (Sections 2, 4), the highest score for SafO staining loss for a given cartilage specimen was recorded.

Total OARSI score was determined as a product of OARSI grade (range: 0–6) and OARSI stage (range: 0–4) [4]. Grade represented the severity of the disease observed in the section and stage represented the extent of the severity of the disease. In general, for majority of the OARSI scoring, we first evaluated the SafO sections and then confirmed our grade and stage using HE sections.

For scoring the PLM images, PLM-Changoor system for repair cartilage was used (the only available scoring system in literature for PLM images) [5]. The scores ranged between 0 (totally disorganized cartilage) and 5 (healthy adult cartilage). A new PLM system developed and published by our team for primary OA cartilage was also used [6]. This scoring system has two score sub-categories, superficial zone PLM (PLM-SZ) and deep zone PLM (PLM-DZ) scores, each ranging between 0 (healthy adult superficial zone and deep zone collagen organization) and 4 (total loss of collagen organization).

## Ethics Statement

This study was approved by the Institutional Review Board committee of the Cleveland Clinic (Protocol:13,641).

## CRediT Author Statement

**Venkata P. Mantripragada:** Conceptualization, Methodology, Validation, Investigation, Formal Analysis, Data Curation, Writing – original draft, Visualization; **Nicolas S. Piuzzi:** Methodology, Validation, Investigation, Formal Analysis, Data Curation, Writing – Review & Editing; **George F. Muschler:** Conceptualization, Methodology, Validation, Investigation, Formal Analysis, Data Curation, Writing – Review & Editing, Funding Acquisition; **Ahmet Erdemir:** Conceptualization, Writing – Review & Editing; **Ronald J. Midura:** Methodology, Validation, Investigation, Formal Analysis, Data Curation, Writing – Review & Editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

## Acknowledgments

The authors would like to acknowledge the histological processing skills of Katrina Fife in the Histochemistry Core Facility in the Biomedical Engineering Department (Lerner Research Institute, Cleveland Clinic). This work was funded by NIH R01 AR063733.

## References

[1] V.P. Mantripragada, N.S. Piuzzi, T. Zachos, N.A. Obuchowski, G.F. Muschler, R.J. Midura, Histopathological assessment of primary osteoarthritic knees in large patient cohort reveal the possibility of several potential patterns of osteoarthritis initiation, Curr. Res. Transl. Med. 65 (2017) 133–139, doi:10.1016/j.retram.2017.09.002.

[2] V.P. Mantripragada, N.S. Piuzzi, T. Zachos, N.A. Obuchowski, G.F. Muschler, R.J. Midura, High occurrence of osteoarthritic histopathological features unaccounted for by traditional scoring systems in lateral femoral condyles from total knee arthroplasty patients with varus alignment, Acta Orthop. 89 (2018) 197–203, doi:10.1080/17453674.2017.1398559.

[3] H.J. Mankin, L. Lippiello, Z. A, Biochemical and metabolic cartilage abnormalities in articular cartilage, J. Bone Jt. Surg. 53 (1971) 533–537.
[4] K.P.H. Pritzker, S. Gay, S.A. Jimenez, K. Ostergaard, J.P. Pelletier, K. Revell, et al., Osteoarthritis cartilage histopathology: grading and staging, Osteoarthr. Cartil. 14 (2006) 13–29, doi:10.1016/j.joca.2005.07.014.
[5] A. Changoor, N. Tran-Khanh, S. Méthot, M. Garon, M.B. Hurtig, M.S. Shive, et al., A polarized light microscopy method for accurate and reliable grading of collagen organization in cartilage repair, Osteoarthr. Cartil. 19 (2011) 126–135, doi:10.1016/j.joca.2010.10.010.
[6] V.P. Mantripragada, W. Gao, N.S. Piuzzi, C.D. Hoemann, G.F. Muschler, R.J. Midura, Comparative assessment of primary osteoarthritis progression using conventional histopathology, polarized light microscopy, and immunohistochemistry, Cartilage (2020) 1–17, doi:10.1177/1947603520938455.