Computational pathology measures OA cartilage damage and recapitulates OARSI scoring in a mouse model of tibial loading-induced OA

Richard D Bell[1], Omar Dervisevic[2], Miguel Otero[1,3], Marjolein C H van der Meulen[2]
[1]Hospital for Special Surgery, New York, NY, [2]Cornell, Ithaca, NY, [3]Weill Cornell Medical College, New York, NY

Disclosures: Richard D Bell (N), Omar Dervisevic (N), Miguel Otero (Jannu Therapeutics, LLC - Ownership Interest), Marjolein van der Meulen (N)

Introduction: Osteoarthritis (OA) is the most prevalent orthopedic condition and a debilitating whole joint disease. Irreversible cartilage changes are a hallmark of OA, characterized by early structural and compositional changes localized in the non-calcified superficial cartilage. Pre-clinical models of OA are essential to understanding mechanisms of cartilage damage and developing strategies that prevent disease initiation or progression. Cartilage damage is generally assessed using specialized stains, such as Safranin-O or Alcian Blue, combined with histopathology scoring using the system recommended by the OARSI for mouse joints[1]. This process can be a significant analytical bottleneck because it requires trained scorers and the assessment of multiple sections per joint, particularly in models that lead to focal damage. Our recent work in inflammatory arthritis[2] demonstrated that computational pathology (CPath) methods, or automated image analysis methods, provide accurate histopathology analyses and can efficiently analyze hundreds of slides. Using these approaches, we measured specific pathologic features, correlated these features with histopathology scores, and improved the sensitivity of outcomes to measure smaller effect sizes[2]. Here we seek to implement our CPath pipeline to perform tissue segmentation and measure cartilage damage in a mouse model of load-induced OA and associate these measures with OARSI scoring.

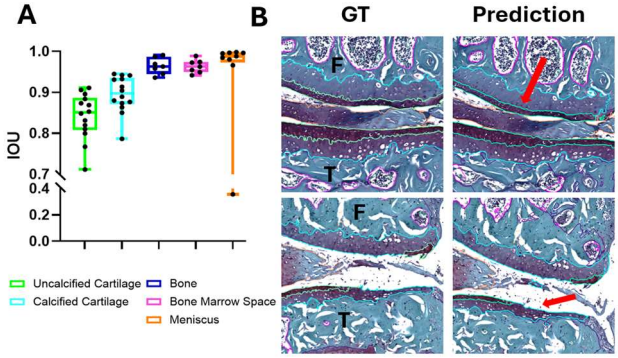

Figure 1. Model Performance. A) IOU of each tissue class and joint. B) Representative images of GT and predictions in healthier (Top) and more disease samples (Bottom). T= Tibia, F= Femur.

Methods: To explore the use of computational histopathology in OA and develop segmentation models, we used a previously published data set of load-induced injury performed with IACUC approval[3]. Briefly, cyclic loading was applied to the mouse tibia (9N, 1200 cycles, 4Hz) for 1, 2, or 6 weeks. After euthanasia, hindlimbs were collected and processed for OARSI scoring, as described[3]. Contralateral limbs were used as controls. A total of n=307 slides from 30 limbs and 26 animals were used for this study. To build our CPath models, 45 slides were hand-annotated for uncalcified cartilage, calcified cartilage, meniscus, subchondral bone, and bone marrow space (ground truth, GT). As high resolution images are often required to appropriately assess damage, we tested 2 different input magnifications (20x and 10x). Segmentation models often benefit from learning adjacent tissues; therefore, we tested both a 2-class model (uncalcified cartilage, calcified cartilage) and a 5-class model (uncalcified cartilage, calcified cartilage, meniscus, subchondral bone, and bone marrow space). This design resulted in 4 different data input structures. A standard 70% training:30% testing data split was performed to re-train, via transfer learning, our previously-developed inflammatory arthritis deep learning UNET++ segmentation model[2]. Model performance was assessed with intersection over union (IOU) and class frequency-weighted IOU. IOU calculates the pixel-wise overlap for each class (intersection) between the GT and predictions and divides by the total area of both the GT and predictions, including the overlap (union), generating a value between 0, no predictive performance, and 1, perfect predictive performance. To validate our segmentation model, we drew a region of interest in the medial tibial compartment and acquired the OARSI scores from the remaining 252 slides. To assess cartilage loss and avoid anatomical location area differences, we calculated an uncalcified-to-total-cartilage-area ratio for each slide. Tibial loading has focal OA damage primarily in the posterior medial tibial compartment; thus, our analyses focused on this anatomic compartment. We averaged uncalcified-to-total-cartilage-area ratio in the posterior-medial compartment. Statistical comparisons were made with Spearman's correlation and ANOVA with a Tukey's post-hoc tests. FDA approval was not required.

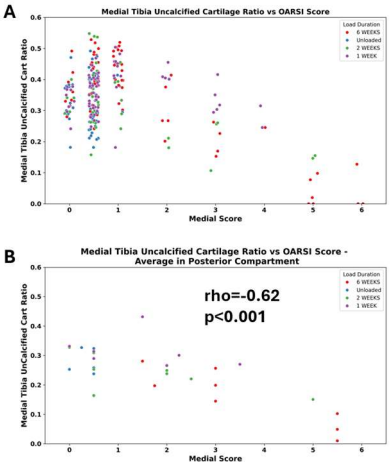

Figure 2. Uncalcified Cartilage Ratio is correlated with OARSI Score. A) Cartilage area was inferred on n=252 slides and plotted against OARSI score. B) The mean in the posterior-medial compartment had a significant negative correlation with the ratio.

Results: While all 4 models performed well with IOU scores ranging from 0.70-0.95, we noticed 1) large amounts of off-target tissue predictions, particularly in the meniscus and at ligament insertion sites, for the 2-class model, and 2) long training/inference times for the high magnification models. Because these off-target predictions and long inference times would be detrimental for downstream analyses, we selected the 5 class - 10x magnification model. This model had an overall excellent class frequency weighted mean IOU of 0.93±0.04 (Fig. 1A). Representative images of GT labels, and predictions had high quality in tracing the tidemark in less damaged samples (Fig. 1B Top, red arrow) and were not confused when the uncalcified cartilage was missing (Fig. 1B Bottom, red arrow). Associations between uncalcified cartilage area ratio and OARSI scores had the expected inverse relationship (Fig. 2A), which reached significance in the posterior compartment (Fig 2B, rho = -0.64, p<0.001). We previously reported an increase in cartilage damage with longer load duration[3]. Our automated CPath model recapitulated this result, with a significant decrease in uncalcified cartilage area ratio at 6-weeks compared to control and 1-week groups (Fig 3).

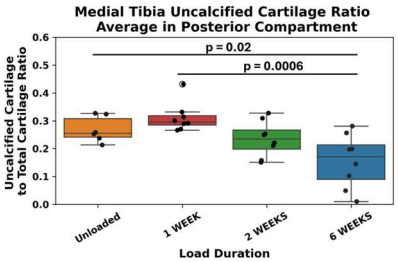

Figure 3. Uncalcified Cartilage Ratio was decreased with longer loading durations, matching results from manual scoring[3].

Discussion: Our automated articular cartilage CPath model was highly performant and assessed damage to uncalcified cartilage equivalent to OARSI scores. However, our composite outcome (uncalcified-cartilage-area-to-total-cartilage ratio) did not fully recapitulate manual scoring and particularly did not differentiate the low score categories (Fig 2A, Fig 3). This limitation likely is because our composite outcome did not include loss of staining (proteoglycan loss) or superficial fibrillation, which contribute to scores associated with early disease. Developing additional metrics using the segmentation outputs is currently ongoing to improve our sensitivity in these low damage categories and improve the accuracy of our models.

Significance/Clinical Relevance: Automated scoring of joint damage will reduce the analytical bottleneck associated with histopathology, decrease the time to hypothesis resolution by improving experimental efficiency and improve reliability and sensitivity of histopathology.

References: 1) Glasson et al, Osteoarthritis Cartilage, 2010; 2) Bell et al, Nat Comms, 2024; 3) Antoinette & Ziemian et al, Sci Advances, 2024