UCSF
UC San Francisco Previously Published Works

Title

Development of a standardized histopathology scoring system for human intervertebral disc degeneration: an Orthopaedic Research Society Spine Section Initiative

Permalink

https://escholarship.org/uc/item/3412g0tm

Journal

JOR Spine, 4(2)

ISSN

2572-1143

Authors

Le Maitre, Christine L
Dahia, Chitra L
Giers, Morgan
et al.

Publication Date

2021-06-01

DOI

10.1002/jsp2.1167

Peer reviewed

SPECIAL ISSUE ARTICLE

JOR Spine OPEN ACCESS

Development of a standardized histopathology scoring system for human intervertebral disc degeneration: an Orthopaedic Research Society Spine Section Initiative

Christine L. Le Maitre[1] 🄳　|　Chitra L. Dahia[2,3] 🄳　|　Morgan Giers[4] 🄳　|
Svenja Illien-Junger[5]　|　Claudia Cicione[6]　|　Dino Samartzis[7,8]　|　Gianluca Vadala[6] 🄳　|
Aaron Fields[9]　|　Jeffrey Lotz[9]

[1]Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, UK

[2]Orthopaedic Soft Tissue Research Program, Hospital for Special Surgery, New York, New York, USA

[3]Department of Cell and Developmental Biology, Weill Cornell Medicine, Graduate School of Medical Sciences, New York, New York, USA

[4]School of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, Oregon, USA

[5]Department of Orthopaedics, Emory University School of Medicine, Atlanta, Georgia, USA

[6]Laboratory of Regenerative Orthopaedics, Department of Orthopaedic and Trauma Surgery, Campus Bio-Medico University of Rome, Rome, Italy

[7]Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, Illinois, USA

[8]International Spine Research and Innovation Initiative, Rush University Medical Center, Chicago, Illinois, USA

[9]Department of Orthopaedic Surgery, University of California at San Francisco, San Francisco, California, USA

Correspondence
Christine L. Le Maitre, Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, S1 1WB, UK.
Email: c.lemaitre@shu.ac.uk

Funding information
Foundation for the National Institutes of Health, Grant/Award Number: S10OD026763; H2020 European Institute of Innovation and Technology, Grant/Award Number: 825925; National Institute of Arthritis and Musculoskeletal and Skin Diseases, Grant/Award Numbers: R01AR065530, R01AR077145; S & L Marx Foundation; Starr Foundation

Abstract

Background: Histopathological analysis of intervertebral disc (IVD) tissues is a critical domain of back pain research. Identification, description, and classification of attributes that distinguish abnormal tissues form a basis for probing disease mechanisms and conceiving novel therapies. Unfortunately, lack of standardized methods and nomenclature can limit comparisons of results across studies and prevent organizing information into a clear representation of the hierarchical, spatial, and temporal patterns of IVD degeneration. Thus, the following Orthopaedic Research Society (ORS) Spine Section Initiative aimed to develop a standardized histopathology scoring scheme for human IVD degeneration.

Methods: Guided by a working group of experts, this prospective process entailed a series of stages that consisted of reviewing and assessing past grading schemes, surveying IVD researchers globally on current practice and recommendations for a new grading system, utilizing expert opinion a taxonomy of histological grading was developed, and validation performed.

Results: A standardized taxonomy was developed, which showed excellent intra-rater reliability for scoring nucleus pulposus (NP), annulus fibrosus (AF), and

---

Abbreviations: AB-PAS, Alcian blue—Periodic Acid Schiff; AF, annulus fibrosus; BEP, boney end plate; CEP, cartilaginous end plate; ECM, extracellular matrix; H&E, hematoxylin and eosin; ICC, interclass correlation coefficient; IVD, intervertebral disc; MRI, magnetic resonance imaging; NP, nucleus pulposus; ORS, Orthopaedic Research Society; UTE, ultrashort echo time.

Christine L. Le Maitre and Chitra L. Dahia contributed equally to this study.

---

cartilaginous end plate (CEP) regions (interclass correlation [ICC] > .89). The ability to reliably detect subtle changes varied by IVD region, being poorest in the NP (ICC: .89-.95) where changes at the cellular level were important, vs the AF (ICC: .93-.98), CEP (ICC: .97-.98), and boney end plate (ICC: .96-.99) where matrix and structural changes varied more dramatically with degeneration.

Conclusions: The proposed grading system incorporates more comprehensive descriptions of degenerative features for all the IVD sub-tissues than prior criteria. While there was excellent reliability, our results reinforce the need for improved training, particularly for novice raters. Future evaluation of the proposed system in real-world settings (eg, at the microscope) will be needed to further refine criteria and more fully evaluate utility. This improved taxonomy could aid in the understanding of IVD degeneration phenotypes and their association with back pain.

KEYWORDS

histopathological scoring, human, intervertebral disc degeneration, standardization

1 | INTRODUCTION

The intervertebral disc (IVD) is a compliant, composite tissue that separates vertebrae within the spine. Its structure and composition are uniquely suited to its biomechanical function, which is to synergize with facet joints, ligaments, and muscles to support spinal compression, shear and torsion forces while facilitating multiaxial motion. IVD degeneration can have a detrimental effect on spinal movement, load sharing with other tissues, catabolic activity, and can ultimately contribute to back pain that can become chronic.[1-5] Furthermore, IVD degeneration may also lead to IVD displacement with subsequent nerve root compression and radiating pain as well as secondary phenotypes of osteophyte formation, endplate abnormalities, Modic changes, IVD space narrowing, facet joint changes, and others. IVD degeneration is often part of the spectrum of degenerative spondylolisthesis and/or spinal stenosis in the older population.[6] IVD degeneration, displacement, and other secondary phenotypes, however, do not just affect the elderly and are common from teenage years into old age. IVD degeneration *per se* has been associated with around 40% of low back pain cases[7]; however, other studies have contended that such IVD changes are purely coincidental with respect to pain. This mismatch further underscores the need to better understand the IVD phenotype that may shed light upon its correlation with clinical features.[8] IVD degeneration is multifactorial and may start at the cellular level, including the formation of nucleus pulposus (NP) cell clusters,[9] senescent,[10,11] or apoptotic cells, caused by, for example, nutrient deprivation due to occlusion of the cartilaginous end plates (CEPs) and boney end plates (BEPs),[12] or could initiate via a structural defect for example, following injury that can cause subsequent cellular changes. The associated extracellular matrix (ECM) degradation can potentially cause a dehydrated NP and weakened annulus fibrosus (AF), which can lead to the formation of fissures and clefts that allow blood vessel and nerve ingrowth,[13-17] and infiltration of inflammatory cells such as macrophages[18] and other immune cells.[19] As such, "discogenic" origins of back

pain are a major socioeconomic concern that affect populations globally and necessitate improved understanding.

Currently, outcomes of chronic back pain management are often unsatisfactory and unpredictable, calling for more precision-based approaches for spine care.[20] In fact, improvement of chronic back pain care is limited by lack of knowledge about degeneration and pain mechanisms at molecular, cellular, and structural levels, further complicated by multiple mechanisms for discogenic pain. Mechanistic insights ultimately form the basis for clinical biomarkers to objectively diagnose painful IVDs, quantify degeneration severity, forecast progression, monitor treatment efficacy, and inform novel therapy development. In this setting, histopathological analyses of IVD tissues from cadaveric spines or surgical samples can be extremely important. However, limitations associated with both tissue sources can restrict the generalizability of findings. For example, cadaveric samples typically lack associated clinical information. Surgically discarded tissues are typically fragments of nucleus, annulus, and bone, from patients with a variety of diagnoses and may not fully represent the back pain population. Additionally, there is an assortment of IVD histopathological methods and classification systems used to assess the severity of conditions and reporting therein. Together, these factors can hinder development of firm conclusions about IVD tissue injury or repair mechanisms. In addition, such limitations of non-standardization can also impact direct comparison of studies due to inherent limitations with language and classification variations.

Previous reports assessing IVD histopathology[12,21-25] have had limitations. For one, previous grading schemes have often been the product of single center investigation, thereby limited in scope with regards to protocol/grade development and external validation. Secondly, reliability of such schemes do not garner community driven consensus. Thirdly, a comprehensive, complete taxonomy of histological features have not been addressed, in particular with a focus of human tissues. In lieu of the above, the Orthopaedic Research Society (ORS) Spine Section Initiative was conceived to address

histopathological phenotyping to facilitate standardization and a common language for widespread utility. This is important in a number of contexts. Standardized and reproducible techniques are critical for confident communication of results and comparisons between studies. Sensitive and reproducible degeneration scoring systems are necessary to clarify disease pathophysiology and progression. Histologic characterization and scoring systems for human IVDs are instrumental for providing context and establishing clinical relevance of pre-clinical studies in animals. The ability to describe degenerative features, particularly those suspected to associate with painful conditions, is fundamental for conceiving new treatment approaches and aligning clinical practice with evidence. As such, the purpose of our following study was to utilize a collaborative process to develop best practice recommendations for consistent processing, identification, nomenclature, and classification of degeneration features within human IVDs. Information obtained could inform models for risk factor identification as well as post-intervention disease progression. Together this will help elaborate on diagnostics, prevention, therapeutics, and outcomes that can further contribute to a more personalized approach to spine care.

To develop a standardized histopathology scoring scheme, our approach was multifaceted (Figure 1). Firstly, an IVD histopathological working group was assembled of recognized key opinion leaders in the field. The group began by reviewing prior classifications systems (stage 1) and surveying the spinal research community who utilized histopathological grading in their research (stage 2). These data were then utilized to develop a taxonomy for histological grading to describe human IVD degeneration (stage 3). We then developed detailed training materials that included descriptions and example images forming 10 "mock" sample IVD image sets (composed of low magnification image of a whole IVD and accompanying high magnification images of features, which could be found in such a representative IVD). These were distributed to a group of spine experts and early career scientists for scoring to provide a preliminary assessment of the new grading system, calculating intra-rater and inter-rater reliability (stage 4), and providing feedback on the usability of the scheme (stage 5).

The resulting scoring system described here is a first step for establishing best practices and methodologies for human IVD grading. We expect this system will undergo continued optimization as it gains use by the wider spine research community, ultimately resulting in a consensus scoring system that can be used worldwide.

2 | STAGE 1: NARRATIVE REVIEW OF HISTORICAL HISTOPATHOLOGIC CLASSIFICATION SYSTEMS OF IVD DEGENERATION

The different IVD sub-tissues, namely the NP, AF, CEP, as well as the adjacent vertebral BEP, each have unique cellular and structural features, differing spatial locations, and varying nutritional and physical stressors. Consequently, the degenerative features vary between these sub-tissues, making it challenging to define one comprehensive grading scheme that incorporates all aspects of IVD degeneration.
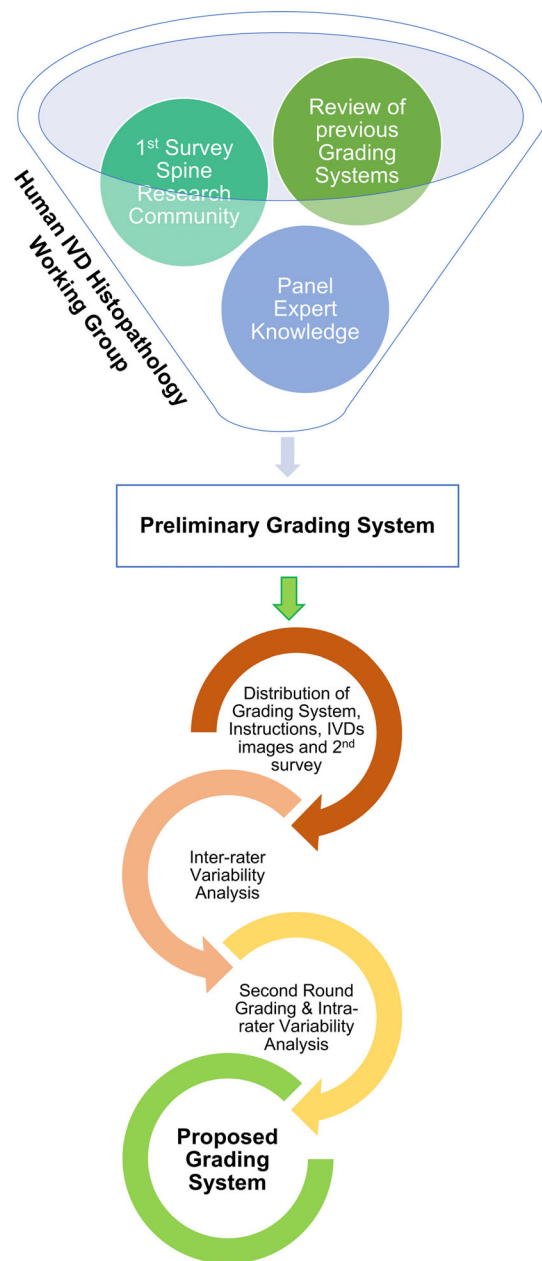
FIGURE 1 Graphical representation of article study design. IVD histopathological working group began reviewing prior classifications, surveying the spinal research community and the knowledge of a panel of expert to develop a preliminary histological grading to describe human IVD degeneration. Detailed training materials, IVDs images, and a second survey were distributed to a group of spine experts. Feedbacks, intra-rater variability, a second-round grading, and intra-rater variability analysis lead to the resulting scoring system for human IVD grading evaluation

2.1 | Methods

Historically used human histopathological grading schemes were identified via a narrative literature search using PubMed and Google Scholar databases. To identify relevant literature, following keywords were used: "intervertebral disc," "grading," "human," "morphology," "surgical," "autopsy." The results where then further refined via

thorough hand evaluation. Only publications that were available in English, had a full text available, and were published in academic journals were included in the study. Articles were excluded when only evaluating micro-CT and/or magnetic resonance imaging (MRI) data, and not describing human IVD morphology, either macroscopically or histologically. After reviewing all selection criteria, six articles developing human IVD grading schemes (Table 1) and nine articles describing morphological changes based on existing human IVD grading schemes (Table 2) were identified.

2.2 | Key findings

Pathological changes of the degenerating IVDs were first reported in 1945[26,27] and since then several grading schemes have been developed to quantify degeneration of human IVD (Table 1). In 1960, Nachemson et al.,[21] reported the first morphologic grading scheme for human IVD autopsy samples at the macroscopic level. Using transverse cut IVDs, the evaluation was based on changes of the NP and AF ranging from grade 1 (no gross changes) to grade 4 (severe structural changes). However, this approach was limited because pathological changes often manifest as horizontal clefts or fissures along the anteroposterior diameter of the IVD and might be missed when assessing the IVD only in the transverse plane.[30] Therefore, degenerative changes are more reliably detected in sagittal sections.[22] In 1990, Thompson et al[12] refined the Nachemson classification based on sagittal plane sections including the CEP and BEP. The Thompson et al classification is still the most widely used method to describe key morphological changes in human IVDs and builds the foundation for several descriptions of morphological features during IVD degeneration (Table 2). Yet, because of limited descriptions of the heterogenous morphological features that associate with degeneration, not all groups adopt previously published grading systems when reporting macroscopic IVD changes (Table 2).

Higher magnification and tinctorial stains are necessary to distinguish between the different IVD components and visualize cells and cell morphology. The first histological grading system was reported by Gries et al,[22] who used hematoxylin and eosin (H&E) staining plus a four-grade classification system, which assessed NP, AF, CEP separately before combining into a single grade. Histological assessment included details about microscopic degenerative changes, such as necrotic cells, chondron formation, changes in ECM composition, invading vascular channels, and minor cleft formation.[22] The disadvantage of this system was that, like the Thompson et al, grading system, it did not fully capture the heterogeneous nature of IVD degeneration (eg, an intact AF but onset of NP degeneration). In 2002, using a combination of several staining methods (H&E; Masson-Goldner; Alcian blue—Periodic Acid Schiff, AB-PAS), Boos et al, described a more detailed scoring system, which scored degeneration of IVD subtissues separately, resulting in separate scoring systems for IVD (0-22) and CEP (0-18).[23] Within the same year, Sieve et al, developed a scoring system specific to NP and AF tissue from surgical samples, which were further profiled at molecular level by *in-situ* hybridization for Sox9, Collagen type II, and immunohistochemistry for Aggrecan.[24] The most recent grading system was described by Rutges et al, in 2013, which utilized three tinctorial stains (H&E, Safranin-O/Fast Green, Picrosirius Red/ Alcian Blue), assessed six features of IVD degeneration separately, and combined them into a single grade by using a scale from 0 to 12. Rudges et al validated their grading system by correlating it to the Boos classification and Thompson grading systems.[25]

While several features are included in all previously published histological grading systems (Figure 2), a consensus about the most appropriate histochemical stain, and a hierarchy of the importance of features to capture the progression of degeneration within each component of the IVD, does not exist. Moreover, only the Boos grading system includes the separate grading of the BEP in their analysis; while none of the grading systems provides a system to grade NP and AF tissue separately, the grading of each region should enable translation to surgical samples where only certain tissues may be present. While there are only four distinct published grading systems,[10,14-16] these share a number of common features (Figure 2), the most

**TABLE 1** Common grading schemes to describe IVD degeneration

Grading classification	Grading range	Method	Stain	Tissue origin	Evaluated tissue	Year	Reference
Nachemson	1 to 4	Macroscopic, unfixed	-	Autopsy (transverse plane)	NP, AF	1960	[21]
Thompson	1 to 5	Macroscopic, unfixed	-	Autopsy (sagittal plane)	NP, AF, CEP, BEP	1990	[12]
Gries	1 to 4	Histological	H&E	Autopsy (sagittal plane)	NP, AF, CEP, BEP	2000	[22]
Boos	IVD: 0 to 22CEP: 0 to 18	Histological	H&E, Masson-Goldner, Alcian blue	Autopsy (sagittal plane) and surgical tissue	NP+AF separate from CEP	2002	[23]
Sive	0–12	Histological	H&E	Surgical tissue	NP+AF	2002	[24]
Rutges	0 to 12	Histological	H&E, SafO, PRAB	Surgical tissue	NP+AF	2013	[25]

Abbreviations: AF, annulus fibrosus; BEP, boney end plate; CEP, cartilaginous end plate; NP, nucleus pulposus.

TABLE 2 Publications that described morphological changes without developing a new grading scheme

Author	Grading method	Method	Stain	Tissue origin	Evaluated tissue	Year	Reference
Coventry	-	Macroscopic	-	Autopsy, sagittal cut	NP, AF, EP	1945	[26]
Friberg & Hirsch	-	Macroscopic, fixed	-	Autopsy, transverse cut	NP, AF	1949	[27]
Vernon-Roberts	-	Macroscopic, fixed	-	Autopsy, sagittal cut	NP, AF, CEP, BEP	1977	[28]
Osti	-	Macroscopic, fixed	-	Autopsy, sagittal cut	NP, AF, EP	1992	[29]
Vernon-Roberts	-	Macroscopic, fixed	-	Autopsy, transverse cut	NP, AF	1997	[30]
Haefeli	Thompson	Macroscopic, fixed	-	Autopsy	NP, AF, EP	2006	[31]
Le Maitre	Sieve	Histological	H&E	Surgical	NP, AF	2005	[32]
Walter	Rutges	Histological	Various stains	Autopsy, transverse cut	NP, AF, EP	2015	[33]
Tomaszewski	Boos	Histological	H&E, Masson-Goldner, Alcian blue-PAS	Autopsy, sagittal, and coronal	NP, AF, EP	2017	[34]

FIGURE 2 Features utilized in published grading systems. Numbers of previously published grading systems for human IVD degeneration (n = 4), which utilize degenerative features. Features classified as whole IVD measures or specific to the nucleus pulposus (NP), annulus fibrosis (AF), cartilaginous end plate (CEP), or the boney end plate (BEP)
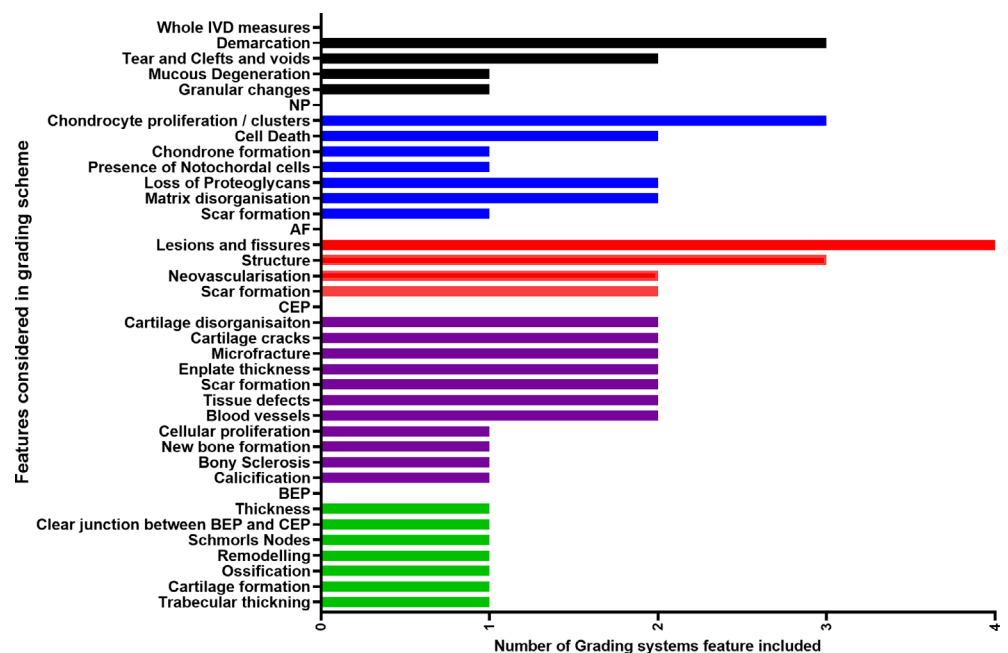

common being presence of lesions or fissures, loss of demarcation between the different tissues of the IVD and the presence of cell clusters within the NP, and changes to the structure of the AF (Figure 2).

2.2.1 | Histopathological features not currently included in prior human IVD scoring systems

In addition to the features identified within prior grading systems, we propose a number of characteristics for the endplate, which is a hard/soft-tissue interface where stresses are concentrated and damage is prevalent.[35] One type of endplate damage is at the annulus/vertebra junction formed by a zone of calcified fibrocartilage (an enthesis) (Figure 3A). During degeneration, the junction between the annulus and fibrocartilage (known as the tidemark) becomes a plane of weakness where clefts can form.[36] These tidemark avulsions are often near innervated, high-intensity zones in the adjacent vertebral rim seen on T2-weighted MRI. Related, the CEP is only loosely adherent to the subchondral bone, and can separate, thereby forming a route of pro-inflammatory crosstalk between the IVD and adjacent vertebra. Bone marrow changes in these areas can be innervated, associated with bone remodeling, be observed on MRI scans (Modic changes), are linked to back pain symptoms, and can be predictive of treatment outcomes.[37-44] Consequently, we have added details to the annulus
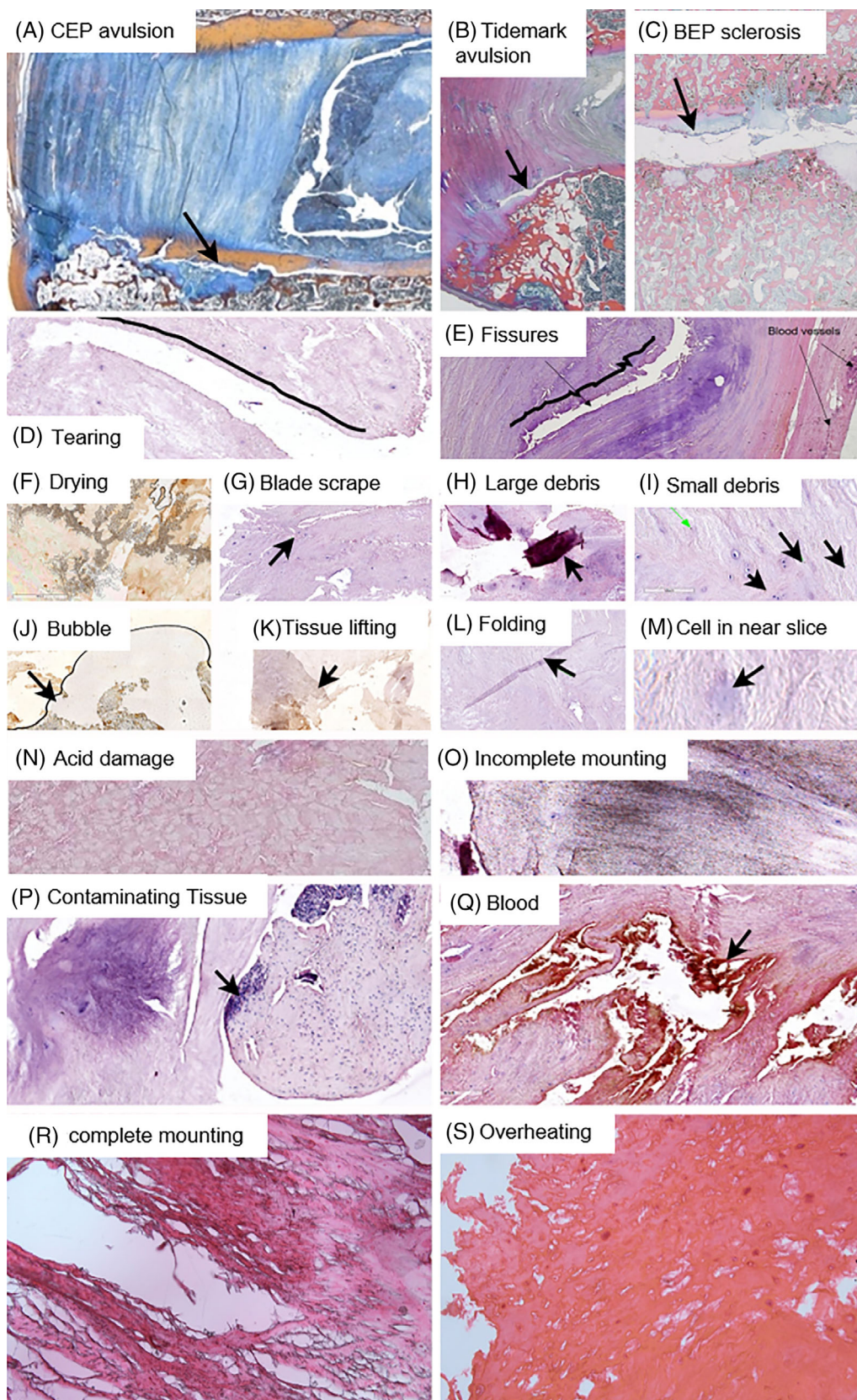
FIGURE 3 Examples of additional characteristics included in the grading system and tissue processing artifacts. A, CEP avulsions are sites where the CEP has separated from the BEP, allowing disc/vertebra cross talk and fibrovascular bone marrow conversion (arrow). B, Tidemark avulsions are clefts at the interface between the annulus and enthesis fibrocartilage (arrow). C, BEP sclerosis refers to densification of subchondral bone and reduction of marrow space. There are many artifacts that may arise during tissue processing. Here are some of common examples including: D, tearing and how to distinguish these from fissures (E); F, drying; G, blade scraping; H, large debris; I, small debris; J, bubbles; K, tissue lifting; L, folding; M, cells in near slices; N, acid damage: O&R, incomplete mounting; P, contaminating tissue; Q, blood; S, overheating. Detailed descriptions of these artifacts can be found in the text

grading to include the tidemark (Figure 3B), and to the CEP and BEP to include avulsions and changes to the bone marrow compartment (Figure 3C).

2.2.2 | Tissue artifacts

Oftentimes, artifacts generated during tissue processing can be misinterpreted as degenerative features. For example, tissue tearing and acid damage could be misinterpreted as degenerative features such as fissures and acellularity (Figure 3D-S). Therefore, it is important to be able to distinguish between real features and those which are introduced during tissue processing and staining.

Tissue artifacts can include:

Tearing vs fissures: Tearing during processing can be mistaken for fissures. When tissue tears during processing or extraction, the edges on either side of the tear will match like a puzzle piece and are both smooth (Figure 3D). In contrast when a fissure occurs the edges of the fissure do not match each other, the edges begin to remodel and become irregular and can often include tissue bridges (Figure 3E). The black line drawn parallel to the edges in each image illustrate the texture difference that is apparent when a tissue begins to remodel (Figure 3E).

Drying: Drying of tissue sections can occur when section of tissue is under a bubble in the resin or when the resin dries out during long-term storage. Drying is particularly prevalent when aqueous mounting medium is used. Dry tissue will appear grey and gravelly (Figure 3F).

Microtome blade scraping vs fissures: During slicing, the microtome blade can occasionally cause a scrape across the tissue. This is visible as a series of small tears in a straight line across the tissue (Figure 3G).

Large debris vs lesion: A region that is out of focus and has a different color than the surrounding tissue, with defined edges, is likely a piece of debris. Lesions will blend into the surrounding tissue and be in focus with the rest of the slice (Figure 3H).

Small debris vs nuclei: There are pieces of small debris and contaminants in most samples. These can be small, dark, or tan spots in the image (Figure 3I). They can be distinguished from cell nuclei, by the lack of lacuna or membrane. Additionally, studying a section of slide that contains no tissue will indicate if the particular slide was particularly dirty.

Bubbles: Bubbles can occur during mounting and appear as out of focus regions surrounded by a black line (Figure 3J).

Tissue lifting: IVD samples can be difficult to adhere to the slide. If a straight edge is seen against a region with much darker stain (similar to a fold), it is an indication that the slice is not adhered to the slide well or that the methods used are causing the tissue to detach (Figure 3K).

Folds: Folds in the tissue can occur during slicing and mounting. This appears as a region with darker staining and unnaturally straight or geometric shape. In addition to the shape, this artifact is distinguishable from color changes due to ECM composition by its defined borders, as opposed to a gradient transition (Figure 3L).

Cells in adjacent slices: Sections are often thin enough that a portion of a cell is visible in the image, but most of the cell is in a serial slice. This is apparent as a region of dark stain that is similar in size to surrounding cells, but contains no nuclei or lacuna (Figure 3M).

Acid damage vs acellularity: In tissue that has undergone acid decalcification, tissue damage is apparent by the presence of many non-nucleated lacunae. This can be distinguished from acellularity due to cell death by the history of the tissue processing and the extent of nuclear absence (Figure 3N).

Contaminating tissue: It is possible to get contaminating tissue in a sample during collection or due to improper cleaning of imbedding and mounting equipment between samples. This could have a variety of appearances (Figure 3P). Samples can also be contaminated by blood during sample collection (Figure 3Q).

Incomplete mounting: When the tissue is not fully mounted, it can lead to a grey appearance. Upon closer examination, small bubbles or protein aggregates can be seen (Figure 3Q&R).

Overheated tissue during processing: Overheating of a tissue sample during processing will lead to small holes in the tissue and ill-defined nuclei and compacted collagen (Figure 3S).

3 | STAGE 2: HUMAN IVD HISTOPATHOLOGICAL SURVEY

We developed a survey in order to capture the needs of the wider scientific community for analyzing human IVD degeneration at the histological level and to garner the communities' opinion on important features that should be incorporated within a grading system, together with an understanding of what groups currently undertake during histology processing.

3.1 | Methods

The distribution and collection of the survey was deemed exempt research by the Corporal Michael J. Crescenz Veterans Affairs (VA) Medical Center Institutional Review Board (Protocol #01862). The study conforms to the US Federal Policy for the Protection of Human Subjects. The survey was based on current published scoring criteria plus potential additional features as described above and was distributed to all ORS spine section members (n ∼ 270) and other spine researchers who were not members of the spine section but have published articles including histological grading (n ∼ 20). We received responses from 38 individuals (note many spine section members do not work with histopathological grading of human tissues and thus were not relevant for this study), representing 29 different institutions from 11 countries and represented the majority of groups publishing within this field. The survey was categorized into sections that included information on the standard operating procedures currently performed within respondents laboratories, together with opinions on what the respondent thought should be utilized in a future grading scheme with particular emphasis on: scoring
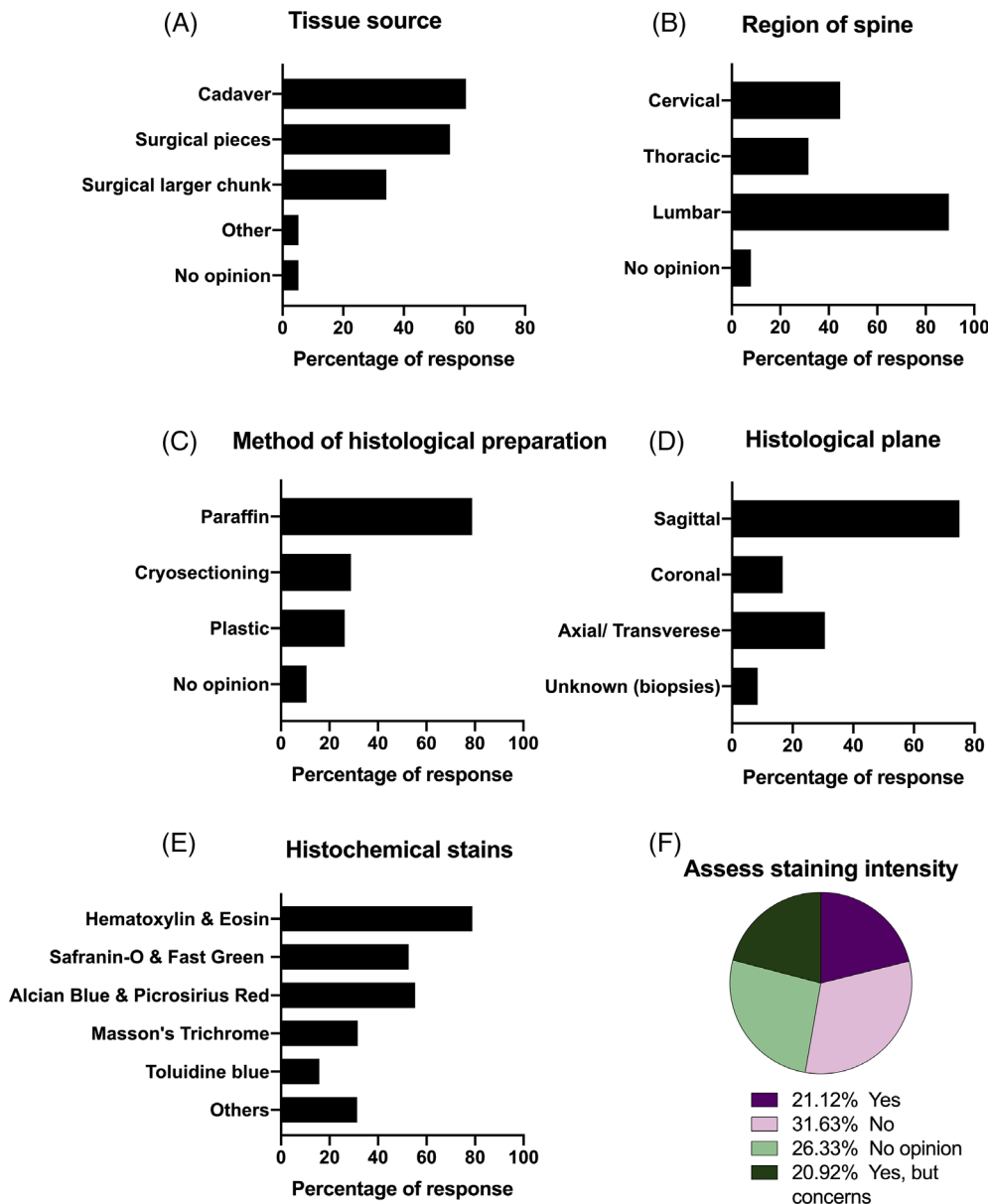
FIGURE 4 SOP for histological preparation of human IVD tissue. Survey data collected from spine researchers (n = 38) show the response in percentage of commonly utilized standard operating procedure for collection and processing of human IVD tissue for histological analysis. Histograms present the response in percentage to multiple-choice question in each category related to source of disc tissue collected (A), region of spine from where tissue is collected (B), methodology for histological preparation of the tissue (C), histological plane in which sections are prepared (D), histochemical staining methods employed for pathological analysis of IVD tissues (E). Pie-chart represents percentage response to close-ended question whether the staining intensity should be assessed for histopathological evaluation (F)

criteria of each IVD sub-tissue (NP, AF, CEP, and BEP); guidance on the scoring range; and whether or not to combine scores from each category to obtain a cumulative score. In addition, sections for additional comments and feedback were also included for each category. The survey data from multiple-choice questionnaires were analyzed for frequencies of response by all survey participants (SPSS 27 (Chicago, Illinois) and Graph pad Prism 9 (San Diego, California).
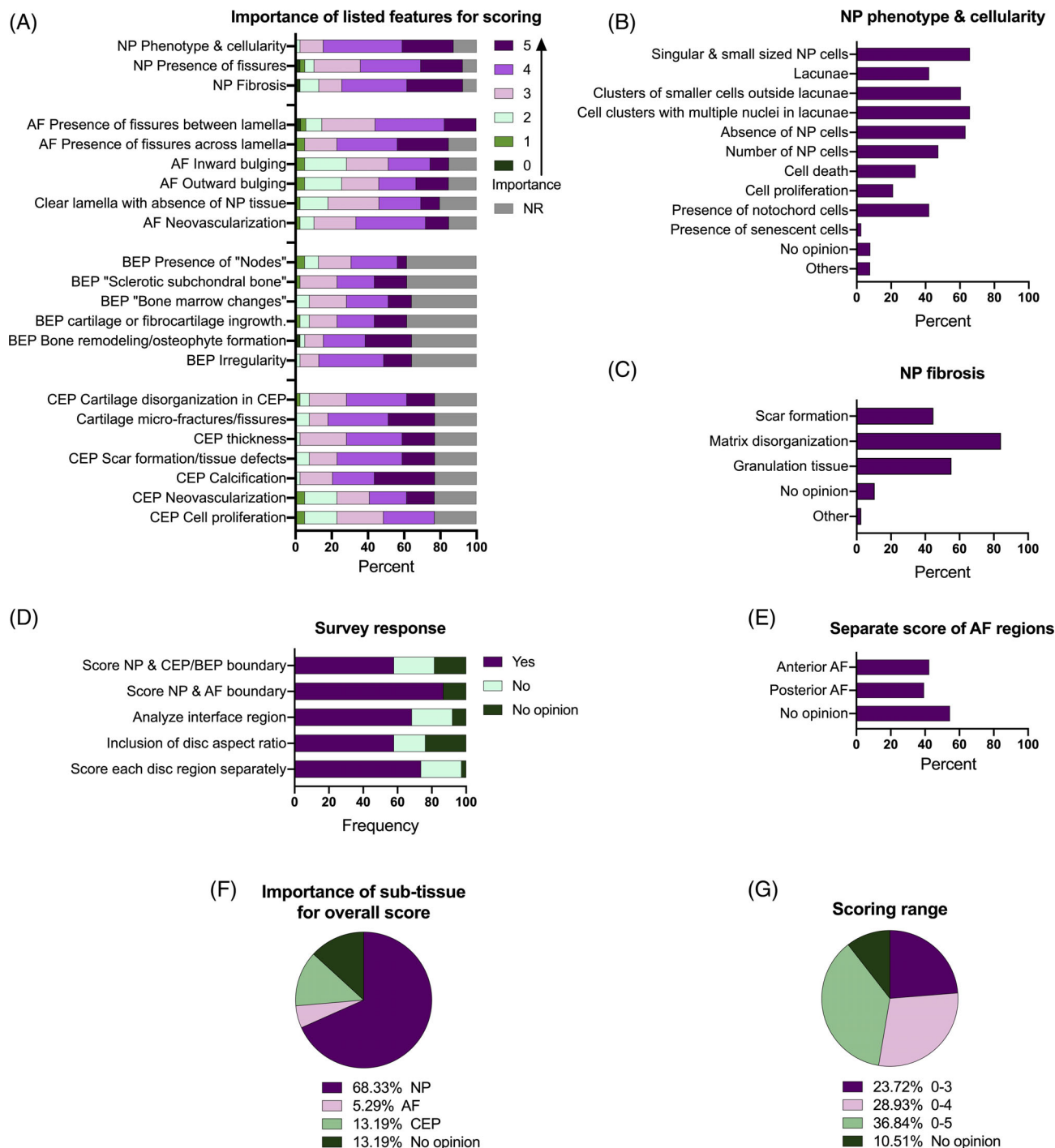
3.2 | Results

Respondents reported that they currently obtained IVD tissue from cadavers (63%) or surgical discard (67%), with 13 individuals reporting access to both tissue sources, 2% did not use IVD tissue, and 2% did not have an opinion (Figure 4). Lumbar IVDs followed by cervical IVD were the most available tissues utilized for research (Figure 4A).

Paraffin embedding followed by cryo-sectioning and finally plastic was utilized for certain applications. Sections between 3 and 10 μm thickness were reported for histological preparation (only one exception of 20 μm) (Figure 4C) The sagittal plane was a prominent choice when analyzing the entire IVD (Figure 4D). H&E was the preferred staining protocol, Safranin-O/Fast green and Alcian blue/Picrosirius Red were other choices for histochemical staining (Figure 4E) (Supplemental file 1—SOPS for staining protocols). Question regarding analysis of the intensity of the histochemical stain for consideration of inclusion in future scoring system was not thought to be a necessary component for histological grading of human IVD tissue (Figure 4F).

The importance of features for histopathological scoring was collected on a six-point Likert scale where least important was scored as 0, and most important was scored as 5. The frequency of response was calculated for each point for all IVD regions; NP, AF, CEP, and BEP (Figure 5A). The features of NP included NP phenotype and

cellularity, "fissures in NP" and "fibrosus of NP" all of which were considered important to include (Figure 5A). Each category was further expanded to capture specific features with most features

considered important to characterize (Figure 5B,C). Seventy-six percent of respondents utilized AF within histological grading systems, with a focus on presence of fissure across and between lamella,

FIGURE 5 Survey of opinion for development of new human IVD scoring system. Survey results show the opinion of spine researchers (n = 38) on the importance of histological features for histopathological assessment of human IVD tissue. Component band chart shows the percentage response for importance of key histological features in NP, AF, CEP, BEP collected on six-point Likert scale from 0 to 5, where 0 represents least important and 5 represents most important (A). Histograms showing the percentage response to multiple choice questions related to grading NP phenotype and cellularity (B) and NP fibrosis (C). Component band chart show the percentage response to close-ended questions for development of the new grading system (D). The percentage response to multiple choice question on grading of AF regions toward pathological scoring (E). Response to multiple choice question regarding importance of IVD sub-tissue (F) and scoring range (G) while development to new histopathological scoring system. The 0 % response to BEP is not plotted in F. NR, not responded in A

neovascularization, discrete lamella with absence of NP tissue, and outward and/or inward AF bulging (Figure 5A). It was also felt that the anterior and posterior AF should be analyzed separately, of course this is applicable to histopathological analysis of the entire IVD (Figure 5E). Sixty-six percent of respondents utilized the CEP within histological grading, with the features for analyzing the histopathological scoring including cartilage disorganization, cartilage microfracture/fissure, thickness, scar formation/tissue defects, calcification, neovascularization, and cell proliferation (Figure 5A). Only 47% of respondents utilized the BEP within histological scoring, with features for histopathological scoring of BEP including sclerotic subchondral bone, bone remodeling, trabecular thickening and osteophyte formation, presence of cartilage or fibrocartilage, bone marrow changes, irregularity of EP, and the presence of nodes (Figure 5A). Further, it was felt important to include features of "Interface regions" to the histopathological scoring including loss of demarcation of NP / AF (87%) and NP and CEP/BEP (60%) boundaries.

Most survey participants (38%) recommended using 0 to 5 for scoring (Figure 5G), although there was a fairly even split in opinion between 0 and 3 (23%), 0 and 4 (29%), and 0 and 5, and it was recommended to separately score each region of the IVD as well as include changes in the IVD aspect ratio. NP tissue was thought to be the most important in quantifying overall IVD degeneration (Figure 5F).

4 | STAGE 3: DEVELOPMENT OF A NEW IVD TAXONOMY FOR HISTOPATHOLOGICAL GRADING

Utilizing the data from the literature review, the survey, and the knowledge and opinions from the authors (Figure 1), a contemporary taxonomy for histological grading of human IVD degeneration was developed that incorporated features that were considered most important in the categorization of human IVD degeneration. IVD regions were separated into the NP, AF, CEP and BEP, and features grouped under the subheadings: (Cellularity, Lesions and ECM structure incorporating the features highlighted in previous scoring systems and ranked important in the survey). A scoring taxonomy was developed for a scoring range of 0 to 3 as the subdivision of features into six criteria as suggested by 38% of survey respondents of 0 to 5 was difficult in practice. Where 0 represents normal morphology and 3 indicates the most severe signs of degeneration (Figures 6–9). Within each grade, descriptive text was developed to describe the features associated with a particular grade. A set of training materials were developed that included the descriptive text plus associated example images, which were submitted from the spine community.

5 | STAGE 4: ASSESSMENT OF THE PROPOSED GRADING SYSTEM

5.1 | Methods

To enable first stage assessment of the proposed grading system, images representing 10 "mock" IVDs were collated using images

supplied of human IVDs collated from the spinal community (Supplementary file 2). Each IVD was represented with a low power image showing the whole IVD and a number of subsequent images to show high magnification regions of the IVD (Figures 10–13). The term "mock" IVD is utilized to highlight that the images provided for each example disc were not necessarily high magnification images of the same IVD but representative of features, which were likely to be identified in such IVDs. These 10 "mock" IVDs together with the grading system and instructions were distributed to 24 spine research labs around the world who distributed the grading system to their students, postdoctoral researchers, technical staff, fellow researchers, and pathologists. All scorers were asked to indicate which images were utilized to score each feature with an overall score provided for each "mock" disc. Independent scoring was completed by 40 observers from 17 different labs around the world with some labs submitting scores from multiple observers. All scorings were performed independently, and no additional training was provided beyond the training materials provided (Figures 6–9). Raters were asked to self-declare themselves as experienced or novel histological grader resulting in 18 experienced graders (eight of which were also authors) and 22 novice graders (one of which is also an author). Data were analyzed according to experience of graders with experienced authors (n = 8), experienced graders (n = 18) and novice graders (n = 22) analyzed independently. In addition, as the method will be used within lab members to analyze data from within labs, the degree of agreement was calculated between raters from the same lab, five cohorts of labs were obtained and analyzed. Inter-rater reliability of the grading criteria and the description of features were tested by interclass correlation coefficient (ICC), confidence intervals, and P-values determined. Type A ICC was calculated using SPSS 27 for an absolute agreement definition and two-way mixed effects model where rater's effects were random and measure effects were fixed, reliability measures were determined as previously reported.[45] For all mock IVDs, the images that were utilized for each grading criteria were recorded and percent scorers utilizing the image plotted (Graph Pad Prism 9). Frequency graphs for submitted grades were generated for each mock IVD using Graph Pad Prism to visually interpret intra-rater reliability and dissect differences between experienced authors (a), all experienced graders (b), and novice graders (c). To assess intra-rater reliability, six raters rescored seven of the mock IVDs, excluding the three IVDs which previously raters were unable to score many features due to lack of images. Intra-rater reliability was assessed using Cohen's Kappa using StatsDirect 3 (Warrington, UK).

5.2 | Results

Initial analysis determined reliability between all raters, and those that were experienced and novice (Table 3). There was excellent reliability (> 0.75) for NP, AF, and CEP regions among all cohorts. However, as the BEP regions were not uniformly scored, the test could not be executed for total and experienced raters. The reliability for BEP was excellent among the novice raters (Table 3). Intra-rater reliability

NP Features	Sub Features included	Score			
		0 Degenerative features: absent	**1** Degenerative features: few, small	**2** Degenerative feature: moderate	**3** Degenerative feature: abundant, large
Cellularity	Single cells in lacunae, cell clusters, necrosis, apoptosis, cell shrinkage, senescence (large cells), acellularity	Single cells in lacunae, no evidence of necrotic, apoptotic or senescent cells	<25% Cells in clusters in lacunae or multiple nuclei in lacunae, may be some evidence of apoptosis (nuclei condensation), cell shrinkage.	25-75% of cells in clusters, May be evidence of apoptosis (cell shrinkage), necrosis (cell debris) and cellular senescence (large irregular cells), may be areas of evidence cell loss (empty lacunae).	>75% in clusters, evidence of empty lacunae, or may be evidence of apoptotic, necrotic and senescent cells, hypercellularity.
Lesions	Tears, clefts and voids	Normal intact NP with intact ECM	Micro fissures present within localised areas of the NP	Micro fissures across the majority of the NP or may be localised larger clefts and voids	Presence of large clefts and voids within the NP and extensive fissures.
ECM structure	Loss of Proteoglycan staining, mucoid degeneration, fibrosis, clear demarcation of the NP or presence of AF or CEP tissues in the NP	Clear ECM structure, Eosin staining present throughout the NP, if using PG stain clear staining can be seen throughout the NP. Matrix structure does not contain lamella or ordered collagens. Clear demarcation seen at junctions of NP and AF and NP and CEP.	May be some loss of demarcation and limited loss of Eosin staining in proximity to the NP cells.	Presence of demarcation within the NP - noted by areas of structured collagens or fibrosis within the NP. Loss of Eosin staining around the proximity of cells but most matrix staining intact.	Substantial demarcation, or fibrosis, may be evidence of mucoid degeneration and loss of eosin staining which is more widespread than the immediate pericellular locations.
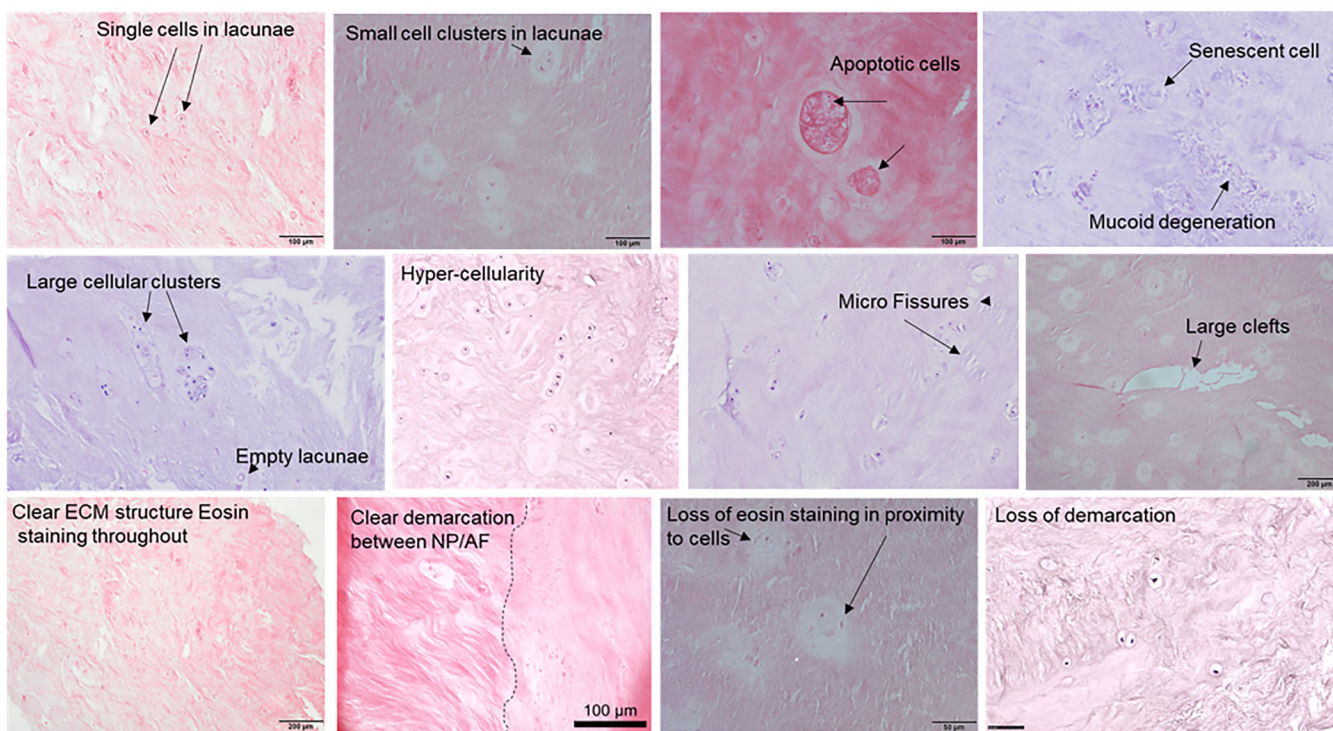
FIGURE 6 Taxonomy of grading for nucleus pulposus features. Descriptive text for features utilized for the grading (0-3) of the nucleus pulposus. Grading criteria broken down into cellularity, lesions and extracellular matrix (ECM) structure. Example images shown to demonstrate: single cells in lacunae, small cell clusters in lacunae, apoptotic and senescent cells, mucoid degeneration, large cellular clusters and hypercellularity, micro fissures and large clefts, clear ECM structure and demarcation between the NP and AF, loss of eosin staining in proximity to cells, and loss of demarcation

within lab members was calculated utilizing five lab cohorts with varying numbers of raters that were either experienced or novice (Table 4). The results indicate excellent reliability (> 0.75) for all features when the experienced raters are more than the novice raters. The ICC was mixed, excellent for some features and moderate (<0.75 and >0.04) to poor (<0.04) for other features when the novice graders were more than the experienced graders in a cohort.

For some IVDs, it was noted that scorers utilized different images for scoring that could in part explain the variation seen with clear examples seen of differential images used linking to poorer grade

AF Features	Sub Features included	Score			
		0 Degenerative features: absent	1 Degenerative features: few, small	2 Degenerative feature: moderate	3 Degenerative feature: abundant, large
Cellularity	Acellularity, cell death, change in the OAF cells from elongated to more rounded, Neovascularisation	Cellular phenotype is normal - i.e. elongated cells within the outer AF progressing to rounded morphology within the IAF, clear transition of cell morphology. Limited blood vessels present which are constrained to the outer 2-3 lamella of the OAF.	May be evidence of blood vessel ingrowth through the OAF, Some change in cellular morphology and potential evidence of small numbers of apoptotic/necrotic cells and empty cell lacunae.	Mixed cell morphologies observed throughout the AF with AF cell proliferation at IAF. Blood vessel ingrowth into the interface region between OAF and IAF. Evidence of apoptosis and necrosis.	Morphology of cells disrupted with mixed cell populations seen through the AF, or a loss of cellularity present with evidence of apoptosis and necrosis and potentially senescent cells. Blood Vessel ingrowth seen into the IAF.
Lesions	Lesions across and between lamella, Enthesis Disruption: Disruption of bone/annulus interface, avulsions	Clear Lamella with no evidence of fissures and lesions, distinct Enthesis and distinct BEP/AF interface region	Some micro fissures between lamella present. Some loss of enthesis organization with presence of cartilage proliferation	Micro fissures between and across lamella, small tidemark avulsions at the enthesis	Micro and macro fissures between and across lamella or extending to BEP. Large tidemark avulsions at the enthesis
ECM Structure	Annular bulging, loss of lamella structure, scaring, fibrosis, disruption of concentric fibres, Matrix disorganisation, demarcation - presence of NP or CEP tissues in the AF	Clear Lamella structure, concentric fibres and clear demarcation between AF and NP / CEP / BEP. IAF is convex.	Some matrix disruption and loss of demarcation, lamella structure may be disrupted in localised areas. Some concavity of IAF	Moderate IAF concavity, matrix disorganisation with evidence of disrupted lamella, and loss of demarcation between NP/IAF, IAF/CEP or OAF/BEP	Large IAF concavity (inward). Extensive loss of lamella structure and clear loss of demarcation
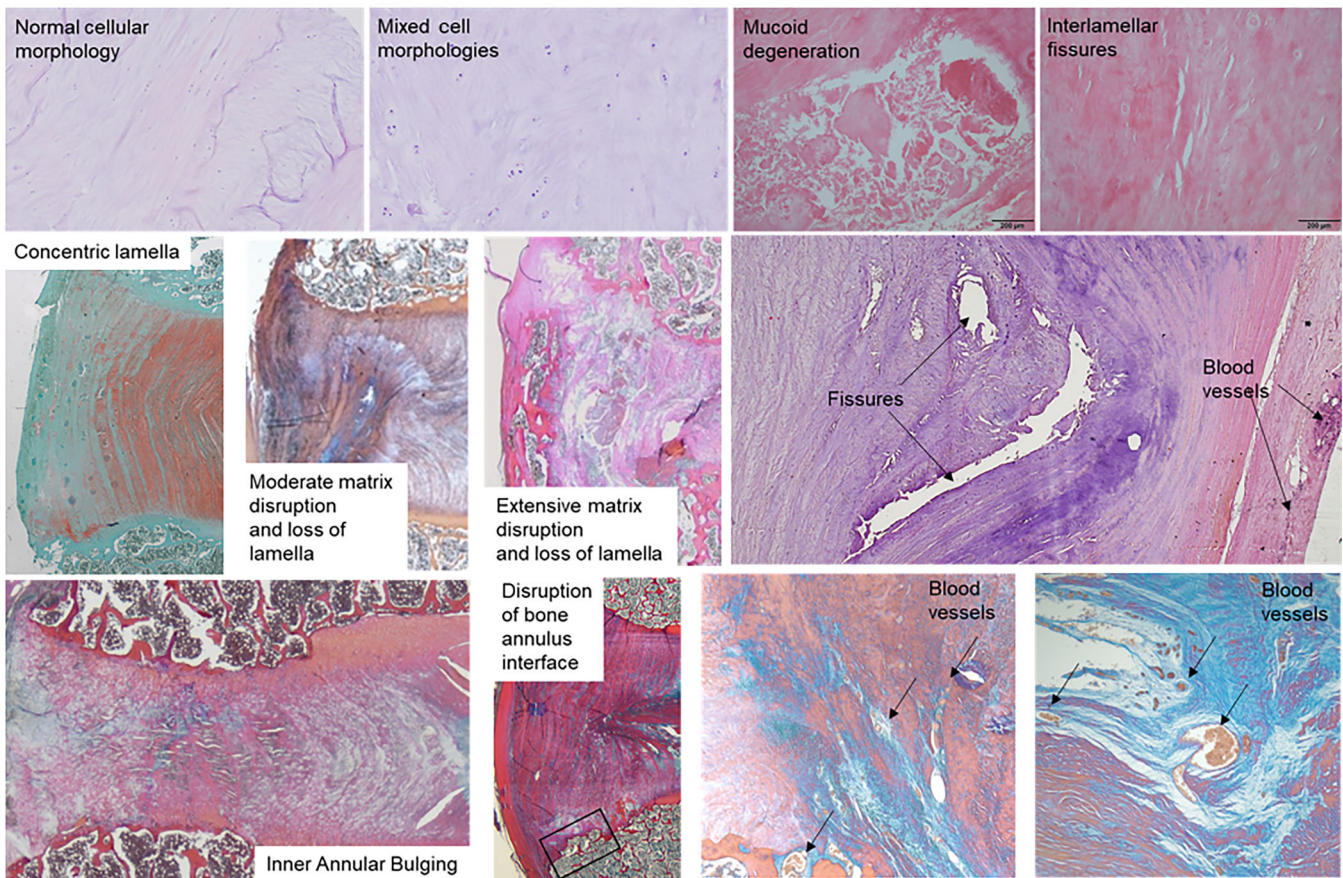
FIGURE 7 Taxonomy of grading for annulus fibrosus features. Descriptive text for features utilized for the grading (0-3) of the annulus fibrosus. Grading criteria broken down into cellularity, lesions and extracellular matrix (ECM) structure. Example images shown to demonstrate: normal cellular morphology, mixed cell morphologies, mucoid degeneration, interlamellar fissures, concentric lamella, disruption of bone/AF interface, extensive matrix disruption and loss of lamella, fissures and blood vessels, inner annular bulging, moderate matrix disruption and loss of lamella

CEP Features	Sub Features included	Score			
		0 Degenerative features: absent	**1** Degenerative features: few, small	**2** Degenerative feature: moderate	**3** Degenerative feature: abundant, large
Cellularity	Cellular proliferation (clusters), Neovascularisation	Single cells in lacunae, no evidence of necrotic, apoptotic or senescent cells	Occasional clone pairs in lacunae or multiple nuclei in lacunae, may be some evidence of apoptosis (nuclei condensation), cell shrinkage.	Dense pairs of clones but not large cell clusters. May be evidence of apoptosis (cell shrinkage), necrosis (cell debris) and cellular senescence (large irregular cells), may be areas of evidence cell loss (empty lacunae). Some neovascularisation.	Large clone clusters, evidence of empty lacunae, or may be evidence of apoptotic, necrotic and senescent cells. Evidence of extensive neovascularisation.
Lesions	Cartilage erosion, focal thinning, fissures, erosions/nodes, avulsion of bone/CEP interface	None	Few sites of focal CEP erosions.	Focal CEP avulsion from BEP. Moderate CEP thinning. CEP fissures and nodes.	Large zones of CEP avulsion from BEP, abundant loss of CEP thickness, large fissures.
ECM Structure	Cartilage thinning, cartilage disorganisation, demarcation, fibrocartilage ingrowth, calcification	Distinct CEP/BEP boundary, uniform CEP thickness.	Few sites of focal CEP thinning, disorganisation.	Moderate CEP calcification, fibrocartilage ingrowth, and disorganization	Abundant CEP calcification, fibrocartilage ingrowth.
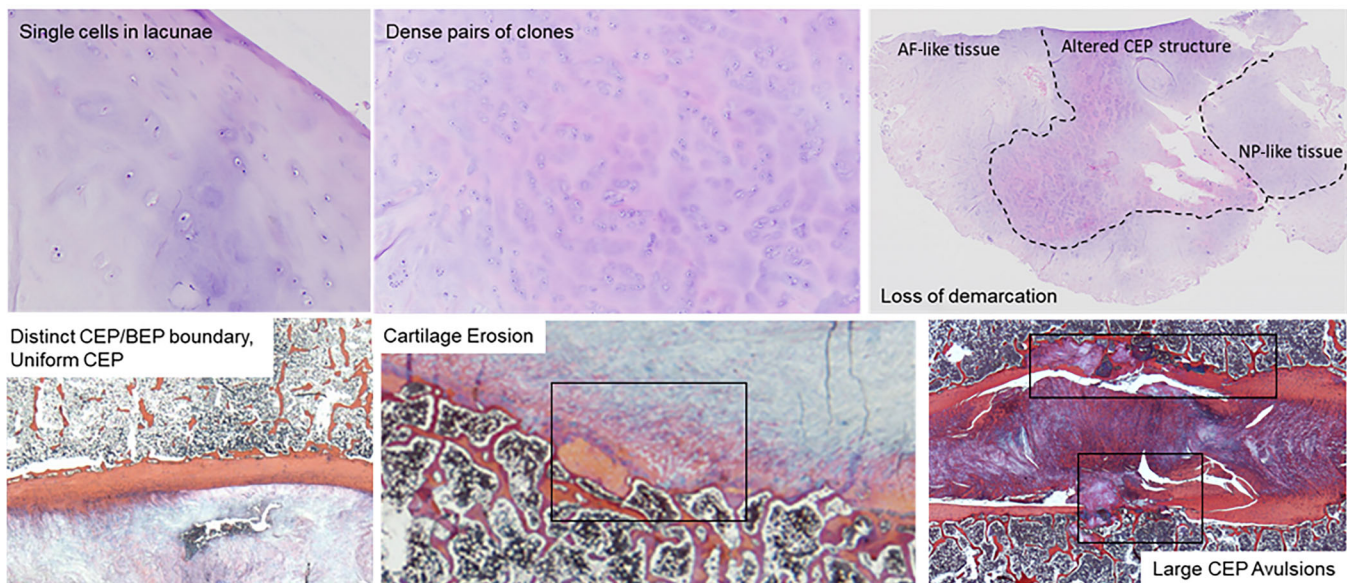
FIGURE 8 Taxonomy of grading for cartilage end plate features. Descriptive text for features utilized for the grading (0-3) of the cartilage end plate (CEP). Grading criteria broken down into cellularity, lesions and extracellular matrix (ECM) structure. Example images shown to demonstrate: single cells in lacunae, dense pairs of clones, loss of demarcation, distinct CEP/BEP boundary and a uniform CEP, cartilage erosion and large CEP avulsions

agreement for NP, with a number of graders reporting using HS_0015 for NP grading while the tissue shown is in fact AF tissue (Figure 10) and thus additional information to describe how to identify NP from AF would have been beneficial, we have now supplemented the training pack with this additional information (Supplementary file 2). While other IVDs with multiple images for some regions demonstrated that not all graders utilized all images of the tissue region to generate the overall grades for that region, examples shown for the AF (Figure 11) and CEP (Figure 12), suggesting that some of the variation seen between raters was due to image selection. While IVDs that had severe degeneration features (Figure 13) showed excellent agreement across raters, although within novice raters there remained

disagreement for some features. For most IVDs, most raters showed single-point disagreement between grades demonstrating general agreement (Figures 10–13).

When scores for each region of the IVD were pooled generating a degeneration grade per region resulting in three classifications of non-degenerate (0-3), mid-grade degeneration (4-6), and severe grade of degeneration (7-9) (Figure 14), inter-rater reliability improved with experience of grader (groups A → C), which was most evident for the NP and BEP demonstrating the need for more training materials or microscope time (Figure 14). Improvements in inter-rater reliability were seen with increasing grade of degeneration (Figure 14). Within non-degenerate IVD, greatest agreement for the region of the IVD

BEP Features	Sub Features included	Score			
		0 Degenerative features: absent	**1** Degenerative features: few, small	**2** Degenerative feature: moderate	**3** Degenerative feature: abundant, large
Cellularity	Bone marrow changes, neovascularisation	Hematopoietic bone marrow	Small zones of fibrovascular and/or fatty bone marrow at sites of CEP damage.	Moderately sized zones of fibrovascular and/or fatty bone marrow extending into the vertebral body	Abundant fibrovascular bone marrow changes
Lesions	Demarcation: Clear junction between CEP and BEP (clear tidemark), Clear Enthesis, erosions/nodes, osteophytes	Clear junction between CEP and BEP	Few irregularities of CEP/BEP interface	Moderate irregularities of CEP/BEP interface	Erosive nodes into vertebra, BEP compression fracture
ECM Structure	Schmorl's nodes, erosive nodes, trabecular remodelling, thickness, cartilage formation, sclerotic subchondral bone, cartilage or fibrocartilage, bone remodelling, osteophytes	Thin subchondral bone plate, distinct vertebral trabecular elements.	Small Schmorl's node rupture of NP through CEP and BEP. Small/few sites of fibrocartilage into subchondral bone with or without sclerotic margin.	Moderately sized Schmorl's nodes, and/or erosions of BEP with adjacent sclerosis or bone marrow fibrovascular changes. Sclerosis of the subchondral bone plate.	Abundant boney sclerosis
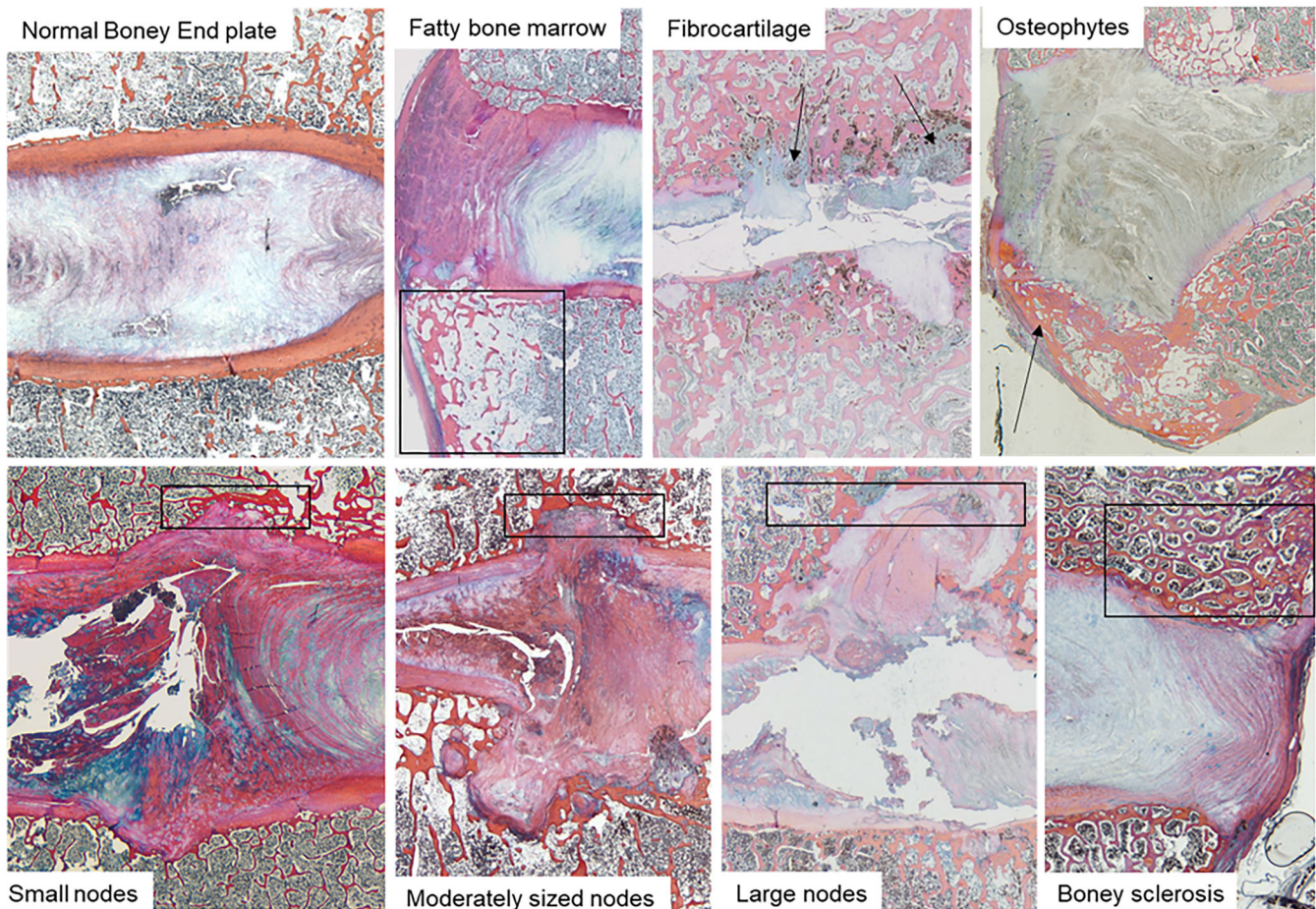
FIGURE 9 Taxonomy of grading for boney end plate features. Descriptive text for features utilized for the grading (0-3) of the boney end plate (BEP). Grading criteria broken down into cellularity, lesions and extracellular matrix (ECM) structure. Example images shown to demonstrate: normal end plate, fibrocartilage, osteophytes, fatty bone marrow, nodes and boney sclerosis

was seen for the CEP and BEP with poorest agreement within the NP region (Figure 14), although the IVDs with poorer agreement also aligned with those IVDs that showed graders utilizing different images to derive their grades. The provision of images for some IVDs did not enable all features to be scored for all regions, particularly the BEP resulting in a number of areas being unscored (Figure 14), of interest however more novice scorers provided scores for all features than experienced and author scorers. The results from the reliability test
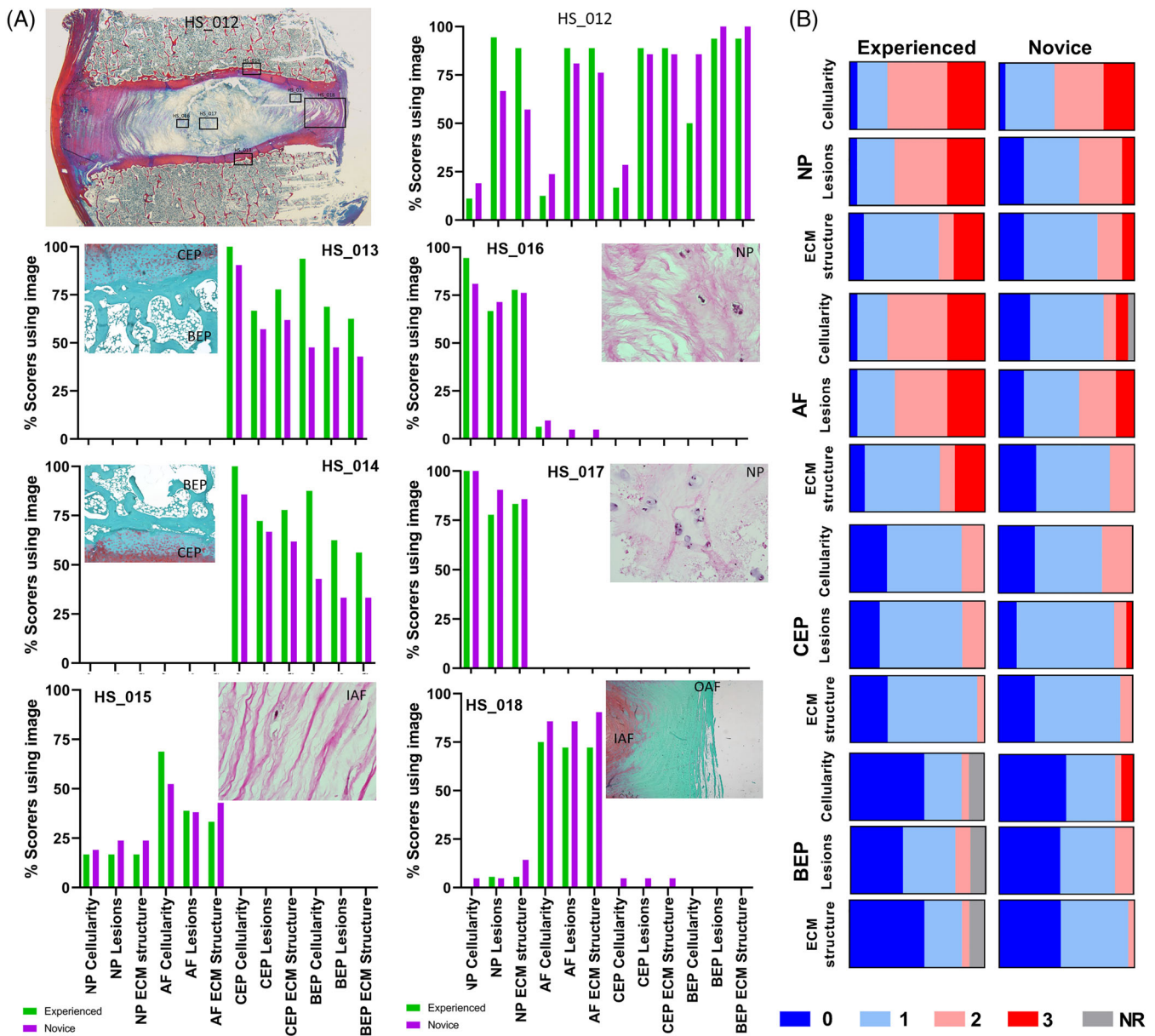
FIGURE 10 Disc 3 images utilized and grades generated following assessment of grading exercise demonstrating differential image use could explain some lack of consensus in nucleus pulposus tissues. Images utilized for mock disc 3 for round robin exercise, percentage scorers for the two groups: Experienced graders (n = 22) and Novice Graders (n = 18) who utilized each image to grade each feature within each disc region (nucleus pulposus (NP), annulus fibrosus (AF), cartilaginous end plate (CEP), and boney end plate (BEP)). Proportionality plots utilized to demonstrate the proportion of raters scoring each feature in each disc region as 0 to 3 or not responded (NR)

indicate that training and experience has an impact in understanding and recognition of the features on microscopic images. A larger number of samples would have impacted the understanding and training of the novice raters to test the reliability of the features.

Intra-rater reliability utilizing six raters demonstrated differential agreement levels between raters with agreement levels between 63.86% and 95.18% (Mean 83.07%), with two of six raters showing moderate agreement (Kappa.47, .51), one of six showing substantial agreement (Kappa .70), and three of six showing almost perfect agreement (Kappa .87, .87, .94) (Table 5).

6 | STAGE 5: POST-GRADING SURVEY

6.1 | Methods

All those who performed grading within the assessment of the scoring system were then asked to complete a post-grading survey, which collected information on grader demographics, and scorers' opinions on whether they agreed with proposed criteria utilized and the usability of the taxonomy of grading, in addition scorers were invited to submit comments via email. While the scoring
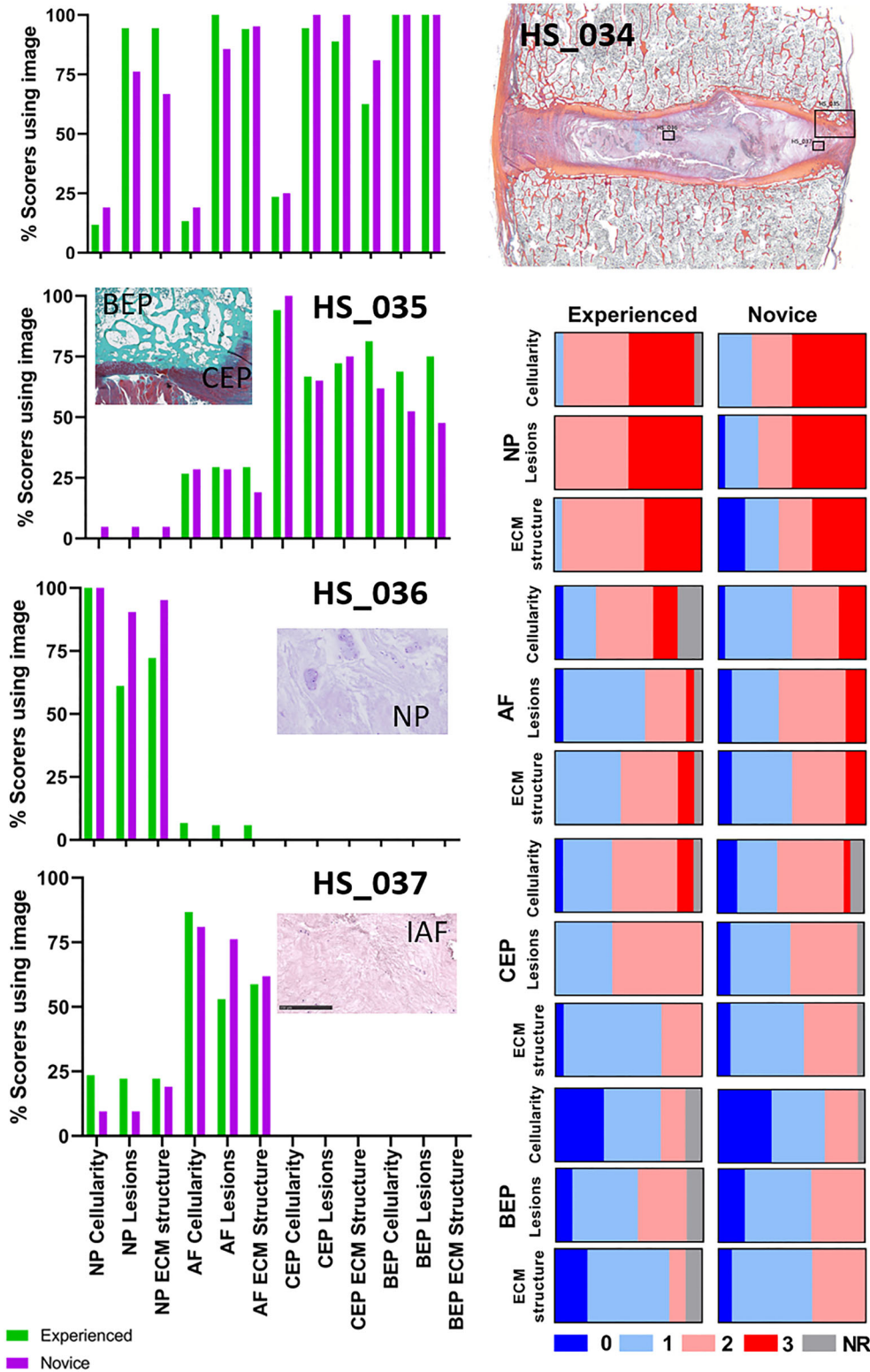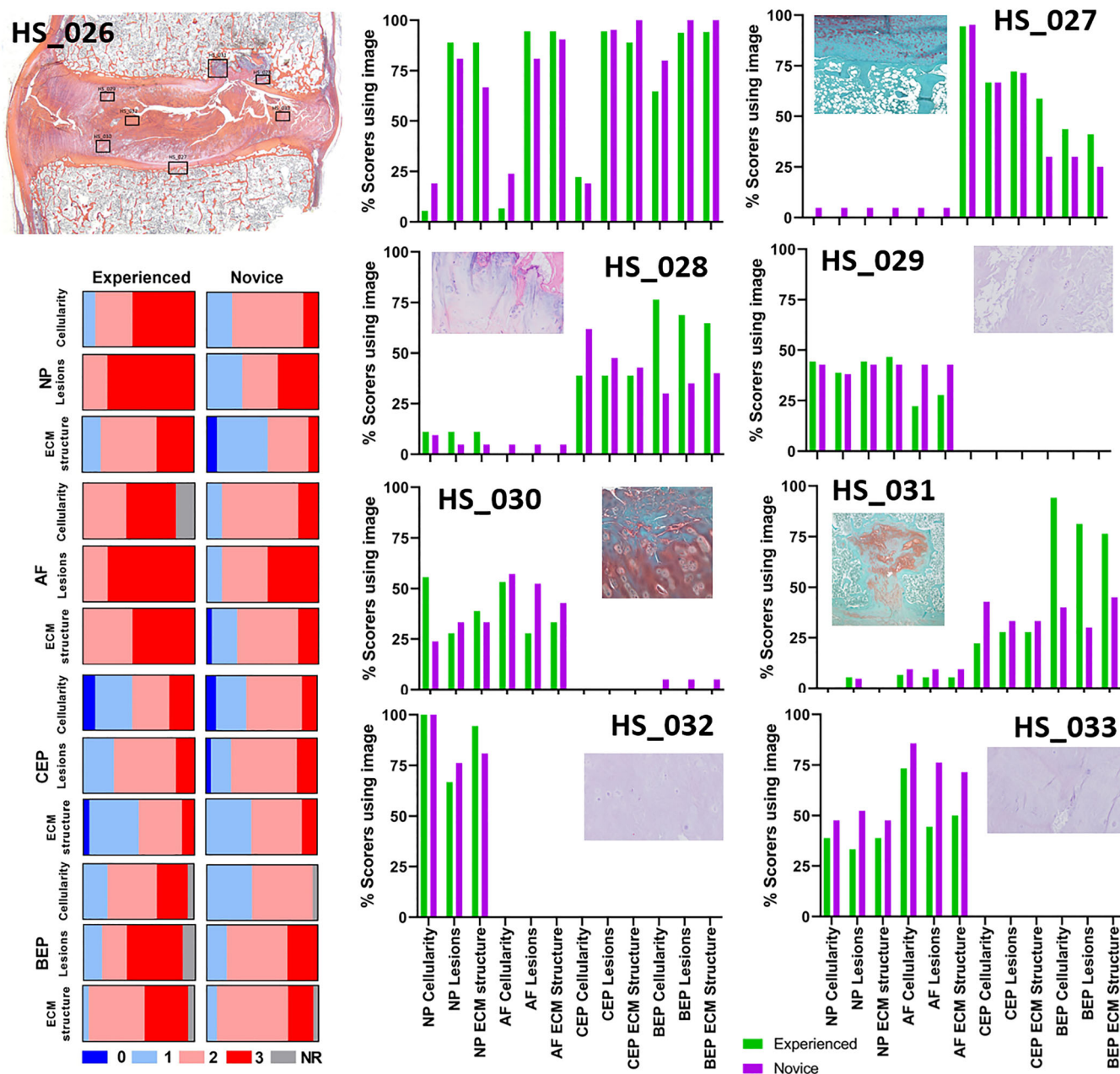
FIGURE 11 Disc 6 images utilized and grades generated following assessment of grading exercise demonstrating differential image use could explain some lack of consensus in annulus fibrosus tissues. Images utilized for mock disc 6 for round robin exercise, percentage scorers for the two groups: Experienced graders (n = 22) and Novice Graders (n = 18) who utilized each image to grade each feature within each disc region (nucleus pulposus (NP), annulus fibrosus (AF), cartilaginous end plate (CEP), and boney end plate (BEP)). Proportionality plots utilized to demonstrate the proportion of raters scoring each feature in each disc region as 0 to 3 or not responded (NR)

criteria were tested by 40 graders: 22 novice and 18 experienced, the post-grading survey was completed by only 28 graders: 13 novice and 15 experienced. The survey results were analyzed using SPSS 27. Using cross-tabulation analysis, the percentages of graders and their response in each category were determined. The survey collected responses on a six-point Likert scale from 0 (disagreement) to 5 (agreement). The percentage response for each point was calculated using SPSS 27, and the data represented as diverging stacked-bar chart, with the lower-half of the six-point response (0-2) for disagreement plotted as negative frequencies, and the upper-half (3-5) for agreement plotted as positive frequencies.

FIGURE 12 Disc 5 images utilized and grades generated following assessment of grading exercise demonstrating differential image use could explain some lack of consensus in cartilaginous end plate tissues. Images utilized for mock disc 5 for round robin exercise, percentage scorers for the two groups: Experienced graders (n = 22) and Novice Graders (n = 18) who utilized each image to grade each feature within each disc region (nucleus pulposus (NP), annulus fibrosus (AF), cartilaginous end plate (CEP), and boney end plate (BEP)). Proportionality plots utilized to demonstrate the proportion of raters scoring each feature in each disc region as 0 to 3 or not responded (NR)

6.2 | Results

Experienced graders included PIs/Postdocs, a master's student and pathologists, while novice scorers included one PI but mainly PhD students and undergraduate students with one technician. The majority of scorers reported were more familiar with the NP tissue (Figure 15). The post-grading survey showed that while the graders were in general agreement with the features described for scoring each IVD region, particularly for NP, AF and CEP, there was mild disagreement in whether these features were easily recognizable in the images and whether it will be easy to adapt for future studies (Figure 16), comments received highlighted the concern of transferability of the full grading system to surgical tissues that do not contain all tissue types.

7 | DISCUSSION

Our goal was to develop a standardized histopathology scoring scheme for histologic evaluation of degenerative features within human IVDs. These recommendations are based on literature review
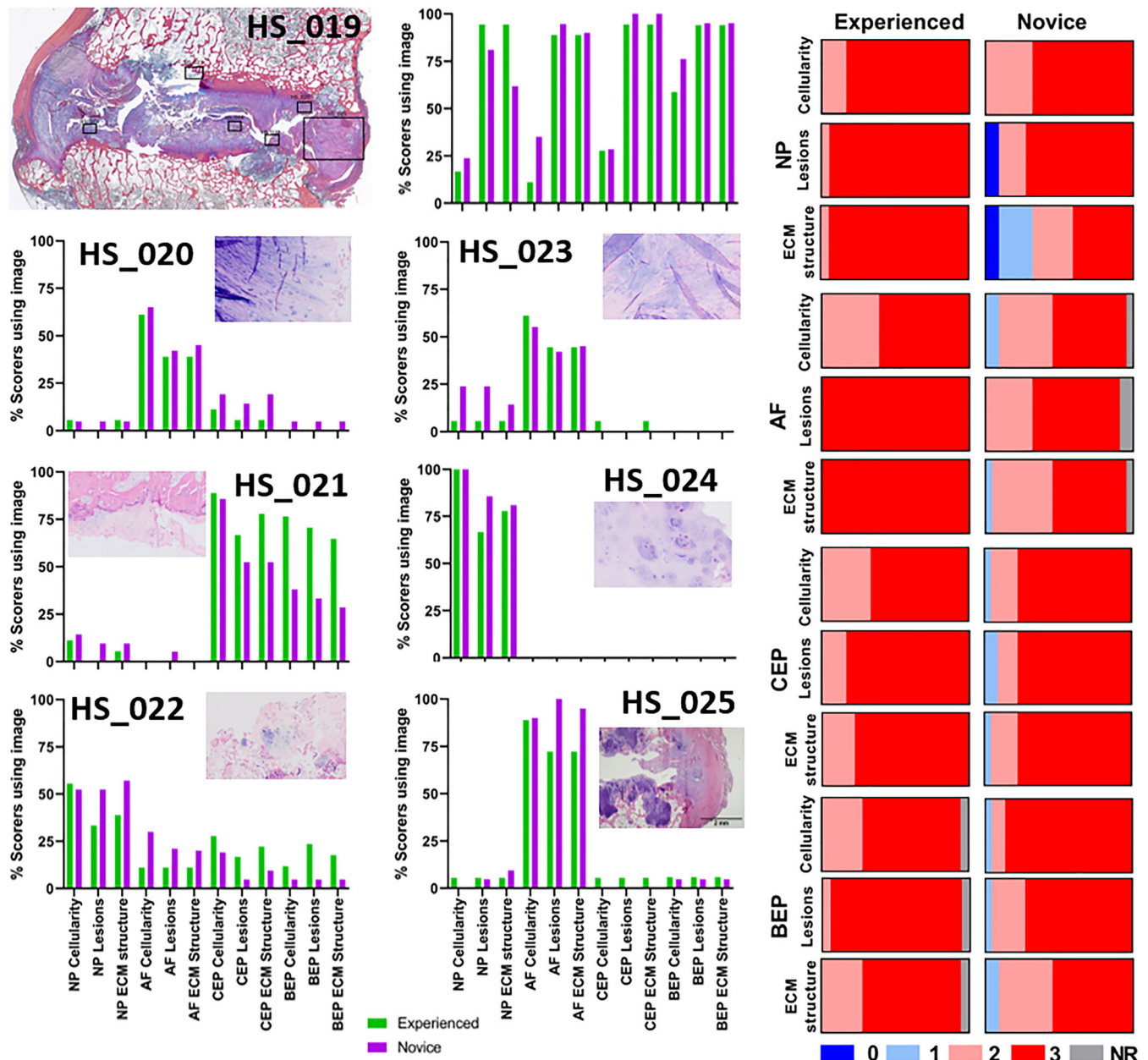
FIGURE 13 Disc 4 images utilized and grades generated following assessment of grading exercise demonstrating severely degenerated disc with good consensus for scoring. Images utilized for mock disc 4 for round robin exercise, percentage scorers for the two groups: Experienced graders (n = 22) and Novice Graders (n = 18) who utilized each image to grade each feature within each disc region (nucleus pulposus (NP), annulus fibrosus (AF), cartilaginous end plate (CEP), and boney end plate (BEP)). Proportionality plots utilized to demonstrate the proportion of raters scoring each feature in each disc region as 0 to 3 or not responded (NR)

and expert opinion, serving as a first step for establishing best practices and methodologies for human IVD grading. This work was motivated by the ongoing challenge to consistently document and report histologic findings across studies, which limits progress toward understanding clinically important changes. We developed a set of visual depictions plus nomenclature to provide a robust system to describe and classify attributes that reliably distinguish IVDs at various stages of degeneration. The implementation of this system requires training materials so raters can improve their recognition for characteristic patterns that associate with degenerative changes. We observed that

inexperienced raters demonstrated poor reliability in scoring, which indicates the need for training methods for both processing tissues and describing findings. This could lead to improved agreement across groups and broader integration of findings.

The proposed scoring system provides a comprehensive evaluation of the main IVD sub-tissues over a range of hierarchical scales: cellular, ECM, and structure. This is because the concept of IVD health includes synergy between sub-tissues at the macroscopic level to achieve region-dependent physical requirements, plus homeostasis at the cellular level to maintain tissue integrity. Results from the IVD

TABLE 3 Interclass correlation coefficient to test the inter-rater reliability for the histopathological features for each disc regions between novice and experienced raters

Features	Total (40 raters)					Experienced (18 raters)					New (22 raters)				
	ICC	LL 95%CI	UL 95%CI	P	Subject (discs)	ICC	LL 95%CI	UL 95%CI	P	Subject (discs)	ICC	LL 95%CI	UL 95%CI	P	Subject (discs)
NP Cellularity	.89[a]	0.75	0.98	.00	7	.81[a]	0.57	0.95	.00	8	.82[a]	0.59	0.96	.00	8
NP lesions	.95[a]	0.88	0.99	.00	8	.93[a]	0.85	0.98	.00	10	.87[a]	0.72	0.97	.00	8
NP ECM structure	.89[a]	0.76	0.97	.00	8	.90[a]	0.79	0.97	.00	10	.69[a]	0.33	0.92	.00	8
AF cellularity	.93[a]	0.78	1.00	.00	4	.94[a]	0.82	0.99	.00	5	.79[a]	0.53	0.95	.00	8
AF lesions	.97[a]	0.92	0.99	.00	8	.95[a]	0.89	0.99	.00	9	.93[a]	0.83	0.98	.00	9
AF ECM structure	.97[a]	0.93	0.99	.00	8	.97[a]	0.92	0.99	.00	9	.92[a]	0.81	0.98	.00	9
CEP cellularity	.97[a]	0.92	0.99	.00	6	.93[a]	0.82	0.99	.00	6	.96[a]	0.89	0.99	.00	7
CEP lesions	.98[a]	0.96	1.00	.00	8	.96[a]	0.91	0.99	.00	9	.97[a]	0.92	0.99	.00	8
CEP ECM structure	.98[a]	0.96	1.00	.00	8	.95[a]	0.90	0.99	.00	9	.97[a]	0.93	0.99	.00	8
BEP cellularity	[b]					[b]					.99[a]	0.94	1.00	.00	2
BEP lesions	[b]					[b]					.97[a]	0.91	1.00	.00	5
BEP ECM structure	[b]					[b]					.96[a]	0.88	1.00	.00	5

Note: Type A intraclass correlation coefficients using an absolute agreement definition. Two-way mixed effects model where people effects are random and measures effects are fixed. The estimator is the same, whether the interaction effect is present or not.

[a]This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

[b]There are too few subjects (N = 0) for the analysis.
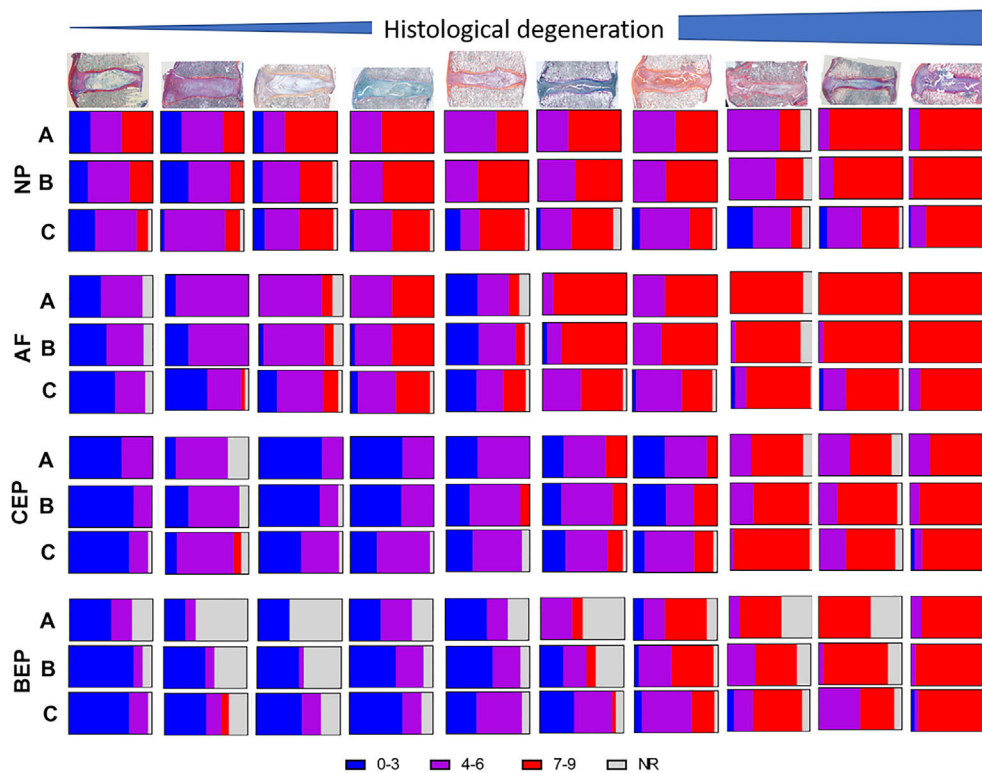
FIGURE 14 Proportionality plots for grades generated following assessment of grading exercise. Ten mock discs were utilized within a beta testing round robin scoring. Each disc region was scored on a scale 0 to 3 for three features and the sum degeneration score calculated for each disc region generating an overall grade for each region of non-degenerate (0-3), medium grade of degeneration (4-6), and severe degeneration (7-9), if any feature was not scored by a rater then the combined degeneration grade was not calculated and shown on plots as not responded (NR). Grading results represented for three groups: A, Experienced authors (n = 8); B, Experienced graders (n = 22); C, Novice Graders (n = 18). Discs shown in order of increasing grade of degeneration, together with the low power image utilized for the grading round

ratings indicate that degenerative changes are observed initially at the cellular level and become more prominent at the matrix and structural level as degeneration progresses. Interestingly we identified that degenerative features were only seen within the BEP, CEP, and AF when degenerative changes were present in the NP, while degenerative changes were seen in the NP regions in the absence of degenerative changes within the AF, CEP, and BEP. This could indicate that the IVD degenerates from the "inside-out," with earliest degenerative features being observed in the large and avascular NP, this requires further investigation. The initial survey of spine researchers indicated interest in scoring the changes related to cellular features for NP, and features related to structure changes in AF and CEP. Most enthusiastic response was received for NP, followed by AF and CEP. There was less response and interest in the BEP, but this may purely be representative of the research interests of the respondents. The post-grading survey demonstrated agreement with features for NP, AF, and CEP.

There was strong inter-rater reliability with more experienced graders, and mild disagreement among all raters when scoring BEP represented by moderate to poor inter-rater reliability test results and a greater number of abstaining graders. This may be because the BEP is an under-studied region of the IVD, and the graders are not familiar with the histology and histopathology of this region. The results from the reliability test indicate that training and experience has an impact in understanding and recognition of the features on microscopic images and it would have been beneficial if more novice graders had completed the post grading survey. A larger number of samples would have impacted the understanding and training of the novice raters to test the reliability of the features. Furthermore, this study was limited by the use of representative images rather than utilizing slides and microscope-based training, the differential use of images to score certain regions demonstrates fundamental training on identification of tissue types is also essential. This also highlighted the need that when grading scorers should review multiple regions and assess average scores to take into account variability in features across the IVD. It is also essential that differential magnifications are utilized to be able to identify certain features, for example, cellular changes can only be visualized at higher magnifications and higher magnification is necessary to determine whether a tissue void is a true fissure or an artifact of tissue processing (Supplementary file 2).

Also, while most participants were enthusiastic about a five- (0-4) to six-point (0-5) scoring range, based on reliability testing spreading the scoring range further would result in poor-agreement, as the ability to distinguish between mild or subtle changes will require a very thorough histopathological training, and may not yield consistent and reproducible results in labs with students and trainees. Hence, a

TABLE 4 Interclass correlation coefficient to test the inter-rater reliability for the histopathological features for each disc regions within labs

Features	Lab-A (2 experienced, 1 novice rater)					Lab-B (2 experienced, 2 novice raters)					Lab-C (1 experienced, 5 novice raters)					Lab-D (1 experienced, 5 novice raters)					Lab-E (4 novice raters)				
	ICC	LL 95 %CI	UL 95 %CI	P	Subject (discs)	ICC	LL 95 %CI	UL 95 %CI	P	Subject (discs)	ICC	LL 95% CI	UL 95% CI	P	Subject (discs)	ICC	LL 95 %CI	UL 95 %CI	P	Subject (discs)	ICC	LL 95 %CI	UL 95 %CI	P	Subject (discs)
NP cellularity	.98[a]	0.94	1.00	.00	9	-.04c	-1.33	0.73	.49	8	.11[a]	-1.07	0.74	.37	10	.56[a]	0.11	0.86	.01	10	.66[a]	0.17	0.90	.01	10
NP lesions	1.0[a]				10	.76c	0.24	0.95	.00	8	.65[a]	0.19	0.90	.01	10	.45[a]	-0.23	0.84	.08	10	.56[a]	0.06	0.86	.01	10
NP ECM structure	1.0[a]				10	.47[a]	-0.43	0.88	.11	8	.18[a]	-0.63	0.74	.28	10	-.65[a]	-1.92	0.41	.86	10	.70[a]	0.28	0.91	.00	10
AF cellularity	.90[a]	0.72	0.97	.00	10	.09[a]	-1.09	0.80	.39	7	.57[a]	0.06	0.88	.02	9	.70[a]	0.33	0.91	.00	10	.76[a]	0.41	0.93	.00	10
AF lesions	.90[a]	0.73	0.97	.00	10	.77[a]	0.36	0.94	.00	9	.83[a]	0.59	0.95	.00	10	.82[a]	0.57	0.95	.00	10	.86[a]	0.60	0.96	.00	10
AF ECM structure	.95[a]	0.85	0.99	.00	10	.63[a]	0.14	0.90	.00	9	.86[a]	0.63	0.96	.00	10	.87[a]	0.69	0.96	.00	10	.77[a]	0.44	0.94	.00	10
CEP cellularity	.82[a]	0.48	0.96	.00	9	.71[a]	0.14	0.93	.02	8	.91[a]	0.76	0.98	.00	9	.85[a]	0.64	0.96	.00	10	.86[a]	0.65	0.96	.00	10
CEP lesions	.97[a]	0.90	0.99	.00	10	.68[a]	0.04	0.93	.02	8	.93[a]	0.83	0.98	.00	10	.96[a]	0.90	0.99	.00	10	.87[a]	0.66	0.96	.00	10
CEP ECM structure	.92[a]	0.76	0.98	.00	10	.60[a]	-0.19	0.91	.05	8	.94[a]	0.87	0.98	.00	10	.87[a]	0.70	0.96	.00	10	.91[a]	0.77	0.97	.00	10
BEP cellularity	.96[a]	0.81	1.00	.00	5	.58[a]	-0.29	0.93	.07	6	.94[a]	0.84	0.99	.00	7	.91[a]	0.76	0.98	.00	8	.82[a]	0.53	0.95	.00	9
BEP lesions	.90[a]	0.73	0.97	.00	10	.85[a]	0.51	0.98	.00	6	.94[a]	0.84	0.99	.00	8	.85[a]	0.64	0.96	.00	10	.85[a]	0.58	0.96	.00	9
BEP ECM structure	.94[a]	0.83	0.98	.00	10	.83[a]	0.45	0.97	.00	6	.95[a]	0.88	0.99	.00	8	.88[a]	0.73	0.97	.00	10	.60[a]	-0.01	0.89	.03	9

Note: Type A intraclass correlation coefficients using an absolute agreement definition. Two-way mixed effects model where people effects are random and measures effects are fixed. The estimator is the same, whether the interaction effect is present or not.

[a]This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Rater	Observed agreement (%)	Kappa	LL 95% CI	UL 95% CI	P-value	Discs
1	91.36	.87	0.77	0.96	.0001	7
2	95.18	.94	0.87	1	.0001	7
3	90.63	.87	0.77	0.97	.0001	7
4	75.68	.47	0.33	0.62	.0001	7
5	63.86	.51	0.38	0.64	.0001	7
6	81.71	.70	0.56	0.84	.0001	7

TABLE 5 Cohen's Kappa (unweighted) to test the intra-rater reliability for the histopathological features across all features and regions within seven discs within selected raters
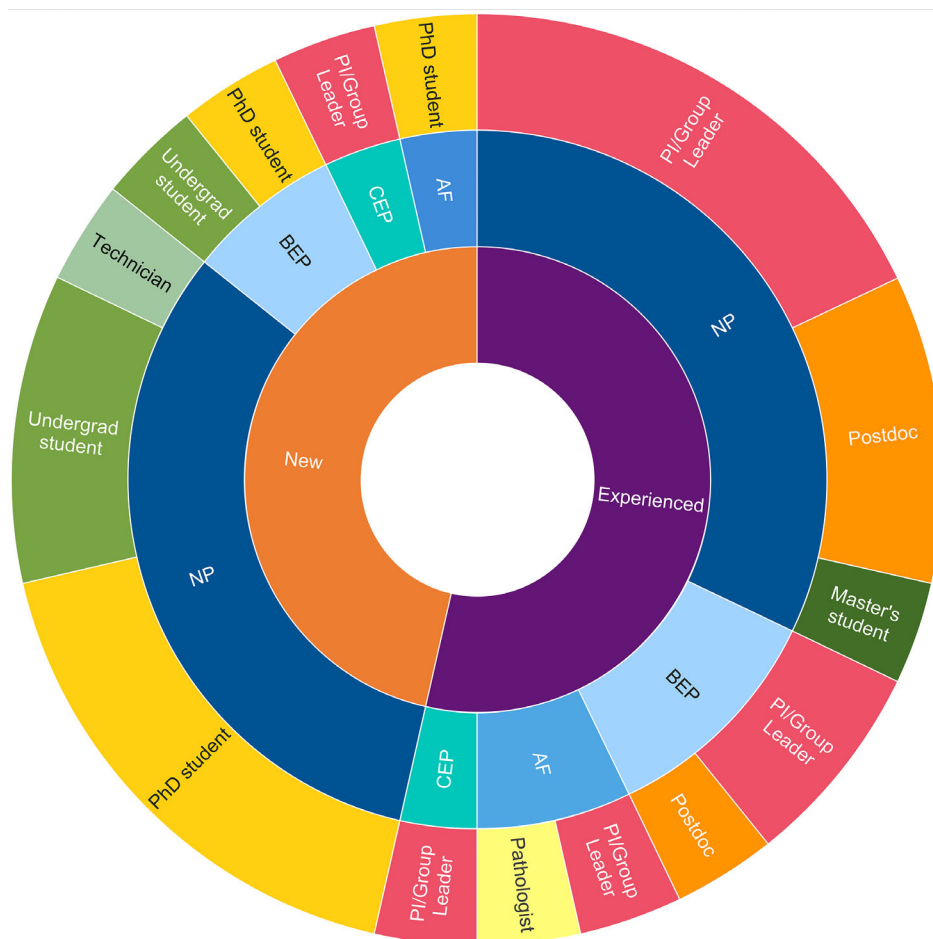
FIGURE 15 Frequency distribution of raters that tested the new scoring system. Multi-layer donut for the cross-tabulation analysis shows the frequency distribution of post-grading survey participants (n = 28) by novice (n = 13) and experienced (n = 15) raters. The layers further show the percentage response for each survey category by novice and experienced graders including: the comfort in grading specific region of IVD tissue, and current academic and training level

scoring range where changes from non-, mild-, moderate-, and severe-degeneration can be easily recognized (four-point scoring range) will be more consistent and reproducible. The combination of scores for regions of the IVD further improved agreement for the overall grading of the IVD region as non-, moderate-, and severe degeneration suggesting that the combined grades for IVD regions would be more reliable than specific grades for each feature.

Intra-rater reliability was excellent in some observers but poorer in others. Those with poorer intra-rater reliability results reported that the discussions on the grading system between scoring had generated improved understanding of features and impacted on differential scores in the subsequent round of scoring. Very few grading systems for the IVD have been assessed as intensively, as studied here, with inter-rater reliability testing limited to within lab users and intra-rater reliability normally only completed with one or two scorers.[12,25] Thompson et al, 1990, validated their scoring system using 136 sections, where two sections were analyzed from the same IVD, and were scored by three independent blinded graders. The reliability of the scoring system was tested using Counter-rater results showed 61% to 88% agreement, with Cohen's kappa between .67 and .94 range.[12] And intra-rater reliability tests showed 85% and 87% agreement, with Cohen's kappa between .87 and .91.[12] Boos et al, 2002, tested the scoring system between two pathologists, who scored 54 samples, and 150 slices. The inter-rater reliability of the Boos grading system was tested using weighted kappa which was reported between .49-.98, while intra-rater reliability was not reported.[25] The inter-rater reliability we observed here showed similar agreement levels across a much broader population of scorers with engagement

FIGURE 16 Opinion of testers on the new scoring system. Diverging stacked bar-chart (A and B) show percentage response to each question on Likert scale of 0 to 5, from disagreement (0) to agreement (5) by testers (n = 28) in a post-grading survey. The lower-half of the six-point response (0-2) for disagreement are plotted as negative frequencies, and the upper-half (3-5) for agreement are plotted as positive frequencies
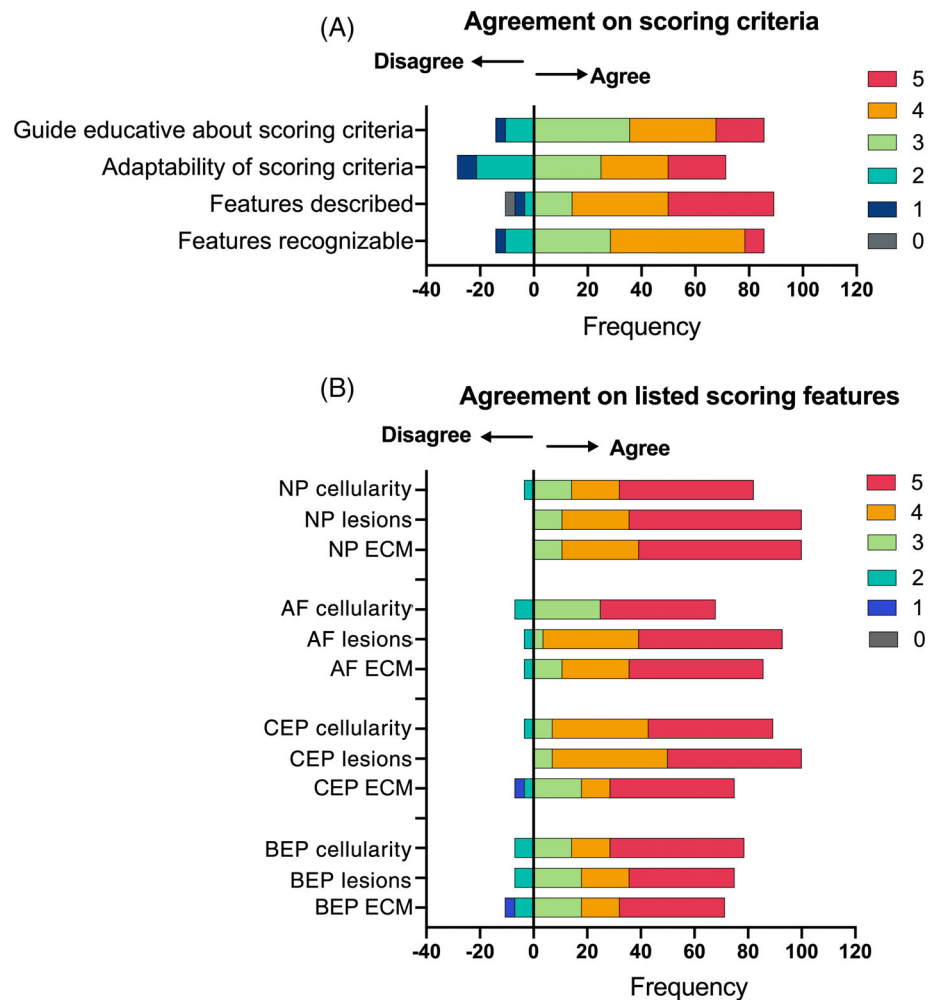
of experienced and inexperienced graders, importantly the inclusion of early-career scientists (undergraduates, PhDs and postdocs) within the scoring is essential as these are the individuals who are required to score for their research studies and thus the usability in the actual individuals who undertake the grading is essential. The current study developed a comprehensive, complete taxonomy of histological features that can be utilized for assessing human IVD degeneration. The testing of this grading system across seventeen labs worldwide brings in a wider perspective rather than single lab development and testing.

Because the pathophysiology of IVD degeneration and chronic back pain is multifactorial,[46] clinical implications drawn from histologic findings can vary widely. Mechanistic studies typically focus on biological features, such as cellular and ECM structure. These were the preferences of the majority of our survey respondents, which may reflect a mechanistic bias. Alternatively, biomechanical researchers may view lesions as important evidence of tissue overload and damage. Pain researchers tend to focus on features that associate with painful clinical conditions, such as inflammatory changes at the endplate and outer annulus where mechanical and chemical sensitization of nerves or the generation of neurotrophic factors can be more prevalent. The ideal IVD grading scheme should be agnostic to the intent of the user and be sufficiently comprehensive so as to

investigate degeneration concepts that bridge these perspectives. This is particularly true when considering the development and evaluation of new therapies.

Histologic assessment of IVD tissues may be the gold standard for judging degenerative changes. However, clinical interpretation of these findings typically relies on the identification of these features in routine medical imaging, such as plain radiographs and traditional T1- and/or T2-weighted MRI.[47,48] This can be difficult owing to these modalities' limited spatial resolution and image contrast (Figure 17). Moreover, cadaveric studies used to characterize histologic features often lack imaging, patient demographics, and clinical profiles, which precludes firm conclusions regarding IVD pathologies as pain generators. Improved characterization and interpretation of IVD pathologies may also shed light on mechanisms for clinical complications of current treatments, such as adjacent segment degeneration/disease, IVD re-herniation, IVD resorption following an initial herniation, resolution or intensity of pain, and pain severity among others.

In spite of these challenges, certain imaging/histological findings do appear to be associated with chronic back pain (Table 6), which supports the clinical relevance of the individual features and also provides rationale for their histologic grading. According to systematic reviews, IVD changes have been found to be related to low back pain

T1-weighted MRI   T2-weighted MRI   Histology
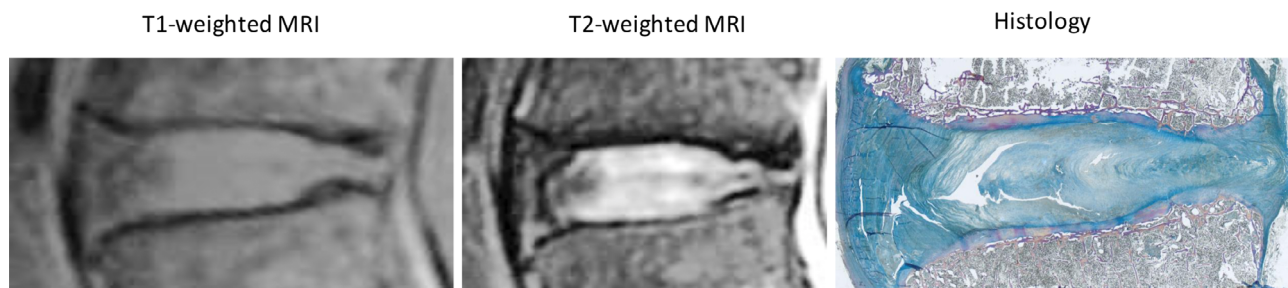
FIGURE 17 Mid-sagittal images of a lumbar, L3/4, intervertebral disc. Left two panels are clinical MRI scans of the intact lumbar spine. Right panel is a histologic section of the intact disc, coincident with the MRI images (decalcified, paraffin-embedded, and stained with Mallory-Heidenhain). Images demonstrate how subtle features of the disc sub-tissues are not apparent with clinical imaging

TABLE 6 Summary of human IVD histopathology studies that reported associations between various features and IVD degeneration severity or low back pain

Feature	Imaging/biopsy	Positive association with IVD degeneration severity (references)	Positive association with low back pain (references)
AF tear/disruption	Imaging/Biopsy	29,49,50	14,17,51-54
IVD height collapse	Imaging	55	56
sGAG loss	Imaging/Biopsy	57-61	59,62,63
NP cell cluster formation And increased catabolic phenotype.	Biopsy	9,10,32,64-66	
CEP damage	Biopsy/Imaging	67,68	69,70
Vertebral endplate bone marrow lesions (Modic changes)	Biopsy/Imaging	42,71,72	42,69,73,74

but the association requires further investigation because of inherent heterogeneity between studies, incomplete assessment of phenotypes that can also be associated with pain, insufficient statistical modeling, and issues related to imaging quality.[2,75] Moreover, recent advances in quantitative MRI (eg, T2, T2*, T1ρ mapping, sodium, UTE, and spectroscopy) enable non-invasive measurement of IVD biochemical composition that can facilitate identification of early IVD changes and identify the symptomatic IVD(s).[76] Furthermore, newer sequences with higher spatial resolution and improved image contrast permit visualization of CEP structure and pathologies at the bone-IVD interface as well as within the IVD itself (ultrashort echo time, UTE).[66,77] In the future, these advanced imaging sequences may make it possible to prospectively validate the clinical relevance of histopathology features observed in IVDs that are difficult to discern on conventional images.

Histopathology studies performed on tissue samples biopsied from chronic back pain patients also provide strong support for the features in the proposed grading scheme. For example, in symptomatic patients, innervation is greater in CEPs with cartilage and subchondral bone damage,[16] perhaps as a chemotactic response to neurotrophin production by IVD cells[64,78,79] and new blood vessels.[15] Innervation is also greater in painful IVDs with annulus fissures,[17] which may provide a chemically and mechanically favorable environment for nerve ingrowth,[51] with nerve fibers found to migrate into the NP associated with loss of proteoglycans and ECM fissures.[17] Likewise, elevated levels of pro-inflammatory cytokines measured in

these IVD and endplate tissues,[32,80] in particular associated with cellular clusters,[81] suggest these cytokines may play an important role in promoting degeneration and pain.[82,83]

When interpreting the histologic features described here, it is also important to distinguish between prevalence vs pathogenesis and between an association vs causation. Indeed, some features may be highly prevalent, although their role in IVD degeneration pathophysiology remains unclear. For example, Schmorl's nodes or structural endplate abnormalities/changes that may vary in size and extent of indentation involvement across the endplate can be associated with IVD degeneration severity and pain.[84,85] Nevertheless, it is challenging to distinguish between endplate changes that are developmental and attributed to neurocentral synchondrosis and improper notochord regression, those that may arise during skeletal development and attributed to a weakened endplate, to those that form traumatically or part of the remodeling process in response to IVD changes and/or mechanical effects from structural spine changes.[86-89] In fact, a hereditary and genetic predisposition has been found to be associated with these endplate phenotypes that may precipitate their manifestation in relation to IVD degeneration and may be an initiator of IVD changes.[90,91] Consequently, results from clinical studies relating endplate abnormalities to symptoms are mixed, and the prevalence of such phenotypes is relatively high in asymptomatic individuals.[85,92,93] However, limitations exist with previous studies, largely attributed to the lack of understanding and definition of the endplate phenotype,

Future consensus studies
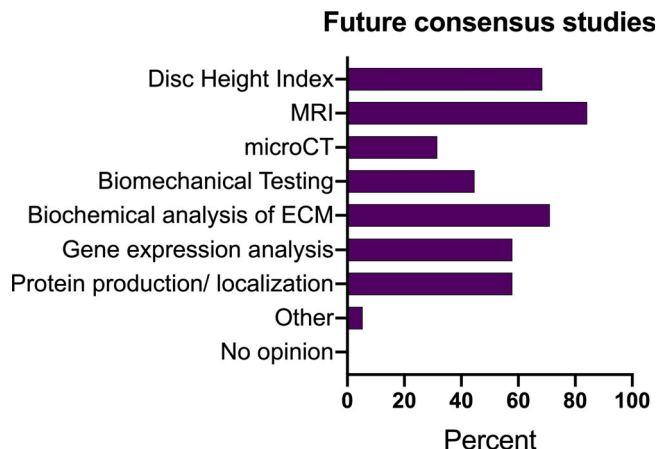
FIGURE 18 Opinion of spine community on future consensus studies. Histogram plotting the opinion of survey responders (n = 38) in percentage for future consensus outcome measure studies related to spine research

its various sub-phenotypes, study design, mode of assessment, and depth/breadth of analyses.[94]

Beyond further work to validate the proposed grading system, our survey noted interest by the community for future consensus papers (Figure 18). These include work to develop and validate simplified radiologic measures of IVD degeneration, such as the IVD height index. Additionally, mechanistic and clinical interpretation of histologic changes may be improved from consensus on concurrent changes in characteristics such as cellular function, matrix composition, and biomechanical behavior. Another interest is the need to have standardization of MRI phenotypes that will help establish clinical importance of structural features of IVD degeneration. As previously mentioned, there is tremendous variation between rater reliability of MRI phenotypes and the definition of such phenotypes.[95-98] This discrepancy may account for the inconsistent association and predictive utility of such phenotypes in relation to the LBP profile and disability.[95] As such, international consortia have been formed to help provide a common language and standardization of MRI and other imaging phenotypes.[94] In addition, machine learning approaches regarding feature phenotype recognition on imaging have been developed to assist with standardization of phenotype assessment, shorten time of assessment and facilitate multicenter studies.[99,100] However, such approaches are based on a truth set and dependent on human interpretation. Again, the need to properly define and understand such phenotypes is critical, further necessitating universal consensus. Such an initiative is further compounded by the need to develop more personalized spine care methods that aim to further incorporate imaging and clinical phenotypes to maximize management, address targeted therapeutics, reinforce predictive modeling algorithms, and further inform preventative measures.[20]

8 | CONCLUSION

This ORS Initiative to advance histopathologic evaluation of the IVD in humans has engaged spine researchers from across the world and at different stages of their careers to develop a robust and comprehensive grading scheme of IVD degeneration. This work focused on the use of a training set of images that were composed of whole cadaveric and magnified regions of tissues to demonstrate features, many of which were derived from surgical samples. The use of these mock images while extremely useful to engage a wide range of potential users did have some issues with mismatch and difficulties experienced, particularly by novice scorers in identification of tissue types. The development of defined grading criteria for each region of the IVD should enable rapid translation to surgical tissues where grading of the tissue types available can be performed (mainly NP and AF tissues) but compared to cadaveric IVDs for these regions. Future studies will further refine, verify, and evaluate the grading system for application to cadaveric and surgical samples, further developing the training materials to enable online training across labs around the world.

The resulting scoring system described here is a first step for establishing best practices and methodologies for human IVD grading. We expect this system will undergo continued optimization as it gains use by the wider spine research community, ultimately resulting in a consensus scoring system that can be used worldwide.

CONFLICT OF INTEREST

The authors have no relevant conflicts of interest to declare in relation to this article.

AUTHOR CONTRIBUTIONS

All authors contributed to study design. C.L.M., S.I.J., and C.L.D. performed detailed analysis of previously published human histopathological grading systems. J.L., M.G., and C.L.M. expanded features for inclusions and potential artifacts from processing. C.L.D. designed both surveys, all authors provided feedback to the development of the surveys, and C.L.D. analyzed the survey data. All authors contributed to discussions and development of the new grading system and participated in the beta testing of the scoring system, C.L.M., C.L.D., J.L., S.I.J., A.F., and C.C. performed second round grading for intra-rater reliability testing. C.L.M. performed data analysis of the assessment of the scoring system with 40 participants. C.L.D. and C.L.M. statistically analyzed the data from inter-rater reliability and intra-rater reliability testing respectively. All authors contributed to the interpretation of the data and the generation of the manuscript. All authors read and approved the final submitted manuscript.

ORCID

Christine L. Le Maitre https://orcid.org/0000-0003-4489-7107
Chitra L. Dahia https://orcid.org/0000-0003-3683-9791
Morgan Giers https://orcid.org/0000-0002-2247-1167
Gianluca Vadala https://orcid.org/0000-0001-7142-1660

Bibliography

1. Cheung KMC, Samartzis D, Karppinen J, Luk KDK. Are "patterns" of lumbar disc degeneration associated with low back pain?: new insights based on skipped level disc pathology. *Spine*. 2012;37:E430-E438. https://doi.org/10.1097/BRS.0b013e3182304dfc

2. Chou D, Samartzis D, Bellabarba C, et al. Degenerative magnetic resonance imaging changes in patients with chronic low back pain: a systematic review. *Spine*. 2011;36:S43-S53. https://doi.org/10.1097/BRS.0b013e31822ef700

3. Livshits G, Popham M, Malkin I, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. *Ann Rheum Dis*. 2011;70:1740-1745. https://doi.org/10.1136/ard.2010.137836

4. Samartzis D, Karppinen J, Mok F, Fong DYT, Luk KDK, Cheung KMC. A population-based study of juvenile disc degeneration and its

5. Takatalo J, Karppinen J, Niinimäki J, et al. Does lumbar disc degeneration on magnetic resonance imaging associate with low back symptom severity in young Finnish adults? *Spine*. 2011;36:2180-2189. https://doi.org/10.1097/BRS.0b013e3182077122

6. Vos T, Allen C, Arora M. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1545-1602. https://doi.org/10.1016/S0140-6736(16)31678-6

7. Luoma K, Riihimäki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. *Spine*. 2000;25:487-492. https://doi.org/10.1097/00007632-200002150-00016

8. Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am*. 1990;72:403-408. https://doi.org/10.2106/00004623-199072030-00013

9. Johnson WE, Eisenstein SM, Roberts S. Cell cluster formation in degenerate lumbar intervertebral discs is associated with increased disc cell proliferation. *Connect Tissue Res*. 2001;42:197-207. https://doi.org/10.3109/03008200109005650

10. Le Maitre CL, Freemont AJ, Hoyland JA. Accelerated cellular senescence in degenerate intervertebral discs: a possible role in the pathogenesis of intervertebral disc degeneration. *Arthritis Res Ther*. 2007;9:R45. https://doi.org/10.1186/ar2198

11. Roberts S, Evans EH, Kletsas D, Jaffray DC, Eisenstein SM. Senescence in human intervertebral discs. *Eur Spine J*. 2006;15(Suppl 3):S312-S316. https://doi.org/10.1007/s00586-006-0126-8

12. Thompson JP, Pearce RH, Schechter MT, Adams ME, Tsang IK, Bishop PB. Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. *Spine*. 1990;15:411-415. https://doi.org/10.1097/00007632-199005000-00012

13. Palmgren T, Grönblad M, Virri J, Seitsalo S, Ruuskanen M, Karaharju E. Immunohistochemical demonstration of sensory and autonomic nerve terminals in herniated lumbar disc tissue. *Spine*. 1996;21:1301-1306. https://doi.org/10.1097/00007632-199606010-00004

14. Freemont AJ, Peacock TE, Goupille P, Hoyland JA, O'Brien J, Jayson MI. Nerve ingrowth into diseased intervertebral disc in chronic back pain. *Lancet*. 1997;350:178-181. https://doi.org/10.1016/s0140-6736(97)02135-1

15. Freemont AJ, Watkins A, Le Maitre C, et al. Nerve growth factor expression and innervation of the painful intervertebral disc. *J Pathol*. 2002;197:286-292. https://doi.org/10.1002/path.1108

16. Binch ALA, Cole AA, Breakwell LM, et al. Nerves are more abundant than blood vessels in the degenerate human intervertebral disc. *Arthritis Res Ther*. 2015;17:370. https://doi.org/10.1186/s13075-015-0889-6

17. Lama P, Le Maitre CL, Harding IJ, Dolan P, Adams MA. Nerves and blood vessels in degenerated intervertebral discs are confined to physically disrupted tissue. *J Anat*. 2018;233:86-97. https://doi.org/10.1111/joa.12817

18. Cunha C, Silva AJ, Pereira P, Vaz R, Gonçalves RM, Barbosa MA. The inflammatory response in the regression of lumbar disc herniation. *Arthritis Res Ther*. 2018;20:251. https://doi.org/10.1186/s13075-018-1743-4

19. Silva AJ, Ferreira JR, Cunha C, et al. Macrophages down-regulate gene expression of intervertebral disc degenerative markers under a pro-inflammatory microenvironment. *Front Immunol*. 2019;10:1508. https://doi.org/10.3389/fimmu.2019.01508

20. Samartzis D, Alini M, An HS, et al. Precision spine care: A new era of discovery, innovation, and global impact. *Global Spine J*. 2018;8:321-322. https://doi.org/10.1177/2192568218774044

21. Nachemson A. Lumbar intradiscal pressure. Experimental studies on post-mortem material. *Acta Orthop Scand Suppl*. 1960;43:1-104. https://doi.org/10.3109/ort.1960.31.suppl-43.01

22. Gries NC, Berlemann U, Moore RJ, Vernon-Roberts B. Early histologic changes in lower lumbar discs and facet joints and their correlation. *Eur Spine J*. 2000;9:23-29. https://doi.org/10.1007/s005860050004

23. Boos N, Weissbach S, Rohrbach H, Weiler C, Spratt KF, Nerlich AG. Classification of age-related changes in lumbar intervertebral discs: 2002 Volvo Award in basic science. *Spine*. 2002;27:2631-2644. https://doi.org/10.1097/01.BRS.0000035304.27153.5B

24. Sive JI, Baird P, Jeziorsk M, Watkins A, Hoyland JA, Freemont AJ. Expression of chondrocyte markers by cells of normal and degenerate intervertebral discs. *Mol Pathol*. 2002;55:91-97. https://doi.org/10.1136/mp.55.2.91

25. Rutges JPHJ, Duit RA, Kummer JA, et al. A validated new histological classification for intervertebral disc degeneration. *Osteoarthr Cartil*. 2013;21:2039-2047. https://doi.org/10.1016/j.joca.2013.10.001

26. Coventry MB, Ghormley RK, Kernohan JW. The intervertebral disc: its microscopic anatomy and pathology. 1945;

27. Friberg S. Anatomical studies on lumbar disc degeneration. *Acta Orthop Scand*. 1948;17:224-230. https://doi.org/10.3109/17453674808988942

28. Vernon-Roberts B, Pirie CJ. Degenerative changes in the intervertebral discs of the lumbar spine and their sequelae. *Rheumatol Rehabil*. 1977;16:13-21. https://doi.org/10.1093/rheumatology/16.1.13

29. Osti OL, Vernon-Roberts B, Moore R, Fraser RD. Annular tears and disc degeneration in the lumbar spine. A post-mortem study of 135 discs. *J Bone Joint Surg Br*. 1992;74:678-682. https://doi.org/10.1302/0301-620X.74B5.1388173

30. Vernon-Roberts B, Fazzalari NL, Manthey BA. Pathogenesis of tears of the anulus investigated by multiple-level transaxial analysis of the T12-L1 disc. *Spine*. 1997;22:2641-2646.

31. Haefeli M, Kalberer F, Saegesser D, Nerlich AG, Boos N, Paesold G. The course of macroscopic degeneration in the human lumbar intervertebral disc. *Spine*. 2006;31:1522-1531. https://doi.org/10.1097/01.brs.0000222032.52336.8e

32. Le Maitre CL, Freemont AJ, Hoyland JA. The role of interleukin-1 in the pathogenesis of human intervertebral disc degeneration. *Arthritis Res Ther*. 2005;7:R732-R745. https://doi.org/10.1186/ar1732

33. Walter BA, Torre OM, Laudier D, Naidich TP, Hecht AC, Iatridis JC. Form and function of the intervertebral disc in health and disease: a morphological and stain comparison study. *J Anat*. 2015;227:707-716. https://doi.org/10.1111/joa.12258

34. Tomaszewski KA, Henry BM, Gładysz T, Głowacki R, Walocha JA, Tomaszewska R. Validation of the intervertebral disc histological degeneration score in cervical intervertebral discs and their end plates. *Spine J*. 2017;17:738-745. https://doi.org/10.1016/j.spinee.2017.01.006

35. Lotz JC, Fields AJ, Liebenberg EC. The role of the vertebral end plate in low back pain. *Global Spine J*. 2013;3:153-164. https://doi.org/10.1055/s-0033-1347298

36. Berg-Johansen B, Jain D, Liebenberg EC, et al. Tidemark avulsions are a predominant form of endplate irregularity. *Spine*. 2018;43:1095-1101. https://doi.org/10.1097/BRS.0000000000002545

37. Baker JD, Harada GK, Tao Y, et al. The impact of modic changes on preoperative symptoms and clinical outcomes in anterior cervical discectomy and fusion patients. *Neurospine*. 2020;17:190-203. https://doi.org/10.14245/ns.2040062.031

38. Dudli S, Fields AJ, Samartzis D, Karppinen J, Lotz JC. Pathobiology of Modic changes. *Eur Spine J*. 2016;25:3723-3734. https://doi.org/10.1007/s00586-016-4459-7

39. Koivisto K, Karppinen J, Haapea M, et al. The Effect of Zoledronic Acid on Serum Biomarkers among Patients with Chronic Low Back Pain and Modic Changes in Lumbar Magnetic Resonance Imaging. *Diagnostics (Basel)*. 2019;9(4):212. https://doi.org/10.3390/diagnostics9040212

40. Määttä JH, Karppinen J, Paananen M, et al. Refined phenotyping of modic changes: imaging biomarkers of prolonged severe low back pain and disability. *Medicine*. 2016;95:e3495. https://doi.org/10.1097/MD.0000000000003495

41. Määttä JH, Karppinen JI, Luk KDK, Cheung KMC, Samartzis D. Phenotype profiling of Modic changes of the lumbar spine and its association with other MRI phenotypes: a large-scale population-based study. *Spine J*. 2015;15:1933-1942. https://doi.org/10.1016/j.spinee.2015.06.056

42. Mok FPS, Samartzis D, Karppinen J, Fong DYT, Luk KDK, Cheung KMC. Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. *Spine J*. 2016;16:32-41. https://doi.org/10.1016/j.spinee.2015.09.060

43. Määttä JH, Wadge S, MacGregor A, Karppinen J, Williams FMK. ISSLS prize winner: vertebral endplate (modic) change is an independent risk factor for episodes of severe and disabling low back pain. *Spine*. 2015;40:1187-1193. https://doi.org/10.1097/BRS.0000000000000937

44. Takatalo J, Karppinen J, Niinimäki J, et al. Association of modic changes, Schmorl's nodes, spondylolytic defects, high-intensity zone lesions, disc herniations, and radial tears with low back symptom severity among young Finnish adults. *Spine*. 2012;37:1231-1239. https://doi.org/10.1097/BRS.0b013e3182443855

45. Fleiss JL, Levin B, Cho PM. *The Measurement of Interrater Agreement. Statistical Methods for Rates and Proportions*. 3rd ed. 2004:598-626 https://onlinelibrary.wiley.com/doi/10.1002/0471445428.ch18. Accessed 2020 December.

46. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet*. 2018;391:2356-2367. https://doi.org/10.1016/S0140-6736(18)30480-X

47. Benneker LM, Heini PF, Anderson SE, Alini M, Ito K. Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. *Eur Spine J*. 2005;14:27-35. https://doi.org/10.1007/s00586-004-0759-4

48. Fields AJ, Liebenberg EC, Lotz JC. Innervation of pathologies in the lumbar vertebral end plate and intervertebral disc. *Spine J*. 2014;14:513-521. https://doi.org/10.1016/j.spinee.2013.06.075

49. Sharma A, Pilgram T, Wippold FJ. Association between annular tears and disk degeneration: a longitudinal study. *AJNR Am J Neuroradiol*. 2009;30:500-506. https://doi.org/10.3174/ajnr.A1411

50. Gruber HE, Hanley EN. Ultrastructure of the human intervertebral disc during aging and degeneration: comparison of surgical and control specimens. *Spine*. 2002;27:798-805.

51. Stefanakis M, Al-Abbasi M, Harding I, et al. Annulus fissures are mechanically and chemically conducive to the ingrowth of nerves and blood vessels. *Spine*. 2012;37:1883-1891. https://doi.org/10.1097/BRS.0b013e318263ba59

52. Peng B, Wu W, Hou S, Li P, Zhang C, Yang Y. The pathogenesis of discogenic low back pain. *J Bone Joint Surg Br*. 2005;87:62-67.

53. Wang Z-X, Hou Z-T, Hu Y-G. Anterior high-intensity zone in lumbar discs: Prevalence and association with low back pain. *Pain Med*. 2020;21:2111-2116. https://doi.org/10.1093/pm/pnaa236

54. Aprill C, Bogduk N. High-intensity zone: a diagnostic sign of painful lumbar disc on magnetic resonance imaging. *Br J Radiol*. 1992;65:361-369. https://doi.org/10.1259/0007-1285-65-773-361

55. Wilke H-J, Rohlmann F, Neidlinger-Wilke C, Werner K, Claes L, Kettler A. Validity and interobserver agreement of a new radiographic grading system for intervertebral disc degeneration: Part I. Lumbar spine. *Eur Spine J*. 2006;15:720-730. https://doi.org/10.1007/s00586-005-1029-9

56. Luoma K, Vehmas T, Kerttula L, Grönblad M, Rinne E. Chronic low back pain in relation to Modic changes, bony endplate lesions, and

disc degeneration in a prospective MRI study. *Eur Spine J*. 2016;*25*: 2873-2881. https://doi.org/10.1007/s00586-016-4715-x

57. Blumenkrantz G, Zuo J, Li X, Kornak J, Link TM, Majumdar S. In vivo 3.0-tesla magnetic resonance T1rho and T2 relaxation mapping in subjects with intervertebral disc degeneration and clinical symptoms. *Magn Reson Med*. 2010;*63*:1193-1200. https://doi.org/10.1002/mrm.22362

58. Johannessen W, Auerbach JD, Wheaton AJ, et al. Assessment of human disc degeneration and proteoglycan content using T1rho-weighted magnetic resonance imaging. *Spine*. 2006;*31*:1253-1257. https://doi.org/10.1097/01.brs.0000217708.54880.51

59. Krug R, Joseph GB, Han M, et al. Associations between vertebral body fat fraction and intervertebral disc biochemical composition as assessed by quantitative MRI. *J Magn Reson Imaging*. 2019;*50*:1219-1226. https://doi.org/10.1002/jmri.26675

60. Rodriguez AG, Slichter CK, Acosta FL, et al. Human disc nucleus properties and vertebral endplate permeability. *Spine*. 2011;*36*:512-520. https://doi.org/10.1097/BRS.0b013e3181f72b94

61. Marinelli NL, Haughton VM, Anderson PA. T2 relaxation times correlated with stage of lumbar intervertebral disk degeneration and patient age. *AJNR Am J Neuroradiol*. 2010;*31*:1278-1282. https://doi.org/10.3174/ajnr.A2080

62. Borthakur A, Maurer PM, Fenty M, et al. T1ρ magnetic resonance imaging and discography pressure as novel biomarkers for disc degeneration and low back pain. *Spine*. 2011;*36*:2190-2196. https://doi.org/10.1097/BRS.0b013e31820287bf

63. Zuo J, Joseph GB, Li X, et al. In vivo intervertebral disc characterization using magnetic resonance spectroscopy and T1ρ imaging: association with discography and Oswestry Disability Index and Short Form-36 Health Survey. *Spine*. 2012;*37*:214-221. https://doi.org/10.1097/BRS.0b013e3182294a63

64. Binch ALA, Cole AA, Breakwell LM, et al. Expression and regulation of neurotrophic and angiogenic factors during human intervertebral disc degeneration. *Arthritis Res Ther*. 2014;*16*:416. https://doi.org/10.1186/s13075-014-0416-1

65. Sharp CA, Roberts S, Evans H, Brown SJ. Disc cell clusters in pathological human intervertebral discs are associated with increased stress protein immunostaining. *Eur Spine J*. 2009;*18*:1587-1594. https://doi.org/10.1007/s00586-009-1053-2

66. Le Maitre CL, Freemont AJ, Hoyland JA. Localization of degradative enzymes and their inhibitors in the degenerate human intervertebral disc. *J Pathol*. 2004;*204*:47-54. https://doi.org/10.1002/path.1608

67. Law T, Anthony M-P, Chan Q, et al. Ultrashort time-to-echo MRI of the cartilaginous endplate: technique and association with intervertebral disc degeneration. *J Med Imaging Radiat Oncol*. 2013;*57*: 427-434. https://doi.org/10.1111/1754-9485.12041

68. Berg-Johansen B, Han M, Fields AJ, et al. Cartilage Endplate Thickness Variation Measured by Ultrashort Echo-Time MRI Is Associated With Adjacent Disc Degeneration. *Spine*. 2018;*43*:E592-E600. https://doi.org/10.1097/BRS.0000000000002432

69. Bailey JF, Fields AJ, Ballatori A, et al. The relationship between endplate pathology and patient-reported symptoms for chronic low back pain depends on lumbar paraspinal muscle quality. *Spine*. 2019; *44*:1010-1017. https://doi.org/10.1097/BRS.0000000000003035

70. Chen L, Battié MC, Yuan Y, Yang G, Chen Z, Wang Y. Lumbar vertebral endplate defects on magnetic resonance images: prevalence, distribution patterns, and associations with back pain. *Spine J*. 2020; *20*:352-360. https://doi.org/10.1016/j.spinee.2019.10.015

71. Jensen TS, Karppinen J, Sorensen JS, Niinimäki J, Leboeuf-Yde C. Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. *Eur Spine J*. 2008;*17*:1407-1422. https://doi.org/10.1007/s00586-008-0770-2

72. Karchevsky M, Schweitzer ME, Carrino JA, Zoga A, Montgomery D, Parker L. Reactive endplate marrow changes: a systematic

73. Fields AJ, Ballatori A, Han M, et al. Measurement of vertebral endplate bone marrow lesion (Modic change) composition with water-fat MRI and relationship to patient-reported outcome measures. *Eur Spine J*. 2021. https://doi.org/10.1007/s00586-021-06738-y

74. Thompson KJ, Dagher AP, Eckel TS, Clark M, Reinig JW. Modic changes on MR images as studied with provocative diskography: clinical relevance – a retrospective study of 2457 disks. *Radiology*. 2009;*250*:849-855. https://doi.org/10.1148/radiol.2503080474

75. Brinjikji W, Diehn FE, Jarvik JG, et al. MRI Findings of Disc Degeneration are More Prevalent in Adults with Low Back Pain than in Asymptomatic Controls: A Systematic Review and Meta-Analysis. *AJNR Am J Neuroradiol*. 2015;*36*:2394-2399. https://doi.org/10.3174/ajnr.A4498

76. Samartzis D, Borthakur A, Belfer I, et al. Novel diagnostic and prognostic methods for disc degeneration and low back pain. *Spine J*. 2015;*15*:1919-1932. https://doi.org/10.1016/j.spinee.2014.09.010

77. Pang H, Bow C, Cheung JPY, et al. The UTE disc sign on MRI: A novel imaging biomarker associated with degenerative spine changes, low back pain, and disability. *Spine*. 2018;*43*:503-511. https://doi.org/10.1097/BRS.0000000000002369

78. Purmessur D, Freemont AJ, Hoyland JA. Expression and regulation of neurotrophins in the nondegenerate and degenerate human intervertebral disc. *Arthritis Res Ther*. 2008;*10*:R99. https://doi.org/10.1186/ar2487

79. Krock E, Rosenzweig DH, Chabot-Doré A-J, et al. Painful, degenerating intervertebral discs up-regulate neurite sprouting and CGRP through nociceptive factors. *J Cell Mol Med*. 2014;*18*:1213-1225. https://doi.org/10.1111/jcmm.12268

80. Krock E, Currie JB, Weber MH, et al. Nerve growth factor is regulated by toll-like receptor 2 in human intervertebral discs. *J Biol Chem*. 2016;*291*:3541-3551. https://doi.org/10.1074/jbc.M115.675900

81. Lama P, Claireaux H, Flower L, et al. Physical disruption of intervertebral disc promotes cell clustering and a degenerative phenotype. *Cell Death Discov*. 2019;*5*:154. https://doi.org/10.1038/s41420-019-0233-z

82. Baumgartner L, Wuertz-Kozak K, Le Maitre CL, et al. Multiscale regulation of the intervertebral disc: achievements in experimental, in silico, and regenerative research. *Int J Mol Sci*. 2021;*22*:703. https://doi.org/10.3390/ijms22020703

83. Risbud MV, Shapiro IM. Role of cytokines in intervertebral disc degeneration: pain and disc content. *Nat Rev Rheumatol*. 2014;*10*: 44-56. https://doi.org/10.1038/nrrheum.2013.160

84. Samartzis D, Mok FPS, Karppinen J, Fong DYT, Luk KDK, Cheung KMC. Classification of Schmorl's nodes of the lumbar spine and association with disc degeneration: a large-scale population-based MRI study. *Osteoarthr Cartil*. 2016;*24*:1753-1760. https://doi.org/10.1016/j.joca.2016.04.020

85. Wang Y, Videman T, Battié MC. Lumbar vertebral endplate lesions: prevalence, classification, and association with age. *Spine*. 2012;*37*: 1432-1439. https://doi.org/10.1097/BRS.0b013e31824dd20a

86. Hassler O. The human intervertebral disc. A micro-angiographical study on its vascular supply at various ages. *Acta Orthop Scand*. 1969;*40*:765-772. https://doi.org/10.3109/17453676908989540

87. Hirsch C, Schajowicz F. Studies on structural changes in the lumbar annulus fibrosus. *Acta Orthop Scand*. 1952;*22*:184-231. https://doi.org/10.3109/17453675208989006

88. Hilton RC, Ball J, Benn RT. Vertebral end-plate lesions (Schmorl's nodes) in the dorsolumbar spine. *Ann Rheum Dis*. 1976;*35*:127-132.

89. Wagner AL, Murtagh FR, Arrington JA, Stallworth D. Relationship of Schmorl's nodes to vertebral body endplate fractures and acute endplate disk extrusions. *AJNR Am J Neuroradiol*. 2000;*21*:276-281.

morphologic and epidemiologic evaluation. *Skeletal Radiol*. 2005;*34*: 125-129. https://doi.org/10.1007/s00256-004-0886-3

90. Munir S, Freidin MB, Rade M, Määttä J, Livshits G, Williams FMK. Endplate defect is heritable, associated with low back pain and triggers intervertebral disc degeneration: a longitudinal study from twinsuk. *Spine*. 2018;43:1496-1501. https://doi.org/10.1097/BRS.0000000000002721

91. Williams FMK, Manek NJ, Sambrook PN, Spector TD, Macgregor AJ. Schmorl's nodes: common, highly heritable, and related to lumbar disc disease. *Arthritis Rheum*. 2007;57:855-860. https://doi.org/10.1002/art.22789

92. Mok FPS, Samartzis D, Karppinen J, Luk KDK, Fong DYT, Cheung KMC. ISSLS prize winner: prevalence, determinants, and association of Schmorl nodes of the lumbar spine with disc degeneration: a population-based study of 2449 individuals. *Spine*. 2010;35:1944-1952. https://doi.org/10.1097/BRS.0b013e3181d534f3

93. Takahashi K, Miyazaki T, Ohnari H, Takino T, Tomita K. Schmorl's nodes and low-back pain. Analysis of magnetic resonance imaging findings in symptomatic and asymptomatic individuals. *Eur Spine J*. 1995;4:56-59. https://doi.org/10.1007/BF00298420

94. Zehra U, Bow C, Lotz JC, et al. Structural vertebral endplate nomenclature and etiology: a study by the ISSLS Spinal Phenotype Focus Group. *Eur Spine J*. 2018;27:2-12. https://doi.org/10.1007/s00586-017-5292-3

95. Fields AJ, Battié MC, Herzog RJ, et al. Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): recommendations from the ISSLS Degenerative Spinal Phenotypes Group. *Eur Spine J*. 2019;28:2266-2274. https://doi.org/10.1007/s00586-019-06119-6

96. Doktor K, Jensen TS, Christensen HW, et al. Degenerative findings in lumbar spine MRI: an inter-rater reliability study involving three raters. *Chiropr Man Therap*. 2020;28:8. https://doi.org/10.1186/s12998-020-0297-0

97. Smith BM, Hurwitz EL, Solsberg D, et al. Interobserver reliability of detecting lumbar intervertebral disc high-intensity zone on magnetic resonance imaging and association of high-intensity zone with pain and anular disruption. *Spine*. 1998;23:2074-2080. https://doi.org/10.1097/00007632-199810010-00007

98. Zook J, Djurasovic M, Crawford C, Bratcher K, Glassman S, Carreon L. Inter- and intraobserver reliability in radiographic assessment of degenerative disk disease. *Orthopedics*. 2011;34. https://doi.org/10.3928/01477447-20110228-07

99. Jamaludin A, Lootus M, Kadir T, et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J*. 2017;26:1374-1383. https://doi.org/10.1007/s00586-017-4956-3

100. Ishimoto Y, Jamaludin A, Cooper C, et al. Could automated machine-learned MRI grading aid epidemiological studies of lumbar spinal stenosis? Validation within the Wakayama spine study. *BMC Musculoskelet Disord*. 2020;21:158. https://doi.org/10.1186/s12891-020-3164-1

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.