Single-cell Atlas Unveils Cellular Heterogeneity and Novel Markers in Human Neonatal and Adult Intervertebral Discs

Wensen Jiang[1,2], Juliane D. Glaeser[1,2,5,6], Khosrowdad Salehi[1,2], Giselle Kaneda[1,2], Pranav Mathkar[2], Anton Wagner[2], Ritchie Ho[2,3,4,#], Dmitriy Sheyn[1,2,5,6,7,#,*]

[1]Orthopaedic Stem Cell Research Laboratory, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[2]Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[3]Center for Neural Sciences and Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[4]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[5]Department of Orthopedics, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[6]Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[7]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA, 90048

[#]These authors share equal seniority

*Corresponding author

Correspondence:

Dmitriy Sheyn, PhD

Board of Governors Regenerative Medicine Institute

Department of Orthopedics

Department of Surgery

Department of Biomedical Sciences

Cedars-Sinai Medical Center

AHSP A8308, 8700 Beverly Boulevard

Los Angeles, CA 90048

Email: dmitriy.sheyn@csmc.edu

SUMMARY

The origin, composition, distribution, and function of cells in the human intervertebral disc (IVD) has not been fully understood. Here, cell atlases of both human neonatal and adult IVDs have been generated and further assessed by gene ontology pathway enrichment, pseudo-time trajectory, histology, and immunofluorescence. Comparison of cell atlases revealed the presence of several sub-populations of notochordal cells (NC) in the neonatal IVD and a small quantity of NCs and associated markers in the adult IVD. Developmental trajectories predicted that most neonatal NCs develop into adult nucleus pulposus cells (NPCs) while some keep their identity throughout adulthood. A high heterogeneity and gradual transition of annulus fibrosus cells (AFCs) in the neonatal IVD was detected and their potential relevance in IVD development was assessed. Collectively, comparing single-cell atlases between neonatal and adult IVDs delineates the landscape of IVD cell biology and may help discover novel therapeutic targets for IVD degeneration.

KEYWORDS

single-cell RNA sequencing; human intervertebral disc; intervertebral disc development; notochordal cells; nucleus pulposus cells; annulus fibrosus cells

INTRODUCTION

At least 30% of total adult population is suffering from lower back pain, which originates in intervertebral disc (IVD) degeneration (Andersson, 1999; de Schepper et al., 2010; Frank et al., 1996; Macfarlane et al., 1999) To date, most treatments of IVD degeneration are limited to invasive surgical interventions, such as disc replacement and spinal fusion, or pain management that is not addressing the underlying cause of IVD degeneration (Knezevic et al., 2017).

The IVD is a shock-absorbing structure that connects two adjacent vertebras and enables spine movement (Buckley et al., 2018; Vergroesen et al., 2015). The human IVD cell composition is highly heterogenous; nucleus pulposus (NP) is the inner core and the annulus fibrosus (AF) is the outer region that is confined by two endplates sandwiching the disc (Sun et al., 2020). It is widely accepted the embryonic notochord develops into the NP and that notochordal cells (NCs) disappear within the first decade of life (Séguin et al., 2018). However, the NC remnants or dormant NC can be found in the adult human IVD (McCann and Séguin, 2016; Wang et al., 2008). Additional study showed that the NCs or NC-like NP cells (NPCs) in human persist throughout life (Risbud and Shapiro, 2011). The eventual fate of human NCs is not conclusive to date, but understanding its development has considerable potential to benefit cell therapies of IVD degeneration. Animals that keep the NCs in adulthood like cats and pigs do not to exhibit disc degeneration (Hunter et al., 2003; Sheyn et al., 2019). NCs hold great potential to rejuvenate a degenerated human IVD and this potential has already been demonstrated in a mini pig model (Sheyn *et al.*, 2019).

Recent advances in single-cell RNA sequencing (scRNA-seq) allow for creation of cell atlases that unveil rare sub-populations, delineate cellular heterogeneity, identify new markers, and predict developmental trajectories (Ho et al., 2021; Kelly et al., 2020; Mathys et al., 2019; Setty et al., 2019; Stark et al., 2019). Unraveling the single-cell atlas and transcriptomes of the human IVD at different ages will largely extend our understanding of its cell biology and development. To date its heterogeneity, particularly in the early development, has not been sufficiently shown due

to the limitation of traditional bulk-sequencing (Fernandes et al., 2020). Previously, multiple bulk proteomics studies delineated the general cell heterogeneity of human IVD (Rodrigues-Pinto et al., 2018; Tam et al., 2020b) and made comparison at different developmental stages (Tam *et al.*, 2020b). Recent studies of single-cell atlases in non-human vertebrates resolved the cellular heterogeneity in bovine caudal IVD (Panebianco et al., 2021) and found stem cells in rat IVD (Wang et al., 2021). Another study provided single-cell analysis of human adult IVD that identified the NPC and AFC populations (Fernandes *et al.*, 2020). A recent study identified chondrocytes as dominating cell population in human IVD and found small NC sub-populations (Gan et al., 2021). Yet no study has investigated cell atlases from neonatal human IVD and thus elaborated the role of NCs in development and hemostasis at a single-cell resolution.

Here, we analyze a human IVD single-cell atlas with direct head-to-head comparison between neonatal and adult human subjects to reveal the identities of all IVD populations and sub-populations of both developmental stages, discover rare cell populations and novel markers thereof, and predict the developmental trajectories from neonatal to adult IVD. We isolated and collected cells from both neonatal and adult individuals postmortem, and analyzed them with scRNA-seq. An in-depth gene ontology term enrichment and pathway analysis was performed and developmental trajectories from neonatal to adult IVD were predicted. We further validated our findings with histology and immunofluorescence.

RESULTS

Human Neonatal and Adult IVD Cells are Comprised of 14 Distinct Sub-populations

Histological analysis of neonatal and adult IVD tissue demonstrated two distinct tissue regions. In the neonatal IVD, region A contained aggregated cells and underdeveloped extracellular matrix (ECM) structures and region B contained sparsely distributed cells and a high ECM content tissue (Fig. 1a). Vessels and red blood cells can be clearly identified in Region A (Fig. 1a). In contrast, the adult IVD does not exhibit different regions. Instead, one uniform and homogenous region contains sparsely distributed cells whose cytoplasm was greatly larger and cell density lower than in neonatal IVD (Fig. 1b).

IVD tissue harvested from three spinal disc levels from a neonatal spinal column (L3-L5), demonstrated a consistent overlap after integration (Fig. S1). indicating consistent quality of the data.

Dimensional reduction based on the unsupervised UMAP method sorted all cells into 14 clusters that could be classified as 14 cell sub-populations residing in both neonatal IVD (Fig. 1c) and adult IVD (Fig 1d). The expression levels of classical IVD markers in each cluster were shown as dot plot for neonatal IVD (Fig. 1e) and adult IVD (Fig. 1f). The selection of classical IVD markers includes NC markers [*MAP1B* (Rodrigues-Pinto *et al.*, 2018)*, SOX4* (Bhattaram et al., 2010)*, SOX17* (D'Amour et al., 2005; Sheyn *et al.*, 2019)*, ITGA6 and BASP1* (Sheyn *et al.*, 2019)*)* NPC markers (*ACAN* (Fernandes *et al.*, 2020; Risbud et al., 2015; Tam et al., 2020a), *COL2A1* (Fernandes *et al.*, 2020; Risbud *et al.*, 2015; Tam *et al.*, 2020a) and *SOX9* (Sheyn *et al.*, 2019)*) ,* AFC markers (*COL1A1* (Tam *et al.*, 2020b; van den Akker et al., 2017), *CALR* (Tam *et al.*, 2020b) and *HSPA6* (Takao and Iwaki, 2002)] as well as other previously shown markers *[MEG3 for non-mesenchymal cells* (Chen et al., 2017)*, CD44* (Schumann et al., 2015) *and CD14* (Ziegler-Heitbrock and Ulevitch, 1993) *for monocytes,* and *HBB* (Talamo et al., 2003) and *HBA1* (Pandey and Rizvi, 2011) *for red blood cells*. These classical markers were mostly expressed in one or more clusters. Based on these markers we annotated each cluster (Fig. 1g). The distribution of cells from neonatal versus adult IVD in each annotated sub-population, as normalized by total cells of their respective age show that neonatal IVD dominantly contributed to the cell

compositions of NC1-2, NP2 IAF2, OAF1, and OAF3 (Fig. 1h). Adult IVD dominantly contributed to the cell compositions of NP1. In the rest of sub-populations, the contribution was more or less evenly distributed.

Human Neonatal Intervertebral Disc Contained Two Distinct Notochordal Cell Sub-Populations and Adult Contained One Sub-Population

Cluster 1 and 2 in the neonatal IVD were identified as NC sub-populations (Fig. 1c-d) based on their expression of NC markers, *MAP1B* (Rodrigues-Pinto *et al.*, 2018) and *SOX4* (Bhattaram *et al.*, 2010), and have been assigned to NC1 and NC2 (Fig. 2a). NC1 and NC2 were distinct from each other: NC1 expressed *BASP1* and *CD44*, while NC2 expressed *SOX17* and *ITGA6* (Fig. 2a). The NC sub-populations were also found in adult IVD (Fig. 2a). Unlike neonatal, the adult IVD only contained the *MAP1B*+ NC1 sub-population (Fig. 2a). The NC2 sub-population in adult IVD, though existing, did not strongly express its markers (Fig. 2a). Immunostaining of the NC markers, *MAP1B* and *SOX17,* confirmed their expression in both neonatal and adult IVDs (Fig. 2b) on the protein level. These proteins showed a higher overall expression in neonatal than in adult IVD. We also observed the NC1 expressed immune-relevant marker *CD44* (Fig. 2a) and confirmed the presence of CD44 proteins in both ages (Fig. 2b).

Pathway networks were further created based on the list of all enriched genes for NC1 and NC2 (Fig. S2a, b), that is, 458 genes for NC1 and 507 genes for NC2. Our results show NC1 preferentially demonstrated the activation of cancer and osteoarthritis-related pathways, inflammatory cytokines (IL2, IL5 and IL15), inflammatory and anti-pathogen pathways (STAT1, STAT3), and stem cell pluripotency factor FGF2 (Fig. S2a). NC2 demonstrated the activation of RNA transcription, the cytoskeleton organization, and the osteoarthritis-related pathways, and the deactivation of Brachyury (*TBXT*), Microphthalmia-associated Transcription Factor (*MITF)*, and SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin (*SMARCD3)*, which is further associated with the deactivation of connective tissue cell pathways and the collagen-relevant Glycoprotein VI (*GP6*) signaling pathway (Fig. S2b). *MITF* was previously shown to be associated with osteoclast activities (Lu et al., 2010) and GP6 was shown to induce collagen deposition (Stegner et al., 2014). The NC2 gene enrichment indicates the

downregulation of osteoclast activity, leading to less connective tissue cell activity and less cell-ECM interactions.

AKR1C1, APOE, and *FABP4* are Novel Markers for Notochordal Cells in Human Intervertebral Disc

We selected several NC markers from top enriched genes as shown in Fig. 2c. Specifically, *AKR1C1* is the NC1 marker for neonatal IVD, *FABP4* is the universal NC marker (NC1 and NC2) for neonatal IVD, and *APOE* is the neonatal/adult dual-age marker for NC1. We compared the expression level of *AKR1C1*, *APOE*, and *FABP4* in neonatal and adult IVD due to their possible age-dependent specificity (Fig. 3a). *AKR1C1* is only specific to NC1 and *FABP4* to NC (both NC1 and NC2) in neonatal IVD, but such specificity is lost in adult IVD (Fig. 3a). Yet *APOE* seems to have good specificity to NC1 sub-population in both neonatal and adult (Fig. 3a). The immunostaining of novel NC markers, *AKR1C*, *FABP4,* and *APOE,* confirmed their expression in both neonatal and adult IVD, but a higher expression was observed in neonatal tissue (Fig. 3b). All markers show overlap with cell nucleus besides *AKR1C* that appears in the cytoplasm. The presence of these proteins confirms the expression of the novel NC markers, *AKR1C1*, *FABP4*, and *APOE* (Fig. 3a)*,* in both neonatal and adult IVD.

We further investigated the role of the new markers in the regulatory networks and biofunctions. We found *AKR1C1* and *AKR1C2* to play a critical role in the IL1 cascade in neonatal NC1 (Fig. 3c). The activation of *AKR1C1* is associated with multiple biological functions as shown in Fig. 3c. Specifically, the upregulation of *AKR1C1* activates multiple tumor and cancer-related functions, as shown in Fig. 3d. Next, the novel *APOE* marker was shown as an upstream regulator in a network in neonatal NC1 that led to the activation of organization of cytoplasm function (Fig. S2c). The NC marker *MAP1B* was also in the same network as an upstream regulator (Fig. S2c). In Fig. 3e, the novel marker *FABP4* was in a more complicated network related to tumor and neoplasm. In this network, *FABP4* was activated by IL4 (Fig. 3e).

NPCs expressing their classical markers, *ACAN* (Fernandes *et al.*, 2020; Risbud *et al.*, 2015; Tam *et al.*, 2020a), *COL2A1* (Fernandes *et al.*, 2020; Risbud *et al.*, 2015; Tam *et al.*, 2020a), and *SOX9* (Sheyn *et al.*, 2019) were detected in both neonatal IVD (Fig. 4a) and adult IVD (Fig. 4b). The entire NPC population contained four different sub-populations labeled with their specific markers: NP1 labeled with *FABP5*, NP2 labeled with *OGN*, NP3 labeled with *FGFBP2*, NP4 labeled with *ANXA1*.The clusters 3, 4, 5, and 6 (Fig. 1c-d) thus have been designated as NP1, NP2, NP3 and NP4 respectively. The expression of these NPC markers is shown in the violin plots (Fig. 4a-b). However, four NPC sub-populations from the same age are barely distinguishable (Fig. 4a, b). The sub-population markers are not very specific as the gene expression distributed on the UMAP did not always match the clustering on the UMAP (Fig. 4a, b). The four NPC sub-populations may have similar transcriptomic profiles. The AFC marker levels are shown as controls in Fig. 4a, b. The AFC marker expression levels are very low among all NPC populations in both the neonatal and adult IVD, which confirmed the NPC clusters are not overlapping with AFCs.

The combined NPC populations showed substantial differences in different ages. Figure S3a-b shows the summary of network and pathways for the four NPC subtypes taken together. Figure S3a shows the deactivation of the differentiation of epithelial tissues, cell proliferation of carcinoma cell lines, and the unfolded protein response (UPR) in neonatal NPCs. The whole network contains key downregulated cytokines like IL2, IL18, and IL17A (Fig. S3a). Figure S3b shows the activation of embryonic cell differentiation, formation of actin stress fibers, development of cytoplasm, and osteoarthritis pathway in adult NPCs. Notably, *TGFB3, BCL6, BMP2, SMARCD3,* and *EPAS1* are the upregulated key regulators that activated the cascade network (Fig. S3b).

NPCs are critical to the hemostasis and development of ECM (Frapin et al., 2019). We also observed the age greatly affected the critical ECM proteins (Fig. 4c). Figure 4c shows the abundant presence of collagen type 1 (*COL1A1*) in the extracellular space and aggrecan in the cytoplasm in the neonatal IVD. The signal of intracellular *COL1A1* appears to be weak in neonatal

IVD (Fig. 4c). Figure 4c shows the opposite trend for adult IVD: the staining of *COL1A1* in the extracellular space of neonatal IVD is negligible, but the staining in cytoplasm is strong; the staining of aggrecan in the extracellular space is strong but in the cytoplasm is weak (Fig. 4c).

Notochordal Cell and Nucleus Pulposus Cell Populations Exhibited Different Gene Ontology Enrichments, Pathways, and Regulators

NC and NPC populations were compared in both neonatal and adult IVDs (Fig. 4d-f). In this analysis, NC1-2 were combined as NCs, and NP1-4 were combined as NPCs (Fig. 4d). We also showed top-scored key regulatory networks in neonatal NCs and adult NPCs in Fig. S3c-d. We found positive regulation of cartilage and connective tissue development in adult NPCs (Fig. S3d). Several inflammatory and tumorigenic factors and pathways (*IL13, Notch, VEGF*, and *HGF, STAT*), as well as angiotensin precursor *AGT* (Lu et al., 2016) were upregulated in NC (Fig. S3e). In NPCs, upregulation of osteogenic factor *BMP2* (Cai et al., 2021), osteoclast regulator *MITF* (Lu *et al.*, 2010), DNA-binding transcription factor-related *BCL6*, NPC and skeletal development marker *SOX9* (Bi et al., 1999; Sheyn *et al.*, 2019; Zhou et al., 2006), and embryonic and notochordal marker *TBXT* (Sheyn *et al.*, 2019; Zhang et al., 2020) was detected (Fig. S3e). The trend seems to be age-independent except for STAT, which was not detected in the adult IVD (Fig. S3e).

Different cell types (NC vs. NPC) demonstrated hugely distinct transcriptomes, but the differences between the ages (neonatal vs. adult) are much smaller, as summarized in the Circos plot (Fig. 4e) where the overlap of expressed genes mostly occurred among the same cell types. NCs showed thrombogenic, inflammatory, and osteoarthritis-like genes and are likely involved in promoted cell activities and the growth of connective tissue (Fig. 4f). NCs in adult IVD showed additional resemblance to immune cell transcriptome when compared to the neonatal (Fig. 4f). NPCs showed genes that are osteogenic and embryonic and are likely involved in the musculoskeletal development and connective tissue function (Fig. 4f). NPCs in adult IVD showed additional involvement in cartilage development, cytoplasm development and actin formation when compared to the neonatal (Fig. 4f).

Annulus Fibrosus Cells in Neonatal IVD is Highly Heterogenous, and their Expression of Collagen-Relevant Genes is Location-Dependent

COL1A1 is a known AFC marker (Tam *et al.*, 2020b; van den Akker *et al.*, 2017). *CALR* is a marker for the outer AF (Tam *et al.*, 2020b) and heat shock proteins (*HSPs*, e.g. *HSPA6*, *HSPA1A*, etc.) are markers for the most outer layer of the AF (Takao and Iwaki, 2002). Thus, the *COL1A1, CALR,* and *HSPA6* were used to identify AFC populations (Fig.5a, b). The cluster 7 and 8 (Fig. 1c-d) have been assigned to inner AF sub-populations, IAF2 and IAF1, due to their expression of *COL1A1* (Fig.5a, b) and the fact that they are close to NPC populations on the UMAP (Fig. 1c, d). The clusters 9, 10, and 11 (Fig. 1c-d) were assigned to OAF1, OAF2, and OAF3 respectively due to their expression of *CALR* (cluster 9, 10) and *HSPA6* (cluster 11, Fig. 5a, b). We have found high degree of heterogeneity among neonatal AFC populations (Fig. 5c). We identified *LGALS1, OGN, HES1, ZNF385D,* and *HSPA1A* as the AFC sub-population markers for IAF1-2 and OAF1-3. Moreover, several collagen markers were also detected in neonatal IVD, specifically *COL1A1, COL3A1, COL5A1, COL5A2, COL6A3, COL12A1* (Fig. 5d). Strikingly, their expression showed a decreasing trend assuming the location of subtypes follows the order of IAF2, IAF1, OAF1, OAF2 and OAF3 from inner core to the outer region (Fig. 5d). The gradient in neonatal IVD appears to be higher than in the adult (See in Fig. S5). Similar trend was observed in the enriched biofunctions (Fig. 5e) and key pathway analysis (Fig. 5f) across the neonatal IAF and OAF populations. Among the pathways listed in Fig. 5f, we show the mechanistic scheme of Integrin signaling pathway and the relevant gene expression in IAF2 in Fig. S7, UPR signaling pathway and the relevant gene expression in OAF2 in Fig. S8. The illustration in Fig. 5g shows the postulated structure transition from neonatal to adult IVD based on these spatially dependent gene expression patterns.

Resolved Single-cell Atlas for Human Neonatal and Adult IVD with Clusters Assigned to Sub-populations

We assigned cluster 12 into non-mesenchymal cells (non-MC) due to their *MEG3* expression, a marker for non-mesenchymal, tumor-suppressing cells (Chen *et al.*, 2017). Cluster 13 is a satellite

cell population on the UMAP (Fig. 1c-d). It shared the *CD44* expression as NC1 but lacked all other NC markers (Fig. 1e, f). Thus, we assigned cluster 13 to immune-like cells (IC) due to its expression of immune cell marker CD14 (Ziegler-Heitbrock and Ulevitch, 1993) and CD44 (Schumann *et al.*, 2015) (Fig. 6a). Cluster 14 was assigned to red blood cells (RBC) due to its expression of hemoglobin markers *HBB* (Talamo *et al.*, 2003) and *HBA1* (Pandey and Rizvi, 2011) (Fig. 6a). In addition to classical markers that were used to assign cell identities, Figure 6a also includes novel markers we found for NC, AFC and NPC populations. A heatmap of the expression levels of top 10 enriched genes for all sub-populations shows the high specificity of our enriched genes and the high cellular heterogeneity in IVD structures (Fig. S6). Collectively, the single-cell atlas of human IVD heterogeneously comprised of NC, NPC, AFC, IC, non-MC, and RBC populations, which can be more precisely classified into 14 different subtypes: NC1-2, NP1-4, IAF1-2, OAF1-3, non-MC, IC, and RBC (Fig. 6a).

The fully resolved single-cell atlases were plotted in three-dimensional (3D) UMPAs (Fig. 6b) for a more accurate and intuitive visualization. In the 3D atlas, NC1, NC2, IC, and non-MC visually appear to be away from the main continents. The NPC sub-populations (NP1-NP4) tended to cluster together with vague boundaries whereas AFC sub-populations (IAF1-2, OAF1-3) are distant from each other with clear boundaries. Comparing adult to neonatal IVD, the IAF2 and OAF3 almost disappeared in adult (Fig. 6b), which is in line with the quantitative results (Fig. 1h). NPC populations occupied a dominantly larger portion in adult than in neonatal IVD (Fig. 6b).

One of the key pathways involved in the IVD is Wnt/β-catenin pathway (Kondo et al., 2011). We found that the Wnt/β-catenin pathway is activated in neonatal NC but downregulated in neonatal and adult NPCs (Fig. 6c). We further expanded the key genes in Wnt/β-catenin pathway and showed the expression of *SOX4* regulator (also the NC marker) was upregulated in neonatal NCs compared to neonatal and adult NPCs, whereas another regulator gene *SOX9* was downregulated in neonatal NCs compared to neonatal and adult NPCs (Fig. 6d).

The Pseudo-time Trajectory Predicted Five Developmental States from Neonatal Cells into Adult Nucleus Pulposus Cells with Some Notochordal Cells Preserved Throughout Adulthood.

The trajectory color-coded by age shows that neonatal cells occupy two branches and adult cells occupy the third branch (Fig. 7a). There is another small side-branch composed of both neonatal and adult cells (Fig. 7a). ECM gene expression has changed over pseudo-time. *ACAN* expression increased over pseudo-time while *COL1A1* decreased. *COL2A1* increased over pseudo-time initially but then decreased dramatically into later pseudo-time in adult IVD (Fig. 7b). The *SOX4,* as a positive regulator of the Wnt/β-catenin canonical pathway (Sinner et al., 2007), decreased over pseudo-time but *SOX9,* as a negative regulator (Sinha et al., 2021), increased (Fig. 7c), which is consistent with Fig. 6d.

NC populations are mainly positioned in the two neonatal branches, which show different cell compositions (Fig. 7d): one contains NCs, IAFCs, and OAFCs; the other one contains NCs, IAFCs, and NPCs. The adult branch mostly consists of NPCs (Fig. 7d). The side-branch consists mostly of the OAFCs and NPCs (Fig. 7d). The NC marker expression in NCs and the NPC marker expression in NPCs also change over the pseudo-time (Fig. 7e).

We found five cell states along the trajectory (Fig. 7f). The annotation of each state depends on the trajectories of age (Fig. 7a) and cell types (Fig. 7d). State 1 has been annotated as Neonatal Outer Region because the cell composition of its branch is dominated by IAFCs and OAFCs that were presumably resided in the Outer Region as defined in Fig. 5g. State 2 has been annotated as Adult Outer Region (IAFC) because it contains mainly IAFCs that were also presumably resided in the Outer Region as defined in Fig. 5g. State 3 is a short intermediate state which did not show any clear preference. State 4 has been annotated as Adult Outer Region because it contains mainly OAFCs that were presumably resided in the Outer Region as defined in Fig. 5g. State 5 has been annotated as Adult Inner Core (NPC) because it mainly contains NPCs that were presumably resided in the Inner Core as defined in Fig. 5g. The change of expression levels of key molecules during the developmental process are also shown over pseudo-time in Fig. 7g, including *SOX4*, *SOX9*, *COL2A1*, *AKR1C1*, and *ACAN*. State 1 might indicate early developmental stage whereas State 2, 4, 5 might indicate later developmental stage. On this ground, we predicted the direction of trajectory started from the neonatal State 1 and branched

into adult State 2 and intermediate State 3, and the State 3 further branched into adult State 4 and 5 (grey arrows, Fig. 7g). The direction from State 2 to State 3 is possible due to the fact that State 5 positioned later in pseudo-time than State 2, but this claim lacks solid evidence (dashed arrow).

DISCUSSION

This study provides the first direct comparison of cell atlases at single-cell level between neonatal and adult human IVDs. With the help of scRNA-seq technology, we discovered *AKR1C1*, *FABP4*, and *APOE* as novel markers for human NCs. Our data also supported the previous assumption of the presence of a small quantity of NCs in the adult. It also models the developmental process from neonatal IVD, composed of various cell types, into adult IVD, composed mainly of NPCs. Lastly, AFCs in neonatal IVD were found to be highly heterogenous with a strong gradient profile that most probably correlated with spatial distribution.

This study identifies *AKR1C1*, *FABP4*, and *APOE* as highly specific markers for NC populations in the human neonatal and adult IVD (Fig. 3a, b). Although our network analysis demonstrated both *AKR1C1* and *FABP4* to be associated with tumorgenicity (Fig. 3c-e), these genes have also been associated with cell metabolism and may be therefore linked to increased cell activity and biosynthesis in early development. For example, *AKR1C1* is an enzyme that catalyzes NAPDH-dependent reductions (Zeng et al., 2017) and is responsible for hormone secretion (Marín et al., 2009). *FABP4* is expressed in embryonic cartilage and bone tissues (Urs et al., 2006). In addition, *AKR1C1* and *FABP4* also mark the inflammatory response in early-stage development, as shown by IL1-activated *AKR1C1* in neonatal NC1 sub-population (Fig. 3c). Similarly, we also demonstrated IL4-activated *FABP4* to be involved in inflammation in early developmental stages (Fig. 3e). In previous research, the IL1 cascade was already considered an inflammatory and catabolic marker for adult and degenerative human IVD (Johnson et al., 2015). To our knowledge, its occurrence in neonatal IVD has not been previously reported. Noteworthy, *FABP4* appears to be a robust NC2 marker in both neonatal and adult IVD, but *AKR1C1* lost its marker specificity in adult IVD (Fig. 3a, b). This may indicate a reduced NC cell metabolism in the adult IVD compared to neonatal.

Our results show that *APOE* is a highly specific NC1 marker in both neonatal and adult IVD (Fig. 3a, b), in contrast to a previous study in mouse IVD that demonstrated *APOE* to be expressed in

both AFCs and NPCs (Zhang et al., 2013). Differences in findings may be due to the different species investigated. Moreover, we noticed the *APOE* may have different roles than *AKR1C1* and *FABP4*. Network analysis showed the involvement of *APOE* in cytoplasm organization, in which *APOE* is a coactivator with classical NC marker *MAP1B* (Fig. S2c). A prior review by Huang et al. suggested the role of intracellular *APOE* in cellular processes including cytoskeletal assembly and stability, mitochondrial integrity, and function, in addition to its well-known role in lipid distribution (Huang and Mahley, 2014). Therefore, *APOE* may play an important role in these processes in NCs.

Isolating and identifying rare NC populations in adult IVD is a challenging task due to their small quantity. Our single-cell atlas detected rare NC populations in adult IVD expressing *MAP1B* marker, and further unveiled two distinct NC sub-populations, NC1 and NC2 in neonatal IVD, and confirmed the presence of NC1 in adult IVD (Fig. 2a). To our knowledge, only one recent single-cell study by Gan et al. reported a NC population in adult IVD. However, they used *TBXT* to identify NCs, whereas we used *MAP1B* (Gan *et al.*, 2021). Other studies, including proteomics (Tam *et al.*, 2020b) and scRNA-seq (Fernandes *et al.*, 2020) of human adult IVD, reported the presence of NPCs and AFCs, but did not report NC populations. The differences in findings might be a result of different samples analyzed or analysis techniques (Fernandes *et al.*, 2020; Tam *et al.*, 2020b). Partly in line with our study, Richardson et al. demonstrated the presence of a sub-population with NC-like phenotype but NPC-like morphology in adult IVD using qPCR and immunohistochemical analysis (Richardson et al., 2017). In contrast to our findings (Fig. 2a), they detected an expression of *CD24* and *TBXT*. In a later study, Rodrigues-Pinto et al. from the same group reported the pre-selected *CD24*+ cells contained NCs expressing MAP1B (Rodrigues-Pinto *et al.*, 2018), in fetal IVD. To our knowledge, our study is the only report demonstrating *MAP1B*+ NC populations in adult IVD. We are also the only report for two distinct NC sub-populations. The difference in results might be due to several reasons: Firstly, the work by Richardson et al. focused on the *CD24*+ sub-population as NC-like phenotype and observed their NP-like morphology (Richardson *et al.*, 2017), while our study is focused *MAP1B*+ NC cells, since interestingly CD24+ cells were not identified by our analysis as a distinct cluster. Therefore, their *CD24*+ sub-populations may be a different, more heterogenous population across our clusters. Secondly,

Rodrigues-Pinto et al. sorted and specifically analyzed *CD24*+ cells (Rodrigues-Pinto *et al.*, 2018), while our study did not include any pre-selection process and unbiasedly analyzed all cells. Lastly, our scRNA-seq technique identified two NC populations, which we divided into NC1 and NC2 (Fig. 2a), which were not detected in previous studies (Fernandes *et al.*, 2020; Gan *et al.*, 2021).

Historically, NPCs are considered to be derived from NC lineage (Risbud et al., 2010). We confirmed the close lineage relations between NCs and NPCs and further unveiled the transition process. Firstly, our developmental trajectories predicted that NCs in the neonatal IVD develop into NPCs in the adult IVD via five distinct cell states (Fig. 7f). To the best of our knowledge, no previous study predicted the developmental trajectories crossing ages (early vs. late development) and cell types (NC versus NPC). One recent study predicted the differentiation of NPCs alone based on pseudo-time trajectories without specifying the actual ages (Gan *et al.*, 2021), however it didn't include neonatal IVD that is rich in NCs. Secondly, we found significant distinctions between neonatal NC and adult NPC populations at molecular, cellular, and tissue levels. We showed that neonatal NCs have some tumor-like properties and are involved in tissue growth (Fig. 4d-f), as suggested by their expression of relevant genes and pathways (*IL13*, *Notch* signaling, *STAT* signaling, and *VEGF*, Fig. S3e). In contrast, adult NPCs appear to be more mature, with closer association to fully differentiated connective tissue and cartilage (Fig. 4d-f), and they expressed *BMP2*, *BCL6*, and *SOX9* (Fig. S3e). Our findings are in line with a previous IVD single-cell transcriptomic analysis showing *BMP2* and *SOX9* expression in some adult sub-populations (Gan *et al.*, 2021). Changes were also observed in cell morphology. The adult NPCs have large cytoplasm, as opposed to the small neonatal NPCs (Fig. 1a, b). The large cytoplasm of NPCs (Fig. 1b) contradicts a previous report stating that NCs are larger than NPCs (Rodrigues-Pinto *et al.*, 2018). The matrix also changed from a fibrous, less-developed appearance to a more mature cartilage-like, (Fig. 1a, b), as well as the ECM gene expression transformed from a collagens-rich (*COL1A1* and *COL2A1*) to aggrecan-rich (*ACAN*) in the later stage (Fig. 7b). It is known that *COL2A1* ensures the removal of notochord during the development (Aszódi et al., 1998). Thus, the decrease of *COL2A1* (Fig. 7b, g) may mark the successful removal of notochord and completion of NC to NPC transition. These distinctions at molecular, cellular, and tissue levels highlight the changes occur in neonatal NC during their development into adult NPC. Though

most NC differentiated into adult NPC, some small quantity of NC kept its identity throughout the entire trajectories (Fig. 7d). The NC quantity is so small (Fig. 6b, NCs constituted only 1.1% of total cell counts in adult IVD and 4.7% in neonatal IVD) that previous studies were not conclusive on their presence (Fernandes *et al.*, 2020; McCann and Séguin, 2016; Risbud and Shapiro, 2011; Séguin *et al.*, 2018; Wang *et al.*, 2008). Our study shows that the NCs kept expressing NC marker *MAP1B, APOE, FABP4, and AKR1C1* in later pseudo-time (Fig. 7e). *AKR1C1* expression level was even elevated in the later pseudo-time (Fig. 7e) and later cell state (Fig. 7g). Note *AKR1C1* expression may originate from other cell populations in adult IVD as it lost marker specificity in the adult (Fig. 3a), but FABP4, APOE, and MAP1B still serve as reliable NC markers in adult IVD (Fig. 2a and 3a). Therefore, this small quantity of NCs kept their identity and keeps expressing embryonic, inflammatory and biosynthesis markers throughout adulthood.

The high heterogeneity of AFCs found in neonatal IVD is in line with recent studies that also described spatially defined AFC populations divided into inner and outer regions (Shi et al., 2015; Tam *et al.*, 2020a). Our study presented a very detailed classification of AFCs into two distinct sub-populations in the inner AFC regions (IAF1-2), and three distinct sub-populations in the outer AFC regions (OAF1-3), which has not been reported in previous relevant studies (Shi *et al.*, 2015; Tam *et al.*, 2020a).

A striking trend of decreasing expression levels of collagen-associated genes (Fig. 5d) appeared when the order of the five AFC sub-populations from neonatal IVD was re-arranged from the inner region to the outer region to IAF2, IAF1, OAF1, OAF2, OAF3 (Fig. 5d). IAF2 are likely located in the most inner region, as they are closer to NPC populations in UMAP (Fig. 6b), while OAF3 is likely in the most outer region considering their *HSPA6* expression (Fig. 6a). Interestingly, another proteomics study by Tam et al. (Tam *et al.*, 2020b) showed ECM proteins in 16- to 17-year-olds increased from the inner NP to the outer AF region of the IVD, which seems to be contradictory to our findings of the decreasing expression of collagen-producing genes (Fig. 5d). The protein content and RNA level may not necessarily be aligned. In line with Tam et al., we found two signaling pathways for cell-ECM interaction (integrin signaling (Fig. S7) and collagen-relevant

GP6 signaling) that show gradient of expression from inner AF to outer AF (Fig. 5f). The OAF and IAF have distinct gene expression profiles possibly due to their different spatially defined roles. AFCs are in a position that connects endplate and inner NP core, as suggested in Fig. 5g. The promoted *UPR* we detected in the outer AF region (Fig. 5f) may be a sign of external stressors (Hetz et al., 2020; Wang et al., 2010). However, the *UPR* also orchestrates the enforcement of adaptive mechanisms to maintain an optimal rate of protein production and rapidly reacts to diverse stimuli, including extracellular responses to hormones, growth factors and small ligands that bind cell-surface receptors (Hetz *et al.*, 2020). *UPR* is also a major contributor of the *CALR* marker for outer region OAF1 and OAF2 sub-populations (Fig. S5). The platelet-derived growth factor (*PDGF)* signaling also follows the increasing trend from inner to outer AF regions (Fig. 5f). *PDGF* expression has been shown to promote the proliferation of AFCs and NPCs (Pratsinis and Kletsas, 2007), suggesting a potential increased cell proliferation in the outer region. This theory was additionally supported by our detection of the expressed *HSPA6* in the outmost OAF3 sub-populations and *CALR* in the OAF1 and OAF2 sub-populations (Fig. 5a). The heat shock proteins (*HSP*s, Fig. 5a) were previously found to be strongly expressed in the outmost area of IVD in response to mechanical stress (Tam *et al.*, 2020a). *HSPs*+ cells were also found in endplate cartilage during gestation and then decreased with aging (Takao and Iwaki, 2002), which is consistent with our finding of larger *HSPA6*+ OAF3 sub-populations in the neonatal IVD than in the adult (Fig. 6b). *CALR* was previously found to be expressed in the OAF as well (Tam *et al.*, 2020b) and it is a key gene in *UPR* pathway (Fig. S8). This evidence points to the elevated stress-induced cell proliferation and remodeling among OAFs. In contrast, our pathway analysis indicates a trend of increased cell migration activity in inner comparing to outer regions (Fig. 5e), suggesting growing inner NP core and the gradual movement of boundary between the NP and AF towards the outer region (Fig. 5g). Collectively, the gradient trend of AFC populations in neonatal IVD, but not in the adult (Fig. S5) is critical to understanding its biology but needs to be confirmed by spatial single-cell analysis in the future.

The present work is not without limitations: the number of tissues obtained from human subjects was limited. Inclusion of additional human IVD tissue samples derived from subjects of different ages will help to unravel the role and regulation of cell sub-populations and genes throughout

development and adulthood. Also, our scRNA-seq lacks spatial resolution. The recently emerging single-cell spatial sequencing technology may be required in the next study.

In conclusion, we identified *AKR1C1*, *FABP4*, and *APOE* as novel markers for human NC populations. We demonstrated the first direct comparison of human neonatal and adult IVD at a single-cell level, finding two distinct NC sub-populations in neonatal and one in adult IVD. A pseudo-time trajectory analysis predicted the developmental process from neonatal IVD, composed of various cell types, into adult IVD, composed of mainly NPCs. We also found that a small quantity of NCs preserves their cell identity into adulthood. Lastly, we demonstrated a high heterogeneity of AFC populations in neonatal IVD and showed their gradual transitions of cell types, gene expressions, functions, and signaling pathways and potential relevance in IVD development that might be explained as spatial gradient. Better understanding of these processes helps to identify the underlying factors contributing to age-associated diseases of the IVD and to develop tailored therapeutics that may delay or reverse these processes.

STAR METHODS

Key Resource Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit Polyclonal Anti-HSPA6	Abcam	ab212044
Mouse Monoclonal Anti-CD44 (Clone Bu52)	Biorad	MCA2504T
Mouse Monoclonal Anti-SOX17	Invitrogen	MA5-24885
Mouse Monoclonal Aldo-keto Reductase 1C1/AKR1C	R&D Systems	MAB6529-SP
Goat Polyclonal Anti-FABP4/A-FABP	R&D Systems	AF3150-SP
Rabbit Monoclonal Anti-Apolipoprotein E [EP1374Y]	Abcam	ab52607
Polyclonal Rabbit Anti-MAP1B	Novus Biologicals	NBP3-04801-20ul
Bacterial and virus strains		
Biological samples		
Human neonatal IVD	NDRI	ND18842
Human adult IVD	Cedars-Sinai Medical Center	Patient Cadaver
Chemicals, peptides, and recombinant proteins		
Chromium Single-cell 3' Reagent Kits	10x Genomics	
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Invitrogen	D1306
Critical commercial assays		
Deposited data		

Experimental models: Cell lines		
Experimental models: Organisms/strains		
Oligonucleotides		
Recombinant DNA		
Software and algorithms		
ImageJ (v2.1.0/1.53c)	(Schneider et al., 2012)	
Fiji	(Schindelin et al., 2012)	
R (v4.0.3)	R Development Core Team	https://www.R-project.org
RStudio (v1.4.1103)	RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC	https://www.rstudio.com/products/rstudio/
Cell Ranger (v3.0.0)	10x Genomics	
Loupe Cell Browser (v3.0.0)	10x Genomics	

Seurat (v4.0.1)	(Butler et al., 2018; Hao et al., 2021; Satija et al., 2015; Stuart et al., 2019)	https://satijalab.org/seurat/
Monocle (v2.18.0)	(Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014)	http://cole-trapnell-lab.github.io/monocle-release/
Other		

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dmitriy Sheyn (dmitriy.sheyn@csmc.edu)

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw sequencing data and original data is accessible in repositories.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The Cedars-Sinai Medical Center Institutional Review Board approved this study (IRB number Pro00052234). Human neonatal IVD and human adult IVD have both been investigated with scRNA-seq to reveal their cell atlas of respective age and to determine the difference between them. IVD tissues harvested from three spinal disc levels of a neonatal spinal column and one spinal disc level of an adult cadaver were collected and analyzed via scRNA-seq. The three neonatal samples were collected from three different spine levels (L2-L3, L3-L4 and L4-L5) from a 6-hour postnatal male provided by the National Disease Research Interchange (NDRI, ND18842), and the adult sample was collected from L2-3 disc of a 76-year-old male cadaver.

METHOD DETAILS

Single-cell Isolation and Sequencing

The IVDs were dissected, nucleus pulposus identified and washed with Phosphate Buffered Saline (PBS). The tissue was manually minced to ~1mm pieces in a sterile environment and enzymatically digested. Tissue was digested for 1 hour at 37°C in 2 mg/mL of Pronase Protease (Millipore, Temecula, CA) in growth media containing 1mM L-glutamine, 1% antibiotic-antimycotic solution (Gibco, Carlsbad, CA) and 10% fetal bovine serum (FBS, Gemini Bioproducts, West Sacramento, CA) in Dulbecco's modified eagle media-F12 media (DMEM/F12, GIBCO, Carlsbad, CA). This was followed by ~18 hour digestion at 37°C in 0.25 mg/mL of Collagenase Type 1S (Sigma Aldrich) in growth media. The resulting sample was pushed through a 70μm cell strainer and cells were isolated via centrifugation at 1200 rpm for ten minutes, The pellet was resuspended in phosphate buffered saline at a concentration of 1,500 cells/μL and 100uL of sample was processed for single-cell RNA sequencing using the Chromium Single-cell 3' Reagent Kits from 10X Genomics.

Chromium Single-cell 3' v2 libraries with ~3,000 cells were prepared on a Chromium Controller with chips and reagents from Single-cell Gene Expression v2 kits following the manufacturer's protocols (10x Genomics). The libraries were then sequenced using paired-end sequencing (28bp Read 1 libraries, and 91bp Read 2) with a single sample index (8bp) on an Illumina NovaSeq. Samples were sequenced to a depth of > 50,000 raw reads per cell, with raw sequencing data analyzed and visualized with pre-release versions of Cell Ranger 3.0.0 and Loupe Cell Browser 3.0.0. Raw sequencing data was demultiplexed and converted to fastq format by using bcl2fastq v2.20 (Illumina, San Diego, CA). Fragment analysis of indexed libraries was performed on the Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA).

Integration, Pre-processing, and Dimensional Reduction of the Single-cell Dataset of Neonatal and Adult IVD

Seurat package (v4.0.1) in R (v4.0.3) has been used to process the data of four samples. We firstly used *CreateSeuratObject* function to transform the loaded data into a Seurat object for each sample. Next, the four Seurat objects were combined into one matrix. *NormalizeData* and *FindVariableFeatures* were used, and 6000 features were found. Afterwards, *FindIntegrationAnchors* and *IntegrateData* were used to integrate the four samples' data together. *RunPCA* function was used to for principal component analysis with PC=50. *DimPlot* function was then used with "UMAP" reduction to obtain the 2D Uniform Manifold Approximation and Projection (UMAP) plot. *FindNeighbors* and *FindClusters* functions further classified all cells in four samples into 14 clusters at resolution of 0.5. The UMAP was then split by their sample type (neonatal or adult IVD) for comparison. The consistency of the three biological repeats was confirmed by a comparison that integrated three different neonatal samples.

Detection of Enriched Genes and Novel Markers for Notochordal Cells

The expression levels of NC relevant genes in neonatal or adult IVD were separately visualized using violin plot in log scale with split view. In the volcano plots, we set the cutoff threshold of FC

to be >0.5 and the cutoff *p* value to be <0.0001. Any genes falling out of this range were not considered as enriched genes or markers. Genes falling into this range were labeled with red color in the volcano plot. Next, top enriched genes from the red colored area were manually selected as novel *in silico* markers for specific cluster from neonatal IVD or adult IVD. Their expression levels were presented quantitatively in violin plots.

Visualization of Single-Cell Atlases with 3D UMAP Plots

The cell coordination of the top three dimensions (PC1, PC2, and PC3) were extracted from the Seurat object integrating neonatal IVD and adult IVD. *FetchData* function was used to create the data frame for the 3D UMAP plot. The *plot_ly* function in *Plotly* package was used to obtain the final 3D UMAP plot. The color for each type was plot as the same as in 2D UMAP.

Gene Ontology Term Enrichment and Pathway Analysis

Gene ontology (GO) term enrichment and pathway analysis was performed using Qiagen Ingenuity Pathway Analysis (IPA, version 65367011). A list of enriched genes for a specific cell sub-population (as compared against all other cells) was firstly generated in R studio using *FindMarkers* function. The specific cell sub-populations include NC1, NC2, IAF2, IAF1, OAF1, OAF2, and OAF3 respectively. In other attempts, we did multiple further analyses: we created enriched gene list for NC by combining NC1 and NC2 cell sub-populations followed by *FindMarkers* function against all other cells; we created enriched gene list for NPC by combining NP1, NP2, NP3, NP4 cell sub-populations followed by *FindMarkers* function against all other cells. All the above enriched gene lists were generated for both neonatal IVD and adult IVD. The adjusted *p* value and fold change for genes with $p<0.0001$ were loaded into the IPA software. The pathway analysis for each list was performed based on fold change. We included the database for human tissue and primary cells but excluded those for cell lines, as our samples were harvested from human samples only. The analyses covered canonical pathways, upstream analysis, diseases and biological functions, and regulatory networks. The heatmaps for enriched

biological functions, enriched pathways, and key genes in the pathways were replotted using GraphPad Prism (v9.2.0). We demonstrated the most relevant results in figures.

The comparison of gene overlap in NCs and NPCs in both neonatal and adult IVD was performed at Metascape.org using Circos plot. The enriched genes with $p<0.0001$ and FC>1 were selected for Circos plot analysis. The enriched genes with $p<0.0001$ were used for biological function enrichment and top regulator comparison.

Histology and Immunofluorescence

IVD samples from the sample were fixed in 10% phosphate buffered formalin for three days, decalcified by incubation in 0.5M EDTA (pH 7.4, Sigma Aldrich) for three weeks, and embedded in paraffin. Five-micron-thick sections were cut from the paraffin blocks. Hematoxylin and eosin (H&E) staining was performed to evaluate the morphological features. Images were captured using a Carl Zeiss Axio Imager Z1 fluorescent microscope (Carl Zeiss, Germany) equipped with ApoTome and AxioCam HRc cameras. Images were analyzed using QuPath software (v0.2.3).

For immunofluorescence, sections were deparaffinized, rehydrated in DPBS, treated with antigen retrieval solution at room temperature for 20 minutes (Dako # S3020, Agilent Technologies). Slides were washed with PBS, blocked with 3% normal donkey serum (Jackson ImmunoResearch Laboratories Inc., West Grove, PA) in 0.3% Triton-X (Sigma Aldrich, St. Louis, MO) and again washed with PBS. Slides were stained with primary antibodies (anti-human *HSPA6, CD44, SOX17, AKR1C1, FABP4, APOE*). The primary antibodies were applied to the slides, after which the slides were incubated at 4°C overnight and washed using PBS; the slides were then incubated with secondary antibodies for 1hr at room temperature. Finally, the slides were stained with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI, 0.28 µg/ml) for 5 min in the dark. Subsequently, sections were washed three times in dPBS and mounted with Prolong Gold with DAPI (Life Technologies).

Pseudo-Time Analysis of Developmental Trajectories

The pseudo-time developmental trajectory analysis was performed using *Monocle* package (v2.18.0). The original integrated Seurat object was subset into a new Seurat object that contains the following cell populations: NC (as the combination of NC1 and NC2), NPC (as the combination of NP1, NP2, NP3, NP4), IAFC (as the combination of IAF1 and IAF2), and OAFC (as the combination of OAF1, OAF2, and OAF3). Low-quality cells were filtered by the following standards: 1) Cells whose total mRNAs $>10^6$ were filtered; 2) Cells whose total mRNAs were out of a range of mean ± 2 $\times$ standard deviations were filtered. The gene marker was also filtered when cells expressing this maker are smaller than 10. The data matrix was then log-transformed and standardized to the same scale. We chose up to 500 top globally enriched genes for each sub-populations to determine cell progress. The dimensions of the dataset were reduced using DDRTree method. The trajectories were plotted using *plot_cell_trajectory* function. The cells in trajectories were colored with either their ages, cell types, or pseudo-time states.

ACKNOWLEDGMENTS

This study was supported by NIH/NIAMS, Grant/Award Number: K01AR071512. The authors wish to acknowledge the Cedars-Sinai Genomics Core for next generation services and the Cedars-Sinai Biobank for human adult cadaver sample. The authors thank National Disease Research Interchange (NDRI) for human neonatal sample and Mrs. Julia Sheyn for the fluorescent images acquisition.

AUTHORS CONTRIBUTIONS

Conceptualization, W.J., D.S., J.G., and R.H.; Methodology, W.J., R.H., D.S., P.M., and A.W.; Software, W.J., R.H., and A.W.; Validation, K.S and G.K.; Formal Analysis, W.J., R.H., and D.S.; Investigation, W.J., D.S., and R.H.; Resources, R.H., D.S. and W.J.; Data Curation, W.J. and R.H.; Writing – Original Draft, W.J.; Writing Review & Editing, J.G., D.S., W.J., R.H., and G.K.; Visualization, W.J., D.S., and R.H.; Supervision, D.S., and R.H.; Funding Acquisition, D.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Cells from Human Neonatal and Adult IVD of Distinct Tissue Morphologies Were Sorted into 14 Unsupervised Clusters based on Their Expression of IVD Markers. H&E staining of **a)** neonatal IVD and **b)** adult IVD. The black dashed line shows the boundary separating two distinct regions in **a)**. Black arrow points to a blood vessel. The white arrows point to NPCs. **c-d)** UMAP shows 14 unsupervised cell clusters that were identified in both neonatal and adult samples. **e-f)** Dot plots showing the expression level of classical IVD markers in each cluster of neonatal and adult IVD. Dot size indicates percent of cells expressing the genes and dot color intensity the average expression level (normalized to -2 to 2). **g)** Combined UMAP for all cells from neonatal and adult IVDs with clusters assigned to sub-populations based on their marker expressions in **e-f)**. **h)** shows the cell distribution between two ages (neonatal and adult) for each sub-population. Numbers were normalized by the total cell counts for each age.

Figure 2. Identification of NC Sub-Populations and Their Pathway Network Analysis. a) The two NC sub-populations in UMAP and their expression of classical markers in both neonatal and adult IVD; two *MAP1B*+/SOX4+ NC sub-populations in the UMAP are colored green, and all other cells are colored grey; violin plots show NC expression levels of classical markers of NCs and stemness for human neonatal and adult IVD. Other clusters refer to the expression level in for all cells in all clusters other than NC1 and NC2. **b)** The immunostaining of classical NC markers in neonatal and adult IVD are shown. Scale bar = 100 μm; original magnification = 20x. **c)** Volcano plots showing the enriched genes for NC1 or NC2 sub-populations in neonatal or adult IVD. The y-axis in the volcano plots shows the $-\log_{10}p$ where the p value was about an enriched gene expression level in a specific sub-population in a specific age (neonatal or adult) against all other cells the same age. The x-axis shows the $\log_2$FC (FC - fold change) of the expression level. The cutoff threshold was set to FC>0.5 and $p<0.0001$. The enriched genes meeting these cutoff thresholds are colored with red. The genes not meeting the cutoff thresholds are colored with grey.

Figure 3. Discovery of Novel NC Markers, *AKR1C1, APOE*, *FABP4,* and Their Pathway/Network Analysis. a) Comparison of the expression level of novel top markers in human neonatal versus adult IVD. **b)** Immunostainings of novel NC markers in neonatal and adult IVD are shown. The scale bar = 100 μm and original magnification = 20x. Qiagen IPA software was used to plot **c)** *IL1* cascade network involving *AKR1C1* marker in NC1 sub-population in neonatal IVD. **d)** Top biological functions activated by *AKR1C1*/*AKR1C2* and ranked by the activation score in NC1 sub-population in neonatal IVD. **e)** An NC regulatory network (combined NC1 and NC2 sub-populations in neonatal IVD) involving *FABP4* marker.

Figure 4. Identification of 4 Similar NPC Sub-Populations and Comparison of Their NPC Gene Expression, Pathways, and Biofunctions with NC Populations in Neonatal and Adult IVD. **a-b)** Violin plots show differences in the expression levels of NPC markers in four NPC sub-populations (NP1, NP2, NP3, NP4) identified among the *ACAN+/COL2A1+/SOX9+* NPC populations in UMAP (colored in blue) in both neonatal (a) and adult IVD (b). AFC markers were shown for comparison. Expression level of each NPC subtype markers is projected via UMAP. **c)** Immunostaining of extracellular matrix (ECM) proteins in neonatal and adult IVD is shown. Scale bar = 100 μm; Original magnification = 20x. **d)** Heatmap shows biological functions enriched from all DEGs of NCs or NPCs in neonatal or adult IVD. The orange indicates the activation, the blue indicates the deactivation, and the grey indicates that data either were not detected or did not pass the filtering. **e)** Circos plot shows the enriched genes overlap between four different enriched gene lists: neonatal NCs, adult NCs, neonatal NPCs, and adult NPCs. Each purple line links the same gene expressed in both enriched gene list. Red and orange show the shared and unique genes. **f)** Descriptive summary comments on the biofunctions enriched from each list. Orange and blue fonts indicate the highlighted enriched functions of NCs and NPCs.

Figure 5 Heterogeneity of Annulus Fibrosus Cells in Human Neonatal and Adult IVDs. a-b) Expression levels of classical AFC markers (*COL1A1, CALR, HSPA6*), classical NPC markers *(COL2A1, ACAN*) and classical NC markers (*MAP1B, SOX4*) are shown for comparison, for both (**a)** neonatal IVD and **(b)** adult IVD. **c)** The AFC populations showed strong heterogeneity in

neonatal IVD by having 5 distinct subtypes and their heterogenous markers. We demonstrated the expression distribution projected on UMAP and their quantitative expression levels. **d)** Among the five heterogenous subtypes detected in **(c)**, the neonatal AFCs exhibited a decreasing trend of expression levels of ECM-relevant, collagen-producing genes from inner core NPCs to the outer region AFCs following the order of IAF2, IAF1, OAF1, OAF2, OAF3. **e-f)** The biological function **(e)** and key canonical pathway **(f)** enrichment results **followed** the same gradient. **g)** The predicted scheme for spatial distribution of the five AFC populations, along with inner NP core and notochord, in neonatal IVD as compared with the classical IVD structure in adult human is shown.

Figure. 6. The Complete Single-cell Atlases in Human Neonatal and Adult IVD. a) Complete dot plots of marker expression of all sub-populations in the neonatal and adult IVD, including classical markers and novel markers found in the present study. The percent expressed is shown by the size of the dot, and the average expression level (normalized to -2 to 2) is shown by the color. **b)** The complete single-cell atlases visualized with 3D UMAP method, for both neonatal and adult IVD. The gene markers in the green rectangle are the novel markers found in the present study. **c)** Activation Z-score rated by Qiagen IPA software showing the canonical Wnt/β-catenin pathway has been differentially regulated in NCs and NPCs in neonatal or adult IVD. **d)** Wnt/β-catenin pathway key molecules' expression levels are shown in the heatmap. The orange indicates the activation, the blue indicates the deactivation, and the grey indicates the data either was not detected or did not pass the filtering.

Figure 7. Pseudo-time Trajectories Predicted IVD Cellular and Molecular Transition During the Development. a) Cell trajectories predicted based on 500 top genes enriched to each sub-populations and color-coded based on different ages. **b-c)** The pseudo-time transition of **b)** ECM markers and **c)** Wnt/β-catenin canonical pathway regulators. **d)** Cell trajectories color-coded by different cell types, including NC, NPC, IAFC, and OAFC, and **e)** the change of classical and novel NC markers in NCs over pseudo-time and the change of classical NPC markers in NPCs over pseudo-time. **f)** Cell trajectories color-coded by different cell states. Solid grey line shows the

predicted developmental direction, and the dashed grey line shows the possible but inconclusively predicted developmental direction. **g)** The pseudo-time transition of molecules of interests are displayed and colored with cell states.

REFERENCES

Andersson, G.B. (1999). Epidemiological features of chronic low-back pain. Lancet *354*, 581-585. 10.1016/S0140-6736(99)01312-4.

Aszódi, A., Chan, D., Hunziker, E., Bateman, J.F., and Fässler, R. (1998). Collagen II Is Essential for the Removal of the Notochord and the Formation of Intervertebral Discs. Journal of Cell Biology *143*, 1399-1412. 10.1083/jcb.143.5.1399.

Bhattaram, P., Penzo-Méndez, A., Sock, E., Colmenares, C., Kaneko, K.J., Vassilev, A., DePamphilis, M.L., Wegner, M., and Lefebvre, V. (2010). Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. Nature Communications *1*, 9. 10.1038/ncomms1008.

Bi, W., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrugghe, B. (1999). Sox9 is required for cartilage formation. Nature Genetics *22*, 85-89. 10.1038/8792.

Buckley, C.T., Hoyland, J.A., Fujii, K., Pandit, A., Iatridis, J.C., and Grad, S. (2018). Critical aspects and challenges for intervertebral disc repair and regeneration—Harnessing advances in tissue engineering. JOR SPINE *1*, e1029. 10.1002/jsp2.1029.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology *36*, 411-420. 10.1038/nbt.4096.

Cai, H., Zou, J., Wang, W., and Yang, A. (2021). BMP2 induces hMSC osteogenesis and matrix remodeling. Mol Med Rep *23*, 125. 10.3892/mmr.2020.11764.

Chen, W.K., Yu, X.H., Yang, W., Wang, C., He, W.S., Yan, Y.G., Zhang, J., and Wang, W.J. (2017). lncRNAs: novel players in intervertebral disc degeneration and osteoarthritis. Cell Prolif *50*. 10.1111/cpr.12313.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nature Biotechnology *23*, 1534-1541. 10.1038/nbt1163.

de Schepper, E.I., Damen, J., van Meurs, J.B., Ginai, A.Z., Popham, M., Hofman, A., Koes, B.W., and Bierma-Zeinstra, S.M. (2010). The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. Spine (Phila Pa 1976) *35*, 531-536. 10.1097/BRS.0b013e3181aa5b33.

Fernandes, L.M., Khan, N.M., Trochez, C.M., Duan, M., Diaz-Hernandez, M.E., Presciutti, S.M., Gibson, G., and Drissi, H. (2020). Single-cell RNA-seq identifies unique transcriptional landscapes of human nucleus pulposus and annulus fibrosus cells. Sci Rep *10*, 15263. 10.1038/s41598-020-72261-7.

Frank, J.W., Kerr, M.S., Brooker, A.S., DeMaio, S.E., Maetzel, A., Shannon, H.S., Sullivan, T.J., Norman, R.W., and Wells, R.P. (1996). Disability resulting from occupational low back pain. Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. Spine (Phila Pa 1976) *21*, 2908-2917. 10.1097/00007632-199612150-00024.

Frapin, L., Clouet, J., Delplace, V., Fusellier, M., Guicheux, J., and Le Visage, C. (2019). Lessons learned from intervertebral disc pathophysiology to guide rational design of sequential delivery systems for therapeutic biological factors. Advanced Drug Delivery Reviews *149-150*, 49-71. https://doi.org/10.1016/j.addr.2019.08.007.

Gan, Y., He, J., Zhu, J., Xu, Z., Wang, Z., Yan, J., Hu, O., Bai, Z., Chen, L., Xie, Y., et al. (2021). Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitors in human intervertebral discs. Bone Research *9*, 37. 10.1038/s41413-021-00163-z.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e3529. 10.1016/j.cell.2021.04.048.

Hetz, C., Zhang, K., and Kaufman, R.J. (2020). Mechanisms, regulation and functions of the unfolded protein response. Nature Reviews Molecular Cell Biology *21*, 421-438. 10.1038/s41580-020-0250-z.

Ho, R., Workman, M.J., Mathkar, P., Wu, K., Kim, K.J., O'Rourke, J.G., Kellogg, M., Montel, V., Banuelos, M.G., Arogundade, O.A., et al. (2021). Cross-Comparison of Human iPSC Motor Neuron Models of Familial and Sporadic ALS Reveals Early and Convergent Transcriptomic Disease Signatures. Cell Systems *12*, 159-175.e159. 10.1016/j.cels.2020.10.010.

Huang, Y., and Mahley, R.W. (2014). Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. Neurobiol Dis *72 Pt A*, 3-12. 10.1016/j.nbd.2014.08.025.

Hunter, C., Matyas, J., and Duncan, N. (2003). The notochordal cell in the nucleus pulposus: a review in the context of tissue engineering. Tissue engineering *9*, 667-677.

Johnson, Z.I., Schoepflin, Z.R., Choi, H., Shapiro, I.M., and Risbud, M.V. (2015). Disc in flames: Roles of TNF-α and IL-1β in intervertebral disc degeneration. European cells & materials *30*, 104-117. 10.22203/ecm.v030a08.

Kelly, N.H., Huynh, N.P.T., and Guilak, F. (2020). Single cell RNA-sequencing reveals cellular heterogeneity and trajectories of lineage specification during murine embryonic limb development. Matrix Biology *89*, 1-10. https://doi.org/10.1016/j.matbio.2019.12.004.

Knezevic, N.N., Mandalia, S., Raasch, J., Knezevic, I., and Candido, K.D. (2017). Treatment of chronic low back pain–new approaches on the horizon. Journal of pain research *10*, 1111.

Kondo, N., Yuasa, T., Shimono, K., Tung, W., Okabe, T., Yasuhara, R., Pacifici, M., Zhang, Y., Iwamoto, M., and Enomoto-Iwamoto, M. (2011). Intervertebral disc development is regulated by Wnt/β-catenin signaling. Spine *36*, E513-E518. 10.1097/BRS.0b013e3181f52cb5.

Lu, H., Cassis, L.A., Kooi, C.W.V., and Daugherty, A. (2016). Structure and functions of angiotensinogen. Hypertension Research *39*, 492-500. 10.1038/hr.2016.17.

Lu, S.-Y., Li, M., and Lin, Y.-L. (2010). Mitf induction by RANKL is critical for osteoclastogenesis. Mol Biol Cell *21*, 1763-1771. 10.1091/mbc.e09-07-0584.

Macfarlane, G.J., Thomas, E., Croft, P.R., Papageorgiou, A.C., Jayson, M.I.V., and Silman, A.J. (1999). Predictors of early improvement in low back pain amongst consulters to general practice:

the influence of pre-morbid and episode-related factors. Pain *80*, 113-119. https://doi.org/10.1016/S0304-3959(98)00209-7.

Marín, Y.E., Seiberg, M., and Lin, C.B. (2009). Aldo-keto reductase 1C subfamily genes in skin are UV-inducible: possible role in keratinocytes survival. Experimental Dermatology *18*, 611-618. https://doi.org/10.1111/j.1600-0625.2008.00839.x.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332-337. 10.1038/s41586-019-1195-2.

McCann, M.R., and Séguin, C.A. (2016). Notochord Cells in Intervertebral Disc Development and Degeneration. Journal of Developmental Biology *4*, 3.

Pandey, K.B., and Rizvi, S.I. (2011). Biomarkers of oxidative stress in red blood cells. Biomedical Papers of the Medical Faculty of Palacky University in Olomouc *155*.

Panebianco, C.J., Dave, A., Charytonowicz, D., Sebra, R., and Iatridis, J.C. (2021). Single-cell RNA-sequencing atlas of bovine caudal intervertebral discs: Discovery of heterogeneous cell populations with distinct roles in homeostasis. The FASEB Journal *35*, e21919. https://doi.org/10.1096/fj.202101149R.

Pratsinis, H., and Kletsas, D. (2007). PDGF, bFGF and IGF-I stimulate the proliferation of intervertebral disc cells in vitro via the activation of the ERK and Akt signaling pathways. Eur Spine J *16*, 1858-1866. 10.1007/s00586-007-0408-9.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017a). Single-cell mRNA quantification and differential analysis with Census. Nat Methods *14*, 309-315. 10.1038/nmeth.4150.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H., and Trapnell, C. (2017b). Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv, 110668. 10.1101/110668.

Richardson, S.M., Ludwinski, F.E., Gnanalingham, K.K., Atkinson, R.A., Freemont, A.J., and Hoyland, J.A. (2017). Notochordal and nucleus pulposus marker expression is maintained by sub-populations of adult human nucleus pulposus cells through aging and degeneration. Sci Rep *7*, 1501. 10.1038/s41598-017-01567-w.

Risbud, M.V., Schaer, T.P., and Shapiro, I.M. (2010). Toward an understanding of the role of notochordal cells in the adult intervertebral disc: From discord to accord. Developmental Dynamics *239*, 2141-2148. https://doi.org/10.1002/dvdy.22350.

Risbud, M.V., Schoepflin, Z.R., Mwale, F., Kandel, R.A., Grad, S., Iatridis, J.C., Sakai, D., and Hoyland, J.A. (2015). Defining the phenotype of young healthy nucleus pulposus cells: Recommendations of the Spine Research Interest Group at the 2014 annual ORS meeting. Journal of Orthopaedic Research *33*, 283-293. https://doi.org/10.1002/jor.22789.

Risbud, M.V., and Shapiro, I.M. (2011). Notochordal cells in the adult intervertebral disc: new perspective on an old question. Critical reviews in eukaryotic gene expression *21*, 29-41. 10.1615/critreveukargeneexpr.v21.i1.30.

Rodrigues-Pinto, R., Ward, L., Humphreys, M., Zeef, L.A.H., Berry, A., Hanley, K.P., Hanley, N., Richardson, S.M., and Hoyland, J.A. (2018). Human notochordal cell transcriptome unveils potential regulators of cell function in the developing intervertebral disc. Sci Rep *8*, 12866. 10.1038/s41598-018-31172-4.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495-502. 10.1038/nbt.3192.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676-682. 10.1038/nmeth.2019.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods *9*, 671-675. 10.1038/nmeth.2089.

Schumann, J., Stanko, K., Schliesser, U., Appelt, C., and Sawitzki, B. (2015). Differences in CD44 Surface Expression Levels and Function Discriminates IL-17 and IFN-γ Producing Helper T Cells. PLoS One *10*, e0132479-e0132479. 10.1371/journal.pone.0132479.

Schwarzer, A.C., Aprill, C.N., Derby, R., Fortin, J., Kine, G., and Bogduk, N. (1994). The relative contributions of the disc and zygapophyseal joint in chronic low back pain. Spine (Phila Pa 1976) *19*, 801-806. 10.1097/00007632-199404000-00013.

Séguin, C.A., Chan, D., Dahia, C.L., and Gazit, Z. (2018). Latest advances in intervertebral disc development and progenitor cells. JOR SPINE *1*, e1030. 10.1002/jsp2.1030.

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nature Biotechnology *37*, 451-460. 10.1038/s41587-019-0068-4.

Sheyn, D., Ben-David, S., Tawackoli, W., Zhou, Z., Salehi, K., Bez, M., De Mel, S., Chan, V., Roth, J., Avalos, P., et al. (2019). Human iPSCs can be differentiated into notochordal cells that reduce intervertebral disc degeneration in a porcine model. Theranostics *9*, 7506-7524. 10.7150/thno.34898.

Shi, R., Wang, F., Hong, X., Wang, Y.T., Bao, J.P., Cai, F., and Wu, X.T. (2015). The presence of stem cells in potential stem cell niches of the intervertebral disc region: an in vitro study on rats. Eur Spine J *24*, 2411-2424. 10.1007/s00586-015-4168-7.

Sinha, A., Fan, V.B., Ramakrishnan, A.-B., Engelhardt, N., Kennell, J., and Cadigan, K.M. (2021). Repression of Wnt/β-catenin signaling by SOX9 and Mastermind-like transcriptional coactivator 2. Science Advances *7*, eabe0849. doi:10.1126/sciadv.abe0849.

Sinner, D., Kordich, J.J., Spence, J.R., Opoka, R., Rankin, S., Lin, S.-C.J., Jonatan, D., Zorn, A.M., and Wells, J.M. (2007). Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. Mol Cell Biol *27*, 7802-7815. 10.1128/MCB.02179-06.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. Nature Reviews Genetics *20*, 631-656. 10.1038/s41576-019-0150-2.

Stegner, D., Haining, E.J., and Nieswandt, B. (2014). Targeting Glycoprotein VI and the Immunoreceptor Tyrosine-Based Activation Motif Signaling Pathway. Arteriosclerosis, Thrombosis, and Vascular Biology *34*, 1615-1620. doi:10.1161/ATVBAHA.114.303408.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e1821. 10.1016/j.cell.2019.05.031.

Sun, Z., Liu, B., and Luo, Z.-J. (2020). The Immune Privilege of the Intervertebral Disc: Implications for Intervertebral Disc Degeneration Treatment. Int J Med Sci *17*, 685-692. 10.7150/ijms.42238.

Takao, T., and Iwaki, T. (2002). A Comparative Study of Localization of Heat Shock Protein 27 and Heat Shock Protein 72 in the Developmental and Degenerative Intervertebral Discs. Spine *27*.

Talamo, F., D'Ambrosio, C., Arena, S., Del Vecchio, P., Ledda, L., Zehender, G., Ferrara, L., and Scaloni, A. (2003). Proteins from bovine tissues and biological fluids: Defining a reference electrophoresis map for liver, kidney, muscle, plasma and red blood cells. PROTEOMICS *3*, 440-460. https://doi.org/10.1002/pmic.200390059.

Tam, V., Chen, P., Yee, A., Solis, N., Klein, T., Kudelko, M., Sharma, R., Chan, W.C., Overall, C.M., Haglund, L., et al. (2020a). DIPPER: a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. bioRxiv, 2020.2007.2011.192948. 10.1101/2020.07.11.192948.

Tam, V., Chen, P., Yee, A., Solis, N., Klein, T., Kudelko, M., Sharma, R., Chan, W.C.W., Overall, C.M., Haglund, L., et al. (2020b). DIPPER, a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. eLife *9*, e64940. 10.7554/eLife.64940.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386. 10.1038/nbt.2859.

Urs, S., Harrington, A., Liaw, L., and Small, D. (2006). Selective expression of an aP2/Fatty Acid Binding Protein 4-Cre transgene in non-adipogenic tissues during embryonic development. Transgenic Res *15*, 647-653. 10.1007/s11248-006-9000-z.

van den Akker, G.G., Koenders, M.I., van de Loo, F.A., van Lent, P.L., Davidson, E.B., and van der Kraan, P.M. (2017). Transcriptional profiling distinguishes inner and outer annulus fibrosus from nucleus pulposus in the bovine intervertebral disc. European Spine Journal *26*, 2053-2062.

Vergroesen, P.P.A., Kingma, I., Emanuel, K.S., Hoogendoorn, R.J.W., Welting, T.J., van Royen, B.J., van Dieën, J.H., and Smit, T.H. (2015). Mechanics and biology in intervertebral disc degeneration: a vicious circle. Osteoarthritis and Cartilage *23*, 1057-1070. https://doi.org/10.1016/j.joca.2015.03.028.

Wang, J., Huang, Y., Huang, L., Shi, K., Wang, J., Zhu, C., Li, L., Zhang, L., Feng, G., Liu, L., and Song, Y. (2021). Novel biomarkers of intervertebral disc cells and evidence of stem cells in the intervertebral disc. Osteoarthritis and Cartilage *29*, 389-401. https://doi.org/10.1016/j.joca.2020.12.005.

Wang, W.L., Abramson, J.H., Ganguly, A., and Rosenberg, A.E. (2008). The surgical pathology of notochordal remnants in adult intervertebral disks: a report of 3 cases. Am J Surg Pathol *32*, 1123-1129. 10.1097/PAS.0b013e3181757954.

Wang, Z., Butler, P., Ly, D., Spiotto, M., Koong, A., and Yang, G. (2010). Activation of the Unfolded Protein Response in Wound Healing. Journal of Surgical Research *158*, 209. 10.1016/j.jss.2009.11.111.

Yamazaki, Y., Zhao, N., Caulfield, T.R., Liu, C.-C., and Bu, G. (2019). Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. Nature Reviews Neurology *15*, 501-518. 10.1038/s41582-019-0228-7.

Zeng, C.-M., Chang, L.-L., Ying, M.-D., Cao, J., He, Q.-J., Zhu, H., and Yang, B. (2017). Aldo-Keto Reductase AKR1C1-AKR1C4: Functions, Regulation, and Intervention for Anti-cancer Therapy. Front Pharmacol *8*, 119-119. 10.3389/fphar.2017.00119.

Zhang, D., Jin, L., Reames, D.L., Shen, F.H., Shimer, A.L., and Li, X. (2013). Intervertebral disc degeneration and ectopic bone formation in apolipoprotein E knockout mice. J Orthop Res *31*, 210-217. 10.1002/jor.22216.

Zhang, Y., Zhang, Z., Chen, P., Ma, C.Y., Li, C., Au, T.Y.K., Tam, V., Peng, Y., Wu, R., Cheung, K.M.C., et al. (2020). Directed Differentiation of Notochord-like and Nucleus Pulposus-like Cells Using Human Pluripotent Stem Cells. Cell Reports *30*, 2791-2806.e2795. https://doi.org/10.1016/j.celrep.2020.01.100.

Zhou, G., Zheng, Q., Engin, F., Munivez, E., Chen, Y., Sebald, E., Krakow, D., and Lee, B. (2006). Dominance of SOX9 function over RUNX2 during skeletogenesis. Proceedings of the National Academy of Sciences *103*, 19004-19009. 10.1073/pnas.0605170103.

Ziegler-Heitbrock, H.W., and Ulevitch, R.J. (1993). CD14: cell surface receptor and differentiation marker. Immunol Today *14*, 121-125. 10.1016/0167-5699(93)90212-4.

Andersson, G.B. (1999). Epidemiological features of chronic low-back pain. Lancet *354*, 581-585. 10.1016/S0140-6736(99)01312-4.

Aszódi, A., Chan, D., Hunziker, E., Bateman, J.F., and Fässler, R. (1998). Collagen II Is Essential for the Removal of the Notochord and the Formation of Intervertebral Discs. Journal of Cell Biology *143*, 1399-1412. 10.1083/jcb.143.5.1399.

Bhattaram, P., Penzo-Méndez, A., Sock, E., Colmenares, C., Kaneko, K.J., Vassilev, A., DePamphilis, M.L., Wegner, M., and Lefebvre, V. (2010). Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. Nature Communications *1*, 9. 10.1038/ncomms1008.

Bi, W., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrugghe, B. (1999). Sox9 is required for cartilage formation. Nature Genetics *22*, 85-89. 10.1038/8792.

Buckley, C.T., Hoyland, J.A., Fujii, K., Pandit, A., Iatridis, J.C., and Grad, S. (2018). Critical aspects and challenges for intervertebral disc repair and regeneration—Harnessing advances in tissue engineering. JOR SPINE *1*, e1029. 10.1002/jsp2.1029.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology *36*, 411-420. 10.1038/nbt.4096.

Cai, H., Zou, J., Wang, W., and Yang, A. (2021). BMP2 induces hMSC osteogenesis and matrix remodeling. Mol Med Rep *23*, 125. 10.3892/mmr.2020.11764.

Chen, W.K., Yu, X.H., Yang, W., Wang, C., He, W.S., Yan, Y.G., Zhang, J., and Wang, W.J. (2017). lncRNAs: novel players in intervertebral disc degeneration and osteoarthritis. Cell Prolif *50*. 10.1111/cpr.12313.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nature Biotechnology *23*, 1534-1541. 10.1038/nbt1163.

de Schepper, E.I., Damen, J., van Meurs, J.B., Ginai, A.Z., Popham, M., Hofman, A., Koes, B.W., and Bierma-Zeinstra, S.M. (2010). The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. Spine (Phila Pa 1976) *35*, 531-536. 10.1097/BRS.0b013e3181aa5b33.

Fernandes, L.M., Khan, N.M., Trochez, C.M., Duan, M., Diaz-Hernandez, M.E., Presciutti, S.M., Gibson, G., and Drissi, H. (2020). Single-cell RNA-seq identifies unique transcriptional landscapes of human nucleus pulposus and annulus fibrosus cells. Sci Rep *10*, 15263. 10.1038/s41598-020-72261-7.

Frank, J.W., Kerr, M.S., Brooker, A.S., DeMaio, S.E., Maetzel, A., Shannon, H.S., Sullivan, T.J., Norman, R.W., and Wells, R.P. (1996). Disability resulting from occupational low back pain. Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. Spine (Phila Pa 1976) *21*, 2908-2917. 10.1097/00007632-199612150-00024.

Frapin, L., Clouet, J., Delplace, V., Fusellier, M., Guicheux, J., and Le Visage, C. (2019). Lessons learned from intervertebral disc pathophysiology to guide rational design of sequential delivery systems for therapeutic biological factors. Advanced Drug Delivery Reviews *149-150*, 49-71. https://doi.org/10.1016/j.addr.2019.08.007.

Gan, Y., He, J., Zhu, J., Xu, Z., Wang, Z., Yan, J., Hu, O., Bai, Z., Chen, L., Xie, Y., et al. (2021). Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitors in human intervertebral discs. Bone Research *9*, 37. 10.1038/s41413-021-00163-z.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e3529. 10.1016/j.cell.2021.04.048.

Hetz, C., Zhang, K., and Kaufman, R.J. (2020). Mechanisms, regulation and functions of the unfolded protein response. Nature Reviews Molecular Cell Biology *21*, 421-438. 10.1038/s41580-020-0250-z.

Ho, R., Workman, M.J., Mathkar, P., Wu, K., Kim, K.J., O'Rourke, J.G., Kellogg, M., Montel, V., Banuelos, M.G., Arogundade, O.A., et al. (2021). Cross-Comparison of Human iPSC Motor Neuron Models of Familial and Sporadic ALS Reveals Early and Convergent Transcriptomic Disease Signatures. Cell Systems *12*, 159-175.e159. 10.1016/j.cels.2020.10.010.

Huang, Y., and Mahley, R.W. (2014). Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. Neurobiol Dis *72 Pt A*, 3-12. 10.1016/j.nbd.2014.08.025.

Hunter, C., Matyas, J., and Duncan, N. (2003). The notochordal cell in the nucleus pulposus: a review in the context of tissue engineering. Tissue engineering *9*, 667-677.

Johnson, Z.I., Schoepflin, Z.R., Choi, H., Shapiro, I.M., and Risbud, M.V. (2015). Disc in flames: Roles of TNF-α and IL-1β in intervertebral disc degeneration. European cells & materials *30*, 104-117. 10.22203/ecm.v030a08.

Kelly, N.H., Huynh, N.P.T., and Guilak, F. (2020). Single cell RNA-sequencing reveals cellular heterogeneity and trajectories of lineage specification during murine embryonic limb development. Matrix Biology *89*, 1-10. https://doi.org/10.1016/j.matbio.2019.12.004.

Knezevic, N.N., Mandalia, S., Raasch, J., Knezevic, I., and Candido, K.D. (2017). Treatment of chronic low back pain–new approaches on the horizon. Journal of pain research *10*, 1111.

Kondo, N., Yuasa, T., Shimono, K., Tung, W., Okabe, T., Yasuhara, R., Pacifici, M., Zhang, Y., Iwamoto, M., and Enomoto-Iwamoto, M. (2011). Intervertebral disc development is regulated by Wnt/β-catenin signaling. Spine *36*, E513-E518. 10.1097/BRS.0b013e3181f52cb5.

Lu, H., Cassis, L.A., Kooi, C.W.V., and Daugherty, A. (2016). Structure and functions of angiotensinogen. Hypertension Research *39*, 492-500. 10.1038/hr.2016.17.

Lu, S.-Y., Li, M., and Lin, Y.-L. (2010). Mitf induction by RANKL is critical for osteoclastogenesis. Mol Biol Cell *21*, 1763-1771. 10.1091/mbc.e09-07-0584.

Macfarlane, G.J., Thomas, E., Croft, P.R., Papageorgiou, A.C., Jayson, M.I.V., and Silman, A.J. (1999). Predictors of early improvement in low back pain amongst consulters to general practice: the influence of pre-morbid and episode-related factors. Pain *80*, 113-119. https://doi.org/10.1016/S0304-3959(98)00209-7.

Marín, Y.E., Seiberg, M., and Lin, C.B. (2009). Aldo-keto reductase 1C subfamily genes in skin are UV-inducible: possible role in keratinocytes survival. Experimental Dermatology *18*, 611-618. https://doi.org/10.1111/j.1600-0625.2008.00839.x.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332-337. 10.1038/s41586-019-1195-2.

McCann, M.R., and Séguin, C.A. (2016). Notochord Cells in Intervertebral Disc Development and Degeneration. Journal of Developmental Biology *4*, 3.

Pandey, K.B., and Rizvi, S.I. (2011). Biomarkers of oxidative stress in red blood cells. Biomedical Papers of the Medical Faculty of Palacky University in Olomouc *155*.

Panebianco, C.J., Dave, A., Charytonowicz, D., Sebra, R., and Iatridis, J.C. (2021). Single-cell RNA-sequencing atlas of bovine caudal intervertebral discs: Discovery of heterogeneous cell populations with distinct roles in homeostasis. The FASEB Journal *35*, e21919. https://doi.org/10.1096/fj.202101149R.

Pratsinis, H., and Kletsas, D. (2007). PDGF, bFGF and IGF-I stimulate the proliferation of intervertebral disc cells in vitro via the activation of the ERK and Akt signaling pathways. Eur Spine J *16*, 1858-1866. 10.1007/s00586-007-0408-9.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017a). Single-cell mRNA quantification and differential analysis with Census. Nat Methods *14*, 309-315. 10.1038/nmeth.4150.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H., and Trapnell, C. (2017b). Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv, 110668. 10.1101/110668.

Richardson, S.M., Ludwinski, F.E., Gnanalingham, K.K., Atkinson, R.A., Freemont, A.J., and Hoyland, J.A. (2017). Notochordal and nucleus pulposus marker expression is maintained by sub-populations of adult human nucleus pulposus cells through aging and degeneration. Sci Rep *7*, 1501. 10.1038/s41598-017-01567-w.

Risbud, M.V., Schaer, T.P., and Shapiro, I.M. (2010). Toward an understanding of the role of notochordal cells in the adult intervertebral disc: From discord to accord. Developmental Dynamics *239*, 2141-2148. https://doi.org/10.1002/dvdy.22350.

Risbud, M.V., Schoepflin, Z.R., Mwale, F., Kandel, R.A., Grad, S., Iatridis, J.C., Sakai, D., and Hoyland, J.A. (2015). Defining the phenotype of young healthy nucleus pulposus cells: Recommendations of the Spine Research Interest Group at the 2014 annual ORS meeting. Journal of Orthopaedic Research *33*, 283-293. https://doi.org/10.1002/jor.22789.

Risbud, M.V., and Shapiro, I.M. (2011). Notochordal cells in the adult intervertebral disc: new perspective on an old question. Critical reviews in eukaryotic gene expression *21*, 29-41. 10.1615/critreveukargeneexpr.v21.i1.30.

Rodrigues-Pinto, R., Ward, L., Humphreys, M., Zeef, L.A.H., Berry, A., Hanley, K.P., Hanley, N., Richardson, S.M., and Hoyland, J.A. (2018). Human notochordal cell transcriptome unveils potential regulators of cell function in the developing intervertebral disc. Sci Rep *8*, 12866. 10.1038/s41598-018-31172-4.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495-502. 10.1038/nbt.3192.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676-682. 10.1038/nmeth.2019.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods *9*, 671-675. 10.1038/nmeth.2089.

Schumann, J., Stanko, K., Schliesser, U., Appelt, C., and Sawitzki, B. (2015). Differences in CD44 Surface Expression Levels and Function Discriminates IL-17 and IFN-γ Producing Helper T Cells. PLoS One *10*, e0132479-e0132479. 10.1371/journal.pone.0132479.

Séguin, C.A., Chan, D., Dahia, C.L., and Gazit, Z. (2018). Latest advances in intervertebral disc development and progenitor cells. JOR SPINE *1*, e1030. 10.1002/jsp2.1030.

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nature Biotechnology *37*, 451-460. 10.1038/s41587-019-0068-4.

Sheyn, D., Ben-David, S., Tawackoli, W., Zhou, Z., Salehi, K., Bez, M., De Mel, S., Chan, V., Roth, J., Avalos, P., et al. (2019). Human iPSCs can be differentiated into notochordal cells that reduce intervertebral disc degeneration in a porcine model. Theranostics *9*, 7506-7524. 10.7150/thno.34898.

Shi, R., Wang, F., Hong, X., Wang, Y.T., Bao, J.P., Cai, F., and Wu, X.T. (2015). The presence of stem cells in potential stem cell niches of the intervertebral disc region: an in vitro study on rats. Eur Spine J *24*, 2411-2424. 10.1007/s00586-015-4168-7.

Sinha, A., Fan, V.B., Ramakrishnan, A.-B., Engelhardt, N., Kennell, J., and Cadigan, K.M. (2021). Repression of Wnt/β-catenin signaling by SOX9 and Mastermind-like transcriptional coactivator 2. Science Advances *7*, eabe0849. doi:10.1126/sciadv.abe0849.

Sinner, D., Kordich, J.J., Spence, J.R., Opoka, R., Rankin, S., Lin, S.-C.J., Jonatan, D., Zorn, A.M., and Wells, J.M. (2007). Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. Mol Cell Biol *27*, 7802-7815. 10.1128/MCB.02179-06.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. Nature Reviews Genetics *20*, 631-656. 10.1038/s41576-019-0150-2.

Stegner, D., Haining, E.J., and Nieswandt, B. (2014). Targeting Glycoprotein VI and the Immunoreceptor Tyrosine-Based Activation Motif Signaling Pathway. Arteriosclerosis, Thrombosis, and Vascular Biology *34*, 1615-1620. doi:10.1161/ATVBAHA.114.303408.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e1821. 10.1016/j.cell.2019.05.031.

Sun, Z., Liu, B., and Luo, Z.-J. (2020). The Immune Privilege of the Intervertebral Disc: Implications for Intervertebral Disc Degeneration Treatment. Int J Med Sci *17*, 685-692. 10.7150/ijms.42238.

Takao, T., and Iwaki, T. (2002). A Comparative Study of Localization of Heat Shock Protein 27 and Heat Shock Protein 72 in the Developmental and Degenerative Intervertebral Discs. Spine *27*.

Talamo, F., D'Ambrosio, C., Arena, S., Del Vecchio, P., Ledda, L., Zehender, G., Ferrara, L., and Scaloni, A. (2003). Proteins from bovine tissues and biological fluids: Defining a reference electrophoresis map for liver, kidney, muscle, plasma and red blood cells. PROTEOMICS *3*, 440-460. https://doi.org/10.1002/pmic.200390059.

Tam, V., Chen, P., Yee, A., Solis, N., Klein, T., Kudelko, M., Sharma, R., Chan, W.C., Overall, C.M., Haglund, L., et al. (2020a). DIPPER: a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. bioRxiv, 2020.2007.2011.192948. 10.1101/2020.07.11.192948.

Tam, V., Chen, P., Yee, A., Solis, N., Klein, T., Kudelko, M., Sharma, R., Chan, W.C.W., Overall, C.M., Haglund, L., et al. (2020b). DIPPER, a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. eLife *9*, e64940. 10.7554/eLife.64940.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386. 10.1038/nbt.2859.

Urs, S., Harrington, A., Liaw, L., and Small, D. (2006). Selective expression of an aP2/Fatty Acid Binding Protein 4-Cre transgene in non-adipogenic tissues during embryonic development. Transgenic Res *15*, 647-653. 10.1007/s11248-006-9000-z.

van den Akker, G.G., Koenders, M.I., van de Loo, F.A., van Lent, P.L., Davidson, E.B., and van der Kraan, P.M. (2017). Transcriptional profiling distinguishes inner and outer annulus fibrosus from nucleus pulposus in the bovine intervertebral disc. European Spine Journal *26*, 2053-2062.

Vergroesen, P.P.A., Kingma, I., Emanuel, K.S., Hoogendoorn, R.J.W., Welting, T.J., van Royen, B.J., van Dieën, J.H., and Smit, T.H. (2015). Mechanics and biology in intervertebral disc degeneration: a vicious circle. Osteoarthritis and Cartilage *23*, 1057-1070. https://doi.org/10.1016/j.joca.2015.03.028.

Wang, J., Huang, Y., Huang, L., Shi, K., Wang, J., Zhu, C., Li, L., Zhang, L., Feng, G., Liu, L., and Song, Y. (2021). Novel biomarkers of intervertebral disc cells and evidence of stem cells in the intervertebral disc. Osteoarthritis and Cartilage *29*, 389-401. https://doi.org/10.1016/j.joca.2020.12.005.

Wang, W.L., Abramson, J.H., Ganguly, A., and Rosenberg, A.E. (2008). The surgical pathology of notochordal remnants in adult intervertebral disks: a report of 3 cases. Am J Surg Pathol *32*, 1123-1129. 10.1097/PAS.0b013e3181757954.

Wang, Z., Butler, P., Ly, D., Spiotto, M., Koong, A., and Yang, G. (2010). Activation of the Unfolded Protein Response in Wound Healing. Journal of Surgical Research *158*, 209. 10.1016/j.jss.2009.11.111.

Zeng, C.-M., Chang, L.-L., Ying, M.-D., Cao, J., He, Q.-J., Zhu, H., and Yang, B. (2017). Aldo-Keto Reductase AKR1C1-AKR1C4: Functions, Regulation, and Intervention for Anti-cancer Therapy. Front Pharmacol *8*, 119-119. 10.3389/fphar.2017.00119.

Zhang, D., Jin, L., Reames, D.L., Shen, F.H., Shimer, A.L., and Li, X. (2013). Intervertebral disc degeneration and ectopic bone formation in apolipoprotein E knockout mice. J Orthop Res *31*, 210-217. 10.1002/jor.22216.

Zhang, Y., Zhang, Z., Chen, P., Ma, C.Y., Li, C., Au, T.Y.K., Tam, V., Peng, Y., Wu, R., Cheung, K.M.C., et al. (2020). Directed Differentiation of Notochord-like and Nucleus Pulposus-like Cells Using Human Pluripotent Stem Cells. Cell Reports *30*, 2791-2806.e2795. https://doi.org/10.1016/j.celrep.2020.01.100.

Zhou, G., Zheng, Q., Engin, F., Munivez, E., Chen, Y., Sebald, E., Krakow, D., and Lee, B. (2006). Dominance of SOX9 function over RUNX2 during skeletogenesis. Proceedings of the National Academy of Sciences *103*, 19004-19009. 10.1073/pnas.0605170103.

Ziegler-Heitbrock, H.W., and Ulevitch, R.J. (1993). CD14: cell surface receptor and differentiation marker. Immunol Today *14*, 121-125. 10.1016/0167-5699(93)90212-4.

The origin, composition, distribution, and function of cells in the human intervertebral disc (IVD) has not been fully understood. Here, cell atlases of both human neonatal and adult IVDs have been generated and further assessed by gene ontology pathway enrichment, pseudo-time trajectory, histology, and immunofluorescence. Comparison of cell atlases revealed the presence of several sub-populations of notochordal cells (NC) in the neonatal IVD and a small quantity of NCs and associated markers in the adult IVD. Developmental trajectories predicted that most neonatal NCs develop into adult nucleus pulposus cells (NPCs) while some keep their identity throughout adulthood. A high heterogeneity and gradual transition of annulus fibrosus cells (AFCs) in the neonatal IVD was detected and their potential relevance in IVD development was assessed. Collectively, comparing single-cell atlases between neonatal and adult IVDs delineates the landscape of IVD cell biology and may help discover novel therapeutic targets for IVD degeneration.