Title: Monocyte-derived fibrocytes contribute to nucleus pulposus fibrosis and intervertebral disc degeneration

Yi Sun[1 #], Zezhuo Su[2 #], Yan Peng[2 #], Kyle, KH SO[2], Qiu-ji Lu[2], Mao-jiang Lyu[1], Jianwei Zuo[1], Yong-can Huang[3], Zhi-ping Guan[3], Kenneth MC Cheung[2], Zhao-min Zheng[4], Xin-tao Zhang[1], Victor YL Leung[2]*

[1]Department of Sports Medicine, Peking University-Shenzhen Hospital, Shenzhen, China

[2]Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong SAR, China

[3]Shenzhen Engineering Laboratory of Orthopaedic Regenerative Technologies, Department of Spine Surgery, Peking University Shenzhen Hospital, Shenzhen, China

[4]Department of Spine Surgery, the First Affiliated Hospital, Sun Yat-sen University, Guangdong, China

[#] These authors contribute equally to this work.

*Corresponding Authors:

Dr. Victor YL Leung, PhD

Department of Orthopedics and Traumatology, Li Ka Shing Faculty of Medicine, The University of Hong Kong

L9-12, Lab Block, 21 Sassoon Road, Hong Kong SAR, China

Tel.: (852) 3917 6985; Fax: (852) 2818 5210

E-mail: vicleung@hku.hk

Abstract

Nucleus pulposus (NP) undergoes fibrosis, leading to structural and mechanical anomalies of intervertebral discs that prone to degeneration. Fibroblastic cells emerge and account for NP fibrosis, yet their nature and origin remain elusive. Degenerative NP cells are highly heterogenous and may contain immune cells. By meta-analysis of single-cell data, we here showed an enrichment of hematopoietic fibrocyte-like cells, defined by CD45 and collagen I dual positivity, in human NP specimens which increases with degeneration severity. We identified fibrocytes subtypes with a myofibroblast differentiation capacity in NP fibrosis. These fibrocytes were recruited in NP fibrosis mediated by injury-induced mouse disc degeneration. Monocyte depletion in CD11b-DTR mice attenuated NP fibrosis and myofibroblast generation, and prevented disc height loss and histomorphological abnormalities. Our findings suggest that fibrocyte-like cells may mediate NP fibrosis and therefore be a potential target for modifying disc degeneration and promoting its repair.

Introduction

The intervertebral discs (IVDs) are the largest avascular connective tissues in the body, playing a central role in spine movement. IVD degeneration (IDD) is associated with fibrosis of the nucleus pulposus (NP), the innermost colloidal core of IVDs, along with a loss of hydration and swelling pressure. The degeneration involves a degradation of hyaline extracellular matrix (ECM), in particular aggrecan and collagen II, and accumulation of fibrotic components including collagen I and small leucin-rich proteoglycans such as biglycan and fibronectin[1]. This leads to reduced tissue mechanical strength and hence disc prolapse under load, ultimately causing back pain and myelopathy[2]. Previous study has shown that transplantation of mesenchymal stromal cells (MSCs) [3] can alleviate IDD via inhibiting NP fibrosis. Understanding and controlling the profibrotic events may therefore hold the key to modifying the degeneration process. However, the mechanism of NP fibrosis and the origin of the fibroblastic population in the IDD remains elusive.

The primitive NP is initially populated by forkhead transcription factor (*FOXA2*)- or brachyury (*TBXT*)-expressing vacuolated cells derived from the notochord[4-6]. In humans, these vacuolated cells are gradually replaced by the chondrocyte-like cells after birth[7]. Tissue fibrosis relates to an excessive fibroblast and myofibroblast activity[8]. Reports have shown an increased number of $\alpha SMA^+/FSP1^+$/collagen I^+ fibroblastic cells in the degenerative IVDs[4,9]. In fact, NP cells have a capacity to become $FSP1^+$ cells when exposed to excessive TGF-β[5,10] as well as undergo fibroblastic/myofibroblastic transition ($\alpha SMA^+/COL1A1^+/FSP1^+/FAP\text{-}\alpha^+$) in an injury-induced IDD model[4]. However, tracing study indicated that not all myofibroblastic cells are derived from local cells[4], implying alternative origins of effector cells for NP fibrosis. In line with this notion, single cell transcriptomic studies have revealed that the NP of human degenerative IVDs[7,11-14] and animal model of IDD[6] contain cell types other than IVD cells. These include endothelial cells, myeloid

granulocytic suppressor cells, CCR2^{+} monocytes and derivatives of macrophages. Consistent with the finding, studies showed that bone marrow cells could migrate into mechanically overloaded IVD explants[15] and that focal enrichment of fibroblastic cells could be found near the infiltrating blood vessels in the degenerative IVDs[16]. These evidences support that access of extrinsic cells to the degenerated IVD may mediate NP fibrosis.

We hypothesized that cell sources other than resident disc cells can contribute to NP fibrosis. In this study, we interrogated the cell hierarchy in the NP of degenerative IVDs through an integrative analysis of the published single cell transcriptomes and characterization in human tissue and cells. Our findings indicated an emergence of hematopoietic populations that possess a fibrocyte-like phenotype and myofibroblastic identity in the degenerative samples. Using a mouse model that labels or depletes monocyte lineage cells, we analyzed the time course of the fibrocyte-like cell appearance and identified their role in NP fibrosis and IDD progression. Our study implies a function of circulatory fibrocytes in NP fibrosis and as a target for modifying IDD.

Results

Identification of myofibroblastic NP cells in disc degeneration and NP fibrosis

We examined the cellular heterogeneity and fibroblastic cell entities by an integrative analysis of 4 published single-cell RNA sequencing (scRNA-seq) data derived from cells isolated from NP tissues[7,12,14] or recognized as NP cells[11]. The datasets consist of 6 non-degenerative human IVDs, 13 mildly (Pfirrmann II-III) and 12 severely (Pfirrmann IV-V) degenerative human IVDs (**Table 1**). A number of existing NP cell phenotypes were previously identified from these datasets, and their marker genes were summarized in **Supplementary Table 1**. The meta-analysis identified 12 cell clusters from a total of 192,184 cells and the cell types were annotated based on their cluster differentially expressed genes (DEGs, Table SI). In addition to 5 NP cell clusters, we also identified

101 neutrophils (*LYZ*+), myeloid derived suppressor cells (MDSC+; *MPO*+), macrophages (*CD68*+), T
102 cells (*TRBC2*+), endothelial cells (*CD34*+ and *PECAM1*+), pericytes (*RGS5*+ and *ACTA2*+), and
103 erythrocytes (*HBB*+) (**Fig. 1a** and **Supplementary Figure S1**). These clusters were mostly present
104 across the four datasets although their proportion and distribution varied (**Fig. 1b**). Clustered NP
105 cells were identified by their expression of *ACAN*, *COL2A1* and *SOX9* (Table SI). The most abundant
106 NP cell population was the chondrogenic cluster (ChonNP) (51%) which expressed *COL2A1*, *ACAN*
107 and *CYR61*. The regulatory cluster (RegNP) (21%) expressed *MMP3*, *GPX3* and *CHI3L2/1*. These
108 RegNP gene markers were also expressed in the inflammatory C1/3 chonNPC and regulatory NPC
109 reported by Han et al[7] and Tu et al[12], respectively (**Supplementary Table S1**). The fibroblastic
110 cluster (FibroNP; 8%) expressed *COL1A1*, *COL3A1*, *POSTN* and *ASPN*, which are known to
111 associate with tissue fibrosis. We found that the proportion of FibroNP significantly increased with
112 the severity of disc degeneration and became one of the major NP cell populations (30%) in severely
113 degenerated NP (**Fig. 1c and Supplementary Figure S2**). The cycling cluster (CyclingNP) (1%)
114 expressed the cell proliferation genes including *TOP2A* and *STMN1*, containing two distinct subsets
115 expressing high level of either collagen I or II gene ($COL1A1^{high}$ and $COL2A1^{high}$). The progenitor
116 cluster (ProgNP) (1%), distinguished by the expression of *PDGFRA* and *PLA2G2A*, was unique to
117 the non-degenerative samples as reported by Gan et al[14]. Endothelial cells (2%) were also found,
118 which highly expressed *PECAM1* and *CD34*. Erythrocytes (1%) were defined by gene expression of
119 *HBB* (**Fig. 1d**).
120
121 Contractile myofibroblasts are primary effector cells of tissue fibrosis[8]. FAPα and FSP1 mark the
122 active fibroblasts or myofibroblast precursors[17], and αSMA marks the mature myofibroblasts[18].
123 These three markers were found enriched in the FibroNP (**Fig. 1e**). More fibrillar collagens,
124 specifically thick collagen I and thin collagen III, were deposited in human degenerative IVDs,
125 which together with reduction of hyaline matrix aggrecan and collagen II, indicated NP fibrosis

126 (**Supplementary Figure S3**). Immunofluorescence showed that the number of cells expressing
127 αSMA (14.0% vs 48%), FAPα (31.4% vs 78.8%) and FSP1 (28.4% vs 65%) in the nNP was far
128 fewer than that in the dNP (**Fig. 1f**), indicating the augmented myofibroblast sprouting in dNP. In
129 line with our single-cell transcriptome analysis, cells from either dNP or nNP tissues were found to
130 contain a lower amount of αSMA^+ cells than $FAP\alpha^+$ or $FSP1^+$ cells, which was further supported by
131 fluorescence-activated cell sorter (FACS) analysis of isolated NP cells (**Fig. 1g**). FACS also
132 demonstrated an increased myofibroblast population in the dNP compared to the nNP. Notably, nNP
133 (88.2%) and dNP (93.1%) cells presented a comparable amount of cells expressing intracellular
134 collagen I (**Supplementary Figure S4**). However, under monolayer expansion, dNP cells became
135 more flattened and morphologically less refractive, and preferentially expressed higher levels of
136 myofibroblastic markers including *ACTA2*, *FAP* and *COL1A1* (**Supplementary Figure S5**). These
137 findings indicated that myofibroblastic cells became the dominant cell type in the degeneration and
138 supported their contribution to NP fibrosis.
139
140

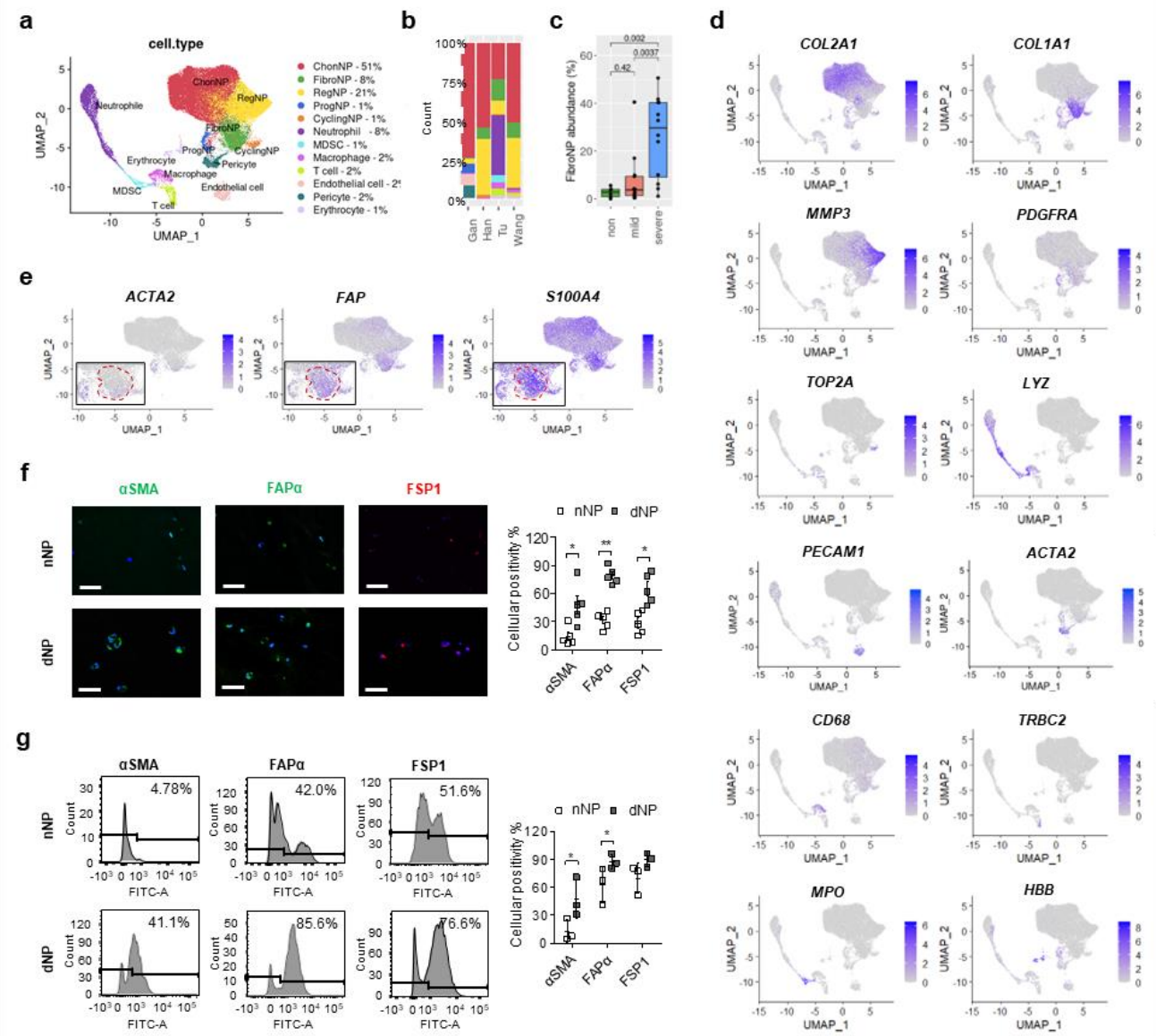

Figure 1. Meta-analysis of single-cell landscape and identification of fibroblastic NPC in human disc degeneration. (a) Identification of 12 distinct cell populations by unsupervised clustering. **(b)** Distribution of the cell populations in individual datasets. **(c)** Degeneration-related FibroNP abundance. **(d)** Signature gene expression for cell clusters defined in A. **(e)** Expression distribution of fibroblastic markers *ACTA2*, *FAP*, and *S100A4*. **(f)** Immunofluorescence and quantification of fibroblastic marker expression in human NP tissues (n=5). Scale bars: 50μm. Nuclei were counterstained with DAPI (Blue). **(g)** Representative flow cytometry plots to identify myofibroblastic NP cells. Percentage of αSMA^{+}, Fap-α^{+} or $Fsp1^{+}$ cells was averaged from three independent experiments. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$ by two-tailed unpaired *t*-tests.

150 *Disc degeneration contains fibroblastic NP cells with hematopoietic features*

151 Previous histological studies suggested that immune cells such as CD68^{+} macrophages and
152 neutrophils may infiltrate into the NP during disc degeneration[19]. In the single cell meta-analysis,
153 immunocyte-like populations could be found at all degeneration grades, expressing markers for
154 neutrophils (*LYZ* and *HLA-DRA*), macrophages (*CD163* and *CD68*), granulocyte-like myeloid
155 derived suppressor cells (G-MDSC) (*ITGAM* and *MPO*) and T cells (*TRAC* and *TRBC2*) (**Table SI**).
156 Interestingly, we found that *COL1A1*, *COL3A1* and *POSTN* were also expressed in the immunocytes,
157 predominantly the neutrophil and macrophage populations (**Fig. 2a**). These *COL1A1*+ immunocytes
158 expressed both myeloid lineage genes (e.g. *S100A9*, *PTPRC* and *ITGB1*) and fibroblastic genes (e.g.
159 *VIM*, *TAGLN2* and *COL1A1*; (**Table SI**). Previous studies reported that monocytes contain a
160 subpopulation that co-expresses collagen I and III in addition to markers of hematopoietic cells
161 (CD11b, CD45 or CD34), being referred as fibrocytes[20-22]. The hematopoietic progenitor-associated
162 marker CD34 was immunolocalized with collagen I and contractile protein αSMA in the NP tissue
163 (**Supplementary Figure S6**). However, CD34 is also highly expressed in other cell types such as
164 endothelial cells and muscle satellite cells. In our meta-analysis, *CD34*$^{+}$*COL1A1*$^{+}$ or *CD34*$^{+}$*ACTA2*$^{+}$
165 cells were in fact mostly enriched in the endothelial cell and pericyte clusters. We further identified
166 *COL1A1*+ subpopulations co-expressing *ITGAM* (coding CD11b) (1094 cells) or *PTPRC* (coding
167 CD45) (1738 cells) (**Fig. 2b**). The abundance of these subpopulations increased with degeneration
168 severity and occupied up to 0.1% cell population dNP tissue (**Fig. 2c**). Immunostaining and flow
169 cytometry verified an increase of CD45^{+}COLI^{+} cells in the dNP tissue (16.4% vs 5.4%, **Fig. 2d**) or
170 dNP-derived cell culture (0.80% vs 0.30%, **Fig. 2e**) compared to the nNP controls. These
171 CD45^{+}COLI^{+} cells adopted an elongated morphology (insert, **Fig. 2d**). Based on the findings, we
172 regarded the CD45^{+}COLI^{+} cells as disc fibrocytes.

173

174 We identified disc fibrocytes within NP cells clusters (NPC), present mostly in the FibroNP (7.5%, as
175 fib-fNP), ChonNP (4.2%, as fib-cNP) and RegNP (3.0%, as fib-rNP), in addition to immunocyte
176 clusters of macrophages (35.4%, as fib-M), neutrophil (34.2%, as fib-N), T cell (11.4%, as fib-T) and
177 G-MDSC (3.0%, fib-G) (**Supplementary Table S2**). By comparing these fibrocyte populations, we
178 identified specific gene markers (**Table SI**) and examined their expression (**Figure 2f**). *GNB1* and
179 gluconeogenesis-related gene *ALDOA*, together with marker genes of *COL1A1* and *ITGAM*, were
180 enriched in both immunocytes- (fib-M/N) and NPC-clustered (fib-cNP/fNP) fibrocytes. Within
181 immunocyte-clustered fibrocyte subpopulations (fib-M/N/T/G), a total of 47 genes (e.g. *PTPRC* and
182 *CAPZA1*) were considerably enhanced and mainly linked to the actin organization and immune
183 process (**Table 2 and Supplementary Figure S7**). *RORA* was dominantly expressed in fib-T. Fib-M-
184 related genes *LGALS1* and *LGALS3BP* have been linked to monocyte-derived fibrocyte
185 differentiation[23]. 43 genes (e.g. *COL3A1*, *ASPN* and *ITGB5*) that are commonly expressed in the
186 three NPC-derived fibrocyte subsets are related to the response to TGF-β (**Supplementary Figure**
187 **S7**). This finding is consistent with the notion that, similar to other fibrotic diseases, disc fibrocyte
188 maturation may involve TGF-β-mediated endothelial- mesenchymal transition[24]. We also performed
189 GO analysis for DEGs specific to the NPC-fibrocytes (**Table 3**). 35 genes distinctly enriched in fib-
190 fNP were mostly related to cell adhesion function. 26 genes enriched in the Fib-cNP were related to
191 matrix organization. Two degeneration-featured genes, *HTRA1*[25] and *ANGPTL4*[26] were distinctly
192 expressed in the NPC-fibrocytes (**Fig. 2g**).
193
194 *ACTA2* was found expressed in *PTPRC*$^{+}$ FibroNP cells (**Fig. 2h**), supporting their potential in
195 myofibroblastic differentiation[23,24]. Consistent with the scRNA-seq finding, co-immunostaining
196 showed a markedly higher expression of CD45 and αSMA in the dNP compared to the nNP (**Fig.**
197 **2h**). We also observed αSMA^{+} cells negative for CD45.

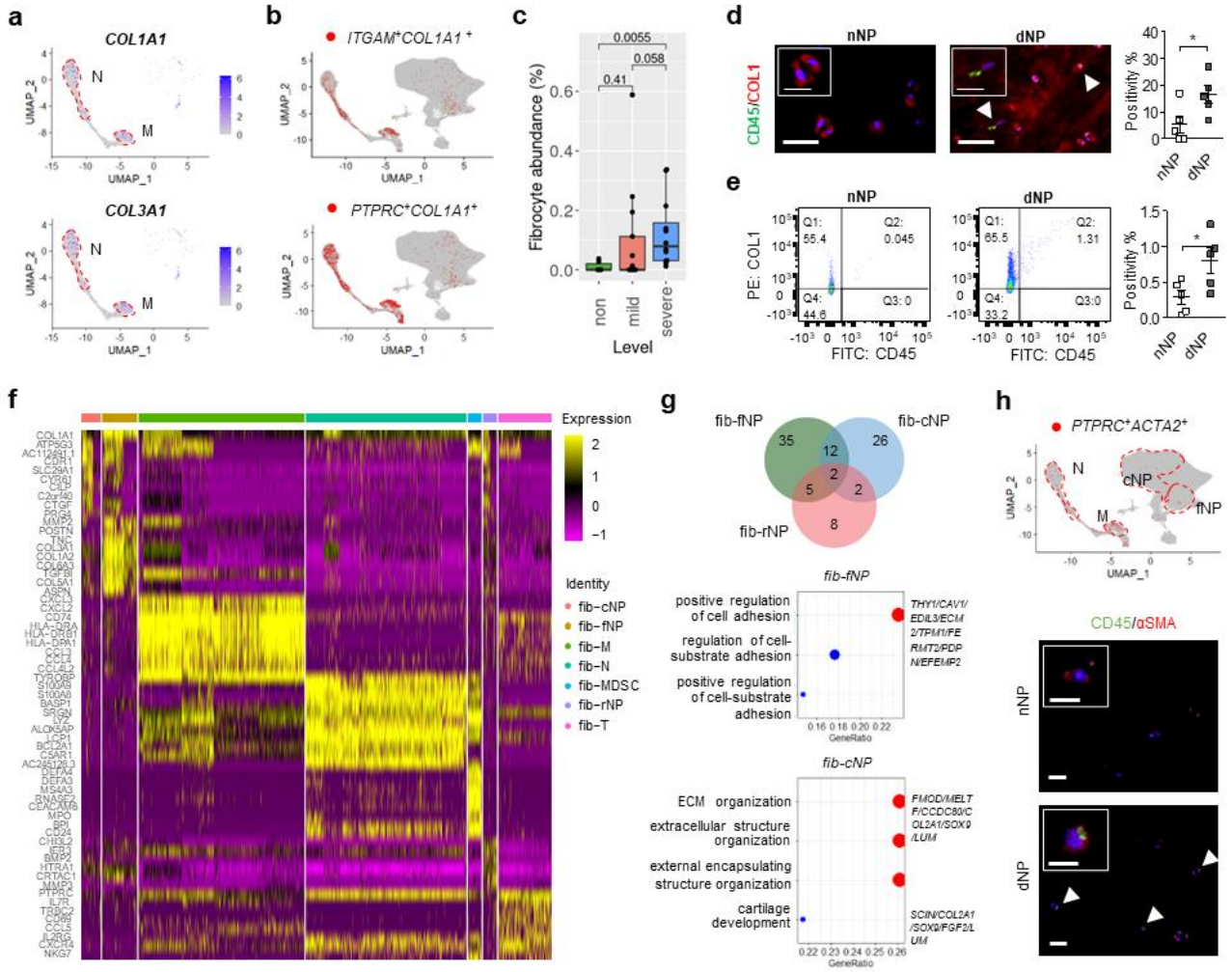

Figure 2. Identification of disc fibrocytes in the disc degeneration. (a) Featured gene plots of fibroblast markers of *COL1A1* and *COL3A1* in disc immunocytes. **(b)** Feature plots of expression distribution for monocyte lineage marker *ITGAM* and pan-hematopoietic marker *PTPRC* with *COL1A1*. **(c)** Fibrocyte abundance in disc degeneration. **(d)** Representative co-immunofluorescence micrographs of CD45 and Collagen I (COL1) in human NP tissues from scoliosis (nNP) and degenerative (dNP) IVDs, and quantification of co-expressing cells (n=5). **(e)** Flow cytometric evaluation of disc fibrocytes in isolated human NP cells. Representative flow cytometry was plotted and the percentages of $CD45^{+}COL1^{+}$ cells were labeled. **(f)** Heatmap showing and highlighting differentially expressed genes for each disc fibrocytes subtype. (**g**) Venn diagram and GO (gene ontology) functions of fibrocytes residing within three NP cells subclusters: FibroNP (fib-fNP), ChonNP (fib-cNP) and RegNP (fib-rNP). **(h)** UMAP distribution and immunodetection of *PTPRC*

209 *(coding CD45)*$^{+}$*ACTA2 (coding* αSMA*)*$^{+}$ cells sin disc NP tissues. Arrowhead: CD45^{+} αSMA^{+} cell;
210 sale bar: 50 μm. Insert: representative staining micrograph showing positive cellular signals; scale
211 bar: 12μm. *$p < 0.05$, **$p < 0.01$ by two-tailed unpaired *t*-tests.
212
213 Given the fibrocytes being a lineage of monocytes expressing CD11b, we studied their
214 temporospatial distribution in the puncture injury induced disc degeneration in CD11b-DTR mice, in
215 which the expression of EGFP-diphtheria toxin (DT) receptor fusion protein is controlled by the
216 *ITGAM* promoter, thereby directing transgene expression and expressing membrane-localized GFP in
217 monocytes/macrophages[22,27] (**Supplementary Figure S8**). Normal mouse NP, which contains
218 predominantly notochordal cells[4], was negative for GFP expression (**Supplementary Figure S8**).
219 Annulus puncture injury led to disc degeneration with NP fibrosis[3,4]. We observed an emergence of
220 GFP^{+} cells in the NP (~15% of NP cells population) by day 3 after puncture, which continued to peak
221 at roughly 35% after 2 weeks and gradually subsided thereafter (**Fig. 3a**). GFP^{+} cells were also
222 observed in the annulus fibrosus (AF) (**Supplementary Figure S8**). We could observe GFP^{+} cells
223 that barely expressed CD11b in the NP at 2 weeks post-puncture (2wpp, **Fig. 3b**), possibly because
224 of a longer turnover of GFP compared to CD11b[27]. Up to 95% of the GFP^{+} NP cells were CD45-
225 positive, indicating their myeloid origin (**Fig. 3c**). In line with previous report of fibrocyte
226 differentiation into myofibroblast, we observed αSMA expression in nearly 60% GFP^{+} cells at 4wpp
227 (**Fig. 3d**). The abundance of these dual-positive cells decreased significantly by 8wpp, indicating that
228 the emergence of fibrocytes might be transient or they underwent differentiation. Furthermore, we
229 detected CD45^{+}COLI^{+} and CD45^{+}αSMA^{+} cells in the punctured disc (**Fig. 4b&c**), and the amount of
230 CD45^{+}αSMA^{+} NP cells showed a two-fold reduction by 8wpp compared to 4wpp. These findings
231 collectively support the notion of fibrocyte infiltration into the degenerating discs and adopting a
232 myofibroblastic phenotype (CD45^{+}αSMA^{+}) during NP fibrosis.

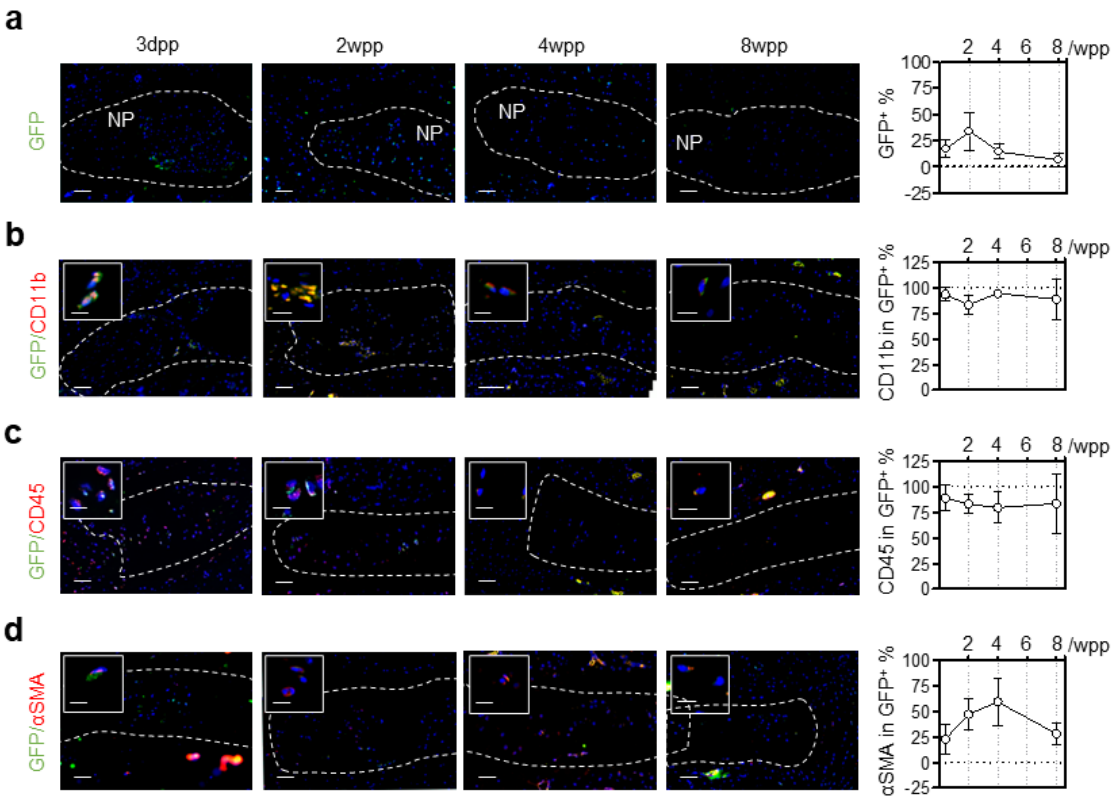

Figure 3. Hematopoietic fibrocytes contribute to myofibroblastic NPCs. (a) Immunostaining and quantification of GFP expression in the degenerative caudal discs of CD11b-DTR in 3 days (dpp) to 8 weeks post-puncture (wpp) window. **(b)** Co-immunostaining of hematopoietic marker CD45 and **(c)** αSMA in $CD11b^+$ cells with GFP with quantification (GFP^+CD34^+, **d**; $GFP^+\alpha SMA^+$, **e**). Cellular positivity counting was conducted in five representative sections, and expressed as the mean ± S.D. Sale bar: 50 μm. * $p<0.05$, One-way ANOVA with Bonferroni post-hoc test.

Fibrocyte depletion rescues NP fibrosis

We tested whether depletion of CD11b-expressing cells by DT administration[27] could lead to a reduction of fibrocytes in the puncture induced IDD. We first examined the frequency and dosage of DT injection to enable extended depletion while minimizing lethality due to compromised immunity (**Supplementary Table S3**). The number of $CD11b^+$ monocytes were largely reduced in peripheral blood (>50%) after twice administration of DT at 10μg/kg (**Supplementary Figure S9**). For the IDD model, the first injection was performed 3 days prior to the puncture surgery (**Fig. 4a**). We found a

247 marked decrease of the GFP^+ cells in the NP at 2 and 8wpp, supporting the effect of DT on the
248 transgene expressing cells (**Supplementary Figure S9**). Interestingly, a lower number of
249 $CD45^+COL1^+$ cells was observed in the NP after the second dose of DT injection by 4wpp when
250 compared to vehicle control (**Fig. 4b**). The amount of $CD45^+\alpha SMA^+$ NP cells also tended to decrease
251 although the difference is not statistically significant (**Fig. 4c**). These findings indicate that DT
252 administration caused a depletion of monocytic fibrocytes in the NP which emerged during injury-
253 induced IDD.
254
255 We questioned if the loss of the myeloid derived fibrocytes leads to a modification of IDD and NP
256 fibrosis. Puncture induced IDD showed a significant disc height loss by 50% at 2wpp and 30% at
257 8wpp (**Fig. 4d**). DT administration could induce a gradual recovery and reduce the disc height loss
258 by >80% at 8wpp. Disc integrity assessed by FAST staining[3] showed a prominent improvement with
259 DT administration at both 4 and 8wpp, including reduced lacune formation, preserved NP-AF
260 boundary, and increased alcian blue intensity in the NP (**Fig. 4e**). Moreover, alcian blue staining in
261 the NP of the DT-injected mice was largely maintained, implying that the loss of proteoglycans was
262 prevented. The amount of myofibroblastic NP cells positive for αSMA steadily increased after
263 puncture and took up to 76% of the whole population by 8wpp (**Fig. 4c**). DT administration could
264 lower its positivity to 41%. It is noteworthy that the majority of αSMA^+ cells detected in the
265 punctured discs, in particular in the DT-treated group, was negative for CD45 (>90%). Picrosirius red
266 staining and immunostaining of collagen I further consolidated a decrease in fibrillar collagens (**Fig.
267 4f-h**). Pericellular expression of collagen I was noted in the NP after DT administration. Collectively,
268 these findings demonstrated that IDD progression is dependent on a function of myeloid derived
269 cells where the $CD45^+COL1A1^+$ fibrocytes and αSMA^+ myofibroblastic cells may mediate NP
270 fibrosis.

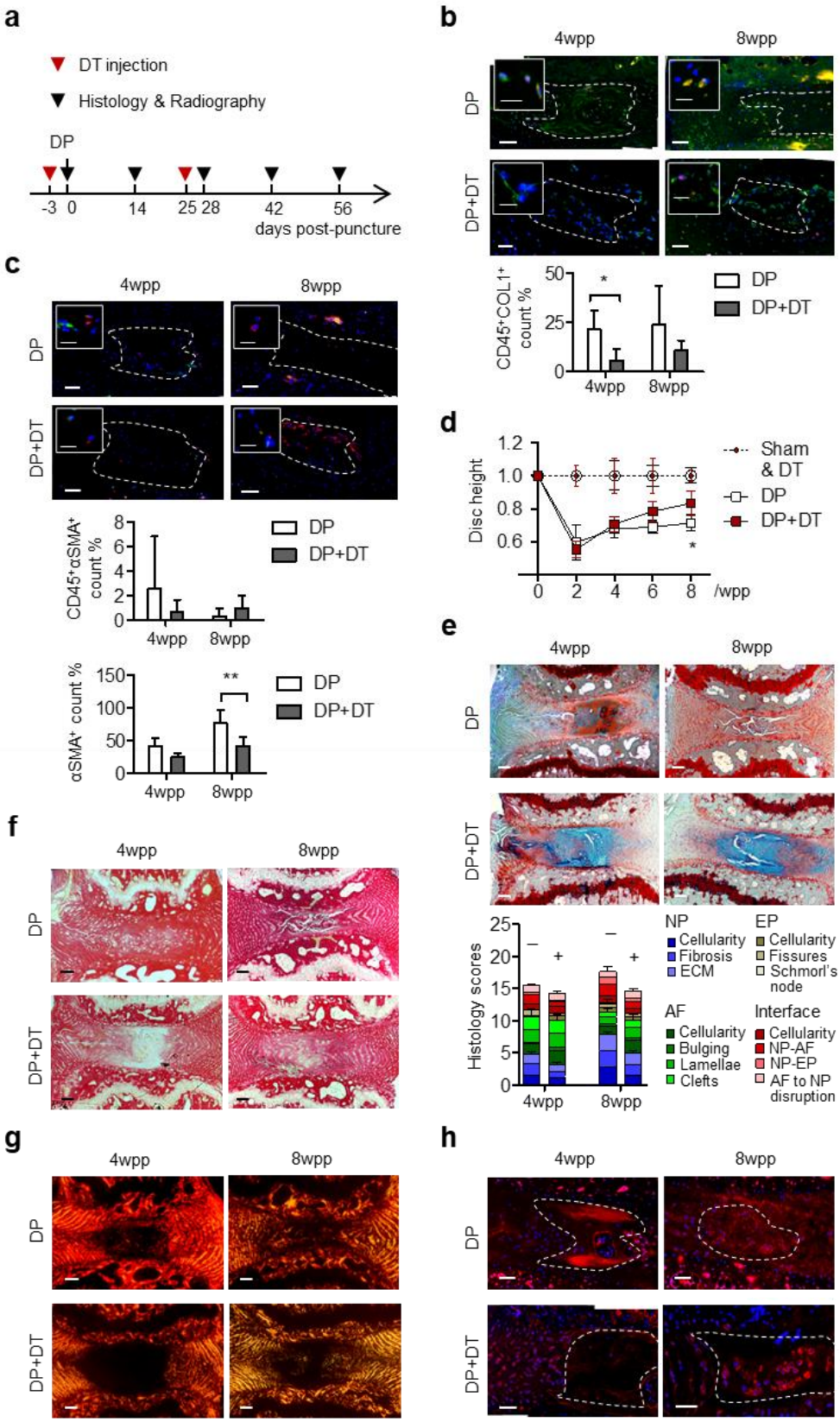

Figure 4. Fibrocyte deficiency suppresses disc degeneration and NP fibrosis. (a) Time-line illustration for DT administration and puncture surgery on CD11b-DTR mice. PBS serves as vehicle control. **(b)** Co-staining of CD45 (red) with COL1 **(**green**)** or αSMA **(c,** green**)** and dual-positive cells

274 counts. αSMA positivity in NP cells was also counted. **(d)** Time course assessment of disc height
275 changes (n=5 per group). **(e)** FAST staining and degeneration scoring based on compartmental
276 deformation in NP, annulus fibrosus (AF), endplate (EP) and interface. **(f)** Picrosirius red staining
277 and **(g)** their polarized microscopy. **(h)** Collagen I immunostaining. In immunofluorescence, nuclei
278 were counterstained by DAPI. Scale bar: 50μm, insert scale bar: 12μm. Dot line: NP region. Wpp:
279 weeks post-puncture; DP: disc puncture. Data are expressed as mean ± S.D. Two-way ANOVA with
280 Bonferroni post-hoc test, * $p<0.05$.
281

282 **Discussion**

283 NP cells are conventionally defined as collagen II-, aggrecan- and Sox9-expressing cells. Recent
284 studies of single-cell transcriptome have indicated cell heterogeneity in the human NP, which
285 contains cell populations expressing notochordal, chondrogenic, and fibroblastic phenotypes[7,11-14].
286 Our meta-analysis of the human disc scRNA-seq datasets has outlined the commonly featured NP
287 cells hierarchy and stratified their fibroblastic subtypes. *COL1A1*$^+$ FibroNP is previously defined as
288 adhesion/fibroNPC[12] or C4 NPC[7]. The FibroNP population was not reported in Wang et al scRNA-
289 seq dataset, possibly because their cells were derived from the whole disc tissue without NP isolation
290 and therefore may have dominated by *COL1A1*$^+$ AF cells[11]. Nonetheless, they defined the FibroNP
291 genes (*POSTN*, *COL1A1*, *COL3A1* and *TMSB4X*) in MK167$^+$ progenitor and NP progenitor cells
292 (**Supplementary Table 1**), implying the fibroblast presence within the two NP subtypes. Han et al
293 found a small fraction of fibrochondrocyte progenitors in the NP, which partially constitute the
294 aforementioned two NP subsets. The regulatory NPC and C1/3 chonNPC subtypes reported by Tu[12]
295 and Han[7] are merged to the RegNP cluster in this study, sharing the expression of *MMP3*, *CHI3L2/1*
296 and *GPX3*. The CyclingNP cluster is a pool of *TOP*2A$^+$ cells fraction from the *COL2A1*$^+$ ChonNP or
297 *COL1A1*$^+$ FibroNP subsets, implying they may be derived from a common replicative progenitor

298 origin. Pseudotime ordering of the NP clusters and velocity analysis may help construct their
299 relationship.
300
301 The combination of collagen production (e.g. collagen I) and the expression of leukocytic common
302 antigen (CD45) or one of the hematopoietic or myeloid antigens (e.g. CD34 or CD11b) is considered
303 as a sufficient criterion for fibrocyte identity[21]. In our meta-analysis, *CD34*$^+$*COL1A1*$^+$ cells are
304 mainly enriched in the endothelial cells and pericytes subpopulations (**Supplementary Figure S6**).
305 *ITGAM* (coding CD11b) or *PTPRC* (coding CD45) is therefore preferrable to CD34 as the marker of
306 fibrocytes in human NP tissues. Our meta-analysis showed a wider UMAP distribution and higher
307 number of *PTPRC*$^+$ than *ITGAM*$^+$ cells, supporting the notion that CD11b is a more selective yet less
308 representative fibrocyte marker compared to the pan-hematopoietic markers Vav-1 or CD45[21].
309 Pronase digestion was commonly deployed for releasing single cells from the NP in scRNA-seq and
310 flow cytometry assays[28]. However, the digestion can also results in extensively membrane proteins
311 degradation, and thereby affect the antibody recognition of surface molecules[29]. This may in part
312 explain the discrepancy in measurements of CD45$^+$COL1A1$^+$ cells between in situ
313 immunohistochemistry and flow cytometry (**Fig. 2d&e**).
314
315 In addition to the NP clusters, *CD11b*$^+$*COL1A1*$^+$ or *CD45*$^+$*COL1A1*$^+$ cells are mostly located within
316 the macrophages, neutrophils and T cells clusters, implying the fibrocytes may be derived from
317 multiple sources or differentiation routes of the immune system. Monocytes are defined as
318 *CD45*$^+$*CD11b*$^+$ cells and constitute the myeloid-derived macrophages (*CD45*$^+$*CD11b*$^+$*F4/80*$^+$),
319 neutrophils (*CD45*$^+$*CD11b*$^+$*HLA-DRA*$^+$) and MDSCs (*CD45*$^+$*CD11b*$^+$*OLR1*$^+$)[12,13,21]. G-MDSCs,
320 shown in cardiac fibrosis with myofibroblast activation capabilities, contribute to T-cell suppression
321 and ROS production in IDD[12]. Monocyte-derived inflammatory macrophages could transit into
322 myofibroblasts[30,31] and were identified in the degenerative NP[19,32]. CD11b$^+$CD45$^+$ monocytes were

323 previously reported to support the proliferation and colony forming ability of MSCs[33]. Whether the
324 monocytic cells infiltrated into the degenerating discs may exert a similar effect on local
325 mesenchymal progenitors to promote fibrogenesis or repair is not clear. Moreover, $CCR2^+$
326 monocytes have been identified in the degenerative IVDs from herniation patients and constitute the
327 majority of $CD11b^+$ and $F4/80^+$ cells[34]. However, as reported in myocardial infarction model[35], loss
328 of CCR2 expression is required for fibrocyte differentiation of the infiltrated monocytes. This is in
329 line with our observation that the disc fibrocytes are negative for *CCR2*. In a mouse model of renal
330 fibrosis where CCR2 is depleted, migration of circulating fibrocytes to the kidney was interfered[27].
331 Whether CCR2 plays a role in monocyte infiltration and hence differentiation into disc fibrocytes
332 await to be determined.
333
334 Fibrocytes could be derived not only from local monocyte differentiation but also migrated from
335 peripheral blood in the form of partially differentiated collagen-producing cells[21,27]. Recruitment of
336 monocytes and peripheral blood fibrocytes may rely on disc neo-vascularization. Normally, blood
337 vessels recede from the outer AF and the cartilaginous endplate (EP) in adults. However, under
338 degenerative conditions the NP can become re-vascularized and blood vessel infiltration is frequently
339 observed and spatially associated with macrophage infiltration[36]. Most of the degenerative NP tissues
340 in this study for scRNA-seq analysis were dissected from extruded NP of herniated discs that
341 normally show pronounced neo-vascularization[37]. The increased number of GFP^+ cells found in the
342 AF and EP of the punctured CD11b-DTR discs may indicate the infiltration through the AF and EP
343 routes. Pericytes ($CD34^+\alpha SMA^+$) were immunodetected in the degenerative human NP
344 (**Supplementary Figure S6)**, supporting the presence of neo-vascularization. In fact, pericytes can
345 contribute to scar-forming myofibroblasts in kidney and subretinal fibrosis[8]. Studying disc
346 vascularization and its relation to the emergence of disc fibrocytes may provide further insights into
347 the pathway of NP fibrosis.

348

349 Trafficking fibrocytes express chemokine receptors CCR5[38], CCR7[21] and CXCR4[21,39]. Studies have
350 reported increased CCR5 expression in AF and NP cells[40] and chemokines CCL2/7 in degenerated
351 NP[41]. CXCR4 is a receptor of SDF1[42] and CXCL12[39], which was found upregulated in degenerated
352 discs and implicated in the disc angiogenesis[42] and cellular apoptosis[43] via PI3K/NF-κB pathway.
353 Bone marrow cells, such as CD146$^+$ MSCs[44], may have a potential to migrate into the NP, possibly
354 by control of SDF1[45]. CCL5 could interact with CCR1/3/5 in NP cells and was up-regulated under
355 inflammatory stimuli[46]. Interestingly, we find that the disc fibrocytes highly express CXCR4 and
356 CCR1 (**Table SI**). Future analysis of ligand-receptor interaction, particularly the CXCR4-SDF1 and
357 CCR1-CCL5 axis in the *CD45*$^+$*COL1A1*$^+$ fibrocytes may gain insights into the regulatory milieu for
358 their recruitment.

359

360 Tissue fibrosis is mediated by myofibroblast from local or extrinsic sources[8] and fibrocytes may
361 serve as one of the progenitors of myofibroblasts[21,24]. Expression of αSMA is the hallmark of mature
362 myofibroblasts and essential to their cellular contractility[18]. Expression of αSMA in NP cells may
363 therefore indicate their myofibroblastic property[24]. FAP-α is a well-known biomarker of
364 (myo-)fibroblasts, and together with FSP1 it can label most of the collagen I-producing cells in the
365 bone[47]. FSP1 marks an active fibroblast population and represents a non-overlapping fibroblast entity
366 to αSMA$^+$ subtype in heart, kidney and skin fibrosis disease[48]. FSP1 is also expressive in atypical
367 myeloid-derived F4/80-positive inflammatory macrophages in liver injury[49]. FSP1 was found
368 expressed in the *COL2A1*$^+$ ChonNP cells in addition to the *COL1A1*$^+$ FibroNP cells.

369

370 Our results demonstrate αSMA expression in the CD45$^+$ and CD11b$^+$ fibrocytes, supporting the
371 conversion of fibrocytes to myofibroblasts in NP fibrosis. Yet, a vast number of αSMA$^+$
372 myofibroblastic NP cells in both human degenerative IVDs and mouse punctured discs are not

373 positive for CD45 and CD11b. In the CD11b-DTR mice where the monocytes and their derived
374 lineages (including fibrocytes) were depleted[27], αSMA expressing cells could still be observed
375 (>55%, 2 wpp in **Fig. 4d**) in the degenerative NP. This may be due to the existence of non-
376 hematopoietic origins of the myofibroblast pool, such as those derived from resident NP cells[4].
377 Alternatively, fibrocytes may lose CD45 and CD11b expression during fibrocyte differentiation[30].
378 We also note the presence of residual $CD45^{+}COL1^{+}$ cells and GFP^{+} cells after DT administration.
379 This likely arises from the incomplete monocyte ablation and a continuation of fibrocyte recruitment.
380
381 Fibrocyte-to-myofibroblast conversion acquires activation of TGF-β/SMAD signaling[50], and
382 interleukin-18 receptor 1[51] and muscarinic receptor M3[52] are shown to promote the fibrocyte
383 formation and contractile function. Moreover, fibrocytes can secret soluble factors such as periostin[53]
384 and TIMP1[54] and semaphorin-7A[55] to regulate myofibroblast activity and collagen expression in
385 pulmonary and intestinal fibrosis[20]. Class 3 semaphorins expression are found associated with disc
386 innervation and angiogenesis[56]. Interestingly, TIMP1[57] and periostin[58] are up-regulated in the
387 degenerated discs and enriched in the FibroNP from our meta-analysis. Whether disc fibrocytes
388 regulate NP fibrosis through a paracrine manner awaits to be addressed. Furthermore, the NPC-
389 enriched disc fibrocytes express *HTRA1* and *ANGPTL4*. HTRA1 is a proteolytic enzyme for
390 degrading the matrix either on its own[59] or by upregulating ADAMTS-5[25] or MMPs, whereas
391 expression of ANGPTL4 is positively associated with IDD severity [26]. Their regulatory role in the
392 fibrocytes and NP fibrosis worths to be investigated.
393
394 In conclusion, this study implicates a function of monocytic fibrocytes in NP fibrosis and IDD
395 progression. Understanding the effects and regulation of the fibrocytes may potentiate the
396 development of new diagnostic and therapeutic strategies for modifying IDD.
397

Methods

scRNA-seq data acquisition and downstream analysis

Four single cell RNA sequencing data sets (GSE160756, GSE165722, CNP0002664, PRJCA014236) were downloaded and processed for next-step analysis. 31 individual samples were rearranged according to Pfirrmann grade, which Pfirrmann grade I samples were classified as non-degenerative NP, Pfirrmann grade II/III as mildly degenerated NP, and Pfirrmann grade IV/V as severely degenerated NP (**Table 1**). Seurat package (version: 3.1.4, https://satijalab.org/seurat/) was used for cell normalization and regression according to the UMI counts of each sample and percent of mitochondria rate to obtain the scaled data. MNN (mutual nearest neighbor) was used to remove potential batch effect. The top 10 principles were used for unsupervised UMAP construction via graph-based cluster method (resolution = 0.8), and marker genes were determined with p-value < 0.01 and log(fold-change) > 0.25 by the FindAllMarkers function. Significant mean and cell communication significance (p-value < 0.05) were calculated based on the interaction and the normalized cell-matrix achieved by Seurat normalization.

Human samples

All procedures described in the present study were approved by the Institutional Ethics Review Board of The University of Hong Kong. IVD tissues were obtained from patients under informed consent. Non-degenerative (n) and degenerative (d) IVDs were harvested from respectively adolescent idiopathic scoliosis (n=7) and lumbar disc degeneration patients (n=9). The age of the patients ranged from 14 to 72 years, and tissue was evaluated with the Pfirrmann classification system (**Supplementary Table 4**).

Cell isolation and culture

Jelly-like NP tissues were carefully dissected from the cartilaginous EP and lamellar-structured AF. In severely degenerated IVDs (grade IV-V), NP tissues were more fibrotic and hardly distinguishable from inner AF, so we harvested the tissue presumably at the central region of the NP. Primary NP cells were extracted by sequential enzymes digestion of pronase (0.25%, Roche Diagnostics) and collagenase II (600U/ml, Worthington Biochemical)[28], and maintained in high-glucose Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (FBS, Biosera) and 1% penicillin/streptavidin (P/S, Invitrogen) at 37 °C and 5% CO_2.

Flow cytometry

Newly extracted NP cells were washed by PBS and re-suspended in PBS. Cells were fixed in 4% paraformaldehyde (PFA) and blocked by 5% BSA. Permeabilization with 0.1% triton X-100 (Sigma-Aldrich) was applied where necessary. Cells were then treated with primary and secondary antibodies, and signals from cells labeled with conjugated fluorophores were detected by using BD FACS CantoII Analyzer (BD Biosciences). The antibodies used for different flow cytometry analysis are listed in **Supplementary Table S5**. Appropriate IgG control fluorescence compensation was applied to avoid false positive signals. Data were further analyzed by BD FACS Diva software (BD Bioscience).

Real-time quantitative PCR

Total RNA was prepared using RNeasy mini kit (Invitrogen) and cDNA synthesis was performed using High-capacity RNA-to-cDNA kit (Applied Biosystems). PCR was performed using PowerUp SYBR green master mix (Applied Biosystems), and data are presented as expression levels relative to *GAPDH* using the $2^{-\Delta CT}$ method. The primer sequences were listed in **Supplementary Table S6**.

Mouse disc puncture model

Tg (ITGAM-DTR/eGFP)34Lan/J mice express membrane localized GFP fluorescence with diphtheria toxin (DT) receptor under the control of CD11b and were used for disc puncture. The animal experiments were approved by The University of Hong Kong Committee on the Use of Live Animals in Teaching and Research (CULATR). CD11b-DTR mice (female, 12 weeks of age, *n*=5 per group and time points) received needle puncture at levels of C5/6 and C7/8 of tail IVDs as described previously[4]. At four days prior to the timepoints and every other week after the puncture surgery, these mice were subjected to intraperitoneal injection of saline or DT (25μg/kg body weight) (Sigma-Aldrich). At assigned end points, mice were euthanized by intravenous injection of sodium pentobarbital (1.2 g/kg), and intact discs with attached vertebral bones were harvested for histochemistry. Peripheral bloods were harvested at days 0, 3, and 7 after DT injection, and number of GFP^+ monocytes was measured by fluorescence-activated cell sorting (FACS).

Disc height measurement

Anterior-to-posterior radiographs of tail discs of CD11b-DTR mice or C57BL/6J were taken using a digital radiography machine (Siemens) with an exposure at 25 kV, for 5 seconds. Disc height index was calculated as previously described [4] and presented as percent relative to the unoperated level of tail disc C6/7. Images were analyzed biweekly in all groups up to 8 weeks.

Histology and immunostaining

Samples from human and CD11b-DTR mice tissues were fixed with 4% paraformaldehyde in phosphate buffer for 24 h. Decalcification was performed for mouse IVDs using mouse solution. Paraffin-embedded tissue blocks of human NP and mice IVDs were sectioned at 7 μm. For histological staining, sections were deparaffinized and rehydrated, and subjected to FAST staining (Alcian blue, Safranin O, Fast green and Tartrazine stain) (Sigma-Aldrich) as previously described[60], or Masson's trichrome stain and Picrosirius red stain following standard procedures. Polarized microscopy was

combined with picrosirius red staining to observe the collagens network in human NP tissue and mouse IVDs[61]. According to the histomophological changes, disc degeneration in CD11b-DTR mice was scored with reference to Tam's scoring system[60].

For immunofluorescence, antigen retrieval was performed following rehydration by incubating with 0.8% hyaluronidase (Sigma-Aldrich) at 37°C for 30 min and antigen retrieval buffer (100 mM Tris, 5% EDTA, pH 9.5) at 95°C for 10 min. The sections were blocked in protein block solution (DAKO) for 60 min before incubating with antibodies. Fluorophore conjugated antibodies were used to enable fluorescence detection. Information for these primary and secondary antibodies were summarized in **Supplementary Table S5**. The sections were mounted with VECTASHIELD Mounting Medium with DAPI for nuclei staining. Isotype-matched mouse, rabbit or rat IgG was used as negative controls. Tissue sections were imaged by fluorescence microscope (Nikon Eclipse E600), and fluorescent positive cells in five fields of view (FOV) were manually counted in ImageJ with reference to total number of cells.

Data analysis

Data are presented as mean ± S.D. of four to six experiments, as indicated in respective figure legends. Unpaired t-test (two tailed) was used to determine significant differences. Comparisons among multiple groups were assessed using one-way ANOVA with Bonferroni post-hoc test. A p value < 0.05 was considered statistically significant.

References

1. Sun, Y., Lyu, M., Lu, Q., Cheung, K. & Leung, V. Current Perspectives on Nucleus Pulposus Fibrosis in Disc Degeneration and Repair. *Int J Mol Sci* **23**(2022).
2. Zhang, S., *et al.* The role of structure and function changes of sensory nervous system in intervertebral disc-related low back pain. *Osteoarthritis Cartilage* **29**, 17-27 (2021).
3. Leung, V.Y., *et al.* Mesenchymal stem cells reduce intervertebral disc fibrosis and facilitate repair. *Stem Cells* **32**, 2164-2177 (2014).
4. Au, T.Y.K., *et al.* Transformation of resident notochord-descendent nucleus pulposus cells

in mouse injury-induced fibrotic intervertebral discs. *Aging Cell* **19**, e13254 (2020).

5. Zhang, Y., *et al.* Directed Differentiation of Notochord-like and Nucleus Pulposus-like Cells Using Human Pluripotent Stem Cells. *Cell Rep* **30**, 2791-2806 e2795 (2020).
6. Zhou, T., *et al.* Spatiotemporal Characterization of Human Early Intervertebral Disc Formation at Single-Cell Resolution. *Adv Sci (Weinh)* **10**, e2206296 (2023).
7. Han, S., *et al.* Single-Cell RNA Sequencing of the Nucleus Pulposus Reveals Chondrocyte Differentiation and Regulation in Intervertebral Disc Degeneration. *Front Cell Dev Biol* **10**, 824771 (2022).
8. Kuppe, C., *et al.* Decoding myofibroblast origins in human kidney fibrosis. *Nature* **589**, 281-286 (2021).
9. Lv, F.J., *et al.* Matrix metalloproteinase 12 is an indicator of intervertebral disc degeneration co-expressed with fibrotic markers. *Osteoarthritis Cartilage* **24**, 1826-1836 (2016).
10. Chen, C., *et al.* Autologous fibroblasts induce fibrosis of the nucleus pulposus to maintain the stability of degenerative intervertebral discs. *Bone Res* **8**, 7 (2020).
11. Wang, D., *et al.* Single-cell transcriptomics reveals heterogeneity and intercellular crosstalk in human intervertebral disc degeneration. *iScience* **26**, 106692 (2023).
12. Tu, J., *et al.* Single-Cell Transcriptome Profiling Reveals Multicellular Ecosystem of Nucleus Pulposus during Degeneration Progression. *Adv Sci (Weinh)* **9**, e2103631 (2022).
13. Ling, Z., *et al.* Single-Cell RNA-Seq Analysis Reveals Macrophage Involved in the Progression of Human Intervertebral Disc Degeneration. *Front Cell Dev Biol* **9**, 833420 (2021).
14. Gan, Y., *et al.* Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitors in human intervertebral discs. *Bone Res* **9**, 37 (2021).
15. Sakai, D., *et al.* Migration of bone marrow-derived cells for endogenous repair in a new tail-looping disc degeneration model in the mouse: a pilot study. *Spine J* **15**, 1356-1365 (2015).
16. Melrose, J., *et al.* Spatial and temporal localization of transforming growth factor-beta, fibroblast growth factor-2, and osteonectin, and identification of cells expressing alpha-smooth muscle actin in the injured anulus fibrosus: implications for extracellular matrix repair. *Spine (Phila Pa 1976)* **27**, 1756-1764 (2002).
17. Esteve, D., *et al.* Lobular architecture of human adipose tissue defines the niche and fate of progenitor cells. *Nat Commun* **10**, 2549 (2019).
18. Hinz, B. & Lagares, D. Evasion of apoptosis by myofibroblasts: a hallmark of fibrotic diseases. *Nat Rev Rheumatol* **16**, 11-31 (2020).
19. Li, X.C., *et al.* Investigation of macrophage polarization in herniated nucleus pulposus of patients with lumbar intervertebral disc herniation. *J Orthop Res* **41**, 1335-1347 (2023).
20. Abe, S., *et al.* Functional analysis of human fibrocytes derived from monocytes reveals their profibrotic phenotype through paracrine effects. *J Med Investig* **67**, 102-112 (2020).
21. Reinhardt, J.W. & Breuer, C.K. Fibrocytes: A Critical Review and Practical Guide. *Front Immunol* **12**, 784401 (2021).

542 22. Ozono, Y., *et al.* Neoplastic fibrocytes play an essential role in bone marrow fibrosis in
543 Jak2V617F-induced primary myelofibrosis mice. *Leukemia* **35**, 454-467 (2021).
544 23. White, M.J., Roife, D. & Gomer, R.H. Galectin-3 Binding Protein Secreted by Breast Cancer
545 Cells Inhibits Monocyte-Derived Fibrocyte Differentiation. *J Immunol* **195**, 1858-1867
546 (2015).
547 24. Abu El-Asrar, A.M., *et al.* Myofibroblasts in proliferative diabetic retinopathy can originate
548 from infiltrating fibrocytes and through endothelial-to-mesenchymal transition
549 (EndoMT). *Exp Eye Res* **132**, 179-189 (2015).
550 25. Li, D., *et al.* HtrA1 upregulates the expression of ADAMTS-5 in HNPCs via the ERK/NF-
551 kappaB/JNK signaling pathway. *Am J Transl Res* **11**, 5114-5121 (2019).
552 26. Liu, F.J., *et al.* Expression of ANGPTL4 in Nucleus Pulposus Tissues Is Associated with
553 Intervertebral Disc Degeneration. *Dis Markers* **2021**, 3532716 (2021).
554 27. Reich, B., *et al.* Fibrocytes develop outside the kidney but contribute to renal fibrosis in a
555 mouse model. *Kidney Int* **84**, 78-89 (2013).
556 28. Basatvat, S., *et al.* Harmonization and standardization of nucleus pulposus cell extraction
557 and culture methods. *JOR Spine* **6**, e1238 (2023).
558 29. Srakocic, S., *et al.* Proposed practical protocol for flow cytometry analysis of microglia
559 from the healthy adult mouse brain: Systematic review and isolation methods' evaluation.
560 *Front Cell Neurosci* **16**, 1017976 (2022).
561 30. Sinha, M., *et al.* Direct conversion of injury-site myeloid cells to fibroblast-like cells of
562 granulation tissue. *Nat Commun* **9**, 936 (2018).
563 31. Meng, X.M., *et al.* Inflammatory macrophages can transdifferentiate into myofibroblasts
564 during renal fibrosis. *Cell Death Dis* **7**, e2495 (2016).
565 32. Burt, K.G., Kim, M.K.M., Viola, D.C., Abraham, A.C. & Chahine, N.O. Nuclear Factor Kappa B
566 Over-Activation in the Intervertebral Disc Leads to Macrophage Recruitment and Severe
567 Disc Degeneration. *bioRxiv* (2023).
568 33. Hong, H.S., Ahn, W. & Son, Y. CD45+/CD11b+monocytes are required for mesenchymal
569 stem cell proliferation. *Tissue Engineering and Regenerative Medicine* **11**, 224-229 (2014).
570 34. Jin, L., *et al.* CCR2 monocytes as therapeutic targets for acute disc herniation and
571 radiculopathy in mouse models. *Osteoarthritis Cartilage* (2023).
572 35. Mouton, A.J., *et al.* Mapping macrophage polarization over the myocardial infarction time
573 continuum. *Basic Res Cardiol* **113**, 26 (2018).
574 36. Zhang, S., *et al.* Inhibiting Heat Shock Protein 90 Attenuates Nucleus Pulposus Fibrosis
575 and Pathologic Angiogenesis Induced by Macrophages via Down-Regulating Cell
576 Migration-Inducing Protein. *Am J Pathol* **193**, 960-976 (2023).
577 37. Yasuma, T., Arai, K. & Yamauchi, Y. The histology of lumbar intervertebral disc herniation.
578 The significance of small blood vessels in the extruded tissue. *Spine (Phila Pa 1976)* **18**,
579 1761-1765 (1993).
580 38. Ishida, Y., *et al.* Essential roles of the CC chemokine ligand 3-CC chemokine receptor 5 axis
581 in bleomycin-induced pulmonary fibrosis through regulation of macrophage and
582 fibrocyte infiltration. *Am J Pathol* **170**, 843-854 (2007).
583 39. Phillips, R.J., *et al.* Circulating fibrocytes traffic to the lungs in response to CXCL12 and

584 mediate fibrosis. *J Clin Invest* **114**, 438-446 (2004).
585 40. Liu, W., *et al.* Annulus fibrosus cells express and utilize C-C chemokine receptor 5 (CCR5)
586 for migration. *Spine J* **17**, 720-726 (2017).
587 41. Phillips, K.L., *et al.* The cytokine and chemokine expression profile of nucleus pulposus
588 cells: implications for degeneration and regeneration of the intervertebral disc. *Arthritis*
589 *Res Ther* **15**, R213 (2013).
590 42. Zhang, H., *et al.* SDF1/CXCR4 axis facilitates the angiogenesis via activating the PI3K/AKT
591 pathway in degenerated discs. *Mol Med Rep* **22**, 4163-4172 (2020).
592 43. Liu, Z., *et al.* SDF-1/CXCR4 axis induces apoptosis of human degenerative nucleus
593 pulposus cells via the NF-kappaB pathway. *Mol Med Rep* **14**, 783-789 (2016).
594 44. Wangler, S., *et al.* CD146/MCAM distinguishes stem cell subpopulations with distinct
595 migration and regenerative potential in degenerative intervertebral discs. *Osteoarthritis*
596 *Cartilage* **27**, 1094-1105 (2019).
597 45. Leite Pereira, C., *et al.* Stromal Cell Derived Factor-1-Mediated Migration of Mesenchymal
598 Stem Cells Enhances Collagen Type II Expression in Intervertebral Disc. *Tissue Eng Part A*
599 (2018).
600 46. Gruber, H.E., *et al.* Production and expression of RANTES (CCL5) by human disc cells and
601 modulation by IL-1-beta and TNF-alpha in 3D culture. *Exp Mol Pathol* **96**, 133-138 (2014).
602 47. Chen, Y., *et al.* Type-I collagen produced by distinct fibroblast lineages reveals specific
603 function during embryogenesis and Osteogenesis Imperfecta. *Nat Commun* **12**, 7199
604 (2021).
605 48. Saraswati, S., Marrow, S.M.W., Watch, L.A. & Young, P.P. Identification of a pro-angiogenic
606 functional role for FSP1-positive fibroblast subtype in wound healing. *Nat Commun* **10**,
607 3027 (2019).
608 49. Osterreicher, C.H., *et al.* Fibroblast-specific protein 1 identifies an inflammatory
609 subpopulation of macrophages in the liver. *Proc Natl Acad Sci U S A* **108**, 308-313 (2011).
610 50. Mitsuhashi, A., *et al.* Identification of fibrocyte cluster in tumors reveals the role in
611 antitumor immunity by PD-L1 blockade. *Cell Rep* **42**, 112162 (2023).
612 51. Niu, X.H., *et al.* IL-18/IL-18R1 promotes circulating fibrocyte differentiation in the aging
613 population. *Inflamm Res* **69**, 497-507 (2020).
614 52. Henrot, P., *et al.* Muscarinic receptor M3 activation promotes fibrocytes contraction. *Front*
615 *Pharmacol* **13**, 939780 (2022).
616 53. Ashley, S.L., Wilke, C.A., Kim, K.K. & Moore, B.B. Periostin regulates fibrocyte function to
617 promote myofibroblast differentiation and lung fibrosis. *Mucosal Immunol* **10**, 341-351
618 (2017).
619 54. Kuroda, N., *et al.* Infiltrating CCR2(+) monocytes and their progenies, fibrocytes,
620 contribute to colon fibrosis by inhibiting collagen degradation through the production of
621 TIMP-1. *Sci Rep* **9**, 8568 (2019).
622 55. Gan, Y., *et al.* Role of semaphorin 7a signaling in transforming growth factor beta1-
623 induced lung fibrosis and scleroderma-related interstitial lung disease. *Arthritis Rheum*
624 **63**, 2484-2494 (2011).
625 56. Binch, A.L., *et al.* Class 3 semaphorins expression and association with innervation and

626 angiogenesis within the degenerate human intervertebral disc. *Oncotarget* **6**, 18338-
627 18354 (2015).
628 57. Vo, N.V., *et al.* Expression and regulation of metalloproteinases and their inhibitors in
629 intervertebral disc aging and degeneration. *Spine J* **13**, 331-341 (2013).
630 58. Zhu, D., *et al.* Periostin: An Emerging Molecule With a Potential Role in Spinal
631 Degenerative Diseases. *Front Med (Lausanne)* **8**, 694800 (2021).
632 59. Furtwangler, T., Chan, S.C., Bahrenberg, G., Richards, P.J. & Gantenbein-Ritter, B.
633 Assessment of the matrix degenerative effects of MMP-3, ADAMTS-4, and HTRA1, injected
634 into a bovine intervertebral disc organ culture model. *Spine (Phila Pa 1976)* **38**, E1377-
635 1387 (2013).
636 60. Tam, V., *et al.* Histological and reference system for the analysis of mouse intervertebral
637 disc. *J Orthop Res* **36**, 233-243 (2018).
638 61. Lattouf, R., *et al.* Picrosirius red staining: a useful tool to appraise collagen networks in
639 normal and pathological tissues. *J Histochem Cytochem* **62**, 751-758 (2014).
640

641 **Acknowledgements**

642 This work was jontly supported by the General Research Fund (17121619) of the Research Grant
643 Council of Hong Kong and the Scientific Research Foundation of PEKING UNIVERSITY
644 SHENZHEN HOSPITAL (KYQD202100X). We thank Dr. Foon Lian Lim for preparing the human
645 tissue sections.

646 **Author contributions**

647 YS, ZS and YP contributed equally to this manuscript. YS contributed to the study design, data
648 collection and interpretation, producing figures and obtaining the fund. ZS, KKS and YS conducted
649 the bioinformatic analysis. YS, YP, QL and ML performed the experiments and helped in data
650 analysis. YS, ZS and VYL wrote the manuscript. JZ, YH, ZG, KMC, ZZ and XZ collected samples
651 and clinical data, and edited the manuscript. KMC, ZZ and XZ reviewed the data and provided
652 critical data interpretation. VYL conceptualized and supervised the whole project, provided supports
653 and coordinated the collaboration. All authors have read and approved the manuscript.

654 **Competing Interests statement**

655 The authors have no competing interest to declare.

656

Table 1 Demographic data. A total of 31 subjects from 4 studies were selected with degeneration graded I-V, and their single cell RNA sequence data were extracted and integrated for further analysis. Samples defined as Pfirrmann grade I were considered as Non-degenerative; grades II/III and IV/V as respectively Mild and Severe degeneration. F: Female; M: Male.

Data availability	Sample ID	Gender	Age	Pfirrmann grade	Diagnosis	Labeling in analysis	Ref
GSE160756	GSM4878538	M	16	I	Spine fracture	Non-degenerative	14
	GSM4878539	M	31	I	Brain death	Non-degenerative	
	GSM4878540	M	31	I	Brain death	Non-degenerative	
CNP0002664	Ctrl	M	37	I	Spinal cord injury	Non-degenerative	7
	NP2	F	65	V	Herniation	Severe	
	NP4	M	59	III	Herniation	Mild	
	NP8	M	50	IV	Herniation	Severe	
	NP9	M	56	III	Herniation	Mild	
	NP10	F	48	III	Herniation/spondylolisthesis	Mild	
GSE165722	S1	M	63	II	Burst fracture	Mild	12
	S2	M	41	II	Burst fracture	Mild	
	S3	F	56	III	Herniation	Mild	
	S4	F	65	III	Herniation	Mild	
	S5	F	64	IV	Herniation	Severe	
	S6	F	53	IV	Herniation	Severe	
	S7	M	54	V	Herniation	Severe	
	S8	M	56	V	Herniation	Severe	
PRJCA014236	D1_1	M	7	I	Scoliosis	Non-degenerative	11

	D1_2	F	6	I	Scoliosis	Non-degenerative
	D2_1	M	19	II	Herniation/spondylolisthesis	Mild
	D2_2	F	46	II	Herniation/spondylolisthesis	Mild
	D2_3	M	15	II	Herniation/spondylolisthesis	Mild
	D3_1	M	37	III	Herniation/spondylolisthesis	Mild
	D3_2	F	31	III	Herniation/spondylolisthesis	Mild
	D3_3	M	30	III	Herniation/spondylolisthesis	Mild
	D4_1	F	44	IV	Herniation/spondylolisthesis	Severe
	D4_2	F	43	IV	Herniation/spondylolisthesis	Severe
	D4_3	M	54	IV	Herniation/spondylolisthesis	Severe
	D5_1	M	68	V	Herniation/spondylolisthesis	Severe
	D5_2	F	50	V	Herniation/spondylolisthesis	Severe
	D5_3	F	57	V	Herniation/spondylolisthesis	Severe

666 **Table 2. Analysis of DEGs for disc fibrocytes.** *PTPRC*$^{+}$*COL1A1*$^{+}$ or *ITGAM*$^{+}$*COL1A1*$^{+}$ NP cells
667 were considered as disc fibrocytes, and subdivided to two major clusters based on their UMAP
668 distribution: 1) immunocytes-clustered fibrocytes containing fib-M (Macrophage), fib-N
669 (Neutrophil), fib-T (T cells) and fib-G (G-MDSC); 2) NP cells (NPC)-clustered fibrocytes containing
670 fib-fNP (FibroNP), fib-cNP (ChonNP), fib-rNP (RegNP). Positively and negatively expressed DEGs
671 of different fibrocytes subsets were demonstrated.

DEGs for immunocytes-clustered fibrocytes: fib-M; fib-N; fib-T; fib-G	
Commonly shared positive genes	*LAPTM5 PTPRC CXCR4 CFLAR HCLS1 PLEKHA2 FAM49B CELF2 SRGN ARHGDIB LCP1 EVI2B FMNL1 COL1A1 GMFG ITGB2 TMSB4X SAMSN1 MT-ND1 RAC2 B2M MT-ND3 SLA WDR26 FAM107B DEDD2 PFN1 MT-ND4 CFL1 FYB1 HLA-B LASP1 WIPF1 PPP1R18 GNB1 HLA-E HLA-A MSN MT-ATP8 MYH9 RAB8B ATP5E BAZ1A ARPC2 CAPZA1 MT-ND6 PABPC1*
Commonly shared negative genes	*GLT8D2 NCKAP1 NGEF FAP TMEM98 SOX5 DIXDC1 CD320 GSTM3 PDGFRL SERPINA1 SERTAD4-AS1 KDELR3 BARX1 TGM2 ZNF385D CFB FKBP7 ECM2 GPC6 CAV2 CHPF CDH19 AOPEP PERP LARP6 CNN3 CCDC85B TUBB2B MIR100HG MPP6 MAGI2-AS3 TIMP4 GFPT2 ARL3 MDFI SCIN SMOC1 VASN DPT LAMB2 CHST3 PPFIBP1 ACKR3 HSPA2 EVA1B GOLM1 METRN GNG11 WWTR1 SGCB ITGA10 MELTF PYURF THY1 FXYD1 LY6E EFEMP2 SNAI2 SSPN VGLL4 ADPRHL1 FHL2 DDR2 LTBP3 CAVIN3 RGS3 CILP2 ANGPTL2 PAX1 FSTL1 FGFRL1 P4HA2 FAM114A1 FAM3C MRC2 ANGPTL5 TMEM45A COX7A1 ANGPTL4 MXRA7 C1R H19 DANCR COL6A3 RHOBTB3 BAMBI TPM1 NFIA TRPS1 FKBP10 MXRA8 COL5A2 MEG3 CRIP2 TCEAL2 RCN3 COPZ2 IFITM3 CDO1 PDPN ID1 NRN1 TCF4 ID4 PDLIM4 SIVA1 ERRFI1 ADIRF ENPP1 PPIC EDIL3 FOXC1 TPM2 FGF2 NENF FERMT2 NOVA1 PCOLCE CHAD APOD ATP5IF1 RCN2 SLC39A14 EMP2 TMEM59 MAGED2 GPRC5A SBDS COL11A2 PRDX2 WDR83OS ISLR CALD1 AK1 CAVIN1 FHL1 COL6A1 BNIP3 TM4SF1 FXYD6 C1S*

	CLEC3A FBXO2 SOX9 TIMP3 RRBP1 CAV1 MYDGF SERPINA5 UAP1 PPDPF TNFRSF11B SPARC ITM2A FN1 CP EID1 LINC00632 RBP4 BEX3 MT1M KRT10 SERPING1 CPE VCAN VKORC1 PLOD2 PDIA6 CAPS COL6A2 SNHG5 KCNMA1 PMEPA1 MFGE8 COL9A3 SOD3 RABAC1 TNFRSF12A CCDC80 C12orf57 CCN1 DST HAPLN1 PRDX4 COL11A1 GPX3 PLAC9 COL2A1 PPIB FMOD HTRA1 CLU S100B NUPR1 TSC22D1 ZFAS1 ABI3BP SERPINA3 PCOLCE2 OGN S100A13 SERPINE2 PRELP IGFBP7 DSTN SSR4 AEBP1 C11orf96 MIF MIA CNMD SELENOM NNMT S100A1 BGN CCN2 SCRG1 MT1G RPS26 FGFBP2 SNORC ACAN ECRG4 LUM CRYAB DCN TIMP1 MGP NDUFA4L2 MT2A MT1X RPLP0 MT1E COMP RPL7
Fib-M specific	*CSTB LGMN ANXA2 DAB2 LGALS1 CTSL PLD3 CST3 NRP2 LMNA MGST3 PDE4DIP*
Fib-T specific	*RORA*
Fib-G specific	*NUCB2 DUT*
Fib-N specific	*N/A*
DEGs for NPC-clustered fibrocytes: fib-cNP; fib-fNP; fib-rNP	
Commonly shared positive genes	*COL1A1 COL3A1 COL1A2 ASPN TGFBI CTSB CTGF MT-ATP8 ATP5E ALDOA CDR1 HTRA1 ITGB5 MATR3 AC112491.1 ATP5G2 CYR61 ATP5J ATP5H HMOX1 NBEAL1 ATP5G3 ITGAM ATP5A1 ATP5F1 ATP5B ANKH LRRC75A ZNF90 TRAM1 ATP5L CERS2 DUSP5 PPP3CA PXDC1 CDK2AP1 OAT CKS2 ANGPTL4 SLC1A4 LDHA ZFAND5 MAN1A2*
Commonly shared negative genes	*RPL21 MALAT1*

678 **Table 3. Analysis of DEGs for NPC-clustered disc fibrocytes.** NP cells (NPC)-clustered fibrocytes
679 were subdivided based on their UMAP distribution: fib-cNP (ChonNP), fib-fNP (FibroNP) and fib-
680 rNP (RegNP). Featured genes of each subset that were negatively expressed by immunocytes-
681 clustered fibrocytes were summarized.

	Featured DEGs of NPC-derived fibrocytes
fib-cNP	*PLOD2 FGFBP2 SPARC FMOD HTRA1 TIMP3 ANGPTL2 DPT MELTF CCDC80 PCOLCE2 FHL2 OGN CILP2 GFPT2 SCIN PRELP SERPINA5 CPE HAPLN1 SCRG1 COL2A1 ANGPTL4 ABI3BP SOX9 COL5A2 FXYD6 P4HA2 SMOC1 FGF2 NRN1 MXRA7 CDO1 FSTL1 PPIC SERPINA1 ID4 TUBB2B TMEM45A LUM CDH19 FAM114A1*
fib-fNP	*COL6A3 COL5A2 COL6A1 DPT FSTL1 SPARC THY1 FAP COL6A2 ISLR CALD1 ANGPTL2 HTRA1 PCOLCE MXRA8 C1R CAV1 EDIL3 CAVIN1 CNN3 ECM2 TPM1 PPIC FKBP10 DDR2 PLOD2 GOLM1 SMOC1 SSPN FAM114A1 PDGFRL FN1 ERRFI1 TIMP3 FERMT2 PDPN GLT8D2 MXRA7 SLC39A14 BGN NCKAP1 CAVIN3 ANGPTL4 DST WWTR1 SGCB DIXDC1 P4HA2 C1S TCF4 RCN3 ACKR3 EFEMP2 IGFBP7*
fib-rNP	*DCN FAP NNMT FN1 MFGE8 GFPT2 CLU SERPINE2 SLC39A14 ANGPTL4 GPX3 ERRFI1 SOD3 C11orf96 CDO1 ACKR3 HTRA1*

682
683
684
685
686
687
688
689
690
691
692