Notochordal differentiation and integrative transcriptomic analysis using human pluripotent stem cells

Yuelin Zhang[1,2,a], Zhao Zhang[2,a], Peikai Chen[3,4,a], Chui Yan Ma[2], Cheng Li[2], Ron Wu[4], Tiffany Y. K. Au[4], Yan Peng[5], Victor Y. Leung[5], Keith Luk[6], Vivian Tam[4], Kenneth Man Chee Cheung[5,6], Danny Chan[4], Kathryn Song Eng Cheah[4], Qizhou Lian[1,2]*

[1]Guangzhou Women and Children's Medical Center, Guangzhou 510623, Guangdong, China;

[2]Department of Medicine, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong;

[3]Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong,;

[4]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong;

[5]Department of Orthopaedics and Traumatology, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong;

[6]Queen Mary Hospital, Hong Kong.

[a]These authors contributed equally to this study.

*__Address correspondence to:__
Qizhou Lian, MD, PhD,
Cardiology Division,
Department of Medicine,
The University of Hong Kong,
Hong Kong
Tel: +852-21899752
Fax: +852-28162095
Email:qzlian@hku.hk

Summary

Progressive loss of nucleus pulposus cells (NPCs) is associated with the onset of intervertebral disc degeneration (IDD). Transplantation of NPCs, derived from human pluripotent stem cells including hESC/iPSCs, may offer a novel therapy for IDD. To date, effective *in vitro* differentiations of notochordal and NP cells remained to be demonstrated. Towards this end, we developed a three-step protocol to directly differentiate hESC/iPSC towards mesodermal, then notochordal and finally NPCs. Our results showed that notochordal-like cells (NCCs) were successfully derived from the first two-steps of the protocol. Furthermore, these cells could be differentiated into NPCs. These NPCs expressed the tyrosine kinase receptor Tie2 (Tie2), disialoganglioside 2 (GD2), collagen II and aggrecan. Genome-wide transcriptomic analyses by sequencing (RNA-seq) revealed the expression of a wide array of known NP markers, extracellular matrix (ECM) genes, up-stream regulators and pathways. Cross-comparison of *in vitro* RNA-seq profiles with *in vivo* human NP data confirmed the *in vitro* NPCs are significantly more similar to *in vivo* NP than hESC/iPSCs. Transplantation of NPCs effectively attenuated disc injury in a rat model of IDD. We utilized CRISPR/Cas9 to seamlessly knock in an enhanced green fluorescent protein (eGFP) to the loci of the Noto gene in ESCs for NCC generation. Our study achieved effective notochordal differentiation and provided transcriptomic insights into the use of human ESC/iPSCs.

Keyword: human induced pluripotent stem cells, intervertebral disc degeneration, nucleus pulposus, notochord

Introduction

Intervertebral disc degeneration (IDD) is a major cause of low back pain, affecting the quality of life of over 60-80% of the global population (Kennon et al., 2018). Traditional treatments such as physiotherapy and anti-phlogistic drugs that aim to provide symptomatic relief are ineffective since they cannot recover the disc and may even exacerbate degeneration (Hanley et al., 2010; Mirza and Deyo, 2007). Surgical intervention is an option for severe IDD and may involve removal of the defective disc, fusion of the vertebral segments or implantation of an artificial disc made from synthetic polymers or metallic compounds. There remains nonetheless the long term problem of wear and failure of artificial discs after transplantation (Hanley et al., 2010). Biological treatments would provide an ideal solution but require far greater knowledge of the process of disc degeneration. There is a need to define a more effective therapeutic strategy of which rapid developments in regenerative medicine offer strong potential.

The intervertebral disc (IVD) comprises three interdependent tissues: the nucleus pulposus (NP), a central hydrated gelatinous core sandwiched between the cartilaginous end-plates (EP) of the adjacent vertebral bodies and surrounded by a tough ring-like fibrous laminar structure, and the annulus fibrosus (AF) (Humzah and Soames, 1988). Together they synthesize a great deal of extracellular matrices (ECMs) such as collagens, proteoglycans, and fibronectin/laminins, that maintain the structural integrity of the disc and its functional homeostasis (Roberts et al., 2006). The NP, which is derived from mid-mesoderm notochordal structures, plays a critical role in

maintaining homeostasis of the NP matrix. Accumulating evidence has demonstrated that loss of nucleus pulposus cells (NPCs) contributes to IDD (Wu et al., 2018; Zhu et al., 2018). Therefore, transplantation of NPCs may be a novel strategy for IDD therapy. Matrix or cell transplantation, including NP-like cells, promotes recovery of IDD (Leung et al., 2014; Perez-Cruet et al., 2018). Nonetheless it is not ethically feasible to extract NPCs from healthy individuals for IDD therapy. Exploring an alternative source of NPCs is of great importance.

Embryonic notochordal cells (NCCs) have been hypothesized to be the precursors of adult NP cells (Stemple, 2005). We and others have shown from lineage tracing in mice that NP cells are of notochordal origin, suggesting that NCCs are the progenitors of NP cells (McCann and Seguin, 2016). A notochord homeobox gene, Noto, has been found to be specifically expressed in notochordal development. Human pluripotent stem cells have been shown to be an important cell source for regenerative medicine. Ideally, we would obtain sufficient NPCs from cell differentiation. In the current study, we developed a three-step protocol to direct human embryonic stem cell/induced pluripotent stem cell (ESC/iPSC) differentiation towards mesodermal, subsequently notochordal lineages and finally NPCs. Due to the lack of surface markers for NCCs, we utilized CRISPR/Cas9 technology to seamlessly knock in an enhanced green fluorescent protein (eGFP) to the loci of the Noto gene in a human pluripotent stem cell line. This Noto-GFP reporter could effectively monitor temporal and spatial expression of the Noto gene, and provided a straightforward strategy for human NCC differentiation and purification.

Results

Early notochord markers are differentiated from ESC/iPSCs

Notochord is derived from chordamesoderm in early embryonic development that is regulated by the Activin/Nodal signaling pathway (Gritsman et al., 2000; Pauklin and Vallier, 2015). Mouse notochordal cells can be induced by a low concentration of Activin A or Nodal (Winzi et al., 2011). The expression of key transcription factors, such as Forkhead A2 (Foxa2), Noto and Brachyury (T), is essential to determine notochord differentiation (Risbud et al., 2010; Sivakamasundari and Lufkin, 2012). In human pluripotent stem cells, early mesodermal differentiation and node formation requires activation of the Wnt signaling pathway (ten Berge et al., 2008). Blockage of the BMP4 signaling pathway combined with Wnt signaling contributes to specification of the notochord (Yasuo and Lemaire, 2001). Furthermore, FGF2 is necessary to promote a mesodermal lineage and acts as an inducer of Xnot that is an ortholog of Noto (von Dassow et al., 1993). Retinoic acid (RA) at physiological levels inhibits notochord differentiation. Therefore, inhibition of RA is a strategy to improve NCC differentiation (Winzi et al., 2011). Based on this knowledge that key signaling pathways determine chordamesodermal lineage, we developed a compounds-based protocol to achieve directed lineage differentiation of human ESC/iPSCs to NP progenitors. The protocol for NCC differentiation was is in Figure 1A. It has been reported that withdrawal of pluripotency conditions for 48 hours can make ESC/IPSCs more sensitive to differentiation signals (Thomson et al., 2011). We used the N2B27 medium with Active A for 48 hours and then used Activin A, CHIR (CHIR99021, Wnt3a activator), Noggin (BMP-4 antagonist), AGN193109 (RA inhibitor) and FGF2 for 2-4 days. We first evaluated mRNA expression in monolayer

culture at an early stage of notochord differentiation. The results of RT-PCR showed that the expression level of Noto was gradually increased and achieved the highest level at day 5 after differentiation in ESC3, ESC9 and IMR-90-iPSCs. Similarly, the expression level of mesoderm markers T and Foxa2, which act upstream of Noto, also exhibited the highest level on day 5 (Figure 1B-i, ii, iii). The expression of Noto, T and Foxa2 was also examined by immunofluorescence staining on day 5 after differentiation in ESC3, ESC9 and IMR-90-iPSCs. The results showed that some Noto positive cells were expressing T and Foxa2 (Figure 1C-i, ii, iii).

Differentiation of NP-like cells from ESC/iPSCs

Since NCCs are thought be the progenitors of NP cells and the TGF-β signaling pathway is essential for NP cell functional maintenance (Tang et al., 2018; Zhou et al., 2015), we continued NPC differentiation using TGF-β3 for an additional 15 days in monolayer culture. The protocol for NP-like cell differentiation is shown in Figure 2A. During differentiation, some cells developed a vacuole-like structure, the major feature of NPCs (Figure 2B-i). It has been reported that the tyrosine kinase receptor Tie2 (Tie2) and disialoganglioside 2 (GD2) are surface markers of NPCs (Sakai et al., 2012). After differentiation, immunohistochemistry analysis showed that some cells were positive for Tie2 and GD2 (Figure 2B-ii). These differentiated cells also expressed Aggrecan and Collagen II that are major makers of NPCs (Figure 2B-ii). Next, we sorted the Tie2/GD2 positive cells using FACS. The results of FACS demonstrated that around 25~30% of the differentiated cells were Tie2/GD2 positive (Figure 2C-i). Subsequently, these Tie2/GD2 positive cells were cultured with NPC medium. As shown in Figure 2C-ii, there were many vacuoles in these sorted cells

(Figure 2C-ii). In addition, immunofluorescence staining showed that these sorted cells were positive for Tie2, GD2, T, Aggrecan and Collagen II but negative for Noto (Figure 2C-iii). Next, we focused on characterizing these NP-like cells. We first cultured these cells in methylcellulose medium. After culturing for 10 days, two types of colonies, colony-forming units-fibroblastic (CFU-F) and colony-forming units-spherical (CFU-S), were formed (Figure 2D-i). Immunofluorescence staining demonstrated that Collagen II and Aggrecan were highly expressed in CFU-S but not in CFU-F (Figure 2D-i). Furthermore, compared with another sub-population of differentiated cells, Tie2/GD2 double positive cells formed more CFU-S (Figure 2D-ii). It has been reported that NPCs can release sonic hedgehog (Shh) (Choi et al., 2012). Therefore, we collected the conditioned medium (CdM) of NP-like cells and co-cultured it with a Light2 cell line, a luciferase-based reporter responsive to sonic hedgehog (Shh) secreted protein. Compared with DMEM, the CdM of NP-like cells greatly enhanced the luciferase activity of Light2 cells in a dose dependent manner, suggesting that ESC/iPSC-derived NP-like cells could release Shh (Figure 2F).

Genome-wide transcriptomic analysis reveals differentiated cells have strong characteristics of the nucleus pulposus

To assess the genome-wide behaviors of the differentiated products, we performed bulk RNA sequencing on the before- (n=3) and after-differentiation samples (n=12) (Methods). Many important genes previously reported to be defining young adult NP [PMID: 25411088] or involved in NP development [PMID: 30902259] show strong up-regulation patterns in the differentiations (Figure 3A; Figures S1-3). Hierarchical

clustering of these 15 RNA profiles revealed the formation of two distinctive clusters: one that comprised the three starting lines, and the other comprising the 12 differentiated samples (Figure 3B). The intra-cluster replicates were 91-97% similar, with only 50-60% similarity between before and after differentiations, suggesting that the lines followed the same differentiation path regardless of their origins. Stem cell markers Oct-4 (POU5F1), Sox2, and Nanog were significantly lowered in the differentiated cells (Figure 3A; Figure S4), indicating that differentiation had proceeded considerably. Downstream targets of the TGF-β pathway (Ito and Miyazono, 2003; Meng et al., 2016), including RUNX1/2, ID1/2/3/4, PITX2 and P15 (CDKN2B), had all been activated in the differentiation products (Figure S5), confirming the efficacy of our NPC protocol.

Secretion of ECMs is a pivotal role of normal human NPCs, and as such many efforts to establish cell markers for NPC have focused on ECM genes (the 'matrisome') [PMID: 25411088; PMID: 20722018; PMID: 24049099]. Overall, 82 of the 275 core matrisome genes (Figures S6-7), and another 148 (Figures S8-9) of the 753 non-core matrisome genes were significantly up-regulated (FDR<0.05) in our NPC differentiations. In particular, these included 14 collagen genes: COL1A1/2, COL2A1, COL3A1, COL5A1/2/3 and COL6A3. Apart from collagen genes, the other two categories of the matrisome, proteoglycans and glycoproteins, were also abundantly up-regulated.

Proteoglycans are essential for the hydro-homeostasis of the inter-vertebral disc (Iatridis et al., 2007) and include three dozen or so genes, many of which were

proposed markers of NP, and were significantly up-regulated in the differentiations, including DCN (decorin), SPOCK1/2/3 (testican), LUM (lumican), BGN (biglycan), OGN (osteoglycin), alpha-2-macroglobulin (A2M), and FMOD (fibromodulin).

Glycoproteins encompass a diverse list of 200 genes that fulfill functions of ECM assembly, cell-adhesions and communications. Many of them were differentially expressed genes (DEGs) that were up-regulated in the differentiations, including laminins (LAMA4/LAMB4), emilins (EMILIN2), periostin (POSTN), thrombospondins (THBS1), cartilage oligomeric matrix protein (COMP), osteopontin (SPP1), fibronectin (FN1), tenascins (TNC), reelin (RELN), and SPARC. Some of the up-regulated glycoproteins are specific to bone and cartilage, including MGP and CILP. The up-regulated genes also include some modifiers of ECMs, including TIMP2/3 and MMP1/2/10/28, and cell-surface integrins interacting with ECMs, including ITGA1/4/8/11, ITGB3, and NCAM1/VCAM1. All up-regulated genes are summarized in supplementary file.

A panel of reported NP markers showed strong up-regulation in the differentiations (Figure 3A; Figure S1-3). Carbonic anhydrase III (CA3) is a highly specific marker for NP (Silagi et al., 2018) that increased 38.6 fold (FDR 1.12×10^{-6}) after differentiation. Galectin-3 (LGALS3) is a marker for young adult NP samples (Silagi et al., 2018), and increased 5.5 fold (FDR: 0.036). Another NP marker, FOXF1, was not statistically significant, but increased in all differentiated products, by an average 8.2 fold. Many bone, skeletal and cartilage genes were also highly expressed. CHST3, which encodes chondroitin 6-O-sulfotransferase 1 and is a replicated risk factor for

lower-back pain (Maxim et al., 2018), is essential for regulating the proteoglycans by means of sulfation. Chondrocyte markers such as SOX9 and COL2A1 are also significantly up-regulated. Other markers of NP, such as CDH2 (N-cadherin) (Lv et al., 2014) and AQP1 (Richardson et al., 2008) are also up- regulated genes.

Apart from the prominent ECM activities, a plethora of upstream regulators, pathways, and transcription factors were activated. A super Venn diagram (Figure S10) visualizes the significantly enriched pathways, from a compendium of sources, including KEGG, Biocarta and Reactome. A cluster of enriched terms including matrisome, focal adhesion, integrin, Wnt (DKK1/2, FZD1, LEF1, NKD1/2, and 15 WNT-family genes), hedgehog (GLI3), angiogenesis and myogenesis can be observed in the top items, and further indicates cellular differentiation towards targeted cell-type. The expression of BMP inhibitors NOG (noggin), GREM1 (Gremlin1) and CHRD (chordin) that are known to be markers of NP cells (Chan et al., 2015) was also significantly increased in the differentiations. Other bone development genes that are up-regulated include runt-related transcription factor 2 (RUNX2), BMP1/4/5/7, TGFBI/2/3, LGR5, GDF5, and SOX5/6 (Smits et al., 2004). Sox5 and Sox6 are transcription factors that are required for ECM sheath formation and cell survival in the notochord, and development of the NP in IVDs (Smits and Lefebvre, 2003). TGF-β1 and GDF5 drive the differentiation of human adipose stromal cells towards NP-like cells (Colombier et al., 2016).

Cross-comparisons of transcriptome profiles with human *in vivo* NP data

To further verify the identity of our *in vitro* cells, we used two human *in vivo* NP datasets for cross comparisons: 1) an in-house set of bulk RNAseq profiles from four NP of three adolescent or young individuals (aged 13/14/33) undergoing discectomy due to scoliosis or burst fracture; and 2) a public microarray data of three 'healthy' human NP in older adults (46-57 years old) (Minogue et al., 2010). In both datasets, genome-wide similarity of our differentiations to *in vivo* NP increased significantly and by a margin of ~5%. Particularly, similarity with the adolescent or young NP increases from 72% to 77% after differentiations (Student's *t*-testing $p=1.29\times10^{-07}$; Figure 3C, Figures S11-12). By employing a statistical approach outlined in the methods, we were able to detect influential genes that significantly contributed to the increase of similarities (Figure 3D). We note that that these influential genes can work either by having lower (the negative influential genes) or higher (the positive influential genes) levels in the *in vitro* differentiated NPCs. But in both cases, they help by having closer values to the *in vivo* levels (moving towards the diagonal line). Among the negative influential genes is SOX2 which is an indication of loss of stemness; and the fact that it is negative influential means it is also low in *in vivo* NP. Among the positive influential genes are a huge number of NP relevant genes discussed in the previous section. Many of them rise from almost zero to ultra-high levels to match with their levels in *in vivo* NP, particularly several proteoglycans (DCN, LUM, and BGN) and some well-known NP markers. This result suggests that the increase in similarities to *in vivo* NP is intrinsically due to the existence of sizeable NP-like characteristics in our *in vitro* differentiations.

We also calculated the DEGs between *in vivo* NP and hESC/iPSCs, and those between *in vitro* NPCs and hESC/iPSCs, and their overlaps with the positive influential genes (Figure 3E). The three sets have a strong overlap (N=148; odds ratio=32.8) and any two sets are in significant overlap (supplementary methods). The 148 genes were highly enriched for matrisome, glycoproteins, proteoglycans, integrins and other regulatory terms (Figure 3F). Figure 3G visualizes the expressions of these 148 genes, many of which are known NP markers discussed earlier.

Transplantation of ESC/iPSC-derived NP-like cells attenuates IDD in rats

Cell therapy is leading the field of regenerative medicine for many tissues and organs, including IDD. The development of effective NPC-based treatment for IDD requires much greater understanding to be derived first from animal models. We used a rat model of needle punch-induced disc injury as reported previously (Leung et al., 2014). Radiographic results revealed that compared with the non-injured disc (L5-6), intervertebral disc height index (DHI) was dramatically decreased in the injured discs (L4-5, L6-7, L7-8) two weeks after punch injury, suggesting that a model of IDD model was successfully established (Figure 4A, B). Furthermore, %DHI of the injured disc without cell transplantation showed no obvious recovery at 10 weeks after injury (Figure 4B). Nonetheless %DHI of the damaged disc was significantly enhanced four weeks after ESC/iPSC-NP-like cell transplantation and persisted at 8 weeks (Figure 4B). Importantly, ESC-NP-like cells were superior to iPSC-NP-like cells in attenuation of IDD (Figure 4B). FAST staining showed that compared with the control disc, the nucleus pulposus tissue had disappeared in the injured disc (Figure

4C). Nevertheless transplantation of ESC/iPSC-NP-like cells resulted in significant recovery of the injured nucleus pulposus tissue (Figure 4C). These results demonstrate that transplantation of ESC/iPSC-NP-like cells attenuated needle puncture-induced IDD in rats.

Cell survival and recovery of IDD

To examine the cell survival post transplantation, ESC/iPSC-NP like cells were labeled with GFP *in vitro*. As shown in Figure 4A, GFP positive cells were detected in the NP at 2, 4 and 8 weeks post transplantation (Figure 5A-i). Nonetheless cell survival density was gradually reduced post transplantation (Figure 5A-ii). To further examine the therapeutic effects of ESC/iPSC-NP-like cells on IDD, we analysed the presence of Collagen II and Aggrecan among the different groups. Compared with the control group, the expression of Collagen II and Aggrecan as determined by immune-fluorescence staining was significantly reduced, indicating injury of the intervertebral disc (Figure 5B-i, ii). Nevertheless ESC/iPSC-NP-like cells greatly improved Collagen II and Aggrecan deposition in the interterritorial matrix of the NP (Figure 5B-i, ii).

Establishing Noto-GFP reporter in hES cells

Since there is no specific surface marker for Noto cells, we generated a Noto reporter in a human ESC line (hESC9). A classic homologous recombination strategy was used to precisely knock in an eGFP coding nucleotide into the loci of the Noto gene (Figure 6A). The binding site of gRNA was designed to be close to the terminal codon

of the Noto gene. Two Noto arms were used in this strategy; the eGFP was fused to the left arm using T2A as the adapter. To easily confirm that eGFP correctly integrated hESC9 colonies, a puromycin selection cassette was utilized that could be removed by activating the LoxP elements after transient introduction of Cre protein. Two pairs of primers were designed to characterize the real positive targeted colonies (Table S1). Among them, primer 1 was located at the flanking sequence of the homologous arms to cover the whole donor. To acquire better cutting efficiency in ESC9, two gRNAs were tested by SURVEYOR assay, with the gRNA1 presenting around 45% cutting efficiency (Figure 6B). After puromycin selection, hundreds of colonies were recovered. Twelve colonies were selected for PCR characterization using Primer 1 (Table S1), and nine colonies were eventually precisely integrated with GFP-LoxP-Puromycin-LoxP (Figure 6C). To remove the selection cassette, Cre protein was transiently induced. Finally, we successfully obtained four Noto-GFP reporter ESC9 cell lines after PCR characterization using Primer 2 (Table S1) (Figure 6D). Sequencing results showed that eGFP was precisely and seamlessly fused to the Noto gene using T2A as the adapter; no frame-shift, nonsense or deletion mutation was detected (Figure 6E).

Derivation and Isolation of NCCs from an hESC line with Noto-GFP reporter

The Noto-GFP reporter cell line was directly cultured for two days on N2B27 basal medium containing Activin A (10ng/mL) followed by another three days of differentiation with addition of the same cell factors (Figure 7A). The GFP positive cells were also observed on the fifth day (Figure 7B). Immune staining revealed that

the GFP could overlap perfectly with Noto at the cell level (Figure 7B). Notably, real-time PCR results demonstrated higher expression of Noto, Foxa2 and T, indicating higher efficiency for NCC differentiation using a monolayer strategy (Figure 7C). To further test our Noto-GFP reporter cell line, we digested differentiating pluripotent stem cells into single cells and sorted the Noto-GFP positive cells by flow cytometry using the undifferentiated reporter cell line as a negative control. Noto-GFP cells were very sensitive to dissociation challenge and quickly died during dissociation from differentiating hES clusters. Although most Noto-GFP cells could not survive after dissociation from differentiating human pluripotent stem cells, sporadic living GFP-positive cells were acquired after sorting (around 0.1%). The sorted cells demonstrated more than 100 times higher expression of the Noto gene compared with the undifferentiated cells (Figure 7D). Further efforts are being made to define the optimum conditions under which to maintain Noto-GFP cells in culture.

Discussion

We developed a strategy to efficiently induce hESC/iPSCs into functional notochordal-like cells and NP-like cells. We have also provided transcriptomic landscapes to understand the molecular basis of notochordal differentiation**.** There are several major findings in the current study. First, NP-like cells and Noto-GFP NCCs can be successfully differentiated from human pluripotent stem cells with a compounds-defined protocol based on the knowledge that key signaling pathways determine a chordamesodermal lineage during embryonic development. Second,

whole transcriptomics analysis and cross-comparisons with public NP data reveal a panel of common and distinct molecular signatures. These molecular signatures could potentially be developed as general identity tags to define NP cells, and more specific markers to define features of different stage NP cells. Third, human ESC/iPSC-derived NP-like cells can functionally attenuate IDD in rats after transplantation. Fourth, we have successfully established a Noto-GFP reporter in a human ES cell line using CRISPR/Cas9 technology, allowing us to further define conditions for NCC differentiation and maintenance *in vitro*. Lastly, despite their poor viability, notochord-like cells can be purified after monolayer differentiation using Noto-GFP-hES cell reporter cells.

IDD is a global condition that is strongly associated with low back pain. It is a complex disorder of ageing compounded by environmental and genetic factors. Current treatment options are limited and sub-optimal. Surgical interventions, including discectomy or implantation of an artificial disc, can alleviate the symptoms but are prohibitively expensive, complex and also prone to failure. The cellular composition of the adult nucleus pulpous varies depending on the condition of the disc and an individual's age. Although the underlying pathogenesis is not fully understood, it has been proposed that exhaustion of NCCs plays an essential role in the etiology of IDD. Genetic lineage mapping studies in mice indicate that progressive loss of NCCs and replacement with hypertrophic chondrocyte-like nucleus pulposus cells (NPCs) are associated with the onset of IDD. Cell-based therapy has emerged as a novel strategy in regenerative medicine for many conditions,

including IDD. Unfortunately, this approach remains in its infancy due to many challenges. One major challenge is the limited availability of NCCs or NPCs. Therapeutically useful progenitor cells generated from pluripotent stem cells offer great potential to treat presently incurable human diseases. Indeed some stem cell-based treatments have achieved encouraging results in certain specific diseases. Several studies have also highlighted the strong potential of NP-like cells in regeneration of intervertebral discs (Perez-Cruet et al., 2018; Zhou et al., 2018). Nevertheless numerous ethical issues and technical limitations currently prevent us from obtaining functional cells from biopsy. Therefore, differentiating NP-like cells from human pluripotent stem cells is vital for basic and clinical research. Since NCCs are the progenitors of NP cells, differentiating NCCs was the first step in the current study. We cultured the human ESC/iPSC cell line in basal medium with Activin A (10ng/mL) for two days to stimulate mesoderm differentiation. Takada *et al.* found that Wnt3a-/- mice presented disrupted notochord development (Takada et al., 1994); and Mina reported that CHIR99021 could promote mesoderm differentiation of hESCs (Gouti et al., 2014). We replaced the DKK1 reported in the mouse NCC differentiation protocol with CHIR to continue cell differentiation. The results of immunostaining and real time PCR showed that transcription factors including Foxa2, T and Noto, which are essential for notochordal development, were greatly enhanced during the differentiation. Furthermore, Foxa2 and T are co-expressed in notochordal-like cells and have been shown to be required for notochord development in ascidians, frogs and mice, suggesting that Foxa2 and T are functionally conserved

in the specification and maintenance of NCCs (Risbud and Shapiro, 2011). Next, we continued this protocol for NPC differentiation using TGF-β3 for 15 days. Using this multistage protocol, ESC/iPSCs were differentiated to an NP-like phenotype, demonstrating vacuolated cell morphology and expression of NP markers including GD2 and Tie2 (Sakai et al., 2012). We also showed that ESC/iPSC-derived NP-like cells can release Shh. Subsequently, we examined the therapeutic effects of ESC/iPSC-derived NP-like cells in a rat model of IDD. Accumulating evidence has shown that administration of MSCs ameliorates the degenerated disc by promoting expression of Collagen II and Aggrecan and inhibits the apoptosis of NPCs (Sakai et al., 2006; Yang et al., 2008). Moreover, transplantation of dedifferentiated fat cells has been shown to promote regeneration of an intervertebral disc by differentiation into NP-like cells (Nakayama et al., 2017). Theoretically, transplantation of ESC/iPSC-derived NP-like cells is superior to other types of cells in attenuation of intervertebral disc degeneration. Our results show that transplantation of ESC/iPSC-derived NP-like cells functionally attenuated intervertebral disc degeneration by improving expression of Collagen II and Aggrecan in the injured disc.

Transcriptomic analyses by RNA-sequencing revealed a large number of NP markers and regulators that were strongly up-regulated in our differentiations. Despite the scarcity of healthy NP samples, we managed to collect two *in vivo* datasets as control NP for cross-comparisons, showing marked increase in similarity to *in vivo* NP after differentiations, to which major contributors were also identified and they show

strong relevance to the known biology of NP. We notice that even compared with adolescent or young *in vivo* NP, the genome-wide similarity only increases from 72% to 77%, suggesting sizeable heterogeneity in both *in vivo* and *in vitro* samples. Another reason we are not seeing higher similarities might be that the reference *in vivo* NP was taken from individuals with scoliosis conditions or burst fracture, whose transcriptomes by themselves may have deviated from the healthy state to a certain degree. Nonetheless the proportion of our cells with strong NP characteristics is projected to have been reasonably high in order to bring out those DEGs and increases in similarities to *in vivo* NP. An ideal and popular tactic may be to perform single-cell transcriptomics of our cells to further delineate their heterogeneity and assess the degree of yield but that is beyond the scope of this report. We are also aware that Tie2/GD2 expression levels in our differentiations were not particularly high. A possible scenario is that they existed in niche cell-populations with stronger progenitor-like capabilities that can be assessed in the near future with single-cell RNA-seq.

Although the Noto gene has been reported to be a tissue-specific marker for NCCs, there remains a substantial amount of work to verify its differentiating status and difficulties in sorting cells since Noto is expressed only transiently in the nucleus. Recently, Winzi et al obtained mouse NCCs from *in vitro* cell differentiation with the support of Noto-GFP reporter mouse ESCs (Winzi et al., 2011). The co-expression of GFP and Noto represented the real-time differentiating status of NCCs, and effectively reduced the workload of establishing the protocol. Moreover,

differentiated NCCs could be purified to define the surface marker using high throughput techniques. We successfully generated a Noto-GFP reporter cell line in hESC H9 to support human NCC differentiation. CRISPR/Cas9 has been widely utilized to achieve genome manipulation in all kinds of species; nonetheless its genome-editing efficiency remains quite low in human PSPs (Zhang et al., 2017). To seamlessly knock-in GFP in hESCs, we used the classic homologous recombination strategy mediated by Cre-LoxP system and optimized the cutting efficiency of gRNAs. Then, a GFP was seamlessly fused to the Noto using T2A as the adapter that contained the left homologous arm (1599bp), selection cassette (LoxP-Puro-LoxP) and the right homologous arm (1561bp). Finally, we achieved GFP Knock-in in ESC9 cells with a high efficiency of 66%. After removing the selection cassette, Sanger sequencing indicated that GFP was precisely and seamlessly knocked in to the ESC9 cells without any mutation. Next, we used this Noto-GFP-reporter ESC9 for NCC differentiation. With the exception of Foxa2 and T expression, the overlap between Noto and GFP at the protein and mRNA level also demonstrated the capacity of this Noto-GFP reporter cell line for NCC differentiation. Since the differentiated cells were extremely fragile and sensitive, we sorted only the targeted cells with an efficiency of around 0.1%. Nonetheless we obtained the GFP positive NCCs using flow cytometry as the Noto expression was more than 100 times higher than that of the control cells.

Several limitations in the current study need to be noted. First, although NP-like cells could be derived from human ESC/iPSCs, some genes of NP cells were not

expressed, thus the current protocol needs to be optimized. Second, although we could derive NCCs from human ESC/iPSCs, their survival and maintenance was not achieved in cell culture conditions. Third, although we found no teratogenic effects of ESC/iPSC-NP-like cells in animal study, their genomic stability needs to be evaluated carefully.

In summary, our study shows that NP-like cells can be effectively and efficiently derived from human pluripotent stem cells via a specific differentiation protocol. Transplantation of ESC/iPSC-NP-like cells can relieve IDD. Therefore, ESC/iPSC-NP-like cells provide a novel resource for the understanding of IDD and for the development of an NPC-based treatment for IDD.

Star★Methods

Cell culture and differentiation

hESC3/ESC9 and IMR90-iPSC were cultured on a six-well culture plate (Corning Cellgro) coated with Geltrex (Thermo Fisher). Essential 8 medium (Thermo Fisher) was changed daily to sustain cell growth. Y-27632 (Sigma) was added after cell passage, thawing and nucleofection (Lonza). For NC-like cell differentiation, ESC3/ESC9/IMR90-iPSC were passaged and seeded on a 0.1% Geltrex-coated 6-well plate at a density of 50,000 cells/cm^2 with Essential 8 medium. One day later, Essential 8 medium was replaced by DMEM/F12 medium supplemented with N2, B27, ITS, 0.1mM NEAA, 2 mM L-glutamine, and 0.1 mM 2-mercaptoethanol (Thermo Fisher) and Activin A (10ng/mL) for two days at step1. Then, other cell

growth factors containing Activin A (10ng/mL), FGF2 (10ng/mL), Noggin (50ng/mL), CHIR (3μM) and AGN193109 (10μM) were added for a further three days differentiation (Cell growth factor from PeproTech, chemicals from Sigma) at step 2. For NP-like cell differentiation, the medium at step 2 was changed to DMEM-HG supplemented with 1% ITS+, 1% NEAA, 1% penicillin/streptomycin, 50μg/ml ascorbic acid-2-phosphate (Sigma), 40μg/ml L-proline (Sigma), 10nM Dexamethasone and 10ng/ml TGF-β3 (R&D) for 15 days at step3.

Immune staining

The differentiated cells were fixed in 4% PFA (Merck Millipore) for 30 minutes. The slides were blocked in 1% horse serum (Thermo Fisher) for one hour after rehydration, and then incubated at 4℃overnight with the primary antibodies: Noto, T, Foxa2, GD2, Tie2, Collagen II, Aggrecan (Santa Cruz Biotechnology). After washing with PBS three times, the slides were incubated with the secondary antibodies in the dark for one hour at room temperature. Finally, cells were mounted with DAPI and photographed using a confocal microscope LSM 700 (Carl Zeiss).

Quantitative RT-PCR

To examine the mRNA expression of Noto, Foxa2 and T, all cells at day 0, 4, 5 and 6 after differentiation were collected to extract the total RNA using an RNeasy mini kit (Qiagen). The cDNA was then synthesized using a Prime ScriptRT reagent kit with gDNA eraser (Clontech) for real-time PCR detection (Thermo Fisher). The primer sequences are listed in Supplementary Table 1.

Flow cytometry

The NP-like cells after differentiation were digested to single cells using Accutase after washing with PBS, and then stained with GD2 and Tie2 antibodies followed by incubation with fluorescently labeled secondary antibodies. Finally, cells were re-suspended in cool PBS containing 0.5% BSA (Sigma) at a concentration of $2x10^6$/ml. Appropriate isotype controls were used. Cells were sorted using a FACSAria II cell sorter system (BD) to determine the percentage of cells positive for GD2 and Tie2 (% cells).

Human *in vivo* samples

A set of four human *in vivo* NP (from three individuals) were used as reference for the target cell types. The three individuals were all females and aged 13, 14 and 33, respectively. The first two were diagnosed with scoliosis, and the third one (33yo) had burst fracture at L1. All three underwent discectomy (two levels at T12/L1 and L1/2, one level at L2/3, and one level at L1/2, respectively) between June 2012 and May 2014 at the Queen Mary Hospital of Hong Kong. Informed consents and Institution Review Board (IRB) approval were obtained.

RNA sequencing and processing

Bulk transcriptomes of both before (ESC3/ESC9/IMR90-iPSC), *in vitro* differentiated cells (four replicates for each differentiation line, twelve samples in total), and four human NP were performed at the Centre for Genomic Sciences (CGS) of the University of Hong Kong.

To assess the degree of heterogeneity after differentiations, four biological replicates per line were sequenced. In all, cDNA libraries of 19 input samples (~100 ng each) were prepared by the KAPA Stranded mRNA-Seq Kit (KR0960-v3.15) according to the manufacturer's instructions. They were then sequenced on the Illumina HiSeq-1500 platform, in five flow-cell lanes, and in paired-end formats (2×101bp). Over 90% of the bases achieved a quality score of Q30 (error calling <1/1000). On average, approximately half a million read-pairs were obtained per sample. The reads were mapped against the human reference genome (GRCh38) using the HISAT2 aligner (Kim et al., 2015) with default parameters. The average mapping rate was 93.6% (min. 90.6%, max. 96.6%) and ~80% of the reads were mapped to unique genomic locations (Figure S13A-B). The resultant bam-files were quantified using both Cufflinks (Trapnell et al., 2012) and HTSeq-count (Anders et al., 2015) with gene annotation files provided by GENCODE (version 25) (Harrow et al., 2012). On average, 14,546 and 27,859 genes per sample with FPKM values greater than 1 and 0, respectively, were recorded.

Analyses of differentially expressed genes (DEGs)

The raw read counts generated by HTSeq-count were further analysed by DESeq2 (Anders and Huber, 2010) to detect DEGs before and after the differentiation. In all, 1691 genes with log2 (fold-change) >2 and FDR q-value <0.05 were detected for genes specifically up-regulated in the *in vitro* differentiated NPCs (Figure S14). We also computed the DEGs among *in vitro* NPCs of different lines, and found that only 12.3% of the DEGs in *in vitro* NPCs varied across the lines (Figure S15A). The

majority were line-independent. There were several hundred DEGs across the lines, mostly due to original line differences before differentiation (Figure S15B-D). This shows that our protocol for NPC differentiation is robust against different starting cell-lines. The DEGs were further analysed with Geneset Enrichment Analyses (GSEA) tools (Subramanian et al., 2005). To overcome the strong overlap among enriched GO terms and pathways, a super Venn diagram was used to visualize their pairwise similarities (Fatemieh, 2006) and subsequent dimensional reduction by t-SNE (Maaten and Hinton, 2008).

Cross-comparison and detection of influential NP-specific markers

First, we identified a reference dataset recognized to be of target tissue/cell-type (in our case, the nucleus pulposus). For a sample in our differentiations, its expression profile is denoted by the vector $\boldsymbol{x} \in R^N$, where N is the number of genes and R denotes real number. A profile in the reference data is denoted as $\boldsymbol{y} \in R^N$. Let r^2 be the coefficient of determination for the linear model $\mathrm{lm}(y{\sim}x)$, then r is also the Pearson Correlation Coefficient (PCC). One gene $i \in [1, N]$ at a time was left out and the coefficient of determination calculated and denoted as $r_{(i)}^2$, and the difference denoted as $\Delta r_{(i)}^2 = r_{(i)}^2 - r^2$. A negative $\Delta r_{(i)}^2$ suggested that removal of gene i decreased the goodness-of-fit. Thus, the more negative $\Delta r_{(i)}^2$ was, the more contribution gene i makes in establishing the similarity between our differentiations and the reference sample. A cut-off was chosen by fitting $\{\Delta r_{(i)}^2 | i = 1,2, \dots, N\}$ on a Gaussian model, and choosing genes falling in the $P < 0.05$ region of the model.

Since there were multiple samples in our data in the public data, the values for $\Delta r_{(i)}^2$ became $\Delta r_{(i)}^2(j,k)$ for samples j in our data and k in the public one. Our data also consisted of before and after differentiations data (n=3 and 12, respectively). We performed *t*-testing of $\Delta r_{(i)}^2(j,k)$ for before and after differentiations, by linking it as a series of observations.

Data availability

All genomics data reported in this study are deposited at the NIH Gene Expression Omnibus (GEO) databases (GSE122429, reviewer's token: adqxysegfpkxzir). An interactive web interface for the data is available at: www.sbms.hku.hk/scopes/NPC.

[The CDS data is being uploaded. Will add the GSE number after I obtained it.]

Colony-forming assay

To evaluate colony formation, a single-cell suspension of $1.0x10^3$ NP-like cells was plated on 35-mm diameter dishes and cultured in 1ml of MethoCult H4230 methylcellulose medium (Stem Cell Technologies) for 10 days. Colonies were counted using an inverted microscope.

Vector construction

For donor construction, the homologous arms, the eGFP and the LoxP-Puro-LoxP selection cassette were amplified separately (Primer see Table S1). Then, they were transformed into competent cells after ligation with pBluescript II KS(-) plasmid using a Gibson kit (NEB). gRNA vectors were designed as described online

(http://crispr.mit.edu/) and constructed into MLM3636 (Addgene, 43860) in accordance with its supplemental cloning protocol. A surveyor mutation detection kit (IDT) was used to evaluate the cutting efficiency and off-targets of gRNAs (Primers see Table S2). Plasmids of hCas9 (Addgene, 41815) were utilized for genome editing.

GFP knock-in

4D nucleofector (Lonza) was utilized to achieve the highest transfection efficiency in H9. Briefly, ESC9 was digested into single cells by accutase (Thermo Fisher) and mixed with the vectors containing gRNA, Cas9 and donor for nucleofection. Then, cells were seeded onto the plate coated with Geltrex and cultured in essential 8 medium. 0.5μg/mL of puromycin (Thermo Fisher) was added to screen the integrated colonies. They were picked for PCR characterization and Sanger sequencing (BGI) using primer 1 (Table S1) when the colonies became visible to the naked eye. Finally, AAV-mediated Cre recombinase (Clontech) was utilized to remove the selection cassette and the targeted colonies characterized using Primer 2 (Table S1) and Sanger sequencing. All Noto-GFP reporter cell lines were evaluated by off-target testing (Primers in Table S2) and three germ layer differentiations.

Rat model of IDD and cell transplantation

All animal experiments were performed in the laboratory according to the Guide for the Care and Use of Laboratory Animals in the University of Hong Kong. The IDD rat model was induced as previously described (Wang et al., 2018). Briefly, after inhalational anesthesia with isoflurane, the tail discs were exposed by a longitudinal

incision and a 21-gauge needle punctured 1.5mm into the disc at the level of caudal intervertebral disc (L4-5, L6-7, L7-8). The intervertebral disc L5-6 remained unpunctured and served as the control. Two weeks after injury, GFP-labeled ESC/iPSC-NP like cells ($5x10^4$/10 μl PBS) were injected into the injured intervertebral disc (L6-7, L7-8). Meanwhile, 10 μl PBS was injected into L4-5 as the IDD group.

Radiographic evaluation

Radiographic assessment of IDD in rats was performed as described previously (Nakayama et al., 2017). Briefly, under inhalational anesthesia, the radiographic images of intervertebral disc levels of rats including L4-5, L5-6, L6-7 and L7-8 were captured using a Toshiba Medical Systems KXO-80C instrument. Intervertebral disc height index (DHI) was analyzed using NIH-Image J software as described previously. The change in DHI before and after ESC/iPSC-NP-like cell transplantation was expressed as %DHI (post injection DHI/pre-injection DHI)(Li et al., 2014).

Histogical evaluation and immunohistochemistry

Eight weeks post cell transplantation, whole discs with the adjacent vertebrae (L4-5, L5-6, L6-7 and L7-8) were collected and dissected. Discs were fixed, paraffin-embedded, and sectioned into 5μm slices. The histological structure of the discs was examined by FAST staining according to the protocol (Leung et al., 2009).

After blocking with 5% bovine serum albumin for 30mins, the sections were incubated overnight at 4°C with primary antibodies including anti-GFP (SC-8334,

Santa Cruz), anti-Collagen II, and anti-Aggrecan. After washing with PBS, samples were incubated with the second antibody, FITC-conjugated anti-mouse IgG (1:1000), anti-rabbit IgG (1:1000) or anti-goat IgG (1:1000), for 1 hr at room temperature. Finally, the sections were washed, mounted with DAPI and photographed using a fluorescent microscope.

Statistical Analysis

Quantitative data are expressed as mean±SEM. Statistical analysis was performed by unpaired Student t test for comparisons between 2 groups and by 1-way ANOVA, followed by Bonferroni test for comparisons between 2 groups. A value of $P<0.05$ was considered statistically significant.

Acknowledgements

This research was supported in part by the NSFC (grants 31270967 and 31571407 to Q.L.), a Shenzhen seed grant (JCYJ20140828163633995), Hong Kong TRS (grant T12-708/12-N), and Hong Kong RGC GRF grants (HKU17113816 and HKU17120517 to Q.L.).

Author Contributions

Experiments and assembly of data, YLZ, ZZ, PKC; Bioinformatics analysis PKC; Experiments with assistance from RW, TKA, PY, CL, CM; Human sample collection and processing by KL, VT, TKA, KMC; Data analysis and explanations VYL, YLZ, ZZ, PKC, HFT, KSC and QZL; Manuscript written YLZ, ZZ, PKC and QZL; Concept and design, manuscript revision, and final approval of the manuscript, QZL.

Conflict of interest

The authors declare that they have no conflict of interest.

Reference:

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Chan, S.C., Tekari, A., Benneker, L.M., Heini, P.F., and Gantenbein, B. (2015). Osteogenic differentiation of bone marrow stromal cells is hindered by the presence of intervertebral disc cells. Arthritis Res Ther *18*, 29.

Choi, K.S., Lee, C., and Harfe, B.D. (2012). Sonic hedgehog in the notochord is sufficient for patterning of the intervertebral discs. Mech Dev *129*, 255-262.

Colombier, P., Clouet, J., Boyer, C., Ruel, M., Bonin, G., Lesoeur, J., Moreau, A., Fellah, B.H., Weiss, P., Lescaudron, L*., et al.* (2016). TGF-beta1 and GDF5 Act Synergistically to Drive the Differentiation of Human Adipose Stromal Cells toward Nucleus Pulposus-like Cells. Stem Cells *34*, 653-667.

Fatemieh, S.O. (2006). A Review of: "Discovering Knowledge in Data: An Introduction to Data Mining": D. T. Larose, New York: Wiley, 2005, ISBN 0-471-66657-2, xv + 222 pp., $69.95, pp. 127-130.

Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. PLoS Biol *12*, e1001937.

Gritsman, K., Talbot, W.S., and Schier, A.F. (2000). Nodal signaling patterns the organizer. Development *127*, 921-932.

Hanley, E.N., Jr., Herkowitz, H.N., Kirkpatrick, J.S., Wang, J.C., Chen, M.N., and Kang, J.D. (2010). Debating the value of spine surgery. J Bone Joint Surg Am *92*, 1293-1304.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S*., et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res *22*, 1760-1774.

Humzah, M.D., and Soames, R.W. (1988). Human intervertebral disc: structure and function. Anat Rec *220*, 337-356.

Hynes, R.O., and Naba, A. (2012). Overview of the matrisome--an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol *4*, a004903.

Iatridis, J.C., MacLean, J.J., O'Brien, M., and Stokes, I.A. (2007). Measurements of proteoglycan and water content distribution in human lumbar intervertebral discs. Spine (Phila Pa 1976) *32*, 1493-1497.

Ito, Y., and Miyazono, K. (2003). RUNX transcription factors as key targets of TGF-beta superfamily signaling. Curr Opin Genet Dev *13*, 43-47.

Kennon, J.C., Awad, M.E., Chutkan, N., DeVine, J., and Fulzele, S. (2018). Current insights on use of growth factors as therapy for Intervertebral Disc Degeneration. Biomol Concepts *9*, 43-52.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357-360.

683 Leung, V.Y., Aladin, D.M., Lv, F., Tam, V., Sun, Y., Lau, R.Y., Hung, S.C., Ngan, A.H., Tang, B.,
684 Lim, C.T.*, et al.* (2014). Mesenchymal stem cells reduce intervertebral disc fibrosis and
685 facilitate repair. Stem Cells *32*, 2164-2177.
686 Leung, V.Y., Chan, W.C., Hung, S.C., Cheung, K.M., and Chan, D. (2009). Matrix remodeling
687 during intervertebral disc growth and degeneration detected by multichromatic FAST
688 staining. J Histochem Cytochem *57*, 249-256.
689 Li, Y.Y., Diao, H.J., Chik, T.K., Chow, C.T., An, X.M., Leung, V., Cheung, K.M., and Chan, B.P.
690 (2014). Delivering mesenchymal stem cells in collagen microsphere carriers to rabbit
691 degenerative disc: reduced risk of osteophyte formation. Tissue Eng Part A *20*, 1379-1391.
692 Lv, F., Leung, V.Y., Huang, S., Huang, Y., Sun, Y., and Cheung, K.M. (2014). In search of
693 nucleus pulposus-specific molecular markers. Rheumatology (Oxford) *53*, 600-610.
694 Maaten, L.J.P.v.d., and Hinton, G.E. (2008). Visualizing High-Dimensional Data Using t-SNE.
695 2579-2605.
696 Maxim, B.F., Yakov, A.T., Melody, P., Lennart, K., Group, C.M.W., Pradeep, S., Yurii, S.A., and
697 Frances, M.K.W. (2018). Genome-Wide Association Summary Statistics For Back Pain
698 (Zenodo).
699 McCann, M.R., and Seguin, C.A. (2016). Notochord Cells in Intervertebral Disc Development
700 and Degeneration. J Dev Biol *4*.
701 Meng, X.M., Nikolic-Paterson, D.J., and Lan, H.Y. (2016). TGF-beta: the master regulator of
702 fibrosis. Nat Rev Nephrol *12*, 325-338.
703 Minogue, B.M., Richardson, S.M., Zeef, L.A., Freemont, A.J., and Hoyland, J.A. (2010).
704 Characterization of the human nucleus pulposus cell phenotype and evaluation of novel
705 marker gene expression to define adult stem cell differentiation. Arthritis Rheum *62*,
706 3695-3705.
707 Mirza, S.K., and Deyo, R.A. (2007). Systematic review of randomized trials comparing lumbar
708 fusion surgery to nonoperative care for treatment of chronic back pain. Spine (Phila Pa 1976)
709 *32*, 816-823.
710 Nakayama, E., Matsumoto, T., Kazama, T., Kano, K., and Tokuhashi, Y. (2017).
711 Transplantation of dedifferentiation fat cells promotes intervertebral disc regeneration in a
712 rat intervertebral disc degeneration model. Biochem Biophys Res Commun *493*, 1004-1009.
713 Pauklin, S., and Vallier, L. (2015). Activin/Nodal signalling in stem cells. Development *142*,
714 607-619.
715 Perez-Cruet, M., Beeravolu, N., McKee, C., Brougham, J., Khan, I., Bakshi, S., and Chaudhry,
716 G.R. (2018). Potential of Human Nucleus Pulposus-Like Cells Derived From Umbilical Cord to
717 Treat Degenerative Disc Disease. Neurosurgery.
718 Richardson, S.M., Knowles, R., Marples, D., Hoyland, J.A., and Mobasheri, A. (2008).
719 Aquaporin expression in the human intervertebral disc. J Mol Histol *39*, 303-309.
720 Riester, S.M., Lin, Y., Wang, W., Cong, L., Mohamed Ali, A.M., Peck, S.H., Smith, L.J., Currier,
721 B.L., Clark, M., Huddleston, P.*, et al.* (2018). RNA sequencing identifies gene regulatory
722 networks controlling extracellular matrix synthesis in intervertebral disk tissues. J Orthop
723 Res *36*, 1356-1369.
724 Risbud, M.V., Schaer, T.P., and Shapiro, I.M. (2010). Toward an understanding of the role of
725 notochordal cells in the adult intervertebral disc: from discord to accord. Dev Dyn *239*,
726 2141-2148.

727 Risbud, M.V., and Shapiro, I.M. (2011). Notochordal cells in the adult intervertebral disc:
728 new perspective on an old question. Crit Rev Eukaryot Gene Expr *21*, 29-41.
729 Roberts, S., Evans, H., Trivedi, J., and Menage, J. (2006). Histology and pathology of the
730 human intervertebral disc. J Bone Joint Surg Am *88 Suppl 2*, 10-14.
731 Sakai, D., Mochida, J., Iwashina, T., Hiyama, A., Omi, H., Imai, M., Nakai, T., Ando, K., and
732 Hotta, T. (2006). Regenerative effects of transplanting mesenchymal stem cells embedded in
733 atelocollagen to the degenerated intervertebral disc. Biomaterials *27*, 335-345.
734 Sakai, D., Nakamura, Y., Nakai, T., Mishima, T., Kato, S., Grad, S., Alini, M., Risbud, M.V., Chan,
735 D., Cheah, K.S.*, et al.* (2012). Exhaustion of nucleus pulposus progenitor cells with ageing and
736 degeneration of the intervertebral disc. Nat Commun *3*, 1264.
737 Silagi, E.S., Batista, P., Shapiro, I.M., and Risbud, M.V. (2018). Expression of Carbonic
738 Anhydrase III, a Nucleus Pulposus Phenotypic Marker, is Hypoxia-responsive and Confers
739 Protection from Oxidative Stress-induced Cell Death. Sci Rep *8*, 4856.
740 Sivakamasundari, V., and Lufkin, T. (2012). Bridging the Gap: Understanding Embryonic
741 Intervertebral Disc Development. Cell Dev Biol *1*.
742 Smits, P., Dy, P., Mitra, S., and Lefebvre, V. (2004). Sox5 and Sox6 are needed to develop and
743 maintain source, columnar, and hypertrophic chondrocytes in the cartilage growth plate. J
744 Cell Biol *164*, 747-758.
745 Smits, P., and Lefebvre, V. (2003). Sox5 and Sox6 are required for notochord extracellular
746 matrix sheath formation, notochord cell survival and development of the nucleus pulposus
747 of intervertebral discs. Development *130*, 1135-1148.
748 Stemple, D.L. (2005). Structure and function of the notochord: an essential organ for
749 chordate development. Development *132*, 2503-2512.
750 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
751 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment
752 analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
753 Proc Natl Acad Sci U S A *102*, 15545-15550.
754 Takada, S., Stark, K.L., Shea, M.J., Vassileva, G., McMahon, J.A., and McMahon, A.P. (1994).
755 Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes Dev *8*, 174-189.
756 Tang, R., Jing, L., Willard, V.P., Wu, C.L., Guilak, F., Chen, J., and Setton, L.A. (2018).
757 Differentiation of human induced pluripotent stem cells into nucleus pulposus-like cells.
758 Stem Cell Res Ther *9*, 61.
759 ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E., and Nusse, R. (2008). Wnt signaling
760 mediates self-organization and axis formation in embryoid bodies. Cell Stem Cell *3*, 508-518.
761 Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011).
762 Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell
763 *145*, 875-889.
764 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L.,
765 Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of
766 RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.
767 von Dassow, G., Schmidt, J.E., and Kimelman, D. (1993). Induction of the Xenopus organizer:
768 expression and regulation of Xnot, a novel FGF and activin-regulated homeo box gene.
769 Genes Dev *7*, 355-366.

Wang, W., Wang, Y., Deng, G., Ma, J., Huang, X., Yu, J., Xi, Y., and Ye, X. (2018). Transplantation of Hypoxic-Preconditioned Bone Mesenchymal Stem Cells Retards Intervertebral Disc Degeneration via Enhancing Implanted Cell Survival and Migration in Rats. Stem Cells Int *2018*, 7564159.

Winzi, M.K., Hyttel, P., Dale, J.K., and Serup, P. (2011). Isolation and characterization of node/notochord-like cells from mouse embryonic stem cells. Stem Cells Dev *20*, 1817-1827.

Wu, H., Shang, Y., Yu, J., Zeng, X., Lin, J., Tu, M., Cheang, L.H., and Zhang, J. (2018). Regenerative potential of human nucleus pulposus resident stem/progenitor cells declines with ageing and intervertebral disc degeneration. Int J Mol Med *42*, 2193-2202.

Yang, S.H., Wu, C.C., Shih, T.T., Sun, Y.H., and Lin, F.H. (2008). In vitro study on interaction between human nucleus pulposus cells and mesenchymal stem cells through paracrine stimulation. Spine (Phila Pa 1976) *33*, 1951-1957.

Yasuo, H., and Lemaire, P. (2001). Role of Goosecoid, Xnot and Wnt antagonists in the maintenance of the notochord genetic programme in Xenopus gastrulae. Development *128*, 3783-3793.

Zhang, Z., Zhang, Y., Gao, F., Han, S., Cheah, K.S., Tse, H.F., and Lian, Q. (2017). CRISPR/Cas9 Genome-Editing System in Human Stem Cells: Current Status and Future Prospects. Mol Ther Nucleic Acids *9*, 230-241.

Zhou, X., Ma, C., Hu, B., Tao, Y., Wang, J., Huang, X., Zhao, T., Han, B., Li, H., Liang, C.*, et al.* (2018). FoxA2 regulates the type II collagen-induced nucleus pulposus-like differentiation of adipose-derived stem cells by activation of the Shh signaling pathway. FASEB J, fj201800373R.

Zhou, X., Tao, Y., Liang, C., Zhang, Y., Li, H., and Chen, Q. (2015). BMP3 Alone and Together with TGF-beta Promote the Differentiation of Human Mesenchymal Stem Cells into a Nucleus Pulposus-Like Phenotype. Int J Mol Sci *16*, 20344-20359.

Zhu, C., Jiang, W., Cheng, Q., Hu, Z., and Hao, J. (2018). Hemeoxygenase-1 Suppresses IL-1beta-Induced Apoptosis Through the NF-kappaB Pathway in Human Degenerative Nucleus Pulposus Cells. Cell Physiol Biochem *46*, 644-653.

Figure legends

Figure 1. Early notochord markers are differentiated from ESC/iPSCs.

A. Schematic diagram illustrating two steps for directed differentiation of human ESC/iPSCs into notochord-like cells.

B. RT-PCR demonstrating the mRNA levels for Noto, Foxa2 and T on different days post differentiation from ESC3 (i), ESC9 (ii) and iPSC (iii). n=3.

C. Immunofluorescence staining showing the expression of Noto, T and Foxa2 on day 5 after differentiation from ESC3 (i), ESC9 (ii) and IMR-90-iPSC (iii), respectively. Scale bars: 50 μm.

All data are represented as mean ± SEM per independent experiments. Difference between means was analyzed using one-way ANOVA followed by Bonferroni's Multiple Comparison Test. $^{**}P < 0.01$.

Figure 2. Differentiation of NP-like cells from hESC/iPSCs

A. Schematic diagram illustrating directed differentiation of human ESC/iPSCs into NP-like cells.

B. Representative images showing the morphology of NP-like cells on Day 14,17, and 19 during the differentiation (i). The vacuoles in the differentiated cells are marked by arrows (i). Representative images of immunofluorescence staining showing the expression of GD2, Tie2 and Collagen II in the differentiated cells (ii). Scale bars: 50 μm.

C. The GD2/Tie2 double positive cells were sorted by FACS (i). The vacuoles were observed in the sorted NP-like cells (ii). Representative images of immunofluorescence staining showing the expression of GD2, Tie2, Collagen II, Aggrecan and T but not Noto in the sorted NP-like cells (iii). Scale bars: 50 μm.

D. Two types of colonies derived from NP-like cells in methylcellulose medium at 10 days. NP-CFU-S (left) and NP-CFU-F (right) (i). Immunofluorescence staining showed Collagen II and Aggrecan were highly expressed in NP-CFU-S but not NP-CFU-F(i). One thousand NP-like cells from each population were subjected to a CFA for 10 days. The frequency of NP-CFU-S was greatly increased in GD2+/Tie2+ sorted cells compared with other population (ii). n=3. Scale bars: 50 μm.

E. CdM of NP-like cells greatly enhanced the luciferase activity of Light2 cells in a dose dependent manner. n=3.

All data are represented as mean±SEM per independent experiments. Difference between means was analyzed using one-way ANOVA followed by Bonferroni's Multiple Comparison Test. $^{**}P < 0.01$.

Figure 3. Transcriptomic analyses of *in vitro* NP cells and their cross-comparisons with *in vivo* NP.

A. A heatmap and bi-clustering of a set of important genes in NP development. Colors indicate the normalized expression levels. Data are gene-wise standardization (zero-mean and unit variance) of $\log_2$FPKM (same in G).

B. A heatmap showing the pairwise similarity among samples before and after differentiations as measure by genome-wide Pearson correlation coefficients. The samples were also clustered by hierarchical clustering based on the similarity matrix.

C. A violin and box-plot showing the increased similarity to a set of four *in vivo* NP samples taken from surgical samples of the human intervertebral discs. The Student's *t* test p-value is 1.29×10^{-07}.

D. A scatter-plot of all genes before (grey, averaged over three cell-lines) and after differentiations (light-green, cyan and pink; averaged over 12 replicates). The most influential genes, whose removal significantly reduce the similarities, were highlighted by an arrow pointing from its expression level in the hESC/iPSCs to its expression in the *in vitro* differentiated NPCs. Left arrows indicate negative influential genes: those contributing to increased similarities by having lower expression in the *in vitro* differentiated NPCs; right arrows indicate positive influential genes: those contributing to increased similarities by having higher expressions in the *in vitro* differentiated NPCs.

E. A Venn diagram showing the overlaps among differentially expressed genes (DEGs) between the *in vivo* differentiated NP and hESCs/iPSCs, and those between *in vitro* differentiated NPCs and hESCs/iPSCs, and the positive influential genes identified in (D).

F. A gene-ontology (GO) analysis of the 148 positive influential genes in overlap shows that they are enriched for various types of ECMs and their regulatory pathways.

G. A heatmap and bi-clustering of the 148 genes in overlap.

Figure 4. Transplantation of ESC/iPSC- NP-like cells attenuates disc injury in a rat model of IDD.

A. Two weeks after intervertebral disc injury, ESC/iPSC- NP-like cells or PBS were injected into the injured intervertebral disc. Representative radiographic images of IDD from the different experimental groups before or after injury.

B. Time course of change to percentage disc height index (% DHI) for the different experimental groups. n=6/group.

C. Representative images of Fast staining showing a damaged intervertebral disc at 8 weeks post cell transplantation.

All data are represented as mean ± SEM per independent experiments. Difference between means was analyzed using one-way ANOVA followed by Bonferroni's Multiple Comparison Test. $^{**}P < 0.01$ vs. PBS group.

Figure 5. Evaluation of GFP-labeled ESC/iPSC- NP-like cell transplantation in the rat model of IDD.

A. Representative immunohistochemical images for GFP showing the survival of ESC/iPSC- NP-like cells at 2, 4 and 8 weeks post cell injection (i). The cell density of

ESC/iPSC-NP-like cells at 2, 4 and 8 weeks post transplantation was calculated (ii). n=6/group. Scale bars: 50 μm.

B. Representative immunohistochemical images for Collagen IIA among the different experimental groups (i). The fluorescence density of Collagen IIA among the different experimental groups was calculated (ii). Representative immunohistochemical images for Aggrecan among the different experimental groups (iii). The fluorescence density of Aggrecan among the different experimental groups was calculated (iv). n=6/group. Scale bars: 100 μm.

All data are represented as mean ± SEM per independent experiments. Difference between means was evaluated using one−way ANOVA followed by Bonferroni's Multiple Comparison Test. $^{**}P < 0.01$.

Figure 6. Generation of a Noto-GFP reporter cell line

A. General sketch for GFP knock-in. gRNA was designed to be close to the stop codon. The donor comprised two homologous arms, exogenous gene and selection cassette, which finally formed a Noto-2A-GFP-LoxP-Puro-LoxP-Noto” structure.

B. SURVEYOR assay of gRNAs. gRNA2 showed higher cutting efficiency in ESC9.

C. Characterization of GFP knock-in. Precise GFP integrated colonies present 6590bp of PCR products.

D. Characterization of selection cassette removal. Longer PCR products (1309bp) indicated that the selection cassette had been successfully removed.

E. Sanger sequencing results. No mutation was detected in the Noto-GFP reporter cell line.

Figure 7. Evaluation of Noto-GFP reporter by differentiation

A. Schematic diagram illustrating two steps for directed differentiation of human ESC/iPSCs into notochord-like cells.

B. Five days after differentiation, GFP cells were observed with a fluorescence microscope (iii, iv) but not in ESC9 (i, ii). The Noto and GFP were shown by immune staining to be co-expressed in the same cell (v-viii). Scale bars: 50 μm.

C. Compared with ESC9, the mRNA levels of Noto, Foxa2 and T were greatly enhanced in the differentiated cells (i-iii). n=3.

D. Differentiating hES cells that were Noto-GFP positive were dissociated into single cells. A small portion of Noto-GFP positive cells were sorted out. The expression level of Noto in sorted out cells was over 100 times greater that of undifferentiated hES cells (i-iii). n=3.

All data are represented as mean±SEM per independent experiments. Difference between means was evaluated using Unpaired *t*-test (C, D). $^{**}P < 0.01$.