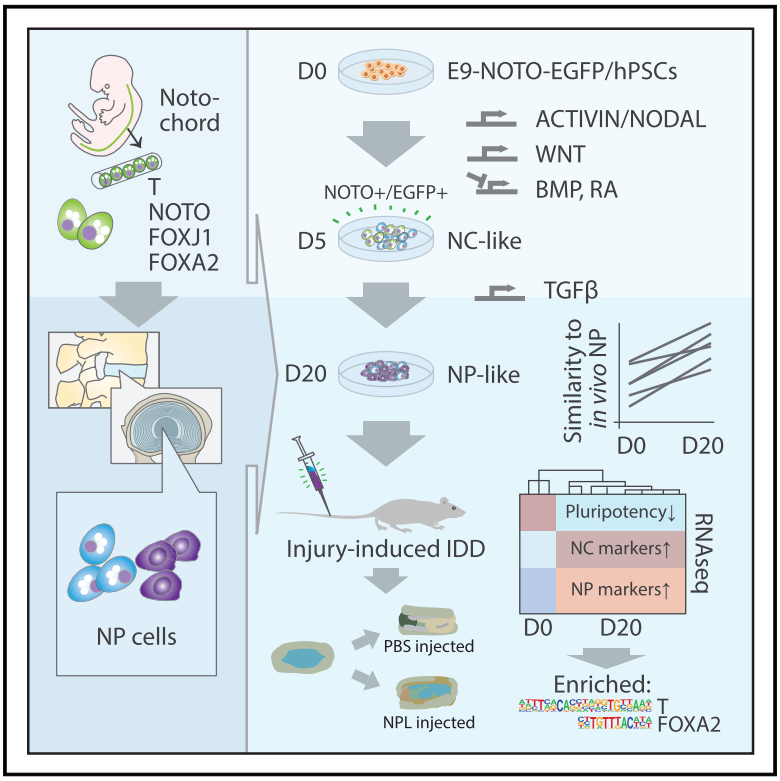
Article

Cell Reports

Directed Differentiation of Notochord-like and Nucleus Pulposus-like Cells Using Human Pluripotent Stem Cells

Graphical Abstract



Highlights

- A NOTO-EGFP reporter in hPSCs is established to monitor differentiation to NCLs
- hPSCs can be induced into NCLs and NP-like cells using a defined protocol
- hPSC-NP-like cells and human NP cells share high similarities in molecular signatures
- Transplantation of hPSC-NP-like cells attenuates injury-induced intervertebral disc degeneration

Authors

Yuelin Zhang, Zhao Zhang, Peikai Chen, ..., Victor Y. Leung, Kathryn S.E. Cheah, Qizhou Lian

Correspondence

kathycheah@hku.hk (K.S.E.C.), qzlian@hku.hk (Q.L.)

In Brief

Zhang et al. report notochord-like and nucleus pulposus (NP)-like cells can be derived from human pluripotent stem cells using a *NOTO-eGFP* reporter system and a compound-defined protocol. These derived NP-like cells share high similarities with adolescent human NP cells and attenuate injury-induced intervertebral disc degeneration after transplantation.

Zhang et al., 2020, Cell Reports *30*, 2791–2806
February 25, 2020
https://doi.org/10.1016/j.celrep.2020.01.100

CellPress

Cell Reports
Article

Directed Differentiation of Notochord-like and Nucleus Pulposus-like Cells Using Human Pluripotent Stem Cells

Yuelin Zhang,[1,2,7] Zhao Zhang,[1,2,7] Peikai Chen,[3,7] Chui Yan Ma,[1] Cheng Li,[1] Tiffany Y.K. Au,[3] Vivian Tam,[3] Yan Peng,[4] Ron Wu,[3] Kenneth Man Chee Cheung,[4] Pak C. Sham,[5] Hung-fat Tse,[1] Danny Chan,[3] Victor Y. Leung,[4] Kathryn S.E. Cheah,[3,*] and Qizhou Lian[1,2,6,8,*]

[1]Department of Medicine, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong
[2]Guangzhou Women and Children's Medical Centre, Guangzhou Medical University, Guangzhou, Guangdong 510080, China
[3]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong
[4]Department of Orthopaedics and Traumatology, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong
[5]Centre for PanorOmic Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong
[6]The State Key Laboratory of Pharmaceutical Biotechnology, the University of Hong Kong, Hong Kong
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: kathycheah@hku.hk (K.S.E.C.), qzlian@hku.hk (Q.L.)
https://doi.org/10.1016/j.celrep.2020.01.100

SUMMARY

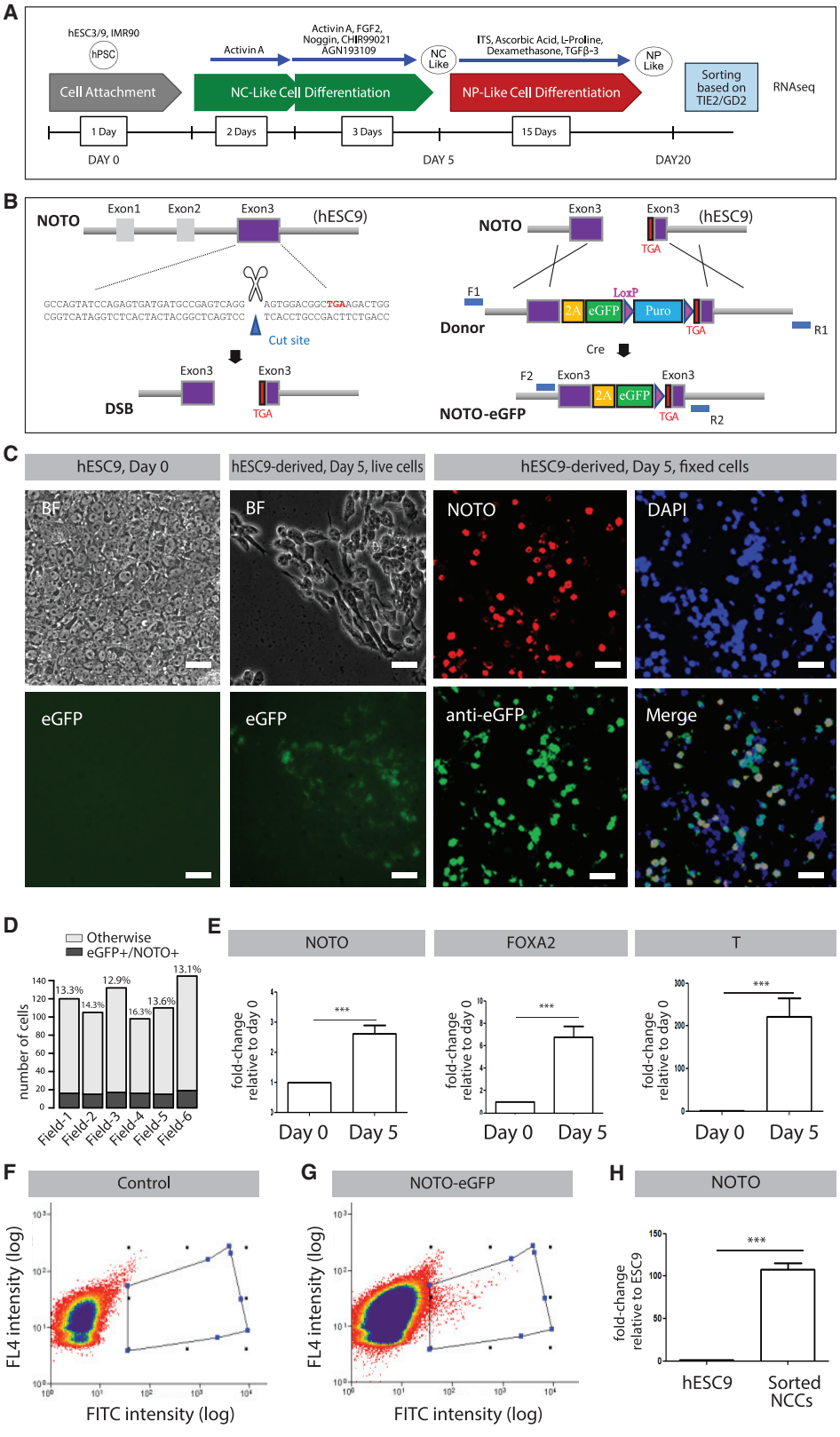
Intervertebral disc degeneration might be amenable to stem cell therapy, but the required cells are scarce. Here, we report the development of a protocol for directed *in vitro* differentiation of human pluripotent stem cells (hPSCs) into notochord-like and nucleus pulposus (NP)-like cells of the disc. The first step combines enhancement of ACTIVIN/NODAL and WNT and inhibition of BMP pathways. By day 5 of differentiation, hPSC-derived cells express notochordal cell characteristic genes. After activating the TGF-β pathway for an additional 15 days, qPCR, immunostaining, and transcriptome data show that a wide array of NP markers are expressed. Transcriptomically, the *in vitro*-derived cells become more like *in vivo* adolescent human NP cells, driven by a set of influential genes enriched with motifs bound by BRACHYURY and FOXA2, consistent with an NP cell-like identity. Transplantation of these NP-like cells attenuates fibrotic changes in a rat disc injury model of disc degeneration.

INTRODUCTION

Intervertebral disc degeneration (IDD) is a major cause of low back pain, impairing the quality of life in more than 50% of the global population (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018). IDD is a complex disorder of aging compounded by environmental and genetic factors. Traditional treatments such as physiotherapy and drugs that aim to provide symptomatic relief do not reverse disc degeneration (Lindbäck et al., 2018; Mirza and Deyo, 2007). Surgical intervention is a common option for severe IDD and involves removal of the defective disc, then fusion of the vertebral segments or implantation of an artificial disc made of synthetic polymers or metallic compounds (Othman et al., 2019). The long-term problem of wear and failure of artificial discs after transplantation nonetheless remains (van Ooij et al., 2007). Alternative therapeutic strategies that can stimulate repair or regeneration of the degenerated disc are needed.

The intervertebral disc comprises three interdependent tissues: the nucleus pulposus (NP), a central hydrated gelatinous core, sandwiched between the cartilaginous endplates of the adjacent vertebral bodies and surrounded by the annulus fibrosis, a ring-like fibrous laminar structure (Happey et al., 1964). Cells in these three compartments synthesize extracellular matrix (ECM) proteins such as collagens, proteoglycans, and fibronectin/laminins, which maintain the structural and functional integrity of the disc (Roberts et al., 2006). During development, the NP forms from the segmentation of the embryonic notochord, a rod-like structure that drives longitudinal growth, producing signals crucial for induction and patterning of the sclerotome, which gives rise to the vertebral bodies (Lawson and Harfe, 2017; Stemple, 2005). Development of the human notochord, as in the rat and mouse, occurs via three morphological processes, involving the formation of the anterior notochord (prechordal plate and anterior head process), the trunk notochord, and the tail notochord (de Bree et al., 2018; Yamanaka et al., 2007). Lineage-tracing studies in mice expressing *cre* recombinase under the control of the Sonic hedgehog (*Shh*) (Choi et al., 2008) or the notochord-specific *Noto* gene (McCann et al., 2012) have shown that all the cells in the NP are derived from precursors in the notochord.

Human fetal NP is characterized by the presence of large vacuolated notochord-like cells (NCLs), while small non-vacuolated chondrocyte-like cells (CLCs) are present in adult NP (Risbud et al., 2015; Rodrigues-Pinto et al., 2014). Recent fate-mapping studies in the mouse show that CLCs in the aging NP are derived from early notochord-like NP cells (Mohanty et al., 2019), raising the possibility that a similar lineage relationship may exist in the ontogeny of human NP cells. Early NP NCLs express Brachyury (*T*) and other genes characteristic of the notochord later in development (such as Keratin 8, 18, and 19)



(legend on next page)

(Rodrigues-Pinto et al., 2016, 2018). Progressive loss of NCLs and replacement with CLCs in the NP are associated with aging and the onset of IDD (Weiler et al., 2010). Whether, as in the mouse, the chondrocyte-like NP cells are late-stage cells derived from NCLs remains to be validated (Mohanty et al., 2019). Although the underlying pathogenesis is not fully understood, exhaustion of NCLs may affect NP function and hence the health of intervertebral disc (Cappello et al., 2006).

Transplantation of mesenchymal stem cells (MSCs) and *in vitro*-differentiated NP-like cells derived from human umbilical cord or human induced pluripotent stem cells (hiPSCs) into the intervertebral disc has been shown to promote recovery from injury-induced IDD in animal models (Henriksson et al., 2009; Leung et al., 2014; Perez-Cruet et al., 2019; Wei et al., 2016). Therefore, transplantation of NCLs and/or healthy young NP cells or their precursors may be a novel strategy for IDD therapy, but it is ethically infeasible and technically challenging to extract NP cells from healthy individuals. An appropriate source of NP precursor cells for cell therapy is urgently needed.

Human pluripotent stem cells (hPSCs) have proved to be an important source for the treatment of various diseases, including spinal cord injury (Okubo et al., 2018) and skeletal degeneration (Sheyn et al., 2016). Efforts to differentiate hPSCs into NP-like cells have been reported (Tang et al., 2018; Xia et al., 2019; Zhu et al., 2017), but several problems remain. First, a key prerequisite is to judge the molecular identity of the derived cells as putative embryonic notochordal cells and NP cells (composed of NCLs and CLCs). In the mouse, the homeobox gene, *Noto*, is detected not only in the axial mesoderm at the anterior tip of the primitive streak (PS) but also in the adjacent endoderm, subsequently localized to the node and its descendant notochord (Yamanaka et al., 2007) and transiently expressed in early notochord development (Abdelkhalek et al., 2004), and has been used for the identification of NCLs in the *in vitro* differentiation of pluripotent mouse and human stem cells (Tang et al., 2018; Winzi et al., 2011). In addition, in the mouse, the early notochord is characterized by the expression of transcription factors essential for its development, such as T (Rashbass et al., 1991; Zhu et al., 2016), Forkhead A2 (*Foxa2)* (Ang and Rossant, 1994), *Sox5*, *Sox6* (Smits and Lefebvre, 2003), and *Sox9* (Barrionuevo et al., 2006). Definition of human notochordal cell and NP cell signatures has been hampered by challenges in obtaining samples and data from very early human embryos, late fetal tissue, from mature NP in which several putative molecular criteria for identifying young healthy human NP cells have been reported (Minogue et al., 2010a; Tang et al., 2016). Keratin 8, 18, and 19 and CD24 have been proposed as human NCL-specific markers during early intervertebral disc development, but they are known to be expressed in other tissues (Fujita et al., 2005; Minogue et al., 2010b). The challenge is further complicated by the inherent cellular heterogeneity in the NP *in vivo*. Until reliable cell surface markers and molecular signatures for the resident NP cells (NCLs and CLCs) are available, the efficiencies of these differentiation strategies cannot be assessed.

In this study, we used the CRISPR/Cas9 technique to precisely knock in a gene for enhanced green fluorescent protein (EGFP) at the stop codon of *NOTO* in hPSCs, thereby allowing us to monitor the differentiation progress to NCLs. We developed a compound-defined protocol to differentiate NCLs from the *NOTO-EGFP* hPSCs and three independent hPSC lines. In a second phase we continued differentiation in the presence of TGF-β3 and assessed the identity of the differentiated cells by global transcriptomic sequencing on the cells before and after differentiation. We generated global bulk transcriptome data from the NP of human adolescent non-degenerated discs and derived their molecular signature, which served as an *in vivo* NP reference transcriptome (with the caveat that this includes a combination of both NCLs and CLCs). We also developed an advanced statistical method to identify genes characteristic of the NP that serve to assess the similarity between the *in vitro*-derived NP-like cells and *in vivo* NP cells. By this approach we identified the characteristic molecular signatures of NP-like cells within the hPSC-derived differentiated cell population. Last, we transplanted these NP-like cells into a rat model of IDD and found that they effectively attenuated injury-induced disc degeneration.

RESULTS

Establishing an hPSC *NOTO*-EGFP Reporter for the Differentiation of NCLs

Developmental studies in the mouse have shown that derivation of the notochord from the chordamesoderm in early embryonic development is regulated by the Activin/Nodal signaling pathway (Gritsman et al., 2000; Pauklin and Vallier, 2015) and also requires the expression of key transcription factors, such as *FOXA2*, *NOTO*, and *T* (Abdelkhalek et al., 2004; Ang and

Figure 1. Generation of *NOTO*-EGFP Reporter hESC9 and Differentiation to NCLs

(A) Schematic diagram showing the NCL/NPL differentiation protocols.

(B) Schematic targeting strategy of EGFP into the *NOTO* gene. Guide RNAs (gRNAs) were designed close to the stop codon of *NOTO*. The donor comprises two homologous arms, exogenous gene (EGFP), and selection cassette to finally form a *NOTO*-2A-EGFP-LoxP-Puro-LoxP-*NOTO* structure. After knockin, EGFP was linked to the stop codon of *NOTO* using T2A as an adaptor.

(C) Immunoanalysis detection of NOTO and EGFP-expressing cells 5 days after hPSC differentiation (leftmost column of right panel) but not in hESC9 (left panel). Note co-expressing cells (other columns of right panel). Scale bars: 50 μm.

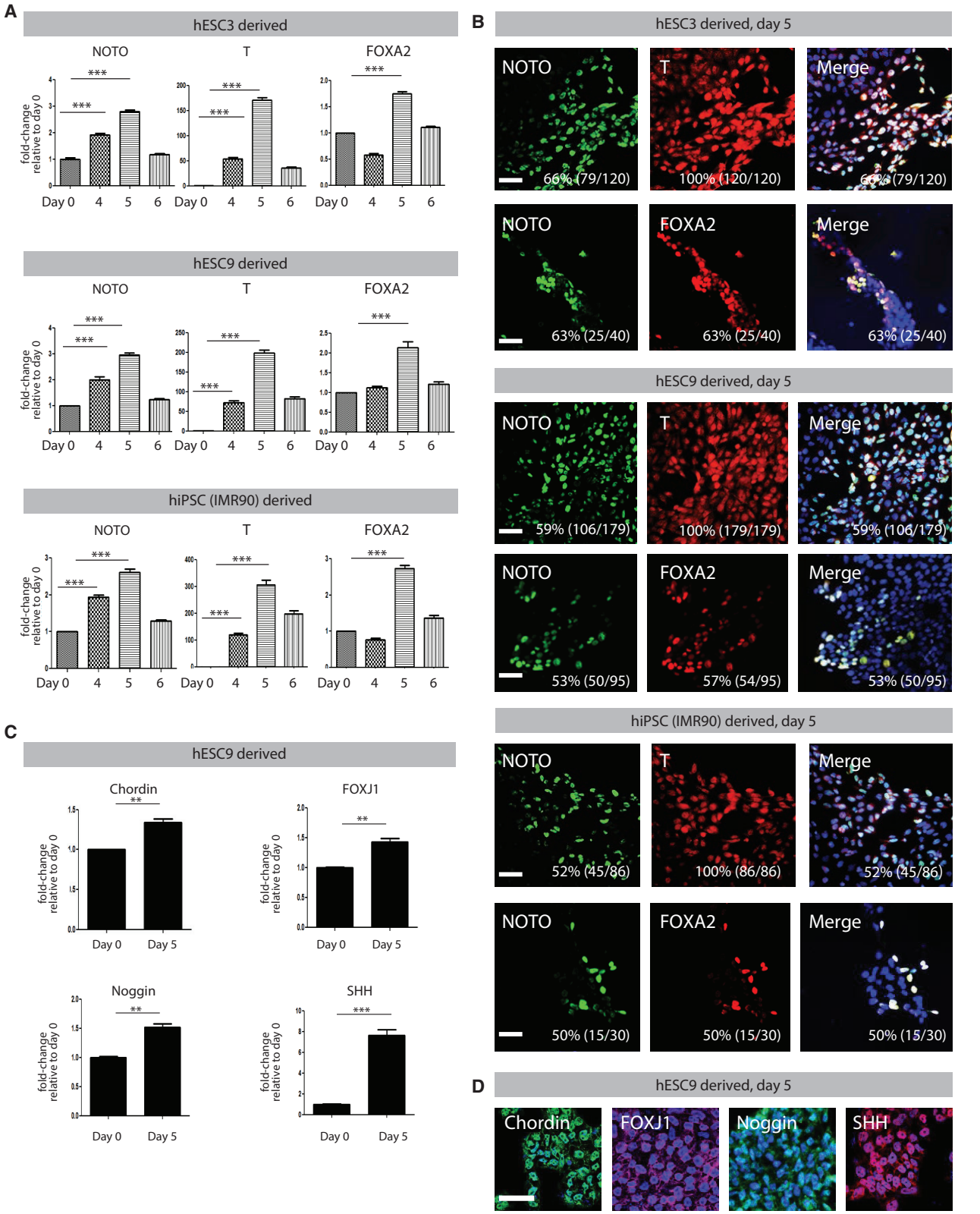
(D) Efficiency of differentiation of $NOTO^+/GFP^+$ cells at day 5. Percentage of $NOTO^+/GFP^+$ cells scored in six fields (top of each bar). Single-positives account for small proportions (see Table S1C).

(E) Differential upregulated expression of notochordal characteristic genes in hESC9-derived cells, the mRNA levels (fold changes relative to day 0, same in F) of *NOTO*, *FOXA2*, and *T* increase by 2.5-, 7-, and more than 200-fold, respectively, in the differentiated cells.

(F and G) Flow cytometry profiles for control (F) and NOTO-EGFP-positive cells (G).

(H) Expression level of *NOTO* in sorted *NOTO*-EGFP-positive cells. Note 100 times higher *NOTO*-EGFP expression than for undifferentiated hESC9.

Error bars: mean ± SEM; n = 3 biological replicates; $^{**}p < 0.01$ and $^{***}p < 0.001$ by Student's *t* test.



(legend on next page)

Rossant, 1994; Herrmann and Kispert, 1994). *NOTO* is a highly conserved homeobox transcription factor, whose expression is restricted to the organizer node and the nascent notochord during gastrulation and axis elongation in mice (McCann et al., 2012; Stein and Kessel, 1995; von Dassow et al., 1993).

To be able to monitor the progress of differentiation of NCLs, we generated *NOTO*-EGFP reporter cells from an hPSC line (hESC9) using CRISPR/Cas9-mediated gene knockin strategy to insert a transgene comprising the 2A self-cleaving peptide (Szymczak et al., 2004) fused to the EGFP sequence, adjacent to the stop codon of *NOTO* (Figure 1B). Four hESC9 colonies were confirmed to have correct integration of the *NOTO*-EGFP reporter by PCR (Table S1A; Figure S1D). Sanger sequencing detected no frameshift, deletion, insertion, or point mutation in the *NOTO*-EGFP reporter cell line (Figure S1E).

In hPSCs, early mesodermal differentiation and node formation require activation of *WNT* signaling (ten Berge et al., 2008). Blockade 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 mesodermal lineage and acts as an inducer of *Xnot* (an ortholog of *Noto*) in the gastrula organizing region of *Xenopus* embryos (von Dassow et al., 1993). Retinoic acid (RA) at physiological concentrations inhibits notochord differentiation (Thomson et al., 2011). Therefore, inhibition of RA is a strategy to improve notochord cell (NC) differentiation (Winzi et al., 2011).

On the basis of the above knowledge, we developed an optimized compound-based protocol (Figure 1A) to achieve directed lineage differentiation of the *NOTO*-EGFP reporter hPSCs toward a notochordal fate. First, we applied N2B27 medium with ACTIVIN A, for 48 h, as it has been reported that withdrawal of pluripotency conditions for 48 h can cause hPSCs to be more sensitive to differentiation signals (Thomson et al., 2011). Next, we administered ACTIVIN A, CHIR (CHIR99021, *WNT3A* activator), NOGGIN (BMP-4 antagonist), AGN193109 (RA inhibitor), and FGF2 for another 2–4 days. EGFP-positive cells were highest on day 5 (Figure 1C, right panel). Immunostaining showed co-expression of EGFP and *NOTO* (Figure 1C, right panel), with 13.9% ± 1.3% of the cells being EGFP/*NOTO* double-positive (Figure 1D; Table S1C). qPCR results indicated high expression levels of *NOTO*, *FOXA2*, and *T* (Figure 1E).

To isolate and purify the NCLs, we digested the differentiated *NOTO*-EGFP reporter hESC9 into single cells and sorted the *NOTO*-EGFP-positive cells by flow cytometry using the undifferentiated cells as the negative control (Figures 1F and 1G). Unfortunately, consistent with a previous report (Winzi et al., 2011), most EGFP-positive cells died during dissociation from differentiating human embryonic stem cell (hESC) clusters, and only ~0.1% living EGFP-positive cells were obtained after sorting. However, those surviving cells did show much higher expression (>100 times) of *NOTO* compared with the undifferentiated cells (Figure 1H).

Differentiation of hPSCs to NCLs

On the basis of these findings, we applied the compound-based protocol to achieve directed lineage differentiation of three hPSC lines to NCLs: two hESC lines (hESC3 and hESC9) and one hiPSC line (IMR90) (Figure 1A). We then evaluated the mRNA expression of several key genes involved in the early stage of notochord differentiation, including *NOTO*, *T*, and *FOXA2*. The results of qPCR showed that the expression of *NOTO* increased gradually, peaking on day 5 after differentiation of hESC3, hESC9, and hiPSC (IMR90) (Figure 2A). The expression of mesoderm markers *T* and *FOXA2*, which act upstream of *NOTO*, also peaked on day 5 (Figure 2A). Immunoanalysis on day 5 showed that 50%–66% of NOTO-positive cells also expressed T or FOXA2 (Figure 2B; Figure S1F).

We examined the expression of reported diagnostic early notochord markers expressed in mouse and human (Rodrigues-Pinto et al., 2018), such as *NOGGIN* (Lustig et al., 1996), *CHORDIN* (Kuroda et al., 2004), and *FOXJ1* (Beckers et al., 2007), in the hESC9 line. *Shh* is expressed in the early mouse notochord and is essential for its development and growth (Chiang et al., 1996; Echelard et al., 1993). Interrogation of microarray data derived from sorted CD24-positive cells isolated from human embryonic and fetal spines (7.5–14 weeks post-conception) (Rodrigues-Pinto et al., 2018) show expression of *SHH*. qPCR results showed the expression of these four genes increased between 1.3 and 7 times (Figure 2C), and immunostaining confirmed the strong expression of their encoded proteins on day 5 (Figure 2D). Overall, the increased expression of these markers suggests that early notochordal cell differentiation was induced.

Further Differentiation of NCLs into Maturing NP-like Cells

Because notochordal cells are precursors of NP cells and the TGF-β pathway is essential for NP cell maintenance (Tang et al., 2018; Zhou et al., 2015), to differentiate the NCLs on day 5 into more mature NP-like cells, we administered TGF-β3 and cultured the cells for an additional 15 days in monolayer culture. Vacuolated cells, a characteristic of NCLs, were observed after 14, 17, and 19 days of treatment (Figure 3A). Immunostaining showed that the TGF-β-treated cells expressed

Figure 2. Directed *In Vitro* Differentiation of hPSC-Derived NCLs

(A) qPCR showing the mRNA levels (fold changes relative to day 0; same in C) for *NOTO*, *FOXA2*, and *T* on different days post-differentiation from hESC3 (upper panel), hESC9 (middle panel), and IMR90 (bottom panel).

(B) Immunostaining showing the expression of NOTO, T, and FOXA2 on day 5 after differentiation from hESC3 (upper panel), hESC9 (middle panel), and IMR90 (bottom panel), respectively. Scale bars: 50 μm.

(C) qPCR showing the mRNA levels for *Chordin*, *FOXJ1*, *Noggin*, and *SHH* on day 0 and day 5 post-differentiation in hESC9, respectively.

(D) Immunofluorescence staining showing the expression of Chordin, FOXJ1, Noggin, and SHH on day 0 and day 5 post-differentiation in hESC9, respectively. Scale bars: 50 μm.

NC-like/NCL, notochord-like; NP-like/NPL, nucleus pulposus-like. Error bars indicate mean ± SEM; n = 3 biological replicates; $^{**}p < 0.01$ and $^{***}p < 0.001$ by one-way ANOVA, followed by Bonferroni corrections.

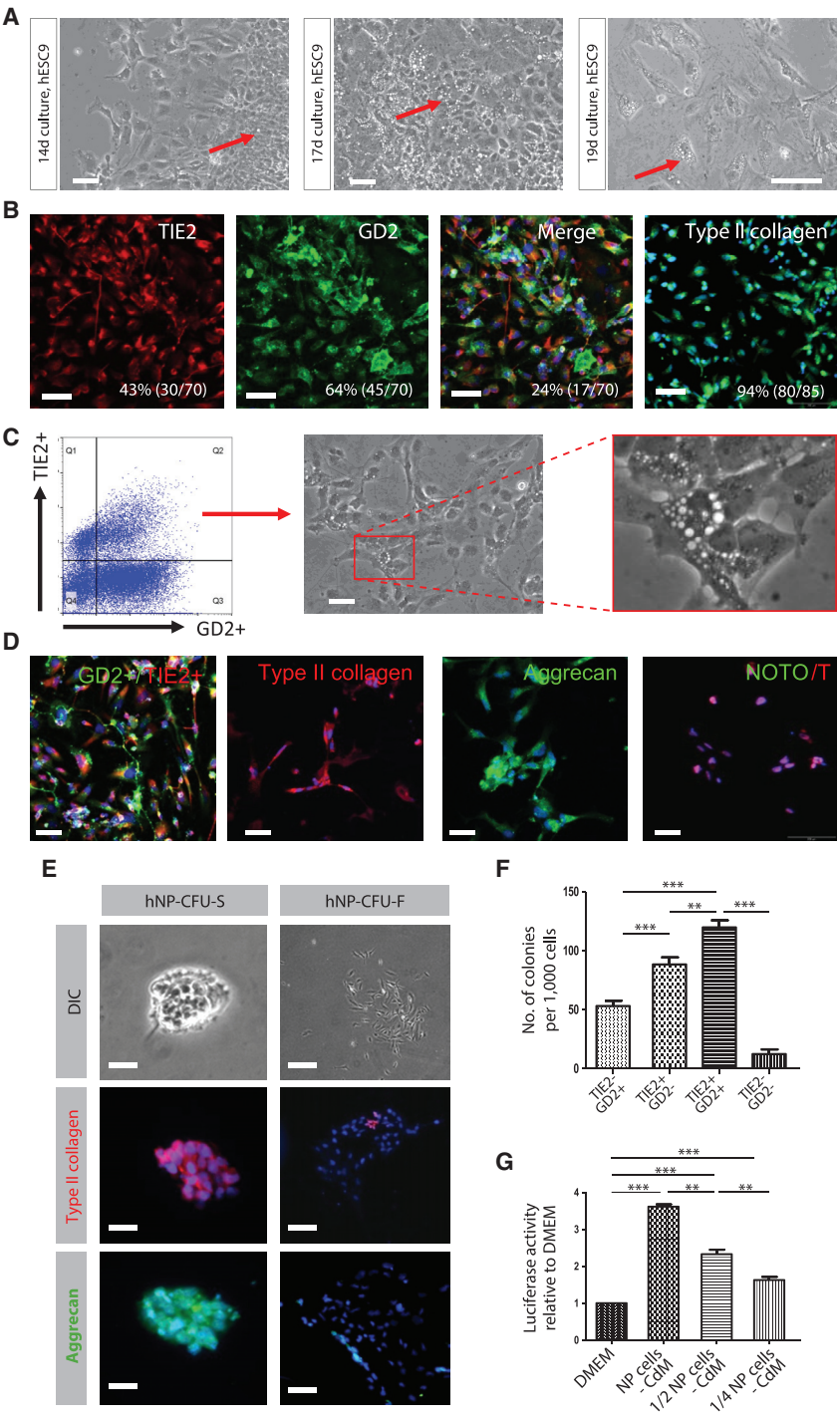


Figure 3. Differentiation of NP-like Cells from Day 5 hPSC-Derived NCLs

(A) Vacuolated morphology of hESC9-derived NP-like cells on days 14, 17, and 19 of differentiation. Arrows mark vacuoles in the representative images of hPSC-derived differentiated cells.

(B) Immunoanalyses of GD2, TIE2, and type II collagen expression in hESC9-derived cells on day 20. Representative images of immunofluorescence staining are shown.

(C) Vacuolar phenotype of GD2/TIE2 FACS sorted double-positive (left) on day 20 hESC9-derived cells (middle and right).

(D) Immunoanalyses of expression of GD2, TIE2, type II collagen, aggrecan, and T but not NOTO in sorted hESC9-derived cells on day 20.

(E) Two types of colonies derived from NP-like cells in methylcellulose medium at 10 days. NP-CFU-S (left) and NP-CFU-F (right). Immunofluorescence staining showed that type II collagen and aggrecan were highly expressed in NP-CFU-S but not NP-CFU-F. One thousand NP-like cells from each population were cultured in methylcellulose medium for 10 days.

(F) The frequency of NP-CFU-S significantly increased in GD2+/TIE2+ sorted cells compared with other populations.

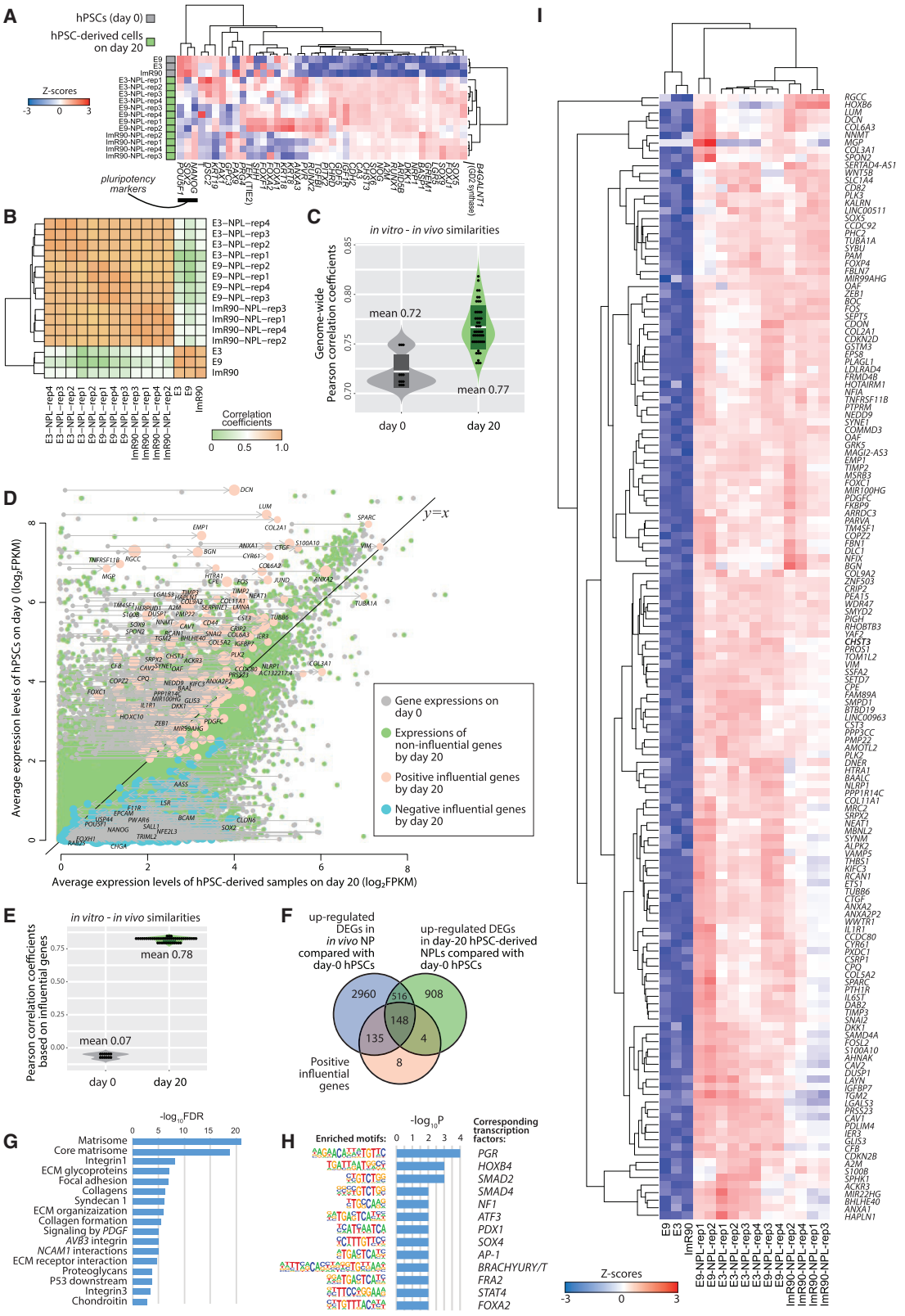
(G) Conditioned medium (CdM) of NP-like cells significantly enhanced the luciferase activity of Light2 cells in a dose-dependent manner.

Scale bars: 50 μm. Error bars indicate mean ± SEM; n = 3 biological replicates; $^{**}p < 0.01$ and $^{***}p < 0.001$ by one-way ANOVA, followed by Bonferroni corrections. CFU, colony-forming unit.

TEK receptor tyrosine kinase (TIE2), and disialoganglioside 2 (GD2), proposed surface markers of disc NP progenitors (Sakai et al., 2012) (Figure 3B). These differentiated cells also expressed aggrecan (*ACAN*) and type II collagen (*COL2A1*), the major chondrocyte markers of NP cells (Figure 3B) (Sive et al., 2002). By fluorescence-activated cell sorting (FACS) analysis, 25%–30% of the differentiated cells were TIE2/GD2 positive (Figure 3C, left). These TIE2/GD2-enriched cell populations, when sorted and cultured in NP cell medium, were vacuolated (Figure 3C, middle and right) and were positive for TIE2, GD2, T, aggrecan, and type II collagen but not NOTO (Figure 3D), suggesting that the differentiation has proceeded beyond the early notochord-like stage.

Culturing hESC9-derived NP-like cells in methylcellulose medium for 10 days led to the formation of two types of colonies, colony-forming units-fibroblastic (CFU-F) and colony-forming units-spherical (CFU-S) (Figure 3E, first row). Type II collagen and aggrecan were highly expressed in CFU-S but not in CFU-F (Figure 3E, bottom two rows). Furthermore, TIE2/GD2 double-positive cells formed more CFU-S than other subpopulations (Figure 3F).

SHH secreted by the notochord is required to induce *Shh* in the floor plate (Fan and Tessier-Lavigne, 1994; Johnson et al., 1994) and pattern the intervertebral discs (Choi and Harfe, 2011). To assess whether the hPSC-derived cells expressed functionally active SHH, we collected the conditioned medium of the hESC9-differentiated NP-like cells and used it to culture a Light2 cell line, which carries a luciferase-based reporter



(legend on next page)

responsive to SHH secreted protein (Taipale et al., 2000; Wang et al., 2010). Compared with DMEM, the conditioned medium greatly enhanced the luciferase activity of Light2 cells in a dose-dependent manner, suggesting that the hESC9-differentiated NP-like cells secrete active SHH (Figure 3G).

Genome-wide Transcriptomic Analysis Shows Rich Characteristics of Nucleus Pulposus Cells in hPSC Differentiation Derivatives

To assess the global transcriptomes of the hPSC-derived cells, for each hPSC line (n = 3, on day 0) we used sequenced RNA (RNA-seq) isolated from cells prior to differentiation and on day 20. For the day 20 hPSC-derived cells, RNA-seq was done on cells expressing TIE2/GD2 singly and in combination (total n = 12; STAR Methods; Figure S1A). Compared with day 0, many characteristic genes previously reported to be specific to young adult NP (Risbud et al., 2015) or involved in NP development (Williams et al., 2019) showed strong upregulation patterns in the hPSC-derived cells on day 20 (Figure 4A; Figures S3 and S4; Table S2). Hierarchical clustering of the 15 *in vitro* RNA profiles revealed two clusters: one comprising the pattern for the three hPSC lines (day 0) and the other comprising the 12 day 20 differentiated samples (Figure 4B). The intra-cluster samples were 91%–97% similar, with only 50%–60% similarity between before and after differentiation, suggesting that the different hPSC lines followed closely similar differentiation paths. Overall, 1,576 genes were significantly upregulated on day 20 compared with day 0, and another 1,460 were downregulated (Figure S2C; Table S2).

Expression of pluripotency markers *POU5F1* (OCT4), *SOX2*, and *NANOG* was significantly higher in the day 0 cluster than in the differentiated cells (Figure 4A; Figure S4B), indicating that differentiation had progressed beyond the pluripotent stage. Compared with day 0, the expression of *B4GALNT1* (GD2 synthase) was increased by more than 6 times on day 20 and at higher levels in GD2+ samples than in GD2− samples, but *TEK* (TIE2) was not significantly different between TIE2+ and TIE2− samples or between day 0 and day 20 samples (Figure 4A; Figure S4C). Expression of downstream targets of the TGF-β pathway (Ito and Miyazono, 2003; Meng et al., 2016), *RUNX1/2*, *ID1/2/3/4*, *PITX2*, and *CDKN2B* (P15), was markedly upregulated in the day 20 hPSC-derived cells (Figure S4D), confirming the efficacy of our protocol.

Secretion of ECM proteins is a core role of normal human NP cells, and many efforts to establish cell markers for the NP have focused on ECM genes (the "matrisome") (Lv et al., 2014; Minogue et al., 2010a; Risbud et al., 2015). Overall, expression of 82 of the 275 core NP matrisome genes (Figures S5A and S5B), and another 148 (Figures S5C and S5D) of the 753 non-core matrisome genes, were upregulated (false discovery rate [FDR] < 0.05) in the day 20 differentiated products. Among these were 14 collagen genes, including *COL2A1*, *COL3A1*, *COL5A1/2/3*, *COL6A3*, *COL9A2*, and *COL11A1*.

Proteoglycans are essential for the homeostasis of the intervertebral disc (Iatridis et al., 2007). Compared with day 0 cells, expression of several proteoglycan genes was upregulated in the hPSC-derived cells, including proteoglycans reported in the NP such as *DCN* (decorin), *SPOCK1/2/3* (testican), *LUM* (lumican), *BGN* (biglycan), *OGN* (osteoglycin), *A2M* (alpha-2-macroglobulin), and *FMOD* (fibromodulin) (Choi et al., 2015). Notably, also included was *CHST3*, which encodes chondroitin 6-o-sulfotransferase 1 and is essential for regulating the proteoglycans by means of sulfation. Variants of *CHST3* are validated genetic risk factors for low back pain and IDD (Maxim et al., 2018; Song et al., 2013).

Many ECM glycoprotein genes expressed in the NP were upregulated in hPSC-derived cells on day 20 (Table S2). Some of these are also characteristic of chondrogenic cells, such as *MGP* and *CILP* (Seki et al., 2005; Wallin et al., 2000). Other upregulated genes encoding glycoproteins include laminin (*LAMA4*/*LAMB4*), emilins (*EMILIN2*), periostin (*POSTN*), thrombospondin (*THBS1*), cartilage oligomeric matrix protein

Figure 4. Comparative Transcriptomic Analyses of hPSC-Derived Cells before and after 20 Days of Differentiation with *In Vivo* NP Cells Reveal Similarities

(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 (fragments per kilobase of transcript per million mapped reads) (same in G). rep1, GD2−/TIE2−; rep2, GD2−/TIE2+; rep3, GD2+/TIE2−; rep4, GD2+/TIE2+.

(B) Heatmap showing the pairwise similarity among samples before and after differentiations as measured by genome-wide Pearson correlation coefficients. The samples were also clustered by hierarchical clustering on the basis of the similarity matrix.

(C) Violin and boxplot showing the increased genome-wide similarity between the hPSC-derived cells to a set of four *in vivo* NP bulk RNA-seq samples taken from surgical samples of young scoliosis patients ($p = 1.29 \times 10^{-7}$, Student's *t* test).

(D) Scatterplot of all genes before (gray; averaged over three cell lines) and after differentiations (light green, cyan, and pink; averaged over 12 replicates). The most influential genes, whose removal significantly reduced the similarities, are highlighted by an arrow pointing from expression level in the hPSCs on day 0 to expression in the hPSC-derived cells on day 20. Left arrows indicate negative influential genes, those contributing to increased similarities by having lower expression in the hPSC-derived cells; right arrows indicate positive influential genes, those contributing to increased similarities by having higher expression in the hPSC-derived cells.

(E) Same *in vitro-in vivo* comparison as in (C) but now using the influential genes identified in (D) only. The similarity now increases from 7% on day 0 to 78% on day 20 ($p = 3.9 \times 10^{-27}$).

(F) Venn diagram showing the overlaps among differentially expressed genes (DEGs) between the four *in vivo* NP and hPSCs and those between hPSC-derived NPLs and hPSCs and the positive influential genes identified in (D).

(G) 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, key for normal NP functioning.

(H) Motif enrichment analysis of the 148 genes identifies motifs and their corresponding binding factors. T, FOXA2, and SOX4 are key known transcription factors in NP development.

(I) Heatmap and bi-clustering of the 148 genes in overlap.

(*COMP*), osteopontin (*SPP1*), fibronectin (*FN1*), tenascins (*TNC*), reelin (*RELN*), and osteonectin (*SPARC*). The upregulated genes also include some modifiers of the ECM, such as *TIMP2/3* and *MMP1/2/10/28* (Lluri and Jaworski, 2005) and integrins that interact with the ECM, including *ITGA1/4/8/11*, *ITGB3*, and *NCAM1/VCAM1*.

A panel of reported NCL or NP markers showed strong upregulation in the hPSC-derived cells on day 20 (Figure 4A; Figures S3 and S4A). Carbonic anhydrase III (*CA3*) is a highly specific marker for NP (Silagi et al., 2018), expression of which increased 38.6 times (FDR = 1.12×10^{-6}) after differentiation. Galectin-3 (*LGALS3*) is a marker for young adult NP samples (Silagi et al., 2018), and its expression was increased 5.5 times (FDR = 0.036). Another NP marker, *FOXF1*, increased in all differentiated products by 8.2 times. Expression of other genes reported to be expressed in the NP, such as *CDH2* (N-cadherin) (Lv et al., 2014) and *AQP1* (Richardson et al., 2008), were also upregulated on day 20.

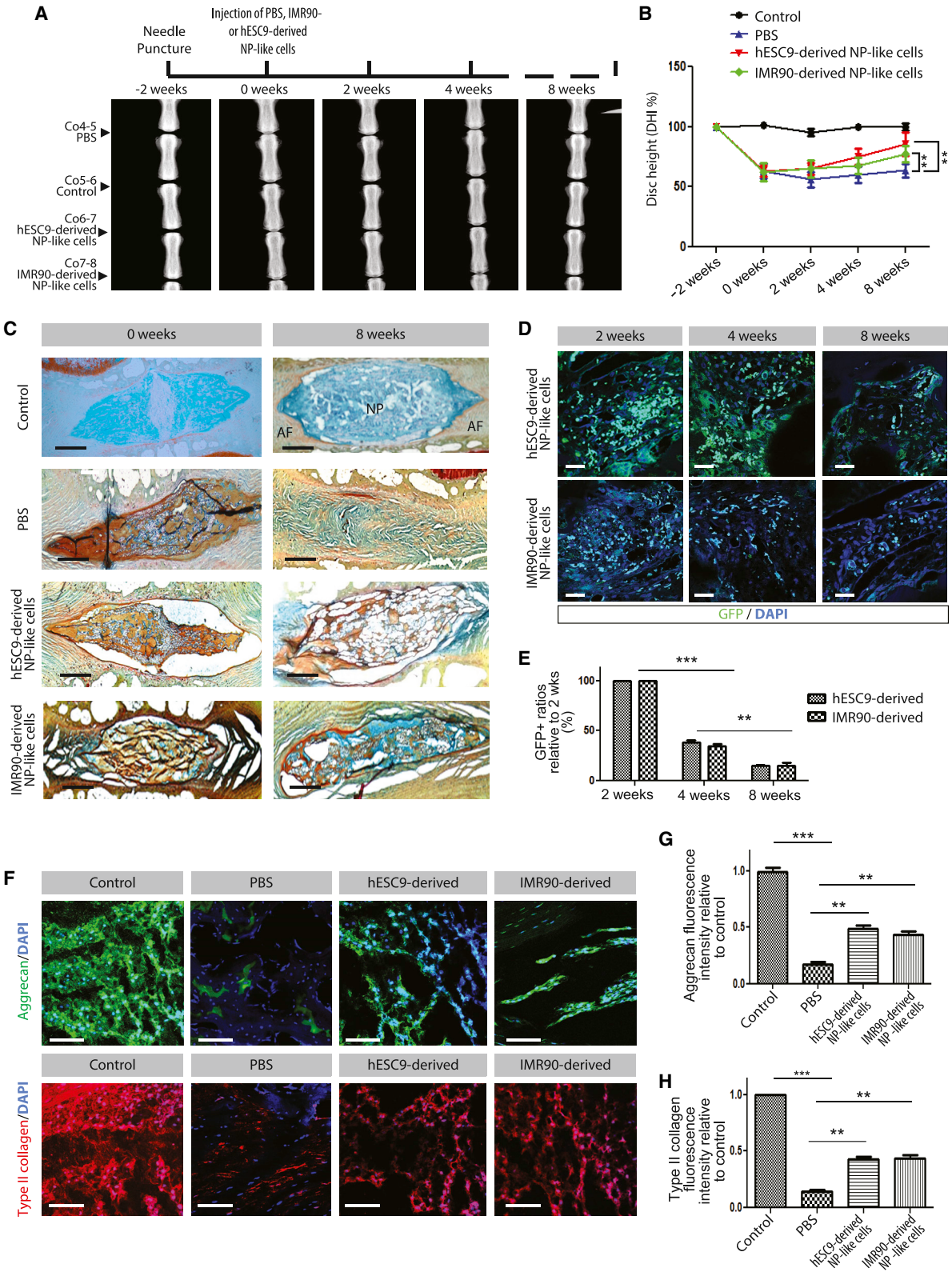
The hPSC-derived cells also showed upregulation of a variety of pathways and transcription factors. A super Venn diagram (Figure S2D) visualizes the significantly enriched pathways. A cluster of enriched terms, including matrisome, focal adhesion, integrin, Wnt (*DKK1/2*, *FZD1*, *LEF1*, *NKD1/2*, and 15 WNT-family genes), hedgehog (*GLI3*), angiogenesis, and myogenesis, were the top-ranked items, indicating that the differentiation was directed toward the targeted cell type. The expression of BMP inhibitors *NOG* (NOGGIN), *GREM1* (GREMLIN 1), and *CHRD* (CHORDIN), known markers of NP cells (Chan et al., 2015), were significantly increased in the hPSC-derived cells. Other cartilage and bone development genes that were upregulated 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 required for formation of the notochordal sheath and cell survival in the notochord, as well as development of the NP (Smits and Lefebvre, 2003). TGF-β1 and GDF5 have been reported to drive the differentiation of human adipose stromal cells toward NP-like cells (Colombier et al., 2016). AT-rich interactive domain 5b (*ARID5B*; increased by 40.6 times) is a co-factor of SOX9 (increased by 2.4 times), and it recruits PHF2 (increased by 1.6 times) to modify promoters of SOX9 targets, leading to facilitated chondrogenesis (Hata et al., 2013). Expression of *ARID5B*, *PHF2*, and *SOX9* were all significantly increased in the derived cells (FDR $q = 1.0 \times 10^{-32}$, 0.003, and 5.9×10^{-5}, respectively). These and the upregulation of *COL2A1* (21.5 times; FDR $q = 2.4 \times 10^{-28}$) may reflect the shared gene expression between CLCs and NCLs (such as *SOX9*, *COL2A1*, *COL9A2*, *COL11A1*, *ACAN*, *SOX5*, and *SOX6*) (Wang et al., 2018a). Since NCLs (evidenced by *T* and *KRTs*; Figure 4A) and CLCs are the two morphologically distinct cell that populate the young/healthy human NP (Risbud et al., 2015), these results show that our protocol for generating hPSC-derived NP-like cells is effective, and may produce cells resembling both NCLs and CLCs.

Shared Molecular Signatures between hPSC-Derived NP-like Cells and Human NP Cells *In Vivo*

To examine the identity of the hPSC-derived cells, we used two human NP cell transcriptome datasets for comparison: (1) an in-house set of bulk RNA-seq profiles from four freshly isolated NP of two adolescent/young females and an adult (aged 13, 14, and 33 years) who underwent discectomy because of scoliosis or herniation following fracture and (2) public microarray data of three "healthy" human NP from older adults (46–57 years; two male, one female) (Minogue et al., 2010a). For both datasets, the similarity between the global transcriptomes, that of hPSC-derived cells versus *in vivo* NP data, increased by about 5% from day 0 to day 20. Specifically, genome-wide similarity between hPSC-derived cell transcriptomes and that of young NP cells (the first dataset) increased from 72% at day 0 to 77% at day 20 ($p = 1.29 \times 10^{-7}$, Student's t test; Figure 4C; Figures S2E and S2F).

By using a statistical approach outlined in STAR Methods, we were able to detect "influential genes" that significantly contributed to the increase in similarity (Figure 4D). These influential genes can work either by having lower (the negative influential genes) or higher (the positive influential genes) expression in the differentiated cells, but in both cases, they help by having values closer to those in NP cells *in vivo*. Among the negative influential genes were *NANOG* and *SOX2*, which are associated with pluripotency, so their downregulation is an indication of loss of pluripotency. Among the positive influential genes were several NP-relevant genes discussed in the previous section. The expression of many of them rose from almost zero before differentiation to high levels comparable with that in native NP cells, particularly several proteoglycans (*DCN*, *LUM*, and *BGN*) and some reported NP markers (e.g., *A2M*, *SOX9*, *COL2A1*) and important regulatory genes (*CTGF* and *HAPLN1*) (Matta et al., 2017). We re-calculated the *in vitro* and *in vivo* similarities using these influential genes only and found that the similarity was 7% on day 0 and 78% on day 20, achieving a substantially larger margin of 71% compared with the 5% margin using all genes (Figure 4E).

We identified differentially expressed genes between hPSCs and human NP cells and those between hPSCs and hPSC-derived cells on day 20 (Figure 4F). The two sets of differentially expressed genes have a strong overlap with the positive influential genes (N = 148; odds ratio = 32.8) and any two sets overlapped substantially. The 148 genes (Figure 4I) were highly enriched for the Gene Ontology (GO) terms matrisome, glycoproteins, proteoglycans, and integrins (Figure 4G). Motif discovery on the these 148 genes showed enrichment for a set of known motifs bound by transcription factors known to be important for the notochord, such as T and FOXA2 (Figure 4H). Enrichment for SMAD2/4 motifs is consistent with the importance of TGF-β signaling in NP cell differentiation. Enrichment for HOXB4 motifs may reflect HOXB4 control of apoptosis (Morgan et al., 2004) and patterning signals that control RA signaling and neural induction associated with the notochord (Begemann et al., 2001; Salehi et al., 2011). The data suggest that the differentiation protocol has not only led to abundant expression of ECM genes and NP markers but has also effectively triggered and driven the regulatory mechanisms in a similar way as the *in vivo* NP. This rich dataset is available at an interactive web link (www.sbms.hku.hk/scopes/npc) as a resource to the research community.



(legend on next page)

Potential of the hPSC-Derived NP-like Cells to Attenuate Injury-Induced IDD

Several types of cells have been tested for their activity in animal models of IDD. Administration of MSCs has been reported to improve the condition of the degenerated disc by promoting expression of type II collagen and aggrecan and inhibiting the apoptosis of NP cells (Sakai et al., 2006; Yang et al., 2008). Transplantation of fat-derived MSC-like cells has been reported to promote regeneration of an intervertebral disc by stimulating differentiation into CD24-positive cells (Nakayama et al., 2017). A recent transplantation study of hiPSC-derived cells expressing COL2A1, COL1A1, ACAN, and CA12 in a similar rat disc degeneration model also demonstrated a positive effect at radiographic and histological levels up to 24 weeks post-transplantation (Xia et al., 2019).

We used an established puncture-induced rat tail disc injury model (Lv et al., 2016) to test whether the hPSC-derived NP-like cells have potential in attenuating the degenerative process and/or promote repair of the NP in the context of injury-induced IDD. To examine the survival of NP-like cells post-transplantation, NP-like cells were transduced with EGFP lentiviral vector prior to transplantation. In this model, puncture injury to coccygeal level discs (Co4–5, Co6–7, and Co7–8) resulted in a reduced disc height index (DHI) 2 weeks post-puncture, compared with the non-operated control disc (Co5–6) (Figures 5A and 5B). FAST staining showed decreased Alcian blue staining and cellular content in the NP (Figure 5C). Compared with PBS injection, transplantation of hPSC-derived NP-like cells into the punctured discs led to an increased trend of DHI by 4 weeks post-transplantation, with an effect up to 8 weeks (Figure 5B). Transplantation of hESC9-derived NP-like cells showed a trend of stronger DHI recovery than transplantation of IMR90-derived NP-like cells (Figure 5B). The cell transplantation resulted in a more preserved NP compartment, whereas PBS-injected discs were severely degenerated, with a loss of distinct NP-annulus fibrosus (AF) boundaries by 8 weeks post-treatment (Figure 5C). EGFP immuno-positivity in the punctured discs after transplantation of EGFP-labeled cells showed some of these cells could survive (Figure 5D), although EGFP-positive cells were reduced over time (Figure 5E). Compared with the non-operated control group, the expression of type II collagen and aggrecan was significantly reduced in the NP of the punctured discs at 8 weeks after cell transplantation in PBS controls, whereas these ECM proteins were still present after transplantation of hPSC-derived NP-like cells (Figures 5F–5H). These results suggest that the day 20 hPSC-derived NP-like cells contribute to facilitating repair of the NP, attenuating the injury-induced IDD.

DISCUSSION

We developed a simplified strategy to efficiently induce hPSCs to form cells with the characteristics of notochord-like and NP-like cells. In so doing, we have also provided information on the transcriptomic landscapes of NP cells *in vivo* from non-degenerated discs, which will aid an understanding of the molecular basis of notochordal differentiation. The present study is important in several aspects. First, we have successfully established a *NOTO*-EGFP reporter in an hESC line using CRISPR/Cas9-mediated genome editing. Using these NOTO-EGFP hPSCs and on the basis of the knowledge of key signaling pathways controlling notochord development, we were able to develop a compound-defined protocol and define conditions for the differentiation of notochord-like *NOTO*-EGFP-expressing cells *in vitro*, with an efficiency of approximately 14%. Second, using this approach, NCLs and NP-like cells were successfully differentiated from three independent hPSC lines. Third, global transcriptomics analysis of differentiated hPSC cells and comparison with *in vivo* transcriptome data of NP cells revealed a panel of common and distinct molecular signatures. These molecular signatures could potentially used to characterize non-degenerated NP cells. Fourth, transplantation of hPSC-derived NP-like cells can attenuate IDD in a rat model of injury-induced disc degeneration.

Cell-based therapy has emerged as a novel strategy in regenerative medicine for many tissues and organs, including IDD (Maidhof et al., 2017; Zhang et al., 2015). Several studies have highlighted the strong potential of NP-like cells in regeneration of intervertebral discs (Perez-Cruet et al., 2019; Zhou et al., 2018). Because no cells are the precursors of NP cells, differentiating to NCLs was the objective of the first step in the present study. Our protocol illustrates the benefits of exploiting knowledge of early embryonic development and has several key features. We cultured the hPSC cell lines in basal medium with Activin A for 2 days to stimulate mesoderm differentiation. Takada et al. (1994) found that Wnt3a−/− mice presented a disruption of notochord development, and CHIR99021 can promote hESC differentiation to mesoderm (Gouti et al., 2014). We replaced DKK1 used in the differentiation protocol for mouse notochord with a chemical inhibitor of Wnt signaling, CHIR99021, in our human differentiation protocol. This resulted in the activation of key transcription factors (*FOXA2*, *T*, and *NOTO*) known to be essential for notochordal development.

Figure 5. Transplantation of hPSC-Derived NP-like Cells Attenuates Disc Injury in a Rat Model of IDD

(A) 2 weeks after intervertebral disc injury, hPSC-derived NP-like cells or PBS was 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.

(C) Representative images of FAST staining showing a damaged intervertebral disc at 8 weeks after cell transplantation. Scale bars: 500 μm.

(D) Representative immunostaining images for EGFP protein showing the survival of hPSC-derived NP-like cells at 2, 4, and 8 weeks after cell injection. Scale bars: 50 μm.

(E) The cell density of hPSC-derived NP-like cells at 2, 4, and 8 weeks post-transplantation.

(F) Representative immunohistochemical images for type II collagen and aggrecan among the different experimental groups, respectively.

(G and H) The fluorescence intensity of staining for type II collagen (H) and aggrecan (G) among the different experimental groups, respectively.

In (B), (E), (G), and (H), error bars indicate mean ± SEM; n = 6 rats; $^{**}p < 0.01$ and $^{***}p < 0.001$ by one-way ANOVA, followed by Bonferroni corrections.

Importantly, FOXA2 and T proteins were co-expressed in NCLs, genes that are functionally conserved in the specification and maintenance of notochordal cells (Ang and Rossant, 1994; Rashbass et al., 1991). Previously, Winzi et al. (2011) reported a protocol for differentiating mouse notochordal cells using *Noto*-EGFP/+ mouse embryonic stem cells (mESCs), obtaining ~10% EGFP+ cells that expressed notochordal genes *Noto*, *Foxa2*, *Shh*, *Noggin*, *Chordin*, and *T*. To maintain *NOTO-EGFP*-expressing cells is challenging *in vitro*. It has been reported that *NOTO-EGFP*-expressing cells could not be maintained, which could be due to the transient nature of NOTO expression (Abdelkhalek et al., 2004) in notochord development and/or could reflect the need for additional conditions for holding notochordal cells at that particular differentiation stage. However, the efficacy of our protocol in activating the expression of key transcription factors known to be important for notochord development in the mouse implicates conservation of molecular pathways and suggests that lessons learned from developmental genetic studies in mice may be applied for future improvements of the current human protocol.

Extending the culture of these cells for another 15 days and adding TGF-β3 was a key step for successful differentiation to an NP cell-like phenotype, with the characteristic vacuolated cell morphology. A previous study showed that a subset of NP cells with progenitor properties express GD2 and TIE2 (an endothelial cell marker) (Sakai et al., 2012). Although we found that TIE2 and GD2 were expressed in some NP-like cells, we found no difference in differential RNA expression profiles between TIE2/GD2-positive and TIE2/GD2-negative populations, suggesting that TIE2/GD2 are expressed in NP-like cells but do not define NP-like cells. Recently a three-step protocol was described for the differentiation of NP-like cells, with some similarities to our protocol in the factors added (e.g., TGF-β-3) but also differences such as the use of an exogenous T-GFP reporter for monitoring differentiation, transfer to 3D conditions and a longer culture period (Tang et al., 2018). In another study (Xia et al., 2019), a three-step protocol of differentiation spanning 30 days was described: starting with the addition of Activin A, WNT3A, and FGF2, followed by NT4, FOLLISTATIN, and BMP4, and finally supplementation with TGF-β1, GDF5, and BMP2 was used for differentiation to NP-like cells. The expression of COL2A1, COL1A1, ACAN, and CA12 was used as indicators of NP identity. Our protocol differs in the inhibition of RA signaling and the use of NOGGIN to inhibit BMP signaling in the first phase followed by promoting TGF-β signaling in the second phase. More important, we used Noto-GFP knockin reporter to trace NCLs and then induce to NP-like cells. Tang et al. (2018) used a transfected T-GFP reporter. Although T is expressed in the notochord, it is also expressed in mesoderm and hence is less specific than NOTO. However, it is difficult to compare the three protocols, because information on the efficiencies of differentiation and the global transcriptomic signatures of the derived NP-like cells compared with *in vivo* NP cells in those two other studies was not reported. To our knowledge, our study is the first demonstration of the identity of *in vitro*-differentiated NP-like cells that benchmarks to *in vivo* human transcriptome data.

A key outcome of our transcriptome analyses is an expanded, *in vivo*-relevant characterization of the NP cell molecular signature. Our study presents a comprehensive assessment of the NP-like identity of the hPSC-derived cells. Transcriptomic analyses revealed many NP markers and regulators that were strongly upregulated in the hPSC differentiated cells. The transcriptomes of the hPSCs became more similar to those of *in vivo* human NP cells after differentiation. We identified 148 genes as major contributors to this process, and they showed strong relevance to the known biology of NP. However, compared with adolescent or young *in vivo* NP, the genome-wide similarity of the hPSCs was 78%. The degree of similarity could be influenced by heterogeneity in both *in vivo* NP and *in vitro* hPSC-derived cell population. Heterogeneity is inherent because the *in vivo* NP cells include both NCLs and CLCs. The similarity coefficient may also be affected, because although the *in vivo* NP was not degenerated, the discs were taken from individuals with scoliosis or burst fracture.

We started with three types of hPSCs, and although they appeared to have differentiated in similar directions, expression of some reported NCL markers, particularly *T*, *Keratin 8*, *18*, and *19*, and *SHH*, was lower in cells derived from hiPSCs (IMR90) than in those derived from hESCs (hESC3/9). This variation may be due to asynchronous differentiation in the different hPSC lines. However, expression of major CLC markers (e.g., *COL2A1*, *SOX9*; Sive et al., 2002) was comparable among cells derived from the three lines. Previously, CD24 has been used to isolate NP-like cells from differentiating mouse iPSCs (Chen et al., 2013) and to isolate human fetal NP cells. Although the hPSC-derived cells expressed CD24, we did not find differential expression, as it was highly expressed in the starting hPSCs. Because CD24 is also expressed by many types of progenitor and cancer stem cells (Al-Hajj et al., 2003; Sagrinati et al., 2006), this marker may not definitively characterize notochordal-derivatives.

The hPSC-derived NCLs also displayed characteristic activity expected of the notochord. The embryonic notochord produces SHH and serves to induce signaling in the surrounding tissues. Consistent with their notochordal property, we found that the hPSC-derived NP-like cells produced SHH that could activate the HH reporter in co-culture assays. The functional capacity of the hPSC-derived NP-like cells is further indicated by their therapeutic effect in a rat model of injury-induced IDD. Our results show that transplantation of hPSC-derived NP-like cells could attenuate the impact of injury-induced IDD, as judged by the partial recovery of disc height and presence of a type II collagen and aggrecan enriched matrix. This attenuation could be the outcome of ameliorated degenerative changes or enhanced regenerative activity.

Similar to earlier reports of cell transplantation studies in rat IDD models (Allon et al., 2010; Crevensten et al., 2004; Wei et al., 2009), the number of hPSC-derived cells decreased with time after transplantation into the discs. This might be due to the influence of the degenerative milieu, such as inflammatory or mechanical factors that affect their viability, or because the limited nutritional supply in the discs cannot adequately support the activity of the transplanted cells.

Interestingly, more hESC9-derived cells persisted compared with those derived from hiPSCs (IMR90), which had displayed less efficient NP-like differentiation. In the future, it would be interesting to make a comparison of the protective or reparative effects of non-differentiated hPSCs and differentiated cells from different hPSCs and determine whether the hPSC-derived NP-like cells perform better than other types of cells, such as MSCs or ESCs. It is also noteworthy that among the 148 NP genes we identified, several have variants that are reported to be genetic risk factors associated with IDD and/or low back pain, such as *COL9A2*, *COL11A1*, *CHST3*, *THBS1*, *ACAN*, and *SOX5* (Annunen et al., 1999; Dornand et al., 1987; Jim et al., 2005; Kawaguchi et al., 1999; Martirosyan et al., 2016; Noponen-Hietala et al., 2003; Song et al., 2013; Suri et al., 2018). The NP cell molecular signatures we have detected will be a valuable resource for testing for variants that may contribute to genetic risk for IDD. In summary, our study shows that NP-like cells can be effectively and efficiently derived from hPSCs using a specific differentiation protocol. These cells provide an important resource for the understanding of disc cell biology and the genetic contribution to degenerative changes, with potential for the development of an NP cell-based treatment for IDD.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Rats
  - Human NP samples
  - Cell cultures
- METHOD DETAILS
  - Cell differentiation
  - Immunostaining
  - Quantitative PCR
  - Flow cytometry
  - RNA sequencing and processing
  - Analyses of differentially expressed genes (DEGs)
  - *In vitro – in vivo* comparisons and detection of influential NP-specific markers
  - Colony-forming assay
  - Vector construction
  - eGFP knock-in
  - Rat model of IDD and cell transplantation
  - Radiographic evaluation
  - Histological evaluation and immunohistochemistry
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.01.100.

ACKNOWLEDGMENTS

We thank Professor Keith Luk for his help in obtaining human disc tissues and helpful discussion and Professor Patrick Tam for advice and critical feedback. This research was supported by grants from the Hong Kong Research Grants Council (T12-708/12-N and HKU17120517), the National Science Foundation of China (NSFC 31571407), and an AOSpine East Asia Research Grant (AR160008).

AUTHOR CONTRIBUTIONS

Y.Z. developed the hPSC differentiation protocol and performed molecular and animal experiments. Z.Z. developed the NOTO-EGFP reporter hPSC cell line and performed molecular experiments. P.C. performed bioinformatics analyses, developed statistical approaches and web interfaces, and was in charge of figure preparation. Y.Z., P.C., and Z.Z. prepared the figures and wrote the original manuscript. K.M.C.C., V.T., and T.Y.K.A. collected and processed human disc samples. T.Y.K.A., Y.P., C.L., R.W., C.Y.M., and V.Y.L. assisted in transcriptome data acquisition, *in vivo* experiments, and data interpretation. D.C., P.S., and H.T. provided expertise and critical feedback and edited the manuscript. K.S.E.C., V.Y.L., and Q.L rewrote and finalized the manuscript. Q.L. and K.S.E.C. jointly supervised experiments, their design, data analyses, and interpretation. K.S.E.C. and Q.L. acquired funding support. Q.L. conceived, designed, and supervised the strategy of hPSC differentiation.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 14, 2018
Revised: October 15, 2019
Accepted: January 28, 2020
Published: February 25, 2020

REFERENCES

Abdelkhalek, H.B., Beckers, A., Schuster-Gossler, K., Pavlova, M.N., Burkhardt, H., Lickert, H., Rossant, J., Reinhardt, R., Schalkwyk, L.C., Müller, I., et al. (2004). The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation. Genes Dev. *18*, 1725–1736.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. U S A *100*, 3983–3988.

Allon, A.A., Aurouer, N., Yoo, B.B., Liebenberg, E.C., Buser, Z., and Lotz, J.C. (2010). Structured coculture of stem cells and disc cells prevent disc degeneration in a rat model. Spine J. *10*, 1089–1097.

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.

Ang, S.L., and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. Cell *78*, 561–574.

Annunen, S., Paassilta, P., Lohiniva, J., Perälä, M., Pihlajamaa, T., Karppinen, J., Tervonen, O., Kröger, H., Lähde, S., Vanharanta, H., et al. (1999). An allele of COL9A2 associated with intervertebral disc disease. Science *285*, 409–412.

Barrionuevo, F., Taketo, M.M., Scherer, G., and Kispert, A. (2006). Sox9 is required for notochord maintenance in mice. Dev. Biol. *295*, 128–140.

Beckers, A., Alten, L., Viebahn, C., Andre, P., and Gossler, A. (2007). The mouse homeobox gene Noto regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. Proc. Natl. Acad. Sci. U S A *104*, 15765–15770.

Begemann, G., Schilling, T.F., Rauch, G.J., Geisler, R., and Ingham, P.W. (2001). The zebrafish neckless mutation reveals a requirement for raldh2 in mesodermal signals that pattern the hindbrain. Development *128*, 3081–3094.

Cappello, R., Bird, J.L., Pfeiffer, D., Bayliss, M.T., and Dudhia, J. (2006). Notochordal cell produce and assemble extracellular matrix in a distinct manner, which may be responsible for the maintenance of healthy nucleus pulposus. Spine *31*, 873–882, discussion 883.

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.

Chen, J., Lee, E.J., Jing, L., Christoforou, N., Leong, K.W., and Setton, L.A. (2013). Differentiation of mouse induced pluripotent stem cells (iPSCs) into nucleus pulposus-like cells in vitro. PLoS ONE *8*, e75548.

Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H., and Beachy, P.A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature *383*, 407–413.

Choi, K.S., and Harfe, B.D. (2011). Hedgehog signaling is required for formation of the notochord sheath and patterning of nuclei pulposi within the intervertebral discs. Proc. Natl. Acad. Sci. U S A *108*, 9484–9489.

Choi, K.S., Cohn, M.J., and Harfe, B.D. (2008). Identification of nucleus pulposus precursor cells and notochordal remnants in the mouse: implications for disk degeneration and chordoma formation. Dev. Dyn. *237*, 3953–3958.

Choi, H., Johnson, Z.I., and Risbud, M.V. (2015). Understanding nucleus pulposus cell phenotype: a prerequisite for stem cell based therapies to treat intervertebral disc degeneration. Curr. Stem Cell Res. Ther. *10*, 307–316.

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-β1 and GDF5 act synergistically to drive the differentiation of human adipose stromal cells toward nucleus pulposus-like cells. Stem Cells *34*, 653–667.

Crevensten, G., Walsh, A.J., Ananthakrishnan, D., Page, P., Wahba, G.M., Lotz, J.C., and Berven, S. (2004). Intervertebral disc cell therapy for regeneration: mesenchymal stem cell implantation in rat intervertebral discs. Ann. Biomed. Eng. *32*, 430–434.

de Bree, K., de Bakker, B.S., and Oostra, R.J. (2018). The development of the human notochord. PLoS ONE *13*, e0205752.

Dornand, J., Sekkat, C., Mani, J.C., and Gerber, M. (1987). Lipoxygenase inhibitors suppress IL-2 synthesis: relationship with rise of [Ca++]i and the events dependent on protein kinase C activation. Immunol. Lett. *16*, 101–106.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417–1430.

Fan, C.M., and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. Cell *79*, 1175–1186.

Fatemieh, S.O. (2006). A review of: "Discovering Knowledge in Data: An Introduction to Data Mining.". J. Biopharm. Stat. *16*, 127–130.

Fujita, N., Miyamoto, T., Imai, J., Hosogane, N., Suzuki, T., Yagi, M., Morita, K., Ninomiya, K., Miyamoto, K., Takaishi, H., et al. (2005). CD24 is expressed specifically in the nucleus pulposus of intervertebral discs. Biochem. Biophys. Res. Commun. *338*, 1890–1896.

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet *392*, 1789–1858.

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.

Happey, F., Johnson, A.G., Naylor, A., and Turner, R.L. (1964). Preliminary observations concerning the fine structure of the intervertebral disc. J. Bone Joint Surg. Br. *46*, 563–567.

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.

Hata, K., Takashima, R., Amano, K., Ono, K., Nakanishi, M., Yoshida, M., Wakabayashi, M., Matsuda, A., Maeda, Y., Suzuki, Y., et al. (2013). Arid5b facilitates chondrogenesis by recruiting the histone demethylase Phf2 to Sox9-regulated genes. Nat. Commun. *4*, 2850.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Henriksson, H.B., Svanvik, T., Jonsson, M., Hagman, M., Horn, M., Lindahl, A., and Brisby, H. (2009). Transplantation of human mesenchymal stems cells into intervertebral discs in a xenogeneic porcine model. Spine *34*, 141–148.

Herrmann, B.G., and Kispert, A. (1994). The T genes in embryogenesis. Trends Genet. *10*, 280–286.

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 *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.

Jim, J.J., Noponen-Hietala, N., Cheung, K.M., Ott, J., Karppinen, J., Sahraravand, A., Luk, K.D., Yip, S.P., Sham, P.C., Song, Y.Q., et al. (2005). The TRP2 allele of COL9A2 is an age-dependent risk factor for the development and severity of intervertebral disc degeneration. Spine *30*, 2735–2742.

Johnson, R.L., Laufer, E., Riddle, R.D., and Tabin, C. (1994). Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. Cell *79*, 1165–1173.

Kawaguchi, Y., Osada, R., Kanamori, M., Ishihara, H., Ohmori, K., Matsui, H., and Kimura, T. (1999). Association between an aggrecan gene polymorphism and lumbar disc degeneration. Spine *24*, 2456–2460.

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

Kuroda, H., Wessely, O., and De Robertis, E.M. (2004). Neural induction in Xenopus: requirement for ectodermal and endomesodermal signals via Chordin, Noggin, beta-Catenin, and Cerberus. PLoS Biol. *2*, E92.

Lawson, L.Y., and Harfe, B.D. (2017). Developmental mechanisms of intervertebral disc and vertebral column formation. Wiley Interdiscip. Rev. Dev. Biol. *6*.

Leung, V.Y., Chan, W.C., Hung, S.C., Cheung, K.M., and Chan, D. (2009). Matrix remodeling during intervertebral disc growth and degeneration detected by multichromatic FAST staining. J. Histochem. Cytochem. *57*, 249–256.

Leung, V.Y., Aladin, D.M., Lv, F., Tam, V., Sun, Y., Lau, R.Y., Hung, S.C., Ngan, A.H., Tang, B., Lim, C.T., et al. (2014). Mesenchymal stem cells reduce intervertebral disc fibrosis and facilitate repair. Stem Cells *32*, 2164–2177.

Li, Y.Y., Diao, H.J., Chik, T.K., Chow, C.T., An, X.M., Leung, V., Cheung, K.M., and Chan, B.P. (2014). Delivering mesenchymal stem cells in collagen microsphere carriers to rabbit degenerative disc: reduced risk of osteophyte formation. Tissue Eng. Part A *20*, 1379–1391.

Liang, X., Ding, Y., Zhang, Y., Chai, Y.H., He, J., Chiu, S.M., Gao, F., Tse, H.F., and Lian, Q. (2015). Activation of NRG1-ERBB4 signaling potentiates mesenchymal stem cell-mediated myocardial repairs following myocardial infarction. Cell Death Dis. *6*, e1765.

Lindbäck, Y., Tropp, H., Enthoven, P., Abbott, A., and Öberg, B. (2018). PREPARE: presurgery physiotherapy for patients with degenerative lumbar spine disorder: a randomized controlled trial. Spine J. *18*, 1347–1355.

Lluri, G., and Jaworski, D.M. (2005). Regulation of TIMP-2, MT1-MMP, and MMP-2 expression during C2C12 differentiation. Muscle Nerve *32*, 492–499.

Lustig, K.D., Kroll, K., Sun, E., Ramos, R., Elmendorf, H., and Kirschner, M.W. (1996). A Xenopus nodal-related gene that acts in synergy with noggin to induce complete secondary axis and notochord formation. Development *122*, 3275–3282.

Lv, F., Leung, V.Y., Huang, S., Huang, Y., Sun, Y., and Cheung, K.M. (2014). In search of nucleus pulposus-specific molecular markers. Rheumatology (Oxford) *53*, 600–610.

Lv, F.J., Peng, Y., Lim, F.L., Sun, Y., Lv, M., Zhou, L., Wang, H., Zheng, Z., Cheung, K.M.C., and Leung, V.Y.L. (2016). Matrix metalloproteinase 12 is an indicator of intervertebral disc degeneration co-expressed with fibrotic markers. Osteoarthritis Cartilage *24*, 1826–1836.

Maaten, L.J.P.v.d., and Hinton, G.E. (2008). Visualizing high-dimensional data using t-SNE. J. Machine Learning Res. *9*, 2579–2605.

Maidhof, R., Rafiuddin, A., Chowdhury, F., Jacobsen, T., and Chahine, N.O. (2017). Timing of mesenchymal stem cell delivery impacts the fate and therapeutic potential in intervertebral disc repair. J. Orthop. Res. *35*, 32–40.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Martirosyan, N.L., Patel, A.A., Carotenuto, A., Kalani, M.Y., Belykh, E., Walker, C.T., Preul, M.C., and Theodore, N. (2016). Genetic alterations in intervertebral disc disease. Front. Surg. *3*, 59.

Matta, A., Karim, M.Z., Isenman, D.E., and Erwin, W.M. (2017). Molecular therapy for degenerative disc disease: clues from secretome analysis of the notochordal cell-rich nucleus pulposus. Sci. Rep. *7*, 45623.

Maxim, B.F., Yakov, A.T., Melody, P., Lennart, K., Group, C.M.W., Pradeep, S., Yurii, S.A., and Frances, M.K.W. (2018). Genome-Wide Association Summary Statistics For Back Pain (Zenodo).

McCann, M.R., Tamplin, O.J., Rossant, J., and Séguin, C.A. (2012). Tracing notochord-derived cells using a Noto-cre mouse: implications for intervertebral disc development. Dis. Model. Mech. *5*, 73–82.

Meng, X.M., Nikolic-Paterson, D.J., and Lan, H.Y. (2016). TGF-β: the master regulator of fibrosis. Nat. Rev. Nephrol. *12*, 325–338.

Minogue, B.M., Richardson, S.M., Zeef, L.A., Freemont, A.J., and Hoyland, J.A. (2010a). Characterization of the human nucleus pulposus cell phenotype and evaluation of novel marker gene expression to define adult stem cell differentiation. Arthritis Rheum. *62*, 3695–3705.

Minogue, B.M., Richardson, S.M., Zeef, L.A., Freemont, A.J., and Hoyland, J.A. (2010b). Transcriptional profiling of bovine intervertebral disc cells: implications for identification of normal and degenerate human intervertebral disc cell phenotypes. Arthritis Res. Ther. *12*, R22.

Mirza, S.K., and Deyo, R.A. (2007). Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. Spine *32*, 816–823.

Mohanty, S., Pinelli, R., Pricop, P., Albert, T.J., and Dahia, C.L. (2019). Chondrocyte-like nested cells in the aged intervertebral disc are late-stage nucleus pulposus cells. Aging Cell *18*, e13006.

Morgan, R., Nalliah, A., and Morsi El-Kadi, A.S. (2004). FLASH, a component of the FAS-CAPSASE8 apoptotic pathway, is directly regulated by Hoxb4 in the notochord. Dev. Biol. *265*, 105–112.

Nakayama, E., Matsumoto, T., Kazama, T., Kano, K., and Tokuhashi, Y. (2017). Transplantation of dedifferentiation fat cells promotes intervertebral disc regeneration in a rat intervertebral disc degeneration model. Biochem. Biophys. Res. Commun. *493*, 1004–1009.

Noponen-Hietala, N., Kyllönen, E., Männikkö, M., Ilkko, E., Karppinen, J., Ott, J., and Ala-Kokko, L. (2003). Sequence variations in the collagen IX and XI genes are associated with degenerative lumbar spinal stenosis. Ann. Rheum. Dis. *62*, 1208–1214.

Nurunnabi, A.A.M., Nasser, M., and Imon, A.H.M.R. (2016). Identification and classification of multiple outliers, high leverage points and influential observations in linear regression. J. Appl. Stat. *43*, 509–525.

Okubo, T., Nagoshi, N., Kohyama, J., Tsuji, O., Shinozaki, M., Shibata, S., Kase, Y., Matsumoto, M., Nakamura, M., and Okano, H. (2018). Treatment with a gamma-secretase inhibitor promotes functional recovery in human iPSC-derived transplants for chronic spinal cord injury. Stem Cell Reports *11*, 1416–1432.

Othman, Y.A., Verma, R., and Qureshi, S.A. (2019). Artificial disc replacement in spine surgery. Ann. Transl. Med. *7* (*Suppl 5*), S170.

Pauklin, S., and Vallier, L. (2015). Activin/Nodal signalling in stem cells. Development *142*, 607–619.

Perez-Cruet, M., Beeravolu, N., McKee, C., Brougham, J., Khan, I., Bakshi, S., and Chaudhry, G.R. (2019). Potential of human nucleus pulposus-like cells derived from umbilical cord to treat degenerative disc disease. Neurosurgery *84*, 272–283.

Rashbass, P., Cooke, L.A., Herrmann, B.G., and Beddington, R.S. (1991). A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras. Nature *353*, 348–351.

Richardson, S.M., Knowles, R., Marples, D., Hoyland, J.A., and Mobasheri, A. (2008). Aquaporin expression in the human intervertebral disc. J. Mol. Histol. *39*, 303–309.

Risbud, M.V., Schoepflin, Z.R., Mwale, F., Kandel, R.A., Grad, S., Iatridis, J.C., Sakai, D., and Hoyland, J.A. (2015). Defining the phenotype of young healthy nucleus pulposus cells: recommendations of the Spine Research Interest Group at the 2014 annual ORS meeting. J. Orthop. Res. *33*, 283–293.

Roberts, S., Evans, H., Trivedi, J., and Menage, J. (2006). Histology and pathology of the human intervertebral disc. J. Bone Joint Surg. Am. *88* (*Suppl 2*), 10–14.

Rodrigues-Pinto, R., Richardson, S.M., and Hoyland, J.A. (2014). An understanding of intervertebral disc development, maturation and cell phenotype provides clues to direct cell-based tissue regeneration therapies for disc degeneration. Eur. Spine J. *23*, 1803–1814.

Rodrigues-Pinto, R., Berry, A., Piper-Hanley, K., Hanley, N., Richardson, S.M., and Hoyland, J.A. (2016). Spatiotemporal analysis of putative notochordal cell markers reveals CD24 and keratins 8, 18, and 19 as notochord-specific markers during early human intervertebral disc development. J. Orthop. Res. *34*, 1327–1340.

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

Sagrinati, C., Netti, G.S., Mazzinghi, B., Lazzeri, E., Liotta, F., Frosali, F., Ronconi, E., Meini, C., Gacci, M., Squecco, R., et al. (2006). Isolation and characterization of multipotent progenitor cells from the Bowman's capsule of adult human kidneys. J. Am. Soc. Nephrol. *17*, 2443–2456.

Sakai, D., Mochida, J., Iwashina, T., Hiyama, A., Omi, H., Imai, M., Nakai, T., Ando, K., and Hotta, T. (2006). Regenerative effects of transplanting mesenchymal stem cells embedded in atelocollagen to the degenerated intervertebral disc. Biomaterials *27*, 335–345.

Sakai, D., Nakamura, Y., Nakai, T., Mishima, T., Kato, S., Grad, S., Alini, M., Risbud, M.V., Chan, D., Cheah, K.S., et al. (2012). Exhaustion of nucleus pulposus progenitor cells with ageing and degeneration of the intervertebral disc. Nat. Commun. *3*, 1264.

Salehi, H., Karbalaie, K., Razavi, S., Tanhaee, S., Nematollahi, M., Sagha, M., Nasr-Esfahani, M.H., and Baharvand, H. (2011). Neuronal induction and regional identity by co-culture of adherent human embryonic stem cells with chicken notochords and somites. Int. J. Dev. Biol. *55*, 321–326.

Seki, S., Kawaguchi, Y., Chiba, K., Mikami, Y., Kizawa, H., Oya, T., Mio, F., Mori, M., Miyamoto, Y., Masuda, I., et al. (2005). A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. Nat. Genet. *37*, 607–612.

Sheyn, D., Ben-David, S., Shapiro, G., De Mel, S., Bez, M., Ornelas, L., Sahabian, A., Sareen, D., Da, X., Pelled, G., et al. (2016). Human induced pluripotent stem cells differentiate into functional mesenchymal stem cells and repair bone defects. Stem Cells Transl. Med. *5*, 1447–1460.

Silagi, E.S., Batista, P., Shapiro, I.M., and Risbud, M.V. (2018). Expression of carbonic anhydrase III, a nucleus pulposus phenotypic marker, is hypoxia-responsive and confers protection from oxidative stress-induced cell death. Sci. Rep. *8*, 4856.

Sive, J.I., Baird, P., Jeziorsk, M., Watkins, A., Hoyland, J.A., and Freemont, A.J. (2002). Expression of chondrocyte markers by cells of normal and degenerate intervertebral discs. Mol. Pathol. *55*, 91–97.

Smits, P., and Lefebvre, V. (2003). Sox5 and Sox6 are required for notochord extracellular matrix sheath formation, notochord cell survival and development of the nucleus pulposus of intervertebral discs. Development *130*, 1135–1148.

Smits, P., Dy, P., Mitra, S., and Lefebvre, V. (2004). Sox5 and Sox6 are needed to develop and maintain source, columnar, and hypertrophic chondrocytes in the cartilage growth plate. J. Cell Biol. *164*, 747–758.

Song, Y.Q., Karasugi, T., Cheung, K.M., Chiba, K., Ho, D.W., Miyake, A., Kao, P.Y., Sze, K.L., Yee, A., Takahashi, A., et al. (2013). Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant. J. Clin. Invest. *123*, 4909–4917.

Stein, S., and Kessel, M. (1995). A homeobox gene involved in node, notochord and neural plate formation of chick embryos. Mech. Dev. *49*, 37–48.

Stemple, D.L. (2005). Structure and function of the notochord: an essential organ for chordate development. Development *132*, 2503–2512.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545–15550.

Suri, P., Palmer, M.R., Tsepilov, Y.A., Freidin, M.B., Boer, C.G., Yau, M.S., Evans, D.S., Gelemanovic, A., Bartz, T.M., Nethander, M., et al. (2018). Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. PLoS Genet. *14*, e1007601.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat. Biotechnol. *22*, 589–594.

Taipale, J., Chen, J.K., Cooper, M.K., Wang, B., Mann, R.K., Milenkovic, L., Scott, M.P., and Beachy, P.A. (2000). Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. Nature *406*, 1005–1009.

Takada, S., Stark, K.L., Shea, M.J., Vassileva, G., McMahon, J.A., and McMahon, A.P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes Dev. *8*, 174–189.

Tang, X., Jing, L., Richardson, W.J., Isaacs, R.E., Fitch, R.D., Brown, C.R., Erickson, M.M., Setton, L.A., and Chen, J. (2016). Identifying molecular phenotype of nucleus pulposus cells in human intervertebral disc with aging and degeneration. J. Orthop. Res. *34*, 1316–1326.

Tang, R., Jing, L., Willard, V.P., Wu, C.L., Guilak, F., Chen, J., and Setton, L.A. (2018). Differentiation of human induced pluripotent stem cells into nucleus pulposus-like cells. Stem Cell Res. Ther. *9*, 61.

ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E., and Nusse, R. (2008). Wnt signaling mediates self-organization and axis formation in embryoid bodies. Cell Stem Cell *3*, 508–518.

Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875–889.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

van Ooij, A., Kurtz, S.M., Stessels, F., Noten, H., and van Rhijn, L. (2007). Polyethylene wear debris and long-term clinical failure of the Charité disc prosthesis: a study of 4 patients. Spine *32*, 223–229.

von Dassow, G., Schmidt, J.E., and Kimelman, D. (1993). Induction of the Xenopus organizer: expression and regulation of Xnot, a novel FGF and activin-regulated homeo box gene. Genes Dev. *7*, 355–366.

Wallin, R., Cain, D., Hutson, S.M., Sane, D.C., and Loeser, R. (2000). Modulation of the binding of matrix Gla protein (MGP) to bone morphogenetic protein-2 (BMP-2). Thromb. Haemost. *84*, 1039–1044.

Wang, J., Lu, J., Bond, M.C., Chen, M., Ren, X.R., Lyerly, H.K., Barak, L.S., and Chen, W. (2010). Identification of select glucocorticoids as Smoothened agonists: potential utility for regenerative medicine. Proc. Natl. Acad. Sci. U S A *107*, 9323–9328.

Wang, F., Zhang, C., Shi, R., Xie, Z.Y., Chen, L., Wang, K., Wang, Y.T., Xie, X.H., and Wu, X.T. (2018a). The embryonic and evolutionary boundaries between notochord and cartilage: a new look at nucleus pulposus-specific markers. Osteoarthritis Cartilage *26*, 1274–1282.

Wang, W., Wang, Y., Deng, G., Ma, J., Huang, X., Yu, J., Xi, Y., and Ye, X. (2018b). 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.

Wei, A., Tao, H., Chung, S.A., Brisby, H., Ma, D.D., and Diwan, A.D. (2009). The fate of transplanted xenogeneic bone marrow-derived stem cells in rat intervertebral discs. J. Orthop. Res. *27*, 374–379.

Wei, J.N., Cai, F., Wang, F., Wu, X.T., Liu, L., Hong, X., and Tang, W.H. (2016). Transplantation of CXCR4 overexpressed mesenchymal stem cells augments regeneration in degenerated intervertebral discs. DNA Cell Biol. *35*, 241–248.

Weiler, C., Nerlich, A.G., Schaaf, R., Bachmeier, B.E., Wuertz, K., and Boos, N. (2010). Immunohistochemical identification of notochordal markers in cells in the aging human lumbar intervertebral disc. Eur. Spine J. *19*, 1761–1770.

Williams, S., Alkhatib, B., and Serra, R. (2019). Development of the axial skeleton and intervertebral disc. Curr. Top. Dev. Biol. *133*, 49–90.

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.

Xia, K., Zhu, J., Hua, J., Gong, Z., Yu, C., Zhou, X., Wang, J., Huang, X., Yu, W., Li, L., et al. (2019). Intradiscal injection of induced pluripotent stem cell-derived nucleus pulposus-like cell-seeded polymeric microspheres promotes rat disc regeneration. Stem Cells Int. *2019*, 6806540.

Yamanaka, Y., Tamplin, O.J., Beckers, A., Gossler, A., and Rossant, J. (2007). Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. Dev. Cell *13*, 884–896.

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 *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, Y., Tao, H., Gu, T., Zhou, M., Jia, Z., Jiang, G., Chen, C., Han, Z., Xu, C., Wang, D., et al. (2015). The effects of human Wharton's jelly cell transplantation on the intervertebral disc in a canine disc degeneration model. Stem Cell Res. Ther. *6*, 154.

Zhou, X., Tao, Y., Liang, C., Zhang, Y., Li, H., and Chen, Q. (2015). BMP3 alone and together with TGF-β promote the differentiation of human mesenchymal stem cells into a nucleus pulposus-like phenotype. Int. J. Mol. Sci. *16*, 20344–20359.

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. Published online June 11, 2018. https://doi.org/10.1096/fj.201800373R.

Zhu, J., Kwan, K.M., and Mackem, S. (2016). Putative oncogene Brachyury (T) is essential to specify cell fate but dispensable for notochord progenitor proliferation and EMT. Proc. Natl. Acad. Sci. U S A *113*, 3820–3825.

Zhu, Y., Liang, Y., Zhu, H., Lian, C., Wang, L., Wang, Y., Gu, H., Zhou, G., and Yu, X. (2017). The generation and functional characterization of induced pluripotent stem cells from human intervertebral disc nucleus pulposus cells. Oncotarget *8*, 42700–42711.

STAR★METHODS

KEY RESOURCES TABLE

Reagents or Resource	Source	Identifier
Antibodies		
Goat Polyclonal anti-Noto	Santa Cruz	Cat#sc-168781
Mouse Polyclonal anti-Brachyury	Santa Cruz	Cat#sc-374321
Mouse Polyclonal anti- FOXA2	Santa Cruz	Cat#sc-6554
Mouse Polyclonal anti-GD2	Santa Cruz	Cat#sc-53831
Rabbit Polyclonal anti- TIE2	Santa Cruz	Cat#sc-324
Mouse Polyclonal anti-COL2A1	Santa Cruz	Cat#sc-52658
Rabbit Polyclonal anti-Aggrecan	Santa Cruz	Cat#sc-25674
Rabbit Polyclonal anti-Shh	Santa Cruz	Cat#sc-9024
Rabbit Polyclonal anti-Noggin	abcam	Cat#ab16054
Mouse Monoclonal anti- Foxj1	Santa Cruz	Cat#sc-53139
Mouse Polyclonal anti- Chordin	Bioss	Cat#bs-11831R
Rabbit Polyclonal anti-eGFP	Santa Cruz	Cat#sc-8334
Chemicals, Peptides, and Recombinant Proteins		
Geltrex	Thermo Fisher	Cat#A1048002
Essential 8	Thermo Fisher	Cat#A1517001
Y-27632	Sigma	Cat#146986-50-7
N2	Thermo Fisher	Cat#17502001
B27	Thermo Fisher	Cat#17504044
ITS	Thermo Fisher	Cat#41400045
NEAA	Thermo Fisher	Cat#11140050
L-glutamine	Thermo Fisher	Cat#25030081
2-mercaptoethanol	Thermo Fisher	Cat#21985023
Activin A	PeproTech	Cat#120-14E
FGF2	PeproTech	Cat#100-18B
Noggin	Sigma	Cat#SRP4675
CHIR99021	Sigma	Cat#SML1046
AGN193109	Sigma	Cat#SML2034
ascorbic acid-2-phosphate	Sigma	Cat#A8960
L-proline	Sigma	Cat#P0380
Dexamethasone	Sigma	Cat#D1756
TGF-β3	R&D	Cat#243-B3-010
BSA	Sigma	Cat#A1933
Critical Commercial Assays		
RNeasy mini kit	QIAGEN	Cat#74104
Prime ScriptRT reagent kit	Clontech	Cat#RR047B
KAPA Stranded mRNaseq Kit	Kapa Biosystems	Cat#KR0960-v3.15
Gibson kit	NEB	Cat#E5510S
Deposited Data		
RNA-sequencing data	GEO	GSE122429 GSE131065
Experimental Models: Cell Lines		
hESC3	This paper	N/A
hESC9	This paper	N/A
IMR90-iPSC	This paper	N/A

(*Continued on next page*)

Continued

Reagents or Resource	Source	Identifier
Experimental Models: Organisms/Strains		
Sprague-Dawley rats	the University of Hong Kong	No. 3438-14
Recombinant DNA		
MLM3636	A gift from Keith Joung	Addgene Cat#43860
Plasmids of hCas9	(Mali et al., 2013)	Addgene Cat#41815
Software and Algorithms		
ImageJ	NIH	https://imagej.nih.gov/ij/
GraphPad Prism 6	Prism	https://www.graphpad.com/scientific-software/prism/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qizhou Lian (qzlian@hku.hk). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Rats

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 (CULATR No. 3438-14). Sprague-Dawley rats (males, 10∼12 weeks old) used in the current study were purchased from laboratory animal unit of the University of Hong Kong. Mice were kept in the animal facility with 12 hours of light and dark cycle with food and water.

Human NP 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, aged 13, 14 and 33 years. The younger two were diagnosed with scoliosis, and the 33 year-old had a burst fracture at L1. All three underwent discectomy (two levels at T12/L1 and L1/2 for the 13 year-old, one level at L2/3, and one level at L1/2, respectively for the other 2 individuals) at Queen Mary Hospital of Hong Kong and under informed consent and Institution Review Board (IRB) approval (UW13-576) of the University of Hong Kong and Hong Kong Hospital Authority West Cluster.

Cell cultures

Three cell lines, hESC3/9 and IMR90, were cultured on a six-well culture plate (Corning) coated with Geltrex (A1048002, Thermo Fisher). Essential 8 medium (A1517001, Thermo Fisher) was changed daily. Y-27632 (146986-50-7, Sigma) was added after cell passage, thawing and nucleofection (Lonza, Switzerland).

METHOD DETAILS

Cell differentiation

For NC-like cell differentiation, hESC3, hESC9 and IMR90 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 (17502001, Thermo Fisher), B27 (17504044, Thermo Fisher), ITS (41400045, Thermo Fisher), 0.1mM NEAA (11140050, Thermo Fisher), 2 mM L-glutamine (25030081, Thermo Fisher) and 0.1 mM 2-mercaptoethanol (21985023, Thermo Fisher) and 10 ng/mL Activin A (120-14E, PeproTech) for two days at step1. Then, other cell growth factors containing 10 ng/mL Activin A, 10 ng/mL FGF2 (100-18B, PeproTech), 50 ng/mL Noggin (SRP4675, Sigma), 3 μM CHIR99021 (SML1046, Sigma) and 10 μM AGN193109 (SML2034, Sigma) were added for a further three days differentiation 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, 50 μg/ml ascorbic acid-2-phosphate (A8960, Sigma), 40 μg/ml L-proline (P0380, Sigma), 10 nM Dexamethasone (D1756, Sigma) and 10 ng/ml TGF-β3 (243-B3-010, R&D) for 15 days at step 3.

Immunostaining

The differentiated cells were fixed in 4% PFA (158127, Sigma) for 30 minutes. The slides were blocked in 1% horse serum (16050130, Thermo Fisher) for one hour after rehydration, and then incubated at 4°C overnight with the primary antibodies: NOTO (1:100,

sc-168781, Santa Cruz), T (1:100, sc-374321, Santa Cruz), FOXA2 (1:100, sc-6554, Santa Cruz), GD2 (1:100, sc-53831, Santa Cruz), TIE2 (1:100, sc-324, Santa Cruz), type II collagen (1:100, sc-52658, Santa Cruz), Aggrecan (1:100, sc-25674, Santa Cruz), Shh (1:100, sc-9024, Santa Cruz), Noggin (1:200, ab16054, abcam), Foxj1 (1:100, sc-53139, Santa Cruz) and Chordin (1:100, bs-11831R, Bioss). 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 PCR

To examine the mRNA expression of *NOTO*, *FOXA2* and *T*, all cells on day 0, 4, 5 and 6 after differentiation were collected to extract the total RNA using an RNeasy mini kit (74104, QIAGEN). The cDNA was then synthesized using a Prime ScriptRT reagent kit with gDNA eraser (RR047B, Clontech) for real-time PCR detection. The primer sequences are listed in Table S1. In all qPCR experiments, three biological replicates per condition were measured.

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 (A1933, Sigma) at a concentration of 2x 10^6/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).

RNA sequencing and processing

For each of the three lines (hESC3/9 and IMR90-iPSC), one sample on day 0, and four samples of sorted cells on day 20 corresponding to TIE2-;GD2-, TIE2+;GD2-, TIE2-;GD2+ and TIE2+;GD2+, were used for RNaseq (Figure S1A). The three day 0 samples serve as base-line biological replicates for comparison with the day 20 samples. In all, 15 *in vitro* samples and four *in vivo* human NP were sequenced at the Centre for Genomic Sciences of the University of Hong Kong.

Briefly, cDNA libraries of these 19 input samples (~100 ng each) were prepared using the KAPA Stranded mRNAseq Kit (KR0960-v3.15, Kapa Biosystems) 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 (Figures S2A and S2B). 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 analyzed by DESeq2 (Anders and Huber, 2010) to detect DEGs between the before- (n = 3) and the after-differentiation (n = 12) samples. In all, 1576 genes with $\log_2$ (fold-change) > 2 and FDR q-value < 0.05 were detected for genes specifically upregulated in hPSC-derived cells on day 20 (Figure S2C; Table S2). We also detected DEGs before and after differentiation per each of the three lines, and found that only 12.3% of them varied across the lines (Figure S5E). The majority were line independent. Several hundred genes were differentially expressed across the lines, mostly due to differences before differentiation (Figures S5F–S5H). This shows that our protocol for NP cell differentiation is robust across different starting cell-lines. The DEGs were analyzed with Geneset Enrichment Analyses tools (Subramanian et al., 2005). To overcome the strong overlap among enriched pathways and gene-ontology (GO) terms, a super Venn diagram was used to visualize their pairwise similarities (Fatemieh, 2006) and a subsequent dimensional reduction pattern by t-SNE (Maaten and Hinton, 2008). HOMER was used in motif enrichment analyses (Heinz et al., 2010).

In vitro – in vivo comparisons and detection of influential NP-specific markers

In correlation and regression analyses, influential data points contribute more to the regression results (Nurunnabi et al., 2016). Here, we outline key steps for *in vitro – in vivo* comparisons, and the identification of influential genes. First, we identified a reference dataset recognized to be of target tissue/cell-type (in our case, the NP). For an *in vitro* sample, denote its expression profile as $\boldsymbol{x} \in R^N$, where N is the number of genes and R denotes a real number. Denote a profile in the reference data 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. Leave one gene $i \in [1, N]$ out and obtain the coefficient of determination denoted as $r^2_{(i)}$, for all genes. Let $\Delta r^2_{(i)} = r^2_{(i)} - r^2$. A negative $\Delta r^2_{(i)}$ suggested that removal of gene i decreased the goodness-of-fit. Thus, the more negative $\Delta r^2_{(i)}$ is, the more gene i contributes to establishing the similarity between *in vitro* differentiations and *in vivo* reference sample. A cut-off was chosen by fitting $\{\Delta r^2_{(i)} \,|\, i = 1, 2, \ldots, N\}$ 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 $\Delta r^2_{(i)}$ became $\Delta r^2_{(i)}(j, k)$ for samples j in *in vivo* data and k in the reference one. Our data also consisted of before

and after differentiation data (n = 3 and 12, respectively). Denote $\delta^B = \{\Delta r^2_{(i)}(j,k) | j \in B\}$ as the values for before-differentiation sample set B; and likewise δ^A for after-differentiation sample set A. We performed t testing between δ^B and δ^A for all genes, thereby detecting those that behaved significantly differently before and after differentiations, hence the influential genes.

Colony-forming assay

To assess colony formation, a single-cell suspension of 1.0×10^3 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 (primers in Table S1A). Then, they were transformed into competent cells after ligation with pBluescript II KS(-) plasmid using a Gibson kit (NEB, E5510S). gRNA vectors were designed as described online (http://zlab.bio/guide-design-resources) and constructed into MLM3636 (Addgene, 43860) in accordance with its supplemental cloning protocol. A surveyor mutation detection kit (IDT, 706025) was used to evaluate the cutting efficiency and off-targets of gRNAs (primers in Table S2). Plasmids of hCas9 (Addgene, 41815) were utilized for genome editing (Mali et al., 2013).

eGFP knock-in

Two guide RNAs (gRNAs) were designed to cut adjacent nucleotides of the *NOTO* TGA stop code, of which gRNA 1 presented higher cutting efficiency in SURVEYOR assay (Figure S1B). Briefly, hESC9 was digested into single cells by accutase (Thermo Fisher) and mixed with the vectors including gRNA, Cas9, and donor for nucleofection. Then, cells were seeded on tissue-treated culture plates coated with Geltrex and cultured in essential 8 medium. 0.5 μg/mL of puromycin (Thermo Fisher) was used to screen the integrated colonies. After puromycin selection, 12 colonies were picked for PCR characterization (Table S1A), and Sanger sequencing (BGI) using primer 1 (Table S1A). 9 positive colonies were identified with eGFP and selection cassette knock-in (Figure S1C). To remove the selection cassette, AAV-mediated Cre recombinase (Clontech) was transiently induced to remove the selection cassette *Cre* protein. 4D-nucleofector (Lonza) was used to achieve the highest transfection efficiency in hESC9. Finally, and the targeted colonies were characterized using Primer 2 (Table S1A) and Sanger sequencing. All *NOTO*-eGFP reporter cell lines were evaluated by off-target testing (primers in Table S1B).

Rat model of IDD and cell transplantation

The IDD rat model was induced as previously described (Wang et al., 2018b). Briefly, after inhalational anesthesia with isoflurane, the tail discs were exposed by a longitudinal incision and a 21-gauge needle punctured at a depth of 1.5mm into the disc via the mid-dorsal annulus at the coccygeal level Co4-5, Co6-7, and Co7-8. Level Co5-6 was not operated to serve as a reference control. Next, the wound was covered with gauze and monitored daily to ensure all rats underwent the standard post-operative procedures. To trace the survival of NP-like cells post transplantation, hESC9/IMR90-derived NP-like cells were transfected with eGFP as previously described (Liang et al., 2015). Two weeks after puncture, eGFP-labeled hESC9/IMR90-derived NP-like cells (5×10^4/5μl PBS) were injected into the injured intervertebral disc (Co6-7, Co7-8). Meanwhile, 5μl PBS was injected into Co4-5 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 Co4-5, Co5-6, Co6-7 and Co7-8 were captured using a Toshiba Medical Systems KXO-80C instrument. Intervertebral disc height index (DHI) was analyzed using NIH-ImageJ software as described previously. The change in DHI before and after hPSC-derived NP-like cell transplantation was expressed as %DHI (post injection DHI/pre-injection DHI) (Li et al., 2014).

Histological evaluation and immunohistochemistry

Eight weeks post cell transplantation, whole discs with the adjacent vertebrae (Co4-5, Co5-6, Co6-7 and Co7-8) 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 30 mins, the sections were incubated overnight at 4°C with primary antibodies including anti-eGFP (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 h at room temperature. Finally, the sections were washed, mounted with DAPI and photographed using a fluorescent microscope. The eGFP-positive cells was counted from different view fields of each sample from 6 rats.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantitative data are expressed as mean ± SEM (standard error of mean). Statistical analysis was performed by unpaired Student's t test for comparisons between two groups and by one-way ANOVA (Analysis of Variance), followed by Bonferroni corrections for multiple testing. A value of (adjusted) $p < 0.05$ was considered statistically significant.

DATA AND CODE AVAILABILITY

All genomic data reported in this study are deposited on the NCBI Gene Expression Omnibus (accession numbers GSE122429 and GSE131065).