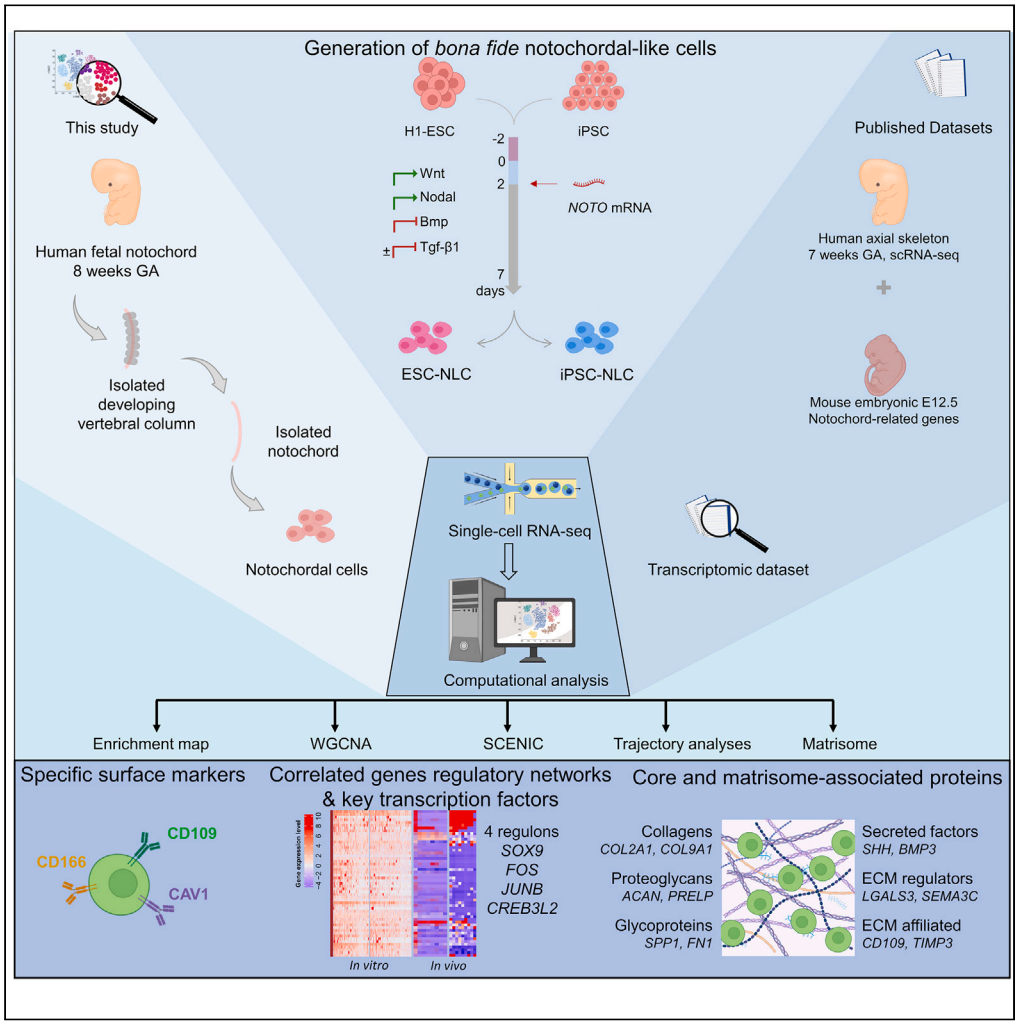
iScience

CellPress
OPEN ACCESS

Article

In vitro and *in vivo* models define a molecular signature reference for human embryonic notochordal cells



Julie Warin, Nicolas Vedrenne, Vivian Tam, ..., Laurent David, Danny Chan, Anne Camus

anne.camus@univ-nantes.fr

Highlights

Stem cell-based model for effective differentiation of notochordal cells

ScRNA-seq uncovers gene regulatory networks linked to the notochordal cell fate

Human stem cell and native notochord integration define reference molecular signature

Secreted and surface markers identification for notochord related disorders modeling

Warin et al., iScience 27, 109018
February 16, 2024
https://doi.org/10.1016/j.isci.2024.109018

iScience

CellPress
OPEN ACCESS

Article

In vitro and *in vivo* models define a molecular signature reference for human embryonic notochordal cells

Julie Warin,[1,9] Nicolas Vedrenne,[1,8,9] Vivian Tam,[2] Mengxia Zhu,[2] Danqing Yin,[2,3] Xinyi Lin,[2,3] Bluwen Guidoux-D'halluin,[1] Antoine Humeau,[8] Luce Roseiro,[1] Lily Paillat,[1] Claire Chédeville,[1] Caroline Chariau,[4] Frank Riemers,[5] Markus Templin,[6] Jérôme Guicheux,[1] Marianna A. Tryfonidou,[5] Joshua W.K. Ho,[2,3] Laurent David,[4,7] Danny Chan,[2] and Anne Camus[1,10,*]

SUMMARY

Understanding the emergence of human notochordal cells (NC) is essential for the development of regenerative approaches. We present a comprehensive investigation into the specification and generation of *bona fide* NC using a straightforward pluripotent stem cell (PSC)-based system benchmarked with human fetal notochord. By integrating *in vitro* and *in vivo* transcriptomic data at single-cell resolution, we establish an extended molecular signature and overcome the limitations associated with studying human notochordal lineage at early developmental stages. We show that TGF-β inhibition enhances the yield and homogeneity of notochordal lineage commitment *in vitro*. Furthermore, this study characterizes regulators of cell-fate decision and matrisome enriched in the notochordal niche. Importantly, we identify specific cell-surface markers opening avenues for differentiation refinement, NC purification, and functional studies. Altogether, this study provides a human notochord transcriptomic reference that will serve as a resource for notochord identification in human systems, diseased-tissues modeling, and facilitating future biomedical research.

INTRODUCTION

Notochordal cells (NC) are a rare cell type present in all vertebrates, arising at gastrulation from the specialized organizer region and contributing to axial elongation.[1–3] Once shaped into a rod-like structure at the midline of the body plan, the notochord ensures signaling function essential to the differentiation of surrounding tissues such as the neuroectoderm and paraxial mesoderm-derived somites.[4] These early interactions mediated by SONIC HEDGEHOG (SHH) morphogen have long been shown to be necessary for the correct regionalization of the central nervous system and for axial skeleton formation.[5–10] Ultimately NC contribute to the formation of the intervertebral disc (IVD), whereafter they reside within the center to progressively disappear during childhood.[11–13] NC have been demonstrated to be a crucial regulator in disc homeostasis maintenance.[14] NC hold great potential for cell-based therapies which would be advanced with the resolution of the prevailing knowledge gaps. The specification of NC fate and co-emerging lineages in human remains poorly understood.[15] Additionally, NC are rarely detected in systems modeling early steps of human embryonic development, e.g., gastruloids or embryoids[16–19] leaving most of our knowledge relying on data generated from *in vivo* or *in vitro* studies in the mouse model.[20–24] Moreover, the disparity of the *in vitro* strategies to guide pluripotent stem cells (PSC) differentiation results in limited production of notochordal-like cells (NLC) on average 5% and to a diverse range of co-emerging cells. The lack of a precise NC molecular signature complicates the interpretation of these heterogeneous populations mostly studied by bulk transcriptome and FACS analyses.[25–28] Here, we propose an improved PSC-based system for the generation of NLC and establish a benchmark for the transcriptome of human fetal notochord. We then combined both *in vitro* NLC differentiation and fetal human notochord transcriptomic profiles to identify gene regulatory networks and the signaling ligands that define the human notochord. Finally, we assemble a comprehensive molecular signature as reference for this tissue that plays many important functions.

[1]Nantes Université, Oniris, CHU Nantes, Inserm, Regenerative Medicine and Skeleton, RMeS, UMR 1229, 44000 Nantes, France
[2]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[3]Laboratory of Data Discovery for Health Limited (D24H), Hong Kong Science Park, Hong Kong SAR, China
[4]Nantes Université, CHU Nantes, Inserm, CNRS, BioCore, 44000 Nantes, France
[5]Department of Clinical Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands
[6]NMI Natural and Medical Sciences Institute, Markwiesenstraße 55, 72770 Reutlingen, Germany
[7]Nantes Université, CHU Nantes, Inserm, CR2TI, 44000 Nantes, France
[8]Inserm, Univ. Limoges, Pharmacology & Transplantation, U1248, CHU Limoges, Service de Pharmacologie, toxicologie et pharmacovigilance, FHU SUPORT, 87000 Limoges, France
[9]These authors contributed equally
[10]Lead contact
*Correspondence: anne.camus@univ-nantes.fr
https://doi.org/10.1016/j.isci.2024.109018

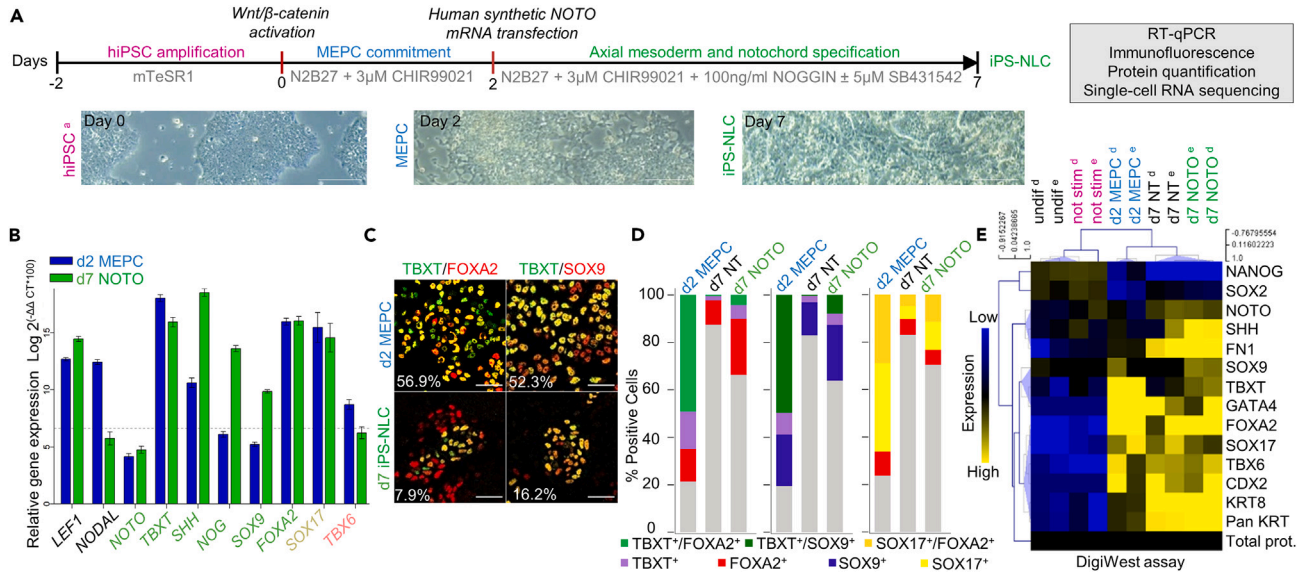


Figure 1. Characterization of hiPSC differentiation toward mesendoderm progenitors and notochordal-like cells

(A) Schematic of the NLC differentiation method. Human iPSC are differentiated into NLC in a three-steps process: Step 1 (hiPSC amplification from day 2 to day 0) comprises the seeding and amplification of the hiPSC; step 2 (MEPC commitment from day 0 to day 2) consists in the activation of the Wnt/β-catenin pathway by CHIR99021 stimulation to obtain mesendoderm progenitors (MEPC); step 3 (axial mesoderm and notochord specification from day 2 to day 7) involves the transfection of the synthetic mRNA encoding *NOTO* transcription factor and culture in medium supplemented with CHIR99021 and NOGGIN (Bmp pathway inhibition). During this last step samples were treated with or without SB431542 (Tgf-β pathway inhibition). Bottom panel: brightfield images illustrate the main morphological changes at each step. Lineage specification and cell identity were monitored by RT-qPCR, immunofluorescence; DigiWest protein quantification and single-cell RNA-sequencing. Scale bars = 50μm.

(B) RT-qPCR assessment of MEPC and NLC differentiation. Relative expression of Wnt and Nodal pathways activation (*LEF1, NODAL, LEFTY1*), notochord (endogenous mRNA for *NOTO, TBXT, SHH, NOGGIN, SOX9, FOXA2*), endoderm (*SOX17*) and paraxial mesoderm (*TBX6*) markers for MEPC and NLC. Mean expression levels are presented as fold changes relative to unstimulated hiPSC and standard error is indicated, for n = 20 biological replicates using hiPSC[a,b,c,d,e]. Dotted line represents the unstimulated hiPSC expression level.

(C) Representative co-immunostaining of TBXT/FOXA2 and TBXT/SOX9 assessing differentiation efficiency for MEPC (day 2) and iPS-NLC (day 7). Double-positive cells are indicated in %. Scale bars = 50μm.

(D) Quantification of TBXT/FOXA2, TBXT/SOX9 and FOXA2/SOX17 immunopositivity of the differentiated cells. Mean values for day 2 (MEPC), day 7 not transfected (d7 NT) and day 7 NOTO transfected (d7 NOTO) cells are represented in %, for n = 9 to 15 biological replicates and n = 3 technical replicates, using hiPSC[a,b,c,d]. Gitlink for detailed quantifications.

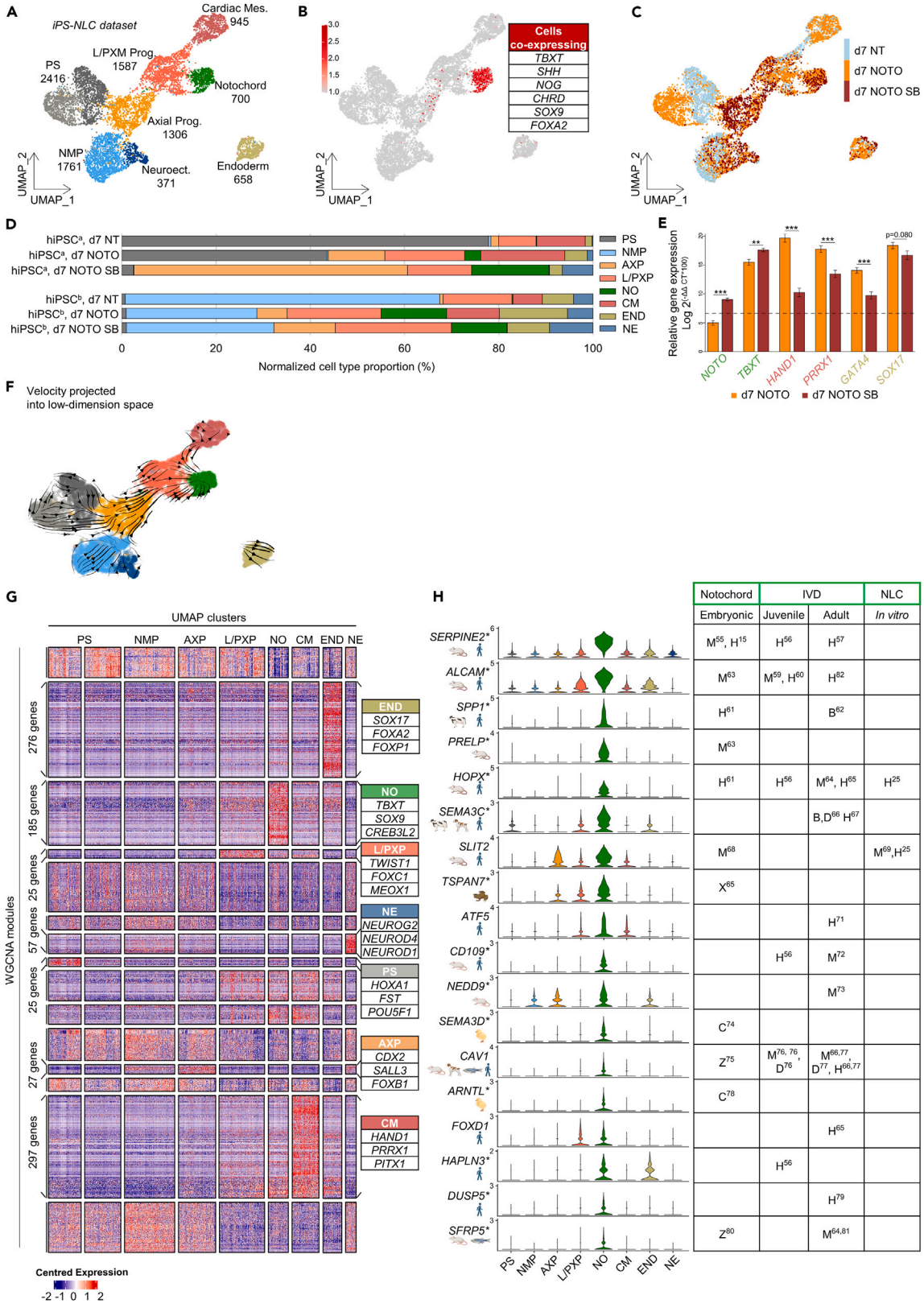
(E) Protein content evaluation at key steps of the NLC differentiation using DigiWest bead-based microarray. The expression levels of the indicated proteins are displayed as a heatmap from lowest (blue) to highest (yellow). Clustering was delineated according to treatment and day of differentiation for hiPSC[d,e]: amplified hiPSC on matrigel (undif), day 2 single-hiPSC seeded on laminin without CHIR stimulation (not stim), CHIR stimulated (MEPC), day 7 not transfected (d7 NT), day 7 NOTO transfected (d7 NOTO). Abbreviations Fig1: d2 MEPC: mesendoderm progenitors at day 2; d7 NOTO: iPS-NLC differentiation at day 7 following *NOTO* mRNA transfection; d7 NT: differentiation at day 7 without transfection; iPS-NLC: notochordal-like cells.

RESULTS

Characterization of human induced pluripotent stem cells differentiation toward mesendoderm progenitors and notochordal-like cells

To delineate the molecular events that characterize the notochord lineage, we set out to replicate, *in vitro*, the early developmental events leading to the induction of bipotent mesendoderm progenitors cells (MEPC)[29–35] and their subsequent commitment toward axial mesoderm to finally become NLC. Building up on knowledge gained from previous reports,[25,27] we defined the fine balance of canonical Wnt/β-catenin, Nodal/Smad2/3 and Bmp signaling necessary to enhance the differentiation of human induced PSC (hiPSC) into NLC (Figure 1A). Efficiency (20% of NLC on average) and robustness were achieved by (1) replacing matrigel by laminin coating, a major constituent of the notochord basement membrane,[36–39] also in the pursuit of designing a GMP-compliant protocol (data not shown); (2) using modified synthetic mRNA[40] to enhance transfection efficiency on day 2 allowing transient expression of the key notochordal transcription factor *NOTO* (Figure S1A); (3) medium supplementation with a selective inhibitor of Tgf-β pathway (SB431542[41]) to limit the emergence of endodermal fate[42,43]; and (4) medium supplementation with the Bmp signaling inhibitor NOGGIN (data not shown) to limit intermediate mesoderm induction.[25,44–46]

Lineage monitoring was performed by RT-qPCR, immunofluorescence and bead based western microarray to assess for MEPC and NLC differentiation. High levels of *LEF1* (key mediator of the Wnt/β-catenin signaling) and *NODAL* transcription were associated with high *TBXT* and *FOXA2* expression in MEPC at day 2. Following *NOTO* mRNA transfection, the upregulation of notochord-related genes, *NOTO, TBXT,*



Notochord	IVD		NLC
Embryonic	Juvenile	Adult	*In vitro*
M[55], H[15]	H[56]	H[57]	
M[63]	M[59], H[60]	H[82]	
H[61]		B[62]	
M[63]			
H[61]	H[56]	M[64], H[65]	H[25]
		B,D[66] H[67]	
M[68]			M[69],H[25]
X[65]			
		H[71]	
	H[56]	M[72]	
		M[73]	
C[74]			
Z[75]	M[76, 76], D[76]	M[66,77], D[77], H[66,77]	
C[78]			
		H[65]	
	H[56]		
		H[79]	
Z[80]		M[64,81]	

Figure 2. Transcriptomic analyses at single-cell resolution of hiPSC differentiation into notochordal-like cells

(A) Unsupervised clustering of iPS-NLC single-cell dataset revealing 9 lineage-associated clusters. Clusters of primitive-streak (PS), neuromesodermal progenitors (NMP), lateral/paraxial mesoderm progenitors (L/PXM Prog.), axial progenitors (Axial Prog.), cardiac mesoderm (Cardiac Mes.), notochord, endoderm, neuroectoderm (Neuroect.) are projected on UMAP embedding. Values indicate the number of cells per cluster.

(B) Distribution of cells co-expressing TBXT, SHH, NOG, CHRD, SOX9, FOXA2 notochord-related markers on UMAP embedding. Expression scores, displayed from lowest (gray) to highest (red), were calculated using the AddModuleScore function.

(C) Distribution of day 7 differentiated cells related to their treatment on UMAP embedding. Treatments day 7 not transfected (d7 NT), day 7 NOTO transfected (d7 NOTO) and day 7 NOTO treated with SB (d7 NOTO SB) are plotted with a distinct color code for iPS-NLC dataset. Note that although the two clusters of primitive-streak have similar transcriptomic profile they are distinct in their composition "NT"/"NOTO".

(D) Cell type differentiation profile related to treatment for two representative hiPSC lines. Normalized values using total number of cells analyzed per treatment in iPS-NLC dataset are shown in %, for day 7 not transfected (d7 NT), day 7 NOTO transfected (d7 NOTO) and day 7 NOTO treated with SB (d7 NOTO SB), for hiPSC[a] and hiPSC[b].

(E) Inhibition of Tgf-β signaling pathway significantly enhances NLC markers expression and reduces off-target differentiation at day 7.

Relative expression of notochord (endogenous mRNA for *NOTO* and *TBXT*), cardiac (*HAND1, PRRX1*) and endoderm (*GATA4, SOX17*) markers by RT-qPCR. Mean expression levels are presented as fold changes relative to unstimulated hiPSC and standard error is indicated, for n = 14 biological replicates using hiPSC[a,b,c,d,e]. Dotted line represents the unstimulated hiPSC expression level. Statistical significance was measured by a student test for each gene marker. **p < 0.001, ***p < 0.0001.

(F) Dynamic velocities projected into UMAP-embedding revealing differentiation trajectories. The spliced and unspliced mRNA dynamics were inferred using scVelo, projected into low dimension space with veloAE and visualized here by arrows. Cells are colored based on cluster assignation.

(G) Identification of major WGCNA modules of co-expressed genes distinguishing cell clusters. Each row represents a WGCNA module, and each column comprises all cells from the annotated clusters. The height of rows indicates the number of genes in the module and the expression level is shown colored from lowest (blue) to highest (red). Three selected lineage-associated genes co-expressed within modules are indicated ordered by intra-connectivity.

(H) Violin plots of selected genes differentially expressed in NO cluster and their corresponding tissue specificities in various species described in the literature. Genes are ordered by Log2FC. Star indicates their belonging to the NO specific WGCNA module. Table on the left recapitulates genes links to lineage, developmental or post-natal time frame and species information curated from the literature. mouse (M), human (H), bovine (B), dog (D), xenopus (X), chick (C), zebrafish (Z). Abbreviations Fig2: AXP: axial progenitors; CM: cardiac mesoderm; END: endoderm; L/PXP: lateral/paraxial mesoderm progenitors; MEPC: mesendoderm progenitors; NE: neuroectoderm; NLC: notochordal-like cells; NMP: neuromesodermal progenitors; NO: notochord; PS: primitive-streak.

SHH, *NOGGIN*, *SOX9*, and *FOXA2* corroborated the differentiation toward the notochordal lineage at day 7. The expression of *SOX17* and *TBX6* evidenced the co-emergence of alternative cell fates such as endoderm and paraxial mesoderm, respectively (Figure 1B). The robustness of the protocol was demonstrated by similar trends of transcriptional expression regardless of inter-cell-line variability (five hiPSC and one hESC lines; Figure S1B). At day 2, over 50% of MEPC co-expressed transcription factors of relevance, i.e., TBXT+/FOXA2+. On day 7, NLC were characterized and quantified by TBXT+/FOXA2+ immunopositivity (7.9% compared to 0.1% in untransfected control, "NT") and TBXT+/SOX9+ immunopositivity (16.2% compared to 0.3% in "NT"), taking into account the possibility that some double-positive cells may express the three factors (Figure 1C). Heterogeneity in cell-types was also featured by the detection of endodermal cells co expressing FOXA2+/SOX17+ (Figure 1D). Finally, protein quantification showed high expression levels of TBXT, FOXA2, SOX17, TBX6, CDX2 in the MEPC while the effect of *NOTO* transfection correlated, at day 7, with higher levels of TBXT, SOX9, and SHH the key signaling protein secreted by the notochord (Figure 1E). Based on this initial characterization, it appears that our differentiation protocol effectively captures the lineage segregation for notochordal fate along with other neighboring lineages and is consistent with the current view of notochord formation during mouse embryonic development.[1,47–50]

Transcriptomic analysis at single-cell resolution of hiPSC differentiation

We subsequently performed single-cell RNA-seq at day 7 to delineate cellular heterogeneity in two distinct iPSC lines for the following three conditions: "NT", NOTO transfected ("NOTO") alone or supplemented with SB431542 from day 2 ("NOTO SB"). Unsupervised clustering revealed 9 discrete subpopulations that we annotated according to expression of lineage-associated genesets (Figure 2A; Gitlink), including two clusters of primitive-streak with similar transcriptomic profile (PS; DEG: *FST, FOXH1, HOXA1, ZIC2*), neuromesodermal progenitors (NMP; DEG: *HOXB9, PAX6, SOX2, SOX3*), axial progenitors (AXP; DEG: *CDX2, FGF8, HOXA7, SLIT2*), lateral and paraxial mesoderm progenitors (L/PXP; DEG: *FOXC1, FOXC2, MEOX1, TWIST1*) and notochord (NO). Cells found in the NO cluster expressed all the marker genes commonly used to define notochordal identity (Figure 2B). The NO cluster comprised 23.8% of cells co-expressing *TBXT*, *FOXA2* and *SOX9*. In addition, non-axial cell fates emerging in the culture were evidenced with the presence of cardiac mesoderm cluster (CM; DEG: *GATA6, HAND1, MSX1, MYL7*)[51,52] as well as 2 clusters related to distinct primary germ layers identified as endoderm (END; DEG: *CLDN3, EPCAM, SOX17, FOXA2*) and neuroectoderm (NE; DEG: *NEUROD4, NEUROG2, POU3F2, TUBB3*). The absence of neural crest cell differentiation was evidenced by the lack of cell cluster expressing *DLX5, SOX10* and *TFAP2C* specific markers. Overall, the observed diversity in cellular outcomes aligned with expected cell fates that are established *in vivo* concomitantly to the process of notochord formation.

Inhibition of Tgf-β signaling pathway significantly enhances NLC markers expression and reduces off-target differentiation

At day 7, UMAP visualization of the spatial relationship of "NT" and *NOTO* transfected ("NOTO") cells alone or supplemented with SB431542 from day 2 ("NOTO SB") highlighted the central role of the transcription factor *NOTO* for the acquisition of the notochordal identity (Figure 2C). Consistently, "NT" cells were over-represented in PS and NMP cluster but absent in the NO cluster. PS and NMP clusters were also composed of "NOTO" cells emphasizing transitory cell states with a lesser degree of maturity within the population at day 7. Addition

of SB431542 reduced endoderm commitment and abolished cardiac derivatives as indicated by a complete lack of "NOTO SB" cells in the CM cluster. In contrast, "NOTO SB" reinforced the other mesodermal subpopulations (AXP, L/PXP, NO). Analysis of cell-type proportions distinguishing the three treatments in the two differentiated hiPSC lines provided evidence that selective inhibition of the Tgf-β/Activin/Nodal pathway substantially reduced heterogeneity, i.e., the diminution of a cell type proportion and the absence of additional cell-type or derivatives (Figure 2D). Indeed, SB431542 treatment reduced off-target differentiation such as cardiac derivatives (highly significant downregulation of key cardiac markers *HAND1* and *PRRX1*) as well as endoderm derivatives (significant downregulation of *GATA4* and a trend to a downregulation of *SOX17* expression) and in contrast promoted NLC differentiation as demonstrated by the significant increase expression of *NOTO* and *TBXT* markers (Figure 2E). Results also revealed one major divergence in the transcriptional profiles of the two differentiated lines with dominant PS or NMP clusters for hiPSC[a] and hiPSC[b], respectively (Figure 2D). Yet, despite intrinsic differences in lineage potency, NOTO and SB treatment guaranteed the emergence of NLC.

Differentiation trajectories and molecular network supporting notochordal lineage

VeloAE analysis to infer cellular transitions from RNA velocity,[53] provided insights into the dynamics and heterogeneity of cellular states *in vitro*. Two notable differentiation trajectories were distinguished (Figure 2F). One predominant path started from PS or NMP cells progressing through the AXP- to NO- cluster that could be attributed to cells, displaying distinct differentiation paths or slower pace of cellular specification, gradually transitioning toward NO cellular identity. A second path revealed transient cellular progression from the L/PXP-to NO- cluster that may indicate off-target differentiation rescued through the forced expression of *NOTO* transcription factor. Trajectories are supported by specific expression patterns of candidate genes associated with these cellular dynamics (Figures S2A and S2B, Gitlink).

To investigate the molecular network supporting the emergence of the notochordal lineage in this human stem-cell-based *in vitro* model, we performed weighted gene correlation network analysis (WGCNA).[54] This revealed seven major modules of co-expressed genes that were biologically relevant and related to END, NO, L/PXP, NE, PS, AXP, and CM clusters. A specific module of 185 co-regulated genes was highly associated with notochord development (Figures 2G and Gitlink). WGCNA and DEG analysis allowed us to identify and compile a comprehensive molecular signature that represents the molecular landscape of *in vitro* generated NLC. It encompasses a wide range of transcription factors, plasma membrane proteins, ECM regulators and secreted proteins. Violin plots depicted the expression distribution of representative selection of these genes (Figure 2H). Interestingly, some have been previously implicated in the regulation of notochord formation in various animal models (mouse, zebrafish, Xenopus, chicken) or the maintenance of IVD homeostasis in humans, dogs or bovine.[15,25,55–82] From this analysis, no lysosome-related organelles, nor vacuole markers were found except for *CAV1* (manually curated gene list from the following ref. [72,83,84] is available on the Gitlink). This *in vitro* model generates embryonic NLC in an immature state (i.e., embryonic state) and does not support vacuole appearance within 7 days of differentiation.[76]

In vitro derived NLC shares gene signature with human fetal notochord

Although hiPSC and hESC share key properties, there are fundamental molecular characteristics that distinguish them, including possible "epigenetic memory" for the former that influence differentiation potential.[85,86] Thus, we characterized notochordal differentiation of human H1-ESC line by scRNA-seq and found comparable cellular heterogeneity and lineage gene expression, except that the differentiation propensity toward definitive endoderm was lower for H1-ESC (Figures S1B and S2C). Comparable efficacy to differentiate toward notochordal lineage was demonstrated for the two pluripotent cell-types following data integration (Int_iPS/ESC-NLC dataset; Figure S2D). Moreover, correlation plots to visualize iPS-NLC and ESC-NLC datasets confirmed a high degree of overlap in their respective notochordal molecular signatures (Figure S2E).

Next, we conducted scRNA-seq gene expression profiling on human fetal notochord to serve as a refence for comparing with NLC generated *in vitro*. UMAP visualization of notochord microdissected at 8 weeks gestational age (GA; NC_Week8 dataset) revealed 5 populations of cells inferred based on cluster-specific expression of established markers (Figures 3A and 3B). These clusters comprised a discrete notochord cluster (NC) of 9 TBXT+ cells, along with other clusters of sclerotome, somitic mesoderm, axial skeleton system and lateral and paraxial mesoderm (L/PXM) (Figure 3C). To gather additional *in vivo* transcriptomic data, we took advantage of the recently published scRNA-seq on microdissected human fetal axial skeleton for a comparative analysis.[61] Taking into account embryonic stage, global molecular profile, and specific expression of *TBXT* transcription factor as consensus marker to discriminate notochordal identity, two out of the four fetal specimens retrieved from week 7, comprising 8 and 9 *TBXT*+ cells respectively, were retained from Zhou et al. study (Noto week 7 samples #3 and #4 dataset renamed here Axial skeleton_Week7). Statistical analysis based on hypergeometric distribution[87] was applied to differentially expressed genes of the iPS/ESC-NLC dataset and both human datasets. Enrichment map analysis of the top100 genes revealed significant overlap of genesets between notochordal clusters of *in vitro* cell differentiation and *in vivo* tissues as well for L/PXM clusters (Figure 4A). Despite the limited number of putative notochord/TBXT+ cells obtained from the three fetal samples, this comparative analysis and convergent data from two independent studies bring confidence to these respective transcriptional profiles.

Description of an extended gene signature of NLC based on comparison with human fetal notochordal cells

In addition, by characterizing a geneset specific to NC in both *in vitro* and *in vivo* systems, we were able to identify surface markers associated with this cell population, i.e., ALCAM (CD166), CD109 and CAV1 (a structural component of the caveolae at the cell membrane); (Figures 2H and S3 for Top100 DEG all notochord clusters). Immunostaining of mouse tissues during development (E11.5-E17.5) and postnatal day 4 stages substantiated their specific distribution at the plasma membrane of the notochord and the nucleus pulposus cells in the forming disc, thereby confirming their potential for cell sorting and enrichment (Figure 4B). We then focused on the extracellular matrix (ECM) which

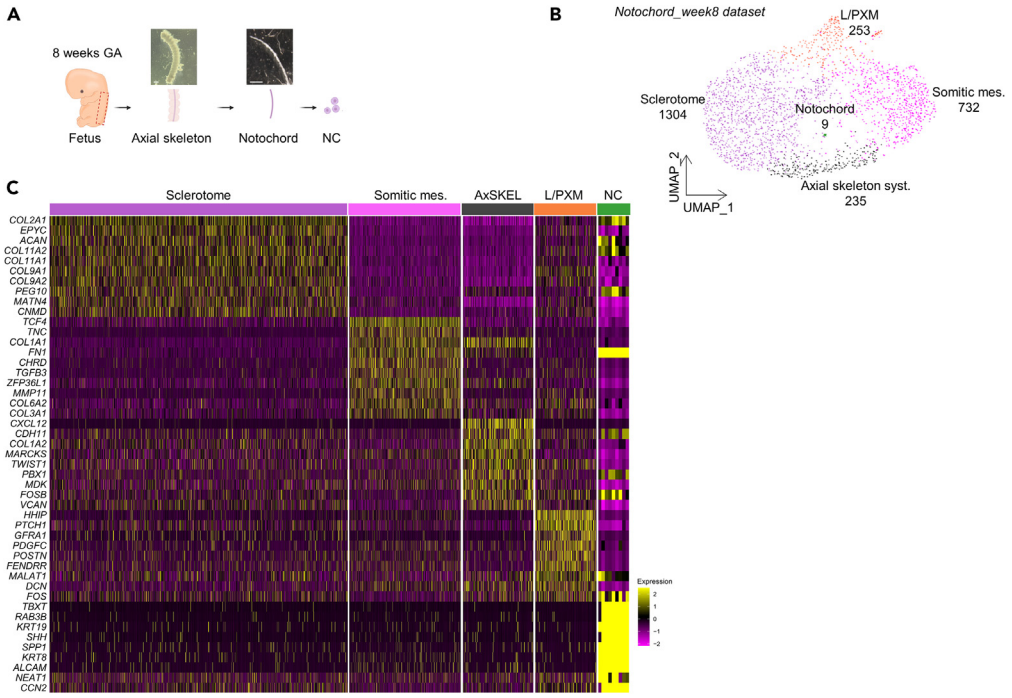


Figure 3. Transcriptomic analyses at single-cell resolution of microdissected human fetal notochord from week 8 gestational age embryo

(A) Schematic workflow of notochord isolation from 8 weeks GA embryo. Notochord was isolated from microdissected axial skeleton and digested prior single-cell processing. Scale bars = 1 mm.

(B) Unsupervised clustering of microdissected human fetal notochord (8 weeks GA) revealing 5 lineage-associated clusters. Clusters of sclerotome, somitic mesoderm (Somitic mes.), lateral/paraxial mesoderm (L/PXM), axial skeleton system development (Axial skeleton syst.) and notochord are projected on UMAP embedding. Values indicate the number of cells per cluster.

(C) Heatmap of top 10 differentially expressed genes corresponding to cell clusters of Notochord_week8 dataset. Expression level is shown colored from lowest (magenta) to highest (yellow). Gitlink for full gene lists. The 5 clusters are sclerotome, somitic mes., axial skeleton system development (AxSKEL), lateral/paraxial mesoderm (L/PXM) and notochord (NC).Abbreviations Fig3: AxSKEL: axial skeleton system development; DEG: differentially expressed genes; GA: gestational age; L/PXM: lateral/paraxial mesoderm; NC: notochordal cells; somitic mes: somitic mesoderm.

in native NC contributes to the environmental factors that support cellular state, IVD development and homeostasis.[14,88,89] Heatmap comparing matrisome-related genes specific to the human embryonic notochord revealed a high degree of similarity with the NO cluster of *in vitro* differentiated cells, very distinct from the END cluster (Figure 4C). This analysis led to the identification of a comprehensive matrisome composition for NC (core matrisome and matrisome-associated proteins). NC was enriched for matrisome proteins already described in the IVD context in human or other animal models such as *COL2A1*, *COL11A2*, *ACAN*, *HAPLN3*, *PRELP*, *SPP1*, *FN1*, *SLIT2*, *SHH*, *BMP3*, *LGALS3*, *SEMA3C*, *CD109*, *TIMP3*.[25,56,61,63,66–69,72,90–94] Interestingly, the analysis highlighted *SEMA3C*, *SULF2*, *HSPG2* and *P4HA1* proteins, not yet described in this developmental context. They have been involved in mediating Hedgehog signaling or in post-translational ECM proteins modifications.[95–97] They may contribute to the establishment of NC cellular niche possibly facilitating cell organization, cell-matrix interaction and cell–cell communication and providing structural and biochemical supports to NC differentiation and tissue development. We also analyzed the iPS-NLC dasaset using CellChat, a tool that quantitatively infers and analyzes cell-cell communication.[98] *COLLAGEN*, *SPP1* (*OSTEOPONTIN*), *SEMA3* and *SHH* ligands were found being predominantly secreted by NLC and involved in signaling interactions with other cell-types (Figure S4). Finally, we showed that a subset of matrisome-related markers identified in the native tissue and the archetypical notochordal markers were associated with highest co-expression scores in the corresponding *in vivo* or *in vitro* notochord clusters (Figure 4D). This underscored the strength of the *in vitro* differentiation protocol in establishing *bona fide* notochordal identity.

Identification of specific regulons activators of NLC differentiation

To assess the gene regulatory networks that would drive the notochordal fate during the differentiation, single-cell regulatory network inference and clustering (SCENIC) analysis[99] was applied on WGCNA-associated gene modules (Figure 2G). Four regulons were particularly associated with the NO cluster: *CREB3L2*, *FOS* and *JUNB* and *SOX9* (Figures 5A and S5). *SOX9* gene is described as an essential regulator of NC identity and ECM deposition.[100–103] Of particular interest, 11 of *SOX9* known targets are showing specific correlated expressions exclusively in the NO cluster of iPS-NLC dataset (Figures 5B, S5 and 2H) and have known functions in survival or in metabolism and ECM regulation (Figure 5C). Ultimately, we analyzed the profile of correlated-genes assigned to notochord (WGCNA) in both the *in vitro* and *in vivo* datasets (Figure S6). Lastly, gene lists

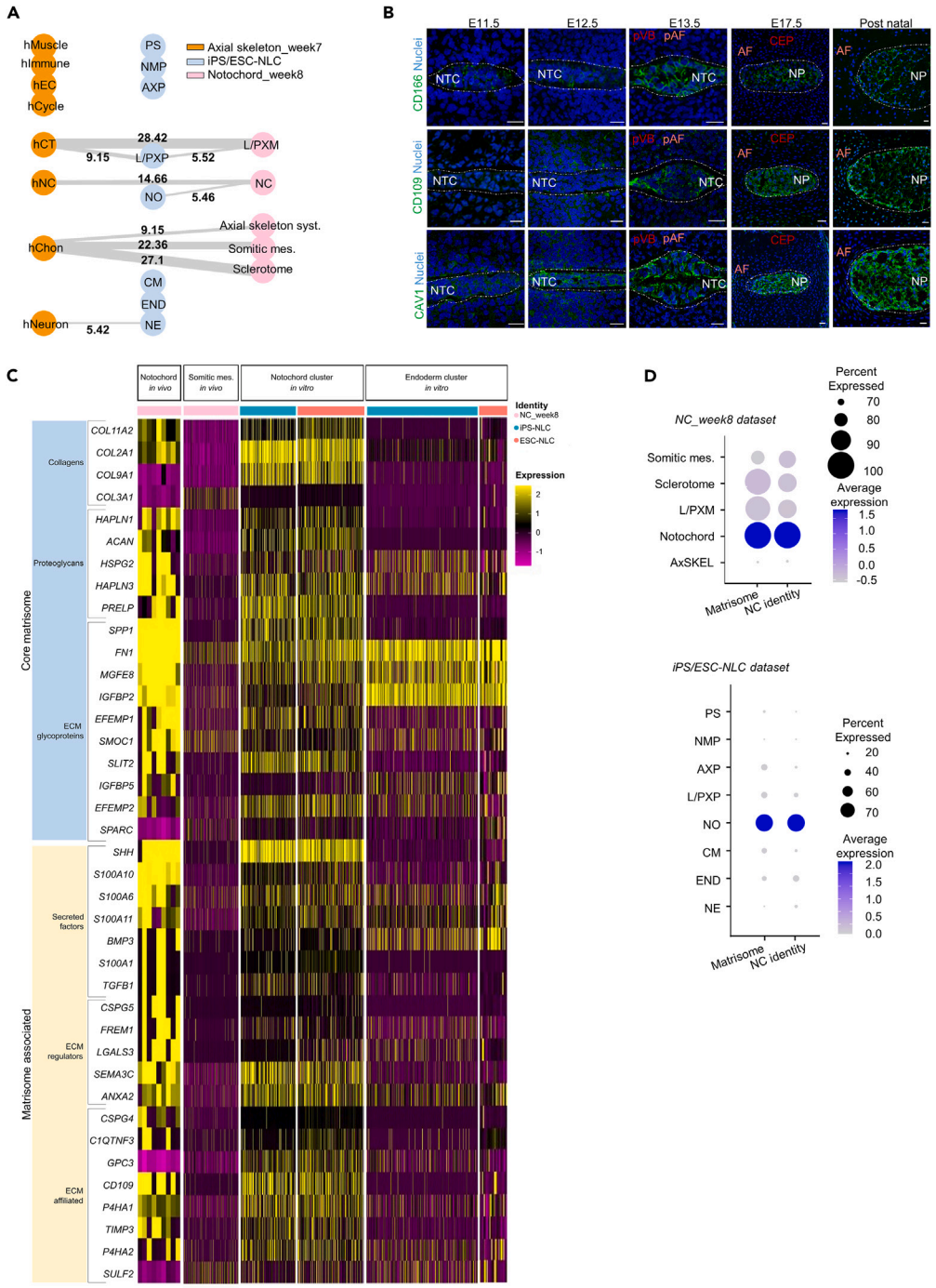


Figure 4. Description of an extended transcriptomic signature of NLC based on comparison with human fetal notochordal cells

(A) Enrichment map showing the connections between the different cell clusters from *in vivo* samples and *in vitro* differentiated cells. The top100 DEG were retrieved for each dataset: Axial skeleton_week7,[56] Notochord_week8 (this study), iPS/ESC-NLC and hypergeometric testing was applied to identify degree of similarities. Only similarities with p value <0.00001 are represented to reveal the significant association between all three datasets. Thicker gray lines represent a higher level of correlation with the numbers indicating the strength of the association as $\log_{10}$(p value). The degree of similarity between hNC and NO does not pass our pre-set stringent significance threshold with a borderline p = 0.14. Nonetheless, the degree of overlap has a reasonable odds ratio of 2. Complete statistical analysis is available on the Gitlink.

(B) Immunostaining of CD166/ALCAM, CD109 and CAV1 showing plasma membrane localization in notochord and nucleus pulposus cells on mouse embryo and postnatal day 4 spine sections. AF = annulus fibrosus; CEP = cartilage endplate; NP = nucleus pulposus; NTC = notochord; pAF = prospective annulus fibrosus; pVB = prospective vertebral body. Nuclei are stained with Hoechst. Scale bars = 25μm.

Figure 4. *Continued*

(C) Heatmap of matrisome differentially expressed genes in human fetal notochord and their expression in somitic mes. Genes are ordered by Log2FC and categorized into core matrisome and matrisome-associated genes. Matrisome of *in vivo* and *in vitro* notochord are remarkably similar, while distinct to that of somitic mes. or endoderm. Expression level is shown colored from lowest (magenta) to highest (yellow). Dataset of origin is indicated by color code.

(D) Co-expression scores of matrisome and archetypical notochord genes in NC_week8 and iPS/ESC-NLC datasets. The scores were calculated using the AddModuleScore function for *SPP1, COL2A1, HAPLN1, HAPLN3, SHH, SEMA3C, ANXA2* and *CD109* (matrisome geneset); and *TBXT, SHH, CHRD, SOX9* and *FOXA2* (NC identity geneset). The size of the dots indicates the percentage of cells expressing the gene set for each cluster, and the color indicates the level of expression from lowest (gray) to highest (purple).Abbreviations Fig4: AxSKEL: axial skeleton system development; DEG: differentially expressed genes; END: endoderm; GA: gestational age; hChon: human chondrocytes; hCT: human connective tissue; hNC: human notochordal cells; L/PXM: lateral/paraxial mesoderm; L/PXP: lateral/paraxial mesoderm progenitors; NC: notochordal cells; NO: notochord; somitic mes: somitic mesoderm.

were curated for positive differentially expressed genes to examine *in vitro* NLC gene signature with human notochord gene signature as well as with node/notochord related gene sets yet published in the mouse model[63,104–106] (Figure 6). While emphasizing commonalities in gene expression profiles, our findings contribute to the construction of an extended molecular signature for human notochordal identity.

DISCUSSION

Investigations into the human notochord development, specifically between 3 and 8 weeks post-conception, are ethically constrained.[15,107,108] In this particular context, combining stem cell models with limited fetal material was critical in confidently characterizing human NC. Here, we describe an extended molecular signature for the notochord and specific lineage markers established from the *in vitro* and *in vivo* transcriptome at single-cell resolution. By integrating transcriptomic profiling from *in vitro* and *in vivo* approaches, this study also validates the improved PSC-based platform allowing for human notochordal fate specification.

Our results show that inhibition of TGF-β signaling after bipotent MEPC specification enhances the yield and homogeneity of notochordal lineage commitment *in vitro*. The conclusion that sustained WNT signaling activation and transient burst of Nodal signaling are essential to promote notochordal fate *in vitro* aligns with other findings recently publicly deposited.[94,109]

This study sets important milestones. It identifies the dynamic transcriptome profiles orchestrating notochordal fate.[110,111] *NOTO* and *SOX9* transcription factors appear to be central regulators of these specification events. While the other regulons *CREB3L2*, *FOS* and *JUNB* have already been associated with the notochord formation and function,[112–114] further investigation is needed to comprehend their pivotal roles as potential co-regulators of various signaling pathways in the cellular context of NLC differentiation. Interestingly, *SOX9* has pivotal role in chondrogenesis and cartilage formation as well as being an essential regulator of NC identity. Further investigations are needed to unravel the specific

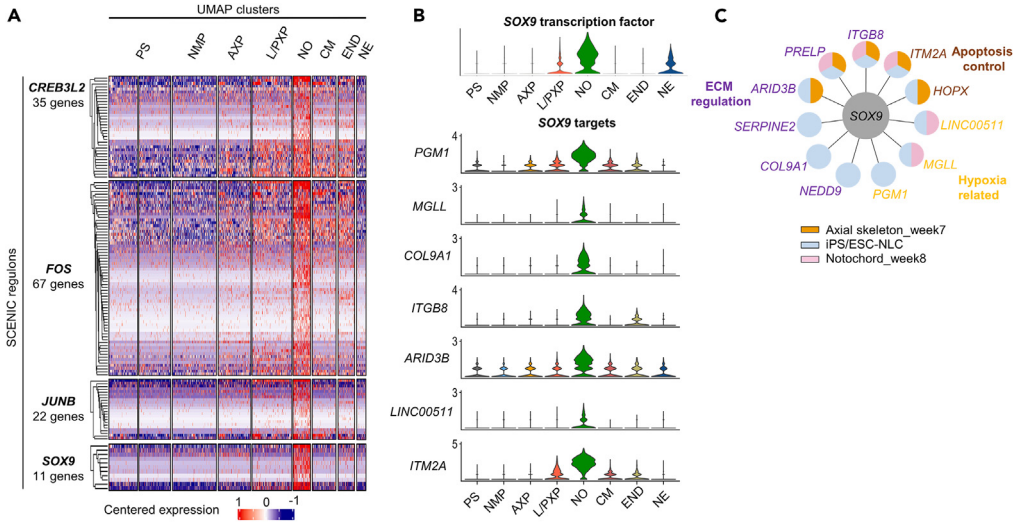


Figure 5. Identification of specific regulons activators of NLC differentiation

(A) Heatmap showing the level of target genes expression of NO regulons identified in iPS-NLC dataset. SCENIC analysis was performed on the NO-specific WGCNA module. Each row represents a regulon, each line indicates a gene, and each column comprises all cells from the annotated clusters. Expression level is shown colored from lowest (blue) to highest (red).

(B) SOX9 positive regulon associated with NLC differentiation. VlnPlot of *SOX9* expression and its targets in iPS-NLC dataset. Of the full list of *SOX9* targets shown in Figure 5A, *PGM1, MGLL, COL9A1, ITGB8, ARID3B, LINC00511, ITM2A* are plotted here and the remaining targets (*SERPINE2, NEDD9, PRELP* and *HOPX*) are already presented in Figure 2H.

(C) Schematic of SOX9 targets categorized according to their function and their expression in different datasets. SOX9 target genes related to apoptosis control, hypoxia and ECM regulation are expressed in the NO-specific WGCNA module identified in iPS-NLC dataset. Color code indicates in which dataset these targets are expressed: orange for Axial skeleton_week8, blue for iPS/ESC-NLC and pink for Notochord_week8.

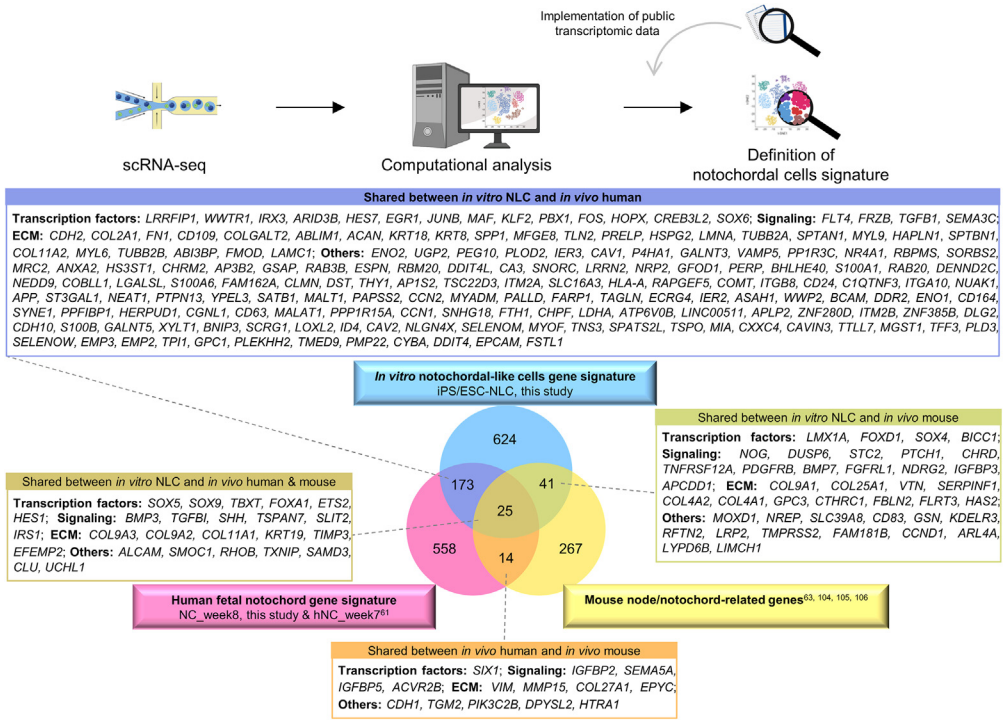


Figure 6. Extended notochordal gene signature based on the comparison of *in vitro* NLC, *in vivo* human and mouse notochord datasets

Mouse gene lists were retrieved from previously published datasets. All gene lists were curated for positive differentially expressed genes. Each inset on the Venn Diagram illustrates the overlapping expressing genes. The full gene list shared between *in vitro* NLC and *in vivo* human notochord is available on the Gitlink.

downstream targets and mechanistic regulations that drive the emergence of these two distinct cell lineages during the early development of the axial skeleton.[115] Note that in the course of the differentiation, our results showed that *SOX9* mRNA increased in differentiating MEPC but a decrease of SOX9 immunopositive cells level was observed at day 7 (from 70% at day 2–30% at day 7). Time course scRNA-seq analysis may provide explanation for these observations and a better understanding of *SOX9* expression pattern related to NC identity.

Another important contribution is the identification of stringent lineage-specific cell-surface markers, *ALCAM* (*CD166*), *CD109* and *CAV1*, as potential tools to allow purification of NLC for further functional study. These markers have been described in juvenile and adult human disc cells.[56,60,72,116] In this study, they characterize the NLC population and chart the corresponding domain of gene expression in rodent embryos. Their specific role in embryogenesis has yet to be explored.

The components of the NLC matrisome *in vitro* were shown to be enriched in the notochordal niche *in vivo*. These insights will be instrumental in guiding the switch from 2D to 3D culture to facilitate further maturation of the NC essential in the process of IVD formation. Moreover, a robust 3D model to investigate the regenerative properties of NC, is yet to be established to advance NC-based therapeutic strategies for disc degeneration, a prominent cause of chronic low back pain.[117] This study also provides fundamental knowledge that can aid in both comprehending and modeling various notochord-related disorders in neural or spine defects.[8,118–120]

Limitations of the study

The roadmap to human notochord development remains to be completed. We present here the initial step by providing an extended molecular signature for embryonic NC and a robust system to investigate related essential developmental processes *in vitro*. It will be necessary to conduct further transcriptome analysis to gain a comprehensive understanding of NLC maturation toward vacuolated NC phenotype and to the ultimate acquisition of their characteristics found in the healthy disc. This work provided insights into the dynamics and heterogeneity of cellular states in an *in vitro* differentiation system where *NOTO* expression is forced. We cannot exclude the possibility that the *in vivo* notochord cell fate trajectory may be distinct. Based on our data, we cannot provide the ultimate proof of the lineage emergence and validation of surface markers. Further investigations involving fluorescence-activated cell sorting analysis on human NC or spatial transcriptomic would be relevant. Addressing these matters prove to be a genuine challenge given the ethical and technical constraints within the human model and the scarcity of notochord cell population.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109018.

ACKNOWLEDGMENTS

This work was funded by European Union's Horizon 2020 research and innovation program (iPSpine; grant agreement number 825925) (A.C., D.C., J.G., M.A.T.); iPSpine funding was not used for the fetal samples and related data collection, CNRS, Inserm, the French Society of Rheumatology (SPHERODISC) (A.C.), the RGC European Union - Hong Kong Research and Innovation Cooperation Co-funding Mechanism (E-HKU703/18) (D.C.) and the Dutch Arthritis Society (LLP22) (M.A.T.) and supported by AIR@InnoHK administered by the Innovation and Technology Commission of Hong Kong (J.W.K.H.). We are most grateful to the Bioinformatics Core Facility of Nantes BiRD, member of Biogenouest, Institut Français de Bioinformatique (ANR-11-INBS-0013) for the use of its resources and for its technical support, to the IBISA MicroPICell facility (Biogenouest), member of the national infrastructure France-Bioimaging supported by the French National Research Agency (ANR-10-INBS-04) and SC3M facility (SFR Francois Bonamy, University of Nantes) and to the Centre for PanorOmic Sciences (CPOS) at the University of Hong Kong, Li Ka Shing Faculty of Medicine for its technical support. We thank Dr Vincent Cheung from Gleneagles Hospital, Hong Kong, for obtaining the fetal sample, the laboratory: UR20218, NEURIT, Limoges, France (Dr. Pierre Antoine Faye) for providing LIMFRi001 cell line (MTA: OV:2021_00845) and Andras Nagy's laboratory for the derivation of the H1 cloaked FailSafe embryonic stem cells originated from Wi Cell (Cell line WA01). We acknowledge Nicolas Antunes-Lardeux, Lucille Degos, Romain Desprat, Virginie Forest, Ben Gantenbein, Boris Halgand, Jean Marc Lemaitre for their valuable technical collaboration and Jerome Jullien for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, A.C. and D.C. with advice from L.D., J.G., J.W.K.H., V.T., M.A.T., N.V., and J.W.; Methodology and data curation, A.C., C.C. (Caroline Chariau), C.C. (Claire Chédeville), L.D., X.L., L.P., L.R., V.T., M.T., N.V., J.W., D.Y., and M.Z.; Project administration, A.C., D.C., L.D., J.G., J.W.K.H., M.A.T., and M.T.; formal analysis and programming, B.G.H., J.W.K.H., A.H., X.L., F.R., V.T., J.W., and D.Y.; writing—original draft preparation, A.C., N.V., and J.W. with the help of D.C., L.D., J.W.K.H., X.L., V.T., and D.Y.; review and editing of the manuscript, A.C., J.G., and M.A.T.; supervision, A.C. All authors contributed to the interpretation of the data and read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 11, 2023
Revised: November 13, 2023
Accepted: January 22, 2024
Published: January 26, 2024

REFERENCES

1. 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. https://doi.org/10.1016/j.devcel.2007.10.016.
2. Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. Genes Dev. *17*, 1646–1662. https://doi.org/10.1101/gad.1100503.
3. Balmer, S., Nowotschin, S., and Hadjantonakis, A.-K. (2016). Notochord morphogenesis in mice: Current understanding & open questions: Notochord Morphogenesis in Mice. Dev. Dynam. *245*, 547–557. https://doi.org/10.1002/dvdy.24392.
4. 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. https://doi.org/10.1038/383407a0.
5. Pourquié, O., Coltey, M., Teillet, M.A., Ordahl, C., and Le Douarin, N.M. (1993). Control of dorsoventral patterning of somitic derivatives by notochord and floor plate. Proc. Natl. Acad. Sci. USA *90*, 5242–5246. https://doi.org/10.1073/pnas.90.11.5242.
6. Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B., and Balling, R. (1994). The role of Pax-1 in axial skeleton development. Development *120*, 1109–1121. https://doi.org/10.1242/dev.120.5.1109.
7. Camus, A., Davidson, B.P., Billiards, S., Khoo, P., Rivera-Pérez, J.A., Wakamiya, M., Behringer, R.R., and Tam, P.P. (2000). The morphogenetic role of midline mesendoderm and ectoderm in the development of the forebrain and the midbrain of the mouse embryo. Development *127*, 1799–1813. https://doi.org/10.1242/dev.127.9.1799.
8. Bagnat, M., and Gray, R.S. (2020). Development of a straight vertebrate body axis. Development *147*, dev175794. https://doi.org/10.1242/dev.175794.
9. Williams, S., Alkhatib, B., and Serra, R. (2019). Development of the axial skeleton and intervertebral disc. Curr. Top. Dev. Biol. *133*, 49–90. https://doi.org/10.1016/bs.ctdb.2018.11.018.
10. Guzzetta, A., Koska, M., Rowton, M., Sullivan, K.R., Jacobs-Li, J., Kweon, J., Hidalgo, H., Eckart, H., Hoffmann, A.D., Back, R., et al. (2020). Hedgehog–FGF signaling axis patterns anterior mesoderm during gastrulation. Proc. Natl. Acad. Sci. USA *117*, 15712–15723. https://doi.org/10.1073/pnas.1914167117.
11. 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. Dynam. *237*, 3953–3958. https://doi.org/10.1002/dvdy.21805.
12. 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. https://doi.org/10.1242/dmm.008128.
13. Jiang, W., Glaeser, J.D., Salehi, K., Kaneda, G., Mathkar, P., Wagner, A., Ho, R., and Sheyn, D. (2022). Single-cell atlas unveils cellular heterogeneity and novel markers in human neonatal and adult intervertebral discs. iScience *25*, 104504. https://doi.org/10.1016/j.isci.2022.104504.
14. Bach, F.C., Poramba-Liyanage, D.W., Riemers, F.M., Guicheux, J., Camus, A., Iatridis, J.C., Chan, D., Ito, K., Le Maitre, C.L., and Tryfonidou, M.A. (2021). Notochordal Cell-Based Treatment Strategies and Their Potential in Intervertebral Disc Regeneration. Front. Cell Dev. Biol. *9*, 780749.
15. Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289. https://doi.org/10.1038/s41586-021-04158-y.
16. Warmflash, A., Sorre, B., Etoc, F., Siggia, E.D., and Brivanlou, A.H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat. Methods *11*, 847–854. https://doi.org/10.1038/nmeth.3016.
17. Martyn, I., Kanno, T.Y., Ruzo, A., Siggia, E.D., and Brivanlou, A.H. (2018). Self-organization of a human organizer by combined Wnt and Nodal signalling. Nature *558*, 132–135. https://doi.org/10.1038/s41586-018-0150-y.
18. Martyn, I., Brivanlou, A.H., and Siggia, E.D. (2019). A wave of WNT signaling balanced by secreted inhibitors controls primitive streak formation in micropattern colonies of human embryonic stem cells. Development *146*, dev172791. https://doi.org/10.1242/dev.172791.
19. Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature *582*, 410–415. https://doi.org/10.1038/s41586-020-2383-9.
20. Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.-C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature *562*, 272–276. https://doi.org/10.1038/s41586-018-0578-0.
21. Turner, D.A., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., Glodowski, C.R., Hayward, P.C., Collignon, J., Gustavsen, C., Serup, P., et al. (2017). Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. Development *144*, 3894–3906. https://doi.org/10.1242/dev.150391.
22. Xu, P.-F., Borges, R.M., Fillatre, J., de Oliveira-Melo, M., Cheng, T., Thisse, B., and Thisse, C. (2021). Construction of a mammalian embryo model from stem cells organized by a morphogen signalling centre. Nat. Commun. *12*, 3277. https://doi.org/10.1038/s41467-021-23653-4.
23. Amadei, G., Handford, C.E., Qiu, C., De Jonghe, J., Greenfeld, H., Tran, M., Martin, B.K., Chen, D.-Y., Aguilera-Castrejon, A., Hanna, J.H., et al. (2022). Embryo model completes gastrulation to neurulation and organogenesis. Nature *610*, 143–153. https://doi.org/10.1038/s41586-022-05246-3.
24. Rossi, G., Broguiere, N., Miyamoto, M., Boni, A., Guiet, R., Girgin, M., Kelly, R.G., Kwon, C., and Lutolf, M.P. (2021). Capturing Cardiogenesis in Gastruloids. Cell Stem Cell *28*, 230–240.e6. https://doi.org/10.1016/j.stem.2020.10.013.
25. Zhang, Y., Zhang, Z., Chen, P., Ma, C.Y., Li, C., Au, T.Y.K., Tam, V., Peng, Y., Wu, R., Cheung, K.M.C., et al. (2020). Directed Differentiation of Notochord-like and Nucleus Pulposus-like Cells Using Human Pluripotent Stem Cells. Cell Rep. *30*, 2791–2806.e5. https://doi.org/10.1016/j.celrep.2020.01.100.
26. 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 Cell. Dev. *20*, 1817–1827. https://doi.org/10.1089/scd.2011.0042.
27. Colombier, P., Halgand, B., Chédeville, C., Chariau, C., François-Campion, V., Kilens, S., Vedrenne, N., Clouet, J., David, L., Guicheux, J., and Camus, A. (2020). NOTO Transcription Factor Directs Human Induced Pluripotent Stem Cell-Derived Mesendoderm Progenitors to a Notochordal Fate. Cells *9*, 509. https://doi.org/10.3390/cells9020509.
28. Diaz-Hernandez, M.E., Khan, N.M., Trochez, C.M., Yoon, T., Maye, P., Presciutti, S.M., Gibson, G., and Drissi, H. (2020). Derivation of notochordal cells from human embryonic stem cells reveals unique regulatory networks by single cell-transcriptomics. J. Cell. Physiol. *235*, 5241–5255. https://doi.org/10.1002/jcp.29411.
29. Burtscher, I., and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. Development *136*, 1029–1038. https://doi.org/10.1242/dev.028415.
30. Arnold, S.J., Hofmann, U.K., Bikoff, E.K., and Robertson, E.J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. Development *135*, 501–511. https://doi.org/10.1242/dev.014357.
31. Dunn, N.R., Vincent, S.D., Oxburgh, L., Robertson, E.J., and Bikoff, E.K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. Development *131*, 1717–1728. https://doi.org/10.1242/dev.01072.
32. Massey, J., Liu, Y., Alvarenga, O., Saez, T., Schmerer, M., and Warmflash, A. (2019). Synergy with TGFβ ligands switches WNT pathway dynamics from transient to sustained during human pluripotent cell differentiation. Proc. Natl. Acad. Sci. USA *116*, 4989–4998. https://doi.org/10.1073/pnas.1815363116.
33. Scheibner, K., Schirge, S., Burtscher, I., Büttner, M., Sterr, M., Yang, D., Böttcher, A., Ansarullah, Irmler, M., Beckers, J., et al. (2021). Epithelial cell plasticity drives endoderm formation during gastrulation. Nat. Cell Biol. *23*, 692–703. https://doi.org/10.1038/s41556-021-00694-x.
34. Lolas, M., Valenzuela, P.D.T., Tjian, R., and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. Proc. Natl. Acad.

Sci. USA *111*, 4478–4483. https://doi.org/10.1073/pnas.1402612111.

35. Legier, T., Rattier, D., Llewellyn, J., Vannier, T., Sorre, B., Maina, F., and Dono, R. (2023). Epithelial disruption drives mesendoderm differentiation in human pluripotent stem cells by enabling TGF-β protein sensing. Nat. Commun. *14*, 349. https://doi.org/10.1038/s41467-023-35965-8.
36. Parsons, M.J., Pollard, S.M., Saúde, L., Feldman, B., Coutinho, P., Hirst, E.M.A., and Stemple, D.L. (2002). Zebrafish mutants identify an essential role for laminins in notochord formation. Development *129*, 3137–3146. https://doi.org/10.1242/dev.129.13.3137.
37. Buisson, N., Sirour, C., Moreau, N., Denker, E., Le Bouffant, R., Goullancourt, A., Darribère, T., and Bello, V. (2014). An adhesome comprising laminin, dystroglycan and myosin IIA is required during notochord development in Xenopus laevis. Development *141*, 4569–4579. https://doi.org/10.1242/dev.116103.
38. Chen, J., Jing, L., Gilchrist, C.L., Richardson, W.J., Fitch, R.D., and Setton, L.A. (2009). Expression of laminin isoforms, receptors, and binding proteins unique to nucleus pulposus cells of immature intervertebral disc. Connect. Tissue Res. *50*, 294–306. https://doi.org/10.3109/03008200802714925.
39. Speer, J.E., Barcellona, M.N., Lu, M.Y., Zha, Z., Jing, L., Gupta, M.C., Buchowski, J.M., Kelly, M.P., and Setton, L.A. (2021). Development of a library of laminin-mimetic peptide hydrogels for control of nucleus pulposus cell behaviors. J. Tissue Eng. *12*, 20417314211021220. https://doi.org/10.1177/20417314211021220.
40. 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. USA *104*, 15765–15770. https://doi.org/10.1073/pnas.0704344104.
41. Callahan, J.F., Burgess, J.L., Fornwald, J.A., Gaster, L.M., Harling, J.D., Harrington, F.P., Heer, J., Kwon, C., Lehr, R., Mathur, A., et al. (2002). Identification of Novel Inhibitors of the Transforming Growth Factor β1 (TGF-β1) Type 1 Receptor (ALK5). J. Med. Chem. *45*, 999–1001. https://doi.org/10.1021/jm010493y.
42. Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M., et al. (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell *14*, 237–252. https://doi.org/10.1016/j.stem.2013.12.007.
43. Faial, T., Bernardo, A.S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G.E., Mascetti, V.L., Trotter, M.W.B., Smith, J.C., and Pedersen, R.A. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. Development *142*, 2121–2135. https://doi.org/10.1242/dev.117838.
44. Bachiller, D., Klingensmith, J., Kemp, C., Belo, J.A., Anderson, R.M., May, S.R., McMahon, J.A., McMahon, A.P., Harland, R.M., Rossant, J., and De Robertis, E.M. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. Nature *403*, 658–661. https://doi.org/10.1038/35001072.
45. McMahon, J.A., Takada, S., Zimmerman, L.B., Fan, C.-M., Harland, R.M., and McMahon, A.P. (1998). Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. Genes Dev. *12*, 1438–1452.
46. Sumi, T., Tsuneyoshi, N., Nakatsuji, N., and Suemori, H. (2008). Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. Development *135*, 2969–2979. https://doi.org/10.1242/dev.021121.
47. Cambray, N., and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. Development *134*, 2829–2840. https://doi.org/10.1242/dev.02877.
48. Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev. Cell *42*, 514–526.e7. https://doi.org/10.1016/j.devcel.2017.07.021.
49. Forlani, S., Lawson, K.A., and Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. Development *130*, 3807–3819. https://doi.org/10.1242/dev.00573.
50. Kinder, S.J., Tsang, T.E., Wakamiya, M., Sasaki, H., Behringer, R.R., Nagy, A., and Tam, P.P. (2001). The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. Development *128*, 3623–3634. https://doi.org/10.1242/dev.128.18.3623.
51. Bardot, E., Calderon, D., Santoriello, F., Han, S., Cheung, K., Jadhav, B., Burtscher, I., Artap, S., Jain, R., Epstein, J., et al. (2017). Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential. Nat. Commun. *8*, 14428. https://doi.org/10.1038/ncomms14428.
52. Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451–467. https://doi.org/10.1016/j.cell.2016.06.011.
53. Qiao, C., and Huang, Y. (2021). Representation learning of RNA velocity reveals robust cell transitions. Proc. Natl. Acad. Sci. USA *118*, e2105859118. https://doi.org/10.1073/pnas.2105859118.
54. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
55. Onuma, Y., Asashima, M., and Whitman, M. (2006). A Serpin family gene, Protease nexin-1 has an activity distinct from protease inhibition in early Xenopus embryos. Mech. Dev. *123*, 463–471. https://doi.org/10.1016/j.mod.2006.04.005.
56. Tam, V., Chen, P., Yee, A., Solis, N., Klein, T., Kudelko, M., Sharma, R., Chan, W.C., Overall, C.M., Haglund, L., et al. (2020). DIPPER, a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. Elife *9*, e64940. https://doi.org/10.7554/eLife.64940.
57. Riester, S.M., Lin, Y., Wang, W., Cong, L., Mohamed Ali, A.-M., Peck, S.H., Smith, L.J., Currier, B.L., Clark, M., Huddleston, P., et al. (2018). RNA sequencing identifies gene regulatory networks controlling extracellular matrix synthesis in intervertebral disk tissues. J. Orthop. Res. *36*, 1356–1369. https://doi.org/10.1002/jor.23834.
58. Ohneda, O., Ohneda, K., Arai, F., Lee, J., Miyamoto, T., Fukushima, Y., Dowbenko, D., Lasky, L.A., and Suda, T. (2001). ALCAM (CD166): its role in hematopoietic and endothelial development. Blood *98*, 2134–2142. https://doi.org/10.1182/blood.v98.7.2134.
59. Tang, X., Jing, L., and Chen, J. (2012). Changes in the Molecular Phenotype of Nucleus Pulposus Cells with Intervertebral Disc Aging. PLoS One *7*, e52020. https://doi.org/10.1371/JOURNAL.PONE.0052020.
60. 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. https://doi.org/10.1002/jor.23244.
61. Zhou, T., Chen, Y., Liao, Z., Zhang, L., Su, D., Li, Z., Yang, X., Ke, X., Liu, H., Chen, Y., et al. (2023). Spatiotemporal Characterization of Human Early Intervertebral Disc Formation at Single-Cell Resolution. Adv. Sci. *10*, 2206296. https://doi.org/10.1002/advs.202206296.
62. Panebianco, C.J., Dave, A., Charytonowicz, D., Sebra, R., and Iatridis, J.C. (2021). Single-cell RNA-sequencing atlas of bovine caudal intervertebral discs: Discovery of heterogeneous cell populations with distinct roles in homeostasis. Faseb. J. *35*, e21919. https://doi.org/10.1096/fj.202101149R.
63. Peck, S.H., McKee, K.K., Tobias, J.W., Malhotra, N.R., Harfe, B.D., and Smith, L.J. (2017). Whole Transcriptome Analysis of Notochord-Derived Cells during Embryonic Formation of the Nucleus Pulposus. Sci. Rep. *7*, 10504. https://doi.org/10.1038/s41598-017-10692-5.
64. Veras, M.A., McCann, M.R., Tenn, N.A., and Séguin, C.A. (2020). Transcriptional profiling of the murine intervertebral disc and age-associated changes in the nucleus pulposus. Connect. Tissue Res. *61*, 63–81. https://doi.org/10.1080/03008207.2019.1665034.
65. Gan, Y., He, J., Zhu, J., Xu, Z., Wang, Z., Yan, J., Hu, O., Bai, Z., Chen, L., Xie, Y., et al. (2021). Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitors in human intervertebral discs. Bone Res. *9*, 37. https://doi.org/10.1038/s41413-021-00163-z.
66. Bach, F.C., Zhang, Y., Miranda-Bedate, A., Verdonschot, L.C., Bergknut, N., Creemers, L.B., Ito, K., Sakai, D., Chan, D., Meij, B.P., and Tryfonidou, M.A. (2016). Increased caveolin-1 in intervertebral disc degeneration facilitates repair. Arthritis Res. Ther. *18*, 59. https://doi.org/10.1186/S13075-016-0960-Y/FIGURES/8.
67. Binch, A.L.A., Cole, A.A., Breakwell, L.M., Michael, A.L.R., Chiverton, N., Creemers, L.B., Cross, A.K., Le Maitre, C.L., Binch, A.L.A., Cole, A.A., et al. (2015). Class 3 semaphorins expression and association with innervation and angiogenesis within the degenerate human intervertebral disc.

Oncotarget *6*, 18338–18354. https://doi.org/10.18632/ONCOTARGET.4274.
68. Medioni, C., Bertrand, N., Mesbah, K., Hudry, B., Dupays, L., Wolstein, O., Washkowitz, A.J., Papaioannou, V.E., Mohun, T.J., Harvey, R.P., and Zaffran, S. (2010). Expression of Slit and Robo Genes in the Developing Mouse Heart. Dev. Dynam. *239*, 3303–3311. https://doi.org/10.1002/dvdy.22449.
69. Yeo, G.H.T., Lin, L., Qi, C.Y., Cha, M., Gifford, D.K., and Sherwood, R.I. (2020). A Multiplexed Barcodelet Single-Cell RNA-Seq Approach Elucidates Combinatorial Signaling Pathways that Drive ESC Differentiation. Cell Stem Cell *26*, 938–950.e6. https://doi.org/10.1016/j.stem.2020.04.020.
70. Kashef, J., Diana, T., Oelgeschläger, M., and Nazarenko, I. (2013). Expression of the tetraspanin family members Tspan3, Tspan4, Tspan5 and Tspan7 during Xenopus laevis embryonic development. Gene Expr. Patterns *13*, 1–11. https://doi.org/10.1016/j.gep.2012.08.001.
71. Song, B.P., Ragsac, M.F., Tellez, K., Jindal, G.A., Grudzien, J.L., Le, S.H., and Farley, E.K. (2023). Diverse logics and grammar encode notochord enhancers. Cell Rep. *42*, 112052. https://doi.org/10.1016/j.celrep.2023.112052.
72. Kudelko, M., Chen, P., Tam, V., Zhang, Y., Kong, O.-Y., Sharma, R., Au, T.Y.K., To, M.K.-T., Cheah, K.S.E., Chan, W.C.W., and Chan, D. (2021). PRIMUS: Comprehensive proteomics of mouse intervertebral discs that inform novel biology and relevance to human disease modelling. Matrix Biol. *12*, 100082. https://doi.org/10.1016/j.mbplus.2021.100082.
73. Markova, D.Z., Kepler, C.K., Addya, S., Murray, H.B., Vaccaro, A.R., Shapiro, I.M., Greg Anderson, D., Albert, T.J., and Risbud, M.V. (2013). An organ culture system to model early degenerative changes of the intervertebral disc II: profiling global gene expression changes. Arthritis Res. *15*, R121. https://doi.org/10.1186/AR4301.
74. Anderson, C.N.G., Ohta, K., Quick, M.M., Fleming, A., Keynes, R., and Tannahill, D. (2003). Molecular analysis of axon repulsion by the notochord. Development *130*, 1123–1133. https://doi.org/10.1242/dev.00327.
75. Garcia, J., Bagwell, J., Njaine, B., Norman, J., Levic, D.S., Wopat, S., Miller, S.E., Liu, X., Locasale, J.W., Stainier, D.Y.R., and Bagnat, M. (2017). Sheath Cell Invasion and Trans-differentiation Repair Mechanical Damage Caused by Loss of Caveolae in the Zebrafish Notochord. Curr. Biol. *27*, 1982–1989.e3. https://doi.org/10.1016/j.cub.2017.05.035.
76. Paillat, L., Coutant, K., Dutilleul, M., Le Lay, S., and Camus, A. (2023). Three-dimensional culture model to study the biology of vacuolated notochordal cells from mouse nucleus pulposus explants. Eur. Cell. Mater. *45*, 72–87. https://doi.org/10.22203/ECM.V045A06.
77. Smolders, L.A., Meij, B.P., Onis, D., Riemers, F.M., Bergknut, N., Wubbolts, R., Grinwis, G.C.M., Houweling, M., Groot Koerkamp, M.J.A., van Leenen, D., et al. (2013). Gene expression profiling of early intervertebral disc degeneration reveals a down-regulation of canonical Wnt signaling and caveolin-1 expression: implications for development of regenerative strategies. Arthritis Res. Ther. *15*, R23. https://doi.org/10.1186/AR4157.
78. Suyama, K., Silagi, E.S., Choi, H., Sakabe, K., Mochida, J., Shapiro, I.M., and Risbud, M.V. (2016). Circadian factors BMAL1 and RORα control HIF-1α transcriptional activity in nucleus pulposus cells: implications in maintenance of intervertebral disc health. Oncotarget *7*, 23056–23071. https://doi.org/10.18632/oncotarget.8521.
79. Fearing, B.V., Speer, J.E., Jing, L., Kalathil, A., P Kelly, M., M Buchowski, J., P Zebala, L., Luhmann, S., C Gupta, M., A Setton, L., et al. (2020). Verteporfin treatment controls morphology, phenotype, and global gene expression for cells of the human nucleus pulposus. JOR spine *3*, e1111. https://doi.org/10.1002/JSP2.1111.
80. Minegishi, K., Hashimoto, M., Ajima, R., Takaoka, K., Shinohara, K., Ikawa, Y., Nishimura, H., McMahon, A.P., Willert, K., Okada, Y., et al. (2017). A Wnt5 Activity Asymmetry and Intercellular Signaling via PCP Proteins Polarize Node Cells for Left-Right Symmetry Breaking. Dev. Cell *40*, 439–452.e4. https://doi.org/10.1016/j.devcel.2017.02.010.
81. Sun, H., Wang, H., Zhang, W., Mao, H., and Li, B. (2023). Single-cell RNA sequencing reveals resident progenitor and vascularization-associated cell subpopulations in rat annulus fibrosus. J. Orthop. Translat. *38*, 256–267. https://doi.org/10.1016/J.JOT.2022.11.004.
82. Risbud, M.V., Guttapalli, A., Tsai, T.T., Lee, J.Y., Danielson, K.G., Vaccaro, A.R., Albert, T.J., Gazit, Z., Gazit, D., and Shapiro, I.M. (2007). Evidence for skeletal progenitor cells in the degenerate human intervertebral disc. Spine *32*, 2537–2544. https://doi.org/10.1097/BRS.0B013E318158DEA6.
83. Ellis, K., Bagwell, J., and Bagnat, M. (2013). Notochord vacuoles are lysosome-related organelles that function in axis and spine morphogenesis. J. Cell Biol. *200*, 667–679. https://doi.org/10.1083/jcb.201212095.
84. Barral, D.C., Staiano, L., Guimas Almeida, C., Cutler, D.F., Eden, E.R., Futter, C.E., Galione, A., Marques, A.R.A., Medina, D.L., Napolitano, G., et al. (2022). Current methods to analyze lysosome morphology, positioning, motility and function. Traffic *23*, 238–269. https://doi.org/10.1111/tra.12839.
85. Puri, M.C., and Nagy, A. (2012). Concise review: Embryonic stem cells versus induced pluripotent stem cells: the game is on. Stem Cell. *30*, 10–14. https://doi.org/10.1002/stem.788.
86. Narsinh, K.H., Plews, J., and Wu, J.C. (2011). Comparison of human induced pluripotent and embryonic stem cells: fraternal or identical twins? Mol. Ther. *19*, 635–638. https://doi.org/10.1038/mt.2011.41.
87. Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. PLoS One *5*, e13984. https://doi.org/10.1371/journal.pone.0013984.
88. Aszódi, A., Chan, D., Hunziker, E., Bateman, J.F., and Fässler, R. (1998). Collagen II Is Essential for the Removal of the Notochord and the Formation of Intervertebral Discs. J. Cell Biol. *143*, 1399–1412.
89. Codsi, E., Brost, B.C., Faksh, A., Volk, A.K., and Borowski, K.S. (2015). Persistent Notochord in a Fetus with COL2A1 Mutation. Case Rep. Obstet. Gynecol. *2015*, e935204. https://doi.org/10.1155/2015/935204.
90. Rebello, D., Wohler, E., Erfani, V., Li, G., Aguilera, A.N., Santiago-Cornier, A., Zhao, S., Hwang, S.W., Steiner, R.D., Zhang, T.J., et al. (2023). COL11A2 as a candidate gene for vertebral malformations and congenital scoliosis. Hum. Mol. Genet. *32*, 2913–2928. https://doi.org/10.1093/hmg/ddad117.
91. Pulina, M., Liang, D., and Astrof, S. (2014). Shape and position of the node and notochord along the bilateral plane of symmetry are regulated by cell-extracellular matrix interactions. Biol. Open *3*, 583–590. https://doi.org/10.1242/bio.20148243.
92. Peskin, B., Norman, J., Bagwell, J., Lin, A., Adhyapok, P., Di Talia, S., and Bagnat, M. (2023). Dynamic BMP signaling mediates notochord segmentation in zebrafish. Curr. Biol. *33*, 2574–2581.e3. https://doi.org/10.1016/j.cub.2023.05.039.
93. Richardson, S.M., Ludwinski, F.E., Gnanalingham, K.K., Atkinson, R.A., Freemont, A.J., and Hoyland, J.A. (2017). Notochordal and nucleus pulposus marker expression is maintained by sub-populations of adult human nucleus pulposus cells through aging and degeneration. Sci. Rep. *7*, 1501. https://doi.org/10.1038/s41598-017-01567-w.
94. Rito, T., Libby, A.R.G., Demuth, M., and Briscoe, J. (2023). Notochord and axial progenitor generation by timely BMP and NODAL inhibition during vertebrate trunk formation. Preprint at bioRxiv. https://doi.org/10.1101/2023.02.27.530267.
95. Martini, D., Giannaccini, M., Guadagni, V., Marracci, S., Giudetti, G., and Andreazzoli, M. (2019). Comparative analysis of p4ha1 and p4ha2 expression during Xenopus laevis development. Int. J. Dev. Biol. *63*, 311–316. https://doi.org/10.1387/ijdb.190067ma.
96. Pinskey, J.M., Hoard, T.M., Zhao, X.-F., Franks, N.E., Frank, Z.C., McMellen, A.N., Giger, R.J., and Allen, B.L. (2022). Plexins promote Hedgehog signaling through their cytoplasmic GAP activity. Elife *11*, e74750. https://doi.org/10.7554/eLife.74750.
97. Danesin, C., Darche-Gabinaud, R., Escalas, N., Bouguetoch, V., Cochard, P., Al Oustah, A., Ohayon, D., Glise, B., and Soula, C. (2021). Sulf2a controls Shh-dependent neural fate specification in the developing spinal cord. Sci. Rep. *11*, 118. https://doi.org/10.1038/s41598-020-80455-2.
98. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.
99. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
100. Barrionuevo, F., Taketo, M.M., Scherer, G., and Kispert, A. (2006). Sox9 is required for notochord maintenance in mice. Dev. Biol. *295*, 128–140. https://doi.org/10.1016/j.ydbio.2006.03.014.
101. Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Günther, T., Schüle, R., Kemler, R., Mallo, M., Kanzler, B., and Scherer, G. (2006). Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9

expression pattern. Dev. Biol. *291*, 382–397. https://doi.org/10.1016/j.ydbio.2005.11.013.

102. Tsingas, M., Ottone, O.K., Haseeb, A., Barve, R.A., Shapiro, I.M., Lefebvre, V., and Risbud, M.V. (2020). Sox9 deletion causes severe intervertebral disc degeneration characterized by apoptosis, matrix remodeling, and compartment-specific transcriptomic changes. Matrix Biol. *94*, 110–133. https://doi.org/10.1016/j.matbio.2020.09.003.
103. Lefebvre, V. (2019). Chapter Seven - Roles and regulation of SOX transcription factors in skeletogenesis. In Current Topics in Developmental Biology Vertebrate Skeletal Development, B.R. Olsen, ed. (Academic Press), pp. 171–193. https://doi.org/10.1016/bs.ctdb.2019.01.007.
104. Tamplin, O.J., Kinzel, D., Cox, B.J., Bell, C.E., Rossant, J., and Lickert, H. (2008). Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. BMC Genom. *9*, 511. https://doi.org/10.1186/1471-2164-9-511.
105. Tamplin, O.J., Cox, B.J., and Rossant, J. (2011). Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord. Dev. Biol. *360*, 415–425. https://doi.org/10.1016/j.ydbio.2011.10.002.
106. Wymeersch, F.J., Skylaki, S., Huang, Y., Watson, J.A., Economou, C., Marek-Johnston, C., Tomlinson, S.R., and Wilson, V. (2019). Transcriptionally dynamic progenitor populations organised around a stable niche drive axial patterning. Development *146*, dev168161. https://doi.org/10.1242/dev.168161.
107. Xu, Y., Zhang, T., Zhou, Q., Hu, M., Qi, Y., Xue, Y., Nie, Y., Wang, L., Bao, Z., and Shi, W. (2023). A single-cell transcriptome atlas profiles early organogenesis in human embryos. Nat. Cell Biol. *25*, 604–615. https://doi.org/10.1038/s41556-023-01108-w.
108. Haniffa, M., Taylor, D., Linnarsson, S., Aronow, B.J., Bader, G.D., Barker, R.A., Camara, P.G., Camp, J.G., Chédotal, A., Copp, A., et al. (2021). A roadmap for the Human Developmental Cell Atlas. Nature *597*, 196–205. https://doi.org/10.1038/s41586-021-03620-1.
109. Garcia, M.R., Thimonier, C., Angoura, K., Ozga, E., MacPherson, H., and Blin, G. (2023). In vitro modelling of anterior primitive streak patterning with hESC reveals the dynamic of WNT and NODAL signalling required to specify notochord progenitors. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.01.543323.
110. Edri, S., Hayward, P., Jawaid, W., and Martinez Arias, A. (2019). Neuro-mesodermal progenitors (NMPs): a comparative study between pluripotent stem cells and embryo-derived populations. Development *146*, dev180190. https://doi.org/10.1242/dev.180190.
111. Tzouanacou, E., Wegener, A., Wymeersch, F.J., Wilson, V., and Nicolas, J.-F. (2009). Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. Dev. Cell *17*, 365–376. https://doi.org/10.1016/j.devcel.2009.08.002.
112. José-Edwards, D.S., Kerner, P., Kugler, J.E., Deng, W., Jiang, D., and Di Gregorio, A. (2011). The identification of transcription factors expressed in the notochord of Ciona intestinalis adds new potential players to the brachyury gene regulatory network. Dev. Dynam. *240*, 1793–1805. https://doi.org/10.1002/dvdy.22656.
113. Behrens, A., Haigh, J., Mechta-Grigoriou, F., Nagy, A., Yaniv, M., and Wagner, E.F. (2003). Impaired intervertebral disc formation in the absence of Jun. Development *130*, 103–109. https://doi.org/10.1242/dev.00186.
114. Tanegashima, K., Zhao, H., Rebbert, M.L., and Dawid, I.B. (2009). Coordinated activation of the secretory pathway during notochord formation in the Xenopus embryo. Development *136*, 3543–3548. https://doi.org/10.1242/dev.036715.
115. Jo, A., Denduluri, S., Zhang, B., Wang, Z., Yin, L., Yan, Z., Kang, R., Shi, L.L., Mok, J., Lee, M.J., and Haydon, R.C. (2014). The versatile functions of Sox9 in development, stem cells, and human diseases. Genes Dis. *1*, 149–161. https://doi.org/10.1016/j.gendis.2014.09.004.
116. Liu, Y., Rahaman, M.N., and Bal, B.S. (2014). Modulating notochordal differentiation of human induced pluripotent stem cells using natural nucleus pulposus tissue matrix. PLoS One *9*, e100885. https://doi.org/10.1371/journal.pone.0100885.
117. Binch, A.L.A., Fitzgerald, J.C., Growney, E.A., and Barry, F. (2021). Cell-based strategies for IVD repair: clinical progress and translational obstacles. Nat. Rev. Rheumatol. *17*, 158–175. https://doi.org/10.1038/s41584-020-00568-w.
118. Briscoe, J., and Thérond, P.P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. Nat. Rev. Mol. Cell Biol. *14*, 416–429. https://doi.org/10.1038/nrm3598.
119. Dubourg, C., Kim, A., Watrin, E., de Tayrac, M., Odent, S., David, V., and Dupé, V. (2018). Recent advances in understanding inheritance of holoprosencephaly. Am. J. Med. Genet. C Semin. Med. Genet. *178*, 258–269. https://doi.org/10.1002/ajmg.c.31619.
120. Giampietro, P.F. (2012). Genetic aspects of congenital and idiopathic scoliosis. Scientifica (Cairo) *2012*, 152365. https://doi.org/10.6064/2012/152365.
121. Faye, P.A., Vedrenne, N., Miressi, F., Rassat, M., Romanenko, S., Richard, L., Bourthoumieu, S., Funalot, B., Sturtz, F., Favreau, F., and Lia, A.S. (2020). Optimized protocol to generate spinal motor neuron cells from induced pluripotent stem cells from charcot marie tooth patients. Brain Sci. *10*, 407–414. https://doi.org/10.3390/brainsci10070407.
122. Kilens, S., Meistermann, D., Moreno, D., Chariau, C., Gaignerie, A., Reignier, A., Lelièvre, Y., Casanova, M., Vallot, C., Nedellec, S., et al. (2018). Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. Nat. Commun. *9*, 360. https://doi.org/10.1038/s41467-017-02107-w.
123. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.
124. Meistermann, D., Bruneau, A., Loubersac, S., Reignier, A., Firmin, J., François-Campion, V., Kilens, S., Lelièvre, Y., Lammers, J., Feyeux, M., et al. (2021). Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. Cell Stem Cell *28*, 1625–1640.e6. https://doi.org/10.1016/j.stem.2021.04.027.
125. Marsh, S., Salmon, M., and Hoffman, P. (2023). Samuel-Marsh/scCustomize: Version 1.1.1. https://doi.org/10.5281/zenodo.7534950.
126. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
127. Gao, C.-H., Yu, G., and Cai, P. (2021). ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. Front. Genet. *12*, 706907.
128. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.
129. Burridge, P.W., Thompson, S., Millrod, M.A., Weinberg, S., Yuan, X., Peters, A., Mahairaki, V., Koliatsos, V.E., Tung, L., and Zambidis, E.T. (2011). A universal system for highly efficient cardiac differentiation of human induced pluripotent stem cells that eliminates interline variability. PLoS One *6*, e18293. https://doi.org/10.1371/journal.pone.0018293.
130. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.
131. Treindl, F., Ruprecht, B., Beiter, Y., Schultz, S., Döttinger, A., Staebler, A., Joos, T.O., Kling, S., Poetz, O., Fehm, T., et al. (2016). A bead-based western for high-throughput cellular signal transduction analyses. Nat. Commun. *7*, 12852. https://doi.org/10.1038/ncomms12852.
132. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296. https://doi.org/10.1186/s13059-019-1874-1.
133. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat polyclonal anti-TBXT	R&D Systems	Cat# AF2085, RRID: AB_2200235
Rabbit polyclonal anti-FOXA2	Cell Signaling Technology	Cat# 8186, RRID: AB_10891055
Rabbit polyclonal anti-SOX9	Millipore	Cat# AB5535, RRID: AB_2239761
Goat polyclonal anti-SOX17	R&D Systems	Cat# AF1924, RRID: AB_355060
Rabbit polyclonal anti-CD109	Abcam	Cat# ab203588, RRID: AB_2936927
Rabbit monoclonal anti-CD166	Abcam	Cat# ab109215, RRID: AB_2936926
Rabbit polyclonal anti-Caveolin 1	GeneTex	Cat# GTX100205, RRID: AB_1240559
Rabbit monoclonal anti-CDX2	Cell Signaling Technology	Cat# #12306, RRID: AB_2797879
Mouse monoclonal anti-Cytokeratin 8	Santa Cruz Biotechnology	Cat# sc-52324, RRID: AB_629847
Mouse monoclonal anti-Cytokeratin Pan (C11)	Cell Signaling Technology	Cat# 4545, RRID: AB_490860
Rabbit monoclonal anti-Fibronectin	Abcam	Cat# ab45688, RRID: AB_732380
Rabbit monoclonal anti-GATA4	Cell Signaling Technology	Cat# 36966, RRID: AB_2799108
Rabbit monoclonal anti-Nanog	Cell Signaling Technology	Cat# 4903, RRID: AB_10559205
Rabbit polyclonal anti-NOTO	Thermo Fisher Scientific	Cat# PA5-69748, RRID: AB_2690559
Rabbit monoclonal anti-Sox2	Cell Signaling Technology	Cat# 3579, RRID: AB_2195767
Rabbit monoclonal anti-Sox9	Abcam	Cat# ab185230, RRID: AB_2715497
Goat polyclonal anti-TBX6	R&D Systems	Cat# AF4744, RRID: AB_2200834
Donkey Anti-Goat IgG (H+L) Antibody, Alexa Fluor 488 Conjugated	Molecular Probes	Cat# A-11055, RRID: AB_2534102
Donkey Anti-Rabbit IgG (H+L) Antibody, Alexa Fluor 568 Conjugated	Molecular Probes	Cat# A10042, RRID: AB_2534017
R-Phycoerythrin AffiniPure F(ab')$_2$ Fragment Goat Anti-Mouse IgG (H+L)	Jackson Immuno-Research Labs	Cat# 115-116-146, RRID: AB_2338629
R-Phycoerythrin AffiniPure F(ab')$_2$ Fragment Donkey Anti-Rabbit IgG (H+L)	Jackson Immuno-Research Labs	Cat# 711-116-152, RRID: AB_2340599
Biological samples		
Human fetal notochord	Gleneagles, Hospital, Hong Kong	F013NCC
Chemicals, peptides, and recombinant proteins		
ESC-qualified Matrigel	Corning	Cat# 354277
Geltrex™ LDEV-Free, hESC-Qualified, Reduced Growth Factor Basement Membrane Matrix	Thermo Fisher Scientific	Cat# A1413302
mTeSR1	Stemcell Technologies	Cat# 85850
mTeSR Plus	Stemcell Technologies	Cat# 100-0276
SB431542	Axon Medchem	Cat# 1661
Lipofectamine RNAiMAX Transfection Reagent	Life Technologies	Cat# 13778075
Opti-MEM™ I Reduced Media	Life Technologies	Cat# 31985062
TrypLE	Life Technologies	Cat# 12604021
DPBS without Ca2+/Mg2+	Life Technologies	Cat# 14190-094
Hanks Balanced Salt Solution	Gibco	Cat# 14175-079
DMEM F-12	Life Technologies	Cat# 11330
MEM non-ess AA	Life Technologies	Cat# 11140

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Glutamax	Life Technologies	Cat# 35050
XF passaging solution	Miltenyi	Cat# 130-104-688
ReLeSr	Stemcell Technologies	Cat# 05872
Dispase	Worthington	Cat# LS02109
Collagenase P	Roche	Cat# 11213865001
Collagenase type II	Worthington	Cat# LS004174
Beta-mercaptoethanol	Life Technologies	Cat# 31350
Dulbecco's modified eagles medium	Gibco	Cat# 21331-020
N2 supplement	Life Technologies	Cat# 17502048
B27 supplement	Life Technologies	Cat# 17504044
KnockOut™ Serum Replacement	Life Technologies	Cat# 10828
Albumin bovine	Sigma	Cat# A1595
Bovine Albumin Fraction V	Thermo Fisher Scientific	Cat# 15260037
PBS Ca2+/Mg2+	Sigma	Cat# D8662
Hybri-Max™ DMSO	Sigma	Cat# D2650
Rock Inhibitor Y-27632 Dihydrochloride	Peprotech	Cat# 1293823
Recombinant Human Noggin	Miltenyi	Cat# 130-103-456
CHIR99021	Axon Medchem	Cat# 1386
Laminin 521	Corning	Cat# 354223
Antifadent Mountant Solution	Cliniscience	Cat# AF1
Hoechst 33342	Life technologies	Cat# H3569
Critical commercial assays		
Nucleospin II RNA Kit	Macherey Nagel	Cat# 740955
TruSeq ChIP Sample Prep Kit	Illumina	IP-202-1012
Chromium Single Cell 3' GEM kit v3	10X Genomics	PN-1000075
Chromium Single Cell 3' GEM kit v3.1	10X Genomics	PN-1000121
Deposited data		
Original raw data for fetal notochord	This paper	Deposited on GEO GSE237542
Original raw data for H1-CFS differentiated NLCs	This paper	Deposited on GEO GSE237546
Original raw data for iPS-NLC	This paper	Deposited on GEO GSE237538
Full code and post-processed single cell data	This paper	https://github.com/linxy29/NCL_differentiation
Experimental models: Cell lines		
hiPSC [a]	Life Technologies	Cat# A18944, Lot Number: 1938075
hiPSC [b]	This paper	hPSCreg Name: REGUi016-D 4F_PBMC_Sv190
hiPSC [c]	Life Technologies	Cat# A18944, Lot Number: 2228383
hiPSC [d]	Faye et al., 2020[121]	hPSCreg Name: LIMFRi001-A
hiPSC [e]	Kilens et al., 2018[122]	N/A
H1-ESC	Prof. Andras Nagy, Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital in Toronto, Ontario	HA01
Experimental models: Organisms/strains		
CD1-Swiss mice	Charles River Laboratory	N/A
Oligonucleotides		
Synthetic NOTO mRNA	Trilink	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TaqMan probe TBXT	Life Technologies	Hs00610080-m1
TaqMan probe FOXA2	Life Technologies	Hs00232764_m1
TaqMan probe GAPDH	Life Technologies	Hs99999905_m1
TaqMan probe LEF1	Life Technologies	Hs01547250-m1
TaqMan probe LEFTY1	Life Technologies	Hs00764128_s1
TaqMan probe NODAL	Life Technologies	Hs00415443_m1
TaqMan probe NOGGIN	Life Technologies	Hs00271352_s1
TaqMan probe NOTO	Life Technologies	Hs01377437_m1
TaqMan probe SHH	Life Technologies	Hs00179843_m1
TaqMan probe SOX17	Life Technologies	Hs00751752_s1
TaqMan probe TBX6	Life Technologies	Hs00365539_m
TaqMan probe NANOG	Life Technologies	Hs04260366_g1
TaqMan probe POU5F1	Life Technologies	Hs00999632_g1
TaqMan probe GATA4	Life Technologies	Hs00171403_m1
TaqMan probe HAND1	Life Technologies	Hs02330376_s1
TaqMan probe PRRX1	Life Technologies	Hs00246567_m1
TBXT primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CTATTCTGACAACTCACCTGCAT R (5′→3′) ACAGGCTGGGGTACTGACT
FOXA2 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) GGAGCAGCTACTATGCAGAGC R (5′→3′) CGTGTTCATGCCGTTCATCC
RPLP0 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) TGGCAGCATCTACAACCCTGAAGT R (5′→3′) ACACTGGCAACATTGCGGACA
LEF1 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) TGCCAAATATGAATAACGACCCA R (5′→3′) GAGAAAAGTGCTCGTCACTGT
LEFTY1 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CTGTGACCCTGAAGCACCAA R (5′→3′) CATCCCCTGCAGGTCAATGT
NODAL primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CTGCTTAGAGCGGTTTCAGATG R (5′→3′) CGAGAGGTTGGAGTAGAGCATAA
NOGGIN primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CCATGCCGAGCGAGATCAAA R (5′→3′) TCGGAAATGATGGGGTACTGG
NOTO primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CTGAGGGCAGCAGTTACAT R (5′→3′) CTTCTGGTTGAGGAGGCTTT
SHH primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CCAAGGCACATATCCACTGCT R (5′→3′) GTCTCGATCACGTAGAAGACCT
SOX17 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) GTGGACCGCACGGAATTTG R (5′→3′) GGAGATTCACACCGGAGTCA
TBX6 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) TACATTCACCCCGACTCTCC R (5′→3′) TGTATGCGGGGTTGGTACTT

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NANOG primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) AAGGTCCCGGTCAAGAAACAG R (5′→3′) CTTCTGCGTCACACCATTGC
POU5F1 primers	SYBR green – IDT (Integrated DNA Technologies)	F (5′→3′) CAAAGCAGAAACCCTCGTGC R (5′→3′) TCTCACTCGGTTCTCGATACTG
Software and algorithms		
ImageJ	NIH Schneider et al., 2012[123]	https://imagej.nih.gov/ij/
Volocity® software	Quorum Technologies Inc.	Version 6.0.0.
Cell Ranger	10x Genomics	Version 6.0.2 and 6.1.2
R	Bioconductor	Version 4.1.0
CellChat	Jin et al. 2021[98]	Version 1.6.0
SCENIC	Aibar et al. 2017[99]	Version 1.2.4
WGCNA	Meistermann et al. 2021[124]	https://gitlab.univ-nantes.fr/E114424Z/WGCNA
scCustomize	Marsh SE, 2021[125]	Version 1.1.1
Seurat	Hao et al., 2021[126]	Version 4.3.0
ggVennDiagram	Gao et al., 2021[127]	Version 1.2.2
VeloAE	Qiao et al., 2021[53]	Version 0.2.0
scVelo	Bergen et al., 2020[128]	Version 0.2.4

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by, Dr. Anne Camus at anne.camus@univ-nantes.fr.

Materials availability

There are restrictions on the availability of the H1-derived cloaked FailSafeTM embryonic stem cell line due to a material transfer agreement. This study did not generate new unique reagents.

Data and code availability

- Single-cell RNA-seq data have been deposited at GEO and are publicly accessible as of the date of publication. Accession numbers are listed in the key resources table.
- The custom scripts for processing and analysing the data were housed at https://github.com/linxy29/NCL_differentiation.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines generation and characterization

hiPSC[a] and hiPSC[c] line derives from different batch (Lot Number: 2228383 and Lot Number: 1938075 repectively) of The Gibco® Episomal hiPSC line which is generated using cord blood-derived CD34$^+$ progenitors and seven factors that are expressed episomally (Oct4, Sox2, Klf4, Myc, Nanog, Lin28, and SV40 T).[129] The hiPSC line was initially cultured on mouse feeder cells and later adapted to a culture condition that is both feeder-free and serum-free. All quality controls are done by the supplier LifeTechnologies.

hiPSC[b] line is generated from peripheral blood mononuclear cells (PBMC) using Virus CytoTune-iPS 2.0 Sendai Reprogramming Kit stemgent StemRNA 3rd Gen Reprogramming Kit (Stemgen).

hiPSC[d], generated from dermal fibroblast cells, were maintained and reprogrammed according to a previously published protocol.[121] Briefly hiPSC were generated following the iStem (INSERM/UEVE UMR861, AFM, Genopole, Evry, France) protocol with three plasmids (Plasmid #6 pCXLE-hOCT3/4 shp53-F Addgene, Plasmid #7 pCXLE-hSK Addgene, Plasmid #8 pCXLE-hUL Addgene) using the Nucleofector II device (Amaxa, Lonza, Basel Switzer-land).

hiPSC[e] is generated by overexpression of *OCT4, KLF4, MYC* and *SOX2* in human fibroblasts from 3 healthy donors, using a non-integrative Sendai virus as describe in Kilens et al.[122]

H1-ESC derived cloaked/fail-safe (CFS) embryonic stem cell line was developed in the laboratory of Andras Nagy, of which the cells originated from the H1 embryonic stem cell line HA01 (Wi Cell, USA).

HiPSC maintenance

Stem cell culture was achieved in line with the ISSCR guidelines (https://www.isscr.org/standards-document). HiPSC lines used in this study were maintained on ESC-qualified Matrigel (Corning), coated plates with mTeSR1 media (Stem cell technology). All cell lines were cultured at 37°C, under normoxic conditions (20% O^2, 5% CO_2) and culture media was daily replaced. HiPSC were passaged every 4 days at a 1:6 split ratio using passaging solution XF (Miltenyi) for 2 min, at room temperature for dissociation into cell clusters. Cells were grown for 10 passages maximum. A new vial of the cell bank was used when cell reach P30. The MycoAlert kit (LONZA) was employed to confirm the absence of mycoplasma in all cell lines.

The H1-ESC line was developed in the laboratory of Andras Nagy, of which the cells originated from the H1 embryonic stem cell line HA01 (Wi Cell, USA). The CFS were cultured in 6-well culture plates coated with Geltrex™ LDEV-Free, hESC-Qualified, Reduced Growth Factor Basement Membrane Matrix (Thermofisher Scientific). Cells were maintained in mTeSR™ Plus (Stemcell Technologies) that was refreshed daily (1.5mL) and passaged every 4-5 days using ReLeSr (Stemcell Technologies) as per manufacturers' instructions.

Clinical specimens

Clinical specimens were obtained with approval by the Institutional Review Board (reference UW13-577) and with informed consent in accordance with the Helsinki Declaration of 1975 (revision 1983). Fetal tissue of unknown gender was obtained from patients undergoing legally induced abortion at the Gleneagles Hospital, Hong Kong. Primary PBMC cells for hiPSC line b generation were obtained with approval of the Ethical Committee of Bern (reference 2019-00097).

Mice

Experiments on mice were conducted according to the French and European regulations on care and protection of laboratory animals (EC Directive 86/609, French Law n°2013-118 issued on February 1st 2013). Embryos and spines for CD166, CD109 and CAV1 protein immunostaining analyses were collected from pregnant mice and newborns from the CD1-Swiss background.

METHOD DETAILS

Differentiation of pluripotent stem cells into NLC

Differentiation protocol follows three-steps: (1) experiment starts with the seeding of single PSC at 20.000 cells/cm^2 on Laminin 521 (Corning)-coated plates in mTseR1 medium supplemented with rock inhibitor for 24 hours; (2) cells are then cultured in specific media for mesendoderm progenitor specification [DMEM/F12 supplemented with 1% N2 (Gibco), 1% B27 (Gibco), 1% non-essential amino acids, 1% GlutaMAX (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 50 mg/ml BSA and 3μM CHIR99021 (Axon Medchem)]; (3) at day 2, mesendoderm progenitor cells are dissociated and transfected with Lipofectamin RNAimax (5:1) in a single cell suspension with 1500 ng of human NOTO synthetic mRNA (vector and co-transcriptional capping technology Trilink) and seeded at 160 000 cells/cm^2. From day 2 up to day 7, cells are maintained in same media supplemented with 100ng recombinant human Noggin (Miltenyi) allowing the emergence of embryonic NLC population. For inhibition of TGF-β1 pathway, medium was supplemented with 5μM SB431542 (Axon Medchem) from day 2 to day 7. The iPS- and ESC- line used for each experiment in each figure panel is summarized in Table S1.

RNA Extraction and RT-qPCR

Total RNA was extracted with the Nucleospin II RNA Kit (Macherey Nagel, 740955) and one microgram was reverse transcribed using Superscript III First Strand synthesis kit (Life technologies, 11752). Quantitative RT-qPCR experiments were performed using TaqMan or SYBR technologies with GAPDH as housekeeping gene and fold change represented using a base 2 logarithm determined by the Livak Method (Relative quantification RQ = $2^{-\Delta\Delta Cq}$).[130]

Immunostaining

Cells were fixed at room temperature using 4% paraformaldehyde for 10 min. Samples were then saturated and permeabilized for 1 hour at room temperature with a phosphate-buffered saline (PBS) containing, 0.1% Triton, 3% bovine serum albumin (BSA). Cells were incubated with primary antibodies overnight at 4 °C in a PBS solution containing 1% BSA. The following antibodies were used anti-TBXT (1:400, R&D), anti-FOXA2 (1:400, Cell Signaling), anti-SOX9 (1:1000, Milipore), anti-SOX17 (1:200, R&D). After washing primary antibodies with PBS, incubation with secondary antibodies was performed using donkey anti-goat IgG (H+L) antibody, Alexa Fluor 488 conjugated (1/1000, Molecular Probes) or donkey anti-rabbit IgG (H+L) antibody, Alexa Fluor 594 conjugated (1/1000, Molecular Probes) for 1 hour at room temperature in PBS containing 1% BSA. After washing secondary antibodies with PBS, cells were counterstained with Hoechst (Life technologies) to visualize nuclei and stored in CitiFluor™ AF1 (Electron Microscopy Sciences, #1179 70-25) at 4°C. Confocal immunofluorescence images were acquired with LFOV FLIM Nikon® confocal microscope.

Embryos and mice spines were embedded in Super Cryoembedding Medium (SCEM) (Section Lab, Hiroshima, Japan) and frozen in isopentane/dry ice, without decalcification. Samples were cut along coronal plane and 7μm sections were mounted on slides (CryoStar NX70,

Thermo Fisher Scientific, Waltham, Massachusetts, USA) and post-fixed in 2% PFA (2 min) for anti-CD166 and anti-CAV1 staining, and in 4% PFA (2min) for anti-CD109 staining. Sections were then permeabilized in 0.2% Triton X-100 in PBS for 20 min and blocked in the blocking solution (10% Foetal Bovine Serum; 4% BSA; 0.1% Triton X-100 in PBS) for 1 h. Next, sections were incubated overnight at 4°C with primary antibodies against CAV1 (1:100, Clinisciences, GTX-100205), CD166 (1:600 or 1:800 depending on the embryonic stage, Abcam), CD109 (1:200, Abcam) followed by a 1h incubation in secondary antibody (either Alexa Fluor-488 or Alexa Fluor-568, 1:1000, Molecular Probes). After washing secondary antibodies with PBS, samples were counterstained with Hoechst (Life technologies) to visualize nuclei and sections were mounted with ProLong™ (Thermofisher, P36934). Confocal immunofluorescence images were acquired with LFOV FLIM Nikon® confocal microscope.

DigiWest

The cells were lysed in radioimmunoprecipitation assay buffer (RIPA) containing proteinase inhibitor cocktails. The lysates were then clarified through centrifugation at 18,000 g for 15 minutes at 4°C. The clarified lysates were boiled in a sample buffer at 95°C for 5 minutes and resolved by SDS-PAGE. DigiWest assays were performed using 15 mg of cellular protein loaded onto a sodium dodecyl sulfate-polyacrylamide gel, followed by size separation via electrophoresis. The size-separated proteins were transferred onto a polyvinylidene fluoride membrane and biotinylated using 50 μM NHS-PEG12-Biotin in phosphate-buffered saline (PBS) supplemented with Tween-20 for 1 hour. After drying of the membrane, the sample lanes were cut into 96 strips, each corresponding to a specific molecular weight fraction. Each strip was placed in a well of a 96-well plate, and 10 μl elution buffer (8 M urea, 1% Triton-X100 in 100 mM Tris-HCl pH 9.5) was added to extract the proteins. The eluted proteins were then diluted with 90μl dilution buffer (5% BSA in PBS, 0.02% sodium azide, 0.05% Tween-20) and incubated with a distinct magnetic color-coded bead population coated with neutravidin. The biotinylated proteins bound to the neutravidin beads, with each bead color representing proteins of a specific molecular weight fraction. The 96 protein-loaded bead populations were mixed to reconstitute the original lane, providing a bead mix for approximately 100 individual antibody incubations. Aliquots of the DigiWest bead-mixes were added to 96-well plates containing an assay buffer supplemented with antibodies. After overnight incubation at 15°C, the bead-mixes were washed, and phycoerythrin-labelled secondary antibodies were added and incubated for 1 hour at 23°C. The beads were washed again prior to readout using a Luminex FlexMAP 3D instrument. For quantification of antibody-specific signals, an Excel-based analysis tool was used to identify peaks corresponding to proteins of appropriate molecular weight and calculate the peak area. The signal intensity was then normalized to the total amount of protein loaded into one lane.[131] Hierarchical clustering was performed with complete linkage method and Pearson correlation for validated gene versus two hiPSC lines from different time point.

Single-cell isolation of fetal notochord and *in vitro* differentiated pluripotent stem cells

Single cell encapsulation and cDNA libraries of fetal notochord and differentiated ESC were prepared at the Centre for PanorOmic Sciences (CPOS) at The University of Hong Kong. Single cell encapsulation and cDNA libraries of differentiated iPSC were prepared at the GenoA platform at Nantes University.

Under the guidance of a stereomicroscope, the fetal tissue was rinsed with Hanks balanced salt solution (HBSS, Gibco) to reveal an intact fetal notochord embedded within the developing vertebral column. The fetal notochord (NC) was dissected out, cut into ~1mm pieces, and digested with 10mL pre-warmed digestion buffer containing 0.5% dispase (5mg/mL), 0.5% collagenase type II (5mg/mL), 0.25mg/ml collagenase P in FBS-free DMEM for 1hr at 37°C on a rocking platform. The digest was gently triturated every 30 min to further aid in the release of cells from the tissue. After digestion, the cells were collected, filtered through a 40μm sieve, and resuspended in 0.04% BSA in phosphate buffered saline (PBS). 20,000 cells in suspension were prepared and loaded into individual wells of 10X Chromium Single Cell chip. Single cells were then encapsulated into Gel Beads-in-emulsion (GEM) by 10X Chromium Single Cell Controller. The Single Cell 3′ Reagent Kit v3 was used to perform downstream steps.

Differentiated iPSC at day 7 were dissociated with TryplE for 7 minutes at 37°C, then $1.5x10^6$ cells from each sample were labelled with a specific lipid-coupled oligonucleotide sequence (CG000388, 10X Multiplexing). After labelling, cells were washed and resuspended at a concentration of 1,200 cells/μL in PBS containing 0.1% BSA. A total of 40,000 cells were loaded into a Chromium Controller (10X Genomics), containing 5,000 not-transfected cells, 10,000 cells NOTO transfected, 5,000 cells NOTO transfected with 5μM SB431542 condition added at day 2, and for the hiPSC lines a and b.

Differentiated ESC at day 7 were processed using the same method as for the notochord, but using the Chromium Next GEM Single Cell 3' Reagent Kit v3.1, Chromium Next GEM Chip G Single Cell Kit and Single Cell 3′ Reagent Kit v3.1.

In brief, single cells were encapsulated into Gel Beads-in-emulsion (GEM) by 10X Chromium Single Cell Controller. Single Cell 3′ Reagent Kit was used to perform reverse transcription, cDNA cleanup and amplification. The double-stranded cDNA underwent enzymatic fragmentation, adapter ligation, index PCR and SPRIselect size selection as per manufacturer's protocol. The library size and concentration were determined by Qubit (1X dsDNA HS Assay Kit) and Bioanalyser assays.

All libraries were sequenced using paired-end sequencing using the Illumina NovaSeq 6000 (#1000000019358v11 Material #20023471) following manufacturer's instructions. Raw sequencing data were processed with Cell Ranger 6.0.2, or 6.1.2 depending on the dataset.

Processing of the scRNAseq data

Primary analysis

Raw sequencing data were processed with Cell Ranger 6.0.2 for iPS-NLC dataset and with Cell Ranger 6.1.2 for ESC-NLC and fetal notochord dataset. CellRanger mkfastq function was used to generate FASTQ files, and FASTQ alignment to the human reference genome (CRCh38-2020-A), filtering, barcode and UMI counting were performed with CellRanger multi. Resulting UMI count tables were further analysed using Seurat v4.[126] For the fetal notochord sample, cells with more than 200 but less than 5000 genes, as well as less than 10% mitochondrial reads were analysed. For the ESC-NLC, those with more than 200 but less than 7500 genes, as well as less than 10% mitochondrial reads were analysed. For the iPS-NLC cells with more than 1000 but less than 8000 genes, with less than 10% mitochondrial reads were analysed. Cell Cycle regression was applied for foetal notochord dataset. The SCTransform method was applied for normalization, variance stabilization and feature selection of the gene expression matrix.[132]

Secondary analysis

Dimensionality reduction on the data was then performed by computing the significant principal components on highly variable genes. We then performed unsupervised clustering by using the FindClusters function in Seurat and clusters were then visualized in a UMAP plot. Differentially expressed genes (DEG) among each cell cluster were determined using the FindAllMarkers function in Seurat, and defined as the positive genes expressed in at least 10% of the cells within the cluster and with a fold change of more than 0.25. Each dataset was annotated based on differential gene expression for markers related to cell lineage. The percentage of cells per clusters was calculated by normalizing the number of cells per cluster per condition to the cells numbers that were included in the analyses.

veloAE was used in each sample separately.[53] We used the parameter 'palette' to specify the colors of each cell type so that the colors are the same as the UMAP we generated using Seurat. The velocity genes are calculated by the function 'rank_velocity_genes' in the scVelo[128] as these values are calculated from scVelo and do not change in veloAE. Except for the parameters we discussed before, all the rest parameters are set as default.

WGCNA was performed on normalized iPS-NLC dataset using a soft threshold of 8 with signed Pearson correlation.[54,124] Resulting modules were manually curated to choose a set of 15 modules that were well represented in data and that have distinct behaviors. The gene matrix expression for each module was used as an input for single-cell regulatory network inference and clustering (SCENIC) analyses. The gene co-expression network was identified using GENIE3. Regulons were identified via RcisTarget. The activity of each regulon for each single cell was determined via the AUC scores using the AUCell R package.[99]

ESC-NLC and iPS-NLC datasets were integrated using the SelectIntegrationFeatures (3000 variable genes) and FindIntegrationAnchors functions using SCT normalization method. The integration object subsequently: runPCA, UMAP, neighbours, and clustered. FindAllMarkers by default was used to retrieve all the markers differentially expressed in each cluster.

Raw data from a previously published study[61] were obtained from Zhou and colleagues and processed (QC, Threshold for DEG and Clustering). This primary analysis allowed the retrieval of two out of four 7 weeks human fetal notochord specimens (samples 3 and 4 noted NOTO, Zhou study) that we filtered based on *TBXT* expressing cells a consensus marker for notochordal identity. Enrichment map was built on hypergeometric testing using the top100 DEG of each dataset to test the similarities between the three datasets of each of their cell type clusters.[87] We used a stringent p-value cut off at 0.00001 in order to reveal the significant associations between all three datasets.

CellChat was used to analyze the expression abundance of ligand-receptor interactions between two cell types on the basis of the expression of a ligand by one cell type and a receptor by another cell type. The communication probability was computed using trimean method. Chord diagram visualization were used to show specific outgoing signaling from the notochord to the other cell clusters.[98] Common gene signature were defined using published gene lists for node and notochord in mouse from Tamplin et al., 2008,[104] Tamplin et al., 2011,[105] Peck et al., 2017,[63] Wymeerch et al., 2019[106] and compared to the gene lists from this work for the notochord cluster. Overlap and visualization was performed using ggVennDiagram package.[127]

ComplexHeatmap was used to compare the gene signatures between human fetal and *in vitro* datasets.[133]

QUANTIFICATION AND STATISTICAL ANALYSIS

Confocal microscopy image analysis

For each sample, 3 pictures 3X3 were taken randomly within the well, using the X20 objective. These images are then analysed using the volocity® software (Quorum technologies), which allows us to determine the positive cells for each marker. The data represent an average of the percentage obtained for 10 independent experiments and derived from 4 hiPSC lines.