



Article

# The Cellular Composition of Bovine Coccygeal Intervertebral Discs: A Comprehensive Single-Cell RNAseq Analysis

Martina Calió <sup>1,2,3</sup>, Benjamin Gantenbein <sup>1,2</sup> , Marcel Egli <sup>3</sup> , Lucy Poveda <sup>4,†</sup> and Fabian Ille <sup>3,\*,†</sup>

<sup>1</sup> Tissue Engineering for Orthopaedics & Mechanobiology (TOM), Department for BioMedical Research (DBMR), University of Bern, 3008 Bern, Switzerland; martina.calio@students.unibe.ch (M.C.); benjamin.gantenbein@dbmr.unibe.ch (B.G.)

<sup>2</sup> Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, 3010 Bern, Switzerland

<sup>3</sup> Space Biology Group, Institute of Medical Engineering, School of Engineering and Architecture, Lucerne University of Applied Sciences and Arts, 6052 Hergiswil, Switzerland; marcel.egli@hslu.ch

<sup>4</sup> Functional Genomics Center Zurich, Swiss Federal Institute of Technology, University of Zurich, 8057 Zurich, Switzerland; lucy.poveda@fgcz.uzh.ch

\* Correspondence: fabian.ille@hslu.ch; Tel.: +41-41-349-36-15

† These authors contributed equally to this work.

**Abstract:** Intervertebral disc (IVD) degeneration and its medical consequences is still one of the leading causes of morbidity worldwide. To support potential regenerative treatments for degenerated IVDs, we sought to deconvolute the cell composition of the nucleus pulposus (NP) and the annulus fibrosus (AF) of bovine intervertebral discs. Bovine calf tails have been extensively used in intervertebral disc research as a readily available source of NP and AF material from healthy and young IVDs. We used single-cell RNA sequencing (scRNAseq) coupled to bulk RNA sequencing (RNAseq) to unravel the cell populations in these two structures and analyze developmental changes across the rostrocaudal axis. By integrating the scRNAseq data with the bulk RNAseq data to stabilize the clustering results of our study, we identified 27 NP structure/tissue specific genes and 24 AF structure/tissue specific genes. From our scRNAseq results, we could deconvolute the heterogeneous cell populations in both the NP and the AF. In the NP, we detected a notochordal-like cell cluster and a progenitor stem cell cluster. In the AF, we detected a stem cell-like cluster, a cluster with a predominantly fibroblast-like phenotype and a potential endothelial progenitor cluster. Taken together, our results illustrate the cell phenotypic complexity of the AF and NP in the young bovine IVDs.

**Keywords:** intervertebral disc; nucleus pulposus; annulus fibrosus; RNA sequencing; bovine



**Citation:** Calió, M.; Gantenbein, B.; Egli, M.; Poveda, L.; Ille, F. The Cellular Composition of Bovine Coccygeal Intervertebral Discs: A Comprehensive Single-Cell RNAseq Analysis. *Int. J. Mol. Sci.* **2021**, *22*, 4917. <https://doi.org/10.3390/ijms22094917>

Academic Editor: Inbo Han

Received: 19 March 2021

Accepted: 27 April 2021

Published: 6 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intervertebral disc (IVD) degeneration is a complex disease for which resolutive therapies are not yet available. Attempts to regenerate the tissue are complicated by the lack of knowledge on the exact cellular composition of the disc. Many studies have tracked and proven the developmental characteristics of IVDs [1–3]. However, little is known about the type of cells populating the mature tissue. IVDs are cushion-like structure between the vertebrae that act as stress and impact absorbers and as a protection for the spinal nerves. The IVD poses extreme conditions for cells to live in, leading to a hyperosmotic, avascular and acidic environment [4]. They provide viscoelastic and mechanical properties with shock-absorbing capacity that allow bending, twisting and loading of the upper body [5–7].

Structurally, the IVD is composed of three main compartments: the cartilaginous endplate (CEP), the annulus fibrosus (AF) and the nucleus pulposus (NP) [8]. The CEP, a thin layer of hyaline cartilage at the interface with the vertebrae, contains chondrocytes embedded in an aggrecan and collagen type II rich extracellular matrix [8,9]. It is composed of a network of collagen fibrils and hydrated proteoglycan molecules. Vascularized by microscopic vessels during development, they disappear once the IVD has reached

maturity [9]. The annulus fibrosus (AF) is a compartment originating from mature somites. It is believed that the AF develops from the mesenchymal cells of the sclerotome that generate the connective tissues of the future axial skeleton [10]. The annulus is connected to the CEP via its lamellae, structures made of fibrous cartilage and mainly composed of water and concentric sheets of collagen fibers [5]. The lamellae, oriented at various angles, enclose the nucleus pulposus (NP).

The NP, which is believed to have derived from the notochord [11–13], is composed of a gel-like material with a loose network of collagen fibers [1,6,14–16]. On a cellular level, the outer human AF is populated by elongated and fusiform cells that extend in the long axis of the fibrils. The cells of the inner AF, instead, are spherical in shape, and many resemble chondrocytes [3]. The annulus fibrosus is believed to retain the capacity to regenerate itself during the neonatal phase, and this is then lost during maturation [17]. Annulus fibrosus cell (APC) markers described in the literature include *Col1a1*, *Col4*, *Lam1* and *Thy1*, among others [10,18–24]. Instead, the young NP presents an abundant extracellular matrix (ECM) that separates NP cells (NPC) [8]. Numerous NPC type populations have been described in the literature. Among them, there are large cells with numerous vacuoles that are believed to be notochordal remnants [25]. NPC progenitors have also been described in the healthy IVD that are TIE2+ and GD2+ positive [26]. Additionally, cells that express a mesenchymal stem cell phenotype have also been reported [27]. The NPC population changes with aging and degeneration leading to the disappearance of the aforementioned progenitor cells and vacuolated cells [27–30]. There are a wide variety of NPC biomarkers such as *Krt8*, *Krt18*, *Krt19*, *Ca12*, *ACAN*, *Col1a2* and *Col2a1*, described in the literature [19,31–35]. Additional, notochordal markers such as *Brachyury*, *FOXA2*, *CHRD*, *NOG* and *LGALS3* are also often used to track NP cells lineage [36–40], together with *NCAM1*,  $\alpha 2$  macroglobulin and desmocollin 2 [36,41–43]. The mature IVD presents a very limited regenerative capacity [44]. However, a comprehensive profile of all of the NPC in this compartment is still missing [2,20,45]. This, in turn, hampers any attempts of regenerative therapies in patients suffering from disc degeneration which is still one of the leading causes of disability worldwide [2,46,47].

The purpose of our study was to deconvolute the cell composition in the AF and NP of healthy bovine IVD. Bovine IVDs are a widely used animal model for spine research as they offer the possibility to examine young and healthy tissues. Furthermore, the bovine model, while not a perfect substitute for the human model, has shown remarkable similarities in tissue specific marker expressions to the human model [41,48]. To reach our goal, we performed 10× Genomics single-cell RNA sequencing together with bulk RNAseq. Single-cell RNA sequencing is a powerful and innovative technique, which allowed the individual transcriptome analysis of 10 k NPCs and 10 k AFCs from two bovine tails.

## 2. Results

### 2.1. Single-Cell RNAseq Experimental Design

Eight samples were processed for whole transcriptome analysis using single-cell RNAseq sequencing (Table 1). Four samples were collected per biological replicate, i.e., the nucleus pulposus and annulus fibrosus from both, proximal and distal intervertebral discs of two bovine tails. Due to the nature of the composition of IVDs entailing a low cell number, each NP or AF sample consisted of a pool of three adjacent IVDs from either the distal or proximal region. To obtain a cell suspension, the ECM was digested before further processing.

**Table 1.** Samples scRNAseq. List of samples used for single-cell sequencing.

Sample N°	Description	Source
1	3 NPs proximal	Animal 1
2	3 AFs proximal	Animal 1
3	3 NPs distal	Animal 1
4	3 AFs distal	Animal 1
5	3 NPs proximal	Animal 2
6	3 AFs proximal	Animal 2
7	3 NPs distal	Animal 2
8	3 AFs distal	Animal 2

### 2.2. Bulk RNAseq Experimental Design

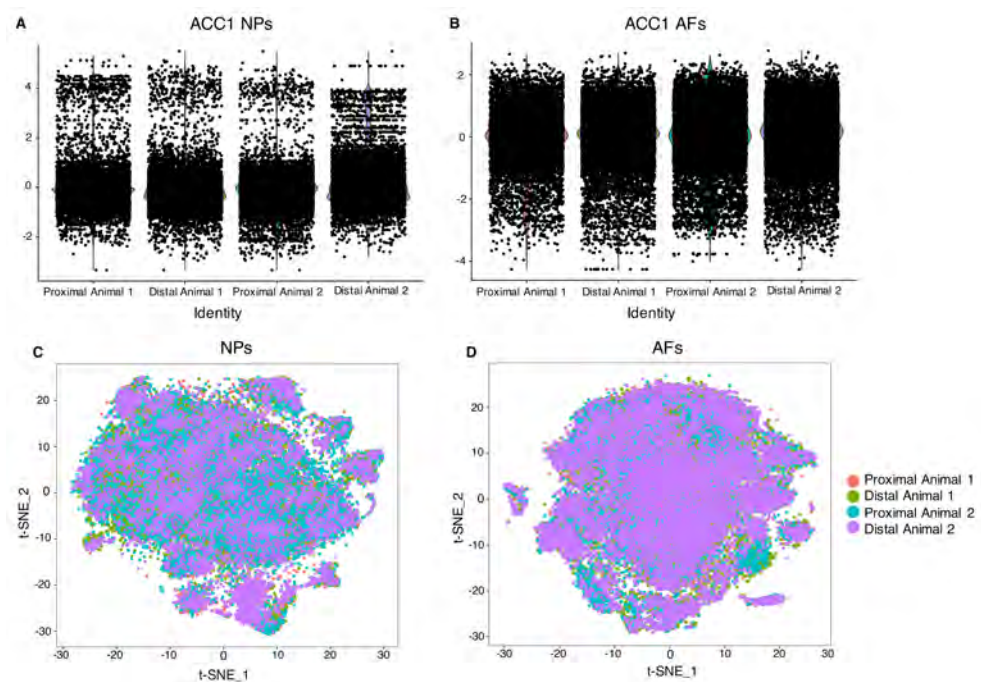
High-throughput transcriptome analysis, also referred to as bulk RNAseq, has become a standard tool for identifying differentially expressed genes between two or more sample groups. scRNAseq is a lot noisier due to technical and biological influences, compared to bulk RNAseq [49,50]. Three biological replicates (three bovine tails) were used to isolate cells for the bulk RNAseq analysis. Similar to the scRNAseq samples, distal and proximal NP and AF samples were processed for each of the three biological replicates. Tissue from two IVDs was pooled for each sample (Table 2), and the ECM was digested overnight to obtain a cell suspension. The cell suspension was then subjected to bulk RNAseq.

**Table 2.** Samples bulk sequencing. List of samples used for bulk RNA sequencing.

Sample N°	Description	Source
1	2 NPs proximal	Animal 3
2	2 AFs proximal	Animal 3
3	2 NPs distal	Animal 3
4	2 AFs distal	Animal 3
5	2 NPs proximal	Animal 4
6	2 AFs proximal	Animal 4
7	2 NPs distal	Animal 4
8	2 AFs distal	Animal 4
9	2 NPs proximal	Animal 5
10	2 AFs proximal	Animal 5
11	2 NPs distal	Animal 5
12	2 AFs distal	Animal 5

### 2.3. Cell Clustering Using Single-Cell RNAseq Data

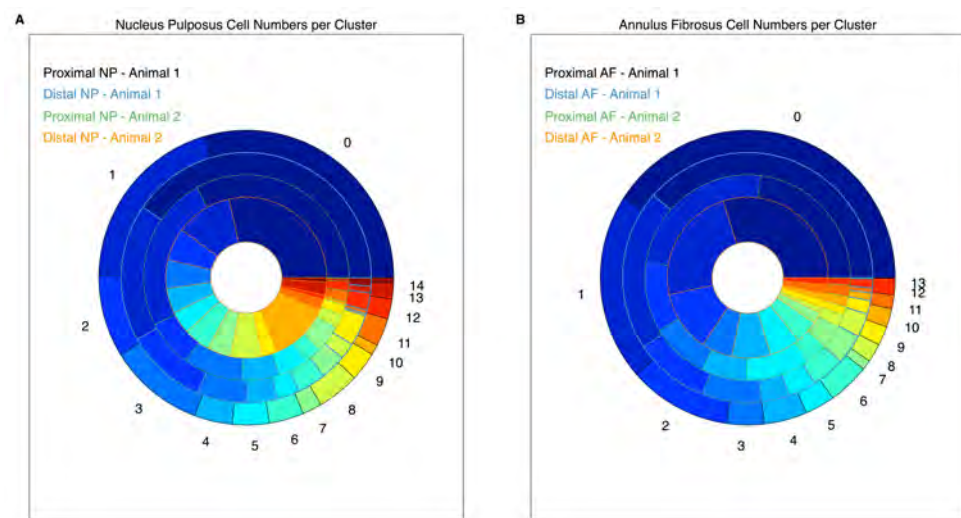
To comprehensively characterize subpopulations of NPs and AFs in either distal or proximal discs, we used the data from the scRNAseq experiments. Cell populations within each of the samples can be clustered according to their single-cell transcriptome profile expression values utilizing graph-based cluster method to acquire unsupervised cell clustering results. To obtain an impression of the overall gene expression variance within the NPs and the AFs, we performed a canonical correlation analysis (CCA) and corrected the datasets by aligning the relevant canonical correlations (CCs). The aligned CCs (ACC) for the NPs (Figure 1A) and the AFs (Figure 1B) exhibit a high correlation between the samples and animals. To better visualize the correlation, we performed t-stochastic neighbor embedding (t-SNE). The overlap of the samples in each t-SNE plot showed a high expression profile correlation among the NP cell populations (Figure 1C) and among the AF cell populations (Figure 1D), within and between animals, respectively.



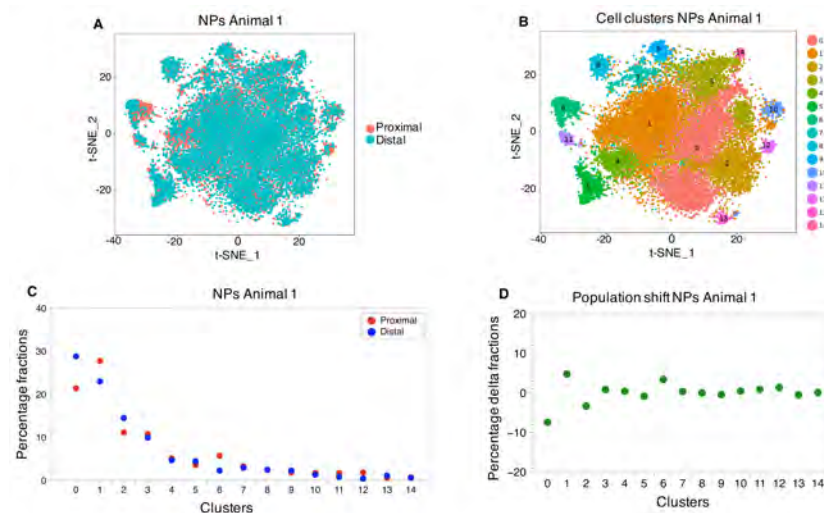
**Figure 1.** Canonical correlation analysis (CCA). CCA identifies common sources of variation between the datasets for NPs (A) and AFs (B). The CCA aligned subspaces ACC1 and ACC2 for each condition are plotted as violin plots with the points reflecting the data distribution. The plots reveal high correlation between proximal and distal NPs and AFs within and between animals. t-SNE plot of four CCA corrected NPs samples (C) and four CCA corrected AFs samples (D). Each dot represents a cell and the different colors are assigned to different samples. The location of a cell is determined by its gene expression profiles, therefore cells that localized in close proximity exhibit a high correlation in their gene expression pattern. In both graphs, the four samples contribute in a comparable way in populating each region of the graph, showing a good overlap among all the NPs (A) and among all the AFs (B).

With the apparent expression profile proximity in the t-SNE plot, we performed a cell cluster analysis and used cell cluster size correlation as a metric to better understand our sample datasets. A size comparison of the cell clusters in the NP samples (Figure 2A) and the AF samples (Figure 2B) served as metric. The graphical representation in Figure 2 shows a high correlation between cell cluster sizes of the NP samples and the AF samples. The correlation coefficients for the NP samples are higher than 0.85 and for the AF samples higher than 0.87. A correlation matrix with hierarchical clustering (Figure S1) shows that the differences of the distal and proximal NPs are higher than the differences between the animals. Conversely, the differences of the AFs between the animals are larger than between the proximal and distal AFs. This may indicate possible temporal shifts in development between NPs and AFs.

Using unsupervised clustering, the NP cell populations (Figure 3A) exhibited 15 distinct cell clusters (Figure 3B). This analysis revealed that proximal and distal NPs identified mostly overlapping cell populations (Figure 3A) resulting from a high correlation of expression profiles and some shifts in the percentage fractions of some of the most abundant cell populations (Figure 3C).

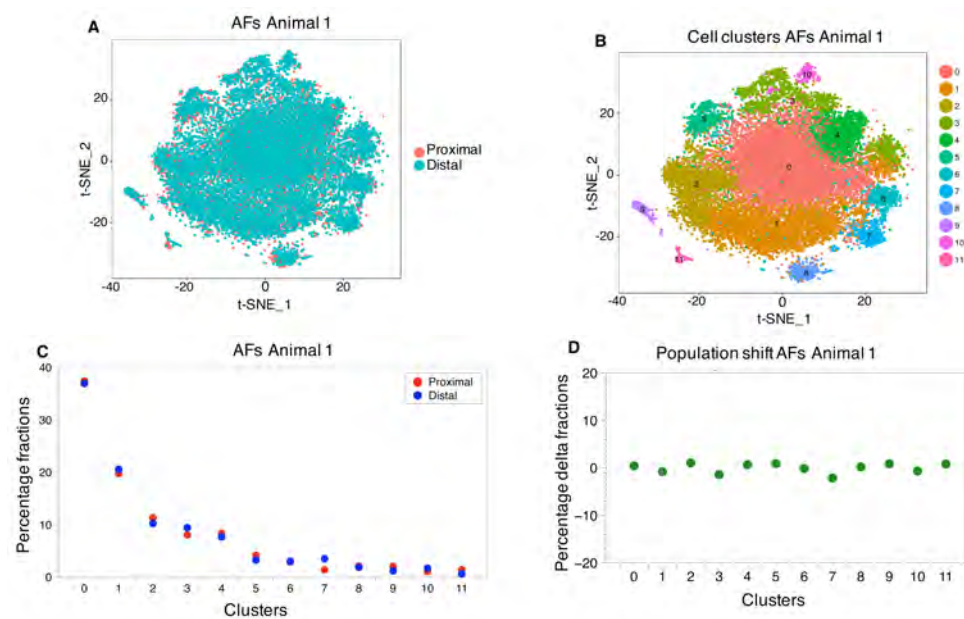


**Figure 2.** Cell cluster size correlation. The graphical alignment of cell cluster sizes within the NPs (A) and the AFs (B) shows a high correlation ( $r > 0.85$ ) within the datasets. The numbers placed around the piechart are cell cluster numbers, with the largest cluster starting at number zero. Each ring represents one sample and each color within the ring represents one cluster.



**Figure 3.** Cell cluster distributions between proximal and distal NPs. The first t-SNE plot (A) provides a visual measure for similarity for the proximal and distal NPs. The graphical overlap stands for the amount of similarity in gene expression profiles of the distal and proximal NPs. The second t-SNE plot (B) shows the identified cell clusters. (C) The population sized within a cell cluster. The first dot plot (C) is a representation of the percentage of the fractions of cells per cluster, normalized to the total amount of cells in each corresponding sample (see Table S1 for the original data). The population shift observed between the proximal and the distal samples, is described in terms of number of cells that compose a cluster in each sample. (D) The difference in percentage between the cell cluster sizes of the proximal and distal NPs, indicating the amount of correlation between the samples. (Animal 2, in Figure S2 and Table S3).

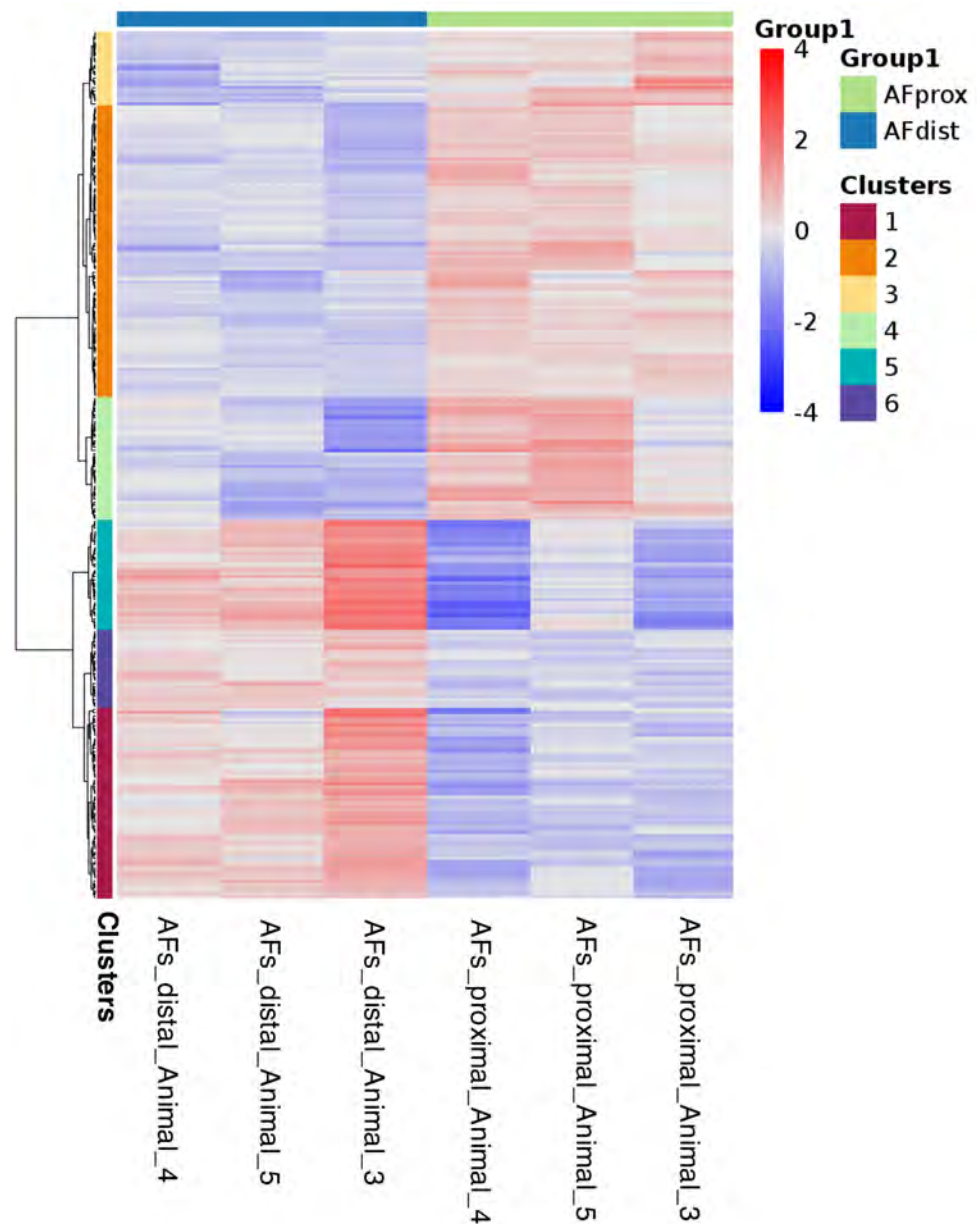
In regards to the AF samples, the overlap between distal and proximal cell cluster populations was even higher (Figure 4A), with 12 main cell clusters identified (Figure 4B) and no apparent shift of cell populations (Figure 4C).



**Figure 4.** Cell cluster distributions between proximal and distal AFs. The first t-SNE plot (A) provides a visual measure for similarity for the proximal and distal AFs. The graphical overlap stands for the amount of similarity in gene expression profiles of the distal and proximal AFs. The second t-SNE plot (B) shows the identified cell clusters. (C) The population sized within a cell cluster. The first dot plot (C) is a representation of the percentage of the fractions of cells per cluster, normalized to the total amount of cells in each corresponding sample (see Table S2 for the original data). The population shift observed between the proximal and the distal samples, is described in terms of number of cells that compose a cluster in each sample. (D) The difference in percentage between the cell cluster sizes of the proximal and distal NPs, possibly indicating developmental transitions in cells. (Animal 2, in Figure S3 and Table S4).

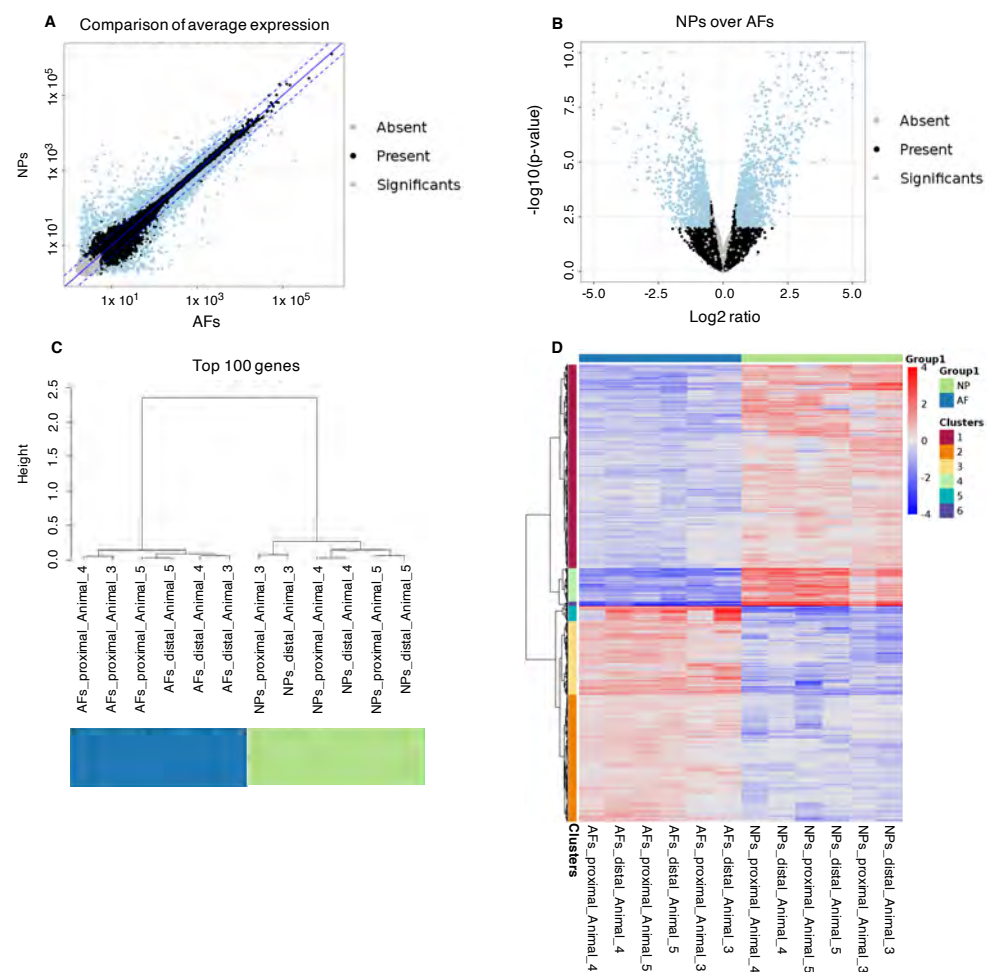
#### 2.4. Differential Gene Expression Analysis from Bulk RNAseq

To further investigate potential phenotypic differences across all samples, we performed differential expression analysis on bulk RNAseq data. This analysis revealed that within the groups, i.e., NPs and AFs, samples are highly correlated and clustered together (Figures S4 and S5) as with the scRNAseq data. Firstly, we investigated other differences along the rostrocaudal axis within the groups, i.e., differential gene expression analysis from the proximal and distal NPs as well as AFs. Hierarchical clustering was applied to the data and plotted as a heat map. There were no significant differences in the proximal and distal NP expression profiles (data not shown). However, the AFs provided distinct expression patterns for the proximal and the distal AFs after gene clustering (Figure 5). Gene Clusters 1, 5 and 6 are upregulated in the distal AFs when compared to the proximal AFs. However, the opposite seems true for the gene Clusters 2–4, which are instead downregulated in the distal AFs.



**Figure 5.** Heat map bulk AFs. Heat map for the differential genes expression for proximal and distal AFs (bulk RNAseq).

In a further analysis, we looked at the significant differential gene expression across all AFs and NPs. A hierarchical clustering was performed on the top 100 expressed genes of all samples and the data was normalized over all datasets. Figure 6A shows a good correlation in expression levels between the AF and NP groups. Figure 6B depicts the statistical significance of each gene against the fold change in expression level ( $\text{Log}_2\text{FoldChange}$ ), whereas Figure 6C shows the expressed genes that are up- or downregulated in either group and are statistically significant, which hints at biologically meaningful tissue profiles. Hierarchical clustering of the differentially expressed genes showed distinct gene clusters in both groups, AFs and NPs (Figure 6D).

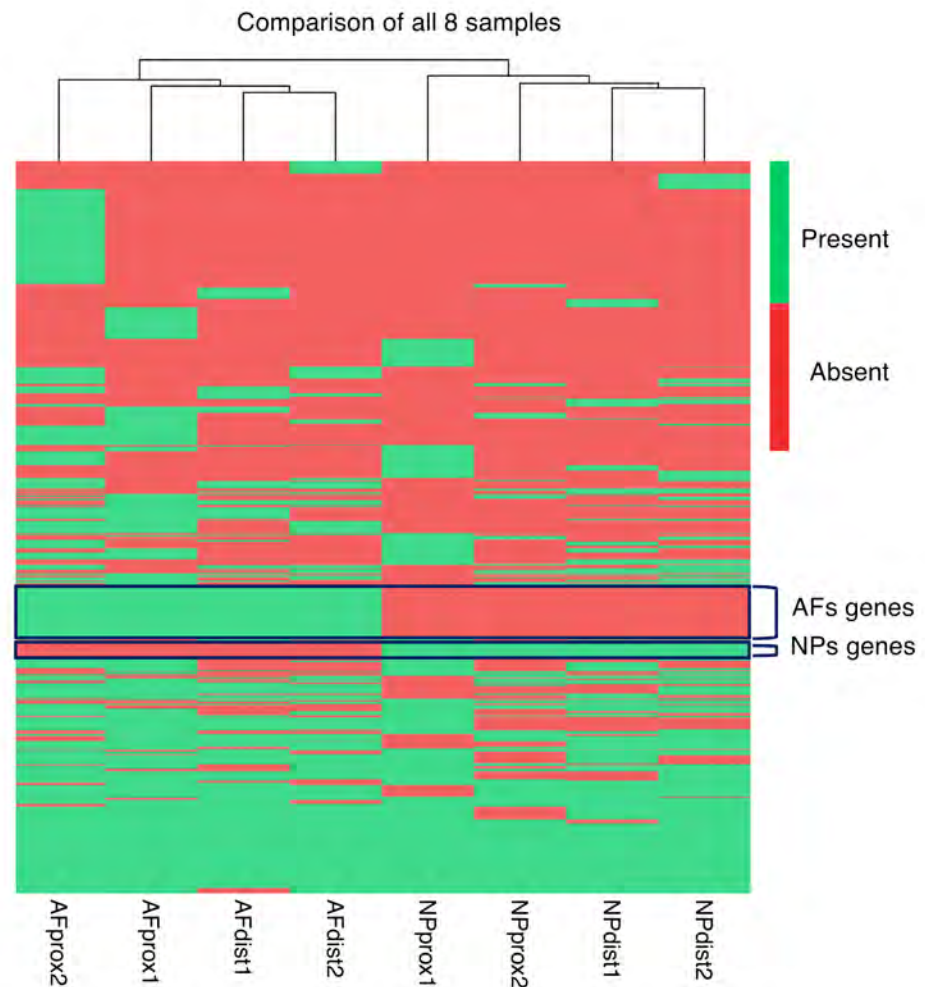


**Figure 6.** Comparison between NPs and AFs in the bulk RNAseq data. In the correlation graphic (A) is present a light correlation in the gene expression between all the (6) NPs and (6) AFs, with relevant differences. The volcano plot (B) shows a high number of significant differentially expressed genes with relevant fold changes between NPs and AFs. The dendrogram (C) compares the top 100 most expressed genes in the NPs and AFs. This shows a clear distinction between the gene expression of the NPs and the AFs, that clustered separately. Intra group, the NPs samples cluster per animal. In the AFs, instead, Animal 5 clustered together, while, for the Animals 3 and 4, the clustering occurs between the two proximal and the two distal samples. The heat map (D) shows a very clear distinction in the gene expression pattern of the NPs over the AFs. Here, the gene Clusters 1, 4 and 6 are upregulated in the NPs, while the gene Clusters 2, 3 and 5 are downregulated.

### 2.5. Integration of the Single-Cell RNAseq Data with the Bulk RNAseq Data

We went back to our scRNAseq data to identify genes that are exclusively expressed in either the NP or the AF samples. The heat map in Figure 7 depicts two sets of genes that are uniquely expressed in either all NPs or all AFs.

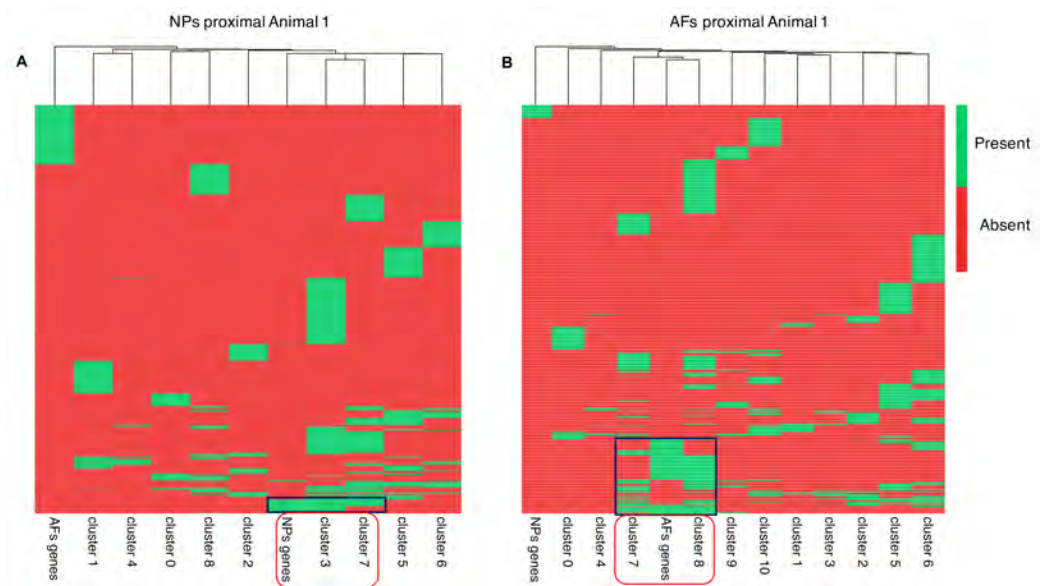




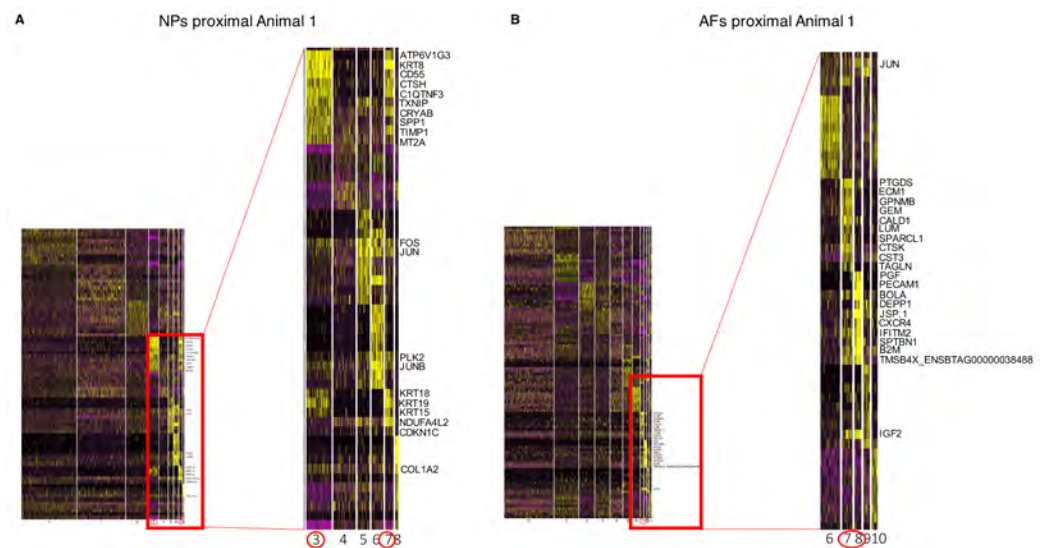
**Figure 7.** Heat map scRNAseq. Heat map depicting the presence and absence of expressed genes (top cluster markers) in the proximal and distal AFs and NPs of Animals 1 and 2, scRNAseq data. The gene clusters that are expressed exclusively in AFs and NPs are of particular interest as potential markers.

The two extracted gene clusters were subsequently mapped to cell clusters to identify the cell populations within the AF and NP that express those exclusive genes (Figure 8). In both sample sets, the AFs and NPs, we identified two cell populations that expressed the previously identified unique expressed genes. In the NPs, cell population/Clusters 3 and 7 showed the highest overlap with the uniquely identified NP genes (Figure 8A), whereas cell population/Clusters 7 and 8 had the highest overlap in the AF unique genes (Figure 8B). Subsequently, we investigated the expression levels of the unique set of AF and NP genes in an expression heat map (Figure 9).

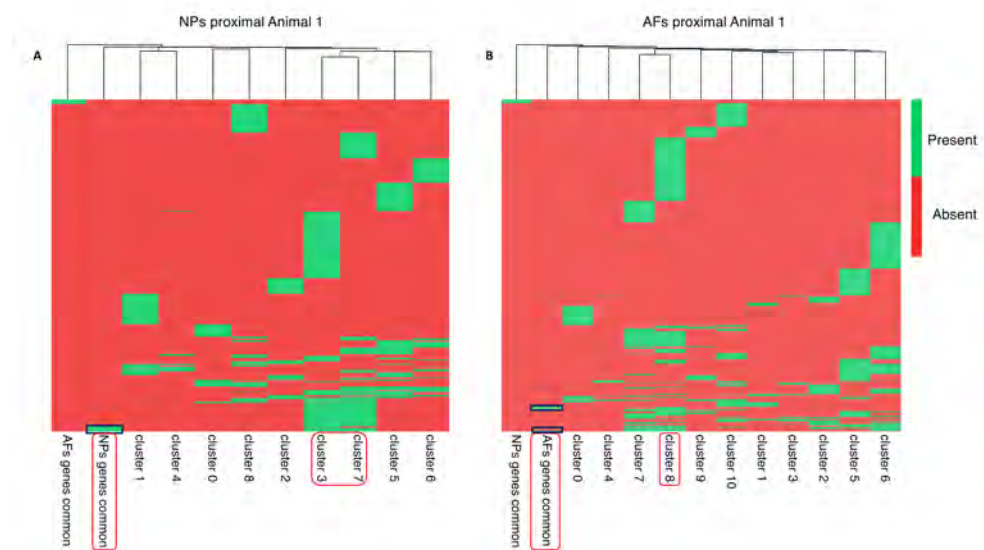
We then proceeded to identify those same genes in the bulk RNAseq data. We intersected the list of unique specific genes from the scRNAseq data with the bulk transcriptome for AFs and NPs and aligned them to the scRNA cell clusters/populations for NPs and AFs. We performed hierarchical clustering and visualized the results in a heat map (Figure 10). In the NPs, the overlap in structure specific (NP) genes for bulk and single-cells data show the previously identified NP cell Clusters 3 and 7 (Figure 10A). The AF specific genes were added as a control set and did not show any overlap. In the AFs, the bulk and single-cell data overlapped in cell Cluster 8 (Figure 10B). The NP specific genes were added as a control. The dendrogram structure at the top of the heat maps provides a measure of commonly expressed genes among the clusters indicating a potential lineage relationship between clusters. All 27 and 24 structure specific genes identified in the NP and AF respectively were found in the scRNAseq and bulkRNAseq data (Table S5).



**Figure 8.** Heat map scRNAseq cluster. Heat map of exclusive AF and NP genes in the scRNAseq data, expressed in the NPs proximal and and AFs proximal of Animal 1. (A) The first heat map depicts the cell clusters that express NP genes. (B) The second heat map depicts cell clusters that express AF genes. The dendrogram at the top is a function of the amount of commonly expressed genes.



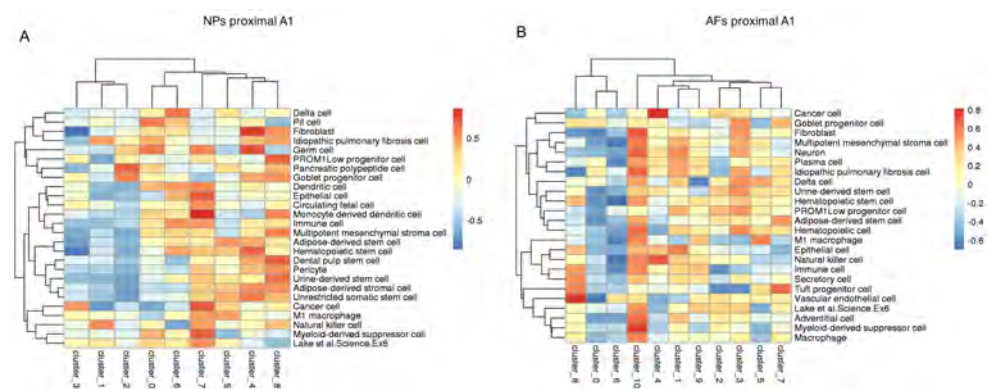
**Figure 9.** Genes that are highly expressed in scRNAseq clusters. Genes that are highly expressed in the clusters of interest in the scRNAseq data. (A) From the NPs proximal Animal 1, the Clusters 3 and 7 previously identified were selected. (B) From the AFs proximal Animal 1, the Clusters 7 and 8 previously identified were selected. The focus was pointed on the genes that are highly expressed in the clusters of interest. Higher expression value are highlighted in yellow.



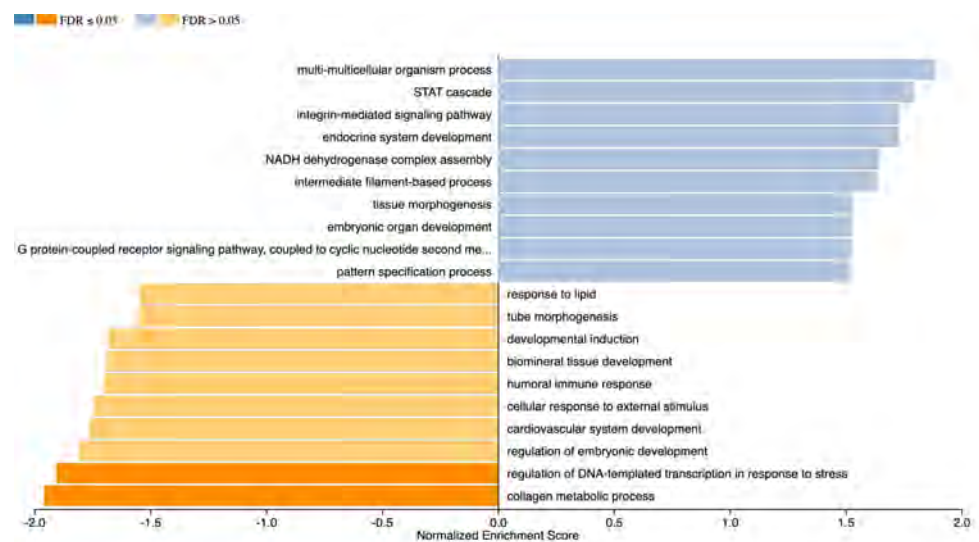
**Figure 10.** Heat map exclusive genes bulk and scRNAseq. Heat map of exclusive AF and NP genes common between the scRNAseq and bulk RNAseq. These genes are mapped against a single sample. (A) The first heat map depicts the cell clusters that express the NP common genes in the sample NPs proximal Animal 1. (B) The second one depicts cell clusters that express AF common genes. The dendrogram at the top is a function of the amount of commonly expressed genes.

### 2.6. Gene Enrichment Analysis

Once we identified cell clusters that express specific tissue genes, we aimed at identifying possible cell types for these clusters. The enrichment analysis on a single-cell level allows for the investigation on the potential characteristic of the cell populations in the NPs and AFs tissue. For the gene enrichment analyses, the genes were translated into their human homologous. The subsequent enrichment analysis of each sample hinted at potential cell types for each identified cell cluster (Figure 11). Such hints are especially interesting for the cell clusters that express the structure specific genes, namely Clusters 3 and 7 for the NPs (Figure 11A) and Clusters 7 and 8 for the AFs (Figure 11B). The results for the NP Cluster 3 show the characteristics closest to a cancer cell, yet with otherwise very little common features to the other cell types. These data suggest a cell type that was not present in the repository. Cell Cluster 7 of the NP, however, has the highest similarity to a monocyte derived dendritic cell, followed by cancer cell characteristics. While most suggested cell types are not sensible, the fibroblast characteristics for Clusters 4 and 8 of the NP are plausible. AF Cluster 8 appears to have the highest similarity to vascular endothelial cells. Similar to the NP suggestions, the AF suggested cell types are mostly not convincing. The cell Cluster 10 appears to possess fibroblast-like features, yet its predicted top feature is Myeloid-derived suppressor cell. The consistent presence of stem cell-like cells in the suggestions might hint at the existence of stem and progenitor cells in the clusters. As a last step, we performed an enrichment analysis. The analysis was performed on the single-cell and bulk RNAseq samples. The enrichment allows for the identification of detectable pathways in our samples. We compared all NPs and all AFs using GSEA GO terms Biological process (Figure 12). There is a significant enrichment in the NPs for the collagen metabolic process and regulation of DNA transcription in response to stress.



**Figure 11.** Heat maps enrichment. Heat maps of the enrichment analysis of all the eight scRNAseq samples. The detected genes for each cell cluster have been translated into the human homologous and run through an enrichment analysis. The pathways that could be characteristic of each cluster are highlighted in red, defining a potential cell type profile. (A) The first heatmap depicts potential cell type profiles for the NP cell clusters. (B) The second heatmap depicts potential cell type profile for the AF cell clusters.



**Figure 12.** Enrichment analysis bulk RNAseq Enrichment analysis of all the NPs over the AFs bulk RNAseq data. GSEA, GO terms Biological Process. The bar chart shows a significant enrichment of the collagen metabolic process and DNA response to stress in the NPs over the AFs.

### 3. Discussion

Single-cell RNAseq enabled the profiling of 92,277 intervertebral individual cell transcriptomes. Our study identified 15 distinct NP cell populations and 12 distinct AF cell populations of the bovine IVDs which supports previous findings on cell population composition of NPs and AFs [34,36,41,43,51–56]. Unsupervised cell clustering methods allow the detection of cell types and differentiation states based on the whole transcriptome in a fine-grained manner. These methods have readily and successfully been used [57–61], even if minor imprecisions may occur at defining cross-species identities of specific cell types and functional differences among identified cell types [62,63]. In our single-cell RNAseq analysis, we used the shared nearest neighbor approach for unsupervised cell clustering to identify distinct cell types in the four AFs and four NPs of the bovine IVDs from two animals [64]. Through hierarchical clustering, we identified structure specific genes in the NPs the AFs. These genes were subsequently matched to the earlier identified cell populations. Interestingly, only two distinct NP cell populations and two distinct AF cell populations matched. A literature research on the top expressed genes of each cell

population identified two distinct notochordal cell-like populations in the NP and two distinct AFC populations, with one putative fibroblast-like population.

The heterogeneity of the cells contributing to either the NP or AF has been extensively described in the literature. Diverse techniques such as immunohistochemistry, qPCR and Western blot, among others, have been used to define phenotypic markers for these two populations, all of which have their obvious disadvantages, especially in regards to sensitivity issues. The aim of this study was to elucidate at a single-cell resolution the individual NPC and AFC populations along the bovine rostrocaudal model. Coupled with highly sensitive bulk RNAseq analysis of these two tissues, we could identify specific NP and AF genes as well as cell populations, which gives unique insight into the biology of this organ. NPC have been reported to have differentiated along the notochordal lineage, clearly distinguishing them from other cell types in the body [11–13]. As a matter of fact, notochordal or notochordal-like cells have been reported in both fetal notochordal immature NP and adult NP. The exact role of these cells remains unclear; however, it is believed that they either act as a regulator or as a progenitor cell for the NP. A broad panel of markers associated with the phenotype of NP notochordal cells [36,38–40] was detected mainly concentrated in cell Clusters 3 and 7 from the single-cell NP samples (Figure S8A); these include *CD24*, galectin-3 (*LGALS3*), cytokeratin 8 (*KRT8*), *KRT18*, *KRT19* and brachyury. Brachyury is a protein that is encoded by the *TBXT* (T-box transcription factor T) and has been described to be necessary for notochordal morphogenesis and patterning. *TBXT* was detected in cell Cluster 7 only, and even some of its binding partners were detected in cell Clusters 3 and 7, such as *SMARCA4*, *STAT3* and *WNT5B*. We additionally found other notochord associated genes expressed, such as *CDH2* and *FN1*. Galectin-3, however, was predominantly expressed in mainly cell Clusters 8 and 0. Moreover, these cell clusters also expressed genes that have been associated with pluripotent stem cells or progenitor cells such as *KRT15* [65], *CD44* and *CD55*. These two cell clusters are in the same trajectory in the slingshot (Figure S8C) which is a cluster-based lineage reconstruction and pseudo-time inference approach indicating their relatedness. These first set of findings, provide a unique insight in the rostrocaudal bovine model for IVDs as we could detect specific cell populations that have a notochord-like phenotype. It is believed that these cell populations decrease as adulthood arrives, as supported by the findings that *CD24* gene expression is reduced in the adult NP [36]. A target of *TBXT* that may be relevant to NPC is the connective tissue growth factor (CTGF/*CCN2*), which is an ECM protein that is necessary for post-natal NP function [66]. *CTGF* was detected in cell Clusters 1, 2 and 5.

There was an additional cell cluster (Cluster 8) that did not express the notochordal associated genes but instead stem cell markers such as *CD63*, *CD44* and galectin-3. This cell cluster additionally expressed *WISP2* (*CCN5*), which is a growth factor member of the *CTGF/CCN* family of secreted, ECM-associated proteins [67] that has been shown to be an inhibitor of angiogenesis [68] and predicted to be an active regulator of notochordal gene expression. *WISP2* was identified in non-chondrodystrophic dog notochordal cell model (together with *CTGF*) and proposed to be key for healthy NP homeostasis. Additionally, *NOV*(*CCN3*) was also found in cell Cluster 8. This secreted extracellular protein is key in mesenchymal stem cell differentiation and regulates differentiation, growth and maturation of osteogenic, chondrogenic and hematopoietic progenitor cells. Wnt signaling has been extensively described in the maintenance of notochord progenitor cells identity [69]. The expression of *CCN2* and *CCN5* are upregulated in the early stages of mesenchymal stem cell differentiation by Wnt3A stimulation [70,71]. These results are not surprising since the Wnt/beta catenin signaling pathway has been shown to be required for the development, organization, senescence and degeneration of the NP [72,73].

Thus, cell Cluster 8 indeed indicates a progenitor-stem cell phenotype, as shown in Figure 11. One of the proposed roles for this cell cluster, which needs further interrogation, is their potential role in regulation of notochordal cells. It has been shown that sustained Wnt signaling is required to maintain the notochordal progenitor cells' notochordal fate [69].

Additionally, cell Clusters 3, 7 and 8 expressed ECM regulators such as *BMP2* and secreted factors (Figure S6), indicating their involvement in NP homeostasis.

Cell Clusters 3 and 7 furthermore expressed *HIGD2A* (HIG1 Hypoxia Inducible Domain Family Member 2A), which is a hypoxia marker, and its expression is induced by the hypoxia-inducible factor-1 (*HIF1*). Cell Cluster 7 additionally had an upregulation of *XBP1* (X-box binding protein), a transcription factor and signal transducer involved in HIF1-alpha mediated hypoxia responses in survival [74], thus further supporting that these cell populations play in potentially detecting the environmental condition, ECM regulators and secreting soluble factors to maintain tissue homeostasis as indicated by the matrisome enrichment (Figure S6) [75–77].

Since the NP is the largest avascular tissue in the body and NPC have developed to thrive in a unique hypoxic and hyperosmotic environment, we looked further at other hypoxic factors other than *HIGD2A* which also contribute to top cell marker clusters. Cell Cluster 5 of the NP had a higher expression of *HIF1A*, which could potentially indicate that these cells may act as environmental/hypoxic regulators. This is of particular importance since *HIF1A* encodes for the alpha subunit of a transcription factor Hypoxia-inducible factor (HIF-1). HIF-1 functions as a master regulator of cellular and systemic homeostatic response to hypoxia by activating transcription of many genes. Our data show an upregulation in HIF-1A-downstream genes reported to be involved in autophagy (*BNIP3* and *BNIP3L*), glycolysis (*LDHA* and *PGK1*), cell migration and adhesion (*CXCR4*), angiogenesis (*VEGFA*) and lipid storage and endoplasmic reticulum homeostasis (*PLIN2*). Furthermore we identified *C1QTNF3* and *TXNIP* in cell Clusters 3 and 7 which are in the same trajectory of the slingshot (Figure S8D). *C1QTNF3* has been described as downstream factor of HIF-1alpha which plays an essential role in cartilage development [78]. *TXNIP* as part of the *TXNIP-NLRP3* inflammasome has been related to IVD degeneration [79], yet it is known for its pleiotropic effects as it plays an important role in the energy metabolism [80] and in cell cycle regulation [81]. Semaphorin 4A (*SEMA4A*) was also expressed in cell Cluster 5. Semaphorins are transmembrane proteins that have been shown to not only lead to repulsive axon guidance but also to play key roles in tissue homeostasis and morphogenesis. These results potentially indicate that cell Cluster 5 is one of the NPC populations involved in NP homeostasis under hypoxic conditions and avoiding nerve ingrowth. It is also a contributor to the ECM by expressing regulators, glycoproteins and proteoglycans (Figure S6). This same Cluster 5 expresses T-cell factor 3 (*TCF3*), a member of the Wnt signaling pathway that has been shown to be a modulator in embryonic stem cells and development [82]. In regards to the ECM, in the NP, we observed that collagens were primarily expressed in cell Clusters 1, 4 and 8 (Figure S6), which accounted for almost a third of all the cells analyzed. Additionally, returning to the point about the effects that hypoxia may have in maintaining tissue homeostasis. HIF1-alpha has been shown to control collagen synthesis in chondrocytes by inducing collagen-modifying enzymes such as lysyl oxidase (*LOX*) and collagen prolyl hydroxylase (*P4H42*), these genes were expressed in cell Cluster 8 and 4, respectively [83–85]. These are key ECM-modifiers that increase crosslinking and stability of the collagen triple helices. Taken together, our data indicate a series of very distinct cell populations that are closely orchestrated living in this unique environment. Our data, whether in the single-cell RNAseq or the bulk RNAseq, show that we could identify most NP specific markers and genes previously reported to be expressed at higher levels in the NP than in the AF [86,87].

From the eight differentially expressed genes in the NP reported in an earlier single-cell experiment from Fernandes et al. 2020, we could detect six genes in our single-cell RNAseq data, namely *COL2A1*, *KRT8*, *CD24*, *COMP*, *FMOD* and *ACAN*, but not *A2M* and *DSC3*. Discrepancies between their single-cell experiments and ours are most likely due to single-cell RNAseq experimental/technique, number of cells analyzed and sequencing differences, especially in regards to sequencing depth. Furthermore, our bulk RNA experiments could only support the overexpression of *KRT8* and *CD24* in the NP, indicating the potential limitations in detecting changes in gene expression in very heterogeneous cell populations.

Novel NP markers such as *CHRD* [88], *CHRD2*, *SCUBE1* [89] and *CLEC3B* were both detected and differentially expressed in both our NP datasets [21]. Some of the novel detected NPC contributing genes are part of important signalling pathways and act as inhibitors of WNT (*FRZB* and *DKK3*) and BMP signaling (*CHRD*, *CHRD2*).

The AF provides a wrapping with strong tensile properties and contains the NP. This leads to a unique architecture of the AF as it is composed of a ring of ligament fibers with a tendon/ligament-like structure [90]. Thus, the cells in the AF reported very specific phenotypic signatures that can enable such a defined biomechanical AF function. In contrast to the NP, where only a couple of cell clusters expressing some collagens, most cell clusters in the AF contributed by expressing of a wide variety of collagens (Figure S7). Among those highly expressed collagens, we found *COL1A1*, *COL1A2*, *COL9A3*, *COL18A1* and *COL4A1*, which have been widely reported [20,21].

Small leucine rich proteoglycans (SLRPs) such as *BGN*, *DCN*, *FMOD*, *OGN* and *PRELP* were highly expressed in the AF and predominantly found to be expressed in Clusters 1 and 6 (Figure 8B). SLRPs are associated with collagen fibrillogenesis and ECM assembly especially in the AF to form concentric lamellae of oriented collagen [22–24].

Another major component of the ECM are the basement membrane proteins and a major noncollagenous constituent of this family are laminins which have been described in both chondrocytes and the IVD. *LAMA3/4/5*, *LAMB1/2/3* and *LAMC1* were top cluster markers for a wide range of cell clusters in the AF.

From the ECM remodeling side, we detected *SPARC* (*SPARC*, also called osteonectin or basement membrane protein 40), which encodes a cysteine-rich acidic matrix-associated protein. This matricellular protein plays a dynamic role in regulating extracellular procollagen processing in the pericellular environment so that it can be efficiently incorporated into the ECM [91–93]. *SPARC* has been proposed to modulate cell-ECM interactions and remodeling of the ECM since it binds collagen Types I–V and VIII, vitronectin and thrombospondin-1 [94–96]. Increased methylation of *SPARC* and its promoter as well as decrease in levels of *SPARC* have been reported in the human IVD with aging and in patients suffering with low back pain [97,98]. In the AF, *SPARC* was highly differentially expressed in the AF over the NP and ubiquitously found in the scRNAseq data for the AF, indicating its pivotal role for a healthy AF environment.

Other matricellular genes involved in collagen fibrillogenesis were detected in our AF data and include *PCOLCE*, *PCOLCE2* and *MXRA5* [21]. Due to load bearing properties of the AF and similarities in mechanical functionalities, it was not surprising to find that AFC expressed tendon-related markers such as Tenascin C, tenomodulin and thrombospondins (*THBS1/3/4*) as well as cartilage intermediate layer proteins (*CILP2* and *CILP*), some of which have reported in previous studies [21,99]. Additionally, a wide variety of heat shock proteins (*HSPA5*, *HSPA6*, *HSPA8*, *HSP90AA1*, *HSP90AB1*, *HSP90B1*, *HSPB1*, *HSPH1* and *HSPA1A*) were mostly detected in Clusters 4 and 5 in the AF, suggesting their role in the stress response to load bearing.

Growth factors such as *CTGF*, *PDGFB*, *VEGFC*, *IGF2* and *FGF2/7* were also detected in the AF mostly concentrated in cell Cluster 8. According to the potential cell type profile (Figure 11), the cells in Cluster 8 have a vascular endothelial and progenitor cell phenotype. We found though that it also expresses genes associated with class II major histocompatibility complex (MHC) such as *CD74* and *BoLA-DR-alpha*. This same cluster highly expresses *CXCR4* that is involved in the recruitment of mesenchymal stem cells in bone fractures and *IGF2* that is related to fibroblast proliferation [100,101].

We detected *TIE1* and *TIE2/TEK* in the AF specifically in cell Cluster 8. Intriguingly, *TIE1* and *TEK* receptor kinases have been reported to be specifically expressed in endothelial cells during embryonic angiogenesis [102]. Through heterodimerisation, they mediate interactions not only with the ECM but also the surrounding mesenchymal cells, thus suggesting a paracrine regulation [103]. VEGF has also been shown to activate *TEK* through *TIE1* [104]. These findings are significant as up to now *TEK/TIE2+* cells have

been uniquely described in progenitor NPC across many species, including bovine and human [26].

Surprisingly, cell Cluster 9 in the AF seems not to be a major contributor to the matrix (Figure S7), although it highly expressed *CD44*, *CD63* and galectin-3, indicating a stem cell-like phenotype which may play key role in tissue homeostasis. Furthermore, *CALD1*, *LUM* and *CTSK*, all of which have been described in fibroblasts of the dermis [105], were highly expressed in cell Clusters 9 and 10, as was *PTGDS*, which has been recently described as a papillary fibroblast marker in the human dermis [106]. The potential cell type profile of Cluster 10 is fibroblast-like cells. *TAGLN* (*SM22*) has similarly been described in fibroblasts and dedifferentiated cultured chondrocytes [107,108]. Overall, our data match spatial proteomics findings from Tam V. 2020 in detecting structural specific genes both in the NP and AF, indicating the relevance and integrity of our data [21]. Discrepancies, however, might be due to the disc composition and remodeling of a sample, especially because with RNAseq data we only capture the cellular information and these only represent a small fraction of the disc volume [109].

Taken together, our results show a complex and heterogeneous cell population in both the NP and AF.

Calf tails have been extensively used in intervertebral disc studies as a ready source of these cell populations [41,110–112]. The use of this model in our study, however, does entail some limitations specially in regards to the age of the calves as there is a variable window of 6 months. Given the rostrocaudal developmental gradient in calves, an age shift can influence the cellular composition of the IVDs at a given level, thus introducing variance in the differential gene expression analysis of the proximal and distal IVDs [44]. However, we may have compensated for the age-related variance due to the pooling of three adjacent IVDs to obtain enough cells for our experiments. Our scRNAseq data suggest, especially for the AF, a significant developmental gradient along the rostrocaudal axis. Indeed, we showed size fluctuations of 10–20% exist in the larger cell cluster fractions of the NPs and AFs. These fluctuations are expected mainly due to age differences when comparing AFCs and NPCs among different animals. Furthermore, the amounts of fluctuation among animals is larger for the NPCs and along the rostrocaudal axis for the AFs. Given that these cell populations may consist of stem cells, transient amplifying progenitors and terminally differentiated cells, the pattern of these fluctuations may indicate possible temporal differences in the development of the two compartments. This would be consistent with the transcripts expressed by certain populations of these intervertebral disc cells. This notion is further supported by the differential gene expression analysis performed on the so-called bulk RNAseq data from three additional calves. While the NP samples showed no distinct patterns across animals along the rostrocaudal axis, the AF samples showed significant differences along the rostrocaudal axis but not across animals. These differences consist of three differentially expressed transcript clusters being upregulated in the proximal AF in all animals and the other three clusters being upregulated in the distal compartment. Furthermore, hierarchical clustering in the bulk RNAseq showed, as expected, major transcriptome profile differences between the AF and NP, as previously published [20]. Overall, our data suggest that the AF exhibits rostrocaudal developments whereas the NP seems to be rather stable.

Another potential drawback from the IVD cell isolation procedure used in our study is the lack of hypoxic conditions during the extracellular matrix digestion to obtain the single-cell suspension required for the scRNAseq experiments. We are aware that this could have influenced the differentiation state of the cells and as a result altered the cell cluster sizes and numbers [113]. However, IVD cells were immediately processed after digestion to ensure a minimal impact on the transcriptome of the IVD cells themselves as well as the maintenance of the original cell numbers. Multiple studies have shown that culturing IVD cells in monolayers after digestion affect IVD cell phenotype and differentiation [113–116].



## 4. Materials and Methods

### 4.1. Samples Collection and Processing

Two bovine tails were used for the scRNAseq, and three tails for the bulk RNAseq. The animals were aged from 6 to 12 months. The cow tails were collected from local abattoir within 2 h post mortem, during which the samples were kept at 4 degrees. After disinfection of the tails with Betadine (Mundipharma Medical Company, Basel, Switzerland), the IVDs were dissected and small portions of the Nuclei Pulposi and the Annuli Fibrosi were extracted. For the scRNAseq, the three most proximal NPs were pooled, as well as the three most distal ones, in order to have a minimum of 30,000 cells per sample (Table 1). The same procedure was applied to isolate the most proximal and distal AFs. As it was previously shown that there are gene expression differences noted depending on the segmental level of the coccygeal discs [117], we decided to pool the most distal and the most proximal IVD tissues. For the bulk RNAseq, only the two most proximal and distal NPs and AFs were used (Table 2). The samples were then cut in small pieces and incubated in flasks with 0.19% Pronase for 1 h at 37° on shaker at lowest speed. After the incubations, the cells were then washed with PBS before placing them again in flasks for an overnight incubation with Collagenase II (260 IU/mg) at 37° on shaker at lowest speed. The next morning, the cells were washed with PBS and then filtered through a 100 µm filter to remove potential undigested tissue aggregates and debris. For the scRNAseq samples, the cells were then counted and diluted in PBS + BSA (0.05%) with an approximate concentration of 1000 cells/µL. For the bulk RNAseq samples, the cells were collected in a pellet and shock frozen.

### 4.2. Bulk RNAseq

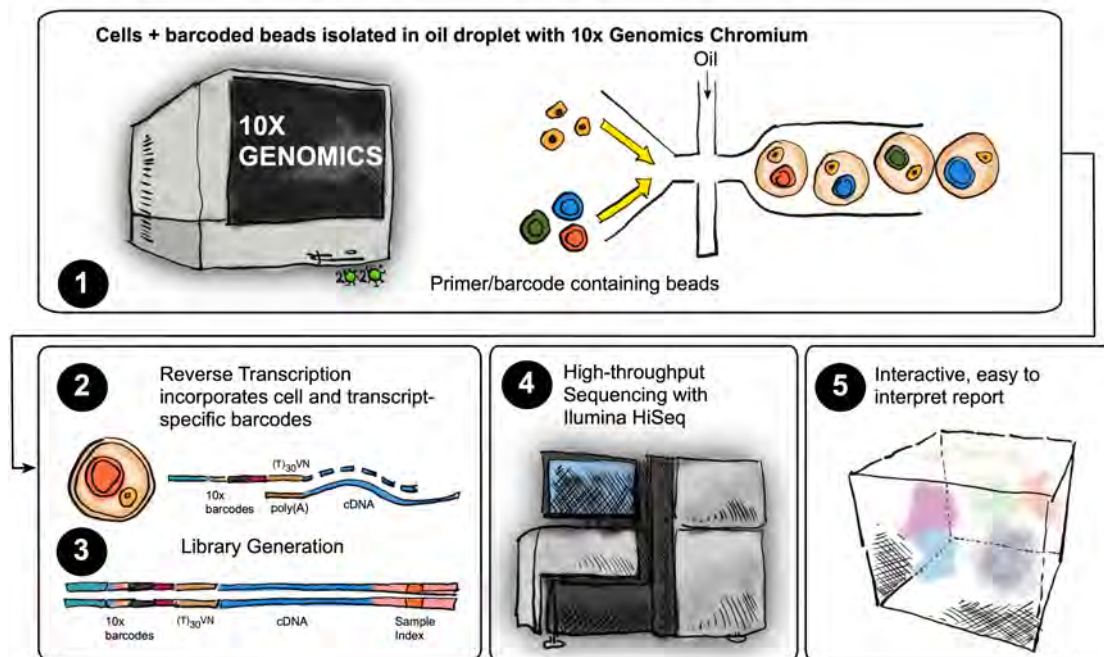
The RNA was extracted following the RNAeasy Mini Kit Qiagen manual. The quality of the RNA was evaluated using TapeStation (Agilent, Waldbronn, Germany). The libraries were prepared using SENSE mRNAseq Library Prep Kit (Lexogen, Vienna, Austria) and sequenced on the NextSeq500 System (Illumina, San Diego, CA, USA) using a 150 cycle kit, single-end 150 bp, following the manufacturer's instructions. In total, 25M reads were obtained per sample.

### 4.3. Bulk RNAseq Data Analysis

The quality of Illumina SE RNAseq reads was evaluated using FastQC v.0.11.7 [118]. FastqScreen v.0.11.1 [119], was used to screen for potential sample contaminations (genomic DNA, rRNA, mycoplasma, etc.) using a custom database including UniVec [120], refseq mRNA sequences, selected genome sequences (human, mouse, bacteria, virus, PHIX, lambda and mycoplasma) [121] and SILVA rRNA sequences [122]. Illumina SE reads were preprocessed using Trimmomatic v.0.36 to trim off sequencing adapters and low-quality ends [123]. Trimmed reads were aligned to the reference genome and transcriptome (FASTA and GTF files, respectively, Ensembl, *Bos taurus*, UMD\_v3.1, Annotation release 92-2018-0530) with STAR [124] version 2.7.3a with default settings for single end reads. PCR duplicates were marked using Picard v.2.18.0. Differentially expressed genes were identified using R/Bioconductor package edgeR (R version: 3.6.1, edgeR version: 3.28.1) [125], in which the normalization factor was calculated by trimmed mean of M values (TMM) method [126].

### 4.4. Single-Cell RNAseq

The quality and concentration of the single-cell preparations were evaluated using a hemocytometer in a Leica DM IL LED microscope and adjusted to 1000 cells/µL. 10,000 cells per sample were loaded in to the 10× Chromium controller (Pleasanton, CA, US) and library preparation was performed according to the manufacturer's indications (single-cell 3' v3 protocol). The resulting libraries were sequenced in an Illumina NovaSeq sequencer according to 10× Genomics recommendations (paired-end reads, R1 = 28, i7 = 8, R2 = 91) to a depth of around 50,000 reads per cell (Figure 13).



**Figure 13.** 10× single-cell sequencing workflow. Single-cell RNA-seq workflow. (1) Isolated single NP or AF cells together with barcoded beads and reverse transcription reagents were isolated in oil droplets using the 10× Genomics Chromium system (2) Within each micro reactor, the single cell were lysed, and reverse transcription of polyadenylated mRNA occurs with the incorporation of cell-specific and transcript specific barcodes or UMIs (Unique Molecular Identifier). (3) As a result, all cDNAs from a single cell had the same barcode, which enabled the sequencing reads to be clustered together. Next generation sequencing (NGS) libraries from these barcoded cDNAs were prepared. (4) NGS was performed in a Novaseq 6000 system. (5) Cell Ranger (v3.0.1) was used to align the reads to the reference genome and generate count tables of UMIs for each gene per cell. Subsequent data analysis steps were performed as described in Section 4.5.

#### 4.5. Single-Cell RNAseq Analysis

The fastq files were aligned to the *Bos taurus* reference sequence (UMD-v3.1 Release-92) taken from Ensembl. After the alignment, each observed barcode, UMI, gene combination was recorded as a UMI count matrix that was then filtered to remove low RNA content cells or empty droplets. All these steps were performed using the CellRanger software (v3.0.1). Starting from this matrix, we used the R package Seurat (version 2.3.4) [127] to perform the following downstream analyses per sample: genes and cells filtering, normalization, feature selection, scaling, dimensionality reduction, clustering and differential expression. We started by filtering out genes that did not obtain at least 1 UMI count in fewer than five cells, and discarded cells for which fewer than 400 genes or more than 4200 genes were detected and also those that had a mitochondrial genome transcript ratio greater than 0.25. After this, the data were normalized using a global-scaling normalization method that normalizes the feature expression measurements for each cell by the total expression, multiplies this by a scale factor (10,000 by default) and log-transforms the result. We next calculated a subset of 2000 features that exhibited high cell-to-cell variation in the dataset. Using as input these variable features, we performed PCA on the scaled data. Since Seurat clusters cells based on their PCA scores, we used a heuristic method called ‘Elbow plot’ to determine how many principal components (PCs) we needed to capture the majority of the signal. In this way, the cells were clustered using an unsupervised graph-based clustering approach using the first 12 PCs and a resolution value that ranged from 0.3 to 0.5. Clusters were visualized using t-distributed Stochastic Neighbor Embedding of the principal components (spectral t-SNE) [128] as implemented in Seurat. We found positive markers that defined clusters compared to all other cells via differential expression. The test we used was the Wilcoxon Rank Sum test, which assesses separation between the

expression distributions of different clusters. Genes with, on average, at least 0.25-fold difference (log-scale) between the cells in the tested cluster and the rest of the cells and an adjusted  $p$ -value  $< 0.05$  were declared as significant. Cell-cycle phases were predicted using a function included in the *scrn* R package [129] that scores each cell based on expression of canonical marker genes for S and G2/M phases. Since *scrn* only includes markers for human and mouse, we performed this analysis using the human homolog genes taken from Ensembl.

#### 4.6. Pathways Analyses

To identify which kind of cell populations were present in the samples, we performed three types of enrichment analyses: singular enrichment analysis (SEA), gene set enrichment analysis (GSEA) and gene set variant analysis (GSVA). GSEA was also performed for the bulk RNAseq data to detect pathways enriched in either the NP or the AF structures. The most traditional enrichment approach, SEA, iteratively tests annotation terms one at a time against a list of interesting genes for enrichment. In this case, we used Gene Ontology 1, KEGG 2 and Wiki pathways 3 as annotation terms and we used the marker genes found in every cluster as our lists of interesting genes. SEA calculates an enrichment  $p$ -value by comparing the observed frequency of an annotation term with the frequency expected by chance; individual terms beyond some cut-off (e.g.,  $p$ -value  $\leq 0.05$ ) were deemed enriched. GSEA method is similar to SEA, but it considers all genes during analysis, not just those deemed as interesting or significant by some metric or threshold. Alternatively, we performed a biological theme comparison among gene clusters with statistical analysis of GO and KEGG pathways. Finally, we used the enrichment method GSVA to find associations between clusters and cell types. GSVA calculates sample-wise gene set enrichment scores as a function of genes inside and outside the gene set, analogously to a competitive gene set test. GSVA is similar to GSEA, but it estimates variation of gene set enrichment over the samples independently of any class label. This time, instead of using GO terms or KEGG pathways as gene sets, we used cell types and gene markers identified in different tissues in human taken from the CellMarker database 4. *Bos taurus* gene symbols were converted to their homolog gene symbols in *Homo sapiens*.

1. <http://geneontology.org> (accessed on 29 April 2019)
2. <https://www.genome.jp/kegg> (accessed on 29 April 2019)
3. <https://www.wikipathways.org> (accessed on 29 April 2019)
4. <http://biocc.hrbmu.edu.cn/CellMarker> (accessed on 29 April 2019)

#### 4.7. Multiple Samples Integration

To find inter- and intra-individual differences between animals, we used a method implemented in Seurat v 3.0.1 [64] that aims to identify shared cell states that are present across different datasets. We initiated this process through dimensional reduction using canonical correlation analysis (CCA), aiming to place datasets in a shared low-dimensional space. Following dimensional reduction, we identified the K-nearest neighbors (KNNs) for each cell within its paired dataset, based on the L2-normalized CCV. Finally, we identified mutual nearest neighbors (MNN; pairs of cells, with one from each dataset, that are contained within each other's neighborhoods). We refer to these pairwise correspondences as "anchors". Filtering and scoring anchors steps were performed to mitigate the effects of any incorrectly identified pairs of cells.

## 5. Conclusions

By integrating the scRNAseq data with the bulk RNAseq data to stabilize the clustering results of our study [130,131], we identified 27 NPC structure/tissue specific genes and 24 AFC structure/tissue specific genes. Developmental changes were detected in the AF along the rostrocaudal axis. From our single-cell RNAseq results, we could deconvolute the heterogeneous cell populations in both the NP and the AF. In the NP, we detected a notochordal-like cell cluster and a progenitor stem cell cluster. In the AF, we also detected

a stem cell-like cell cluster, a cluster with a predominantly fibroblast-like phenotype and an endothelial progenitor cell like cluster.

**Supplementary Materials:** The following are available at <https://www.mdpi.com/article/10.3390/ijms22094917/s1>.

**Author Contributions:** Conceptualization was performed by M.C., L.P. and F.I. Data Curation was performed by M.C. and F.I. Formal analysis was performed by M.C. and F.I. Funding acquisition was performed by M.E. and F.I. The investigation was performed by M.C. Project administration was performed by M.C. Resources were provided by M.E., L.P. and F.I. Supervision was provided by F.I. Visualization was provided by M.C., B.G. and F.I. The manuscript was written by M.C., L.P. and F.I. The manuscript was reviewed and edited by M.C., B.G., M.E., L.P. and F.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study as the tissue was collected from the local abattoire, hence from the food chain.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** <https://www.ebi.ac.uk/ena/browser/view/PRJEB44744> (accessed on 19 March 2021).

**Acknowledgments:** We thank Dayme Gonzalez Rodrigues, Emilio Yángüez and Hubert Rehrauer from the FGCZ for their support with the data analysis. We thank Daniel Sánchez at DMBR of the University of Bern for his technical support with Single R.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AF	Annulus Fibrosus
AFC	Annulus Fibrosus Cell
CCA	Canonical Correlation Analysis
CEP	Cartilage Endplate
GSEA	Gene Set Enrichment analysis
GSVA	Gene Set Variant Analysis
IVD	Intervertebral disc
KNN	K-nearest neighbors
MNN	Mutual Nearest Neighbors
NP	Nucleus Pulposus
NPC	Nucleus Pulposus Cell
PCA	Principal Component Analysis
RISH	Hybridization on paraffin sections
SEA	Singular Enrichment Analysis
t-SNE	t-stochastic neighbor embedding
UMI	Unique Molecular Identifier

## References

1. Risbud, M.V.; Schaer, T.P.; Shapiro, I.M. Toward an Understanding of the Role of Notochordal Cells in the Adult Intervertebral Disc: From Discord to Accord. *Dev. Dyn.* **2010**, *239*, 2141–2148. [[CrossRef](#)]
2. Tessier, S.; Risbud, M.V. Understanding Embryonic Development for Cell-Based Therapies of Intervertebral Disc Degeneration: Toward an Effort to Treat Disc Degeneration Subphenotypes. *Dev. Dyn.* **2020**, dvyd.217,
3. Cox, M.K.; Serra, R. Development of the Intervertebral Disc. In *The Intervertebral Disc: Molecular and Structural Studies of the Disc in Health and Disease*; Shapiro, I.M., Risbud, M.V., Eds.; Springer: Vienna, Austria, 2014; pp. 33–51. [[CrossRef](#)]
4. Madhu, V.; Boneski, P.K.; Silagi, E.; Qiu, Y.; Kurland, I.; Guntur, A.R.; Shapiro, I.M.; Risbud, M.V. Hypoxic Regulation of Mitochondrial Metabolism and Mitophagy in Nucleus Pulposus Cells Is Dependent on HIF-1 $\alpha$ -BNIP3 Axis. *J. Bone Miner. Res.* **2020**, jbmr.4019,
5. Baldit, A. Micromechanics of the intervertebral disk. In *Multiscale Biomechanics*; Elsevier: Oxford, UK, 2018; pp. 455–467. [[CrossRef](#)]

6. Shapiro, I.M.; Risbud, M.V. Transcriptional Profiling of the Nucleus Pulposus: Say Yes to Notochord. *Arthritis Res. Ther.* **2010**, *12*, 117. [[CrossRef](#)]
7. D'Este, M.; Eglin, D.; Alini, M. Lessons to Be Learned and Future Directions for Intervertebral Disc Biomaterials. *Acta Biomater.* **2018**, *78*, 13–22. [[CrossRef](#)] [[PubMed](#)]
8. Cassidy, J.J.; Hiltner, A.; Baer, E. Hierarchical Structure of the Intervertebral Disc. *Connect. Tissue Res.* **1989**, *23*, 75–88. [[CrossRef](#)] [[PubMed](#)]
9. Moore, R.J. The Vertebral Endplate: Disc Degeneration, Disc Regeneration. *Eur. Spine J.* **2006**, *15*, 333–337. [[CrossRef](#)] [[PubMed](#)]
10. Gruber, H.E.; Riley, F.E.; Hoelscher, G.L.; Ingram, J.A.; Bullock, L.; Hanley, E.N. Human Annulus Progenitor Cells: Analyses of This Viable Endogenous Cell Population: Progenitor Cells in the Human annulus. *J. Orthop. Res.* **2016**, *34*, 1351–1360. [[CrossRef](#)]
11. Choi, K.S.; Cohn, M.J.; Harfe, B.D. Identification of Nucleus Pulposus Precursor Cells and Notochordal Remnants in the Mouse: Implications for Disk Degeneration and Chordoma Formation. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **2008**, *237*, 3953–3958. [[CrossRef](#)]
12. McCann, M.R.; Tamplin, O.J.; Rossant, J.; Séguin, C.A. Tracing Notochord-Derived Cells Using a Noto-Cre Mouse: Implications for Intervertebral Disc Development. *Dis. Model. Mech.* **2012**, *5*, 73–82. [[CrossRef](#)]
13. Purmessur, D.; Guterl, C.C.; Cho, S.K.; Cornejo, M.C.; Lam, Y.W.; Ballif, B.A.; Laudier, J.C.I.; Iatridis, J.C. Dynamic Pressurization Induces Transition of Notochordal Cells to a Mature Phenotype While Retaining Production of Important Patterning Ligands from Development. *Arthritis Res. Ther.* **2013**, *15*, R122. [[CrossRef](#)]
14. Shapiro, I.M.; Risbud, M.V. Introduction to the Structure, Function, and Comparative Anatomy of the Vertebrae and the Intervertebral Disc. In *The Intervertebral Disc: Molecular and Structural Studies of the Disc in Health and Disease*; Shapiro, I.M., Risbud, M.V., Eds.; Springer: Vienna, Austria, 2014; pp. 3–15. [[CrossRef](#)]
15. Risbud, M.V.; Shapiro, I.M. Notochordal Cells in the Adult Intervertebral Disc: New Perspective on an Old Question. *Crit. Rev. Eukaryot. Gene Expr.* **2011**, *21*, 29–41. [[CrossRef](#)]
16. Shapiro, I.M.; Risbud, M.V. (Eds.) *The Intervertebral Disc: Molecular and Structural Studies of the Disc in Health and Disease*; Springer: Wien, Austria, 2014. [[CrossRef](#)]
17. Torre, O.M.; Mroz, V.; Benitez, A.R.M.; Huang, A.H.; Iatridis, J.C. Neonatal Annulus Fibrosus Regeneration Occurs via Recruitment and Proliferation of Scleraxis-Lineage Cells. *NPJ Regen. Med.* **2019**, *4*, 23. [[CrossRef](#)]
18. Clouet, J.; Grimandi, G.; Pot-Vaucel, M.; Masson, M.; Fellah, H.B.; Guigand, L.; Cherel, Y.; Bord, E.; Rannou, F.; Weiss, P.; et al. Identification of Phenotypic Discriminating Markers for Intervertebral Disc Cells and Articular Chondrocytes. *Rheumatology* **2009**, *48*, 1447–1450. [[CrossRef](#)]
19. Li, K.; Kapper, D.; Mondal, S.; Lufkin, T.; Kraus, P. Quantitative Single-Cell Transcript Assessment of Biomarkers Supports Cellular Heterogeneity in the Bovine IVD. *Vet. Sci.* **2019**, *6*, 42. [[CrossRef](#)]
20. van den Akker, G.G.H.; Koenders, M.I.; van de Loo, F.A.J.; van Lent, P.L.E.M.; Blaney Davidson, E.; van der Kraan, P.M. Transcriptional Profiling Distinguishes Inner and Outer Annulus Fibrosus from Nucleus Pulposus in the Bovine Intervertebral Disc. *Eur. Spine J.* **2017**, *26*, 2053–2062. [[CrossRef](#)]
21. Tam, V.; Chen, P.; Yee, A.; Solis, N.; Klein, T.; Kudelko, M.; Sharma, R.; Chan, W.C.; Overall, C.M.; Haglund, L.; et al. DIPPER: A Spatiotemporal Proteomics Atlas of Human Intervertebral Discs for Exploring Ageing and Degeneration Dynamics. *bioRxiv* **2020**. [[CrossRef](#)]
22. Reed, C.C.; Iozzo, R.V. The Role of Decorin in Collagen Fibrillogenesis and Skin Homeostasis. *Glycoconj. J.* **2002**, *19*, 249–255. [[CrossRef](#)]
23. Chen, S.; Birk, D.E. The Regulatory Roles of Small Leucine-Rich Proteoglycans in Extracellular Matrix Assembly. *FEBS J.* **2013**, *280*, 2120–2137. [[CrossRef](#)]
24. Hayes, A.J.; Isaacs, M.D.; Hughes, C.; Caterson, B.; Ralphs, J.R. Collagen Fibrillogenesis in the Development of the Annulus Fibrosus of the Intervertebral Disc. *Eur. Cells Mater.* **2011**, *22*, 226–241. [[CrossRef](#)]
25. Trout, J.J.; Buckwalter, J.A.; Moore, K.C. Ultrastructure of the Human Intervertebral Disc: II. Cells of the Nucleus Pulposus. *Anat. Rec.* **1982**, *204*, 307–314. [[CrossRef](#)] [[PubMed](#)]
26. Sakai, D.; Schol, J.; Bach, F.C.; Tekari, A.; Sagawa, N.; Nakamura, Y.; Chan, S.C.W.; Nakai, T.; Creemers, L.B.; Frauchiger, D.A.; et al. Successful Fishing for Nucleus Pulposus Progenitor Cells of the Intervertebral Disc across Species. *JOR Spine* **2018**, *1*, e1018. [[CrossRef](#)]
27. Brisby, H.; Papadimitriou, N.; Brantsing, C.; Bergh, P.; Lindahl, A.; Barreto Henriksson, H. The Presence of Local Mesenchymal Progenitor Cells in Human Degenerated Intervertebral Discs and Possibilities to Influence These In Vitro: A Descriptive Study in Humans. *Stem Cells Dev.* **2013**, *22*, 804–814. [[CrossRef](#)]
28. Hunter, C.J.; Matyas, J.R.; Duncan, N.A. Cytomorphology of Notochordal and Chondrocytic Cells from the Nucleus Pulposus: A Species Comparison. *J. Anat.* **2004**, *205*, 357–362.
29. Risbud, M.V.; Guttapalli, A.; Tsai, T.T.; Lee, J.Y.; Danielson, K.G.; Vaccaro, A.R.; Albert, T.J.; Gazit, Z.; Gazit, D.; Shapiro, I.M. Evidence for Skeletal Progenitor Cells in the Degenerate Human Intervertebral Disc. *Spine* **2007**, *32*, 2537–2544. [[CrossRef](#)] [[PubMed](#)]
30. Leung, V.Y.L.; Aladin, D.M.K.; Lv, F.; Tam, V.; Sun, Y.; Lau, R.Y.C.; Hung, S.C.; Ngan, A.H.W.; Tang, B.; Lim, C.T.; et al. Mesenchymal Stem Cells Reduce Intervertebral Disc Fibrosis and Facilitate Repair. *Stem Cells* **2014**, *32*, 2164–2177. [[CrossRef](#)] [[PubMed](#)]

31. Kraus, P.; Yerden, R.; Kocsis, V.; Lufkin, T. RNA in Situ Hybridization Characterization of Non-Enzymatic Derived Bovine Intervertebral Disc Cell Lineages Suggests Progenitor Cell Potential. *Acta Histochem.* **2017**, *119*, 150–160. [[CrossRef](#)]
32. Kraus, P.; Yerden, R.; Sipes, D.; Sur, S.; Lufkin, T. A Quantitative and Qualitative RNA Expression Profiling Assay for Cell Culture with Single Cell Resolution. *Cytotechnology* **2018**, *70*, 185–192. [[CrossRef](#)]
33. Kraus, P.; Li, K.; Sipes, D.; Varden, L.; Yerden, R.; Henderson, A.; Sur, S.; Lufkin, T. *Single-Cell Phenotyping of Complex Heterogeneous Tissue*; Springer Nature: Singapore, 2019; p. 17.
34. De Luca, P.; Castagnetta, M.; de Girolamo, L.; Coco, S.; Malacarne, M.; Ragni, E.; Viganò, M.; Lugano, G.; Brayda-Bruno, M.; Coviello, D.; et al. Intervertebral Disc and Endplate Cell Characterisation Highlights Annulus Fibrosus Cells as the Most Promising for Tissue-Specific Disc Degeneration Therapy. *Eur. Cells Mater.* **2020**, *39*, 156–170. [[CrossRef](#)]
35. Veras, M.A.; McCann, M.R.; Tenn, N.A.; Séguin, C.A. Transcriptional Profiling of the Murine Intervertebral Disc and Age-Associated Changes in the Nucleus Pulposus. *Connect. Tissue Res.* **2020**, *61*, 63–81. [[CrossRef](#)]
36. Rodrigues-Pinto, R.; Ward, L.; Humphreys, M.; Zeef, L.A.H.; Berry, A.; Hanley, K.P.; Hanley, N.; Richardson, S.M.; Hoyland, J.A. Human Notochordal Cell Transcriptome Unveils Potential Regulators of Cell Function in the Developing Intervertebral Disc. *Sci. Rep.* **2018**, *8*, 12866. [[CrossRef](#)]
37. Richardson, S.M.; Kalamegam, G.; Pushparaj, P.N.; Matta, C.; Memic, A.; Khademhosseini, A.; Mobasheri, R.; Poletti, F.L.; Hoyland, J.A.; Mobasheri, A. Mesenchymal Stem Cells in Regenerative Medicine: Focus on Articular Cartilage and Intervertebral Disc Regeneration. *Methods A Companion Methods Enzymol.* **2016**, *99*, 69–80. [[CrossRef](#)] [[PubMed](#)]
38. Weiler, C.; Nerlich, A.G.; Schaaf, R.; Bachmeier, B.E.; Wuertz, K.; Boos, N. Immunohistochemical Identification of Notochordal Markers in Cells in the Aging Human Lumbar Intervertebral Disc. *Eur. Spine J. Off. Publ. Eur. Spine Soc. Eur. Spinal Deform. Soc. Eur. Sect. Cerv. Spine Res. Soc.* **2010**, *19*, 1761–1770. [[CrossRef](#)]
39. Rodrigues-Pinto, R.; Berry, A.; Piper-Hanley, K.; Hanley, N.; Richardson, S.M.; Hoyland, J.A. Spatiotemporal Analysis of Putative Notochordal Cell Markers Reveals CD24 and Keratins 8, 18, and 19 as Notochord-Specific Markers during Early Human Intervertebral Disc Development. *J. Orthop. Res.* **2016**, *34*, 1327–1340. [[CrossRef](#)] [[PubMed](#)]
40. Liu, Z.; Zheng, Z.; Qi, J.; Wang, J.; Zhou, Q.; Hu, F.; Liang, J.; Li, C.; Zhang, W.; Zhang, X. CD24 Identifies Nucleus Pulposus Progenitors/Notochordal Cells for Disc Regeneration. *J. Biol. Eng.* **2018**, *12*, 35. [[CrossRef](#)]
41. Minogue, B.M.; Richardson, S.M.; Zeef, L.A.; Freemont, A.J.; Hoyland, J.A. Transcriptional Profiling of Bovine Intervertebral Disc Cells: Implications for Identification of Normal and Degenerate Human Intervertebral Disc Cell Phenotypes. *Arthritis Res. Ther.* **2010**, *12*, 1–20. [[CrossRef](#)]
42. Rutges, J.; Creemers, L.; Dhert, W.; Milz, S.; Sakai, D.; Mochida, J.; Alini, M.; Grad, S. Variations in Gene and Protein Expression in Human Nucleus Pulposus in Comparison with Annulus Fibrosus and Cartilage Cells: Potential Associations with Aging and Degeneration. *Osteoarthr. Cartil.* **2010**, *18*, 416–423. [[CrossRef](#)]
43. Peck, S.H.; McKee, K.K.; Tobias, J.W.; Malhotra, N.R.; Harfe, B.D.; Smith, L.J. Whole Transcriptome Analysis of Notochord-Derived Cells during Embryonic Formation of the Nucleus Pulposus. *Sci. Rep.* **2017**, *7*, 1–14. [[CrossRef](#)]
44. Sakai, D.; Nakamura, Y.; Nakai, T.; Mishima, T.; Kato, S.; Grad, S.; Alini, M.; Risbud, M.V.; Chan, D.; Cheah, K.S.E.; et al. Exhaustion of Nucleus Pulposus Progenitor Cells with Ageing and Degeneration of the Intervertebral Disc. *Nat. Commun.* **2012**, *3*, 1211–1264. [[CrossRef](#)]
45. Richardson, S.M.; Ludwinski, F.E.; Gnanalingham, K.K.; Atkinson, R.A.; Freemont, A.J.; Hoyland, J.A. Notochordal and Nucleus Pulposus Marker Expression Is Maintained by Sub-Populations of Adult Human Nucleus Pulposus Cells through Aging and Degeneration. *Sci. Rep.* **2017**, *7*, 1501. [[CrossRef](#)]
46. Schol, J.; Sakai, D. Cell Therapy for Intervertebral Disc Herniation and Degenerative Disc Disease: Clinical Trials. *Int. Orthop.* **2019**, *43*, 1011–1025. [[CrossRef](#)] [[PubMed](#)]
47. Feng, C.; Liu, H.; Yang, M.; Zhang, Y.; Huang, B.; Zhou, Y. Disc Cell Senescence in Intervertebral Disc Degeneration: Causes and Molecular Pathways. *Cell Cycle* **2016**, *15*, 1674–1684. [[CrossRef](#)] [[PubMed](#)]
48. Lakstins, K.; Arnold, L.; Gunsch, G.; Khan, S.; Moore, S.; Purmessur, D. Characterization of Bovine and Canine Animal Model Cartilage Endplates and Comparison to Human Cartilage Endplate Structure, Matrix Composition, and Cell Phenotype. *JOR Spine* **2020**, *3*. [[CrossRef](#)]
49. Dal Molin, A.; Baruzzo, G.; Di Camillo, B. Single-Cell RNA-Sequencing: Assessment of Differential Expression Analysis Methods. *Front. Genet.* **2017**, *8*. [[CrossRef](#)]
50. Soneson, C.; Robinson, M.D. Bias, Robustness and Scalability in Single-Cell Differential Expression Analysis. *Nat. Methods* **2018**, *15*, 255–261. [[CrossRef](#)]
51. Chen, J.; Yan, W.; Setton, L.A. Molecular Phenotypes of Notochordal Cells Purified from Immature Nucleus Pulposus. *Eur. Spine J.* **2006**, *15*, 303–311. [[CrossRef](#)]
52. Kim, J.H.; Deasy, B.M.; Seo, H.Y.; Studer, R.K.; Vo, N.V.; Georgescu, H.I.; Sowa, G.A.; Kang, J.D. Differentiation of Intervertebral Notochordal Cells Through Live Automated Cell Imaging System In Vitro. *Spine* **2009**, *34*, 2486–2493. [[CrossRef](#)]
53. Oguz, E.; Tsai, T.T.; Di Martino, A.; Guttapalli, A.; Albert, T.J.; Shapiro, I.M.; Risbud, M.V. Galectin-3 Expression in the Intervertebral Disc: A Useful Marker of the Notochord Phenotype? *Spine* **2007**, *32*, 9–16. [[CrossRef](#)]
54. Poiraudau, S.; Monteiro, I.; Anract, P.; Blanchard, O.; Revel, M.; Corvol, M.T. Phenotypic Characteristics of Rabbit Intervertebral Disc Cells. Comparison with Cartilage Cells from the Same Animals. *Spine* **1999**, *24*, 837–844. [[CrossRef](#)]

55. Rastogi, A.; Thakore, P.; Leung, A.; Benavides, M.; Machado, M.; Morschauser, M.A.; Hsieh, A.H. Environmental Regulation of Notochordal Gene Expression in Nucleus Pulposus Cells. *J. Cell. Physiol.* **2009**, *220*, 698–705. [[CrossRef](#)]
56. Tang, X.; Jing, L.; Chen, J. Changes in the Molecular Phenotype of Nucleus Pulposus Cells with Intervertebral Disc Aging. *PLoS ONE* **2012**, *7*, e52020. [[CrossRef](#)]
57. He, S.; Wang, L.H.; Liu, Y.; Li, Y.Q.; Chen, H.T.; Xu, J.H.; Peng, W.; Lin, G.W.; Wei, P.P.; Li, B.; et al. Single-Cell Transcriptome Profiling of an Adult Human Cell Atlas of 15 Major Organs. *Genome Biol.* **2020**, *21*. [[CrossRef](#)]
58. Montoro, D.T.; Haber, A.L.; Biton, M.; Vinarsky, V.; Lin, B.; Birket, S.E.; Yuan, F.; Chen, S.; Leung, H.M.; Villoria, J.; et al. A Revised Airway Epithelial Hierarchy Includes CFTR-Expressing Ionocytes. *Nature* **2018**, *560*, 319–324. [[CrossRef](#)]
59. Plasschaert, L.W.; Žilionis, R.; Choo-Wing, R.; Savova, V.; Knehr, J.; Roma, G.; Klein, A.M.; Jaffe, A.B. A Single-Cell Atlas of the Airway Epithelium Reveals the CFTR-Rich Pulmonary Ionocyte. *Nature* **2018**, *560*, 377–381. [[CrossRef](#)]
60. Villani, A.C.; Satija, R.; Reynolds, G.; Sarkizova, S.; Shekhar, K.; Fletcher, J.; Griesbeck, M.; Butler, A.; Zheng, S.; Lazo, S.; et al. Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors. *Science* **2017**, *356*. [[CrossRef](#)]
61. Pal, B.; Chen, Y.; Vaillant, F.; Jamieson, P.; Gordon, L.; Rios, A.C.; Wilcox, S.; Fu, N.; Liu, K.H.; Jackling, F.C.; et al. Construction of Developmental Lineage Relationships in the Mouse Mammary Gland by Single-Cell RNA Profiling. *Nat. Commun.* **2017**, *8*, 1627. [[CrossRef](#)]
62. Kiselev, V.Y.; Andrews, T.S.; Hemberg, M. Challenges in Unsupervised Clustering of Single-Cell RNA-Seq Data. *Nat. Rev. Genet.* **2019**, *20*, 273–282. [[CrossRef](#)]
63. Lähnemann, D.; Köster, J.; Szczurek, E.; McCarthy, D.J.; Hicks, S.C.; Robinson, M.D.; Vallejos, C.A.; Campbell, K.R.; Beerenwinkel, N.; Mahfouz, A.; et al. Eleven Grand Challenges in Single-Cell Data Science. *Genome Biol.* **2020**, *21*, 31. [[CrossRef](#)] [[PubMed](#)]
64. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21. [[CrossRef](#)] [[PubMed](#)]
65. Fuchs, E.; Coulombe, P.A. Of Mice and Men: Genetic Skin Diseases of Keratin. *Cell* **1992**, *69*, 899–902. [[CrossRef](#)]
66. Bedore, J.; Sha, W.; McCann, M.R.; Liu, S.; Leask, A.; Séguin, C.A. Impaired Intervertebral Disc Development and Premature Disc Degeneration in Mice with Notochord-Specific Deletion of CCN2. *Arthritis Rheum.* **2013**, *65*. [[CrossRef](#)]
67. Russo, J.W.; Castellot, J.J. CCN5: Biology and Pathophysiology. *J. Cell Commun. Signal.* **2010**, *4*, 119–130. [[CrossRef](#)]
68. Lake, A.C.; Bialik, A.; Walsh, K.; Castellot, J.J. CCN5 Is a Growth Arrest-Specific Gene That Regulates Smooth Muscle Cell Proliferation and Motility. *Am. J. Pathol.* **2003**, *162*, 219–231. [[CrossRef](#)]
69. Ukita, K.; Hirahara, S.; Oshima, N.; Imuta, Y.; Yoshimoto, A.; Jang, C.W.; Oginuma, M.; Saga, Y.; Behringer, R.R.; Kondoh, H.; et al. Wnt Signaling Maintains the Notochord Fate for Progenitor Cells and Supports the Posterior Extension of the Notochord. *Mech. Dev.* **2009**, *126*, 791–803. [[CrossRef](#)]
70. Abreu, J.G.; Ketpura, N.I.; Reversade, B.; De Robertis, E.M. Connective-Tissue Growth Factor (CTGF) Modulates Cell Signalling by BMP and TGF-Beta. *Nat. Cell Biol.* **2002**, *4*, 599–604. [[CrossRef](#)]
71. Si, W.; Kang, Q.; Luu, H.H.; Park, J.K.; Luo, Q.; Song, W.X.; Jiang, W.; Luo, X.; Li, X.; Yin, H.; et al. CCN1/Cyr61 Is Regulated by the Canonical Wnt Signal and Plays an Important Role in Wnt3A-Induced Osteoblast Differentiation of Mesenchymal Stem Cells. *Mol. Cell. Biol.* **2006**, *26*, 2955–2964. [[CrossRef](#)]
72. Kondo, N.; Yuasa, T.; Shimono, K.; Tung, W.; Okabe, T.; Yasuhara, R.; Pacifici, M.; Zhang, Y.; Iwamoto, M.; Enomoto-Iwamoto, M. Intervertebral Disc Development Is Regulated by Wnt/ $\beta$ -Catenin Signaling. *Spine* **2011**, *36*, E513–E518. [[CrossRef](#)]
73. Hiyama, A.; Sakai, D.; Risbud, M.V.; Tanaka, M.; Arai, F.; Abe, K.; Mochida, J. Enhancement of Intervertebral Disc Cell Senescence by WNT/ $\beta$ -Catenin Signaling-Induced Matrix Metalloproteinase Expression. *Arthritis Rheum.* **2010**, *62*, 3036–3047. [[CrossRef](#)]
74. Xia, Z.; Wu, S.; Wei, X.; Liao, Y.; Yi, P.; Liu, Y.; Liu, J.; Liu, J. Hypoxic ER Stress Suppresses  $\beta$ -Catenin Expression and Promotes Cooperation between the Transcription Factors XBP1 and HIF1 $\alpha$  for Cell Survival. *J. Biol. Chem.* **2019**, *294*, 13811–13821. [[CrossRef](#)]
75. Arkesteijn, I.T.M.; Potier, E.; Ito, K. The Regenerative Potential of Notochordal Cells in a Nucleus Pulposus Explant. *Glob. Spine J.* **2017**, *7*, 14–20. [[CrossRef](#)]
76. Erwin, W.M.; Islam, D.; Inman, R.D.; Fehlings, M.G.; Tsui, F.W.L. Notochordal Cells Protect Nucleus Pulposus Cells from Degradation and Apoptosis: Implications for the Mechanisms of Intervertebral Disc Degeneration. *Arthritis Res. Ther.* **2011**, *13*, R215. [[CrossRef](#)] [[PubMed](#)]
77. Matta, A.; Karim, M.Z.; Isenman, D.E.; Erwin, W.M. Molecular Therapy for Degenerative Disc Disease: Clues from Secretome Analysis of the Notochordal Cell-Rich Nucleus Pulposus. *Sci. Rep.* **2017**, *7*, 45623. [[CrossRef](#)] [[PubMed](#)]
78. Okada, K.; Mori, D.; Makii, Y.; Nakamoto, H.; Murahashi, Y.; Yano, F.; Chang, S.H.; Taniguchi, Y.; Kobayashi, H.; Semba, H.; et al. Hypoxia-Inducible Factor-1 Alpha Maintains Mouse Articular Cartilage through Suppression of NF- $\kappa$ B Signaling. *Sci. Rep.* **2020**, *10*. [[CrossRef](#)] [[PubMed](#)]
79. Chen, F.; Jiang, G.; Liu, H.; Li, Z.; Pei, Y.; Wang, H.; Pan, H.; Cui, H.; Long, J.; Wang, J.; et al. Melatonin Alleviates Intervertebral Disc Degeneration by Disrupting the IL-1 $\beta$ /NF- $\kappa$ B-NLRP3 Inflammation Positive Feedback Loop. *Bone Res.* **2020**, *8*, 1–13. [[CrossRef](#)]
80. Alhawiti, N.M.; Al Mahri, S.; Aziz, M.A.; Malik, S.S.; Mohammad, S. TXNIP in Metabolic Regulation: Physiological Role and Therapeutic Outlook. *Curr. Drug Targets* **2017**, *18*, 1095–1103. [[CrossRef](#)]

81. Kamitori, K.; Yamaguchi, F.; Dong, Y.; Hossain, A.; Katagi, A.; Noguchi, C.; Hirata, Y.; Tsukamoto, I.; Hatano, N.; Tokuda, M. Both Ser361 Phosphorylation and the C-Arrestin Domain of Thioredoxin Interacting Protein Are Important for Cell Cycle Blockade at the G1/S Checkpoint. *FEBS Open Bio* **2018**, *8*, 1804–1819. [[CrossRef](#)]
82. Young, R.A. Control of Embryonic Stem Cell State. *Cell* **2011**, *144*, 940–954. [[CrossRef](#)]
83. Stegen, S.; Laperre, K.; Eelen, G.; Rinaldi, G.; Fraisl, P.; Torrekens, S.; Van Looveren, R.; Loopmans, S.; Bultynck, G.; Vinckier, S.; et al. HIF-1 $\alpha$  Metabolically Controls Collagen Synthesis and Modification in Chondrocytes. *Nature* **2019**, *565*, 511–515. [[CrossRef](#)]
84. Yamauchi, M.; Sricholpech, M. Lysine Post-Translational Modifications of Collagen. *Essays Biochem.* **2012**, *52*, 113–133. [[CrossRef](#)]
85. Myllyharju, J.; Kivirikko, K.I. Collagens, Modifying Enzymes and Their Mutations in Humans, Flies and Worms. *Trends Genet. TIG* **2004**, *20*, 33–43. [[CrossRef](#)]
86. Fernandes, L.M.; Khan, N.M.; Trochez, C.M.; Duan, M.; Diaz-Hernandez, M.E.; Presciutti, S.M.; Gibson, G.; Drissi, H. Single-Cell RNA-Seq Identifies Unique Transcriptional Landscapes of Human Nucleus Pulposus and Annulus Fibrosus Cells. *Sci. Rep.* **2020**, *10*, 15263. [[CrossRef](#)]
87. Risbud, M.V.; Schoepflin, Z.R.; Mwale, F.; Kandel, R.A.; Grad, S.; Iatridis, J.C.; Sakai, D.; Hoyland, J.A. Defining the Phenotype of Young Healthy Nucleus Pulposus Cells: Recommendations of the Spine Research Interest Group at the 2014 Annual ORS Meeting. *J. Orthop. Res. Off. Publ. Orthop. Res. Soc.* **2015**, *33*, 283–293. [[CrossRef](#)]
88. Anderson, R.M.; Lawrence, A.R.; Stottmann, R.W.; Bachiller, D.; Klingensmith, J. Chordin and Noggin Promote Organizing Centers of Forebrain Development in the Mouse. *Development* **2002**, *129*, 4975–4987. [[CrossRef](#)]
89. Johnson, J.L.F.; Hall, T.E.; Dyson, J.M.; Sonntag, C.; Ayers, K.; Berger, S.; Gautier, P.; Mitchell, C.; Hollway, G.E.; Currie, P.D. Scube Activity Is Necessary for Hedgehog Signal Transduction in Vivo. *Dev. Biol.* **2012**, *368*, 193–202. [[CrossRef](#)]
90. Nakamichi, R.; Kataoka, K.; Asahara, H. Essential Role of Mohawk for Tenogenic Tissue Homeostasis Including Spinal Disc and Periodontal Ligament. *Mod. Rheumatol.* **2018**, *28*, 933–940. [[CrossRef](#)]
91. Harris, B.S.; Zhang, Y.; Card, L.; Rivera, L.B.; Brekken, R.A.; Bradshaw, A.D. SPARC Regulates Collagen Interaction with Cardiac Fibroblast Cell Surfaces. *Am. J. Physiol. Heart Circ. Physiol.* **2011**, *301*, H841–H847. [[CrossRef](#)]
92. McCurdy, S.; Baicu, C.F.; Heymans, S.; Bradshaw, A.D. Cardiac Extracellular Matrix Remodeling: Fibrillar Collagens and Secreted Protein Acidic and Rich in Cysteine (SPARC). *J. Mol. Cell. Cardiol.* **2010**, *48*, 544–549. [[CrossRef](#)]
93. Bradshaw, A.D.; Sage, E.H. SPARC, a Matricellular Protein That Functions in Cellular Differentiation and Tissue Response to Injury. *J. Clin. Investig.* **2001**, *107*, 1049–1054. [[CrossRef](#)]
94. Brekken, R.A.; Sage, E.H. SPARC, a Matricellular Protein: At the Crossroads of Cell-Matrix Communication. *Matrix Biol. J. Int. Soc. Matrix Biol.* **2001**, *19*, 816–827. [[CrossRef](#)]
95. Giudici, C.; Raynal, N.; Wiedemann, H.; Cabral, W.A.; Marini, J.C.; Timpl, R.; Bächinger, H.P.; Fardale, R.W.; Sasaki, T.; Tenni, R. Mapping of SPARC/BM-40/Osteonectin-Binding Sites on Fibrillar Collagens. *J. Biol. Chem.* **2008**, *283*, 19551–19560. [[CrossRef](#)]
96. Sasaki, T.; Göhring, W.; Mann, K.; Maurer, P.; Hohenester, E.; Knäuper, V.; Murphy, G.; Timpl, R. Limited Cleavage of Extracellular Matrix Protein BM-40 by Matrix Metalloproteinases Increases Its Affinity for Collagens. *J. Biol. Chem.* **1997**, *272*, 9237–9243. [[CrossRef](#)] [[PubMed](#)]
97. Gruber, H.E.; Ingram, J.A.; Leslie, K.; Hanley, E.N.J. Cellular, but Not Matrix, Immunolocalization of SPARC in the Human Intervertebral Disc: Decreasing Localization with Aging and Disc Degeneration. *Spine* **2004**, *29*, 2223–2228. [[CrossRef](#)] [[PubMed](#)]
98. Tajerian, M.; Alvarado, S.; Millecamps, M.; Dashwood, T.; Anderson, K.M.; Haglund, L.; Ouellet, J.; Szyf, M.; Stone, L.S. DNA Methylation of SPARC and Chronic Low Back Pain. *Mol. Pain* **2011**, *7*. [[CrossRef](#)] [[PubMed](#)]
99. Subramanian, A.; Schilling, T.F. Thrombospondin-4 Controls Matrix Assembly during Development and Repair of Myotendinous Junctions. *eLife* **2014**, *3*, e02372. [[CrossRef](#)] [[PubMed](#)]
100. Kitaori, T.; Ito, H.; Schwarz, E.M.; Tsutsumi, R.; Yoshitomi, H.; Oishi, S.; Nakano, M.; Fujii, N.; Nagasawa, T.; Nakamura, T. Stromal Cell-Derived Factor 1/CXCR4 Signaling Is Critical for the Recruitment of Mesenchymal Stem Cells to the Fracture Site during Skeletal Repair in a Mouse Model. *Arthritis Rheum.* **2009**, *60*, 813–823. [[CrossRef](#)]
101. Zhang, Q.; Wang, Q.; Gong, J.; Du, J.; Zhang, Y.; Zhao, X. Yak IGF2 Promotes Fibroblast Proliferation Via Suppression of IGF1R and PI3KCG Expression. *Genes* **2018**, *9*, 169. [[CrossRef](#)]
102. Patan, S. TIE1 and TIE2 Receptor Tyrosine Kinases Inversely Regulate Embryonic Angiogenesis by the Mechanism of Intussusceptive Microvascular Growth. *Microvasc. Res.* **1998**, *56*, 1–21. [[CrossRef](#)]
103. Leppänen, V.M.; Saharinen, P.; Alitalo, K. Structural Basis of Tie2 Activation and Tie2/Tie1 Heterodimerization. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4376–4381. [[CrossRef](#)]
104. Singh, H.; Milner, C.S.; Aguilar Hernandez, M.M.; Patel, N.; Brindle, N.P.J. Vascular Endothelial Growth Factor Activates the Tie Family of Receptor Tyrosine Kinases. *Cell. Signal.* **2009**, *21*, 1346–1350. [[CrossRef](#)]
105. Leavitt, T.; Hu, M.S.; Borrelli, M.R.; Januszzyk, M.; Garcia, J.T.; Ransom, R.C.; Mascharak, S.; desJardins-Park, H.E.; Litzemberger, U.M.; Walmsley, G.G.; et al. Prrx1 Fibroblasts Represent a Pro-Fibrotic Lineage in the Mouse Ventral Dermis. *Cell Rep.* **2020**, *33*, 108356. [[CrossRef](#)]
106. Solé-Boldo, L.; Raddatz, G.; Schütz, S.; Mallm, J.P.; Rippe, K.; Lonsdorf, A.S.; Rodríguez-Paredes, M.; Lyko, F. Single-Cell Transcriptomes of the Human Skin Reveal Age-Related Loss of Fibroblast Priming. *Commun. Biol.* **2020**, *3*, 1–12. [[CrossRef](#)] [[PubMed](#)]



107. Guerrero-Juarez, C.F.; Dedhia, P.H.; Jin, S.; Ruiz-Vega, R.; Ma, D.; Liu, Y.; Yamaga, K.; Shestova, O.; Gay, D.L.; Yang, Z.; et al. Single-Cell Analysis Reveals Fibroblast Heterogeneity and Myeloid-Derived Adipocyte Progenitors in Murine Skin Wounds. *Nat. Commun.* **2019**, *10*, 650. [CrossRef] [PubMed]
108. Parreno, J.; Raju, S.; Wu, P.H.; Kandel, R.A. MRTF-A Signaling Regulates the Acquisition of the Contractile Phenotype in Dedifferentiated Chondrocytes. *Matrix Biol.* **2017**, *62*, 3–14. [CrossRef] [PubMed]
109. Fortelny, N.; Overall, C.M.; Pavlidis, P.; Freue, G.V.C. Can We Predict Protein from mRNA Levels? *Nature* **2017**, *547*, E19–E20. [CrossRef]
110. Beckstein, J.C.; Sen, S.; Schaer, T.P.; Vresilovic, E.J.; Elliott, D.M. Comparison of Animal Discs Used in Disc Research to Human Lumbar Disc. *Spine* **2008**, *8*. [CrossRef]
111. O'Connell, G.D.; Vresilovic, E.J.; Elliott, D.M. Comparison of Animals Used in Disc Research to Human Lumbar Disc Geometry. *Spine* **2007**, *32*, 328–333. [CrossRef]
112. Showalter, B.L.; Beckstein, J.C.; Martin, J.T.; Beattie, E.E.; Orías, A.A.E.; Schaer, T.P.; Vresilovic, E.J.; Elliott, D.M. Comparison of Animal Discs Used in Disc Research to Human Lumbar Disc: Torsion Mechanics and Collagen Content. *Spine* **2012**, *37*, E900–E907. [CrossRef]
113. Chen, J.W.; Li, B.; Yang, Y.H.; Jiang, S.D.; Jiang, L.S. Significance of Hypoxia in the Physiological Function of Intervertebral Disc Cells. *Crit. Rev. Eukaryot. Gene Expr.* **2014**, *24*, 193–204. [CrossRef]
114. Risbud, M.V.; Shapiro, I.M. Microenvironmental Control of Disc Cell Function: Influence of Hypoxia and Osmotic Pressure. In *The Intervertebral Disc: Molecular and Structural Studies of the Disc in Health and Disease*; Shapiro, I.M., Risbud, M.V., Eds.; Springer: Vienna, Austria, 2014; pp. 93–108. [CrossRef]
115. Li, H.; Tao, Y.; Liang, C.; Han, B.; Li, F.; Chen, G.; Chen, Q. Influence of Hypoxia in the Intervertebral Disc on the Biological Behaviors of Rat Adipose- and Nucleus Pulposus-Derived Mesenchymal Stem Cells. *Cells Tissues Organs* **2013**, *198*, 266–277. [CrossRef]
116. Yang, S.H.; Hu, M.H.; Sun, Y.H.; Lin, F.H. Differential Phenotypic Behaviors of Human Degenerative Nucleus Pulposus Cells under Normoxic and Hypoxic Conditions: Influence of Oxygen Concentration during Isolation, Expansion, and Cultivation. *Spine J. Off. J. N. Am. Spine Soc.* **2013**, *13*, 1590–1596. [CrossRef]
117. Wiseman, M.A.; Birch, H.L.; Akmal, M.; Goodship, A.E. Segmental Variation in the in Vitro Cell Metabolism of Nucleus Pulposus Cells Isolated from a Series of Bovine Caudal Intervertebral Discs. *Spine* **2005**, *30*, 505–511. [CrossRef]
118. Simon, A. Babraham Bioinformatics—FastQC A Quality Control Tool for High Throughput Sequence Data. 2021. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 29 April 2019).
119. Steven, W. Babraham Bioinformatics—FastQ Screen. 2021. Available online: [https://www.bioinformatics.babraham.ac.uk/projects/fastq\\_screen/](https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/) (accessed on 29 April 2019).
120. The UniVec Database. Available online: <https://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/> (accessed on 29 April 2019).
121. The UniVec Database—Refseq. Available online: <https://www.ncbi.nlm.nih.gov/refseq/> (accessed on 29 April 2019).
122. Silva. Available online: <https://www.arb-silva.de/> (accessed on 29 April 2019).
123. Hansen, K.D.; Brenner, S.E.; Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* **2010**, *38*. [CrossRef]
124. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
125. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]
126. Robinson, M.D.; Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol.* **2010**, *11*, R25. [CrossRef]
127. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species. *Nat. Biotechnol.* **2018**, *36*, 411–420. [CrossRef] [PubMed]
128. Van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
129. Lun, A.T.; McCarthy, D.J.; Marioni, J.C. A Step-by-Step Workflow for Low-Level Analysis of Single-Cell RNA-Seq Data with Bioconductor. *F1000Research* **2016**, *5*, 2122. [CrossRef] [PubMed]
130. Wu, A.R.; Neff, N.F.; Kalisky, T.; Dalerba, P.; Treutlein, B.; Rothenberg, M.E.; Mburu, F.M.; Mantalas, G.L.; Sim, S.; Clarke, M.F.; et al. Quantitative Assessment of Single-Cell RNA-Sequencing Methods. *Nat. Methods* **2014**, *11*, 41–46. [CrossRef]
131. Jew, B.; Alvarez, M.; Rahmani, E.; Miao, Z.; Ko, A.; Garske, K.M.; Sul, J.H.; Pietiläinen, K.H.; Pajukanta, P.; Halperin, E. Accurate Estimation of Cell Composition in Bulk Expression through Robust Integration of Single-Cell Information. *Nat. Commun.* **2020**, *11*, 1971. [CrossRef]