Research Article

The temporal transcriptomic signature of cartilage formation

Roland Á. Takács[1,*], Judit Vágó[1,*], Szilárd Póliska[2], Peter N. Pushparaj[3], Patrik Kovács[1], Eun-Jung Jin[4], Richard Barrett-Jolley[5], Róza Zákány[1], Csaba Matta[1,†]

[1]*Department of Anatomy, Histology and Embryology, Faculty of Medicine, University of Debrecen, Debrecen, H-4032, Hungary*

[2]*Genomic Medicine and Bioinformatic Core Facility, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, H-4032, Hungary*

[3]*Center of Excellence in Genomic Medicine Research, Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia*

[4]*Department of Biological Sciences, College of Natural Sciences, Wonkwang University, Iksan, Chunbuk, 570-749, South Korea*

[5]*Department of Musculoskeletal and Ageing Science, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, L7 8TX, United Kingdom*

*These authors contributed equally to this work.

[†]Corresponding author:

Csaba Matta – e-mail: matta.csaba@med.unideb.hu

Abstract

Chondrogenesis is a multistep process whereby cartilage progenitor cells generate a tissue with distinct structural and functional properties. Although several approaches to cartilage regeneration rely on the differentiation of implanted progenitor cells, the temporal transcriptomic landscape of *in vitro* chondrogenesis in different models has not been reported. Using RNA sequencing, we examined the differences between gene expression patterns during early cartilage formation in micromass cultures of embryonic limb bud-derived progenitors. Principal component analysis indicated a progressively different and distinct transcriptome during chondrogenesis. Differentially expressed genes (DEGs) based on pairwise comparisons of samples from consecutive days were classified into clusters and analysed. We confirmed the involvement of the top DEGs in chondrogenic differentiation with pathway analysis. We discovered several chondrogenesis-associated transcription factors and new collagen subtypes not previously linked to cartilage formation. These results provide, for the first time, a detailed insight into the molecular mechanism of *in vitro* chondrogenesis and provide a new gold standard for the temporal transcriptomic landscape of chondrogenic differentiation that may serve as a platform for new approaches in cartilage tissue engineering.

Keywords

chondrogenesis / RNASeq / next generation sequencing (NGS) / transcription factors / collagen / hyaline cartilage

List of abbreviations

DEG, differentially expressed gene; DMMB, dimethyl methylene blue; ECM, extracellular matrix; GO, gene ontology; HDC, high density culture; iPSC, induced pluripotent stem cell; LMP, limb bud mesenchymal progenitor; MSC, mesenchymal stem cell; NGS, next generation sequencing; OA, osteoarthritis; PCA, principal component analysis; QC, quality check; RNASeq, RNA sequencing; TF, transcription factor; WGCNA, weighted gene correlation network analysis

Highlights

- Chondroprogenitors derived from chicken embryonic limb buds were differentiated into chondrocytes.
- RNA-sequencing revealed expression levels of 24,145 transcripts.
- We examined the global expression patterns during chondrogenesis and defined 6 clusters.
- Using WGCNA analysis, we created a module of genes that had similar patterns to the chondrogenic transcription factor *SOX9*.
- We created a list of transcription factors with a hitherto unexplored role in chondrogenesis.

Introduction

During the morphogenesis of the vertebrate appendicular skeleton, progenitor cells derived from embryonic mesoderm undergo specification, proliferation, condensation, and nodule formation (1). The cells in these nodules, termed limb bud mesenchymal progenitors (LMPs), specify themselves to the osteochondrogenic lineage and form the supporting and connective tissues of the developing limb: bone, tendon, and cartilage. Young chondroblasts continue to proliferate and form the structures of the future skeleton. During this process, chondroblasts begin producing a cartilage-specific extracellular matrix (ECM) that gradually changes as the tissue matures. Collagen is the most abundant macromolecule in the ECM, and accounts for approximately two-thirds of the dry weight of cartilage (2). Collagens stabilise the cartilage matrix and provide tensile and shear strength. Many different types of collagen molecules are expressed in articular cartilage but the backbone polymeric framework during development is a copolymer of collagen types II, IX, and XI. Over 90% of the collagens in hyaline cartilage ECM is collagen type II (3). However, very little is known about other types of collagens synthesised by differentiating chondrocytes during chondrogenesis. Knowledge of the molecular structure of the collagen framework of articular cartilage and its development, remodelling, and maturation through its various pericellular, territorial, and interterritorial domains is critical to understanding the mechanisms of its degradation in disease.

The canonical pathway for chondrogenesis is terminal chondrocyte differentiation, followed by hypertrophy and apoptosis. This default pathway is blocked in articular cartilage, resulting in permanent cartilage (4). While its unique composition allows joints to move with little friction and acts as a shock absorber, articular cartilage has little or no ability to regenerate following injury or disease. Therefore, diseases or disorders affecting articular cartilage often result in progressive long-term pain and disability (5). For this reason, significant efforts have been made to develop novel approaches, often based on a combination of biomaterials and stem

cells, to enhance intrinsic cartilage repair, or regenerate new cartilage to repair focal defects or restore the joint surface (6).

Mesenchymal stem cells (MSCs) are promising candidates for cartilage tissue engineering (7,8). However, increasing evidence suggests that the phenotype of chondrocytes differentiated from MSCs for cartilage repair is unstable (9). It is plausible that the canonical differentiation pathway of bone marrow-derived MSCs is toward a hypertrophic phenotype, raising concerns about their applicability in tissue engineering of cartilage, as hypertrophy of chondrocytes in neocartilage ultimately leads to apoptosis and ossification (10). This seems to indicate that adult MSCs are primed for endochondral ossification and the chondrogenic state is transient. Therefore, to improve cartilage tissue engineering strategies, it might be logical to use embryonic progenitor cells (11).

The chicken limb bud-derived 3-dimensional micromass assay uses embryonic LMPs and represents a widely used, adaptable, and relatively simple *in vitro* model of the early stages of skeletal development (11). LMPs in high density micromass cultures self-organise to produce a nodular pattern, which is a pre-requisite for chondrogenic differentiation (12). This model is particularly useful for studying the early events of chondrogenesis. Several molecular pathways are involved in this complex process (13). However, the time course of transcriptomic events involved in the regulation of these pathways during *in vitro* chondrogenesis has not been fully explored. Recent studies have investigated this process by differential gene expression analysis on microarray platforms (14) and next-generation sequencing (NGS) in adult MSCs differentiated into chondrocytes in pellet cultures (6). However, as described above, the primary micromass model, established from embryonic LMPs, more consistently recapitulates the chondrogenesis of hyaline cartilage *in vivo*. To this end, we performed a quantitative study of the transcriptomic landscape of chondrogenesis using this model. We performed RNA sequencing (RNASeq) at six key time points (between culturing days 0 and 6, the latter

representing mature chondrocytes) during *in vitro* chondrogenesis in micromass cultures established from freshly isolated LMPs. Bioinformatic analysis of the transcriptome data revealed several dominant gene signatures associated with chondrogenesis. We identified the transcripts of several "minor" collagen types in differentiating micromass cultures not previously reported in cartilage and found that the LMP-based micromass model was superior to the chondrogenic MSC cultures in terms of cartilage ECM gene expression profiles. We also detected the expression of potentially novel chondrogenic transcription factors (TFs). The results of this study provide a deeper understanding of the biology of chondrocytes during key stages of chondrogenesis, which is essential for future therapeutic interventions.

Materials and Methods

Experimental Design

We used transcriptome profiling by RNASeq to identify changes in global gene expression, as chicken primary limb bud mesenchymal progenitors (LMPs) differentiate into mature chondrocytes (Fig. 1). The limb bud-derived chondrifying micromass model is a well-established method in our laboratory (15,16). Samples for RNASeq were collected at key days of chondrogenic differentiation (days 0, 1, 2, 3, 4, and 6), and subjected to high-throughput mRNA sequencing analysis on an Illumina sequencing platform. Raw sequencing data were aligned to the chicken reference genome version GRCg6a. Genes with a Benjamini-Hochberg adjusted *p*-value of less than 0.05 and a fold change cut-off at 2.0 were considered differentially expressed. Network analysis and enrichment of Gene Ontology (GO) terms were performed using PANTHER and Ingenuity Pathway Analysis (IPA) software (Qiagen, USA).

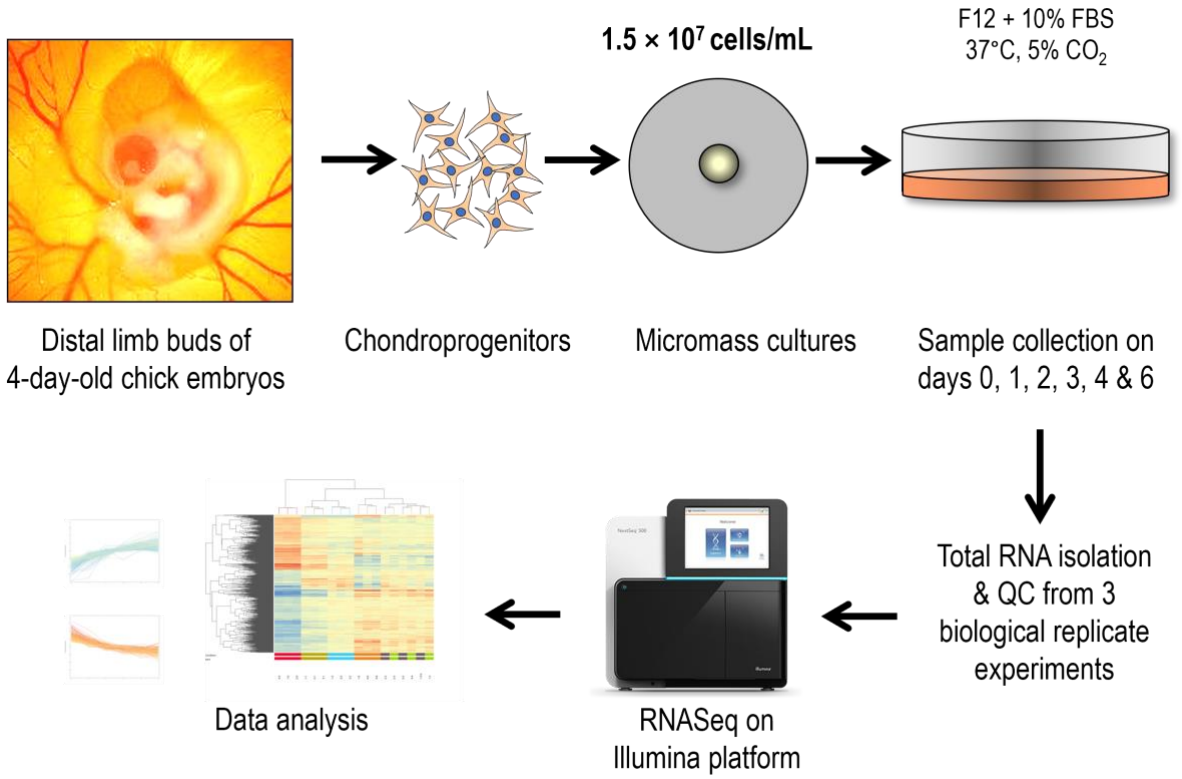

Figure 1. Experimental workflow. QC, quality check.

Cell Cultures

Primary chondrifying micromass cell cultures (high density cultures, HDCs) were used as an *in vitro* experimental model for hyaline cartilage formation (17). LMP cells in HDCs first proliferate and then differentiate into matrix-producing chondroblasts during the first 3 days of culture, and a well-detectable hyaline cartilage-specific ECM was formed on the 6th day of culture. HDCs were established as previously described (15,18,19). Briefly, chondroprogenitor cells were isolated from the developing limb buds of chicken embryos at Hamburger–Hamilton stages 22–24. The *in vitro* work on early-stage (4.5-day-old) chicken embryos does not require approval from the Ethics Committee of the University of Debrecen. Distal parts of the fore and hind limb buds of embryos were removed and dissociated in 0.25% trypsin-EDTA (Sigma-Aldrich, St. Louis, MO, USA; pH 7.4) at 37 °C for 1 h. The digestion of limb buds was terminated by the addition of an equal volume of foetal bovine serum (FBS; Gibco, Gaithersburg, MD, USA). Single-cell suspensions of chondrogenic LMP cells were obtained

by filtering through a 20-µm pore size plastic filter (Millipore, Billerica, MA, USA). After centrifugation of the suspension at 800 × *g* for 10 min at room temperature, cells were resuspended in Ham's F12 culture medium (Euroclone, Pero, Italy) supplemented with 10% FBS at a concentration of 1.5×10^7 cells/mL. Droplets (100 µL) were inoculated into Petri dishes (Eppendorf, Hamburg, Germany) and the cells were allowed to attach to the surface for 2 hours in a CO_2 incubator (5% CO_2 and 90% humidity). Finally, 1.5 mL of Ham's F12 supplemented with 10% FBS, 0.5 mM L-glutamine, and 1% penicillin/streptomycin (TEVA, Debrecen, Hungary) was added to the colonies. The day of inoculation was considered as day 0 of culturing. Cultures were maintained at 37 °C in a CO_2 incubator for 0, 1, 2, 3, 4, or 6 days. The culture medium was changed every two days.

Monitoring Culture Morphology, ECM Production, and Differentiation

Primary chondrifying cell cultures were inoculated onto the surface of 30-mm round cover slips (Menzel-Gläser, Menzel GmbH, Braunschweig, Germany) placed in 35-mm Petri dishes. Photomicrographs of the native micromass cultures were taken using an inverted microscope, and after staining with haematoxylin and eosin as described previously (20). Hyaline cartilage-specific ECM production was qualitatively analysed as follows: dimethyl methylene blue dye (DMMB; pH 1.8; Sigma-Aldrich) was applied to cultures from day 0 to day 6 as previously described (19). Photomicrographs of the stained specimens were obtained using an Olympus BX53 camera on a Nikon Eclipse E800 microscope (Nikon Corporation, Tokyo, Japan). Optical density values of DMMB-stained specimens were determined from HDCs in 3 independent, biological replicate experiments using a MATLAB image analysis application: cartilage nodules rich in metachromatic cartilage ECM were defined by an approximate range of values in the RGB colour space and the pixels were counted. Values were

normalised to day 6 values. Photomicrographs shown are visual representations of three cultures on each culturing day in three independent experiments.

Total RNA Isolation

On the designated days of culturing, HDCs were washed twice with physiological NaCl and then stored at –80 °C. For total RNA isolation, the micromass cultures were dissolved in TRI Reagent (Applied Biosystems, Foster City, CA, USA). Samples were mixed with 20% chloroform and centrifuged at 4 °C at 10,000 × *g* for 20 min. After incubation at −20 °C for 1 h in 500 µL of RNase-free isopropanol, the pellet of total RNA was dissolved in RNase-free water (Promega, Madison, WI, USA) and stored at −80 °C.

RNA Sequencing

To obtain global transcriptome data, high-throughput mRNA sequencing analysis was performed using an Illumina sequencing platform. The quality of total RNA samples was checked on an Agilent BioAnalyzer (Santa Clara, CA, USA) using the Eukaryotic Total RNA Nano Kit according to the manufacturer's protocol. Samples with an RNA integrity number (RIN) value of >7 were used for library preparation. RNASeq libraries were prepared from total RNA using the Ultra II RNA Sample Prep Kit (New England BioLabs, Ipswitch, MA, USA) according to the manufacturer's protocol. Briefly, poly-A RNAs were captured with oligo-dT conjugated magnetic beads then the mRNAs were eluted and fragmented at 94 °C. First-strand cDNA was generated by random priming reverse transcription, and the double-stranded cDNA was generated after the second strand synthesis step. After repairing ends, A-tailing, and adapter ligation steps, adapter-ligated fragments were amplified by enrichment PCR, and finally, sequencing libraries were generated. Sequencing runs were executed on an Illumina

NextSeq500 instrument using single-end 75 sequencing cycles, generating an average of 20 million raw reads per sample.

RNAseq Data Analysis

Raw sequencing data (fastq) were aligned to the chicken reference genome version GRCg6a using the HISAT2 algorithm, and BAM files were generated. Downstream analysis was performed using the StrandNGS software (www.strand-ngs.com). BAM files were imported into the software, and the DESeq algorithm was used for normalisation. To identify differentially expressed genes (DEGs) between conditions, an ANOVA test with Benjamini-Hochberg FDR for multiple testing correction and Tukey post hoc test was used. 6 clusters were created using the K-means algorithm. The entire dataset was deposited and published in the BioProject database (http://www.ncbi.nlm.nih.gov/bioproject/). BioProject ID: PRJNA817177. For each time point and cluster, entities were further analysed using Cytoscape (21) to identify key driving molecules in the networks, and network visualisation.

Pathway Analyses

Pathway enrichment analysis was performed on the differentially expressed genes (DEGs) between consecutive culturing days, as identified by RNASeq using the PANTHER (http://pantherdb.org) and STRING (https://string-db.org) databases to identify significantly enriched pathways. The output was restricted to pathways with a multiple testing p-value <0.05 and an overlap of at least two genes between the pathway term gene list and differentially expressed gene list. The list of TFs was retrieved using a GO term-based search (GO:0003700: 'transcription factor activity, sequence-specific DNA binding'). Genes involved in chondrogenesis, cartilage development, and homeostasis were retrieved from the following GO terms: GO:1990079 cartilage homeostasis, GO:0051216 cartilage development, GO:0060536

cartilage morphogenesis, GO:0001502 cartilage condensation, and GO:0061975 articular cartilage development.

The differentially expressed genes determined by RNASeq analysis, calculated based on fold change (cut-off value at 2) between successive culturing days, were used for the Ingenuity Pathway Analysis (IPA) software (Qiagen, USA). The core analysis module in IPA was used to decipher differentially regulated canonical pathways, upstream regulators, diseases and biofunctions, and novel gene networks based on Fisher's exact test (*p*-value cut-off at 0.05) (22,23).

Weighted Gene Correlation Network Analysis

Co-expression data were analysed in R (4.0) using the Weighted Gene Correlation Network Analysis (WGCNA) tool (24). Our workflow followed that of Huynh *et al.* (6). The trait data *SOX9*, *ACAN*, and *COL2A* were derived from our transcriptomic datasets. The single highest correlating module ("blue") with approximately 2000 genes was taken forward for further analysis. Edge data were exported to Cytoscape (21) for visualisation of the network.

Statistical Analysis

Statistical analysis was performed using the built-in statistical module of StandNGS. To identify DEGs between conditions (time points), an ANOVA test with Benjamini-Hochberg false discovery rate (FDR) for multiple testing correction and a Tukey post hoc test was used. The *p*-value cut-off to determine statistical significance was 0.05.

Results

1. Global Gene Expression Analysis in the Early Stages of Chondrogenesis

To identify the genes involved in the early stages of chondrogenesis in limb bud-derived LMPs, we performed a global transcriptomic analysis on micromass cultures at six different time points, representing the key stages of *in vitro* chondrogenesis (see the workflow in Fig. 1). The chondrogenic differentiation of the micromass cultures was confirmed by the increased metachromatic staining pattern of the cartilage ECM (Fig. 2).

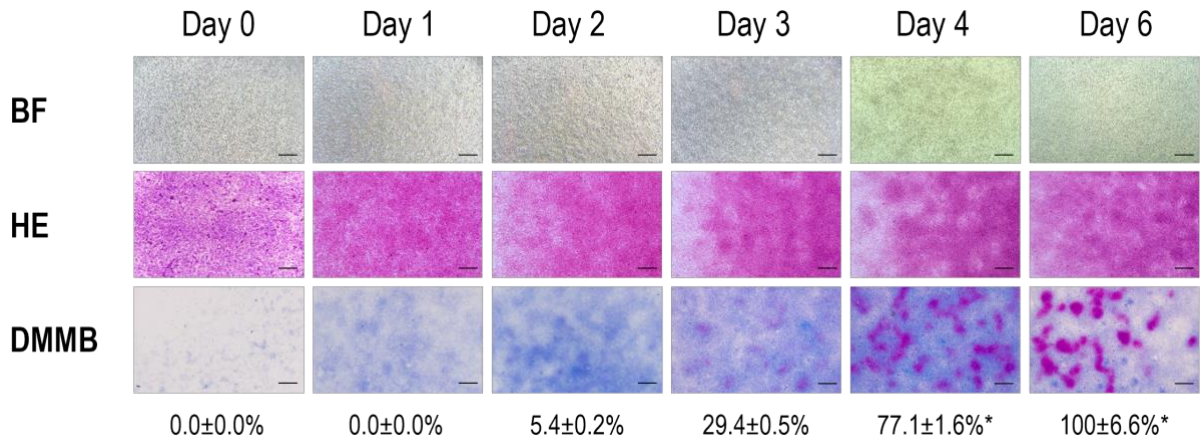

Figure 2. Light microscopy analysis of micromass cultures during the course of chondrogenic differentiation. Photomicrographs of unstained (BF, brightfield), haematoxylin-eosin (HE) stained and dimethyl methylene blue (DMMB) stained cultures are shown. Original magnification was 10×. Scale bar, 200 μm. Values below images of DMMB stained cultures reflect the results obtained using a MATLAB-based image analysis of metachromatic areas. Data are expressed as mean ± SEM, compared to day 6 (100%). Statistical significance (*$p < 0.05$) between subsequent culturing days is indicated by asterisks.
Representative data out of 3 independent biological replicates.

We performed principal component analysis (PCA) of the normalised RNASeq data to explore their interrelationships and to visualise the correlation between the samples (Fig. 3). All biological replicates were clustered closely together at the different time points, showing marked changes in the transcriptome over time in culture as the chondrogenic cells underwent spontaneous chondrogenesis in the micromass cultures. The first component (PC1) on the *x*-axis accounted for 63% of the variance. Day 0 samples containing undifferentiated progenitor cells clustered separately from the other five sample groups in the first principal component.

The transcriptomic changes during this transition period are at least partially influenced by the cell isolation procedure. The second component (PC2) on the *y*-axis explained 16% of the variance. In the second principal component, samples obtained on day 6 (i.e., mature chondrocytes) clustered separately from the other sample groups. There was an overall trend from less mature (undifferentiated) cells at day 1 to mature (differentiated) micromass cultures at day 6, indicating a progressively different transcriptome during chondrogenesis. However, a much closer relationship between day 1 and day 6 time points was observed in the PC1 hyperplane. The slightly greater scatter of day 6 samples along the *y*-axis may indicate that the biological replicates are slightly out of phase at this stage of chondrogenic differentiation, resulting in more inter-culture and inter-cellular variability. However, the PCA analysis demonstrated that the variance between time points was still much greater than the variance between biological replicates.

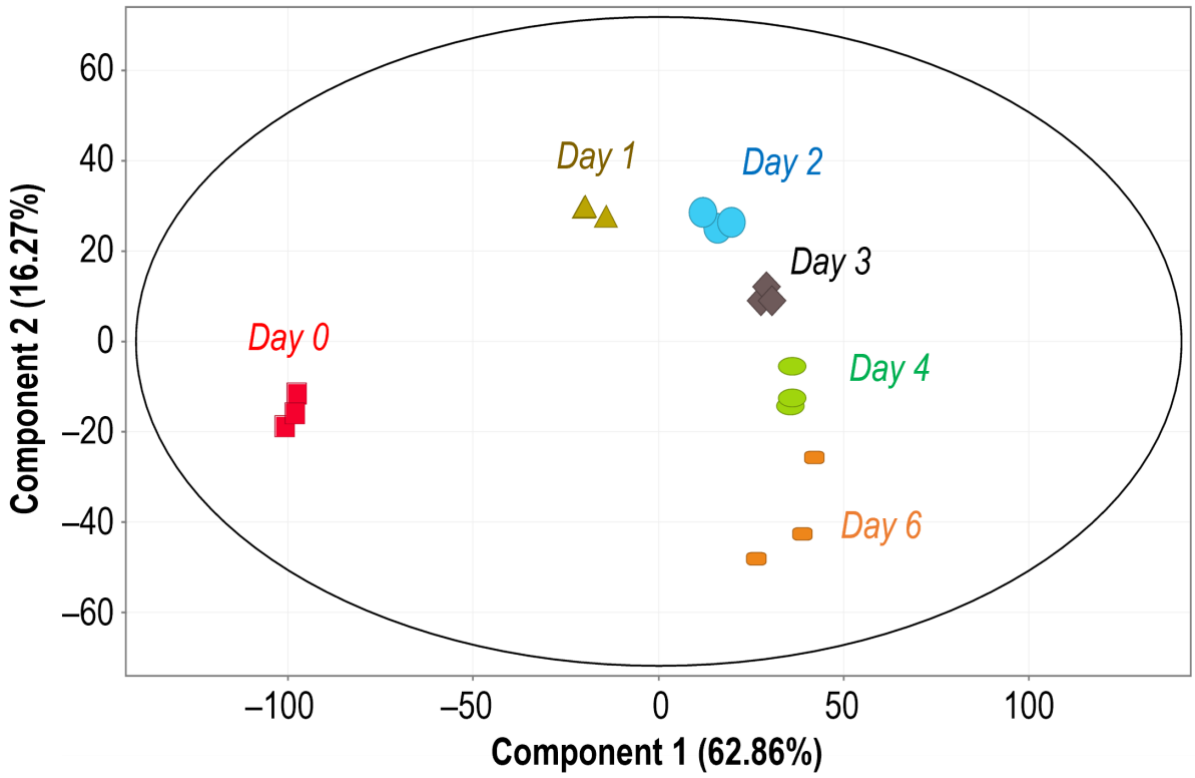

Figure 3. Principal component analysis was performed on the normalised logarithmic transformed read counts. The first component explains 63% of variability, and the second component explains 16%. Each point represents an experimental sample; colours and shapes indicate different time points. Samples were closely related by time points (culturing days), and there was minor variation between biological replicates.

Hierarchical clustering based on the global transcriptome expressed in chondrogenic micromass cultures also revealed the segregation of samples primarily according to the time points (Fig. 4A). The transcriptomic signatures in samples from the early stages of chondrogenesis (day 0, 1, and 2) clustered effectively both between the biological replicates and between the different time points. These samples possessed the very early (days 0 and 1) and early modulated genes (day 2). 6-day-old micromass cultures with mature chondrocytes expressing the late modulated genes also clustered together. However, the 3- and 4-day-old colonies predominantly containing differentiating chondrocytes were not characterised by distinct transcriptomic profiles and clustered together, indicating that differentiating chondroblasts and chondrocytes shared a very similar transcriptome on these two culturing days. When we compared all the time points to day 0, the fold changes (both up- and downregulated entities) gradually increased (Fig. 4B), indicating that the transcriptome of more differentiated cultures became gradually more distinct compared to undifferentiated (day 0) colonies.

2. The Top 20 Most Abundantly Expressed and Differentially Expressed Genes on Each Culturing Day Include Transcripts Known to be Involved in Chondrogenic Differentiation

The top 20 most abundantly expressed genes for each culturing day are listed in the Supporting information (Table S1). Genes involved in chondrogenesis and cartilage ECM synthesis (*COL1A1*, *COL1A2*, *COL2A1*, *COL9A1*, *COL11A1*, *FN1*, *ACAN*), cytoskeleton (*ACTB*, *ACTC2L*, *TUBA1A1*, *TUBB*), ribosomal proteins, and RNAs (*NCL*, *RPL4*, *RPS2*, *RPS3A*, *RPS6*, *RPSA*), glycolysis and energy metabolism (*GAPDH*, *ENO1*), transcription factors (*YBX1*), and lncRNA transcripts (LOC101750014, LOC112530942) were enriched in the lists for each culturing day.

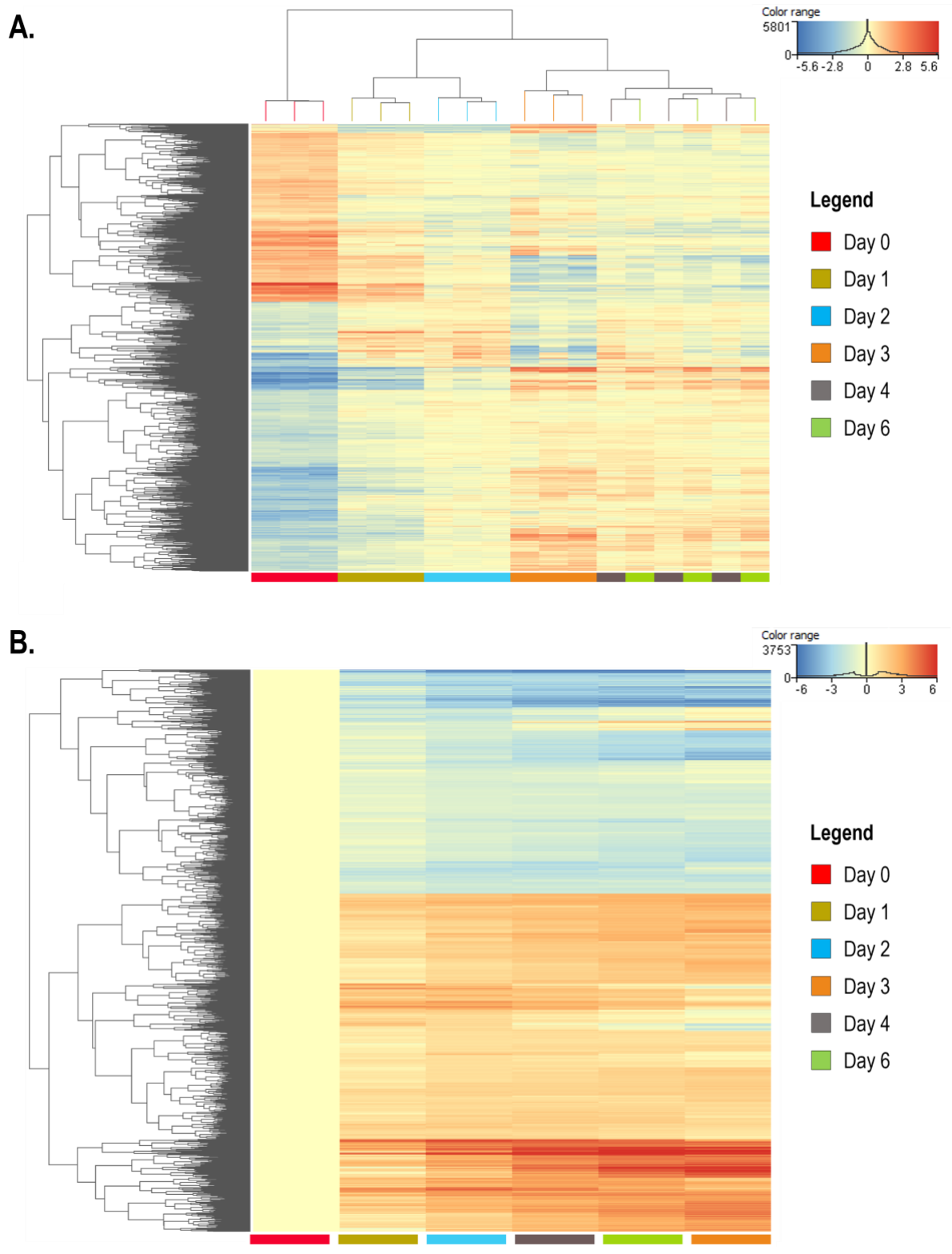

Figure 4. **A.** Hierarchical clustering of the global transcriptome of chondrogenic cells during the differentiation programme (days 0-6) following normalisation and ANOVA analysis. Data for all 3 biological replicates are shown. Whilst most time points clustered separately (days 0, 1, 2, and 6), days 3 and 4 clustered together, indicating a similar transcriptome at these time points. **B.** Heatmap showing differentially expressed genes (DEGs) normalised to day 0. Average expression between biological replicates were used for calculating FC values. A gradual increase in up- and downregulation was observed during the course of differentiation.

Table 1. **A.** Summary of differentially expressed genes (DEGs) by comparing consecutive time points during chondrogenesis. Total and unique entities (and their percentage) between pairwise comparisons, and the proportion of upregulated DEGs are shown. **B.** An overview of the number of down- and upregulated genes between consecutive time points, along with total and unique DEGs (and their percentage) between pairwise comparisons.

A.

Comparison	Total entities	Unique entities	%	Upregulated DEGs
Day 0 vs. 1	1876	1636	87.21%	67.64%
Day 1 vs. 2	267	27	10.11%	64.42%
Day 1 vs. 2	267	212	79.40%	64.42%
Day 2 vs. 3	151	96	63.58%	73.51%
Day 2 vs. 3	151	126	83.44%	73.51%
Day 3 vs. 4	96	71	73.96%	50.00%
Day 3 vs. 4	96	58	60.42%	50.00%
Day 4 vs. 6	240	202	84.17%	37.50%

B.

Comparison	Total downregulated entities	Unique downregulated entities	%
Day 0 vs. 1	607	555	91.43%
Day 1 vs. 2	95	43	45.26%
Day 1 vs. 2	95	82	86.32%
Day 2 vs. 3	40	27	67.50%
Day 2 vs. 3	40	34	85.00%
Day 3 vs. 4	48	42	87.50%
Day 3 vs. 4	48	28	58.33%
Day 4 vs. 6	150	130	86.67%

Comparison	Total upregulated entities	Unique upregulated entities	%
Day 0 vs. 1	1269	1129	88.97%
Day 1 vs. 2	172	32	18.60%
Day 1 vs. 2	172	134	77.91%
Day 2 vs. 3	111	73	65.77%
Day 2 vs. 3	111	92	82.88%
Day 3 vs. 4	48	29	60.42%
Day 3 vs. 4	48	32	66.67%
Day 4 vs. 6	90	74	82.22%

Many genes were significantly differentially regulated, as was revealed by comparing the transcriptomic profiles of consecutive culturing days. An overview of the downregulated and upregulated genes for each comparison, and the genes that were common between

consecutive days, is shown in Table 1. The top 20 DEGs (fold change cut-off: 2.0) obtained from a pairwise comparison between consecutive culturing days (i.e., day 0 *vs*. 1, day 1 *vs*. 2, etc.), in terms of adjusted *p*-value, are shown in Table 2. Full lists of DEGs between consecutive culturing days are shown in the Supporting information (Supplemental data S1 – *full-list-of-DEGs.xlsx*). The number of DEGs gradually declined between pairwise comparisons of consecutive culturing days, indicating a very high abundance of very early modulated genes in the transcriptome. For example, of the genes that were differentially regulated between days 0 and 1 *vs*. days 1 and 2, 240 were common between the two comparisons. There were fewer common DEGs between the pairwise comparisons of the later stages of chondrogenesis, but the number of late-modulated genes increased again in mature chondrocytes (day 6).

Table 2. Top 20 significant DEGs in pairwise comparisons between different time points during *in vitro* chondrogenesis, ranked according to the adjusted *p*-value. Fold change (FC) cut-off: 2.0. Gene name in **bold** have a known role during cartilage development (GO:0051216).

Gene Symbol	*p* ([0] vs [1])	FC ([0] vs [1])	Regulation ([0] vs [1])
ELOVL4	0.001	-2.76	down
ADAMTS12	**0.001**	**2.01**	**up**
GPRC5B	0.001	3.93	up
RDH10	0.001	4.44	up
CYR61	0.001	-3.00	down
PPCDC	0.001	2.10	up
SLC25A48	0.001	2.41	up
MGAM	0.001	-3.53	down
DUT	0.001	-2.08	down
LOC420807	0.001	2.04	up
LOC107055643	0.001	4.72	up
LOC101747310	0.001	5.29	up
FHL5	0.001	3.31	up
DMD	0.001	-2.01	down
ATP6V0C	0.001	-2.00	down
DSEL	0.001	-2.20	down
GAREM2	0.001	3.79	up
PCSK1	0.001	-4.85	down
LOC421415	0.001	-3.75	down
LOC107054051	0.001	2.48	up

Gene Symbol	*p* ([1] vs [2])	FC ([1] vs [2])	Regulation ([1] vs [2])
CSRP2	0.001	2.14	up
TFAP2C	0.001	-3.74	down
RGR	0.001	-5.20	down
ITIH3	0.001	-7.01	down
LAMB2	0.001	2.41	up
TMEM72	0.001	-4.69	down
VWDE	0.001	-2.22	down
CHD7	0.001	-2.99	down
P2RY10	0.001	-2.32	down
COL14A1	0.001	5.02	up
LOC107049082	0.001	-3.45	down
FKBP10	0.002	1.70	up
BTBD6	0.002	2.12	up
LRRC56	0.002	1.60	up
RSPO3	0.002	2.89	up
LOC771545	0.002	2.10	up
IL16	0.002	3.02	up
SULT1E1	0.002	2.35	up
LRFN5	0.002	-2.97	down
CLDN8	0.002	3.90	up

Gene Symbol	*p* ([2] vs [3])	FC ([2] vs [3])	Regulation ([2] vs [3])
ASPN	0.001	5.39	up
GCHFR	0.001	2.36	up
CAPN9	0.001	3.98	up
LRRC56	0.001	1.62	up
LIN28B	0.002	-2.29	down
PTPRVP	0.002	1.98	up
COL16A1	0.002	1.84	up
TMEM132B	0.002	2.62	up
LOC112530970	0.002	1.81	up
GLT8D2	0.002	2.05	up
COBL	0.002	4.86	up
CLEC3A	0.002	19.14	up
CCDC3	0.002	-2.56	down
FAM196A	0.003	2.28	up
CA12	0.003	8.96	up
ARHGEF28	0.003	2.59	up
IGSF10	0.003	2.32	up
LOC112530488	0.003	-1.5	down
SOSTDC1	0.003	1.90	up
CHIA-M31	0.003	5.43	up

Gene Symbol	*p* ([3] vs [4])	FC ([3] vs [4])	Regulation ([3] vs [4])
CHAD	0.001	4.19	up
BLM	0.002	-1.50	down

RFC3	0.002	-1.54	down
SH3BGRL2	0.002	-2.18	down
PLA2G4B	0.002	1.85	up
LOC426385	0.003	-1.46	down
CR1L	0.003	-1.50	down
RSPH3	0.003	-2.02	down
TRPA1	0.003	-3.77	down
TACR1	0.003	-6.97	down
STOX1	0.003	-2.04	down
CA12	0.004	8.24	up
EBF1	0.004	-1.45	down
SV2C	0.0044	-2.47	down
HBBA	0.005	13.06	up
FAM210B	0.005	-1.39	down
LTK	0.007	-2.41	down
PATJ	0.007	-1.63	down
KIF4A	0.007	-1.60	down
RPH3A	0.008	2.24	up

Gene Symbol	***p* ([4] vs [6])**	**FC ([4] vs [6])**	**Regulation ([4] vs [6])**
ST18	0.001	2.80	up
SMC2	0.001	-1.99	down
LRRTM3	0.001	-3.57	down
LOC107057363	0.001	6.19	up
INHBA	0.001	2.32	up
CBLN4	0.001	-4.88	down
KCND3	0.001	5.48	up
SCN9A	0.001	-3.59	down
LOC112530488	0.001	-1.56	down
WNK4	0.002	-2.20	down
AVD	0.002	6.35	up
LOC101749287	0.002	-2.53	down
FOS	0.002	2.62	up
EHD2	0.002	2.02	up
EMILIN3	0.002	-2.03	down
NCKAP5	0.002	-2.85	down
EMB	0.002	-1.92	down
ASB1	0.002	-1.86	down
ECM2	0.002	1.61	up
IL15	0.002	3.80	up

Analysis of the DEGs at different time points compared to the subsequent ones revealed a higher number of genes upregulated at days 0 *vs*. 1, 1 *vs*. 2, and 2 *vs*. 3 (67.6%, 64.4% and 73.5%, respectively). There was an equal proportion of down- and upregulated DEGs at day 3

vs. 4, and a higher number of genes were down-regulated at day 4 *vs*. 6 (62.5% of all DEGs at day 6) (Tables 1 and 2).

Next, we looked at the DEGs in more detail between each comparison and checked those genes that were differentially expressed (down- or upregulated) in two sequential comparisons (Supporting information, Tables S2 and S3). Of the 940 downregulated entities, 91 (9.7%) were common between the two consecutive comparisons. There was a similar proportion (12.6%) of the common upregulated entities (213 entities of 1690 upregulated DEGs). The full lists of common DEGs between each consecutive culturing day are shown in the Supporting Information (Supplemental data S1 – *full-list-of-DEGs.xlsx*).

We exported the top 1000 most abundant transcripts at each time point to Cytoscape. The analysis results (including nodes, edges, node degrees, clustering coefficients) are summarised in Table 3. Interestingly, day 0 was characterised by the highest number of edges (connections) followed by a steady decline in this parameter during chondrogenic differentiation. On each culturing day, transcripts with chaperone (*CCT5*, *CCT2*), ribosomal (*RPS27A*, *RPS3*, *RPL4*, *RPS16*, *RPSA*, *RPS2*, *RPLP0*), translational (*EIF2S1*, *EIF2*), cytoskeletal (*ACTB*), ubiquitination (*UBB*), and glycolytic (*GAPDH*) functions had the highest number of connections and closeness centrality values. While in the chondroprogenitor phase (days 0–2), chaperons or ribosomal proteins were ranked with the highest numbers of connections, from day 3, GAPDH had, by far, the highest closeness centrality value.

Table 3. The overall analysis of nodes and edges of networks amongst the top 1000 most abundant transcripts at each time point using Cytoscape.

	Day 0	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 6**
number of nodes (transcripts/proteins):	892	890	868	880	873	861
number of edges (connections):	21.888	20.044	19.324	18.299	17.234	16.623
average node degree:	49.1	45	44.5	41.6	39.5	38.6
avg. local clustering coefficient:	0.39	0.398	0.411	0.406	0.404	0.401

3. GO and IPA Analysis of the Most Significant DEGs Revealed Pathways Related to Chondrogenic Differentiation

By analysing the most abundant transcripts, as well as the most highly regulated DEGs between the different time points, we identified genes that were already known to be involved in chondrogenesis and/or expressed in chondrocytes according to the literature (see Table 2, and Tables S1, S2, S3). Using the Gene Ontology (GO) molecular function and biological pathway tools of the Panther classification system, we first analysed the top 20 most abundant transcripts for their GO terms at each time point (ranked according to FDR; $p < 0.05$). The subset of the most abundant genes on days 0 and 1, mainly containing undifferentiated chondrogenic LMPs, were enriched with general terms such as "regulation of gene expression," "regulation of protein metabolic process," "regulation of translation" and "nuclear-transcribed mRNA catabolic process." The major enriched pathways on day 2 included "cellular response to growth factor stimulus" and "ossification." However, from day 3, pathways related to chondrogenic differentiation, cartilage ECM production and organisation, skeletal system development and endochondral ossification were significantly over-represented, with notable terms such as "cartilage development involved in endochondral bone morphogenesis," "endochondral bone morphogenesis," "collagen and proteoglycan metabolic process," "biomineral tissue development," "skeletal system morphogenesis," "extracellular matrix organization," and "skeletal system development," this was especially true for more mature cultures (days 4 and 6). For details, see the Supporting Information (Supplemental Table S4).

When we observed the subsets of genes significantly differentially regulated between the different time points, the down-regulated genes in these comparisons were enriched with the following GO biological pathways: "Wnt signalling pathway," "TGF-beta signalling pathway," "cadherin signalling pathway," and "p38 MAPK pathway." The upregulated DEGs in these comparisons were enriched with terms such as "cytoskeletal regulation by Rho

GTPase", "integrin signalling pathway" or "interleukin signalling pathway." However, caution should be exercised when relying on these GO term lists, as many of these terms were associated with both the upregulated and downregulated genes. For details, see the Supporting information (Supplemental Table S5).

We submitted pairwise comparisons of the entire dataset for the Qiagen IPA analysis and examined the significant non-directional networks. The over-represented networks are summarised in the Supporting Information (Supplemental Table S6). When we compared the early time points (day 0 vs. day 1), the following networks were identified: "Cell Cycle, DNA Replication, Recombination, and Repair, Endocrine System Disorders," "Carbohydrate Metabolism, Small Molecule Biochemistry, Visual System Development and Function," and "Cell-To-Cell Signaling and Interaction, Developmental Disorder, Hereditary Disorder". We also found that the "Embryonic Development, Organismal Development, Tissue Morphology;" and the "Cellular Development, Connective Tissue Development and Function, Tissue Development" networks were over-represented in the gene list in this comparison.

When comparing the DEGs between days 1 and 2, the following networks were over-represented: "Molecular Transport, RNA Post-Transcriptional Modification, RNA Trafficking," "Cellular Assembly and Organization, Cellular Function and Maintenance, Molecular Transport," "Connective Tissue Disorders, Developmental Disorder, Hereditary Disorder," "Cell Morphology, Cellular Assembly and Organization, Cellular Function and Maintenance," and "Cancer, Cellular Development, Tissue Development".

In addition to the more generic terms, the terms "Connective Tissue Disorders, Protein Synthesis, RNA Post-Transcriptional Modification," "Connective Tissue Disorders, Developmental Disorder, Hereditary Disorder," and "Cell Cycle, Connective Tissue Development and Function, DNA Replication, Recombination, and Repair," which are more closely related to chondrogenesis, were identified in the day 2 *vs*. day 3, day 3 *vs*. day 4, and

day 4 *vs*. day 6 comparisons. The term "Cell Signalling, Cellular Assembly and Organization, Post-Translational Modification" was only present in the comparison of the last two time points, likely indicating that the genes involved in mediating intracellular signalling pathways were enriched.

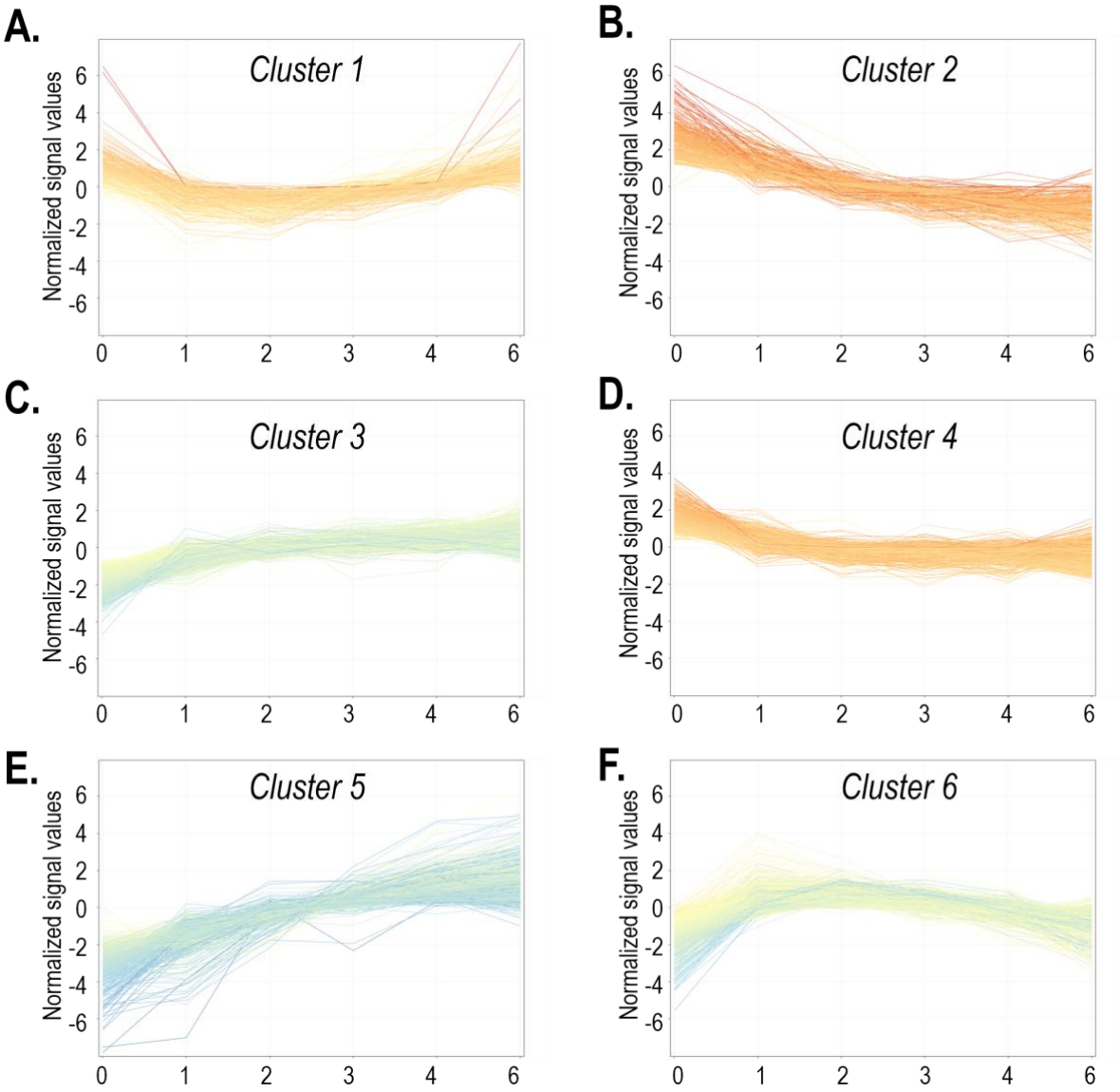

Figure 5. Unsupervised clustering analysis of genes based on their normalised expression values and expression patterns using the K-means algorithm defined 6 groups of genes. The expression dynamics of each cluster are visible in panels A-F (clusters 1–6, respectively). *y*-axis, time points (days of culturing)

4. Unsupervised Clustering Analysis

To identify clusters of genes according to their expression patterns, six groups were defined using the K-means algorithm (Fig. 5). The gene lists for each cluster are available in

the Supporting Information (Supplemental Data S2 – *clusters.xlsx;* Table S7). A detailed analysis of each cluster using Cytoscape (nodes, edges, node degrees, clustering coefficients) is shown in Table 4. Cluster 4 contained significantly more edges (4038) than the other clusters, and the average node degree was 10.1. Cluster 1 comprised 251 genes, which displayed higher transcript levels at the early (day 0) and the late (day 6) time points, and lower expression levels in between. A total of 201 entities were assigned a GO term. Most of the entities were involved in signalling pathways (the following GO terms were over-represented: "Regulation of Protein Phosphorylation" (15/201); "G Protein-Coupled Receptor Signaling Pathway" (15/201); "Cell Surface Receptor Signaling Pathway" (27/201); and "Signal Transduction" (49/201). Entities with the highest numbers of edges and closeness centrality values include the transcription factor subunits *FOS* and *FOSL2*, several growth factors and their receptors (*HGF*, *FGF14*, *FLT1*, *HBEGF*, *PDGFB*, *PGF*), and the matrix metalloproteinases *MMP1* and *MMP27*. This cluster also contains the forkhead transcription factors *FOXE1* and *FOXS1*, which are known to be expressed in chondrocytes.

Table 4. The overall analysis of nodes and edges of networks in each unsupervised cluster of transcripts using Cytoscape.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
number of nodes (transcripts/proteins):	213	263	966	803	321	463
number of edges (connections):	209	416	2349	4038	533	991
average node degree:	1.96	3.16	4.86	10.1	3.32	4.28
avg. local clustering coefficient:	0.368	0.383	0.254	0.321	0.323	0.32

Cluster 2 contained 308 genes, of which 271 were annotated to a GO term. Initially showing high expression levels, the genes in this cluster were characterised by a trend of downregulation towards the later time points. The following general GO terms were over-represented in this cluster: "anatomical structure development" (59/271), "cell differentiation" (33/271), "regulation of gene expression" (61/271), and "cellular macromolecule biosynthetic process" (57/271). Transcripts with the highest number of edges and closeness centrality values

include the morphogens *SHH*, *WNT3A*, *WNT6*, *WNT7A*, *WNT9B*; the transcription factors *MYOD1*, *PAX3*, *PAX7*, *OLIG2*, *MYF5*, *GBX2*, *TCF7*, *TFAP2A*; and protein kinases (*MAPK11*, *PLK4*, *TIE1*, *EPHA4*). Of the genes relevant to chondrocytes, *ADAM20*, *COL2A1*, *HOX*, and *WNT* entities were present in this cluster.

In contrast, the genes that displayed increasing levels over time with the highest expression on day 6 were classified in cluster 3. There are 1254 entities in this cluster, making it the largest group. A total of 953 genes were assigned a GO term, of which the "extracellular matrix organization" (22/953), "ion transport" (51/953), "signaling" (139/953), and "cellular macromolecule metabolic process" (152/953) terms were significantly enriched. This group of genes contained the SOX trio (*SOX5*, *SOX6*, and *SOX9*), RUNX transcription factors (*RUNX1*, *2* and *3*), GDF (*GDF3*, *6*, and *15*), ADAMTS (e.g., *ADAMTS2*, *6*, *9*), SIRT genes (*SIRT4* and *5*), and several genes encoding the alpha chains of various collagens (e.g., *COL3A1*, *COL4A3*, *COL5A1*). Overall, this cluster contained the highest number of genes (29) with a known function in chondrogenesis according to the cartilage development GO:0051216 list. The following entities have the highest number of edges and closeness centrality values in this group: *FN1* (fibronectin); growth factors and receptors (*EGF* and *EGFR*, *FGF7* and *FGFR2*, *FGFR3*, *FGFR4*, *TGFB2*); kinases (*PIK3R1*, *PIK3R2*, *PTK2B*); key transcription factors (*SOX9*, *RUNX2*); and cartilage ECM components (*COL5A1*, *COL5A2*, *ADAMTS2*).

Genes in cluster 4 were characterised by an expression pattern similar to those in cluster 2; those were the only genes that generally showed a more moderate trend of downregulation (they reached a plateau after day 2). There were 907 entities in this cluster, 769 of which were assigned a GO category. The following GO terms were over-represented in this cluster: "cell cycle" (46/769), "cell differentiation" (59/769), "developmental process" (88/769), "regulation of cellular metabolic process" (170/769), and "cellular metabolic process (271/769)". This cluster includes the following genes relevant to chondrogenic differentiation: members of the

forkhead box (FOX) transcription factor family (*FOXG1*, *FOXJ1*, *FOXK2*, *FOXN1*, *FOXO6*), insulin-like growth factor-1 receptor (*IGF1R*), S100 genes (*S100A14* and *S100Z*), and some of the SOX genes (*SOX14* and *18*). As described earlier, this cluster contains the highest number of edges; key nodes with the highest closeness centrality values in this group are as follows: cell cycle regulation (*CDK1*, *PLK1*, *CDC20*, *MAD2L1*, *CCNA2*, *PCNA*); kinases (*TTK*, *CHEK1*, *PIK3CA*); transcription factors (*HMGB2*, *E2F3*); and integrins (*ITGB3*, *ITGA9*).

The entities in cluster 5 displayed a robust pattern of upregulation. Overall, these genes were characterised with a similar trend to that seen in cluster 3, but with larger amplitude changes. Of the 386 genes in this cluster, 322 were assigned a GO biological process term. The following GO terms were enriched in this cluster: “extracellular matrix organization” (19/332), “macromolecule metabolic process” (53/332), and “gene expression” (26/332). Genes encoding cartilage ECM-specific collagens (*COL2A1*, *COL6A1*, *COL9A1*, *COL11A1*), other minor collagens (*COL8A1*, *COL12A1*, *COL15A1*, *COL16A1*, *COL20A1*, *COL22A1*, *COL27A1*), matrilin (*MATN1* and *MATN4*), FOX genes (*FOXA1*, *FOXC1*, *FOXD1*, *FOXD2*), interleukins and their receptors (*IL1R1*, *IL12RB1*, *IL13RA1*, *IL15*, *IL16*, *IL22RA2*, and *IL2RB*) were also present in this cluster. The very high number of entities from the GO:0051216 list (26) in this group was second only to cluster 4. Entities with the highest known connections and closeness centrality values included the key matrix components *COL6A1*, *COL6A3*, *ACAN*, *COL2A1*, *COL11A1*, *COL9A3*, *COL8A1*; *ADIPOQ*; transporters (*SLCO1B1*, *SLC51A*, *SLC22A18*, *SLC10A2*); and *IHH*.

The last cluster (6) contains genes that are characterised by a low expression profile on day 0, then display the highest expression on days 1, 2, or 3, and then show a general decline from there. This group contains genes that likely play key roles in chondrogenic differentiation. Of the 572 transcripts in this cluster, 428 were assigned a GO biological process term. The significantly enriched GO terms in this list include: “regulation of canonical Wnt signaling

pathway" (8/428), "ion transmembrane transport" (30/428), "cell communication" (73/428), macromolecule metabolic process" (62/428). The known chondrogenesis-related genes in this cluster included genes encoding bone morphogenetic protein receptor type 1B (*BMPR1B*), collagen alpha subunits (*COL4A4*, *COL25A1*, *COL26A1*), growth differentiation factor 5 (*GDF5*), homeobox D1 (*HOXD1*), and Wnt family member 16 (*WNT16*). Among these, the following entities were characterised by the highest number of known connections and closeness centrality values: cytoskeletal components (*UNC45B*, *MYL1*, *ACTC1*, *ACTN2*, *MYH1D*); as well as ion channels and receptors (*CACNG1*, *CHRNG*, *CHRNB4*, *KCNQ1*, *SNC3B*).

5. The Temporal Expression Pattern of Chondrogenic Genes and Collagens During Chondrogenic Differentiation

We performed a targeted *in silico* analysis of the NGS data to determine the profiles of common chondrocyte-specific markers in chondrifying cultures. This list was complemented with the contents of the following GO categories: GO:1990079 (cartilage homeostasis); GO:0051216 (cartilage development); GO:0060536 (cartilage morphogenesis); GO:0001502 (cartilage condensation); GO:0061975 (articular cartilage development) (see Supporting Information, Supplemental Data S3 – *chondrogenic-subset.xlsx;* Fig. 6A). As expected, the "classical" hyaline chondrocyte marker genes showed different temporal patterns between the time points. *COL2A1* had by far the highest normalised expression level by day 6 (almost 350,000), following a steady increase along with chondrogenesis (see cluster 5). The same temporal expression pattern was true for some of the other ECM constituents such as collagen alpha chains (*COL1A1*, *COL9A1*, and *COL11A1*), *PRG4*, *ACAN*, *MATN1*, and *CHST11*. Other highly abundant (>3,000) transcripts included *FGFR3*, *HAPLN1*, *NCAM1*, *NFATC1*, *3*, and *5*, *SOX9*, and *SMAD6*. In contrast, some of the "classical" cartilage markers, such as the growth

differentiation factor (GDF) genes, generally showed low (<1,000) normalised expression levels. Most of the genes in this group displayed a trend of upregulation at later time points, but some other key factors (e.g., *BMP7*, *GDF11*, *NCAM1*, *PAX7*, *SMAD1*, and *WNT7A*) showed the opposite trend.

Since various collagen types have been described as major or accessory components of cartilage ECM, we were curious as to which collagen-coding transcripts were expressed in chondrifying micromass cultures (Table 5, Fig. 6B). We detected genes coding for the alpha chains of all known collagen types (I–XXVIII) except for type XXIX (epidermal collagen). While the genes coding for collagen types displayed variable expression patterns, the majority were grouped in cluster 5, indicating an obvious trend for upregulation. As expected, *COL2A1* had the largest overall (~119,000) and absolute (349,000) expression values. In addition to *COL1A1*, transcripts of the "classic" cartilage-specific collagens (*COL9A2*, *COL11A1*, and *COL6A3*) also had high relative normalised expression. The genes coding for minor but still abundant (<10,000 relative expression level) collagen types included *COL5A1*, *COL8A1*, *COL12A1*, *COL14A1*, *COL16A1*, and *COL27A1*. Other collagens such as *COL4A5*, *COL4A6*, *COL17A1*, and *COL18A1* displayed a clear downregulation trend towards the late time points, whereas *COL25A1* and *COL26A1* exhibited a peak-like expression pattern (Fig. 6B).

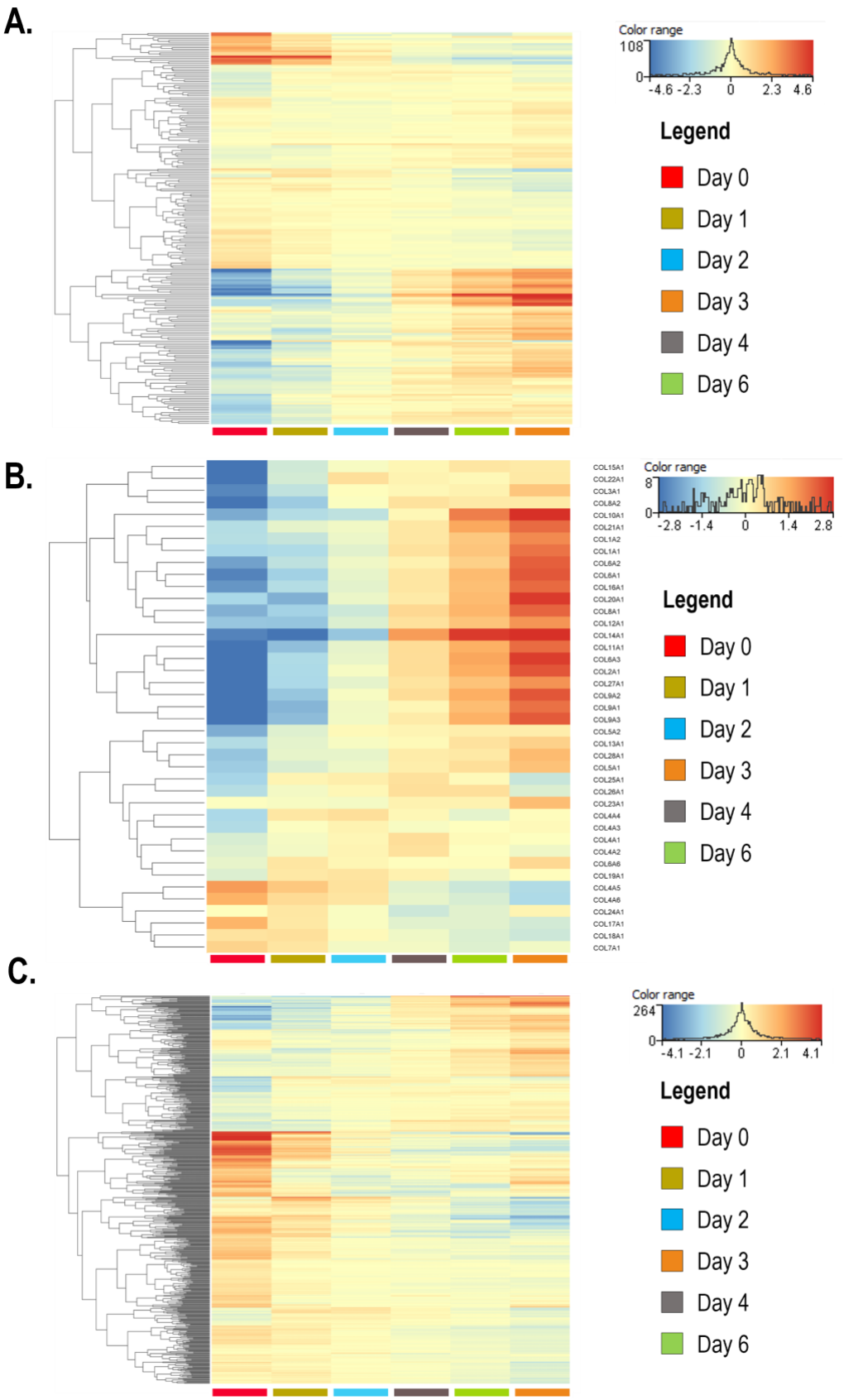

Figure 6. Hierarchical clustering of the (**A**) chondrogenic subset, (**B**) collagens, and (**C**) transcription factors in differentiating cells of micromass cultures undergoing chondrogenesis during days 0-6. Average values for the 3 biological replicates are shown.

Table 5. List of collagen transcripts (normalised expression values) in chondrogenic cultures

Gene symbol	Day 0	Day 1	Day 2	Day 3	Day 4	Day 6
COL1A1	15760.14	16783.08	29794.53	59374.47	83813.15	175109.52
COL1A2	16108.15	19230.82	30259.20	49691.93	72133.11	130630.48
COL2A1	**7442.02**	**25822.98**	**60104.40**	**99841.36**	**173157.34**	**349043.66**
COL3A1	680.82	1931.62	4099.06	4363.38	4647.22	7878.62
COL4A1	1296.68	1851.30	2285.26	3052.18	2180.60	2030.95
COL4A2	909.48	1159.94	1491.09	1891.14	1288.80	1098.22
COL4A3	16.68	39.59	46.98	37.56	41.02	43.34
COL4A4	6.75	23.24	25.60	19.19	12.87	18.11
COL4A5	2431.30	1542.43	1124.61	568.59	440.03	326.30
COL4A6	1491.55	917.59	813.32	453.84	378.91	235.23
COL5A1	4135.03	8000.05	10146.98	15139.36	18244.42	25692.97
COL5A2	4762.46	13143.21	20539.81	21653.75	23396.74	23675.59
COL6A1	489.46	1040.84	2251.97	3978.70	6689.59	15310.95
COL6A2	483.04	1032.32	1820.74	2571.81	3833.58	10601.71
COL6A3	647.34	1853.53	3641.74	6918.79	13034.94	29297.54
COL6A6	6.72	11.81	9.82	8.93	9.07	13.92
COL7A1	524.43	422.68	271.33	273.87	212.58	268.27
COL8A1	118.38	163.72	258.54	730.14	1100.11	2163.65
COL8A2	959.09	2470.48	7143.45	10503.18	9497.00	9553.54
COL9A1	519.88	3192.94	10107.95	13207.23	26196.69	50172.74
COL9A2	856.66	3702.70	10737.15	17682.17	30427.69	61044.56
COL9A3	227.30	2136.43	6379.65	9406.09	18698.01	37738.12
COL10A1	11.81	20.27	23.10	52.86	186.27	782.90
COL11A1	1499.95	3538.83	9008.11	17962.01	30429.23	52635.81
COL12A1	3601.83	3546.07	7596.80	16177.85	23013.66	36610.96
COL13A1	174.18	306.29	379.29	448.09	585.02	721.50
COL14A1	378.33	143.31	719.76	6937.37	14811.47	17633.53
COL15A1	225.73	1747.20	2791.14	3380.38	4071.07	3883.88
COL16A1	894.81	1937.96	3322.55	6120.20	9833.97	20280.25
COL17A1	143.26	78.31	55.94	42.38	42.33	35.30
COL18A1	2127.99	2204.87	1626.42	1284.55	1018.67	850.31
COL19A1	127.48	246.67	259.59	222.47	153.73	193.12
COL20A1	4.67	3.14	9.85	15.92	27.92	80.41
COL21A1	41.79	73.19	84.17	131.28	273.18	437.08
COL22A1	13.21	142.59	342.76	270.03	254.34	306.64
COL23A1	36.15	34.03	32.89	43.44	43.10	81.54
COL24A1	395.79	512.85	361.95	215.21	273.55	443.13
COL25A1	416.17	1296.23	1387.54	1625.92	1206.03	596.14
COL26A1	1494.92	3086.07	3852.59	5092.88	5088.52	2315.01
COL27A1	552.33	2296.59	4434.06	8204.75	12311.92	19412.37
COL28A1	5.63	13.81	13.81	22.57	23.50	40.69

6. Transcription Factors in Chondrogenesis

Using a GO term-based search (GO:0003700: 'transcription factor activity, sequence-specific DNA binding'), we searched for TFs in our dataset. Of the 1175 genes in this GO category, 600 transcripts were identified in our dataset (Fig. 6C; see list in Supporting Information, Supplemental Data 4 – *transcription factors.xlsx*). We then excluded transcripts

from further analysis that did not meet the following criteria: fold-change > ±2; average normalised expression values > 1,000 at least at one time point; p value < 0.05 (Supporting Information, Table S8). We matched the entries on this list containing 52 differentially expressed TFs with the entities of the GO term 'cartilage development' (GO:0051216). Only the following 6 entries (~12%) of our list of TFs were previously annotated under the GO:0051216 term: *NFIB*, *MEF2C*, *SOX9*, *SRF*, *MYCN*, and *MSX2*. With the strictness of our selection criteria in mind, we advocate that the remaining 88% of TFs in our list are also worthy of consideration for further exploration and possibly as an addition to the list of genes involved in cartilage development as either positive or negative regulators. It should also be noted that 12 (~23%) out of these 52 genes (*EBF1*, *ATOH8*, *NFIB*, *MEF2C*, *SOX9*, *BHLHE40*, *FOS*, *MAFF*, *HES1*, *EGR1*, *MYC*, and *MECOM*) are also differentially expressed during the chondrogenic differentiation of human MSC cultures (hMSC data were retrieved from Gene Expression Omnibus, GSE109503), and only 3 of these (*NFIB*, *MEF2C*, *SOX9*) were previously annotated under the GO term cartilage development (GO:0051216).

When looking separately at our selection criteria for TFs (see lists in Supporting Information, Supplemental Data S4 – *transcription factors.xlsx*), 255 TF genes have an mRNA expression that reaches a normalised relative expression of > 1,000 at least at one time point during chondrogenesis of the chicken micromass cultures; 29 (~11%) of these entries are annotated under GO:0051216. Some TFs are steadily expressed during the entire cartilage differentiation period and are necessary for non-stage-dependent processes of chondrogenesis. We identified 96 entries in the chicken model that have a persistent normalised expression value of > 1,000 through the entire experimental period. 14 of them (~15%) are also annotated under the 'cartilage development' GO term (GO:0051216). If our full list of TFs (with 600 entries) is aligned against the list for cartilage development (GO:0051216), 45 entries (~8%) emerge as matches. The above results signify a relative enrichment of the genes with chondrogenic

associations among TFs with a higher normalised read count (~15% vs. ~8%). In hMSC cultures undergoing chondrogenic differentiation, a very similar portion of differentially expressed TFs (under the same criteria defined above) is annotated under the GO:0051216 term (12 entries from the total of 110, ~11%; *GLI3*, *MEF2C*, *NFIB*, *RARG*, *RELA*, *SCX*, *SMAD3*, *SNAI1*, *SOX5*, *SOX6*, *SOX9*, and *TRPS1*), but the total number of entries that fit our criteria is twice as much (110 *vs*. 52 in chicken HDC).

Of the 9 genes that are differentially expressed in both chicken micromass cultures and chondrogenic hMSC cultures (*EBF1*, *ATOH8*, *BHLHE40*, *FOS*, *MAFF*, *HES1*, *EGR1*, *MYC*, and *MECOM*), and are not annotated under the GO:0051216 term, only *BHLHE40* is upregulated during the culturing period in both models. The rest are either downregulated in both (*MAFF*, *HES1*, *EGR1*, *MYC*, *FOS*, and *MECOM*), or upregulated in the chicken model, but downregulated in hMSC chondrogenic cultures (*FOS*, *EBF1*, and *ATOH8*). (*FOS* undergoes both a significant up- and downregulation during the examined period in the chicken micromass cultures.)

As hMSC chondrogenic cultures are more inclined towards terminal differentiation and subsequent osteogenesis, it is of particular interest to identify TFs that have a different expression pattern in the two models. Therefore, in the next steps, we turned our focus to *FOS*, *EBF1*, and *ATOH8*, which are the of particular interest due to their lack of inclusion under the GO:0051216 term and the fact that they only undergo significant upregulation in chicken micromass cultures, but not in hMSC chondrogenic cultures.

We have performed an *in silico* signalling network analysis for all three genes using the SIGNOR 2.0 public repository of the Tor Vergata University of Rome (25) mainly looking for curated interactions where they are present as regulators not restricted to any species in particular. Data were only available for *FOS*. The three target partners *CYP19A1*, *HSD3B2*, and *STAR* are all defined to be under positive transcriptional regulation of our gene of interest. Out

of the three, only *STAR* has a non-negligible expression level, the pattern of which displays a strong resemblance to that of *FOS*, with an early downregulation and a marked late (days 4 to 6) upregulation (fold-change: ~2.6 for *FOS* and ~1.5 for *STAR*).

We also evaluated the STRING interaction networks of *FOS*, *EBF1*, and *ATOH8*. Following the extraction of the top 25 interaction partners for each gene (retrieved from *Homo sapiens* data for best data abundance), their expression levels were analysed and matched to the original gene of interest. Only those interacting partners with normalised expression levels of > 1,000 at least at one time point were considered for analysis. All fold change values at individual time points were compared to that of the gene of interest, and partners with the lowest standard deviation of FC ratios were considered as best matches. Three interacting partners with the most similar expression pattern for each gene (*JUND*, *RB1*, and *JUN* for *FOS*; *ZNF423*, *HECW2*, and *RNF8* for *EBF1*; and *TWIST2*, *SMARCB1*, and *CCNG2* for *ATOH8*) are visualised in Fig. 7.

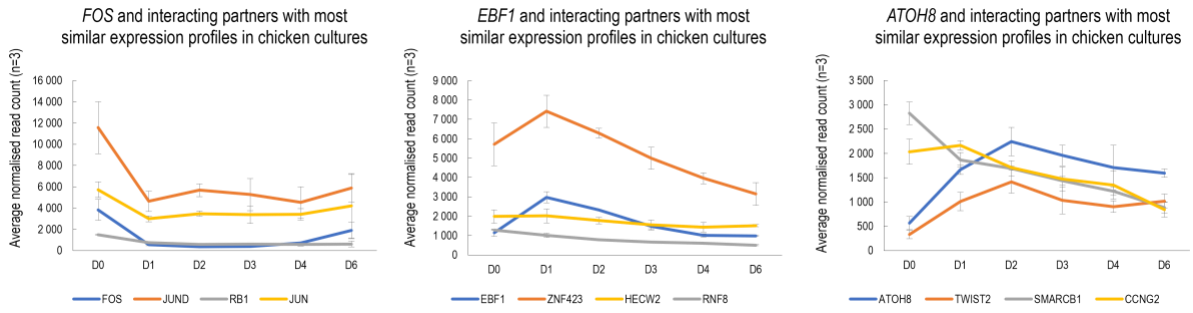

Figure 7. 3 interacting partners with the most similar expression patterns to the transcription factors *FOS*, *EBF1*, and *ATOH8*, as retrieved from the STRING database.

7. Gene Co-Expression Networks

Using signed weighted gene correlation network analysis (WGCNA), we first generated a dendrogram and a trait heatmap of the samples using the three classic chondrogenic markers *SOX9*, *COL2A1*, and *ACAN* (Fig. 8). The hierarchical clustering showed a clear separation between groups of biological replicates according to age and the three marker genes, except days 3 and 4, which formed a common group. We then searched for subsets of genes that highly correlated with the expression patterns of the three marker genes, referred to as 'traits'. To this end, genes were clustered into 48 modules of correlated expression patterns, with each module designated an arbitrary colour (Fig. 9). We identified the *blue* module for detailed analysis from WGCNA, as it showed the highest positive correlation with all four investigated traits. This module contained 1682 genes (see Supporting Information, Supplemental Data S5 – *WGCNA.xlsx*). GO pathway analysis performed on this list revealed that the following terms relevant to chondrogenesis and ECM production were significantly enriched: "regulation of developmental process (GO:0050793)," "collagen fibril organization (GO:0030199)," "extracellular matrix organization (GO:0030198)," "regulation of cytosolic calcium ion concentration (GO:0051480)," and "ion transport (GO:0006811)." Since these pathways are highly relevant to chondrogenic differentiation, the *blue* module can be thus classified as the chondrogenic functional module.

We then extracted the top ~50 most highly correlated genes from the *blue* module and exported the edge data to Cytoscape. To identify key driving molecules in the network, the genes were sorted according to closeness centrality values, and the connections between them were visualised (Fig. 8). *GTF2IRD1*, *COL5A1*, *PRNP*, *MGP*, and *KLF13* were the top 5 genes with the highest numbers of connections, but other relevant genes such as *BMP7*, *COL9A1*, *COL27A1*, *CSGALNACT1*, and *ECM2* were also present.

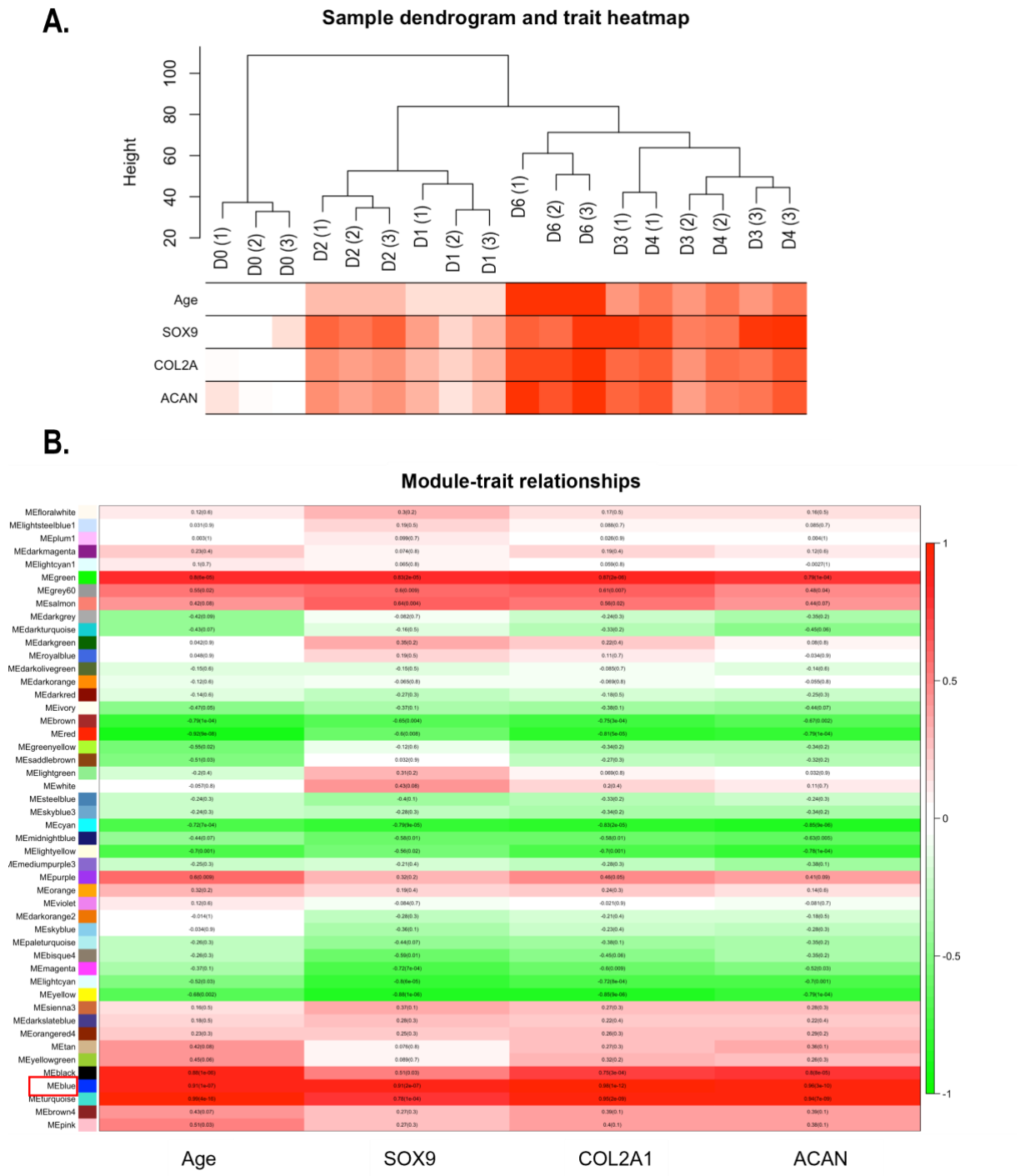

Figure 8. Signed WGCNA was used to identify subsets of genes that were highly correlated with the following traits: time points (age), *SOX9*, *COL2A1*, and *ACAN* expression patterns. Genes were then clustered into modules designated with arbitrary colours. **A.** Dendrogram of RNASeq samples (codes: D, day; numbers in brackets indicate biological replicates) and corresponding changes in traits. The lowest values are shown in white; the highest values are depicted in red. **B.** Modules whose eigengenes are highly correlated with age (days in culture) and the expression patterns of *SOX9*, *COL2A1*, and *ACAN*; their corresponding Pearson correlation values are shown.

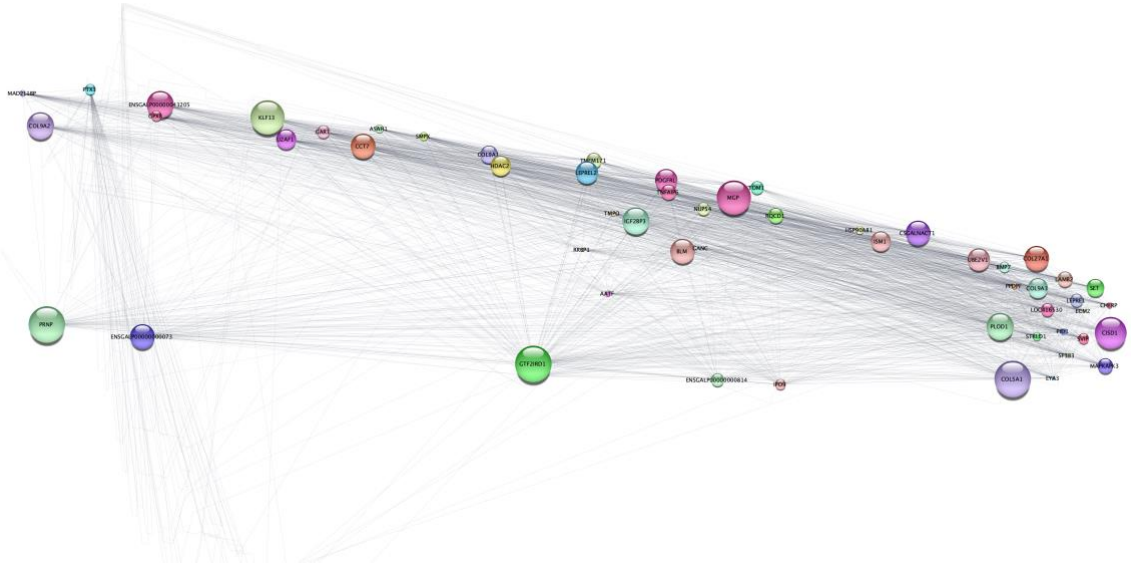

Figure 9. The edge data of the top ~50 genes from the *blue* module from the WGCNA analysis were exported to Cytoscape. The genes were then sorted according to closeness centrality values, and the connections between them were visualised. Node size indicates closeness centrality values; edge length represents the strength of the correlation between the respective nodes.

Discussion

In this work, we examined the differences between gene expression patterns during early cartilage formation in micromass cultures of embryonic limb bud-derived progenitor cells using RNA sequencing. We detected a progressively different and distinct transcriptome during key stages of chondrogenesis. We confirmed the involvement of the top DEGs in chondrogenic differentiation using pathway analysis. We discovered several chondrogenesis-associated TFs and new collagen subtypes not previously linked to cartilage formation.

In situ hyaline cartilage formation in the embryo is a multistep, dynamic, and sequential process that is controlled by several TFs, soluble mediators, ECM, cell-cell, and cell-matrix interactions, as well as epigenetic and miRNA-mediated mechanisms (26-28). In stark contrast, the standard *in vitro* culture conditions for the differentiation of MSCs into cartilage are quite simple and largely unchanged since the original description (29). Current cartilage regeneration options, even the “gold standard” method of using a combination of growth factors, scaffolds,

and MSCs, do not reliably produce hyaline cartilage in favour of biomechanically inferior fibrous cartilage (30). Therefore, there is a critical need for more effective hyaline cartilage regeneration therapies, for which a deep understanding of the molecular events determining chondrogenesis is essential. To our knowledge, this is the first study to examine the global temporal transcriptome landscape of *in vitro* chondrogenesis by high-throughput RNA sequencing using a well-established embryonic limb bud-derived 3D micromass system, which closely recapitulates the process *in vivo*.

High-throughput transcriptomic technologies such as microarrays and RNA sequencing are tools for studying the expression patterns of thousands of transcripts, allowing the identification of differentially expressed genes between two or more conditions, and describing altered biological pathways. Literature on unbiased global transcriptomic profiling of the key regulators involved in chondrogenesis is scarce, and those that are available are carried out on different model systems. Microarrays have previously been employed to examine temporal gene expression patterns during *in vitro* chondrogenic differentiation in a mouse micromass culture system (31). Global gene expression profiling analysis of micro-dissected chondrogenic tissues derived from tibial and fibular pre-condensed mesenchyme from mouse hind limbs using microarrays is the first *in vivo* transcriptomic study on cartilage development (32). Temporal changes in key molecular components during the formation of hyaline-like cartilage from infrapatellar fat pad-derived MSCs in a micromass culture were studied using microarray gene expression analysis (33). Microarrays have also been employed to analyse and compare global gene expression (34) and miRNA expression profiles of chondrocytes derived from human induced pluripotent stem cells (hiPSCs) (35). However, microarrays are limited in their ability to quantify the expression levels of a set of pre-determined transcripts printed on a given chip. Therefore, recent research employs ultrahigh-throughput RNASeq, which has a number of advantages over conventional microarrays.

Detailed unbiased NGS data on chondrogenic differentiation are scarce. Recently, a data set of high-throughput RNA sequencing of pellet cultures generated from primary human bone marrow-derived MSCs induced towards the chondrogenic lineage at six different time points (day 0 to day 21) was published (6). The authors identified a chondrogenic gene subset containing 1172 entities, whose functional characterisation promises to harness the potential of MSCs for cartilage tissue engineering. The same laboratory has recently published a paper on chondrogenic differentiation of hiPSCs using bulk and single-cell RNA sequencing and identified gene regulatory networks regulating chondrogenesis (36). Temporal changes in the transcriptome of human embryonic stem cells (ESCs) during key stages of chondrogenic differentiation have also been recently published (37).

All the above studies were conducted on cells that do not spontaneously generate cartilage *in vitro*, and existing protocols rely on the addition of growth factors and other compounds such as insulin-transferrin-selenium, dexamethasone, ascorbic acid, L-proline, TGF-β3, GDF5, FGF2, and NT4 to achieve chondrogenesis *in vitro*. The main advantage of the embryonic limb bud progenitor-derived micromass model is that, unlike most other differentiation techniques, our protocol does not require additional growth factors and other reagents to achieve high-quality chondrogenic differentiation and ECM similar to hyaline cartilage.

Chondrogenic Gene Expression Profile

Most studies addressing MSC-derived cartilage regeneration rely on the analysis of a selected subset of common cartilage marker genes. To this end, we prepared a custom-made list of “classical” markers for hyaline cartilage and completed that list with the genes annotated with relevant GO terms. We found that all genes on these lists were expressed in the micromass model but with different temporal patterns (i.e., were assigned to all six unsupervised clusters)

and expression levels. For example, *SOX9*, a pivotal TF in chondrogenesis and mature cartilage, along with its partners *SOX5* and *SOX6*, followed a typical "chondrogenic" pattern, displaying gradually higher transcript levels in more mature cultures. In addition to securing and maintaining the commitment of skeletogenic progenitor cells to the chondrocyte lineage, SOX9 is critical to ensure adult articular cartilage maintenance and function by transcriptionally activating genes (i.e. *COL2A1* and *ACAN*) essential for cartilage ECM, and by repressing factors and pathways that favour non-chondrocyte lineages (38). In contrast to an earlier microarray-based study, which only managed to detect *SOX4*, *SOX5*, *SOX8*, *SOX9*, and *SOX11* in normal and OA cartilage (39), we identified transcripts for 13 SOX transcription factors, with *SOX9*, *SOX6*, *SOX5*, *SOX11*, *SOX8*, and *SOX4* displaying the highest read numbers, which confirms the much higher sensitivity of the RNASeq-based methodology. Of these, only the SOX-trio (*SOX9*, *SOX5*, and *SOX6*) followed the typical "chondrogenic" pattern. We also checked the profiles of the TFs which are known to control *SOX9* expression (Supplemental Table S9). Among these TFs, *HIF1A*, *JUND*, *HDAC2*, *RAD21* and *CHD2* were the most abundant; moreover, *EBF1*, *FOS*, and *MYC* have also been identified as TFs of key importance during our detailed analysis.

The most abundantly expressed transcripts in the chondrogenic cultures included ribosomal proteins (indicating preparation to intense protein synthesis), chaperones, cytoskeletal components (members of the actin cytoskeleton, the microtubular system, and vimentin intermediate filaments), ECM components, members of the canonical TGFβ/BMP signalling, multifunctional heterogeneous nuclear ribonucleoproteins (hnRNPs) mediating mRNA metabolism, transcription factors, and glycolytic enzymes (*GAPDH*, *ENO1*, *ALDOC*, and *PKM*). Since chondrocytes are highly glycolytic cells, the fact that these genes were found in high abundance corroborates the scientific literature (40). The above functional classification

of the highly abundant transcripts was also confirmed by the GO and IPA analyses. These are in perfect agreement with published data (13,41).

When we looked at the DEGs in pairwise comparisons between time points, we identified several genes with important roles in skeletal development. From day 0 to day 1 (transition from chondroprogenitors to chondroblasts) *IHH*, *GDF5*, *FGF8*, and *WNT16* were amongst the top DEGs, which are all key components of chondrogenesis (13). We also noted a robust differential expression in transcripts involved in ECM production such as chondrolectin (*CHODL*), chondroitin sulfate N-acetylgalactosaminyltransferase 1 (*CSGALNACT1*), *MMP27*, *COL14A1*, *COL16A1*, matrilin-1 (*MATN1*), and chondroadherin (*CHAD*), also in line with the literature (42). At the same time, myogenic factor 5 (*MYF5*) was massively downregulated by day 2, indicating the gradual loss of the myogenic lineage from the micromass cultures.

To further characterise genes directly involved in chondrogenic differentiation, we employed a gene co-expression analysis using WGCNA, based on the assumption that co-expressed genes are functionally related and co-regulated (6). The rationale behind choosing the three classical chondrogenic markers *SOX9*, *COL2A1*, and *ACAN* as traits in our WGCNA analysis was that the genes showing a highly correlated expression pattern would likely be involved in governing chondrogenic differentiation and ECM production. The chondrogenic functional module (*blue*) contained 1682 genes with a very high correlation to the above traits. Genes in this module included voltage-gated calcium and potassium channel subunits and other transporters, cadherins, carbohydrate sulfotransferases, collagen alpha 1 chains, growth factors, TFs with a known involvement in chondrogenesis, as well as lncRNAs and miRNAs. Specifically, we picked up TFs including *FOXD3*, *FOXF1*, *FOXI2*, *FOXO6*, *FOXP4*, *HMGB1*, *HMGB2*, *HOXA11*, *HOXB4*, *HOXB5*, *HOXB6*, *SOX5*, *SOX7*, and *SOX10*; and signalling molecules such as FGFs and FGFRs, *IHH*, GDFs, semaphorins (*SEMA3D*, *SEMA3G*, *SEMA5A*), *TGFBI*, and WNTs (*WNT2B*, *WNT3A*, *WNT5B*, *WNT9B*, *WNT10A*). These are in

perfect agreement with published data (42,43). When we compared the genes between our chondrogenic module and those obtained in the hMSC-based chondrogenic model using a similar methodology (6), 88 common entities were identified (Supporting Information, Supplemental Data S4 – *WGCNA.xlsx*). These genes represent the 'core' chondrogenic subset between the two chondrogenic models, which include characteristic markers for cartilage ECM (*ACAN*, *COL2A1*, *COL3A1*, *COL9A1*, *COL27A1*, *CSPG4*), transporters and channels (*KCNKH*, *GTR10*, *S35C2*, *TRPV3*), the *SOX5* transcription factor, and cell surface receptors (*GRK5*, *INAR2*, *NTRK2*, *PE2R2*).

Collagen Expression Profile

Since SOX9 is a key regulator of genes coding for ECM components during chondrogenic differentiation, we specifically examined the gene expression profiles of collagens. The importance of collagen in cartilage ECM cannot be overstated (44). The biological properties of cartilage primarily rely on its unique and extensively cross-linked collagen network, which exhibits regional differences in articular cartilage (45). Given the complex ultrastructure created during cartilage development, there appears to be little capacity for chondrocytes or induced MSCs to recapitulate the original collagen architecture following injury (2). For this reason, detailed knowledge of the collagen subtypes, including minor collagens, expressed by *in vitro* models of chondrogenesis is critical.

Numerous collagen subtypes have been identified in articular cartilage, such as type II, III, IX, X, XI, VI, XII, XIII, and XIV collagen (2). Collagen fibrils in articular cartilage mostly consist of type II collagen accompanied by a lower number of minor collagens, which provide cartilage with tensile strength and contribute to the physical properties of the matrix. In our model, we identified the transcripts of alpha chains for almost all known collagen types (I–XXVIII), except for collagen type XXIX (epidermal collagen). Initially, *COL1A1* and *COL1A2*

had the highest transcript numbers, but as early as day 1, *COL2A1* took over and maintained a steadily increasing expression level throughout chondrogenesis, reaching almost 300,000 read numbers by day 6. In contrast to our results, when bone marrow-derived MSCs were induced toward the chondrogenic lineage, the gene coding for the alpha2 chain of collagen type I (*COL1A2*) had the highest expression levels throughout, and *COL2A1* was only the second most abundant collagen transcript by day 21 (mature chondrocyte stage) (6). Additionally, genes coding for collagen types XVII, XIX, XX, and XXIII were not expressed in the MSC-derived chondrocytes. The genes encoding the alpha chain of type III collagen were the third most abundant collagen gene type by day 21 in that model, and the substantial high levels of collagen type X (4th highest level) indicated the formation of hypertrophic chondrocytes. In contrast, type X collagen remained downregulated by day 6 in the limb bud-derived micromass model.

Looking at the collagen expression profile of iPSCs during chondrogenesis, *COL1A1* had, by far, the highest levels by day 60, followed by *COL1A2* and *COL3A1*, with relatively low *COL2A1* levels (36). Interestingly, iPSCs also did not express the genes encoding collagen types XVII and XX. This dataset also included gene expression profiles for adult, juvenile, embryonic, and 17-week cartilage, which showed variable collagen gene expression profiles. One of the adult cartilage samples expressed *COL12A1* most strongly, and *COL3A1* levels were also quite high. The adolescent cartilage samples showed very high *COL6A1* expression, while *COL2A1* was only moderately expressed. A high number of *COL2A1* transcripts was mainly present for embryonic and week-17 cartilage samples (36).

In addition to the major collagens, several minor collagens were detected at high (>1000) read numbers in our chondrogenic model. Collagen type IV is network-forming collagen, which is known to be limited to the pericellular matrix of articular cartilage, where it may be involved in maintaining the chondrocyte phenotype and viability (45). Its expression peaked in differentiating chondrocytes (day 3). The alpha chain of collagen type V, a dominant

regulator of collagen fibrillogenesis, has been documented to accumulate as articular cartilage matures (46). Furthermore, the α1(V) chains were described to be cross-linked to α1(XI) chains in cartilage ECM to form V/XI polymers (46). Collagen type I can form fibrils by itself, but it can also form heterofibrils with collagen types III and V (44). Little is known about the role of the non-fibrillar, short chain collagen type VIII in cartilage, other than that it is present in cartilage ECM (47). Collagen XIV, a FACIT collagen, is a member of the cross-linked collagen network of cartilage ECM, which is predominantly expressed in late embryonic and differentiated cartilage (44,45). Collagen type XV, which belongs to the family of multiplexin collagens, is known to be expressed in cartilage primordia, with stronger signals in the perichondrium (48). Type XXVII collagen is prominently located at the cartilage–bone interface and surrounding proliferative chondrocytes in the epiphyseal growth plate; it plays a role in the transition of cartilage to bone during skeletogenesis (45). However, collagen type XXVI has not previously been implicated in cartilage matrix.

Overall, the chondrogenic model derived from primary limb buds produced a more hyaline cartilage-like ECM compared with hMSCs, at least at the transcriptional level. Although the specific role of minor collagens in chondrogenesis has not been explored, targeted studies may contribute to a deeper understanding of the dynamics of cartilage matrix production, which could facilitate the development of new biomarkers of joint health and drug development in OA.

Transcription Factor Expression Profile

The human genome encodes more than 3000 TFs, many of which are known to control the chondrocyte phenotype at the genomic level (43). To identify TFs that have not been associated with chondrogenic differentiation, we narrowed down the TFs positively identified in chondrogenic cells (GO:0003700) based on a set of criteria (fold-change > 2; average

normalised expression values > 1,000 at least at one time point). We then matched the chicken data with the chondrogenic human MSC transcriptome (retrieved from GSE109503). Out of the 9 TF genes that are differentially expressed in both models and are not annotated under the GO:0051216 'cartilage development' term, only *BHLHE40* (DEC1; differentially expressed in chondrocytes 1) was upregulated during the culturing periods studied. Not only was that gene not annotated under the cartilage development GO term, but very few publications also discuss *BHLHE40* in the context of chondrogenesis. It appears to promote both early and terminal stage chondrocyte differentiation (49), and it is also an important player in the regulation of the molecular clock (50), the proper functioning of which promotes early chondrogenesis (15). Therefore, *BHLHE40/DEC1* appears to be an important candidate for more targeted studies. While some of the TFs picked up in this study that are known interacting partners of *FOS*, *EBF1*, and *ATOH8* have well-established roles in chondrocytes, some of them have not yet been described in the context of chondrogenesis. DNA damage signals transducer RING finger protein 8 (*RNF8*) transduces DNA double-strand break (DSB) signalling pathways. *RNF8* promotes proliferation by upregulating c-Myc expression via the Wnt/β-catenin pathway, one of the fundamental signaling cascades in development and homeostasis (51). SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1 (*SMARCB1*) is a tumor suppressor which regulates ATG5, an essential autophagy-related gene that plays a vital role in autophagosome formation (52). During preadipocyte proliferation and differentiation, *FoxO6* directly targets and induces expression of cyclin G2 (*CCNG2*) (53). Given that these factors are among the top interacting partners of key chondrogenic TFs, unveiling the specific role(s) of *RNF8*, *SMARCB1*, and *CCNG2* in hyaline cartilage differentiation warrant further studies.

Strengths of the Study

The study of human cartilage development is difficult for ethical reasons. Therefore, we used a chicken embryonic limb bud-derived micromass system to model chondrogenesis *in vitro*, which is considered to better recapitulate the normal development of permanent cartilage, in contrast to MSCs. Most of the published global high-throughput transcriptome data on the regulation of human chondrogenesis are from MSC or iPSC differentiation, which has the potential disadvantage of the type of cartilage produced (fibrous cartilage) and the lack of spontaneous chondrogenic differentiation. The advantage of the embryonic LMP-based micromass model system is that it does not require additional growth factors and other reagents to achieve better chondrogenic differentiation compared to MSCs or iPSCs. Having performed a detailed analysis of the genes identified in this dataset, we did not only confirm known components of chondrogenesis in this model, but provide an integrated view of refined inter-regulatory networks between the chondrogenic subset of genes.

Limitations of the Study

There are known differences between chicken and mammalian cartilage (54), that should be considered when applying these data to mammalian or human chondrogenesis.

Conclusions

Here, for the first time, we have provided an in-depth quantitative transcriptomic landscape of *in vitro* chondrogenesis in avian embryonic limb bud-derived micromass cultures using RNA sequencing. We identified several subsets of genes associated with chondrogenesis, including potentially novel TFs. In addition to reveal transcriptomic signatures already known to control chondrogenic differentiation in other models, we provide a refined, integrated view of inter-regulatory networks operating in addition to the conventional early stage chondrogenic

pathways. The data presented could be used to advance the knowledge of chondrogenic differentiation and cartilage pathology and contribute to better cartilage regeneration techniques.

Perspectives

The ultimate goal for successful tissue engineering of cartilage is to produce cartilaginous tissue with characteristics reminiscent of articular cartilage. This is a complex challenge: in addition to finding the right cell source, appropriate and time-dependent stimuli with exogenous factors are also required. However, these cannot be successfully applied until the complex regulatory networks that control the formation of hyaline cartilage are fully explored. The results of this study provide a solid basis for fine-tuning and improving current chondrogenic differentiation protocols.

Data Availability

The entire dataset was deposited and published in the BioProject database (http://www.ncbi.nlm.nih.gov/bioproject/). BioProject ID: PRJNA817177. All other data generated or analysed during this study are included in this published article (and its supplementary information files).

Acknowledgements

The authors are thankful to Mrs. Krisztina Bíróné Barna for her technical assistance. We wish to thank Dr. Peter Nagy at the Department of Biophysics and Cell Biology, University of Debrecen, for developing the MATLAB-based image analysis software. We wish to thank Dr Gergely Nagy at the University of Debrecen for useful comments and discussions and

critical feedback on the manuscript before submission. We are indebted to Ms. Ebeid Rana Abdelsattar Mansour for proofreading the manuscript.

Funding

CM was supported by the Premium Postdoctoral Research Fellowship of the Eötvös Loránd Research Network (ELKH), and the Young Researcher Excellence Programme (grant number: FK-134304) of the National Research, Development and Innovation Office, Hungary. CM was also supported by the EFOP-3.6.3-VEKOP-16-2017-00009 project co-financed by the EU and the European Social Fund. Project no. TKP2020-NKA-04 was implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 funding scheme.

Conflicts of Interest

The authors declare no conflicts of interest. This paper was written by the authors within the scope of their academic and research positions. None of the authors have any relationships that could be construed as biased or inappropriate. The funding bodies were not involved in the study design, data collection, analysis, and interpretation. The decision to submit the paper for publication was not influenced by any funding bodies.

References

1. Reinhardt, R., Gullotta, F., Nusspaumer, G., Unal, E., Ivanek, R., Zuniga, A. and Zeller, R. (2019) Molecular signatures identify immature mesenchymal progenitors in early mouse limb buds that respond differentially to morphogen signaling. *Development*, **146**.
2. Eyre, D.R., Weis, M.A. and Wu, J.J. (2006) Articular cartilage collagen: an irreplaceable framework? *Eur Cell Mater*, **12**, 57-63.
3. Bhosale, A.M. and Richardson, J.B. (2008) Articular cartilage: structure, injuries and review of management. *Br Med Bull*, **87**, 77-95.
4. Chaly, Y., Hostager, B., Smith, S. and Hirsch, R. (2020) The Follistatin-like Protein 1 Pathway Is Important for Maintaining Healthy Articular Cartilage. *ACR Open Rheumatol*, **2**, 407-414.
5. Hunter, D.J. and Bierma-Zeinstra, S. (2019) Osteoarthritis. *Lancet*, **393**, 1745-1759.

6. Huynh, N.P.T., Zhang, B. and Guilak, F. (2019) High-depth transcriptomic profiling reveals the temporal gene signature of human mesenchymal stem cells during chondrogenesis. *FASEB J*, **33**, 358-372.
7. Fellows, C.R., Matta, C., Zakany, R., Khan, I.M. and Mobasheri, A. (2016) Adipose, Bone Marrow and Synovial Joint-Derived Mesenchymal Stem Cells for Cartilage Repair. *Front Genet*, **7**, 213.
8. Richardson, S.M., Kalamegam, G., Pushparaj, P.N., Matta, C., Memic, A., Khademhosseini, A., Mobasheri, R., Poletti, F.L., Hoyland, J.A. and Mobasheri, A. (2016) Mesenchymal stem cells in regenerative medicine: Focus on articular cartilage and intervertebral disc regeneration. *Methods*, **99**, 69-80.
9. Pelttari, K., Winter, A., Steck, E., Goetzke, K., Hennig, T., Ochs, B.G., Aigner, T. and Richter, W. (2006) Premature induction of hypertrophy during in vitro chondrogenesis of human mesenchymal stem cells correlates with calcification and vascular invasion after ectopic transplantation in SCID mice. *Arthritis Rheum*, **54**, 3254-3266.
10. Somoza, R.A., Welter, J.F., Correa, D. and Caplan, A.I. (2014) Chondrogenic differentiation of mesenchymal stem cells: challenges and unfulfilled expectations. *Tissue Eng Part B Rev*, **20**, 596-608.
11. Klumpers, D.D., Mooney, D.J. and Smit, T.H. (2015) From Skeletal Development to Tissue Engineering: Lessons from the Micromass Assay. *Tissue Eng Part B-Re*, **21**, 427-437.
12. Rolfe, R.A., Shea, C.A. and Murphy, P. (2022) Geometric analysis of chondrogenic self-organisation of embryonic limb bud cells in micromass culture. *Cell Tissue Res*.
13. Humphreys, P.A., Mancini, F.E., Ferreira, M.J.S., Woods, S., Ogene, L. and Kimber, S.J. (2021) Developmental principles informing human pluripotent stem cell differentiation to cartilage and bone. *Semin Cell Dev Biol*.
14. Somoza, R.A., Correa, D., Labat, I., Sternberg, H., Forrest, M.E., Khalil, A.M., West, M.D., Tesar, P. and Caplan, A.I. (2018) Transcriptome-Wide Analyses of Human Neonatal Articular Cartilage and Human Mesenchymal Stem Cell-Derived Cartilage Provide a New Molecular Target for Evaluating Engineered Cartilage. *Tissue Eng Part A*, **24**, 335-350.
15. Alagha, M.A., Vago, J., Katona, E., Takacs, R., van der Veen, D., Zakany, R. and Matta, C. (2020) A Synchronized Circadian Clock Enhances Early Chondrogenesis. *Cartilage*, 1947603520903425.
16. Matta, C., Fodor, J., Szijgyarto, Z., Juhasz, T., Gergely, P., Csernoch, L. and Zakany, R. (2008) Cytosolic free Ca2+ concentration exhibits a characteristic temporal pattern during in vitro cartilage differentiation: a possible regulatory role of calcineurin in Ca-signalling of chondrogenic cells. *Cell Calcium*, **44**, 310-323.
17. Ahrens, P.B., Solursh, M. and Reiter, R.S. (1977) Stage-related capacity for limb chondrogenesis in cell culture. *Dev Biol*, **60**, 69-82.
18. Juhasz, T., Matta, C., Somogyi, C., Katona, E., Takacs, R., Soha, R.F., Szabo, I.A., Cserhati, C., Szody, R., Karacsonyi, Z. *et al.* (2014) Mechanical loading stimulates chondrogenesis via the PKA/CREB-Sox9 and PP2A pathways in chicken micromass cultures. *Cell Signal*, **26**, 468-482.
19. Matta, C., Juhasz, T., Fodor, J., Hajdu, T., Katona, E., Szucs-Somogyi, C., Takacs, R., Vago, J., Olah, T., Bartok, A. *et al.* (2019) N-methyl-D-aspartate (NMDA) receptor expression and function is required for early chondrogenesis. *Cell Commun Signal*, **17**, 166.
20. Takacs, R., Matta, C., Somogyi, C., Juhasz, T. and Zakany, R. (2013) Comparative analysis of osteogenic/chondrogenic differentiation potential in primary limb bud-derived and C3H10T1/2 cell line-based mouse micromass cultures. *Int J Mol Sci*, **14**, 16141-16167.
21. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**, 2498-2504.
22. Jafri, M.A., Kalamegam, G., Abbas, M., Al-Kaff, M., Ahmed, F., Bakhashab, S., Rasool, M., Naseer, M.I., Sinnadurai, V. and Pushparaj, P.N. (2019) Deciphering the Association of Cytokines, Chemokines, and Growth Factors in Chondrogenic Differentiation of Human Bone Marrow Mesenchymal Stem Cells Using an ex vivo Osteochondral Culture System. *Front Cell Dev Biol*, **7**, 380.

23. Bahlas, S., Damiati, L.A., Al-Hazmi, A.S. and Pushparaj, P.N. (2020) Decoding the Role of Sphingosine-1-Phosphate in Asthma and Other Respiratory System Diseases Using Next Generation Knowledge Discovery Platforms Coupled With Luminex Multiple Analyte Profiling Technology. *Front Cell Dev Biol*, **8**, 444.
24. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
25. Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., Peluso, D., Calderone, A., Castagnoli, L. and Cesareni, G. (2020) SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res*, **48**, D504-D510.
26. Goldring, M.B. (2012) Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. *Ther Adv Musculoskelet Dis*, **4**, 269-285.
27. Green, J.D., Tollemar, V., Dougherty, M., Yan, Z., Yin, L., Ye, J., Collier, Z., Mohammed, M.K., Haydon, R.C., Luu, H.H. *et al.* (2015) Multifaceted signaling regulators of chondrogenesis: Implications in cartilage regeneration and tissue engineering. *Genes Dis*, **2**, 307-327.
28. Michigami, T. (2014) Current understanding on the molecular basis of chondrogenesis. *Clin Pediatr Endocrinol*, **23**, 1-8.
29. Yoo, J.U., Barthel, T.S., Nishimura, K., Solchaga, L., Caplan, A.I., Goldberg, V.M. and Johnstone, B. (1998) The chondrogenic potential of human bone-marrow-derived mesenchymal progenitor cells. *J Bone Joint Surg Am*, **80**, 1745-1757.
30. Kuo, C.K., Li, W.J., Mauck, R.L. and Tuan, R.S. (2006) Cartilage tissue engineering: its potential and uses. *Curr Opin Rheumatol*, **18**, 64-73.
31. James, C.G., Appleton, C.T., Ulici, V., Underhill, T.M. and Beier, F. (2005) Microarray analyses of gene expression during chondrocyte differentiation identifies novel regulators of hypertrophy. *Mol Biol Cell*, **16**, 5316-5333.
32. Cameron, T.L., Belluoccio, D., Farlie, P.G., Brachvogel, B. and Bateman, J.F. (2009) Global comparative transcriptome analysis of cartilage formation in vivo. *BMC Dev Biol*, **9**, 20.
33. Felimban, R., Ye, K., Traianedes, K., Di Bella, C., Crook, J., Wallace, G.G., Quigley, A., Choong, P.F. and Myers, D.E. (2014) Differentiation of stem cells from human infrapatellar fat pad: characterization of cells undergoing chondrogenesis. *Tissue Eng Part A*, **20**, 2213-2223.
34. Stelcer, E., Kulcenty, K., Rucinski, M., Jopek, K., Richter, M., Trzeciak, T. and Suchorska, W.M. (2018) Chondrogenic differentiation in vitro of hiPSCs activates pathways engaged in limb development. *Stem Cell Res*, **30**, 53-60.
35. Stelcer, E., Kulcenty, K., Rucinski, M., Jopek, K., Richter, M., Trzeciak, T. and Suchorska, W.M. (2019) The Role of MicroRNAs in Early Chondrogenesis of Human Induced Pluripotent Stem Cells (hiPSCs). *Int J Mol Sci*, **20**.
36. Wu, C.L., Dicks, A., Steward, N., Tang, R., Katz, D.B., Choi, Y.R. and Guilak, F. (2021) Single cell transcriptomic analysis of human pluripotent stem cell chondrogenesis. *Nat Commun*, **12**, 362.
37. Griffiths, R., Woods, S., Cheng, A., Wang, P., Griffiths-Jones, S., Ronshaugen, M. and Kimber, S.J. (2020) The Transcription Factor-microRNA Regulatory Network during hESC-chondrogenesis. *Sci Rep*, **10**, 4744.
38. Lefebvre, V., Angelozzi, M. and Haseeb, A. (2019) SOX9 in cartilage development and disease. *Curr Opin Cell Biol*, **61**, 39-47.
39. Haag, J., Gebhard, P.M. and Aigner, T. (2008) SOX gene expression in human osteoarthritic cartilage. *Pathobiology*, **75**, 195-199.
40. Mobasheri, A., Rayman, M.P., Gualillo, O., Sellam, J., van der Kraan, P. and Fearon, U. (2017) The role of metabolism in the pathogenesis of osteoarthritis. *Nat Rev Rheumatol*, **13**, 302-311.
41. Yu, D.A., Han, J. and Kim, B.S. (2012) Stimulation of chondrogenic differentiation of mesenchymal stem cells. *Int J Stem Cells*, **5**, 16-22.
42. Halper, J. (2021) Basic Components of Connective Tissues and Extracellular Matrix: Fibronectin, Fibrinogen, Laminin, Elastin, Fibrillins, Fibulins, Matrilins, Tenascins and Thrombospondins. *Adv Exp Med Biol*, **1348**, 105-126.

43. Liu, C.F., Samsa, W.E., Zhou, G. and Lefebvre, V. (2017) Transcriptional control of chondrocyte specification and differentiation. *Semin Cell Dev Biol*, **62**, 34-49.
44. Bielajew, B.J., Hu, J.C. and Athanasiou, K.A. (2020) Collagen: quantification, biomechanics, and role of minor subtypes in cartilage. *Nat Rev Mater*, **5**, 730-747.
45. Luo, Y., Sinkeviciute, D., He, Y., Karsdal, M., Henrotin, Y., Mobasheri, A., Onnerfjord, P. and Bay-Jensen, A. (2017) The minor collagens in articular cartilage. *Protein Cell*, **8**, 560-572.
46. Wu, J.J., Weis, M.A., Kim, L.S., Carter, B.G. and Eyre, D.R. (2009) Differences in chain usage and cross-linking specificities of cartilage type V/XI collagen isoforms with age and tissue. *J Biol Chem*, **284**, 5539-5545.
47. Kapoor, R., Sakai, L.Y., Funk, S., Roux, E., Bornstein, P. and Sage, E.H. (1988) Type VIII collagen has a restricted distribution in specialized extracellular matrices. *J Cell Biol*, **107**, 721-730.
48. Muona, A., Eklund, L., Vaisanen, T. and Pihlajaniemi, T. (2002) Developmentally regulated expression of type XV collagen correlates with abnormalities in Col15a1(-/-) mice. *Matrix Biol*, **21**, 89-102.
49. Shen, M., Yoshida, E., Yan, W., Kawamoto, T., Suardita, K., Koyano, Y., Fujimoto, K., Noshiro, M. and Kato, Y. (2002) Basic helix-loop-helix protein DEC1 promotes chondrocyte differentiation at the early and terminal stages. *J Biol Chem*, **277**, 50112-50120.
50. Nakashima, A., Kawamoto, T., Honda, K.K., Ueshima, T., Noshiro, M., Iwata, T., Fujimoto, K., Kubo, H., Honma, S., Yorioka, N. *et al.* (2008) DEC1 modulates the circadian phase of clock gene expression. *Mol Cell Biol*, **28**, 4080-4092.
51. Ren, L., Zhou, T., Wang, Y., Wu, Y., Xu, H., Liu, J., Dong, X., Yi, F., Guo, Q., Wang, Z. *et al.* (2020) RNF8 induces beta-catenin-mediated c-Myc expression and promotes colon cancer proliferation. *Int J Biol Sci*, **16**, 2051-2062.
52. Li, M., Shen, Y., Xiong, Y., Wang, S., Li, C., Bai, J. and Zhang, Y. (2021) Loss of SMARCB1 promotes autophagy and facilitates tumour progression in chordoma by transcriptionally activating ATG5. *Cell Prolif*, **54**, e13136.
53. Abdalla, B.A., Chen, X., Li, K., Chen, J., Yi, Z., Zhang, X., Li, Z. and Nie, Q. (2021) Control of preadipocyte proliferation, apoptosis and early adipogenesis by the forkhead transcription factor FoxO6. *Life Sci*, **265**, 118858.
54. Eyre, D.R., Brickley-Parsons, D.M. and Glimcher, M.J. (1978) Predominance of type I collagen at the surface of avian articular cartilage. *FEBS Lett*, **85**, 259-263.