

UNIVERSIDADE ESTADUAL DE CAMPINAS  
SISTEMA DE BIBLIOTECAS DA UNICAMP  
REPOSITÓRIO DA PRODUÇÃO CIENTÍFICA E INTELECTUAL DA UNICAMP

**Versão do arquivo anexado / Version of attached file:**

Versão do Editor / Published Version

**Mais informações no site da editora / Further information on publisher's website:**

<https://www.sciencedirect.com/science/article/pii/S266591312200084X?via%3Dihub#gs2>

**DOI: 10.1016/j.ocarto.2022.100316**

**Direitos autorais / Publisher's copyright statement:**

©2022 by Elsevier. All rights reserved.

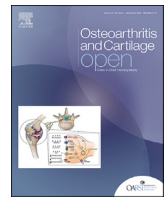
DIRETORIA DE TRATAMENTO DA INFORMAÇÃO

Cidade Universitária Zeferino Vaz Barão Geraldo

CEP 13083-970 – Campinas SP

Fone: (19) 3521-6493

<http://www.repositorio.unicamp.br>



# Differential gene co-expression network analyses reveal novel molecules associated with transcriptional dysregulation of key biological processes in osteoarthritis knee cartilage<sup>☆</sup>

I. Buzzatto-Leite<sup>a</sup>, J. Afonso<sup>b</sup>, B. Silva-Vignato<sup>b</sup>, L.L. Coutinho<sup>b</sup>, L.E. Alvares<sup>a,\*</sup>

<sup>a</sup> Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, Brazil

<sup>b</sup> Department of Animal Science, College of Agriculture "Luiz de Queiroz", University of São Paulo (ESALQ/USP), Piracicaba, Brazil

## ARTICLE INFO

### Keywords:

Osteoarthritis  
Knee cartilage  
PCIT  
RNA-Seq  
Epigenetics  
Transcription regulation

## ABSTRACT

**Objectives:** To compare co-expression networks of normal and osteoarthritis knee cartilage to uncover molecules associated with the transcriptional misregulation compromising biological processes (BPs) critical for cartilage homeostasis.

**Design:** Normal and osteoarthritis human knee cartilage RNA-seq GSE114007 dataset was obtained from the Gene Expression Omnibus database. Partial Correlation and Information Theory (PCIT) algorithm was used to build co-expression networks containing all nodes connecting to at least one differentially expressed gene (DEG) in normal and osteoarthritis networks. Hub and hub centrality genes were used to perform functional enrichment analysis. Enriched BPs known to be associated with both healthy and diseased cartilage were compared in depth.

**Results:** Differential co-expression network analyses allowed the identification of *DDX43* and *USP42* as exclusively co-expressed with DEGs in normal and osteoarthritis networks, respectively. The top hub and hub centrality genes of these networks were *HIST1H3A* and *SNHG12* (normal) and *TAF9B* and *OTUD1* (osteoarthritis). Enrichment analysis revealed several shared BPs between the contrasting groups, which are well-known in osteoarthritis pathogenesis. Protein-protein interaction network analysis for these BPs showed a global down-regulation of transcription factors in osteoarthritis. Specific transcription factors were identified as pleiotropic mediators in articular cartilage maintenance since they take part in several BPs. In addition, chromatin organisation and modification proteins were found relevant for osteoarthritis development.

**Conclusion:** Differential gene co-expression analysis allowed the identification of novel and high priority therapeutic candidate genes that may drive modifications in the transcriptional "status" of cartilage in osteoarthritis.

## 1. Introduction

Osteoarthritis (OA) is characterised by joint cartilage loss leading to direct contact between bones, causing swelling and pain [1]. Such damage is accompanied by modification in gene expression and biological processes (BPs) crucial to tissue homeostasis maintenance [2–4].

Differential gene expression analysis between healthy and diseased tissue is widely used to identify molecular changes associated with different pathologies, including osteoarthritis [5]. However, when used alone, this approach disregards the multiple molecular interactions performed by differentially expressed genes (DEGs). To overcome this

limitation, mathematical models are used to construct gene co-expression networks based on significant correlation values among DEGs, which allows for identifying interconnected genes with putative functional associations [6]. Therefore, differential gene co-expression analysis is valuable for investigating disparities in the expression patterns of gene clusters between healthy and diseased tissues [7,8].

Partial Correlation and Information Theory (PCIT) is a renowned algorithm that identifies gene co-expression patterns [9]. This algorithm tests gene co-expression by analysing the correlation between their expression values. First, the algorithm calculates the linear relationship strength between every two genes that make up each possible trio,

<sup>☆</sup> This study used a bioinformatics approach without experimental verification.

\*Corresponding author. Department of Biochemistry and Tissue Biology, University of Campinas – UNICAMP, Rua Monteiro Lobato 255, Cx. Postal 6109, CEP 13083-862, Campinas, SP, Brazil.

E-mail addresses: [igorbuzzatto@gmail.com](mailto:igorbuzzatto@gmail.com) (I. Buzzatto-Leite), [juafonsobio@gmail.com](mailto:juafonsobio@gmail.com) (J. Afonso), [barbarasilva@usp.br](mailto:barbarasilva@usp.br) (B. Silva-Vignato), [llcoutinho@usp.br](mailto:llcoutinho@usp.br) (L.L. Coutinho), [lealvare@unicamp.br](mailto:lealvare@unicamp.br) (L.E. Alvares).

<https://doi.org/10.1016/j.ocarto.2022.100316>

Received 9 March 2022; Received in revised form 9 September 2022; Accepted 17 October 2022

2665-9131/© 2022 The Authors. Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International (OARSI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

independent of the third, using a partial correlation calculation based on gene expression values. Then the algorithm defines a significance threshold for each tested correlation based on information theory, calculating the average ratio of partial and direct correlations for each trio. This calculation of the correlation-specific significant threshold is the main difference between PCIT and other algorithms used to calculate co-expression [10]. This PCIT feature identifies putative biological significant interactions between genes even when gene pairs are highly correlated over a small expression range [9].

The work performed by Fisch et al. (2018) evidenced a transcriptional dysregulation in OA-affected knee cartilage using RNA-seq data analyses. Firstly, the authors identified differentially expressed genes between normal and osteoarthritic tissue and, among them, those that encode transcription factors (TFs). Subsequent analyses were focused entirely on differentially expressed TFs whose binding sites were enriched in the DEG promoters. As a result, a small repertoire of TFs was identified as central mediators of abnormal gene expression in osteoarthritis, which intriguingly were downregulated in the disease. Despite the importance of these findings, the molecular basis of the global dysregulation of gene expression in osteoarthritic cartilage remains undefined.

Here, we used PCIT to build and compare co-expression networks of normal and osteoarthritis knee cartilage using the dataset generated by Fisch et al. (2018). Our main goal was to uncover molecules related to the transcriptional misregulation previously identified in the disease. For that, our data analysis approach encompassed all cartilage-expressed genes significantly connected to at least one DEG rather than focusing entirely on TFs. In addition, we determined hub and hub centrality genes and established their co-expression relationship to biological processes known to be dysregulated during osteoarthritis onset and progression. Our results revealed that genes involved in chromatin organisation/modification might regulate the transcriptional status in late-stage osteoarthritis and that critical biological processes are affected by a massive down-regulation of TFs.

## 2. Materials and methods

### 2.1. Data collection

The knee cartilage RNA-sequencing data used in this work are available under accession number GSE114007 in the Gene Expression Omnibus (GEO) public genomic data repository (<http://www.ncbi.nlm.nih.gov/geo/>). Samples from the normal and osteoarthritis groups were obtained from tissue banks ( $n = 18$ ) and knee replacement surgeries ( $n = 20$ ), respectively. The healthy group is composed of five female and 13 male samples (age 18–61, mean 38), whereas the osteoarthritis group includes 12 female and eight male samples (age 52–82, mean 66). The Fisch et al. (2018) dataset was selected for our analysis as it contained detailed information about gene expression, clinical data and a considerable sample size.

### 2.2. Filtering of sequencing data

GSE114007 sequencing data were already normalised to counts per million and log 2 transformed (log2CPM) as described elsewhere [11]. We considered for our analyses genes showing log2CPM values  $> 3.0$  in one or more samples, resulting in a list of 13,102 transcripts.

### 2.3. DEGs and gene co-expression networks

The PCIT algorithm was used to determine the differential gene co-expression between sample groups [9]. Initially, the correlation between the expression of all 13,102 transcripts was tested separately between samples from the normal and osteoarthritis knee cartilage groups. Secondly, we filtered the correlations so that only those containing at least one DEG, among the 1332 DEGs previously described elsewhere [5], were considered for the differential co-expression analysis. Cytoscape

software (<https://cytoscape.org/>) was used to visualize the co-expression networks constructed with filtered correlations [12]. The Network Analyzer tool of Cytoscape was used to obtain the connectivity degree and the betweenness centrality measures of each gene in the networks [13]. These values were used to identify hub genes, those with more correlations within networks, and hub centrality genes, which interconnect more groups of correlated genes within networks. The hubs were identified from the mean of the network's connectivity degree values plus three times the standard deviation value (mean + 3SD), while the hubs centrality were obtained from the mean of the network's betweenness centrality values plus three times the standard deviation value (mean + 3SD). Since several genes were simultaneously categorised as hub and hub centrality, a list without redundancy was generated. The differences between the normal and osteoarthritis networks concerning co-expression with DEGs were assessed [5]. In addition, we also identified which hubs and hubs centrality were unique or shared between the contrasting groups. We used the curated list of human TF from Lambert et al., 2018 to identify TFs within each network.

### 2.4. BPs identification

We performed functional enrichment analysis using STRING v11.0 software (<https://string-db.org>) [14] to infer the biological importance of the hubs and hubs centrality to osteoarthritis. BPs exhibiting false discovery rate (FDR)  $< 0.05$  were considered significant. The REVIGO algorithm (Reduce & Visualize Gene Ontology; <http://revigo.irb.hr>) [15] was used to summarise the redundant lists of GO terms and identify “key” BPs of each group. The biological processes that clustered other GO terms by semantic similarity were defined as “key” BP.

Protein-protein interaction networks of BPs shared between normal and osteoarthritis groups previously associated with osteoarthritis.

Based on the list of key BPs generated by REVIGO, we selected those previously associated with osteoarthritis's development and progression in the literature [3]. For further comparative analysis, we build protein-protein interaction (PPI) networks for key BPs enriched in both contrasting groups using STRING. Transcription factors, node's expression status (up or down-regulated) and the exclusive or shared nodes between the networks were manually highlighted. We also indicated related functional sub-processes identified in these networks by individually analysing gene functional enrichment in each key BP. Finally, hubs of the BPs “regulation of cell proliferation” and “regulation of apoptosis” in the PPI networks were further categorised as positive or negative regulators of these processes using STRING.

## 3. Results

### 3.1. Gene co-expression networks

To study genetic interactions involved in osteoarthritis development at the transcriptional level, we analysed the public GSE114007 RNA-seq dataset [5]. In total, 13,102 genes were considered after filtering the data. Among them, 1085 are known human TFs, whereas 1332 genes were previously published as DEGs between normal and osteoarthritis [5, 16]. Among the TFs, 109 were also DEGs, a higher number than that found by Fisch et al. (2018), probably because we used a more recent list of known human TFs for comparison [16].

Differential gene co-expression analyses were performed with 13,102 genes, revealing 9,460,332 and 14, 289, 490 significant correlations in the normal and osteoarthritis groups, respectively. Among these correlations, 11.6% (normal) and 12.45% (OA) involve at least one DEG. To refine our differential co-expression analysis, the subsequent analyses were performed considering only the repertoire of co-expressions involving at least one DEG. After using this filter, the osteoarthritis co-expression network revealed 1.6 times more connectivity than the normal group.

### 3.2. Co-expression networks involving connections with DEGs

#### 3.2.1. Networks exclusivities

Two genes were identified as exclusive in the contrasting groups: *DDX43* in normal and *USP42* in osteoarthritis (Fig. 1A). *DDX43* and *USP42* connect with 27 and 67 genes in their networks, respectively (Tables S1 and S2). Among these, only *ZFP69* connects to both genes (Fig. 1A). Functional enrichment analyses revealed that only genes co-expressed with *USP42* are enriched in BPs and Reactome signalling pathways, as summarised in Table S3.

### 3.3. Identification of hub and hub centrality genes

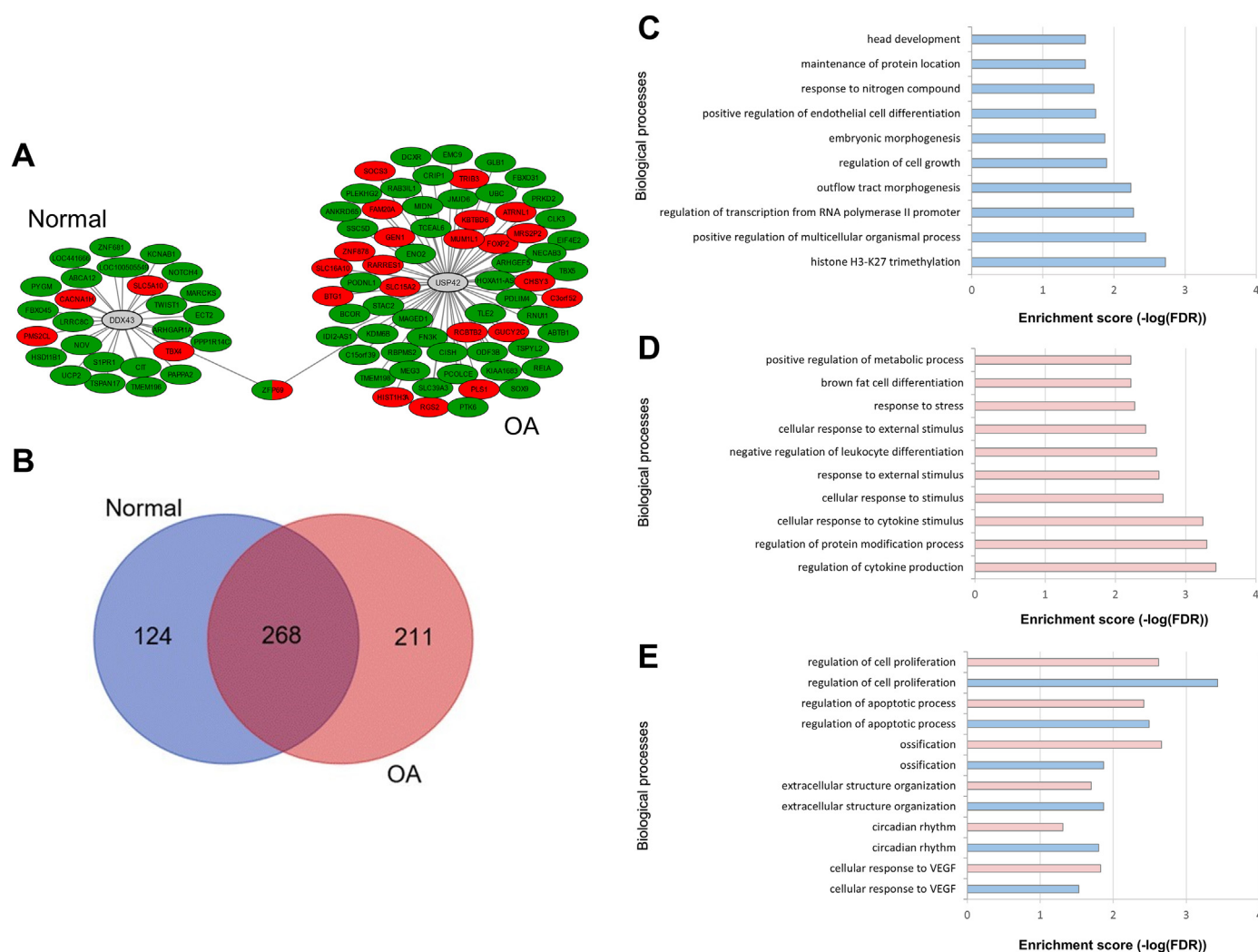
The normal and osteoarthritis co-expression networks are formed by 13,101 genes each, all correlated with at least one DEG. To identify the main components of the networks, we looked for the hub and hub centrality genes. In the normal group, 318 hubs and 365 hubs centrality were found, while in osteoarthritis, 372 hubs and 464 hubs centrality were

identified (Tables S4 and S5). In this regard, the comparison of the gene lists showed that more than half of the hubs and hubs centrality are shared between the groups (Fig. 1B, Table S6).

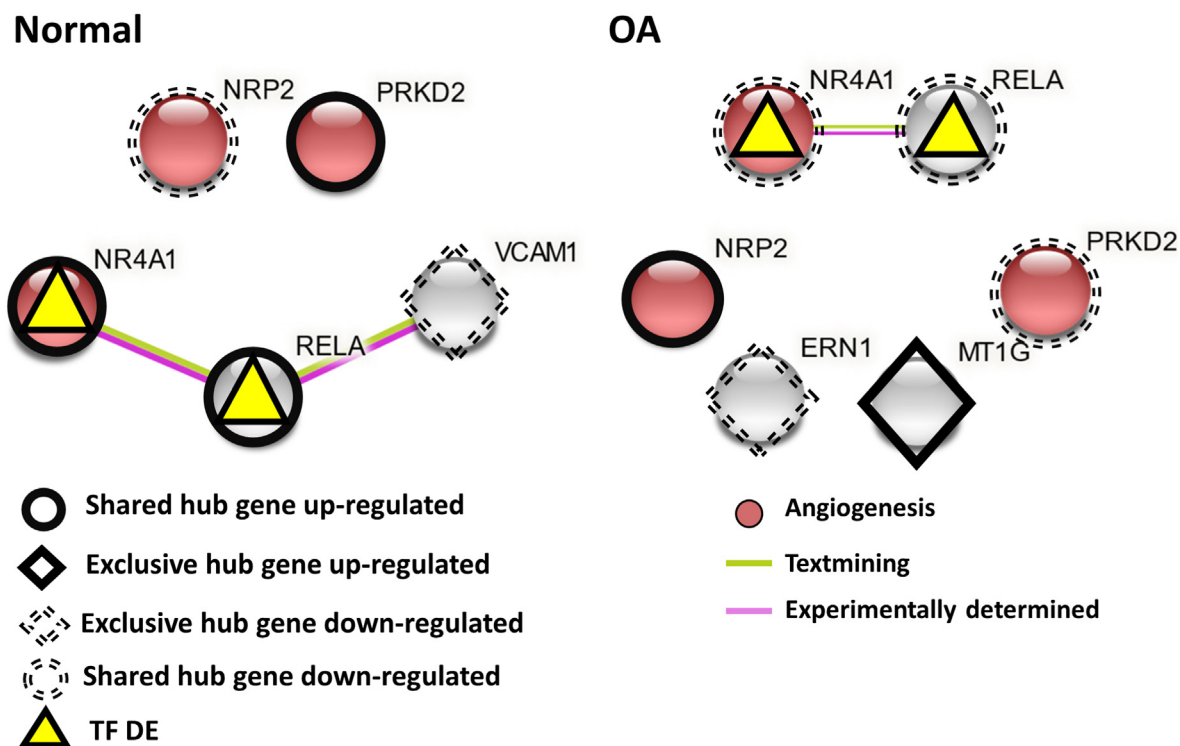
### 3.4. Enrichment analyses of hub and hub centrality genes

Our analyses identified several enriched biological processes in the normal and osteoarthritis groups (Fig. 1C and D; Tables S7-S10). Among these, the BPs: “histone H3–K27 trimethylation”, “positive regulation of multicellular organismal process”, and “regulation of transcription from RNA polymerase II promoter” are exclusive and showed the highest enrichment scores in the normal group (Fig. 1C). On the other hand, the BPs: “regulation of cytokine production”, “regulation of protein modification process” and “cellular response to cytokine stimulus” were found to be exclusive and have the highest enrichment scores in the osteoarthritis group (Fig. 1D).

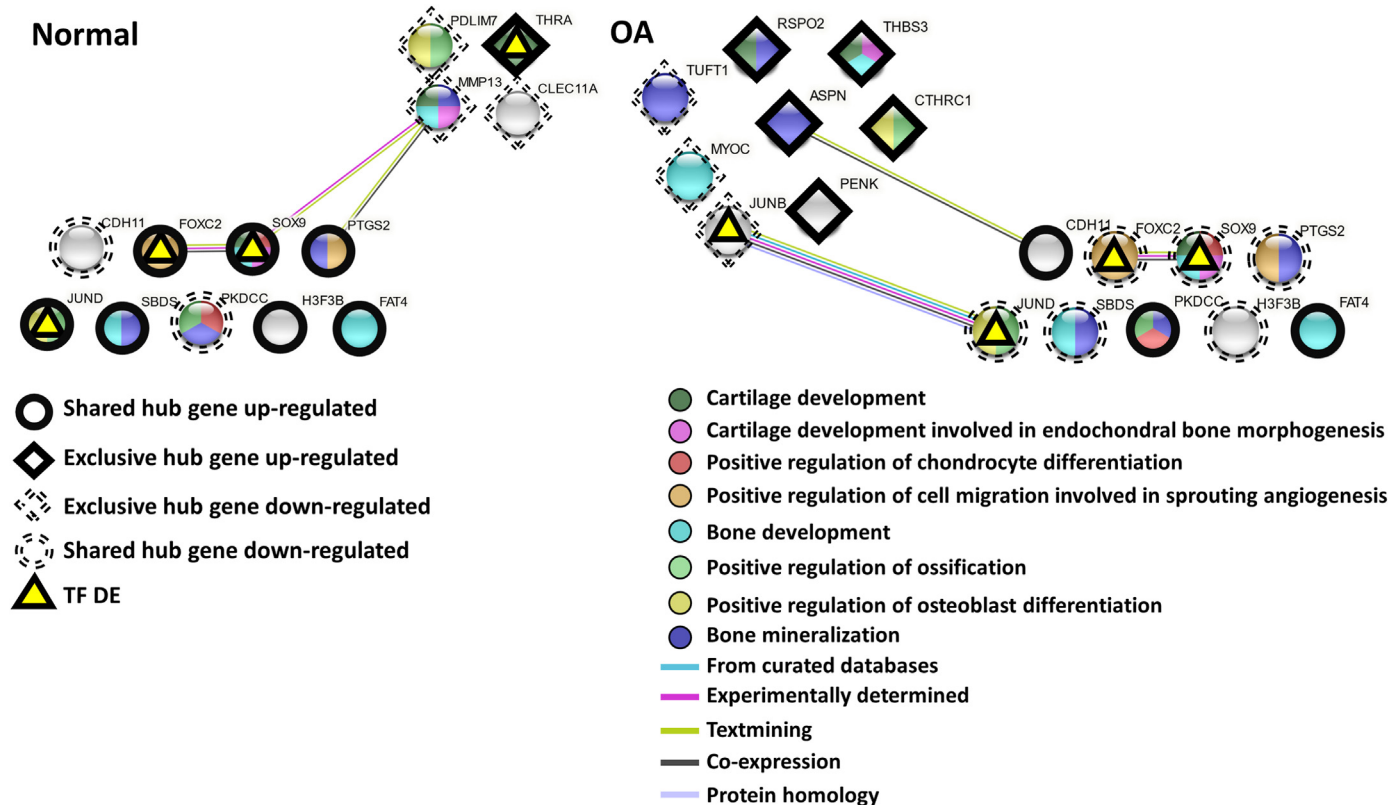
Additionally, we found several BPs enriched in both groups, denominated as “shared” BPs (Table S11). Interestingly, among them,



**Fig. 1.** Gene co-expression network and functional enrichment analysis for hub and hub centrality genes from normal and osteoarthritis groups. A) *DDX43* and *USP42* are exclusively co-expressed with differentially expressed genes in normal and osteoarthritis networks. Genes in red and green denote negative and positive correlations between genes, respectively. B) Venn diagram showing the number of exclusive and shared hub and hub centrality genes between the contrasting groups. C) Bar graph representing key biological processes exclusively enriched to healthy articular cartilage. D) Bar graph representing key biological processes exclusively enriched to osteoarthritis articular cartilage. E) Bar graph representing the key shared biological processes enriched in both normal and osteoarthritis cartilage known to be involved in disease onset and progression, which were used for in-depth investigations. Enrichment values were calculated using  $-\log_{10}$  false discovery rate - blue bars: Normal group; pink bars: osteoarthritis group. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 2.** Protein-protein interaction networks for the biological process “cellular response to vascular endothelial growth factor (VEGF) stimulation” (GO0035924). Genes that participate in related biological sub-processes were annotated in the network. Red nodes indicate genes associated with the angiogenesis sub-process, while white nodes indicate the absence of annotation in the listed sub-process. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 3.** Protein-protein interaction networks obtained for the biological process “ossification” (GO:0001503). Genes that participate in biological sub-processes that refer to specific stages of endochondral ossification (GO:0001958) were noted in the network in different colours. White nodes indicate absence of annotation in the listed sub-processes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



there are many relevant processes involved in the initiation and progression of osteoarthritis (Fig. 1E and Tables S12-13), including cellular “response to VEGF stimulus”, “circadian rhythm”, “ossification”, “extracellular structure organisation”, “regulation of cell proliferation and regulation of the apoptotic process” [3]. Therefore, to put into evidence co-expression alterations in these shared BPs, we focused our analyses on this subset of BPs. For that, PPI networks were constructed based on known interactions between each group's annotated hub and hub centrality genes.

### 3.5. Construction of PPI networks for the key shared BPs

The PPI networks for “cellular response to VEGF stimulation” are shown in Fig. 2. Some of the genes were also annotated for the functional subcategory of angiogenesis (Table S14). The networks contain four shared hubs, which include two TFs (*NR4A1* and *RELA*), both down-regulated in osteoarthritis. Regarding singularities, the healthy and diseased groups have one and two unique genes, respectively. However, there are no condition-exclusive TFs annotated.

The PPI networks for “ossification” are shown in Fig. 3. Each protein in the networks was colour-coded to indicate functional subcategories that represent specific stages of endochondral ossification (Table S14). These networks share nine hubs, among them, *SOX9*, *FOXC2*, and *JUND* are TFs down-regulated in osteoarthritis. As for exclusive genes, the normal group has four hubs, while osteoarthritis has twice as many exclusive hubs. *THRA* encodes the only exclusive TF in the normal group, which is up-regulated. Regarding OA-exclusive genes, *JUNB* encodes another down-regulated TF. Importantly, most of the exclusive hubs in the pathological state are annotated in processes that trigger the replacement of cartilage tissue by bone.

The PPI networks for “extracellular structure organisation” are shown in Fig. 4. Some genes were also annotated in functional subcategories related to the extracellular matrix environment (Table S14). The PPI networks from both groups share eight genes, including the TFs *FOXC2*, *SOX9*, and *NFKB2*, all down-regulated in osteoarthritis. Concerning exclusive genes, a similar amount was identified between the contrasting groups. Among these genes, it is important to highlight *ADAMTS5*, which is up-regulated in osteoarthritis.

The PPI networks for “circadian rhythm” are shown in Fig. 5. Functional subcategories related to the transcriptional and metabolic regulation of the circadian rhythm were also highlighted in the networks (Table S14). The PPI networks of the contrasting groups share seven hubs, including *JUND*, *SREBF1*, *NFIL3* and *NR1D1*, which encode TFs down-regulated in osteoarthritis. In the normal group, we found three unique genes, among them, the TFs *JUN* and *KLF9* were up-regulated. Conversely, in osteoarthritis, we identified three exclusive genes, including the TFs *PROX1* and *ARNTL*, down-regulated in this condition.

The PPI networks for “regulation of cell proliferation” are shown in Fig. 6. Although the osteoarthritis network has a greater number of hubs, more than half of the nodes are shared with the normal group, including the TFs *CEBPB*, *SOX9*, *RARA*, *BCL6*, *JUND*, *NR1D1*, *NR4A1* and *RELA*, which are down-regulated in the pathological condition. Regarding exclusivities, the normal group has 23 exclusive hubs, among which *KLF9*, *TFAP4*, *ATOH8*, *KLF11*, *HES1*, *TCF7* and *MXI1* encode up-regulated TFs. In turn, the osteoarthritis group has a greater number of exclusive hubs, *TFDP1* being the only up-regulated TF. Given the implications of positive and negative regulators in the articular cartilage maintenance mechanism, they were identified in the PPI networks of the contrasting groups (Table S14). Overall, our results showed a predominance of up-regulated positive (Fig. 6, blue nodes) and negative (Fig. 6, red nodes) regulators of

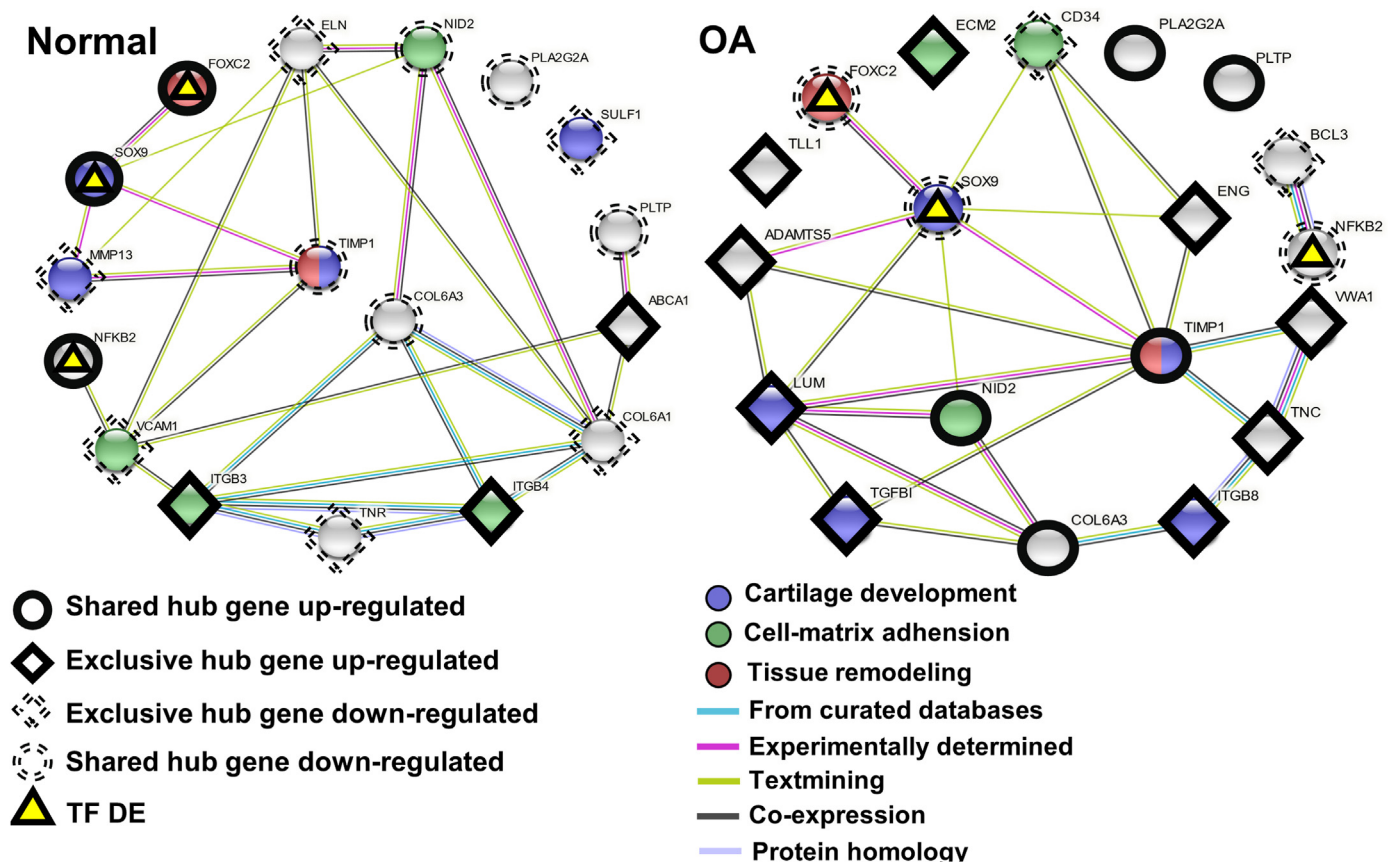
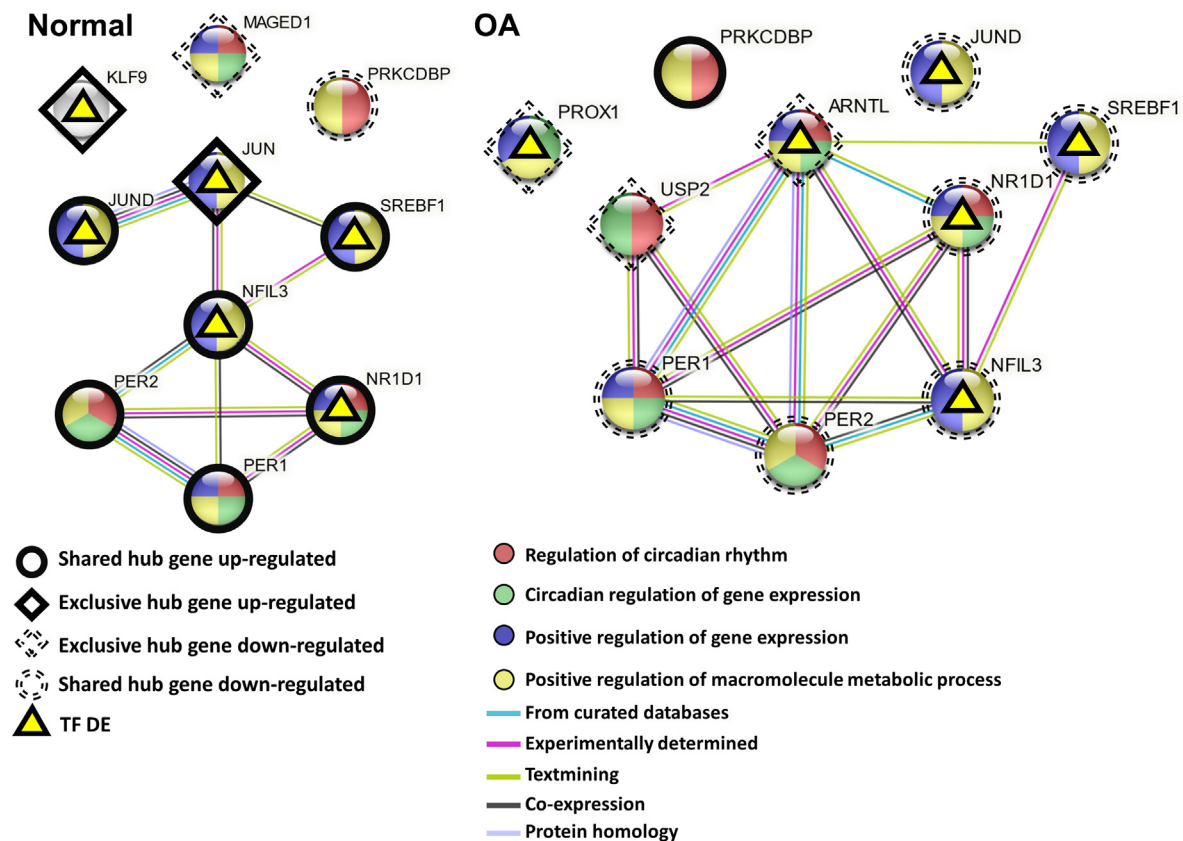


Fig. 4. Protein-protein interaction networks obtained for the biological process “extracellular structure organisation” (GO:0043,062). Genes that participate in other related biological sub-processes were noted on the network in different colours. White nodes indicate the absence of annotation in the listed sub-processes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 5.** Protein-protein interaction networks obtained for the biological process “circadian rhythm” (GO:0007623). Genes that participate in other related biological sub-processes were noted on the network in different colours. White nodes indicate the absence of annotation in the listed sub-processes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

cell proliferation in the normal group. In addition, we noted a third subcategory of genes as being able to act in both positive and negative control of proliferation (ambivalent). These genes were also mostly up-regulated in the normal group compared to osteoarthritis.

The PPI networks for “regulation of the apoptotic process” are shown in Fig. 7. The networks share more than half of the nodes, including the TFs *FOXC2*, *RARA*, *SOX9*, *RELA*, *CEBPB*, *BCL6* and *NKX3-2*, among which only the last one is up-regulated in osteoarthritic cartilage. Regarding the uniqueness of each network, the normal group has 16 exclusive hubs, including the TFs *TCF7*, *THRA*, *KLF11*, *TFAP4* and *JUN*, all up-regulated. Conversely, the osteoarthritis network contains more exclusive hubs, among which *TFDP1* is the only up-regulated TF. As described in the previous BP, the subcategories of positive and negative regulators of apoptosis were annotated in the networks. Our results revealed that nodes annotated as positive, negative or ambivalent regulators are mostly down-regulated in osteoarthritis compared to normal (Fig. 7; Table S14).

Interestingly, 41 genes were associated with more than one of the “shared” BPs (Table S15). Among these, 13 hubs that encode TFs must be highlighted, given their molecular implication on osteoarthritis development and progression (Fig. S1). *SOX9*, *FOXC2*, *RELA*, and *JUND* stand out from the other TFs showing connections to at least half of the BPs evaluated in this work, and also because they are all down-regulated in osteoarthritis. On the other hand, *TFDP1* stands out as the only up-regulated TF and for its two-side effect on cell cycle regulation. Moreover, there is a set of down-regulated TFs connected to cell proliferation and apoptosis (Fig. S1).

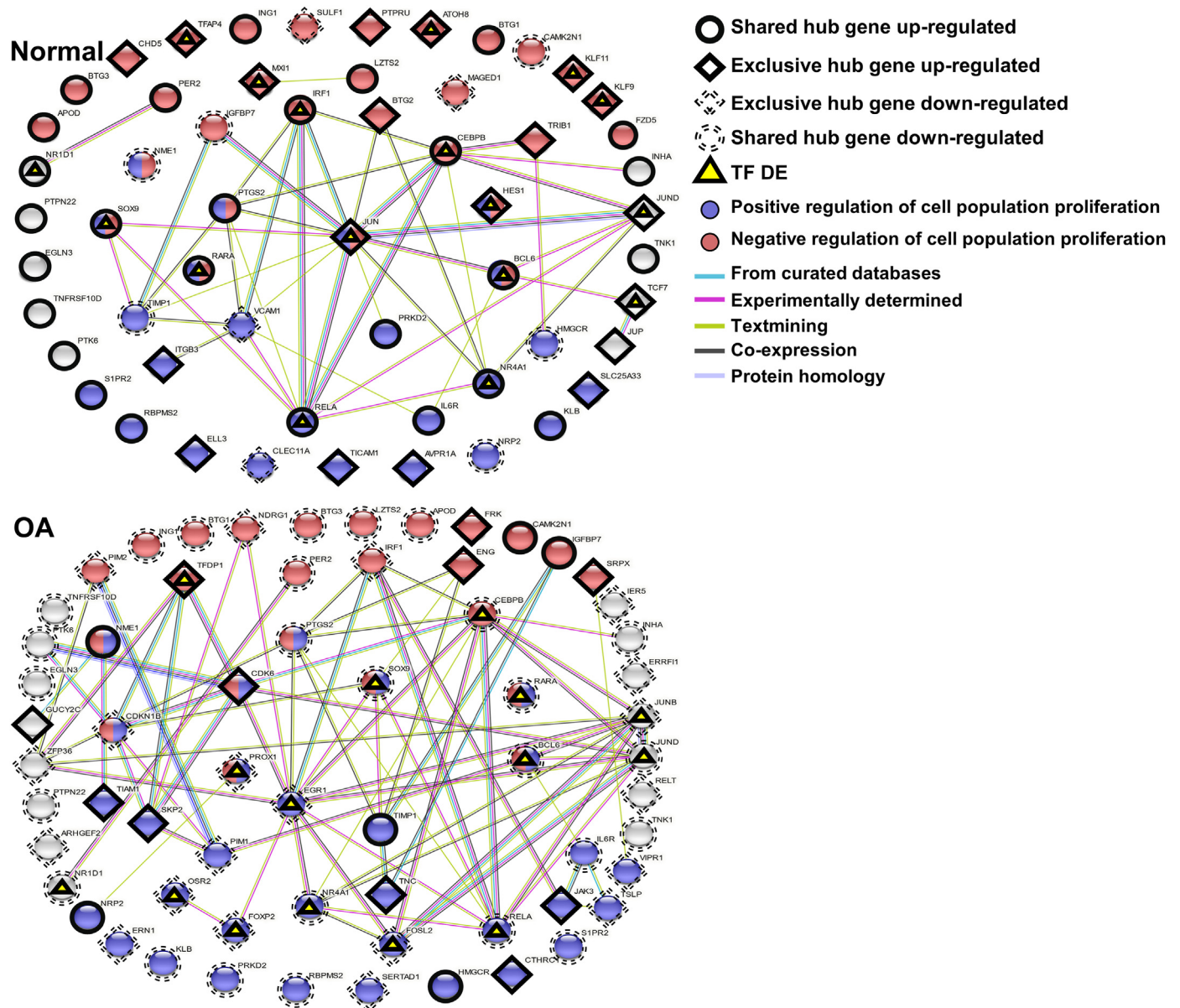
#### 4. Discussion

Here we compared gene co-expression patterns of healthy and

osteoarthritic knee cartilage using the PCIT algorithm. Multiple analyses were performed to identify changes in the co-expression networks and enriched BPs in healthy and osteoarthritis samples, to improve the knowledge about disease onset and progression. Our findings unveiled potential regulators involved in the misregulation of critical TFs known to occur in osteoarthritis cartilage [5]. We also show that some of these dysregulated TFs are connected to BPs involved in cartilage homeostasis, which are impaired in osteoarthritic cartilage.

The comparison of co-expression networks constructed around the DEGs of normal and osteoarthritis samples points to *DDX43* and *USP42* as candidate genes correlated with cartilage homeostasis versus osteoarthritis pathogenesis since their unique connection to DEGs in normal and osteoarthritis groups, respectively. Since *DDX43* is a DNA/RNA helicase and *USP42* is a deubiquitinase enzyme able to deubiquitinate histones, both molecules play critical roles in regulating transcriptional activity, affecting many biological processes [17,18]. In addition, *USP42* is associated with chondrocyte apoptosis and bone remodelling [19,20]. In the osteoarthritis co-expression network generated by our analysis, the genes connected to *USP42* participate in crucial signalling pathways related to disease aggravation. Therefore, we hypothesise that *USP42* exerts its functions in osteoarthritis by controlling gene expression and by modulating chondrocyte death, signalling pathways and inflammatory response, which are well-known promoters of osteoarthritis [21] [–] [23].

Analysis of differential gene co-expression networks approaches pathological states as a set of disrupted modules in a given network. Hub and hub centrality genes are essential constituents of co-expression networks, providing important information about their regulation and topology [24]. While hubs are the most connected genes, hub centrality genes work as bridges between separate clusters of nodes in a network. Focusing on these genes simplifies the comparison between contrasting



**Fig. 6.** Protein-protein interaction networks obtained for the biological process “regulation of cell proliferation” (GO:0042,127). Genes that participate in the positive (blue) and negative (red) control of proliferation were noted on the network. White nodes indicate the absence of annotation in control of proliferation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

groups. In addition, hub and hub centrality genes are more likely to be associated with the disease [25].

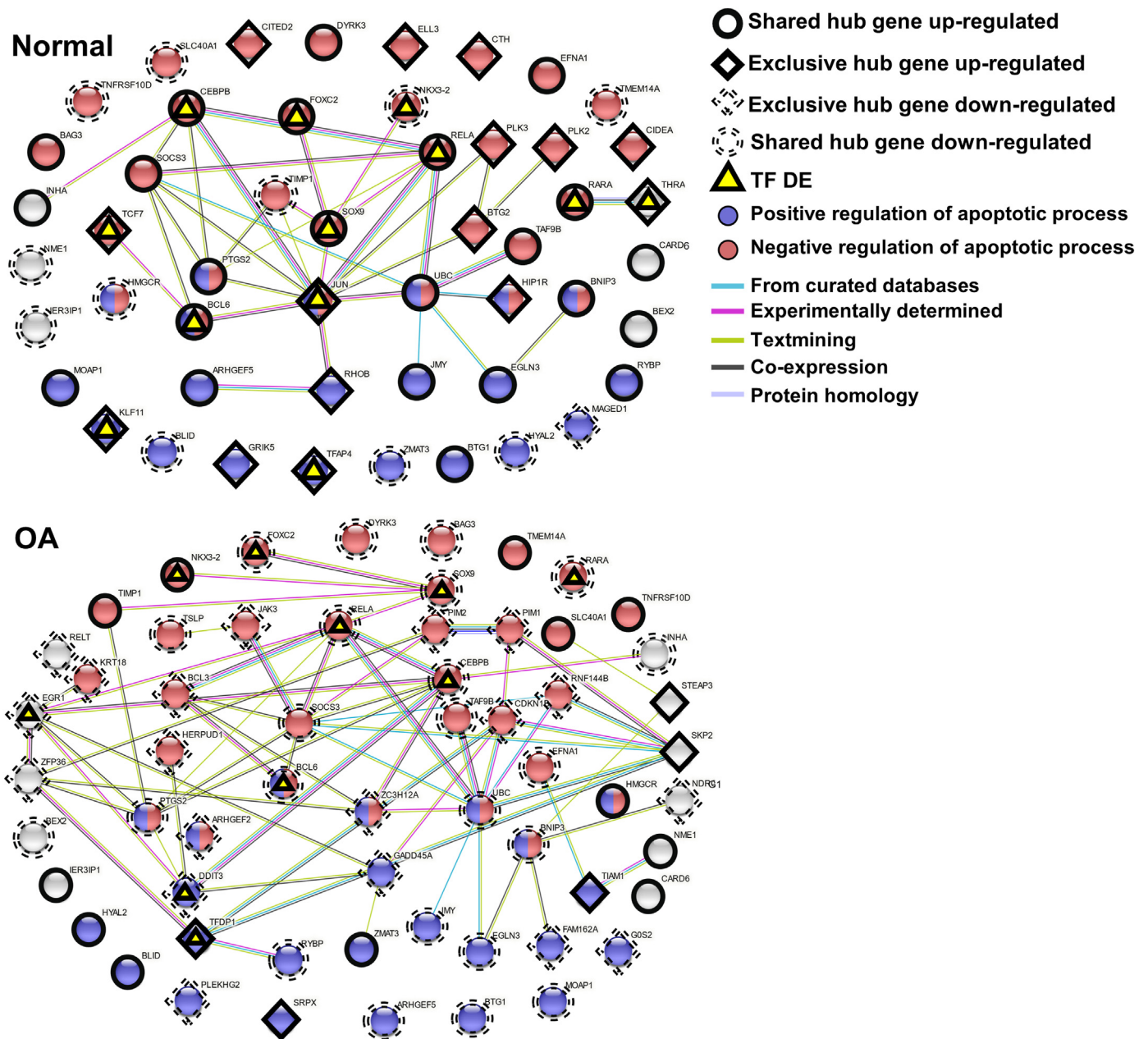
The top hubs for the normal and osteoarthritis groups are *HIST1H3A* and *TAF9B*, while the top hubs centrality are *SNHG12* and *OTUD1*, respectively. Concerning the top hubs, *HIST1H3A* encodes histone H3, a protein that makes up the nucleosome histone octamer, a fundamental component in chromatin compaction [26]. On the other hand, *TAF9B* encodes a subunit of the TBP-associated factors that, together with TATA-binding protein, form the TFIID complex, which is essential for RNA polymerase II assembly and transcription initiation [27]. In addition, *TAF9B* was shown to interact with the SAGA-like complex, which is involved in histone acetylation and deubiquitination [28]. Concerning the top hubs centrality, *SNHG12* encodes a long non-coding RNA that acts as a molecular sponge to a plethora of microRNAs, and it was recently found to down-regulating the miR-16-5p, consequently promoting osteoarthritis onset [29,30]. In turn, *OTUD1* encodes a deubiquitinase enzyme that cleaves ubiquitin linkages to prevent protein degradation [31]. Importantly, all of these genes are down-regulated in

the osteoarthritis group. Given that *HIST1H3A* and *TAF9B* are pivotal molecules in controlling gene transcription, whereas *SNHG12* and *OTUD1* target microRNAs and proteins, respectively, our findings point to impairment at different levels of gene expression control in osteoarthritic cartilage.

As our data suggested that changes in diseased cartilage are related to a global misregulation of gene expression, we wondered whether this might lead to differential gene co-expression patterns between normal and osteoarthritis samples. To assess this, we compared the PPI networks of shared BPs involved in cartilage homeostasis and osteoarthritis onset and progression [3,32].

Starting with “cellular response to VEGF stimulation”, our results showed that two TFs (*NR4A1* and *RELA*) shared between normal and osteoarthritis networks are down-regulated in the latter. *NR4A1* induces chondrocyte apoptosis through different signalling pathways [33]. In addition, this molecule is an endogenous inhibitor of chondrocyte inflammation that conversely is inhibited by chronic inflammation, which leads to collagenase up-regulation and cartilaginous extracellular





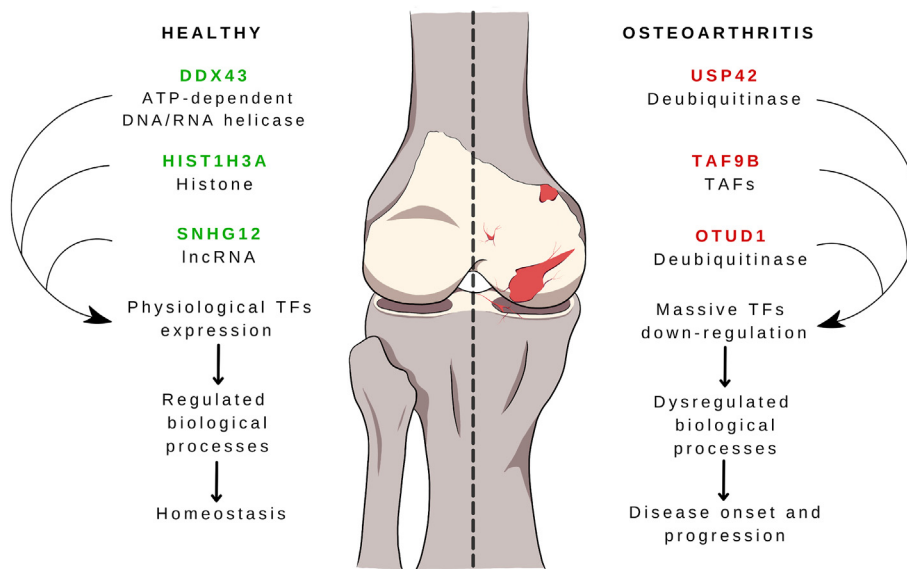
**Fig. 7.** Protein-protein interaction networks obtained for the biological process “regulation of apoptotic process” (GO:0042,127). Genes that participate in the positive (blue) and negative (red) control of proliferation were noted on the network. White nodes indicate the absence of annotation in control of apoptosis. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

matrix degradation [34]. In turn, RELA protein (p65) constitutes the NF- $\kappa$ B heterodimer and was previously reported as a high-priority candidate for therapeutic intervention since it regulates many DEG in osteoarthritis [5,35]. Classic activation of the NF- $\kappa$ B signalling pathway has been reported to have antagonistic effects on osteoarthritis, suppressing or increasing cartilage joint destruction [35]. Given RELA properties, its down-regulation in osteoarthritis possibly favours the mechanisms that lead to cartilage structural alterations.

Concerning “ossification”, although it is not ordinarily observed in healthy articular cartilage, several DEGs belonging to this BP are also crucial for tissue maintenance [36]. Among the shared genes, *SOX9*, *FOXC2* and *JUND* encode TFs down-regulated in the osteoarthritis group. *SOX9* is an essential TF that controls chondrogenesis and postnatal cartilage tissue maintenance, and its sustained expression inhibits endochondral ossification [37]. Therefore, *SOX9* down-regulation in the

osteoarthritis group is expected and has already been reported [38]. In turn, *FOXC2* may act as a pioneer TF that opens the DNA on super-enhancers of cartilage-specific genes facilitating the access of *SOX9* and *SOX5/6* [39]. Finally, *JUND* encodes an AP-1 subunit, and its deficiency causes a pro-osteogenic phenotype in mutant mice through the induction of pro-osteogenic genes [40]. Accordingly, *JUND* binding sites were found in promoters of many DEGs in osteoarthritic cartilage [5]. Altogether, our data point to *SOX9*, *FOXC2* and *JUND* as key molecules involved in the regulatory imbalance associated with ectopic ossification of articular cartilage. Therefore, these genes are clinically relevant candidates to be prioritised for therapeutic interventions.

Regarding “extracellular structure organisation”, our findings reinforce that the transcriptional core governed by *SOX9* and *FOXC2* is compromised in osteoarthritis, as these two TFs are down-regulated in the pathological condition. One of the genes known to be regulated by



**Fig. 8.** Novel candidate genes associated with osteoarthritis and its biological implications. A) In the healthy cartilage, DDX43 DNA/RNA helicase and HIST1H3A histone contribute to maintaining tissue homeostasis by opening the road to proteins involved in fine tuning of gene expression, such as transcription factors. The SNHG12 long non-coding RNA seems to provide an additional gene expression level control by targeting specific miRNAs involved in regulation of mRNA stability and protein translation in cartilage. B) In the disease, the USP42 and OTUD1 deubiquitinating enzymes and TAF9B/SAGA complex may promote epigenetic modifications implicated in the global down-regulation of transcription factors and molecules related to osteoarthritis onset and progression.

*SOX9* is *ADAMTS5* ( $FC > 1.9$ ), a gene unique to the osteoarthritis network, previously described as clinically relevant for degrading extracellular matrix aggrecans [41]. During the onset of osteoarthritis, *ADAMTS5* is repressed by increased levels of *SOX9*. However, *SOX9* levels decay during disease evolution and, consequently, promotes *ADAMTS5* expression [40].

The “circadian rhythm” is also enriched in both groups. *ARNTL*, which initiates the circadian rhythm by activating target genes, is an exclusive hub to the osteoarthritis group and is down-regulated. *ARNTL* down-regulation was already reported for this disease [42] and may have important effects on circadian rhythm in cartilage by affecting regulatory genes that act on the negative limb of this process, such as *PER1-PER2* and *NR1D1* [43]. Another down-regulated molecule in osteoarthritis is *NFIL3*, which competes with activating proteins for the D-box site in promoters of target genes, thus inhibiting their binding. One of these target genes is *ROR*, its activator, which also participates in *ARNTL* activation [44]. In this context, the *PROX1* TF, exclusive to the osteoarthritis network, interacts with *RORα/γ* to inhibit the transcription of central circadian clock components [45]. Overall, our findings uncover a critical down-regulation of master regulators of circadian rhythm in osteoarthritis, which possibly impairs chondroprotective pathways.

Concerning “regulation of cell proliferation” and “regulation of apoptotic process”, we observed that the networks of normal and osteoarthritis groups present several common hubs, including TFs. As far as proliferation is concerned, the osteoarthritis network also presents many unique TFs, mostly down-regulated. On the other hand, apoptosis shows less exclusive TFs in the osteoarthritis network, which are also mostly down-regulated. Among them, we highlight *EGR1*, a gene that has been suggested as clinically relevant for osteoarthritis prevention and treatment [46]. *EGR1* is essential for mitogenesis and is expressed in all zones of healthy articular cartilage, whereas osteoarthritis significantly reduces its expression [46]. Our findings showed that *EGR1* is down-regulated in the osteoarthritis group ( $FC < -1.79$ ) and, most importantly, uncovered the role of this TF as a unique central hub for both proliferation and apoptosis osteoarthritis networks. Although the positive role of *EGR1* in proliferation is well established, its participation in apoptosis remains to be further evaluated. Another noteworthy gene is the transcription factor *TFDP1*. Its encoded protein does not have transcriptional action by itself, being a co-modulator of E2F family proteins, forming heterodimeric complexes with different functions [47,48]. Interestingly, in the absence of E2F1, *TFDP1* becomes polyubiquitinated and accumulates in the cytoplasm along with other proteins. This accumulation delays cell cycle advancement until these protein clusters are fully degraded [49]. In

apoptosis, the E2F1-*TFDP1* complex induces pro-apoptotic events dependent and independent of the p53 protein [48]. Our work found that *TFDP1* is the only TF that is an exclusive and up-regulated hub of the proliferation and apoptosis osteoarthritis networks. Importantly, this TF is a negative regulator of the proliferation process and, as expected, is a positive regulator of apoptosis. Finally, we highlight the TF *DDIT3*, which was already associated with osteoclastogenesis. Its absence enhances osteoclast formation and aggravation of bone resorption, a hallmark in osteoarthritic subchondral bone [50]. We found that this gene is exclusive to the apoptosis osteoarthritis network, indicating that *DDIT3* might play an important role in the cartilage-to-bone transition. Our findings further corroborate that the dysregulation of *EGR1*, *TFDP1* and *DDIT3* may have clinical relevance for osteoarthritis due to their role in cartilage proliferation and apoptosis.

From the analysis of all the BPs evaluated herein, *SOX9*, *FOXC2*, *RELA*, and *JUND* transcription factors emerged as remarkable candidates for further investigations in osteoarthritis, given that they integrate multiple BPs and, therefore, may have broader impacts on cartilage homeostasis.

Remarkably, we evidenced an additional layer of transcriptional regulation in which TFs may not be the only players. The novel candidate genes identified here seem to participate in a putative epigenetic mechanism, leading to the massive gene down-regulation in late-stage osteoarthritis. Among these molecules are proteins important for chromatin organisation and modification, as summarised in Fig. 8. Therefore, our work points to molecules whose involvement in osteoarthritis has not been addressed before.

Regarding the limitations of our work, it is noteworthy that some differences in gene co-expression identified here may be owing to cartilage ageing, given the inherent average age difference between the contrasting groups in the dataset. Also, since our work was focused on bioinformatic analysis, it is important to evaluate gene and protein expression in healthy and diseased cartilage to validate the involvement of the novel candidate genes identified here. In addition, *in vitro* models for the osteoarthritis study could be used to confirm the coordination of gene expression predicted herein by *in silico* analysis. If this co-expression relationship is confirmed, these genes should be considered strong targets for therapeutic interventions, given that the modulation of one of them will result in an overarching effect on expression status. Finally, genetically modified model organisms (knockin and knockout) may be a valuable resource to investigate whether these molecules play a role in epigenetic events that modulate gene transcription in diseased cartilage and establish their therapeutic potential.

In conclusion, our work corroborates previous reports and provides important and new information about the molecular basis of osteoarthritis pathogenesis, expanding the repertoire of known candidate genes for therapeutic interventions. In addition, our findings open perspectives for future research that will provide subsidies for developing preventive and less invasive therapies, which may replace the surgical interventions used today as the only available definitive treatment for this disease.

## Author contributions

IBL participated in the gene functional enrichment analysis, data interpretation and manuscript writing. JA participated in the dataset selection in the PCIT and differential gene co-expression network analyses. BSV participated in the selection of the dataset and gene functional enrichment analyses. LLC participated in the design of the study and coordination of bioinformatic analysis. LEA participated in the design, data analysis, manuscript writing and coordination of the study. All authors revised and approved the final manuscript.

## Declaration of competing interest

The authors have declared no conflict of interest.

## Acknowledgements

This research received financial support from the National Council for Scientific Technological Development (CNPq, Brasília, Brazil), which provided fellowships to Igor Buzzatto-Leite and Luiz Lehmann Coutinho (grant numbers: 131275/2019–4 and 304353/2019–1, respectively); the Coordination for the Improvement of Higher Education Personnel (CAPES, Brasília, Brazil), which provided fellowship to Juliana Afonso (grant number: 88887.473735/2020–00), and the São Paulo Research Foundation (FAPESP, São Paulo, Brazil), which provided fellowship to Bárbara Silva-Vignato (grant number: 2019/18385–2).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ocarto.2022.100316>.

## References

- [1] A.A. Pitsillides, F. Beier, Cartilage biology in osteoarthritis—lessons from developmental biology, *Nat. Rev. Rheumatol.* 7 (2011) 654–663.
- [2] S. Grässel, A. Aszodi, Osteoarthritis and cartilage regeneration: focus on pathophysiology and molecular mechanisms, *Int. J. Mol. Sci.* 20 (2019) 6156.
- [3] J. Martel-Pelletier, et al., Osteoarthritis, *Nat. Rev. Dis. Prim.* 2 (2016) 1–18.
- [4] Y.A. Rim, Y. Nam, J.H. Ju, The role of chondrocyte hypertrophy and senescence in osteoarthritis initiation and progression, *Int. J. Mol. Sci.* 21 (2020) 2358.
- [5] K.M. Fisch, et al., Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis, *Osteoarthritis Cartilage* 26 (2018) 1531–1538.
- [6] A.-L. Barabási, Z.N. Oltvai, Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.* 5 (2004) 101–113.
- [7] A. de la Fuente, From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases, *Trends Genet.* TIG 26 (2010) 326–333.
- [8] S. van Dam, U. Vösa, A. van der Graaf, L. Franke, J.P. de Magalhães, Gene co-expression analysis for functional classification and gene-disease predictions, *Briefings Bioinf.* 19 (2018) 575–592.
- [9] A. Reverter, E.K.F. Chan, Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks, *Bioinformatics* 24 (2008) 2491–2497.
- [10] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (2008) 559.
- [11] D.J. McCarthy, Y. Chen, G.K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation, *Nucleic Acids Res.* 40 (2012) 4288–4297.
- [12] P. Shannon, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504.
- [13] Y. Assenov, F. Ramírez, S.-E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, *Bioinforma. Oxf. Engl.* 24 (2008) 282–284.
- [14] D. Szklarczyk, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.* 47 (2019) D607–D613.
- [15] F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of gene ontology terms, *PLoS One* 6 (2011), e21800.
- [16] S.A. Lambert, et al., The human transcription factors, *Cell* 172 (2018) 650–665.
- [17] M.K. Ghosh, DEAD box RNA helicases crucial regulators of gene expression and oncogenesis, *Front. Biosci.* 21 (2016) 225–250.
- [18] A.K. Hock, A.M. Vigneron, K.H. Vousden, Ubiquitin-specific peptidase 42 (USP42) functions to deubiquitylate histones and regulate transcriptional activity, *J. Biol. Chem.* 289 (2014) 34862–34870.
- [19] S. Hashimoto, et al., Role of p53 in human chondrocyte apoptosis in response to shear strain, *Arthritis Rheum.* 60 (2009) 2340–2349.
- [20] Y.-C. Guo, S.-W. Zhang, Q. Yuan, Deubiquitinating enzymes and bone remodeling, *Stem Cell. Int.* 2018 (2018), 3712083.
- [21] K. Hou, et al., Overexpression and biological function of ubiquitin-specific protease 42 in gastric cancer, *PLoS One* 11 (2016), e0152997.
- [22] H.-B. Park, J.-W. Kim, K.-H. Baek, Regulation of wnt signaling through ubiquitination and deubiquitination in cancers, *Int. J. Mol. Sci.* 21 (2020).
- [23] F. Wang, Z. Guo, Y. Yuan, STAT3 speeds up progression of osteoarthritis through NF- $\kappa$ B signaling pathway, *Exp. Ther. Med.* 19 (2020) 722–728.
- [24] P. Goymer, Why do we need hubs? *Nat. Rev. Genet.* 9 (2008), 651–651.
- [25] T. Arodz, D. Bonchev, R.F. Diegelmann, A network approach to wound healing, *Adv. Wound Care* 2 (2013) 499–509.
- [26] J.L. Workman, R.E. Kingston, Alteration of nucleosome structure as a mechanism of transcriptional regulation, *Annu. Rev. Biochem.* 67 (1998) 545–579.
- [27] M. Frontini, et al., TAF9b (formerly TAF9L) is a bona fide TAF that has unique and overlapping roles with TAF9, *Mol. Cell Biol.* 25 (2005) 4638–4649.
- [28] F.J. Herrera, T. Yamaguchi, H. Roelink, R. Tjian, Core promoter factor TAF9B regulates neuronal gene expression, *Elife* 3 (2014), e02559.
- [29] H. Zhang, W. Lu, LncRNA SNHG12 regulates gastric cancer progression by acting as a molecular sponge of miR-320, *Mol. Med. Rep.* (2017), <https://doi.org/10.3892/mmr.2017.8143>.
- [30] X. Yang, et al., LncRNA SNHG12 promotes osteoarthritis progression through targeted down-regulation of miR-16-5p, *Clin. Lab.* 68 (2022).
- [31] Z. Zhang, et al., Breast cancer metastasis suppressor OTUD1 deubiquitinates SMAD7, *Nat. Commun.* 8 (2017) 2116.
- [32] D. Primorac, et al., Knee osteoarthritis: a review of pathogenesis and state-of-the-art non-operative therapeutic considerations, *Genes* 11 (2020) 854.
- [33] X. Shi, H. Ye, X. Yao, Y. Gao, The involvement and possible mechanism of NR4A1 in chondrocyte apoptosis during osteoarthritis, *Am. J. Transl. Res.* 9 (2017) 746–754.
- [34] Y. Xiong, et al., Reactivation of NR4A1 restrains chondrocyte inflammation and ameliorates osteoarthritis in rats, *Front. Cell Dev. Biol.* 8 (2020) 158.
- [35] H. Kobayashi, et al., Biphasic regulation of chondrocytes by RelA through induction of anti-apoptotic and catabolic target genes, *Nat. Commun.* 7 (2016), 13336.
- [36] K.A. Staines, A.S. Pollard, I.M. McGonnell, C. Farquharson, A.A. Pitsillides, Cartilage to bone transitions in health and disease, *J. Endocrinol.* 219 (2013) R1–R12.
- [37] S.P. Henry, S. Liang, K.C. Akdemir, B. de Crombrughe, The postnatal role of Sox9 in cartilage, *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* 27 (2012) 2511–2525.
- [38] J. Haag, P.M. Gebhard, T. Aigner, SOX gene expression in human osteoarthritic cartilage, *Pathobiol. J. Immunopathol. Mol. Cell Biol.* 75 (2008) 195–199.
- [39] C.-F. Liu, V. Lefebvre, The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis, *Nucleic Acids Res.* 43 (2015) 8183–8203.
- [40] A. Kawamata, et al., JunD suppresses bone formation and contributes to low bone mass induced by estrogen depletion, *J. Cell. Biochem.* 103 (2008) 1037–1045.
- [41] Q. Zhang, et al., SOX9 is a regulator of ADAMTS-induced cartilage degeneration at the early stage of human osteoarthritis, *Osteoarthritis Cartilage* 23 (2015) 2259–2268.
- [42] M. Dudek, et al., The chondrocyte clock gene Bmal1 controls cartilage homeostasis and integrity, *J. Clin. Invest.* 126 (2016) 365–376.
- [43] R. Ikeda, et al., REV-ERB $\alpha$  and REV-ERB $\beta$  function as key factors regulating Mammalian Circadian Output, *Sci. Rep.* 9 (2019), 10171.
- [44] A. Patke, M.W. Young, S. Axelrod, Molecular mechanisms and physiological importance of circadian rhythms, *Nat. Rev. Mol. Cell Biol.* 21 (2020) 67–84.
- [45] Y. Takeda, A.M. Jetten, Prospero-related homeobox 1 (Prox1) functions as a novel modulator of retinoic acid-related orphan receptors  $\alpha$ - and  $\gamma$ -mediated transactivation, *Nucleic Acids Res.* 41 (2013) 6992–7008.
- [46] F.-L. Wang, et al., Differential expression of Egr-1 in osteoarthritic compared to normal adult human articular cartilage, *Osteoarthritis Cartilage* 8 (2000) 161–169.
- [47] L.R. Bandara, V.M. Buck, M. Zamanian, L.H. Johnston, N.B. La Thangue, Functional synergy between DP-1 and E2F-1 in the cell cycle-regulating transcription factor DRTF1/E2F, *EMBO J.* 12 (1993) 4317–4324.
- [48] M.R. Hitchens, P.D. Robbins, The role of the transcription factor DP in apoptosis, *Apoptosis Int. J. Prog. Cell Death* 8 (2003) 461–468.
- [49] J. Magae, S. Illenye, Y.C. Chang, Y. Mitsui, N.H. Heintz, Association with E2F-1 governs intracellular trafficking and polyubiquitination of DP-1, *Oncogene* 18 (1999) 593–605.
- [50] B. Yang, et al., DNA damage-inducible transcript 3 restrains osteoclast differentiation and function, *Bone* 153 (2021), 116162.