



Universiteit
Leiden
The Netherlands

Elucidating epigenetic regulation by identifying functional cis-acting long noncoding RNAs and their targets in osteoarthritic articular cartilage

Hoolwerff, M. van; Metselaar, P.I.; Tuerlings, M.; Suchiman, H.E.D.; Lakenberg, N.; Ramos, Y.F.M.; ... ; Meulenbelt, I.

Citation

Hoolwerff, M. van, Metselaar, P. I., Tuerlings, M., Suchiman, H. E. D., Lakenberg, N., Ramos, Y. F. M., ... Meulenbelt, I. (2020). Elucidating epigenetic regulation by identifying functional cis-acting long noncoding RNAs and their targets in osteoarthritic articular cartilage. *Arthritis And Rheumatology*, 72(11), 1845-1854. doi:10.1002/art.41396






Version: Publisher's Version

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3182204>

Note: To cite this publication please use the final published version (if applicable).

Elucidating Epigenetic Regulation by Identifying Functional *cis*-Acting Long Noncoding RNAs and Their Targets in Osteoarthritic Articular Cartilage

Marcella van Hoolwerff,  Paula I. Metselaar, Margo Tuerlings,  H. Eka D. Suchiman, Nico Lakenberg, Yolande F. M. Ramos, Davy Cats, Rob G. H. H. Nelissen, Demiën Broekhuis,  Hailiang Mei, Rodrigo Coutinho de Almeida,  and Ingrid Meulenbelt 

Objective. To identify robustly differentially expressed long noncoding RNAs (lncRNAs) with osteoarthritis (OA) pathophysiology in cartilage and to explore potential target messenger RNA (mRNA) by establishing coexpression networks, followed by functional validation.

Methods. RNA sequencing was performed on macroscopically lesioned and preserved OA cartilage from patients who underwent joint replacement surgery due to OA ($n = 98$). Differential expression analysis was performed on lncRNAs that were annotated in GENCODE and Ensembl databases. To identify potential interactions, correlations were calculated between the identified differentially expressed lncRNAs and the previously reported differentially expressed protein-coding genes in the same samples. Modulation of chondrocyte lncRNA expression was achieved using locked nucleic acid GapmeRs.

Results. By applying our in-house pipeline, we identified 5,053 lncRNAs that were robustly expressed, of which 191 were significantly differentially expressed (according to false discovery rate) between lesioned and preserved OA cartilage. Upon integrating mRNA sequencing data, we showed that intergenic and antisense differentially expressed lncRNAs demonstrate high, positive correlations with their respective flanking or sense genes. To functionally validate this observation, we selected *P3H2-AS1*, which was down-regulated in primary chondrocytes, resulting in the down-regulation of *P3H2* gene expression levels. As such, we can confirm that *P3H2-AS1* regulates its sense gene *P3H2*.

Conclusion. By applying an improved detection strategy, robustly differentially expressed lncRNAs in OA cartilage were detected. Integration of these lncRNAs with differential mRNA expression levels in the same samples provided insight into their regulatory networks. Our data indicate that intergenic and antisense lncRNAs play an important role in regulating the pathophysiology of OA.

INTRODUCTION

Osteoarthritis (OA) is an age-related, heterogeneous, degenerative disease of the articular joints, characterized in part by cartilage degeneration and remodeling of subchondral bone, which results in stiff and painful joints and decreased mobility (1). Despite the fact that OA is the most globally prevalent joint disease, no effective treatment is currently available (2). It has been demonstrated that OA pathophysiology in cartilage is marked by altered gene expression regulation in chondrocytes (3,4). This alteration

of gene expression regulation could be triggered by adaptation processes occurring due to aging, genetic predisposition, or environmental stimuli, and is in part caused by aberrant epigenetic mechanisms. These mechanisms include DNA methylation, histone modifications, and expression of microRNAs (<22 nucleotides) (4–6). More recently, long noncoding RNAs (lncRNAs; >200 nucleotides) have been shown to play an important role in the homeostasis of the extracellular matrix of cartilage (5,7–10).

lncRNAs are defined as RNA transcripts with little or no protein-coding potential and are known to regulate transcription

Supported by the Dutch Scientific Research Council (ZonMW Vici grant 91816631/528), the Dutch Arthritis Society (grants DAA-10-1-402 and DAF-16-1-405), and TREAT-OA (grant 200800 from the European Commission Seventh Framework Program).

Marcella van Hoolwerff, MSc, Paula I. Metselaar, MSc, Margo Tuerlings, MSc, H. Eka D. Suchiman, MSc, Nico Lakenberg, BSc, Yolande F. M. Ramos, PhD, Davy Cats, BSc, Rob G. H. H. Nelissen, MD, PhD, Demiën Broekhuis, MD, Hailiang Mei, PhD, Rodrigo Coutinho de Almeida, PhD,

Ingrid Meulenbelt, PhD: Leiden University Medical Center, Leiden, The Netherlands.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Ingrid Meulenbelt, PhD, Leiden University Medical Center, Department of Biomedical Data Sciences, Section of Molecular Epidemiology, LUMC Post-zone S-05-P, PO Box 9600, 2300 RC Leiden, The Netherlands. Email: i.meulenbelt@lumc.nl.

Submitted for publication April 8, 2020; accepted in revised form June 4, 2020.

and translation by numerous mechanisms, such as chromatin remodeling, messenger RNA (mRNA) stabilization, microRNA modulation, and recruitment of scaffolding proteins. One classification type of lncRNAs is based on the genomic location with respect to protein-coding genes, so-called biotypes, including antisense RNAs, sense RNAs, pseudogenes, and long intergenic noncoding RNAs (lincRNAs). Another type of classification is based on the location at which the lncRNA functions relative to its transcription site, which can be in *trans* or *cis* (11–13). *Cis*-acting lncRNAs comprise a considerable portion of known lncRNAs and can be positioned at various distances and orientations relative to their target genes, such as lincRNAs around transcription factor start sites, as well as sense and antisense lncRNAs that overlap with their sense genes (13,14). Potentially, lncRNAs could be candidate targets in OA treatment, since their expression can be highly tissue specific (9).

RNA sequencing (RNA-Seq) has improved the ability to detect lncRNAs, but mapping and annotating lncRNAs remains challenging. These challenges arise from the fact that they are usually expressed at very low levels and their sequence–function relationship is still poorly understood. Moreover, recent findings from studies on ribosome profiling and bioinformatics suggest that a large proportion of transcripts has unknown coding potential (15). Recent studies on OA have focused on intergenic lncRNAs, even though the proportion of genic and intergenic lncRNAs can be similar depending on the investigated tissue (15,16). To determine the complete lncRNA transcriptome, we used an in-house pipeline to robustly capture lncRNAs in a previously assessed RNA-Seq data set of lesioned and preserved OA cartilage samples (4). Subsequently, lncRNAs associated with OA pathophysiology were identified, and potential interactions with OA-specific mRNAs were investigated.

MATERIALS AND METHODS

Sample collection. Macroscopically lesioned and preserved articular cartilage samples were obtained from participants in the Research Osteoarthritis and Articular Cartilage (RAAK) study described by Ramos et al (3). In the present study, a total of 98 samples were used (65 knees, 33 hips) (see Supplementary Table 1, on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). Ethical approval was obtained from the medical ethics committee of the Leiden University Medical Center (no. P08.239/P19.013), and informed consent was obtained from all participants.

RNA sequencing. Total RNA from articular cartilage was isolated using a Qiagen RNeasy Mini Kit. Paired-end 2 × 100-bp read RNA sequencing (Illumina TruSeq RNA Library Prep Kit, Illumina HiSeq2000, and Illumina HiSeq4000) was performed. Strand-specific RNA-Seq libraries were generated, which yielded a mean of 20 million reads per sample. Quality control was per-

formed as previously described (4), and reads were subsequently aligned to the GRCh38 reference genome with an RNA-Seq aligner STAR (version 2.6.0) (17). Thereafter, aligned reads were processed into individual transcripts using StringTie (version 1.3.4) (18). lncRNAs were identified by mapping the transcripts to GENCODE (version 29) (11) and Ensembl (version 94) (19).

In order to filter transcripts with unknown protein-coding potential, we integrated 2 sources of evidence: 1) predictions from the alignment-free Coding Potential Assessment Tool (CPAT, version 1.2.2) (20), and 2) predictions from the LncFinder R package (version 1.1.3) (21). CPAT is a machine learning–based method that analyzes the sequence features of transcript open-reading frames (ORFs) using a logistic regression model built from ORF size, Fickett TESTCODE statistic, and hexamer usage bias. In CPAT, a transcript with a coding probability of ≥ 0.364 was considered to be a coding sequence. LncFinder predicts lncRNAs using heterologous features and a machine learning model (21). Transcripts with protein-coding potential predicted by both tools were removed from the data set.

Differential expression analysis and replication.

Differential expression analysis was performed on 32 paired samples (25 knees and 7 hips) (Supplementary Table 1A, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>), using the DESeq2 R package (version 1.24) (22). A general linear model assuming a negative binomial distribution was applied, followed by a paired Wald's test comparing lesioned OA cartilage samples and preserved OA cartilage samples, with the preserved samples as the referent. *P* values less than 0.05 (after Benjamini-Hochberg correction) were considered significant and are reported as the false discovery rate (FDR). This analysis was repeated for knee and hip samples separately.

Furthermore, to validate the results, 5 significantly differentially expressed lncRNAs were selected and measured by reverse transcription–quantitative polymerase chain reaction (RT-qPCR) in 10 paired cartilage samples overlapping with the RNA-Seq samples (Supplementary Table 1B), and replication was performed in an independent cohort of 10 paired cartilage samples (Supplementary Table 1C). Total RNA was isolated using an RNeasy Mini Kit, followed by complementary DNA (cDNA) synthesis using 100-ng RNA with a First Strand cDNA Synthesis kit according to the instructions of the manufacturer (Roche Applied Science). Expression levels of *AC025370.1*, *AC090877.2*, *MEG3*, *P3H2-AS1*, *TBILA*, and *GAPDH* were determined using FastStart SYBR Green Master reaction mix (Roche Applied Science). Primer sequences are shown in Supplementary Table 2 (<http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). Relative gene expression levels were calculated with the $2^{-\Delta\Delta C_t}$ method, using *GAPDH* as internal control. A paired *t*-test was performed on the $-\Delta C_t$ values, and *P* values less than 0.05 were considered significant.

lncRNA–mRNA interactions. To identify potential interactions, correlations were calculated between the identified differentially expressed lncRNAs and the previously reported differentially expressed protein-coding genes in the same samples. lncRNA expression data were normalized and variance stabilizing transformed using the DESeq2 R package (version 1.24) (22), and batch effect was removed using the limma R package (version 3.40.6) (23). Our previously published mRNA data (4) were equally normalized and transformed, and batch effect was removed. Subsequently, Spearman’s correlations were calculated between the significantly differentially expressed lncRNAs identified in the combined analysis of knee and hip samples and the differentially expressed protein-coding genes previously published (4), using the Hmisc R package (version 4.2.0) for OA cartilage samples (Supplementary Table 1D). Correlations with *P* values less than 0.05 were considered significant. Network visualization was performed using the RedeR package (version 3.10) (24).

In vitro down-regulation of lncRNA using locked nucleic acid (LNA) GapmeRs. Primary chondrocytes were isolated from 3 independent donors and passaged twice or thrice, as previously described (25). Chondrocytes were transfected in duplo with antisense LNA GapmeR (Qiagen) targeting *P3H2-AS1* (TGAGCAACTAGGTGTA) or GapmeR negative control (AACACGTCTATACGC) at 10 nM final concentration using Lipofectamine RNAiMax Transfection Reagent according to instructions of the manufacturer (Invitrogen). Cells were lysed 30 hours posttransfection with TRIzol reagent (Thermo Fisher Scientific) for RNA isolation, which was done using an RNeasy Mini Kit. Synthesis of cDNA was performed with 150 ng of total RNA using a First Strand cDNA Synthesis kit according to the instructions of the manufacturer. Expression levels of *P3H2-AS1*, *P3H2*, and *GAPDH* were determined using FastStart SYBR Green Master reaction mix. Primer sequences are shown in Supplementary Table 2 (<http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). Relative gene expression levels were calculated with the $2^{-\Delta\Delta C_t}$ method, using *GAPDH* as internal control. A paired *t*-test was performed on the $-\Delta C_t$ values, and *P* values less than 0.05 were considered significant.

Data availability statement. FASTQ files are available on ArrayExpress E-MTAB-7313.

RESULTS

Characterization of lncRNAs in OA cartilage. To characterize lncRNAs in OA cartilage, we used our previously assessed RNA-Seq data on 32 paired samples (25 knees, 7 hips) of lesioned and preserved OA cartilage (4) (Supplementary Table 1A, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). Our in-house pipeline was applied to capture lncRNAs from 2 databases (GENCODE

and Ensembl). As shown in Figure 1, 30,354 lncRNAs were initially detected in our data set. To filter out possible transcripts of unknown coding potential, we integrated results from 2 machine learning approaches (CPAT [20] and Lncfinder [21]). After removing these transcripts, 29,219 lncRNAs remained in the data set and were considered for further analyses. To robustly detect lncRNAs expressed in OA cartilage, a cutoff of an average of ≥ 2 counts per lncRNA was applied, resulting in a total of 5,053 lncRNAs expressed in cartilage (Figure 1). Classification of these lncRNAs based on biotype showed that 1,989 were antisense RNAs (39.4%), 249 were sense RNAs (4.9%), 1,532 were pseudogenes (30.3%), and 900 were lincRNAs (17.8%) (Figure 2).

Differential expression of lncRNAs between lesioned and preserved OA cartilage. To identify lncRNAs associated with the OA process, differential expression analysis was performed on paired lesioned and preserved OA cartilage samples, resulting in 191 significantly differentially expressed lncRNAs (FDR < 0.05; Figure 1). Of these, 65 were antisense RNAs (34.0%), 10 were sense RNAs (5.2%), 33 were pseudogenes (17.3%), and 66 were lincRNAs (34.6%) (Figure 2 and Supplementary Table 3, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). When comparing the biotypes of the total expressed lncRNAs to the biotypes of the differentially expressed lncRNAs (Figure 2), we observed an increase of lincRNAs and a decrease

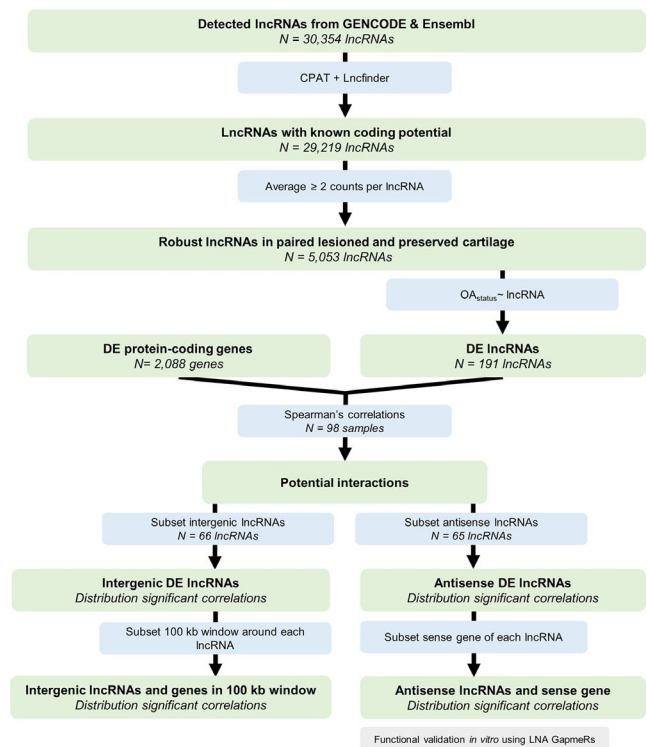


Figure 1. Overview of applied strategy. Numbers of genes or long noncoding RNAs (lncRNAs) represent significantly differentially expressed (DE) genes or lncRNAs (according to false discovery rate). OA = osteoarthritis; LNA = locked nucleic acid.

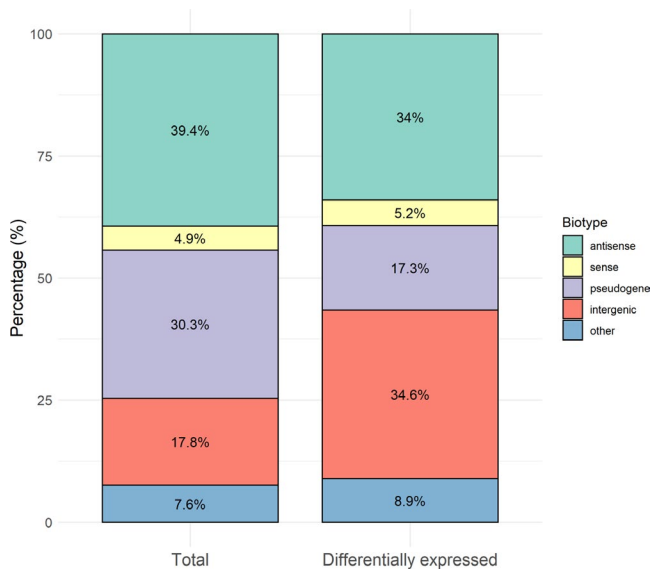


Figure 2. Distribution of biotypes of total long noncoding RNAs (lncRNAs) expressed in total cartilage compared to lncRNAs that were significantly differentially expressed (according to the false discovery rate) between lesioned osteoarthritis (OA) cartilage and preserved OA cartilage.

of pseudogenes. The most significantly differentially expressed lncRNA was lincRNA *AL139220.2* (fold change 2.2, FDR 2.0×10^{-10}). As depicted in Figure 3, 114 lncRNAs were down-regulated and 77 were up-regulated, with a fold change ranging from 0.3 (*AC100782.1*, FDR 6.5×10^{-4}) to 4.5 (*LINC01411*, FDR 2.6×10^{-6}).

The 191 lncRNAs identified in this study included several previously found to be associated with OA, such as *MEG3*

(fold change 0.6, FDR 8.8×10^{-3}), *PART1* (fold change 1.8, FDR 1.7×10^{-4}), and *LINC01614* (fold change 2.6, FDR 9.5×10^{-3}) (16,26), as well as novel OA-associated lncRNAs, including *P3H2-AS1* (fold change 2.7, FDR 4.1×10^{-4}) and *AC090877.2* (fold change 0.3, FDR 6.2×10^{-5}). Notably, previously identified lncRNAs such as *MALAT1* (fold change 1.3, FDR 0.4) (27), *TUG1* (fold change 1.1, FDR 0.7) (28), *HOTAIR* (fold change 0.8, FDR 0.5), and *GAS5* (fold change 1.1, FDR 0.8) (29) were not found to be significantly differentially expressed in the present study.

To validate the differential expression results, we selected 5 lncRNAs (*AC025370.1*, *AC090877.2*, *MEG3*, *P3H2-AS1*, and *TBILA*) based on the highest absolute fold change and genomic location, using RT-qPCR in a cohort consisting of 10 paired samples (Supplementary Table 1B, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>), overlapping with the RNA-Seq samples. All 5 lncRNAs were detected using RT-qPCR with equal direction of effect as those found in the RNA-Seq analysis (Supplementary Table 4, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). Furthermore, replication was performed in an independent cohort of 10 paired cartilage samples (Supplementary Table 1C), which also showed comparable effect sizes and directions (Supplementary Table 4).

To explore whether joint-specific lncRNAs could be detected, stratified analyses were performed for knee samples (25 pairs) and hip samples (7 pairs). Upon performing differential expression analysis on the knee samples, 90 significantly differentially expressed lncRNAs were identified (Supplementary

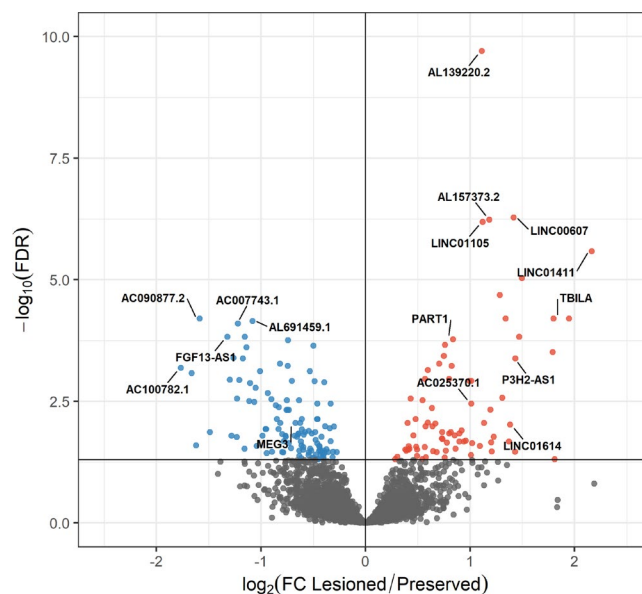


Figure 3. Differential expression analysis of long noncoding RNAs (lncRNAs) between lesioned osteoarthritis (OA) cartilage and preserved OA cartilage. Volcano plot shows differentially expressed lncRNAs, with down-regulated lncRNAs represented by blue circles and up-regulated lncRNAs represented by red circles. Top differentially expressed lncRNAs are labeled, as are known and novel OA-associated lncRNAs. FDR = false discovery rate; FC = fold change.

Table 5A, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>), of which 12 were not found in the previous combined analysis and therefore were unique to knee cartilage (Supplementary Table 6A, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/abstract>). In the hip samples, 31 lncRNAs were significantly differentially expressed (Supplementary Table 5B), of which 13 were unique to hip cartilage (Supplementary Table 6B). The most significantly differentially expressed lncRNA unique to the knee was *MSL3P1* (fold change 1.5, FDR 1.49×10^{-2}), while one of the most significantly differentially expressed lncRNAs unique to the hip was *PAPPA-AS1* (fold change 9.4, FDR 2.77×10^{-4}). Notably, the most up-regulated lncRNA in the hip, *AP001515.1* (fold change 21.5, FDR 2.8×10^{-4}), was also unique to the hip, while the most up-regulated lncRNA in the knee, *LINC01411*

(fold change 5.8, FDR 6.1×10^{-6}), was not unique to the knee.

Potential interactions between lncRNAs and mRNAs relevant to OA pathophysiology.

We next aimed to investigate whether mRNAs associated with the OA process are regulated by differentially expressed lncRNAs. Based on the assumption that interactions between lncRNAs and mRNAs likely show coexpression (30) among lesioned and preserved OA cartilage samples, correlations were calculated between our previously reported differentially expressed protein-coding genes (4) and differentially expressed lncRNAs (Supplementary Table 1D), as shown in Figure 1. This resulted in 343 significant correlations ($r > 0.8$) (Supplementary Table 7, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/>), comprising 47 unique lncRNAs, of which 17 were antisense (36%) and 14

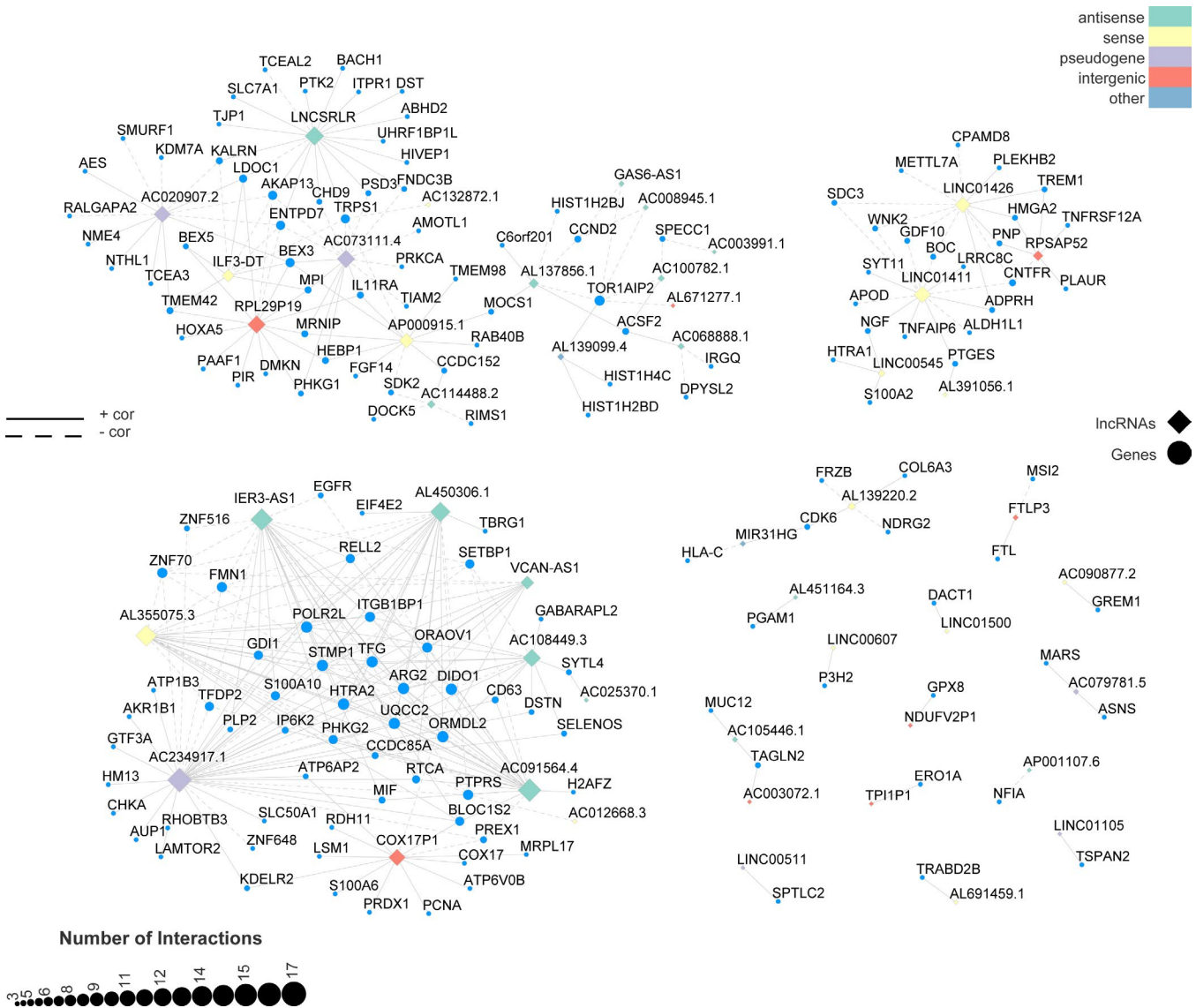


Figure 4. Osteoarthritis (OA)-specific long noncoding RNA (lncRNA)-mRNA coexpression network. Network of differentially expressed lncRNAs and mRNAs with a correlation (cor) of >0.8 between lesioned OA cartilage and preserved OA cartilage is shown.

were intergenic (30%) (Supplementary Table 8, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/>). This distribution is comparable to that found among all differentially expressed lincRNAs (Figure 2), supporting the notion that lincRNAs regulate mRNAs, independent of biotype. Notably, the most significantly differentially expressed lincRNA, *AL139220.2* (fold change 2.2, FDR 2.0×10^{-10}), showed one of the highest correlations with *COL6A3* ($r = 0.8$, $P = 2.2 \times 10^{-16}$), encoding a type VI collagen chain.

To visualize these interactions, an OA-specific lincRNA–mRNA coexpression network was generated. As shown in Figure 4, 3 relatively large clusters of interacting lincRNAs and mRNAs were observed. One cluster was characterized as being highly interlinked with a cluster of the same genes (e.g., *ITGB1BP1* correlated with the 6 lincRNAs *IER-AS1*, *AL355075.3*, *AC234917.1*, *AC091564.4*, *AC108449.3*, and *AL450306.1*), whereas the other 2 were characterized by lincRNAs interlinked with mostly unique genes (e.g., *LNCSRLR* with 18 genes). In addition to the clusters, there were a number of singular interlinked lincRNAs, such as *AC090877.2* (fold change 0.3, FDR 6.2×10^{-5}) with *GREM1* ($r = 0.9$, $P = 2.2 \times 10^{-16}$), which encodes a cytokine of the bone morphogenetic protein antagonist family (Figure 4). Interestingly, *GREM1* was the gene located closest to *AC090877.2*, suggesting that this lincRNA *cis*-regulates this gene.

One of our objectives in the present study was to generalize the identification of potential *cis*-regulation of differentially expressed lincRNAs (Figure 1). As shown in Figure 5A, we compared the distribution of significant correlations between differentially expressed lincRNAs and all genes and between differentially expressed lincRNAs and genes that lie within a 100-kb window of the transcription start site. The proportion of significant correlations >0.5 with all differentially expressed genes was 11%, but this increased to 44% when we only considered the 100-kb window. Since the percentage of differentially expressed antisense lincRNAs (34%) was comparable to that of intergenic lincRNAs (34.6%), we also aimed to identify potential *cis*-regulation of antisense lincRNAs. To this end, we compared the distribution of correlations between differentially expressed antisense lincRNAs and all protein-coding mRNAs and between differentially expressed antisense lincRNAs and their sense genes (Figure 5B). The percentage of correlations >0.5 was 10% with all genes and 61% with only the sense genes, showing that there is an enrichment for higher, positive correlations between antisense lincRNAs and their sense gene. Taken together, these data suggest that both intergenic and antisense lincRNAs are prone to regulate mRNAs in *cis* in OA cartilage.

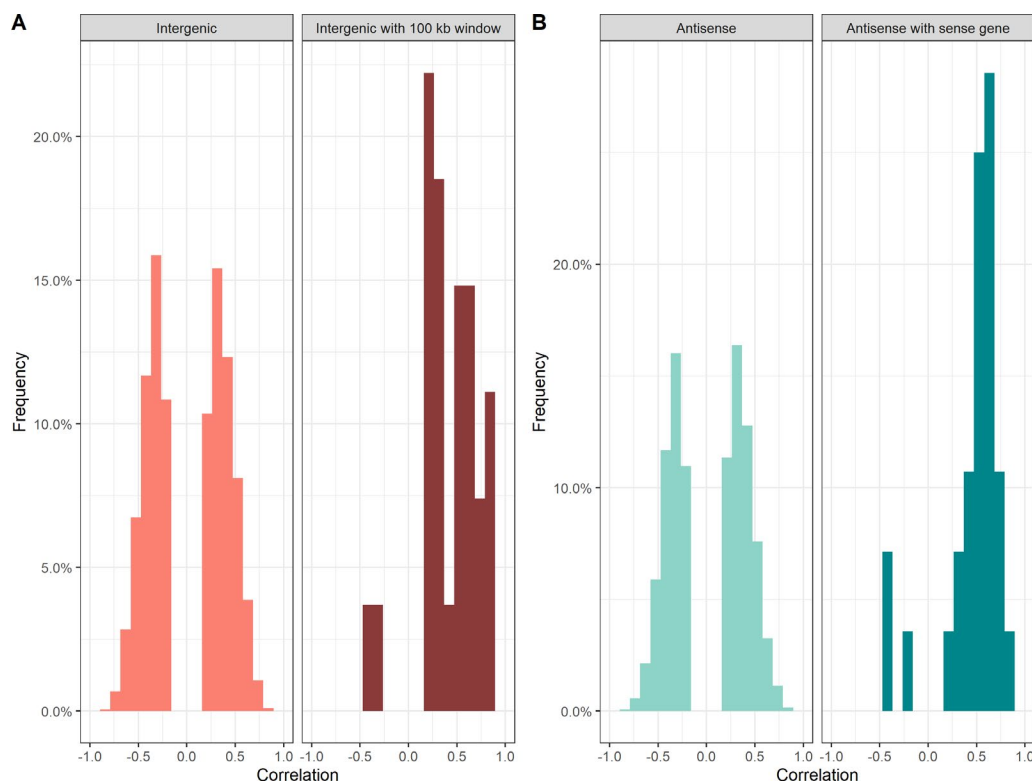


Figure 5. Distribution of significant correlations between intergenic differentially expressed long noncoding RNAs (lincRNAs) and previously identified differentially expressed protein-coding genes or protein-coding genes in a 100-kb window (A), and between antisense differentially expressed lincRNAs and differentially expressed protein-coding genes or their sense genes (B). Correlations between lincRNA and mRNA data were calculated from the same osteoarthritis cartilage samples ($n = 98$).

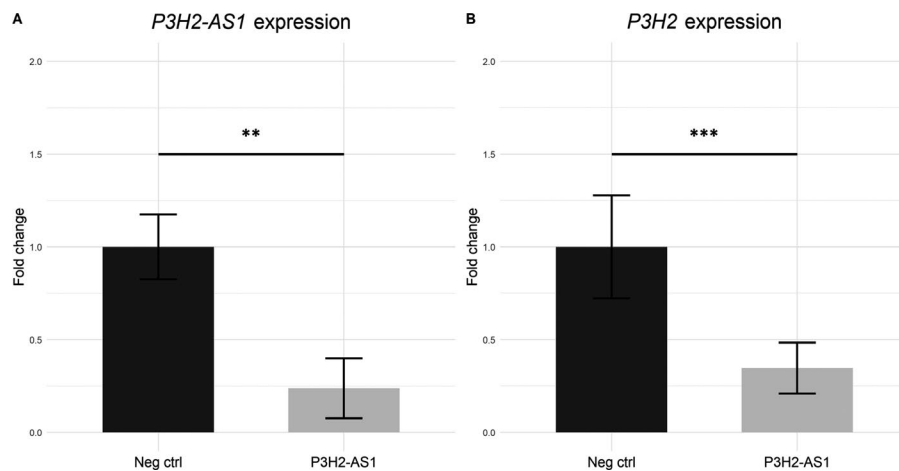


Figure 6. Expression of long noncoding RNA (lncRNA) *P3H2-AS1* and gene *P3H2* in primary chondrocytes transfected with *P3H2-AS1*-targeting antisense locked nucleic acid (LNA) GapmeRs compared to nontargeting LNA GapmeRs. **A**, *P3H2-AS1* expression was significantly down-regulated by the *P3H2-AS1*-targeting LNA GapmeRs. **B**, *P3H2* expression was significantly down-regulated in chondrocytes transfected with *P3H2-AS1*-targeting LNA GapmeRs. Bars show the mean \pm SD. ** = $P < 0.01$; *** = $P < 0.001$, by paired t -test ($n = 3$ donors).

Down-regulation of lncRNA expression using LNA GapmeRs. To validate whether the previously identified *cis*-regulation between lncRNAs and their surrounding genes is caused by a direct effect, *P3H2-AS1* was selected as a proof of concept for functional validation. *P3H2-AS1* is an antisense lncRNA, which was found to be highly up-regulated in lesioned OA cartilage (fold change 2.7, FDR 4.1×10^{-4}) (4), and the highest correlation was with its sense gene *P3H2* ($r = 0.63$, $P = 1.0 \times 10^{-13}$) (Supplementary Figure 1, <http://onlinelibrary.wiley.com/doi/10.1002/art.41396/>). To this end, primary chondrocytes were transfected with a *P3H2-AS1*-targeting LNA GapmeR. As shown in Figure 6A, this resulted in a significant down-regulation of *P3H2-AS1* compared to a nontargeting LNA GapmeR (fold change 0.28, $P = 0.0035$). Subsequently, *P3H2* expression levels were measured, which showed that *P3H2* expression was significantly down-regulated compared to cells transfected with nontargeting control LNA GapmeRs (fold change 0.36, $P = 0.001$) (Figure 6B).

DISCUSSION

To our knowledge, we are the first to report on robust differential expression of lncRNAs as related to OA pathophysiology, while integrating them with data on differential mRNA expression levels of the same samples using RNA sequencing. As a result, our new in-house pipeline identified 5,053 lncRNAs that were robustly expressed, 191 of which were significantly differentially expressed (according to FDR) between lesioned and preserved OA cartilage. Notably, we observed an increase in the percentage of lincRNAs, highlighting their general involvement in the OA pathophysiology process. The directions of effect of *AC025370.1* (fold change 2.0, FDR 3.5×10^{-3}), *AC090877.2* (fold change 0.3, FDR 6.2×10^{-5}), *MEG3* (fold change 0.63, FDR 8.8×10^{-3}),

P3H2-AS1 (fold change 2.7, FDR 4.1×10^{-4}), and *TBILA* (fold change 3.5, FDR 1.1×10^{-7}) were validated and replicated by RT-qPCR, indicating robustness of our lncRNA mapping strategy. Correlations were calculated to identify potential interactions between expression levels of differentially expressed lncRNAs and differentially expressed protein-coding genes (4) in the same OA cartilage samples. As a result, both intergenic and antisense differentially expressed lncRNAs showed an enrichment for higher, positive correlations with their respective flanking or sense genes compared to the total data set. Validating this *cis*-regulation in vitro, *P3H2-AS1* levels were down-regulated in primary chondrocytes, which resulted in down-regulation of the sense gene *P3H2* expression levels, thereby confirming that *P3H2-AS1* regulates its sense gene *P3H2*.

We identified 29,219 lncRNAs that were expressed in OA cartilage. However, after applying a filter with a cutoff of an average of ≥ 2 counts per lncRNA, the detected lncRNAs were reduced by $\sim 83\%$ to 5,053. Since lncRNAs are known to be expressed at very low levels this was to be expected, yet lncRNAs expressed at low levels can still be functional (12). To allow exploratory analyses with lncRNAs expressed at such low levels, deeper sequencing would be necessary, with a read depth of ~ 50 million reads per sample. Additionally, to be able to report on valid lncRNAs in OA articular cartilage and their potential target mRNAs, we prioritized reporting known lncRNAs with a predicted non-protein-coding potential. Nonetheless, by focusing on these known lncRNAs, we may have disregarded compelling novel OA-relevant lncRNAs.

Given that we had a (within-patient) paired lesioned cartilage-preserved cartilage study design, with pairs sequenced on the same batch, we applied a paired Wald's test as implemented in the DESeq2 R package. Since our data set also included lncRNAs expressed at low levels, the addition of a random effect to compensate for technical errors may have been a better, yet more

conservative, approach. As such, the lncRNAs in our data set, particularly those with low read counts, could be subject to false positive results and therefore require replication and verification.

We observed a particular enrichment of lincRNAs in the differential expression analysis compared to the total data set (34.6% versus 17.8%) (Figure 2), showing that lincRNAs indeed play an important role in OA pathophysiology, as seen in previous studies (8,16,30). Nonetheless, in comparison to the fraction of significantly differentially expressed lincRNAs reported by Pearson et al (8), this proportion is still relatively small. However, Pearson and colleagues performed RNA-Seq on samples from isolated chondrocytes in contrast to the RNA isolated from cartilage in our study and focused on profiling lncRNAs up-regulated by interleukin-1 β . The activation of chondrocyte proliferation in tissue culture will likely induce expression of RNAs involved in transcriptional regulation, compared to the transcriptome of maturational arrested chondrocytes residing in cartilage.

Of the 191 lncRNAs that were significantly differentially expressed between lesioned and preserved OA cartilage (Figure 3), multiple lncRNAs have been previously identified, including *MEG3*, *LINC01614*, and *PART1* (16,26). However, there were also examples of lncRNAs previously associated with OA (27–29), which were not significantly differentially expressed, such as *MALAT1*, *HOTAIR*, *GAS5*, and *TUG1*. A possible explanation could be that they were found to be differentially expressed between preserved OA and healthy cartilage, as opposed to our comparison between lesioned OA cartilage and preserved OA cartilage (7). The cross-sectional study design comparing OA cartilage and healthy cartilage provides insight into which lncRNAs are involved in the early phase of OA pathophysiology and are therefore potentially causal in the process and which lncRNAs are specific to healthy cartilage; this was not possible with our study design. Nonetheless, the paired analysis allowed for detection of lncRNA expression changes specific to the OA pathophysiologic process, independent of confounding factors such as sex and age. At least 35 differentially expressed lncRNAs in our data set were previously found to be associated with OA (10,16,30), but the most significantly differentially expressed lncRNA, *AL139220.2*, and the most up- and down-regulated differentially expressed lncRNAs, *LINC01411* and *AC100782.1*, respectively, have not previously been found to be associated with OA (3), showing that a paired study design allows for the detection of many more lncRNAs involved in the OA pathophysiologic process.

Previous studies have demonstrated differences in dysregulated pathways between knee and hip OA cartilage and epigenetic differences based on DNA methylation (8,16,31,32); thus, we aimed to identify joint-specific lncRNAs. Differential expression analysis in knee samples resulted in a higher number of significantly differentially expressed lncRNAs ($n = 90$) than in hip samples ($n = 31$), which could be due to the smaller sample size of the hip samples (25 knee pairs versus 7 hip pairs). However, the number of unique lncRNAs per joint site was similar: 12 unique

knee lncRNAs and 13 unique hip lncRNAs. This suggests that there is more heterogeneity in the processes in the knee, which could be due in part to anatomic joint site-specific differences. This is also supported by the fact that the average fold change of the up-regulated lncRNAs unique to the hip was 8.3, while it was 1.5 for knee. The unique lncRNAs with the highest fold change in the knee (*AC068768.1* fold change 1.6, FDR 2.3×10^{-2}) and hip (*AP001615.1* fold change 21.5, FDR 2.8×10^{-4}) were not previously found to be associated with OA. The identification of these joint-specific lncRNAs is useful for follow-up studies to determine potential joint-specific therapeutic targets.

Unlike conserved microRNAs, it is difficult to predict the function of lncRNAs based solely on nucleotide sequence, due to their lack of conservation of the primary sequence (15). To explore potential regulatory interactions between lncRNAs and mRNAs in cartilage, correlations were calculated between differentially expressed lncRNAs and differentially expressed protein-coding mRNAs (Figure 4). At the transcriptional level, lncRNAs can exert their function in *trans* or *cis* (13), both of which we observed in this study. The most significantly differentially expressed lncRNA, *AL139220.2*, showed one of the highest correlations with *COL6A3* ($r = 0.8$, $P = 2.2 \times 10^{-16}$), encoding one of the type VI collagen chains as part of the complete type VI collagen molecule, which is mostly present in the pericellular matrix of cartilage. *AL139220.2* is located on chromosome 1 and, at present, little is known about its function. Since *COL6A3* is located on chromosome 2, it seems likely that *AL139220.2* regulates *COL6A3* expression in *trans*. Notably, *AC090877.2* showed the highest correlation with its sense gene *GREM1* ($r = 0.9$, $P = 2.2 \times 10^{-16}$), suggesting that this lncRNA *cis*-regulates this gene. In previous studies, it has been shown that lincRNAs often regulate flanking mRNAs in *cis* in OA, in which a positive correlation was found between the expression of mRNA-flanking lincRNAs and their nearest coding mRNA (8,30). This observation was confirmed by our findings, as the percentage of higher, positive correlations ($r > 0.5$) was considerably larger between lincRNAs and the differentially expressed genes that lie within a 100-kb window (44%) than with all differentially expressed genes (11%) (Figure 5A).

Furthermore, it is known that antisense lncRNAs can regulate their overlapping sense genes in *cis* (14), which has not previously been investigated in OA. We found an enrichment for higher, positive correlations between antisense differentially expressed lncRNAs and their sense genes ($r > 0.5$ in 61%) compared to correlations between antisense differentially expressed lncRNAs and all differentially expressed genes ($r > 0.5$ in 10%), suggesting that indeed antisense lncRNAs often regulate their sense genes in *cis* (Figure 5B). Therefore, to completely understand the transcriptional regulation of lncRNAs in the OA process, the total lncRNA transcriptome should be considered, not solely the lincRNAs. Of importance is the notion that these correlations are not yet proof of a (direct) downstream effect of lncRNAs on the mRNAs.

Given these observations, we selected the antisense lncRNA *P3H2-AS1* as proof of principle to establish whether it regulates its sense gene. Down-regulation of *P3H2-AS1* resulted in a significant down-regulation of *P3H2* expression levels (Figure 6), thereby confirming that *P3H2-AS1* regulates its sense gene in *cis*. *P3H2* encodes an enzyme that catalyzes posttranslational 3-hydroxylation of proline residues and plays a critical role in collagen chain assembly, stability, and crosslinking and was recently found to be highly up-regulated in lesioned OA cartilage, and therefore likely involved in the OA process (4). Antisense lncRNAs can affect biogenesis or mobilization of target RNA on multiple levels, such as transcription, splicing, and translation (14). To elucidate the exact mechanism of *P3H2-AS1* regulating *P3H2* and investigate whether *P3H2-AS1* can be used as a potential preclinical target by modulating *P3H2* expression levels via *P3H2-AS1*, complementary functional studies employing clustered regularly interspaced short palindromic repeat/Cas9, RNA fluorescence in situ hybridization, or crosslinked immunoprecipitation are necessary (33).

In conclusion, our improved detection strategy resulted in the characterization of lncRNAs robustly expressed in OA cartilage. Our data signify that intergenic as well as antisense lncRNAs play an important role in regulating the pathophysiology of OA. Moreover, we observed that in addition to a previous finding that intergenic lncRNAs function in *cis*, antisense lncRNAs can exert their function in *cis*, which we confirmed in vitro. Future studies regarding lncRNAs and OA should be complemented by functional validation, e.g., by modulating lncRNA expression levels using antisense LNA GapmeRs, in order to confirm whether a correlation signifies a biologic relationship between lncRNA and mRNA.

ACKNOWLEDGMENTS

We thank all the participants of the RAAK study (supported by Leiden University Medical Center). We thank Evelyn Houtman, Enrike van der Linden, Robert van der Wal, Peter van Schie, Shaho Hasan, Maartje Meijer, Daisy Latijnhouwers, and Geert Spierenburg for their contribution to the collection of the joint tissue. Data were generated within the scope of the Medical Delta Programs Regenerative Medicine 4D: Generating Complex Tissues With Stem Cells and Printing Technology and Improving Mobility with Technology.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Ms van Hoolwerff had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Van Hoolwerff, Metselaar, Ramos, Coutinho de Almeida, Meulenbelt.

Acquisition of data. Van Hoolwerff, Metselaar, Tuerlings, Suchiman, Lakenberg, Ramos, Cats, Nelissen, Broekhuis, Mei, Coutinho de Almeida, Meulenbelt.

Analysis and interpretation of data. Van Hoolwerff, Metselaar, Tuerlings, Ramos, Mei, Coutinho de Almeida, Meulenbelt.

REFERENCES

1. Goldring SR, Goldring MB. Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone crosstalk [review]. *Nat Rev Rheumatol* 2016;12:632–44.
2. Tonge DP, Pearson MJ, Jones SW. The hallmarks of osteoarthritis and the potential to develop personalised disease-modifying pharmacological therapeutics [review]. *Osteoarthritis Cartilage* 2014; 22:609–21.
3. Ramos YF, den Hollander W, Bovee JV, Bomer N, van der Breggen R, Lakenberg N, et al. Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study. *PLoS One* 2014;9:e103056.
4. Coutinho de Almeida RC, Ramos YF, Mahfouz A, den Hollander W, Lakenberg N, Houtman E, et al. RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. *Ann Rheum Dis* 2019;78:270–7.
5. Coutinho de Almeida RC, Ramos YF, Meulenbelt I. Involvement of epigenetics in osteoarthritis [review]. *Best Pract Res Clin Rheumatol* 2017;31:634–48.
6. Den Hollander W, Ramos YF, Bomer N, Elzinga S, van der Breggen R, Lakenberg N, et al. Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. *Arthritis Rheumatol* 2015;67:2108–16.
7. Jiang SD, Lu J, Deng ZH, Li YS, Lei GH. Long noncoding RNAs in osteoarthritis [review]. *Joint Bone Spine* 2017;84:553–6.
8. Pearson MJ, Philp AM, Heward JA, Roux BT, Walsh DA, Davis ET, et al. Long intergenic noncoding RNAs mediate the human chondrocyte inflammatory response and are differentially expressed in osteoarthritis cartilage. *Arthritis Rheumatol* 2016;68: 845–56.
9. Sun H, Peng G, Ning X, Wang J, Yang H, Deng J. Emerging roles of long noncoding RNA in chondrogenesis, osteogenesis, and osteoarthritis. *Am J Transl Res* 2019;11:16–30.
10. Liu Q, Zhang X, Dai L, Hu X, Zhu J, Li L, et al. Long noncoding RNA related to cartilage injury promotes chondrocyte extracellular matrix degradation in osteoarthritis. *Arthritis Rheumatol* 2014;66:969–78.
11. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;47:D766–73.
12. Jarroux J, Morillon A, Pinsky M. History, discovery, and classification of lncRNAs. *Adv Exp Med Biol* 2017;1008:1–46.
13. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long noncoding RNAs [review]. *Nat Rev Genet* 2020;21:102–17.
14. Villegas VE, Zaphiropoulos PG. Neighboring gene regulation by antisense long noncoding RNAs. *Int J Mol Sci* 2015;16:3251–66.
15. Uszczyńska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long noncoding RNA transcriptome [review]. *Nat Rev Genet* 2018;19:535–48.
16. Ajekigbe B, Cheung K, Xu Y, Skelton AJ, Panagiotopoulos A, Soul J, et al. Identification of long noncoding RNAs expressed in knee and hip osteoarthritic cartilage. *Osteoarthritis Cartilage* 2019;27: 694–702.
17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
18. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a

- transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;33:290–5.
19. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. Ensembl 2019. *Nucleic Acids Res* 2019;47:D745–51.
 20. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* 2013;41:e74.
 21. Han S, Liang Y, Ma Q, Xu Y, Zhang Y, Du W, et al. LncFinder: an integrated platform for long noncoding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. *Brief Bioinform* 2019;20:2009–27.
 22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
 23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
 24. Castro MA, Wang X, Fletcher MN, Meyer KB, Markowitz F. RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. *Genome Biol* 2012;13:R29.
 25. Bomer N, den Hollander W, Suchiman H, Houtman E, Slieker RC, Heijmans BT, et al. Neo-cartilage engineered from primary chondrocytes is epigenetically similar to autologous cartilage, in contrast to using mesenchymal stem cells. *Osteoarthritis Cartilage* 2016;24:1423–30.
 26. Chen K, Zhu H, Zheng MQ, Dong QR. LncRNA MEG3 inhibits the degradation of the extracellular matrix of chondrocytes in osteoarthritis via targeting miR-93/TGFBR2 axis. *Cartilage* 2019. E-pub ahead of print.
 27. Zhang Y, Wang F, Chen G, He R, Yang L. LncRNA MALAT1 promotes osteoarthritis by modulating miR-150-5p/AKT3 axis [review]. *Cell Biosci* 2019;9:54.
 28. Tang LP, Ding JB, Liu ZH, Zhou GJ. LncRNA TUG1 promotes osteoarthritis-induced degradation of chondrocyte extracellular matrix via miR-195/MMP-13 axis. *Eur Rev Med Pharmacol Sci* 2018;22:8574–81.
 29. Xing D, Liang JQ, Li Y, Lu J, Jia HB, Xu LY, et al. Identification of long noncoding RNA associated with osteoarthritis in humans. *Orthop Surg* 2014;6:288–93.
 30. Fu M, Huang G, Zhang Z, Liu J, Zhang Z, Huang Z, et al. Expression profile of long noncoding RNAs in cartilage from knee osteoarthritis patients. *Osteoarthritis Cartilage* 2015;23:423–32.
 31. Den Hollander W, Ramos YF, Bos SD, Bomer N, van der Breggen R, Lakenberg N, et al. Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. *Ann Rheum Dis* 2014;73:2208–12.
 32. Loughlin J, Reynard LN. Osteoarthritis: epigenetics of articular cartilage in knee and hip OA [review]. *Nat Rev Rheumatol* 2015;11:6–7.
 33. Charles Richard JL, Eichhorn PJA. Platforms for investigating LncRNA functions [review]. *SLAS Technol* 2018;23:493–506.