

Analysis of RNA polyadenylation in healthy and osteoarthritic human articular cartilage

Phaedra Winstanley-Zarach¹, Gregor Rot², Shweta Kuba^{1, 3}, Aibek Smagul¹, Mandy J Peffers¹ & Simon R Tew¹

¹ Centre for integrated Research into Musculoskeletal Ageing (CIMA), Department of Musculoskeletal and Ageing Sciences, Institute of Life Course and Medical Sciences, University of Liverpool, William Henry Duncan Building, West Derby Street, Liverpool, L7 8TX, United Kingdom

² Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, Winterthurerstrasse 190, 8057 Zurich, Switzerland

³ School of Health and Life Sciences, National Horizons Centre, Teeside University, Darlington, DL1 1HG, United Kingdom

ABSTRACT

An important transcript structural element is its 3' polyadenylated (polyA) tail, which defines the 3' boundary of the transcript's genetic information and is necessary for transcript stability. The position of the polyA tail can vary, with multiple alternatively polyadenylated (APA) transcripts existing for a single gene. This can lead to different length transcripts which can vary in their 3' regulatory domains and even by inclusion or exclusion of protein-coding introns. The distribution of polyA tail location on articular chondrocyte transcripts has not been examined before and this study aimed to be the first to define polyadenylation events in human chondrocytes using age-matched healthy and osteoarthritic knee articular cartilage samples. Total RNA was isolated from frozen tissue samples and analysed using the QuantSeqReverse 3' RNA Sequencing approach, where each read runs 3' to 5' from within the polyA tail into the transcript and will contain a distinct polyA site. Initial analysis of differential expression of overall transcript abundance identified by the reads showed significant disruption to transcript levels when healthy samples were compared to osteoarthritic ones. As we expected, differentially regulated genes were enriched with functionalities that were strongly associated with joint pathology. As part of this analysis, we also identified a substantial number of differentially expressed long non-coding RNAs that had not been linked to osteoarthritis before. Subsequent examination of polyA site data allowed us to define the extent of site usage across all the samples. This included identification of chondrocyte genes that exhibited the greatest amount polyA site variation. When comparing healthy and osteoarthritic samples, we found that differential use of polyadenylation sites was modest. However, of the small number of genes affected, there was clear potential for the change in polyadenylation site usage elicited by pathology to have functional relevance. We examined two genes, OSMR and KMT2A, in more detail, defining how APA affects transcript turnover and then, in the case of OSMR, identifying that APA is sensitive to inflammatory cytokine stimulation. Overall, we have characterised the polyadenylation landscape of human knee articular chondrocytes but can conclude that osteoarthritis does not elicit a widespread change in their polyadenylation site usage. This finding differentiates knee osteoarthritis from pathologies such as cancer where APA is more commonly observed.

INTRODUCTION

Osteoarthritis is a multifactorial disease in which the initiation and progression is driven by a combination of genetic, environmental and mechanical factors. The functionality of the chondrocytes within the articular cartilage of the joint is significantly altered by the late stages of the disease, and this is associated with altered gene expression patterns that have been examined in a range of earlier transcriptomic studies [1].

The processes regulating gene expression changes during osteoarthritis are a source of ongoing study. Genetic association studies have identified numerous polymorphisms that are linked to osteoarthritis incidence [2], which can affect the processes that control transcriptional regulation [3]. Mechanisms of transcriptional regulation of known effectors of articular cartilage degradation, such as extracellular proteases have also been determined [4]. Transcriptomic analysis of osteoarthritic cartilage in a variety of species has demonstrated that widespread changes in gene expression is induced by the disease [5].

A growing number of studies have implicated post-transcriptional gene regulation in the disruption of transcript abundance during osteoarthritis. Mouse genetic models have identified significant disruption to skeletal development when regulators of mRNA decay or translation such as microRNAs and RNA binding proteins are disrupted [6-8], indicating the importance of these factors in chondrocyte function. Our previous studies have demonstrated that mRNA decay rates are affected in osteoarthritic human chondrocytes, with an increased number of short-lived mRNAs evident in osteoarthritic cells [9].

One aspect of post-transcriptional gene regulation that has received little attention in cartilage biology is the alternative polyadenylation (APA) of transcripts. It is well established that a high proportion of human genes display evidence of APA, leading to functional and structural

heterogeneity in the transcript pool [10]. By using different polyadenylation sites, transcripts from the same gene can be produced which have altered 3' untranslated region sequence lengths. This can result in the inclusion or exclusion of protein and small RNA interacting regions, generating pools of transcripts that code for the same protein but have different regulatory responses [11]. Functionally, altered 3'UTR length induced by APA influences the decay, localisation and translation of mRNAs. Furthermore, the use of polyadenylation sites upstream of the 3'UTR, often contained within intronic sequences, can lead to transcripts where the protein coding region is altered resulting in the production of truncated proteins with altered functionality [12, 13].

Evidence now suggests that modulation of the transcriptome by APA can be dynamic, is driven by differentiation state and can be tissue specific [14]. Furthermore, links between APA and pathologies are beginning to emerge. Global shortening of 3'UTRs because of APA has been demonstrated to be a feature of cancer pathology [15, 16] whilst impaired APA has been implicated in the development of alcohol induced fatty liver disease [17]. To date however, there has been no study into how chondrocytes utilise this method of transcript mRNA regulation or whether it is affected by degenerative diseases such as osteoarthritis.

In this study we used the QuantSeq Reverse 3' end sequencing approach, which produces reads that run 3' to 5' from the polyA tail into the 3' transcript sequence [18], to examine RNA purified from healthy or osteoarthritic human knee articular chondrocytes. By using age-matched human samples this data allows complementation of existing studies by analysis of how transcript levels change in the disease. This data also allows us to define how chondrocytes polyadenylate their transcripts and how this is affected by disease. We determine that chondrocytes, like many other cell types, make widespread use of APA to generate diversity in their transcript pool. We also find that disease associated changes in APA are observed at a relatively low level but that there is potential to produce significant alterations to both gene function and regulatory kinetics where it occurs.

MATERIALS AND METHODS

Age matched healthy and osteoarthritic articular cartilage

Osteoarthritic human articular cartilage tissue was obtained from the knee joints of OA patients undergoing total knee arthroplasty (Clatterbridge Hospital, Wirral, UK). This material was obtained with informed consent and following national research ethics committee approval (IRAS ID 242434). Age-matched, control human articular cartilage tissue was obtained from Articular Engineering (Northbrook, IL). Six cartilage samples from each group were analysed and all were obtained from male donors. All tissue was harvested from load bearing regions of the femoral condyle. Samples were snap frozen and 10 μ m thick full depth sections of the tissue were cut using a cryostat and collected prior to RNA isolation.

QuantSeq Reverse analysis

Total RNA was isolated from tissue cryosections using a Trizol extraction (Thermo Fisher Scientific, Horsham UK) with subsequent purification using the RNeasy spin column (Qiagen, Manchester UK) purification procedure including on-column DNase digestion [19]. QuantSeq Reverse library preparation and NextSeq 75cyc High Output sequencing with V2chemistry, was performed by Lexogen (Vienna, Austria) [18]. For gene expression analysis the in-house Lexogen pipeline was used, consisting of read trimming, alignment to GRCh38 using STAR aligner [20] and then differential gene expression determined using DESeq2. Correction for multiple hypothesis testing was performed via the Benjamini-Hochberg procedure [21]. Differentially Expressed Genes, defined as \log_2 fold change ≥ 1 and ≤ -1 and adjusted p-values (≤ 0.05), were used to distinguish between downregulated and upregulated genes. Raw FastQ files from QuantSeq Reverse analysis have been deposited in ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) with the accession number E-MTAB-12184.

Alternative polyadenylation analysis

We constructed a reference polyA database by considering genome alignments of sequence reads from all twelve experiments. To account for internal-priming events (reads mapping to genomic polyA sequences) we filtered out detected polyA sites without an upstream polyadenylation signal (PAS). After obtaining read coverage for each detected polyA site separately for each experiment, we ran DEXSeq analysis comparing six healthy vs. six osteoarthritis samples to identify alternatively polyadenylated (APA) genes. Further polyadenylation analysis details and methodology can be explored at [expressRNA](#) [33].

Gene ontology and pathway analysis

Enriched KEGG [22] and Reactome [23] pathways were identified using the clusterProfiler package [24] and visualised using the pathview package [25]. For Gene Set Enrichment Analysis (GSEA), a pre-ordered list of gene-named log₂ fold change values, and for Over Representation Analysis (ORA), an input list of differentially expressed genes was used to obtain enriched pathways. Up-regulated pathways were defined by a normalised enrichment score (NES) > 0 and the down-regulated pathways were defined by an NES < 0. Pathways with BH adjusted-P value ≤ 0.05 were chosen as significantly enriched. RcisTarget v1.9.1 [26] was applied on differentially expressed genes to identify overrepresented potential transcription factor regulation. We examined up to 500bp upstream of the transcription start site, and TF_highConf (TFs, annotated to the motif according to 'motifAnnot_highConfCat' of RcisTarget) were considered associated with TFs of interest.

Human articular chondrocyte isolation

Osteoarthritic tissue samples were obtained as described above. Chondrocytes were isolated from finely diced cartilage tissue by digestion in 0.08% collagenase type II (Worthington Biochemical, Lakewood NJ) in growth medium (Dulbecco's modified Eagle's medium, containing 10% foetal bovine serum, 100 units/ml penicillin and 100 units/ml streptomycin, all from Thermo Fisher Scientific) overnight at 37°C with agitation. Isolated cells were washed three times in growth medium without collagenase, plated at 20,000 cells/cm², and grown for 2 passages at 37°C in 5% CO₂ with 1:2 split ratios before use in experiments.

Cell culture

For analysis of the ratios of alternatively polyadenylated transcripts the following cell types were screened: SW1353 chondrosarcoma, AC10 ventricular cardiomyocytes, MG63 osteosarcoma, NTM5 trabecular network cells, A549 lung carcinoma and MDA-MB231 breast adenocarcinoma cells. All cell lines were cultured in growth medium at 37°C in 5% CO₂. For mRNA decay analysis, confluent cell monolayer cultures of the cell lines of interest were cultured in standard media supplemented with 1ng/ml actinomycin for 0, 2, 4, 6 and 8 hours before lysis using TRIzol reagent. Purification of cultured cell total RNA was performed by chloroform-based phase separation and subsequent precipitation with isopropanol. RNA was redissolved in RNAse free H₂O and stored at -80°C until required [9]. Cytokine stimulation of human articular chondrocytes was performed on cells at confluence by adding 10ng/ml interleukin-1β (IL-1β, Peprotech, London UK) to standard media and stimulating cells for up to 6 hours.

qRT-PCR analysis

cDNA was synthesized using 1μg total RNA with 200 units moloney murine leukemia virus reverse transcriptase (Promega, Southampton UK), primed with 2μM oligo dT (Thermo Fisher Scientific). qPCR was performed on a Roche Lightcycler 96 using Takyon ROX SYBR MasterMix (Eurogentec, Seraing Belgium) using transcript specific primer pairs at a concentration of 300nM per primer.

Primer efficiency for each assay was established by as previously described and assays were only used where efficiencies were within 10% of those of the reference gene [27]. Primer sequences for each assay can be found in Table 1. The $2^{-\Delta Ct}$ method was used to calculate gene expression levels [28] with RPS13 used as the reference expression gene [29].

RESULTS

Analysis of transcript abundance in control and osteoarthritis articular cartilage using QuantSeq Reverse

We used the QuantSeq Reverse approach to create 3' RNA sequence libraries that provide transcriptome-wide read coverage across chondrocyte polyadenylation sites. Total RNA isolated directly from snap frozen tissue was used for this and mean sample age for healthy tissues was 75 years with standard deviation (SD) of +/- 6 (n=6) and for osteoarthritic tissues mean age was 70 years with SD +/- 5 (n=6). Analysis with an unpaired t-test indicated no statistical difference between the ages of the groups (p = 0.16). Sequencing resulted in median sample reads of 4.4×10^7 with highest number of reads in a sample being 6.0×10^7 and the lowest 3.7×10^7 reads. We began our analysis by examining differential expression of steady state transcript levels between healthy and osteoarthritic samples. Principal components analysis of the data demonstrated that the first principal component (PC) was responsible for 40% of the experimental variance and the second principal component 13 % of the variance. There was a clear separation between the healthy and osteoarthritic samples across PC1 (Figure 1A). Analysis of differential expression revealed that osteoarthritic tissue had 561 transcripts that were differentially up regulated and 630 transcripts that were differentially down regulated (where $\log_2 fc > 1$ and adjusted p value was < 0.05) (Figure 1B and supplemental table 1). Functional annotation enrichment analysis revealed a strong association of regulated genes with extracellular structure/matrix organisation, skeletal development and cell motility (Figure 1C). Examination of differentially regulated genes using RcisTarget allowed us to identify enriched transcriptional regulatory sequences associated with OA (Figure 1D). This identified that genes with promoter sequences containing Smad3 and IKZF1 binding motifs were over-represented in genes that were regulated during osteoarthritis. Regulated genes also consisted of a significant number of long non-coding RNAs that were differentially regulated in osteoarthritis (Figure 1E). We noted that very few small regulatory RNAs such as microRNAs (miRNAs) or small nucleolar RNAs (snoRNAs) were represented in the data set. When examining data from across both control and osteoarthritic samples, only 4 miRNAs (miR125b-1, miR320d-1, miR3671, miR7851) and 11 snoRNAs (SNORA11F, SNORA12, SNORA59B, SNORA72, SNORA79B and SNORD104, SNORD13, SNORD13E, SNORD38C, SNORD38-1, SNORD38-2) were detected across more than 50% of the samples analysed. No significant effect of osteoarthritis was observed on the levels of the small RNAs detected with SNORD13 which exhibited a \log_2 fold decrease of 2.0 in osteoarthritic samples with an adjusted p-value of 0.093, being the only one to exhibit a trend toward regulation by disease.

Human articular chondrocytes make widespread use of alternative polyadenylation to modify their transcripts

Making use of the Quantseq-Rev data's property of sequencing 3' – 5' from the polyA tail, we were able to use the dataset to identify chondrocyte polyadenylation site usage. Aligned reads were analysed using expressRNA to identify polyadenylation site usage in each sample. To understand general chondrocyte polyadenylation site usage the healthy and osteoarthritic samples were examined together as a complete chondrocyte dataset. We found that 90% of chondrocyte transcripts have <5 polyadenylation sites with 37% having just one isoform (Figure 2A). No transcripts were observed with more than 14 alternatively polyadenylated isoforms in our chondrocyte data set. We performed similar analysis on a human data set obtained from the polyA_DB database version 3.2 (https://exon.apps.wistar.org/polya_db/v3/), which is comprised of pooled polyadenylation site data from the analysis of 107 tissue and cell samples [30]. This polyA_DB data represents a reference of polyadenylation site usage across a wide variety of human cellular transcriptomes. When we compared chondrocyte polyadenylation site usage to that of the reference PolyA_DB data

for 10551 genes common to each dataset, it was evident that chondrocytes utilise a narrow band of the potential polyadenylation sites that have been so far defined in human cell and tissues (Figure 2B). Our chondrocyte data contained 1114 genes that exhibited >5 alternative polyadenylation sites. We performed ontological analysis of these highly alternatively polyadenylated chondrocyte genes to determine whether they were enriched for any specific characteristics. We found that for genes in chondrocytes that utilised this many polyadenylation sites were strongly enriched for functions involved in mRNA processing and splicing as well as organelle formation and function (Figure 2C).

Characterisation of highly expressed, multi-polyadenylated genes in human articular chondrocytes

To examine the implication of alternative polyadenylation events in chondrocytes, we examined the ten most strongly expressed, multi polyadenylated genes across the healthy and osteoarthritic chondrocyte data (Table 2). We found that these genes would be affected in varying ways based upon the polyadenylation site usage identified. Some genes exhibited simple truncations of the 3'UTR when more proximal sites were used (DCN, SPARC, PRELP, CILP, EGR1). Some of these genes (DCN, PRELP, CILP) did not always exhibit a distal polyadenylation site coincident with those present in the genomic annotation, indicating that in chondrocytes all of their transcripts are truncated relative to these canonical sites. Other genes exhibited alternative polyadenylation that led to significant changes in mRNA structure. For instance, the chondrocyte data showed that two sites are used for CFH, the more proximal of which is a well characterised alternatively spliced variant that leads to a shorter protein coding sequence. Less well characterised is the chondrocyte site usage for the CLU gene. Here, six sites were identified, three of which were close to existing annotated sites and are utilised in the majority of samples submitted to polyA_DB . Two of the other three sites are identified in a small proportion (<10%) of the samples used by polyA_DB . These three sites occur at the 3' end of CLU exon 3 and represent around a quarter of all CLU reads in chondrocytes.

Osteoarthritis does not result in widespread changes in chondrocyte polyadenylation patterns

We used the expressRNA tool to determine the differential usage of polyadenylation sites between healthy and osteoarthritic cartilage samples. We found that disease specific polyadenylation is not a widespread process with just 20 genes identified as being differentially regulated in this way in the diseased samples (where adjusted p values for both distal and proximal sites relaxed to <0.1). These genes are presented in Table 3 with details of the alternative polyadenylation sites that are affected between healthy and osteoarthritic samples. As expected, many of these APA events result in a truncated 3'UTR (KMT2A, SMAD5, SLC25A37, TMOD1) however there are examples of other events which would lead to more substantial changes to transcript structure. For instance, we found that the use of a previously identified intronic polyadenylation site in the OSMR gene [31] is significantly downregulated in osteoarthritis. Similarly, the ratio between the use of an established 3'UTR site and an intronic site in the NT5E gene is disrupted in osteoarthritis. One of the most strongly regulated polyadenylation events was found for the lncRNA NEAT1. Healthy chondrocytes primarily use a distal NEAT1 polyadenylation site, encoded at chr11:65,426,254, with a lower-level usage of sites 110bp and 549bp upstream of this. Osteoarthritic chondrocytes retain the same lower level usage of each upstream site usage whilst significantly reducing the use of the distal site.

Intronic polyadenylation of OSMR mRNA leads to altered transcript decay rate and can be influenced by inflammatory cytokine stimulation.

Focussing firstly on examples of genes that exhibited intronic polyadenylation we were able to produce qPCR assays to assess expression of variants of the OSMR and ANOS1 transcripts. We examined read counts for proximal and distal polyadenylation sites for these two genes and found that osteoarthritis led to a reduced use of the proximal, intronic site for OSMR whilst inducing the use of a distal site for ANOS1 (Figure 3A). We next examined the expression and rates of mRNA decay of the isoforms of these two transcripts in osteoarthritic human chondrocytes (Figure 3B). For the OSMR gene, the use of the intronic polyadenylation site, which results in a truncated protein coding domain, also led to significant change in post-transcriptional regulation compared to the full-

length transcript. The shorter transcript was less stable with a half-life of 9 hours whereas the full-length transcript's decayed slowly (half-life > 24 hours). Interestingly, despite originating from osteoarthritic cartilage, the cultured chondrocytes expressed higher levels of the proximal isoform. For ANOS1, each isoform was very stable (>24-hour half-life) and no difference was observed in their decay rates (Figure 3B). We examined whether the inflammatory cytokine IL-1 β , which is known to strongly regulate many osteoarthritis associated genes, affected the use of the short and long isoforms of each gene in HACs (Figure 3C). We found that there was a significant reduction in the ratio of short to long isoform usage for OSMR over a 6-hour exposure to IL-1 β , consistent with the shift in ratio observed between healthy and osteoarthritic tissue. The ratio of ANOS1 isoforms was not significantly affected by the IL-1 β treatment.

mRNA decay is affected by polyadenylation status of the KMT2A transcript.

Several transcripts that were differentially polyadenylated in osteoarthritis were characterised by differences in length of the 3'UTR. To analyse these isoforms with qPCR we were only able to specifically identify longer, distally polyadenylated isoforms because they contain a specific region of 3'UTR sequence. We therefore intended to use ratios of "distal" to "proximal and distal" site usage to allow us to differentiate between APA isoforms and allow us to determine how they affected transcript decay rates. For many targets, designing discriminatory assays with satisfactory amplification efficiencies was challenging. We were, however, able to identify primer pairs with comparable amplification efficiencies that could be used to analyse of the KMT2A and SLC25A37 isoforms. Each of these genes exhibited significantly reduced use of a proximal polyadenylation site in osteoarthritic samples (Figure 4A). Using qPCR, we found it difficult to differentiate between expression levels of proximal and distal isoforms for the genes KMT2A, and SLC25A37 in both HAC and SW1353 cells (Figure 4B). Because these were genes that exhibited lower proximal site usage in OA this indicated that in cultured human chondrocytes the predominant isoform uses the distal polyadenylation site. Unsurprisingly, the results of actinomycin D chase experiments in chondrocytes led to similar results with each primer pair, as the distal isoform was likely to be predominately detected. We still wanted to understand how the APA of these genes might affect mRNA decay so we screened a range of cell types of different tissue origins in an attempt to identify a model system where a higher ratio of short and long isoforms were expressed (Figure 4C). For SLC25A37, we found that a ratio of 1 in qPCR products was generally consistent across all of the cell types we analysed, again indicating mostly proximal site usage. However, for KMT2A we found that two of the cell lines screened, AC10 and MDA-MB-231, exhibited short to long isoform ratios of 2 and above indicating that they were also expressing the proximally polyadenylated form. We also measured APA usage for OSMR and ANOS1 in the cell lines to determine whether there was any common regulation of the osteoarthritic APA sites across different cell types. This did not seem to be the case: ANOS1 chiefly exhibited preponderance to proximal site usage across all of the lines. In contrast, OSMR site usage varied between cell type, with A549 and NTMS cells making less use of the intronic proximal site compared to the other four cell lines. Based upon our cell line data, we chose to use the AC10 cells to examine whether APA affected mRNA decay rates for KMT2A (Figure 4D). Using our proximal and distal qPCR assays we found that the product of the distal, long isoform specific PCR was more stable than the product of the proximal PCR which detects both long and short isoforms. We can infer from this that the shorter KMT2A transcript is less stable than the longer one.

DISCUSSION

This is the first study to examine in detail the use of polyadenylation sites by articular chondrocytes and the effect of osteoarthritis on this process. It has confirmed that chondrocytes make widespread use of APA to create heterogeneity in their transcriptomes. However, change in the pattern of APA is not a widespread feature of the transcriptomes of osteoarthritic chondrocytes.

QuantSeq reverse leads to 3' end reads only and is particularly well suited to studying polyadenylation patterns because each read sequences from the polyA tail into the 3'UTR of a transcript. This approach makes the technique less efficient than other 3' sequencing approaches for transcript expression profiling but does not preclude such analyses from being conducted.

Encouragingly, we have found that our dataset allowed us to identify many transcripts with altered abundance between osteoarthritic and age matched normal cartilage tissue. Many of the highly regulated genes observed have previously been identified in other transcriptional profiling studies examining the disease [5]. Furthermore, gene ontology analysis shows that regulated genes are enriched for processes associated with the skeleton, joints and connective tissue, areas that are all known to be significantly affected by osteoarthritis. Our use of age matched human control and diseased samples allows us to build on previous transcriptomic studies. Our healthy samples were sourced from a commercial supplier, whilst we collected osteoarthritic tissue from a local hospital following knee arthroplasty procedures. PCA analysis of the transcriptome data showed that there was significantly less variance in the osteoarthritic samples compared with the healthy, a finding that we have observed in other RNA-sequencing studies [32]. This may reflect a greater homogeneity of phenotype among late-stage osteoarthritic samples, in line with a growing opinion that differing age related mechanisms can converge into a common end stage form of osteoarthritis [33].

A characteristic property of the QuantSeq Reverse technique is that sequences originate within the polyA tail and proceed 3'-5' into the transcript, meaning that every read specifically identifies a transcript polyadenylation site. This has allowed for fresh insights into transcript regulation in osteoarthritic chondrocytes at the post-transcriptional level by identifying patterns of chondrocyte polyadenylation site usage as well as determining the extent of alternative polyadenylation that occurred between age matched healthy and osteoarthritic cartilage samples. We confirmed that chondrocytes make widespread use of alternative polyadenylation to enhance the diversity of their transcript pool although the number of sites used is substantially lower than those identified across multiple human tissue and cell types in the polyA_DB database [34]. It is therefore likely that the chondrocyte employs a specific pattern of polyadenylation usage, some aspects of which contribute to phenotype-specific gene functionality. An example of this would be the clusterin gene where a quarter of the chondrocytes' reads indicate relatively high use of polyadenylation sites towards the 3' end of exon 3, which have been previously identified in this region but are not commonly observed in other cell types. Exon 3 terminating transcripts have been annotated as CLU-204 and CLU-214 and are predicted to code a truncated protein, although they would be polyadenylated directly on the protein coding region with no 3'UTR. The polyadenylation sites identified in the chondrocytes would actually lead to transcripts between 20 to 83bp shorter again than those annotations. It is currently not understood what the function of these truncated clusterin transcripts is or whether they are even translated.

It was instructive that chondrocytes from osteoarthritic cartilage do not exhibit widespread altered usage of polyadenylation sites, indicating that a systematic change in the processes controlling APA is not a significant feature of the disease. This distinguishes the processes occurring in osteoarthritis from other pathologies. For instance, many cancers exhibit global regulation of alternative polyadenylation events [15, 35]. Despite osteoarthritis lacking these widespread changes, a small number of alternatively polyadenylated transcripts emerged from our analysis, which potentially have a capacity to contribute to the disease mechanism. One of the most interesting is the oncostatin-M receptor OSMR. The use of an intronic polyadenylation site has already been demonstrated for this gene, which results in a truncated, secreted form of the receptor that can act as an oncostatin-M antagonist [31]. We have shown here that the use of the intronic site is decreased in osteoarthritic cartilage, which could lead to reduced levels of this antagonistic protein isoform and increased oncostatin-M signalling in osteoarthritic tissue. It was interesting to note that stimulation of chondrocytes with IL-1 β led to an decrease in the ratio of polyadenylation usage at the intronic site versus the full length form, indicating that inflammatory conditions are able to promote a similar reduction in levels of the oncostatin-M antagonistic form. Inflammatory cytokine stimulation can promote many of the catabolic processes that lead to extracellular matrix destruction in osteoarthritis and oncostatin M has recently been associated with inflammatory subtypes of osteoarthritis [36]. Understanding how loss of oncostatin-M antagonism might contribute to this warrants further investigation.

The gene KMT2A encodes a lysine methyl transferase, and its expression was down-regulated overall in osteoarthritic tissue. It has previously been linked to a potential causal role in altered

cartilage thickness through genome wide association studies [37]. It also exhibited an almost complete absence of the use of a proximal polyadenylation site within its 3'UTR in osteoarthritis, which could influence the dynamics of post-transcriptional processing. Because our chondrocyte primary cell models are from osteoarthritic tissue we were unable to identify whether this alteration in 3'UTR length affected mRNA decay because it appears that the cells only express the longer version. However, following a screen of a variety of cell lines we found that AC10 cells appear to produce both long and shorter isoforms of the KMT2A transcript and that shortening of its 3'UTR leads to a less stable mRNA. This polyadenylation event may therefore play an important role in maintaining KMT2A levels in healthy cartilage and it would be useful to elucidate the mechanisms that lead to its use being suppressed in osteoarthritis.

Long noncoding RNAs have been increasingly studied in relation to osteoarthritis recently and the DeSeq2 analysis of the QuantSeq Rev data identified 18 significantly regulated lncRNAs that had not previously been associated with joint pathology. Furthermore, an alternative polyadenylation event was identified for the lncRNA NEAT1, which has been extensively studied in a variety of tissues and has been found to function as microRNA sponge chondrocytes [38]. The functional significance of this 110bp difference is not understood although the sequence in this region is more strongly conserved the rest of the 3' half of most NEAT1 transcript isoforms. A limitation of the QuantSeq-Reverse approach appears to have been a poor coverage of small non-coding RNAs such as microRNAs and snoRNAs. We did not specifically prepare our RNA samples to enrich for small RNAs, which may explain this. These molecules are known to perform important regulatory roles in chondrocytes and can be regulated in osteoarthritis. They are polyadenylated and so our inability to fully characterise them in our samples represent a limitation of the study. Interestingly, among the small RNAs that we were able to detect were potential biomarkers for osteoarthritis, such as miR-320d [39] and SNORD13 [40].

Overall, this study has demonstrated that altered polyadenylation site usage is not a widespread characteristic of osteoarthritis, but that a small number of genes are affected and the implications of the changes could have significant potential to affect their functions. Further understanding of how to control polyadenylation in chondrocytes across the transcriptome is therefore unlikely to be required to combat osteoarthritis but specific polyadenylation events could still represent important targets for modulation of the disease.

Assay name	Forward Primer sequence (5'-3')	Reverse primer sequence (5'-3')
RPS13	GGCTTTACCCTATCGACGCA	GTCAGATGTCAACTTCAACCAAGTG
OSMR proximal	GCAGTAGGTTGTCTGGGTCA	TGAGCCTGGAACAAACAGCA
OSMR distal	TCATTACAGCGGAGGTGAG	GCGTGCATCCATGAGGAGAA
ANOS1 proximal	CCAGATCCTGCCTTCCGTAG	AGTGCAGCACACAGAATGAC
ANOS1 distal	ATGCAGATGCCTGGCCATT	AGGCTCTGTGGAGTACAGTGA
KMT2A proximal	CACCTACAGCGTCTGTGCGAA	CACCGGAGGTGCTAGGAATC
KMT2A distal	GGGGTTCCACTAGTGTCTGC	TTCAGGACCCATCAGTGCAT
SLC25A37 proximal	GGTGCATCTTACCGAGGAGG	GTGCATGGTCCCAAGATGGA
SLC25A37 distal	GTGTGCTTGTGCGTGTCTAC	TGTTGCCTTTTCGTTACCTG

TABLE 1: Sequences of primers used in qRT-PCR assays.

Gene Name, (Gene Symbol), <i>Ensembl Gene ID</i>	Gene Locus (Strand)	Sites (Proximal to Distal)	Site Distance from Proximal Site (bp)	Site Mean Read Count +/- SD
Decorin, (DCN), <i>ENSG00000011465</i>	chr12:9114048 3-91182823 (Antisense)	91145968	0	12270 +/- 8494
		91145870	98	19760 +/- 13585
		91145526	442	51032 +/- 39570
noncoding nuclear-enriched abundant transcript 2, (MALAT1), <i>ENSG00000251562</i>	chr11:6549768 7-65506515 (Sense)	65499339	0	28517 +/- 21758
		65499659	320	21804 +/- 17950
		65499698	359	100707 +/- 98832
		65499736	397	99063 +/- 96843
		65499776	437	50554 +/- 37148
		65499794	455	5531 +/- 4083
		65500522	1183	6120 +/- 6761
		65500558	1219	5446 +/- 5383
Complement Factor H, (CFH), <i>ENSG00000000971</i>	chr11:6549768 7-65506515 (Sense)	196701562	0	38800 +/- 68040
		196747501	45939	8984 +/- 9034
Clusterin, (CLU), <i>ENSG00000120885</i>	chr8:27596916- 27614699 (Antisense)	27609020	0	2143 +/- 1472
		27608995	25	6215 +/- 3978
		27608957	63	7438 +/- 4695
		27597996	11024	33471 +/- 22377
		27597813	11207	8339 +/- 9719
		27597728	11292	21976 +/- 21562
Osteonectin, (SPARC), <i>ENSG00000113140</i>	chr5:15166109 5-151686974 (Antisense)	151662435	0	22006 +/- 29354
		151661449	986	4341 +/- 5672
60S ribosomal protein L7a, (RPL7A), <i>ENSG00000280858</i>	chr9:13334821 7-133351425 (Sense)	133351412	0	4060 +/- 2684
		133351424	12	18479 +/- 12447
Prolargin, (PRELP), <i>ENSG00000188783</i>	chr1:20347580 5-203491351 (Sense)	203487161	0	17465 +/- 13603
		203489613	2452	16419 +/- 13500
Cartilage intermediate layer protein, (CILP), <i>ENSG00000138615</i>	chr15:6519475 9-65211472 (Antisense)	65196529	0	3135 +/- 4161
		65196251	278	13156 +/- 16572
		65196000	529	4965 +/- 5868
H3 histone family member 3B, (H3F3B), <i>ENSG00000132475</i>	chr17:7577643 3-75785892 (Antisense)	75778250	0	2376 +/- 2438
		75778023	227	2255 +/- 2841
		75777456	794	12733 +/- 12977
Early Growth Response 1, (EGR1), <i>ENSG00000120738</i>	chr5:13846547 8-138469302 (Sense)	138469224	0	2641 +/- 2780
		138469240	16	729 +/- 649
		138469301	77	12484 +/- 16106

TABLE 2: The 10 most highly expressed alternatively polyadenylated genes in articular chondrocytes regardless of disease state.

Gene Locus	Strand	ENSEMBL gene_id	Gene Symbol	Proximal/distal polyA sites	Proximal/distal feature	Proximal log2fc	Distal log2fc	Proximal adjusted p value	Distal adjusted p value
chr11:11848273-7-118526831	+	ENSG00000118058	KMT2A	118523915, 118526830	utr3/utr3	2.06168	-1.67404	0.01625	0.00558
chr11:10812073-10858795	-	ENSG00000236287	ZBED5	10852797, 10852747	utr3/utr3	2.29148	-2.94875	0.01812	0.0418
chr9:97501179-97601742	+	ENSG00000136842	TMOD1	97600155, 97601252	utr3/utr3	1.51617	-1.24927	0.02117	0.01554
chr21:15730024-15880068	+	ENSG00000155313	USP25	15878905, 15880062	utr3/utr3	-1.52865	3.44074	0.00356	0.054
chr11:65422773-65445539	+	ENSG00000245532	NEAT1	65426414, 65426524	exon/exon	-0.82173	1.04492	0.06951	0.08359
chr19:45349836-45370917	-	ENSG00000104884	ERCC2	45351594, 45349837	utr3/utr3	-1.70459	2.32543	0.0418	0.07641
chr11:11154027-9-111561744	+	ENSG00000204381	LAYN	111560953, 111561510	utr3/utr3	-1.46532	2.04648	0.06024	0.09604
chr5:136132844-136188746	+	ENSG00000113658	SMAD5	136179501, 136182732	utr3/utr3	-3.68803	1.67826	0.08576	0.00919
chr9:129634603-129642168	-	ENSG00000148331	ASB6	129636893, 129634604	utr3/utr3	-1.45783	1.29424	0.09604	0.08311
chr12:51186935-51217707	-	ENSG00000184271	POU6F1	51189184, 51186936	utr3/utr3	1.70409	-1.32788	0.08359	0.0527
chr8:23528955-23575462	+	ENSG00000147454	SLC25A37	23572547, 23575460	utr3/utr3	0.68296	-1.24319	0.06512	0.07641
chr2:44316280-44361861	-	ENSG00000138078	PREPL	44318763, 44317605	utr3/intron	-1.54952	2.27364	0.01646	0.04608
chr5:38845857-38945595	+	ENSG00000145623	OSMR	38886653, 38935628	intron/utr3	0.7451	-1.81506	0.03228	0.00233
chr5:132556018-132646348	+	ENSG00000113522	RAD50	132616045, 132642473	utr3/utr3	3.07214	-5.49643	0.00683	0.04908
chr3:48241099-48299252	+	ENSG00000164048	ZNF589	48270986, 48273225	utr3/intron	-0.70227	0.75939	0.04908	0.05682
chr7:77798791-77957502	+	ENSG00000006576	PHTF2	77940457, 77957223	utr3/utr3	3.12642	-2.27133	0.05189	0.01812
chr2:85751343-85791382	+	ENSG00000168874	ATOH8	85775209, 85791376	exon/utr3	3.02271	-3.80491	0.01812	0.05189
chr6:85449583-85495790	+	ENSG00000135318	NT5E	85452154, 85495783	intron/utr3	1.14747	-1.70338	0.04908	0.0273
chrX:8528873-8732136	-	ENSG00000011201	ANOS1	8536187, 8528874	intron/utr3	3.04985	-3.07918	0.00008	0.00008
chr12:19404044-19720800	+	ENSG00000139154	AEBP2	19441779, 19520222	intron/utr3	3.11654	-1.15156	0.09497	0.00558

TABLE 3: Genes affected by alternative polyadenylation in osteoarthritic chondrocytes with corrected p values for distal and proximal sites set to <0.1.

FIGURE LEGENDS

FIGURE 1: (A) Principal components analysis following determination of differential gene expression in QuantSeq-Rev dataset using DeSeq 2. Red points are control samples and blue points are osteoarthritic samples. (B) Volcano plot showing differential expression of genes between datasets, with genes significantly unregulated in osteoarthritis highlighted in red and genes significantly downregulated in osteoarthritis highlighted in blue (C) Gene ontology analysis of genes that were identified as being differentially regulated in osteoarthritic samples. The colour of each circle represents the p value associated with the enrichment whilst circle size is associated with the number of genes associated with the ontological term. (D) Transcription factor binding motifs enriched in the promoters of genes that are differentially regulated by osteoarthritis. (E) Heatmap illustrating the expression of lncRNAs that are differentially expressed in osteoarthritic samples compared to controls.

FIGURE 2: (A) The number of cartilage genes exhibiting different numbers of polyadenylation sites plotted as a percentage all genes. (B) Counts for the number of genes with different numbers of polyadenylation sites identified across control and osteoarthritis cartilage samples combined (blue circles) and human data from polyA_DB (orange circles). (C) Gene ontology analysis of genes where > 5 polyadenylation sites have been identified across all the cartilage samples analysed. The colour of each circle represents the p value associated with the enrichment whilst circle size is associated with the number of genes associated with the ontological term.

FIGURE 3: (A) QuantSeq-Rev read numbers for proximal (intronic) and distal (3'UTR) polyadenylation sites for the OSMR and ANOS1 genes, which were differentially regulated between control and osteoarthritic (OA) samples. OSMR proximal site is at CHR5:38886653, OSMR distal site is at CHR5:38935628, ANOS1 proximal site is at CHRX:8536187, ANOS1 distal site is at CHRX:8528874. (B) RT-qPCR analysis of mRNA decay rates of OSMR and ANOS1 transcripts that use intronic or canonical 3'UTR polyadenylation sites in human articular chondrocytes treated with actinomycin D. Mean and standard error shown, n = 3. (C) RT-qPCR analysis of the ratio of intronic to canonical 3'UTR transcript levels of OSMR and ANOS1 in HAC treated with 10ng/ml interleukin-1 β (IL-1 β), individual data points are presented (n = 4) whilst line represents mean value, p values represent result of one way analysis of variance.

FIGURE 4: (A) QuantSeq-Rev read numbers for proximal and distal polyadenylation sites for the KMT2A and SLC25A37 genes, usage of which leads to different length 3'UTRs and is differentially regulated between control and osteoarthritic (OA) samples. KMT2A proximal site is at CHR11:118523915, KMT2A distal site is at CHR11:118526830, SLC25A37 proximal site is at CHR8:23572547, SLC25A37 distal site is at CHR8:23575460. (B) RT-qPCR analysis of mRNA decay rates of KMT2A and SLC25A37 using primer pairs that are specific to only longer 3'UTR isoforms (long only) or both proximal and ideally polyadenylated isoforms (both) in human articular chondrocytes treated with actinomycin D. Mean and standard error shown, n = 3. (C) Heat map showing ratio of expression of intronic to canonical 3'UTR (for OSMR and ANOS 1) or long only to both for (KMT2A and SLC25A37) in a panel of different cell lines. (D) RT-qPCR analysis of mRNA decay rates of long only or both polyadenylated isoforms of KMT2A in AC10 cells. Mean and standard error shown, n = 3.

REFERENCES

1. Katsoula, G., P. Kreitmaier, and E. Zeggini, *Insights into the molecular landscape of osteoarthritis in human tissues*. *Curr Opin Rheumatol*, 2022. **34**(1): p. 79-90.
2. Boer, C.G., et al., *Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations*. *Cell*, 2021. **184**(24): p. 6003-6005.
3. Reynard, L.N., et al., *Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation*. *Hum Mol Genet*, 2011. **20**(17): p. 3450-60.
4. Hirata, M., et al., *C/EBPbeta and RUNX2 cooperate to degrade cartilage with MMP-13 as the target and HIF-2alpha as the inducer in chondrocytes*. *Hum Mol Genet*, 2012. **21**(5): p. 1111-23.
5. Soul, J., et al., *Stratification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage*. *Ann Rheum Dis*, 2018. **77**(3): p. 423.
6. Katsanou, V., et al., *The RNA-binding protein Elavl1/HuR is essential for placental branching morphogenesis and embryonic development*. *Mol Cell Biol*, 2009. **29**(10): p. 2762-76.
7. Kobayashi, T., et al., *Dicer-dependent pathways regulate chondrocyte proliferation and differentiation*. *Proc Natl Acad Sci U S A*, 2008. **105**(6): p. 1949-54.
8. Son, Y.O., et al., *RNA-binding protein ZFP36L1 regulates osteoarthritis by modulating members of the heat shock protein 70 family*. *Nat Commun*, 2019. **10**(1): p. 77.
9. Tew, S.R., et al., *Transcriptome-wide analysis of messenger RNA decay in normal and osteoarthritic human articular chondrocytes*. *Arthritis Rheumatol*, 2014. **66**(11): p. 3052-61.
10. Derti, A., et al., *A quantitative atlas of polyadenylation in five mammals*. *Genome Res*, 2012. **22**(6): p. 1173-83.
11. Hoffman, Y., et al., *3'UTR Shortening Potentiates MicroRNA-Based Repression of Proliferation Genes in Proliferating Human Cells*. *PLoS Genet*, 2016. **12**(2): p. e1005879.
12. Davis, M.J., et al., *Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units*. *PLoS Genet*, 2006. **2**(4): p. e46.
13. Vorlova, S., et al., *Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation*. *Mol Cell*, 2011. **43**(6): p. 927-39.
14. Zhou, R., et al., *SCAPE: a mixture model revealing single-cell polyadenylation diversity and cellular dynamics during cell differentiation and reprogramming*. *Nucleic Acids Res*, 2022.
15. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells*. *Cell*, 2009. **138**(4): p. 673-84.
16. Xia, Z., et al., *Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types*. *Nat Commun*, 2014. **5**: p. 5274.
17. Jobbins, A.M., et al., *Dysregulated RNA polyadenylation contributes to metabolic impairment in non-alcoholic fatty liver disease*. *Nucleic Acids Res*, 2022. **50**(6): p. 3379-3393.
18. Moll, P., et al., *QuantSeq 3' mRNA sequencing for RNA quantification*. *Nature Methods*, 2014. **11**(12): p. i-iii.
19. Flannery, C.R., et al., *Articular cartilage superficial zone protein (SZP) is homologous to megakaryocyte stimulating factor precursor and is a multifunctional proteoglycan with potential growth-promoting, cytoprotective, and lubricating properties in cartilage metabolism*. *Biochem Biophys Res Commun*, 1999. **254**(3): p. 535-41.
20. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner*. *Bioinformatics*, 2013. **29**(1): p. 15-21.
21. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing*. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 1995. **57**(1): p. 289-300.
22. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs*. *Nucleic Acids Res*, 2017. **45**(D1): p. D353-D361.
23. Fabregat, A., et al., *The Reactome pathway Knowledgebase*. *Nucleic Acids Res*, 2016. **44**(D1): p. D481-7.
24. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters*. *OMICS*, 2012. **16**(5): p. 284-7.

25. Luo, W. and C. Brouwer, *Pathview: an R/Bioconductor package for pathway-based data integration and visualization*. Bioinformatics, 2013. **29**(14): p. 1830-1.
26. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering*. Nat Methods, 2017. **14**(11): p. 1083-1086.
27. Martin, I., et al., *Quantitative analysis of gene expression in human articular cartilage from normal and osteoarthritic joints*. Osteoarthritis Cartilage, 2001. **9**(2): p. 112-8.
28. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method*. Methods, 2001. **25**(4): p. 402-8.
29. de Jonge, H.J., et al., *Evidence based selection of housekeeping genes*. PLoS One, 2007. **2**(9): p. e898.
30. Wang, R., et al., *A compendium of conserved cleavage and polyadenylation events in mammalian genes*. Genome Res, 2018. **28**(10): p. 1427-1441.
31. Diveu, C., et al., *Molecular and functional characterization of a soluble form of oncostatin M/interleukin-31 shared receptor*. J Biol Chem, 2006. **281**(48): p. 36673-82.
32. Peffers, M.J., et al., *SnoRNA signatures in cartilage ageing and osteoarthritis*. Sci Rep, 2020. **10**(1): p. 10641.
33. Dell'Isola, A. and M. Steultjens, *Classification of patients with knee osteoarthritis in clinical phenotypes: Data from the osteoarthritis initiative*. PLoS One, 2018. **13**(1): p. e0191045.
34. Wang, R., et al., *PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes*. Nucleic Acids Res, 2018. **46**(D1): p. D315-D319.
35. Xia, Z., et al., *Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3' UTR landscape across seven tumour types*. Nature Communications, 2014. **5**.
36. Garcia, J.P., et al., *Association between Oncostatin M Expression and Inflammatory Phenotype in Experimental Arthritis Models and Osteoarthritis Patients*. Cells, 2021. **10**(3).
37. Castano-Betancourt, M.C., et al., *Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis*. PLoS Genet, 2016. **12**(10): p. e1006260.
38. Tu, Y., et al., *MicroRNA-377-3p alleviates IL-1beta-caused chondrocyte apoptosis and cartilage degradation in osteoarthritis in part by downregulating ITGA6*. Biochem Biophys Res Commun, 2020. **523**(1): p. 46-53.
39. Ali, S.A., et al., *Circulating microRNAs differentiate fast-progressing from slow-progressing and non-progressing knee osteoarthritis in the Osteoarthritis Initiative cohort*. Ther Adv Musculoskelet Dis, 2022. **14**: p. 1759720X221082917.
40. Castanheira, C., et al., *Equine synovial fluid small non-coding RNA signatures in early osteoarthritis*. BMC Vet Res, 2021. **17**(1): p. 26.

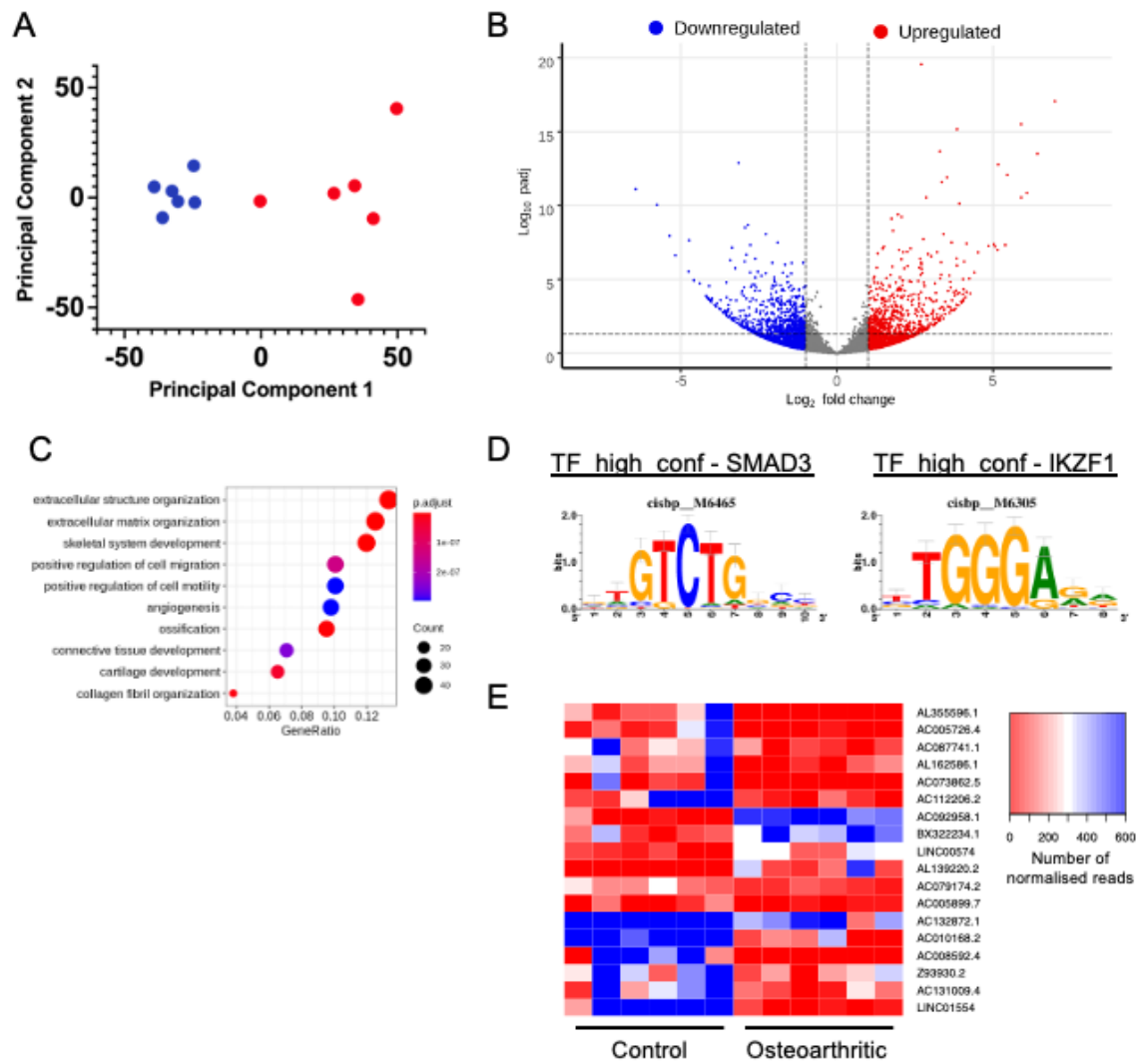


FIGURE 1

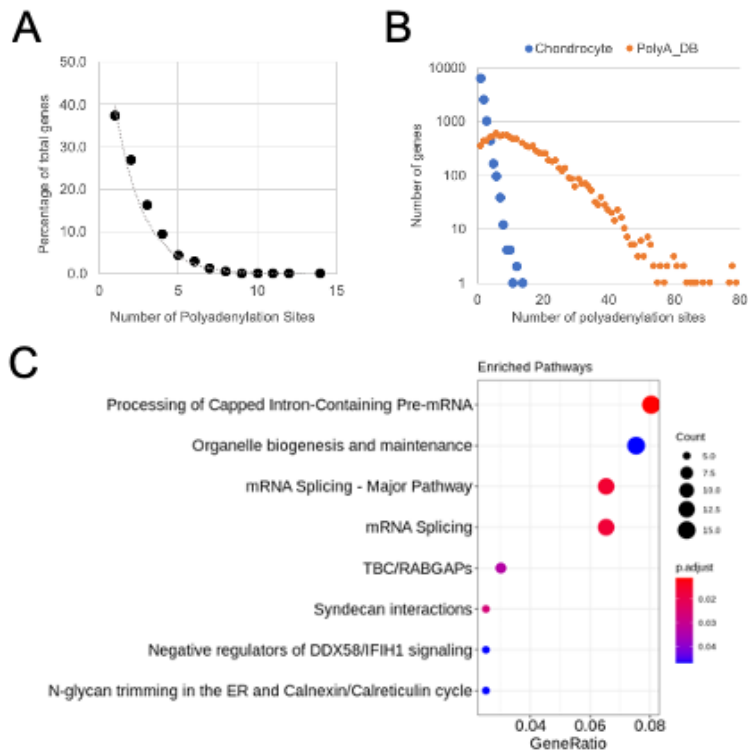


FIGURE 2

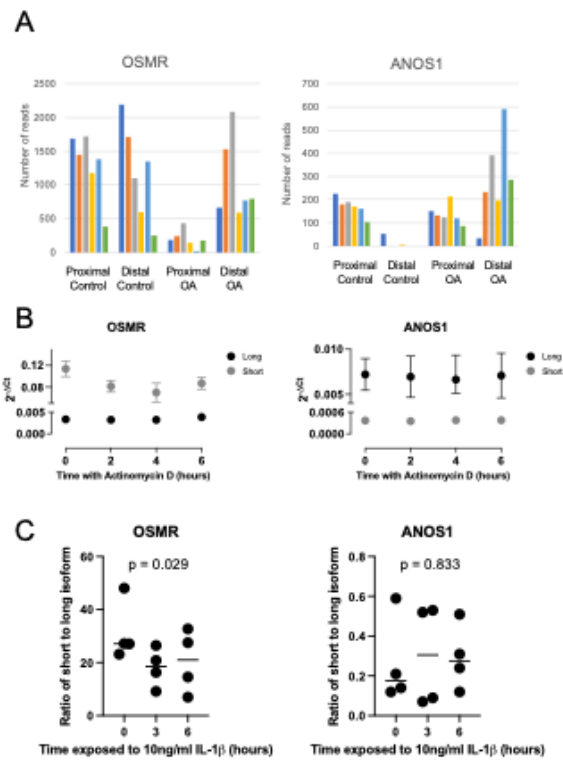


FIGURE 3

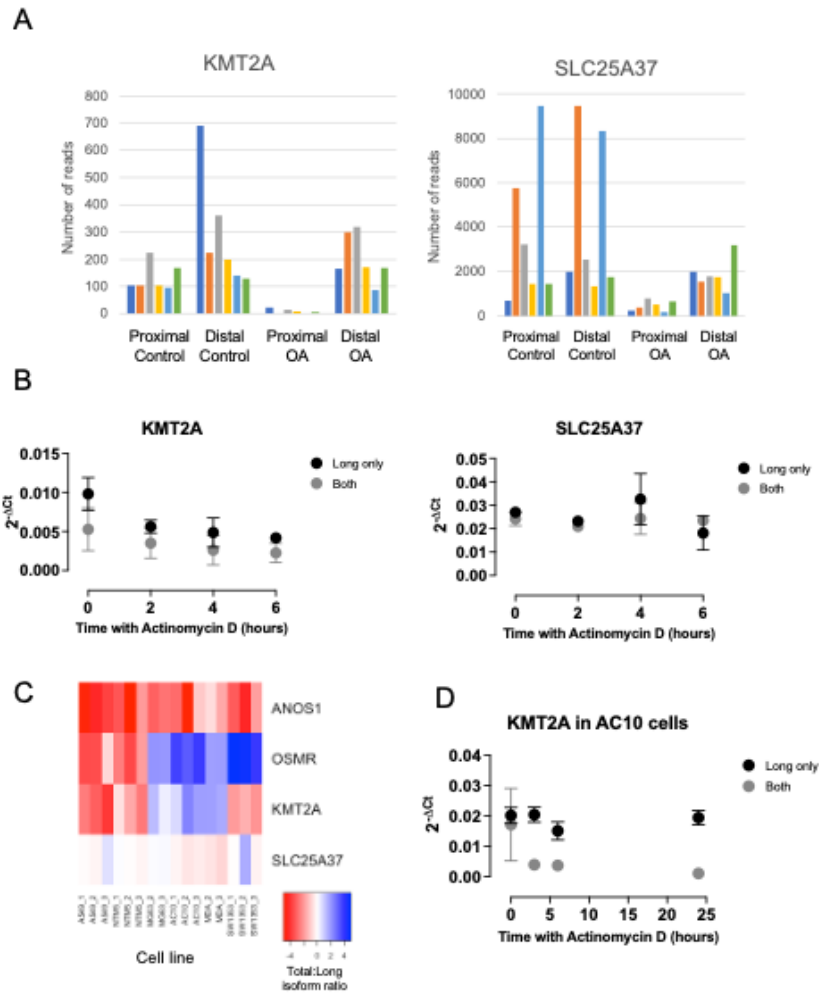


FIGURE 4

