

RESEARCH ARTICLE

Open Access

# Transcriptome analysis of ageing in uninjured human Achilles tendon

Mandy Jayne Peffers<sup>1\*</sup>, Yongxiang Fang<sup>2</sup>, Kathleen Cheung<sup>3</sup>, Tim Koh Jia Wei<sup>4</sup>, Peter David Clegg<sup>1</sup> and Helen Lucy Birch<sup>5</sup>

## Abstract

**Introduction:** The risk of tendon injury and disease increases significantly with increasing age. The aim of the study was to characterise transcriptional changes in human Achilles tendon during the ageing process in order to identify molecular signatures that might contribute to age-related degeneration.

**Methods:** RNA for gene expression analysis using RNA-Seq and quantitative real-time polymerase chain reaction analysis was isolated from young and old macroscopically normal human Achilles tendon. RNA sequence libraries were prepared following ribosomal RNA depletion, and sequencing was undertaken by using the Illumina HiSeq 2000 platform. Expression levels among genes were compared by using fragments per kilobase of exon per million fragments mapped. Differentially expressed genes were defined by using Benjamini-Hochberg false discovery rate approach ( $P < 0.05$ , expression ratios  $1.4 \log_2$  fold change). Alternative splicing of exon variants were also examined by using Cufflinks. The functional significance of genes that showed differential expression between young and old tendon was determined by using ingenuity pathway analysis.

**Results:** In total, the expression of 325 transcribed elements, including protein-coding transcripts and non-coding transcripts (small non-coding RNAs, pseudogenes, long non-coding RNAs and a single microRNA), was significantly different in old compared with young tendon ( $\pm 1.4 \log_2$  fold change,  $P < 0.05$ ). Of these, 191 were at higher levels in older tendon and 134 were at lower levels in older tendon. The top networks for genes differentially expressed with tendon age were from cellular function, cellular growth, and cellular cycling pathways. Notable differential transcriptome changes were also observed in alternative splicing patterns. Several of the top gene ontology terms identified in downregulated isoforms in old tendon related to collagen and post-translational modification of collagen.

**Conclusions:** This study demonstrates dynamic alterations in RNA with age at numerous genomic levels, indicating changes in the regulation of transcriptional networks. The results suggest that ageing is not primarily associated with loss of ability to synthesise matrix proteins and matrix-degrading enzymes. In addition, we have identified non-coding RNA genes and differentially expressed transcript isoforms of known matrix components with ageing which require further investigation.

## Introduction

The increasing number of people reaching old age provides huge challenges to society, as whereas life span increases, life quality faced by many individuals in old age is poor [1]. Whereas muscle, bone, and joint age-related disease is well recognised, the fibrous connecting tendon

tissue has received little attention, despite representing a very common site of pain and dysfunction. Epidemiological studies have revealed a clear link between age and increasing incidence of tendon injury [2,3], suggesting that the mechanical integrity of tendon declines during the ageing process.

Although it is generally accepted that a degenerative process precedes gross tendon injury, the aetiology of this process remains elusive and the definition of degeneration is poorly defined. Histological examination of painful Achilles tendon [4], dysfunctional posterior tibialis

\* Correspondence: peffs@liv.ac.uk

<sup>1</sup>Comparative Musculoskeletal Biology, Institute of Ageing and Chronic Disease, University of Liverpool, Leahurst, Chester High Road, Neston, Wirral CH64 7TE, UK

Full list of author information is available at the end of the article

[5], and supraspinatus tendon collected from cadavers [6] has revealed pathological changes, including signs of collagen fibre disruption, increased staining for glycosaminoglycan, hypercellularity, and cell shape change to a more chondroid appearance. Similar changes have been observed in macroscopically abnormal equine flexor tendon [7], another common site of age-associated tendon injury. Histological abnormalities are more often observed in older individuals [6], although the relationship with ageing and the apparent change in cell function is not clear.

Ageing is generally associated with a decline in protein synthesis [8] and a loss of cell functionality [9]. It has been suggested that early degenerative changes in tendon result from an accumulation of micro-damage within the extracellular matrix (ECM) due to an imbalance between anabolic and catabolic pathways [10]. Recent work on equine flexor tendon identified an accumulation of partially degraded collagen within the ECM of old tendons, and it was hypothesised that an inability to remove partially degraded collagen may account for reduced mechanical competency [11]. Another study found that flexor tendon explants from older horses were more susceptible to fatigue damage following cyclical loading *in vitro* than explants from young horses and that this was a cell-mediated process involving the matrix metallo-proteinases (MMPs) [12].

Cell ageing has been associated with a decreased ability to modulate inflammation resulting in a chronic low-level inflammation termed 'inflamm-aging' [13]. Recent work by Dakin and colleagues [14] measured prostaglandin E2 in injured equine flexor tendons and found that levels increased with increasing horse age but that levels of formyl peptide receptor 2/ALX, a receptor responsible for suppressing the inflammatory response, were significantly reduced. These findings intimate that aged individuals exhibit a reduced capacity to resolve inflammation and that ageing may contribute to deregulated tendon repair through these pathways.

Quantitative analysis of gene expression changes with age may help the understanding of ageing mechanisms and their interactions with age-related diseases such as tendinopathies [15]. Although microarray technology has been employed to investigate gene expression changes following tendon injury [16], in tendinopathic tissue [17], in response to cyclical strain [18] or a single loading event [19], and effect of loading on tendon healing [20], no comprehensive analysis of alterations in gene expression with age has been undertaken in tendon.

RNA-Seq can capture the whole transcriptome, including coding RNAs, isoforms produced by alternative splicing, long non-coding RNAs (lncRNAs) (the importance of which is becoming apparent in disease [21] and ageing [22-24]), and short non-coding RNAs. We have previously used RNA-Seq successfully on equine cartilage

tissue and identified an over-representation of genes with reduced expression relating to ECM, degradative proteases, matrix synthetic enzymes, cytokines, and growth factors in ageing cartilage [24].

In this study, we used RNA-Seq to comprehensively identify the human Achilles tendon transcriptome for the first time and then examine changes that occur with ageing. We hypothesised that ageing results in reduced expression of ECM-related proteins and matrix-degrading enzymes. In addition, we sought to identify previously unrecognised splice variants and non-coding RNAs associated with tendon ageing in a 'bottom-up' inductive approach.

## Methods

### Sample collection and preparation

All human Achilles tendons used in this study—RNA-Seq and quantitative real-time polymerase chain reaction (qRT-PCR)—were harvested from limbs amputated during surgical procedures to treat sarcomas at the Royal National Orthopaedic Hospital, Stanmore. Tissue collection was carried out through the Stanmore Musculoskeletal Bio-Bank, which has ethical approval from the Cambridgeshire 1 Research Ethics Committee (REC reference 09/H0304/78) to collect tissue for research into musculoskeletal conditions. All patients gave consent for their tissue to be used for musculoskeleton-related research. Local research-and-development approval for this project was given by the UCL/UCLH/RF Joint Research Office (reference number 11/0464). For RNA-Seq, tendons were collected from donors who were  $69.4 \pm 7.3$  years old (old group,  $n = 5$ , 3 female, 2 male) and donors who were  $19 \pm 5.8$  years old (young group,  $n = 4$ , 4 male). Tendon tissue was collected within 24 hours of limb removal, except for one sample in which tissue was collected within 48 hours. The Achilles tendon was dissected free from the limb. Only tendons with a normal macroscopic appearance were used for this study. A section of tissue approximately 1 cm in length was taken from the mid region of the tendon between the musculotendinous junction and the insertion site. Outer tissue (paratenon) was removed and the remaining tendon tissue placed into RNAlater (Ambion, Warrington, UK) in accordance with the instructions of the manufacturer.

### RNA extraction

Tendon was pulverising into a powder with a dismembranator (Mikro-S; Sartorius, Melsungen, Germany) following freezing in liquid nitrogen. Immediately, 20 volumes of Tri Reagent (Ambion) was added to the powdered tendon tissue and the RNA extracted and purified as described by Peffers *et al.* [25] (2013). RNA was quantified by using a Nanodrop ND-100 spectrophotometer (Labtech, Uckfield, East Sussex, UK) and assessed for purity by UV absorbance measurements at 260 and 280 nm.

### RNA-Seq analysis: cDNA library preparation and sequencing

Total RNA was analysed by the Centre for Genomic Research, University of Liverpool, for RNA-Seq library preparation and sequencing by using the Illumina HiSeq 2000 platform. Total RNA integrity was confirmed by using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Ribosomal RNA (rRNA) was depleted from 9 total RNA samples by using the Ribo-Zero™ rRNA Removal Kit (Human/Mouse/Rat; Epicentre, Madison, WI, USA) in accordance with the instructions of the manufacturer. cDNA libraries were prepared with the ScriptSeq v2 RNA-Seq library preparation kit (Epicentre) by using 50 ng rRNA depleted RNA as starting material in accordance with manufacturer protocols as previously described [24]. The final pooled library was diluted to 8 pmol before hybridisation. The dilute library (120 µL) was hybridised on one lane of the HiSeq 2000 at 2 × 100-base pair (bp) paired-end sequencing with v3 chemistry.

### Data processing

The sequence libraries for each sample were processed by using CASAVA version 1.8.2 to produce 100-bp paired-end sequence data in fastq format. The fastq files were processed by using Cutadapt version 1.2.1 [26] with option '-O 3' to trim adapter from any read if it matched the adapter sequence for 3 bp or more at the 3' end. In addition, a quality trimming was performed by using Sickle version 1.200.

The trimmed R1-R2 read pairs, for each sample, were aligned to reference sequence [27] by using TopHat2 version 2.0.10 [28] with default settings, except for the option -g 1. Read counts were obtained from the mapping results by using HTSeq-count and genome annotation [29].

The differential gene expression analysis was performed on R platform by using the edgeR package [30] and focused on the contrast of old and young donors. The count data were normalised across libraries by using trimmed mean M (TMM) values of the default methods edgeR. The tagwise dispersions were estimated and then used for logFC (log<sub>2</sub> fold change) estimating and testing. Differentially expressed genes (DEGs) were extracted by applying the threshold false discovery rate (FDR) of less than 0.05 to adjusted *P* values, which were generated by using Benjamini and Hochberg approach [31]. In addition, FPKM (fragments per kilobase of exon per million fragments mapped) values were converted from count values for comparing expression levels among genes. All sequence data produced in this study have been submitted to National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) under Array Express accession number E-MTAB-2449.

### Analysis of splice variants

Trimmed paired reads were aligned to a reference human transcriptome (Ensembl iGenomes build GRCh37) by

using Bowtie2 [32]. The alignments (BAM files) were converted into sorted SAM files by using SAMtools [33]. Parameters for TopHat were estimated by using a Picard tool (CollectInsertSizeMetrics.jar) [34]. Reads were aligned to the reference genome (Ensembl build GRCh37) by using TopHat [35], specifying mate inner distance (mean inner distance between mate pairs) and standard deviation for each sample. Mapped reads were then assembled into complete transcripts by using the splice junction mapping tool Cufflinks [36] with option -G, which uses the Ensembl reference gene track to improve mapping. Cuffmerge was used to merge the assembled transcripts into a consensus gene track from the all of the mapped samples. Cuffdiff was used to identify DEGs and differentially expressed transcripts between young and old tendon. Genes and transcripts were identified as being significantly differentially expressed with *q* values of less than 0.05, calculated by the Benjamini and Hochberg FDR correction [31].

Downstream analysis and visualisation of results, including quality control of the samples, was undertaken by using the cummeRbund package in R. Graphs were generated by using cummeRbund and the ggplot2 package [37].

### Functional analysis

To systematically determine networks, functional analyses, and canonical pathways that the DEGs might involve, we performed the pathway/network enrichment analysis using the ingenuity pathway analysis (IPA) tool from the Ingenuity Systems [38] by using a list of DEGs with values-adjusted *P* value of less than 0.05 and ±1.4 log<sub>2</sub> fold regulation. Gene symbols were used as identifiers and the Ingenuity Knowledge Base gene was used as a reference for a pathway analysis. For network generation, a data set containing gene identifiers and corresponding expression values was uploaded. Default settings were used to identify molecules whose expression was significantly differentially regulated. These molecules were overlaid onto a global molecular network contained in the Ingenuity Knowledge Base. Networks of 'network-eligible molecules' were then algorithmically generated based on their connectivity. The functional analysis identified the biological functions and diseases that were most significant to the data set. Right-tailed Fisher's exact test was used to calculate *P* values. Canonical pathways analysis identified the pathways from the IPA library that were most significant to the data set.

For isoform analysis, the Database for Annotation, Visualization and Integrated Discovery (DAVID) (DAVID bioinformatics resources 6.7) was used [39]. The web-based functional annotation tool enabled functional clustering of gene. The functional clustering tool was used for functional enrichment for DEG isoforms with values-adjusted *P* value of less than 0.05 and ±1.4 log<sub>2</sub> fold regulation.

### Real-time polymerase chain reaction

Samples of RNA from both the same pools used for the RNA-Seq analysis and an additional independent cohort harvested in the same manner ( $n=4$  young;  $16.7 \pm 2.8$  years old and  $n=4$  old;  $73.2 \pm 6.5$  years old) were used for qRT-PCR. To validate results from differentially expressed isoforms, the independent cohort was used. Moloney murine leukaemia virus (M-MLV) reverse transcriptase and random hexamer oligonucleotides (both from Promega, Southampton, UK) were used to synthesize cDNA from 1  $\mu\text{g}$  RNA in a 25  $\mu\text{L}$  reaction. PCR was performed on 1  $\mu\text{L}$  10 $\times$  diluted cDNA by employing a final concentration of 300 nM of each primer in 20  $\mu\text{L}$  reaction volumes on an ABI 7700 Sequence Detector using Primer-Design 2X Precision<sup>TM</sup> SYBR Green Mastermix (Primer Design, Southampton, UK). qRT-PCR was undertaken by using gene-specific primers (for protein-coding genes these were exon-spanning). Primers used had been validated in previous publications [40,41] and supplied by Eurogentec (Seraing, Belgium) or were designed and validated commercially (Primer Design). Steady-state transcript abundance of potential endogenous control genes was measured in the RNA-Seq data. Assays for four genes—glucose-6-phosphate isomerase (GPI), beta-actin (ACTB), ribosomal protein 13 (RSP13), and ribosomal protein 16 (RPS16)—were selected as potential reference genes as their expression was unaltered. Stability of this panel of genes was assessed by applying a gene stability algorithm [42]. RSP16 was selected as the most stable endogenous control gene. Relative expression levels were normalised to RPS16 and calculated by using the  $2^{-\Delta\text{Ct}}$  method [43]. Primers pairs used in this study are listed (Table 1). qRT-PCR analysis data was  $\log_{10}$ -transformed to ensure normal distribution and then analysed by using Student's *t* test.

### Statistical analysis

The analyses were undertaken by using edgeR [30], S-Plus (version 7.0; Tibco Software Inc., Palo Alto, CA, USA), SPSS (version 20; IBM, Portsmouth, Hampshire, UK), and Excel (2007; Microsoft, Redmond, WA, USA) software.

## Results

### Overview of RNA-Seq data

An average of 32.1 million pairs of 100-bp paired-end reads per sample were generated that aligned to the reference sequence of the human genome. Using pooled R1 and R2 files for all samples in Trimmed data gave 95.1% of called bases with of Phred score of more than 30 [44]. (See Table 2 for summary of mapping results.) Of the 63,152 human genes, between 40.5% and 47.4% had at least one read aligned; 20,322 of the genes had no reads aligned from any of the nine samples. This is similar to the output of other RNA-Seq sequencing studies [24,45].

These reads were used to estimate transcript expression of all nine samples using FPKM in order to identify the most abundant genes in tendon. Table 3 demonstrates the 25 most highly expressed genes in young and old tendon (the entire data set is in Additional file 1).

### Identification of differentially expressed genes and isoforms

A principal component analysis (PCA) plot of  $\log_2$  gene expression data indicated that the effect of age on gene expression was distinct as data were clustered in two gene groups (Figure 1A). Within the young group, two samples clustered together and two were independent of each other indicating more variability between young donors. Alterations in gene expression between young and old tendon demonstrated significant age-related changes. In total, the expression of 325 transcribed elements, including protein-coding transcripts and non-coding transcripts, small non-coding RNAs (snoRNAs), pseudogenes, lncRNAs, and a single microRNA, was significantly different in old compared with young tendon ( $\pm 1.4 \log_2$  fold change, FDR-adjusted *P* value of less than 0.05) (Figure 1B). Of these, 191 were at higher levels in the older tendon and 134 were at lower levels in the older tendon. The top 10 genes most DEG (increased and decreased) during tendon ageing are given in Table 4. The entire list of significantly DEG transcripts is presented in Additional file 2. NCBI GEO under accession number E-MTAB-2449 contains a complete list of all genes mapped. Of the 191 transcripts expressed at a higher level in old donors, 148 were known protein-coding genes. The remaining 43 genes contained 34 lncRNAs, one snoRNA, and eight pseudogenes (Table 5). Within the group where gene expression was lower in old compared with young tendon, 112 were known protein-coding genes. The remaining 22 genes contained 16 lncRNAs, one snoRNA, four pseudogenes, and a single microRNA (miRNA) (Table 6). Thus, 325 genes were input into IPA for downstream analysis, and 273 of these were mapped.

The analysis identified a number of transcript isoforms expressed in tendon, some of which were differentially expressed between young and old groups of tendon (Figure 2). In total, 183,660 isoforms were detected in young and 191,673 isoforms were detected in old tendon. Among these, 21,193 isoforms were detected only in young and 29,206 isoforms only in old. Sixty-three known isoforms were upregulated in old tendon, with 80 downregulated with an FDR-adjusted *P* value of less than 0.05 and  $\pm 1.4 \log_2$  fold regulation. The top 10 up- and down-regulated isoforms are presented in Table 7. The entire list of significantly DEG isoforms is presented in Additional file 3.

### Age-related changes in transcription factors

There was an increase in the expression of 12 transcription factors in tendon derived from older donors

**Table 1 Gene primer sequences used in RNA-Seq validation**

Name	Gene type	Ensemble gene ID/transcript ID	Primer sequence
<i>RPS16</i>	Protein-coding	ENSG00000105193	F: GAAATCCTACCGATAAGCCCA R: TTTCTTGAAACTTTAAAATCCCTCAA
<i>MMP3</i>	Protein-coding	ENSG00000149968	F: ATTCATGGAGCCAGGCTTTC R: CATTGGGTCAAACCTCAACTGTG
<i>COL1A1<sup>a</sup></i>	Protein-coding	ENSG00000108821	F: GGAGGAGAGTCAGGAAGG R: GCAACACAGTTACACAAGG
<i>ACAN<sup>b</sup></i>	Protein-coding	ENSG00000108821	F: TCGAGGACAGCGAGGCC R: TCGAGGGTGTAGCGTGTAGAGA
<i>POU3F4</i>	Protein-coding	ENSG00000196767	F: GTTCGCTCGCTCTCTCGTA R: GAAGGAAGGGAAGGGGAAA
<i>MYF5</i>	Protein-coding	ENSG00000111049	F: TGAACATAATTTCTGGTCTATATGAC R: TGTACATGTTGCTTGGTTTGGG
<i>IGF1</i>	Protein-coding	ENSG00000017427	F: CACTATGGACAGATGTAAAAGAACTA R: ACACACTGGGGACAAGAAATAAA
<i>MMP16</i>	Protein-coding	ENSG00000156103	F: ACCCGTGAACCCCTTGAGA R: AACCTGAACCTCTTGAACCTGTG
<i>TGFB3</i>	Protein-coding	ENSG00000119699	F: GTAAGAAAGTGTGGGTTTGGTTA R: AACATCTCAACTTACCATCCCTTT
<i>EGF</i>	Protein-coding	ENSG00000138798	F: ACAGGAGGCTTCGGAGTT R: AATCAGGCAATTTACTTACAATCTT
<i>COL3A1-001</i>	Protein-coding	ENST00000304636	F: CAGGTCCCAGCGGTCT R: CCTTTTGGTCCAGACACTCC
<i>COL3A1-201</i>	Protein-coding	ENST00000317840	F: GGTAGCCCTGGTGAGAGAG R: TGCCAGGAGGTCCAAAGAG
<i>FGF10-001</i>	Protein-coding	ENST00000264664	F: TGCCGTCAAAGCCATTAACA R: CATTTCCTCTATCTCTCTCTCA
<i>CRTAC1-001</i>	Protein-coding	ENST00000370597	F: ATCTTCTTCAACAACATTGCCTAC R: GGGTCTCCGTGCTCTCTAC
<i>IGF1-001</i>	Protein-coding	ENST00000337514	F: CAGCAGTCTTCCAACCCAAT R: AAGAGATGCGAGGAGGACAT
<i>XIST</i>	Lnc	ENSG00000229807	F: TCCCAGAGAATGCCTAATACTTT R: GTAGAAGAGATACGGAGTAGGAAT
<i>LINC00957</i>	Lnc	ENSG00000235314	F: GAGAGTAAGCAGACCTGGGT R: ACCTTGTCGAGTCCATCT
<i>RP11.308 N19.1</i>	Lnc	ENSG00000234323	F: GCCTCTTTCATCACTGCCGA R: TAGCAGCAGTTGGGGTGT

<sup>a</sup>[40] and <sup>b</sup>[41] denote primer pairs published previously. F, forward; Lnc, long non-coding; R, reverse.

compared with young: ALX homeobox (ALX1), insulin gene enhancer protein 1 (ISL1), lim homeobox 9 (LHX9), myocardin (MYOCD), POU domain, class 3, transcription factor 4 (POU3F4), POU3F3, paired box transcription factor 3 (PAX3), PAX6, PAX9, thyroid transcription factor 1 (NKX2-1), NKX6-1, and dachshund family transcription factor 2 (DACH2). In comparison, there was an increase in only three in younger donors: basic helix-loop-helix

family, member e40 (BHLHE40), early growth response 2 (EGR2), and zinc finger of the cerebellum 3 (ZIC3).

#### Differentially expressed genes and network analysis

DEGs (325) and differentially expressed transcript isoforms (143) associated with ageing were analysed together in IPA with the following criteria: *P* value of less than 0.05 and 1.4 log<sub>2</sub> fold change. Network-eligible

**Table 2 Summary of sequence alignment to the human genome**

Sample	Reads to align <sup>a</sup>	Reads aligned to genome	Percentage <sup>b</sup>	Reads properly paired	Percentage <sup>c</sup>	Reads with mate unmapped	Percentage <sup>d</sup>	Percentage of 'expressed' genes <sup>e</sup>
Young 1	61,516,106	55,995,223	91.03	51,821,068	92.55	4,174,155	7.45	46.65
Young 2	74,068,430	67,565,304	91.22	62,477,226	92.47	5,088,078	7.53	45.98
Young 3	65,954,168	60,771,423	92.14	56,659,046	93.23	4,112,377	6.77	47.19
Young 4	59,947,540	53,381,917	89.05	48,909,710	91.62	4,472,207	8.38	40.58
Young 5	89,707,092	80,798,459	90.07	74,516,846	92.23	6,281,613	7.77	45.01
Old 1	62,712,522	57,345,261	91.44	53,413,212	93.14	3,932,049	6.86	46.42
Old 2	46,590,704	41,788,780	89.69	38,331,156	91.73	3,457,624	8.27	40.46
Old 3	61,364,402	55,643,829	90.68	51,415,610	92.40	4,228,219	7.60	46.36
Old 4	55,361,082	49,725,122	89.82	45,615,238	91.73	4,109,884	8.27	42.44
Maximum	89,707,092	80,798,459	92.14	74,516,846	93.23	6,281,613	8.38	47.4
Mean	65,591,734	59,482,205	90.65	54,953,876	92.36	4,528,330	7.64	44.92
Minimum	46,590,704	41,788,780	89.05	38,331,156	91.62	3,457,624	6.77	40.46

The table shows the number and percentage of reads mapped to the human reference sequences. <sup>a</sup>Sum of R1 and R2 reads used in the alignment. <sup>b</sup>Percentage of reads used in the alignment that align to the reference genome. <sup>c</sup>Percentage of reads used in the alignment that align to the reference genome in the correct relative orientation to their mate. <sup>d</sup>Percentage of reads used in the alignment that align to the reference genome but whose mate does not align. <sup>e</sup>Percentage of 63,152 annotated human genes with at least one read aligned.

molecules were overlaid onto molecular networks based on information from the ingenuity pathway knowledge database and networks generated based on connectivity. (See Additional file 4 for all identified networks and their respective molecules.)

The top four scoring networks for genes differentially expressed with tendon age were from cellular function and maintenance, cellular growth and proliferation, cellular cycling, and cellular development (Figure 3). Significant IPA canonical pathways are demonstrated in Table 8, and the associated molecules of the top canonical pathways identified are in Additional file 5. These include hepatic fibrosis, oestrogen biosynthesis, and transcriptional regulatory networks in embryonic stem cells. Interestingly, skeletal and muscular disorders were identified as one of the top diseases associated with the gene set (Additional file 6).

#### Functional annotation of up- and down-regulated isoforms

There was a reduction in the DEG isoforms of 32 genes (representing 15% of the data set) relating to the ECM, degradative proteases, cytokines, and growth factors in tendon derived from older donors compared with young donors. In comparison, there was an increase in only two ECM genes (representing 1.3% of the data set) in older donors (data not shown). DAVID identified significant gene ontology (GO) terms in the upregulated and downregulated set of transcript isoforms (Table 9) with only two terms 'secreted' and 'signal' overlapping between the two groups. Interestingly, other terms are strikingly different between the upregulated and downregulated isoform data sets. Several of the top GO terms

identified in downregulated isoforms in old tendon relate to collagen and post-translational modification of collagen (for example, hydroxylation, hydroxylysine, hydroxyproline, and triple helix).

#### Confirmation of DEG by using qRT-PCR measurements of selected genes

To validate the RNA-Seq technology, selected gene expression differences noted in the RNA-Seq analysis were re-measured by using reverse transcription and qRT-PCR. This was performed on the original RNA from all donors used to perform the RNA-Seq experiment (Table 10) and an independent cohort (Additional file 7A). All genes were found to have comparable results with RNA-Seq data; for instance, genes identified as having an increase in expression in older samples in the RNA-Seq experiment also gave increased expression relative to RPS16 following qRT-PCR. Statistical significance was tested by using Student's *t* test. Two genes whose expressions were not significantly altered in RNA-Seq results—aggrecan (ACAN) and MMP3—were also unaltered when assessed with qRT-PCR. Gene expression analysis using qRT-PCR of an independent cohort found similar results. Validation of differential isoform expression by using qRT-PCR was in general concordance with RNA-Seq (Additional file 7B). In all cases, the level of expression varied between the two platforms.

#### Discussion

Ageing is recognised as a significant risk factor for tendon injury; however, knowledge of changes to the transcriptome of tendon cells has previously been limited to that gained from quantitative PCR [5,48,49] and microarray

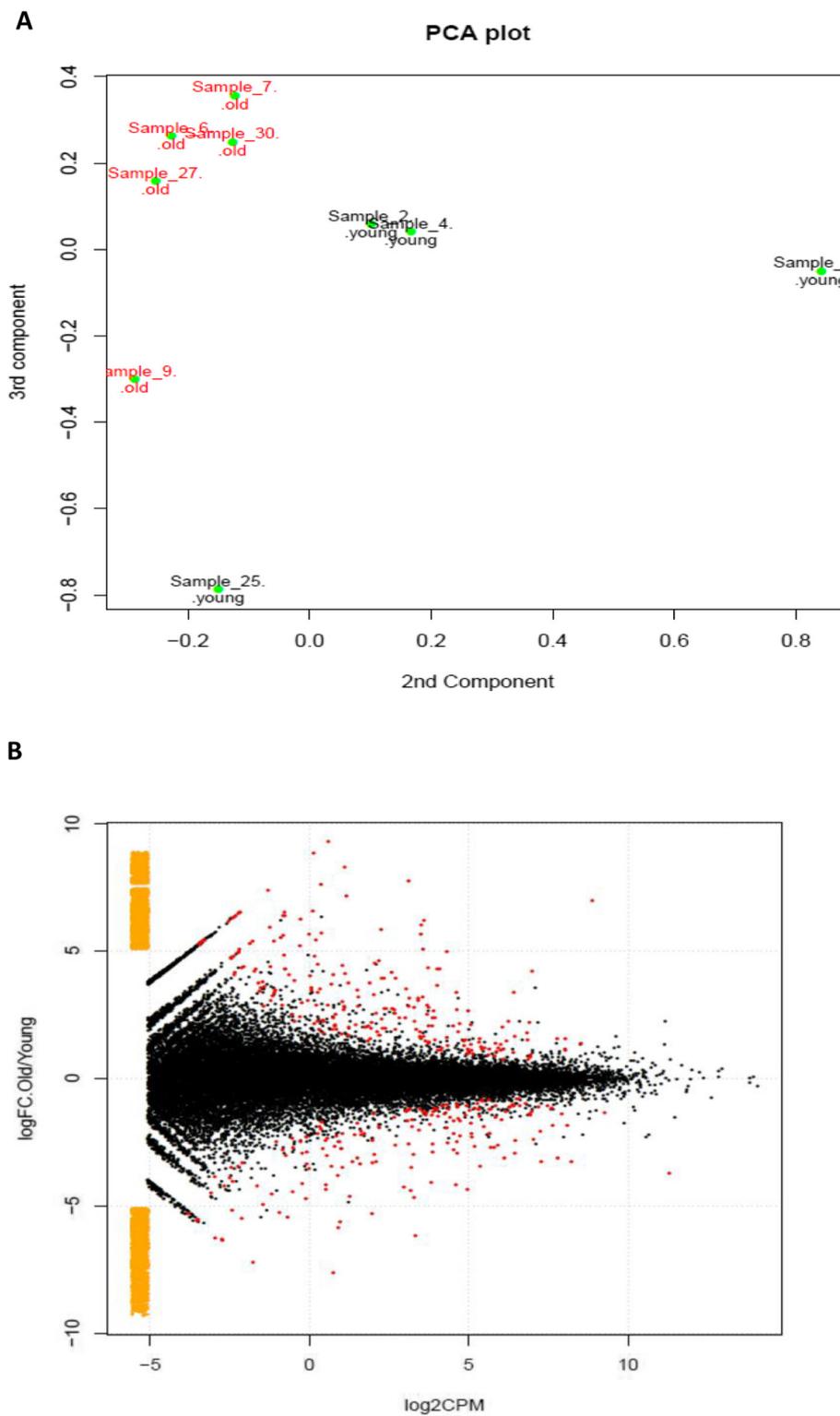
**Table 3 Top 25 genes**

Gene name	Name	Log fold change	FDR-adjusted	Mean FPKM
<i>RN7SK</i>	7SK small nuclear	0.1	1.0	22,323.9
<i>RN7SL2</i>	RNA, 7SL, cytoplasmic 2	-0.5	0.7	12,212.2
<i>RN7SL4P</i>	RNA, 7SL, cytoplasmic, pseudogene 1	-0.6	0.6	5,813.8
<i>RN7SL1</i>	RNA, 7SL, cytoplasmic 2	-1.1	0.5	3,846.0
<i>ANGPTL7</i>	Angiopoietin-like 7	0.4	0.8	3,529.0
<i>RNA28S5</i>	Ribosomal RNA 18S 6	-0.3	0.9	3,374.0
<i>RN7SKP255</i>	7SK small nuclear pseudogene 255	0.1	1.0	2,898.5
<i>RN7SKP203</i>	7SK small nuclear pseudogene 203	-0.1	1.0	2,836.7
<i>S100A6</i>	S100 calcium binding protein A6	0.1	0.9	1,950.3
<i>MALAT1</i>	Metastasis associated lung adenocarcinoma transcript 1	0.0	1.0	1,801.5
<i>DCN</i>	Decorin	0.0	1.0	1,791.5
<i>RN7SL5P</i>	RNA, 7SL, cytoplasmic, pseudogene 2	-1.3	0.4	1,774.0
<i>THBS4</i>	Thrombospondin 4	-0.8	0.6	1,368.2
<i>MT-CO1</i>	Mitochondrial cytochrome c oxidase III	-0.6	0.6	1,230.9
<i>TMSB10</i>	Thymosin beta 10	-0.3	0.7	1,220.1
<i>ASPN</i>	Asporin	-1.6	0.5	1,145.3
<i>MMP3</i>	Matrix metalloproteinase 3	2.2	0.2	1,126.7
<i>RNA18S5</i>	Ribosomal RNA 28S 5	-0.2	0.9	1,125.5
<i>CLU</i>	Clusterin	0.3	0.9	1,106.6
<i>MT-ND3</i>	Mitochondrial NADH dehydrogenase 3	-0.5	0.6	1,036.1
<i>LUM</i>	Lumican	0.0	1.0	978.9
<i>VIM</i>	Vimentin	0.2	0.9	977.1
<i>CILP</i>	Cartilage intermediate layer protein	0.3	0.9	920.8
<i>EEF1A1</i>	Eukaryotic translation elongation factor 1 alpha-like 7	0.0	1.0	891.4
<i>IGFBP6</i>	Insulin-like growth factor-binding protein 6	0.7	0.6	859.0

The table demonstrates the 25 most highly expressed genes in young and old tendon in terms of transcript expression as determined by using fragments per kilobase of exon per million fragments mapped (FPKM). FDR, false discovery rate.

studies on tendinopathic human [50,51] and rat tissue [17,52]. In this study, we report for the first time the use of the RNA-Seq technique to undertake deep transcriptome profiling of young and old macroscopically normal human Achilles tendon. Importantly, validation studies using qRT-PCR demonstrated high correlation between methodologies and demonstrated reproducibility using a different donor set. One of the many advantages of RNA-Seq over microarrays is that it enables *de novo* analysis of transcripts, including novel transcripts. In this study, we were able to identify and quantify protein-coding transcripts, alternatively spliced isoforms, lncRNAs, pseudogenes, and small regulatory RNAs, including small nucleolar RNAs (snoRNA) and an miRNA. The age of the donor accounted for most of the variability in the data, although PCA identified more variability between young donors. We did not have access to detailed medical history and lifestyle factors for the patients in this study, so we are unable to determine whether other factors explain the variability more precisely.

Tendon is characterised by a large amount of ECM interspersed around a relatively sparse population of cells. The main component of the matrix is the fibril-forming type I collagen, which composes about 70% of the dry weight of the matrix [53]. Minor collagen types include other fibril-forming collagens, type III and V; fibril-associated collagens, type XII and XIV; and type VI collagen. As expected, these were the main collagen genes we identified in the transcriptome of the Achilles tendon tissue, albeit at relatively low levels. The non-collagenous component of tendon is rich in small leucine-rich proteoglycans (SLRPs), including decorin, biglycan, fibromodulin, and lumican, and the glycoproteins COMP, lubricin, tenomodulin, and tenascin C [54]. Interestingly, the results of this study show that decorin was by far the most highly expressed ECM gene across the samples in comparison with relatively low levels of collagen transcripts. Lumican was the next most highly expressed ECM protein followed by fibromodulin and COMP. These results are in line with our recent proteomics study in which decorin was the



**Figure 1** Principal component analysis (PCA) and volcano plot of differentially abundant transcripts. **(A)** PCA plot of  $\log_2$  gene expression data showed the greatest variability in RNA-Seq data was due to the age of the donor. The second and third components are also shown as changes in the first and second components are due to differing sample sizes between young and old groups. **(B)** Volcano plot represents  $\log_2$  FC  $\times$   $\log_2$  CPM plot (counts per million mapped reads) CPM plot. Low expression genes ( $\log_2$  CPM of less than  $-5$ ) are coloured orange. Significant differentially expressed genes (DEGs) are coloured in red. A set of DEGs between young and old tendon was identified. With the common dispersion in edgeR [30], 325 DEGs were identified with a  $P$  value of less than 0.05 (red).

**Table 4 Top 10 genes with the highest and lowest log<sub>2</sub> fold change when comparing young and old tendon**

Condition increased differential expression	Gene symbol	Gene name	Location	Function	Log <sub>2</sub> fold change
Old	<i>CYP24A1</i>	Cytochrome P450, family 24, subfamily A,1	Cytoplasm	Enzyme	9.3
	<i>FOLH1B</i>	Folate hydrolase 1B	Cytoplasm	Enzyme	8.8
	<i>POU3F4</i>	POU class 3 homeobox 4	Nucleus	Transcription regulator	8.3
	<i>FOLH1</i>	Folate hydrolase 1	Plasma membrane	Peptidase	7.7
	<i>BRS3</i>	Bombesin-like receptor 3	Plasma membrane	G-protein coupled receptor	7.6
	<i>XIST</i>	X (inactive)-specific transcript (non-protein coding)	Variable	Long non-coding RNA	6.9
	<i>EGF</i>	Epidermal growth factor	Extracellular space	Growth factor	6.5
	<i>BRINP3</i>	Bone morphogenetic protein/retinoic acid inducible neural-specific 3	Secreted/ mitochondria	Osteoblast differentiation factor	6.5
	<i>NXPH2</i>	Neurexophilin 2	Extracellular space	Signalling molecules	6.5
	<i>MYF5</i>	Myogenic factor 5	Nucleus	Transcription regulator	6.4
Young	<i>CYP1A1</i>	Cytochrome P450, family 1, subfamily A, polypeptide 1	Cytoplasm	Enzyme	-7.6
	<i>CLEC3A</i>	C-type lectin domain family 3, member A	Secreted	Cell adhesion	-6.1
	<i>DIRAS1</i>	DIRAS family, GTP-binding RAS-like 1	Plasma membrane	Enzyme	-5.9
	<i>SYT13</i>	Synaptotagmin XIII	Plasma membrane	Transporter	-5.6
	<i>GPR143</i>	G protein-coupled receptor 143	Plasma membrane	G-protein coupled receptor	-5.6
	<i>SLC7A14</i>	Solute carrier family 7 member 14	Plasma membrane	Transporter	-5.4
	<i>HIST 1H3A</i>	HISTONE 1H3A	Nucleus	Nucleosome	-5.4
	<i>PYCR1</i>	Pyrraline-5-carboxylate reductase 1	Cytoplasm	Enzyme	-5.3
	<i>UTS2R</i>	Urotensin 2 receptor	Plasma membrane	G-protein coupled receptor	-5.2
	<i>HBG1</i>	Hemoglobin, gamma A	Cytoplasm	Fe carrier	-4.9

Log<sub>2</sub> fold change and q value (adjusted *P* value) were determined in edgeR. A logarithm to the base 2 of 9 is approximately a linear fold change of 3.2. Shown are the 10 genes with highest and lowest expression in old compared with young tendon samples.

second most abundant ECM protein in a guanidine soluble extract of equine flexor tendon [55]. Degradation of the ECM is accomplished by a family of MMPs along with other proteases, and we identified expression of collagenases, stromelysins, gelatinases, and aggrecanases, although in general the levels of expression were low. An exception to this was MMP3, a stromelysin responsible for proteoglycan degradation, which was one of the most abundant transcripts, again supporting the finding of a higher turnover of non-collagenous proteins.

Ageing results in changes to the tendon ECM composition, although these are poorly defined at present and the impact on tendon mechanical properties is not clear as some studies report increased stiffness with ageing [56,57] whereas others report a decrease [58,59]. A recent study using equine flexor tendon found that, although the mechanical properties of the gross structure and the component fascicles did not change with age,

the inter-fascicular matrix became stiffer. Given this finding, we expected to find differential expression of ECM transcripts, particularly those enriched in the inter-fascicular matrix. The differential gene expression analysis showed no regulation of proteins likely to be enriched in the inter-fascicular matrix or inter-fibrillar proteins [54]. The alpha 1 chain of type I collagen and alpha 1 chain of type III collagen were identified as having reduced expression in the old age group, although this lost statistical significance when measured by qRT-PCR on a larger sample set. For the most part, these data do not support our original hypothesis that tendon ageing results in reduced expression of genes relating to ECM, degradative proteases, cytokines, and growth factors, unlike changes evident in ageing cartilage [24].

Tendon disease, which has a clear association with ageing, has been the focus of several gene expression studies. Generally, findings in these studies are in

**Table 5 Non-coding RNAs significantly overexpressed in old tendon**

Gene accession	Category	Type of transcript	Log <sub>2</sub> fold change	Description/Aliases	Function
RNVU1-6	SnoRNA	Small nuclear	4.9	Variant U1 small nuclear 6	Splicesomal
RP11-578O24.2	Pseudogene	Pseudogene	6.2		Unknown
AP003041.1	Pseudogene	Pseudogene	5.3		Unknown
MKRN7P	Pseudogene	Pseudogene	4.8	Makorin ring finger protein 7	Unknown
RPS4XP22	Pseudogene	Pseudogene	4.1	Ribosomal protein S4X pseudogene 22	Unknown
RP11-346 M5.1	Pseudogene	Pseudogene	2.8		Unknown
RN7SKP234	Pseudogene	Pseudogene	2.4	RNA,7SK small nuclear pseudogene 243	Unknown
CTD-2114 J12.1	Pseudogene	Pseudogene	2.1		Unknown
AL021068.1	Pseudogene	Pseudogene	1.9		Unknown
CTD-2083E4.4	LncRNA	Novel processed	2.3		Unknown
RP3-326 L13.3	LncRNA	Novel lncRNA	7.1		Unknown
RP11-377D9.3	LncRNA	Novel lncRNA	6.3	lnc-KIAA1467-2	Unknown
RP11-71E19.2	LncRNA	Novel lncRNA	5.4	lnc-KLF15-2	Unknown
AC073636.1	LncRNA	Novel lncRNA	5.1	lnc-NFE2L2-3	Unknown
RP11-279 F6.1	LncRNA	Novel lncRNA	4.7	lnc-RPLP1-1	Unknown
AC004510.3	LncRNA	Novel lncRNA	4.1		Unknown
RP11-500B12.1	LncRNA	Novel lncRNA	4.1	lnc-TLR4-1	Unknown
AC007405.6	LncRNA	Novel lncRNA	3.8	lnc-AC007405.7.1-1	Unknown
RP11-399D6.2	LncRNA	Novel lncRNA	3.5	lnc-DMRTA1-2	Unknown
RP11-966I7.1	LncRNA	Novel lncRNA	3.5	lnc-PRKD1-8	Unknown
RP11-815 M8.1	LncRNA	Novel lncRNA	3.1	lnc-DUSP10-3	Unknown
AC003090.1	LncRNA	Novel lncRNA	2.4	lnc-NPVF5	Unknown
RP11-79H23.3	LncRNA	Novel lncRNA	2.2	lnc-FAM164A-1	Unknown
RP11-4 F5.2	LncRNA	Novel lncRNA	2.0	lnc-MCTP2-1	Unknown
RP5-1024G6.8	LncRNA	Novel lncRNA	1.9		Unknown
NKX2-1-AS1	LncRNA	Novel antisense	7.4	NKX2-1 antisense RNA 1	Unknown
RP3-326 L13.2	LncRNA	Novel antisense	6.5		Unknown
RP11-464O2.2	LncRNA	Novel antisense	5.4	lnc-LHPP-1	Unknown
AC091633.3	LncRNA	Novel antisense	5.4	lnc-MUC20-2	Unknown
RP11-711G10.1	LncRNA	Novel antisense	5.3		Unknown
SATB2-AS1	LncRNA	Novel antisense	4.2	SATB2 antisense RNA 1	Unknown
FEZF1-AS1	LncRNA	Novel antisense	4.1	FEZF1 antisense RNA 1	Unknown
SLC26A4-AS1	LncRNA	Novel antisense	3.7	SLC26A4 antisense	Unknown
RP11-160A10.2	LncRNA	Novel antisense	3.5	lnc-CLVS2	Unknown
RP4-803 J11.2	LncRNA	Novel antisense	3.4	lnc-RAB4A-1	Unknown
RP11-402 J6.1	LncRNA	Novel antisense	2.5	lnc-ALPK1-1	Unknown
RP11-710C12.1	LncRNA	Novel antisense	2.0		Unknown
XIST	LncRNA	Known lncRNA	7.0	X inactive specific transcript	X chromosomal inactivation
LINC00261	LncRNA	Known lncRNA	5.9	LINC RNA 261 lnc FOXA2-2	Cancer
TSIX	LncRNA	Known lncRNA	3.9	TSIX transcript, XIST antisense RNA	Antisense during X chromosomal inactivation
LINC00461	LncRNA	Known lncRNA	2.5	lnc MEF2C-2	Unknown
DLX6-AS1	LncRNA	Known antisense	2.7	Embryonic ventral forebrain-1	Gene expression; hippocampus

Genes here have at least a  $\pm 1.4$  log<sub>2</sub> fold change and false discovery rate-adjusted *P* value of less than 0.05 as determined by edgeR. lncRNA, long-coding RNA; SnoRNA, small non-coding RNA.

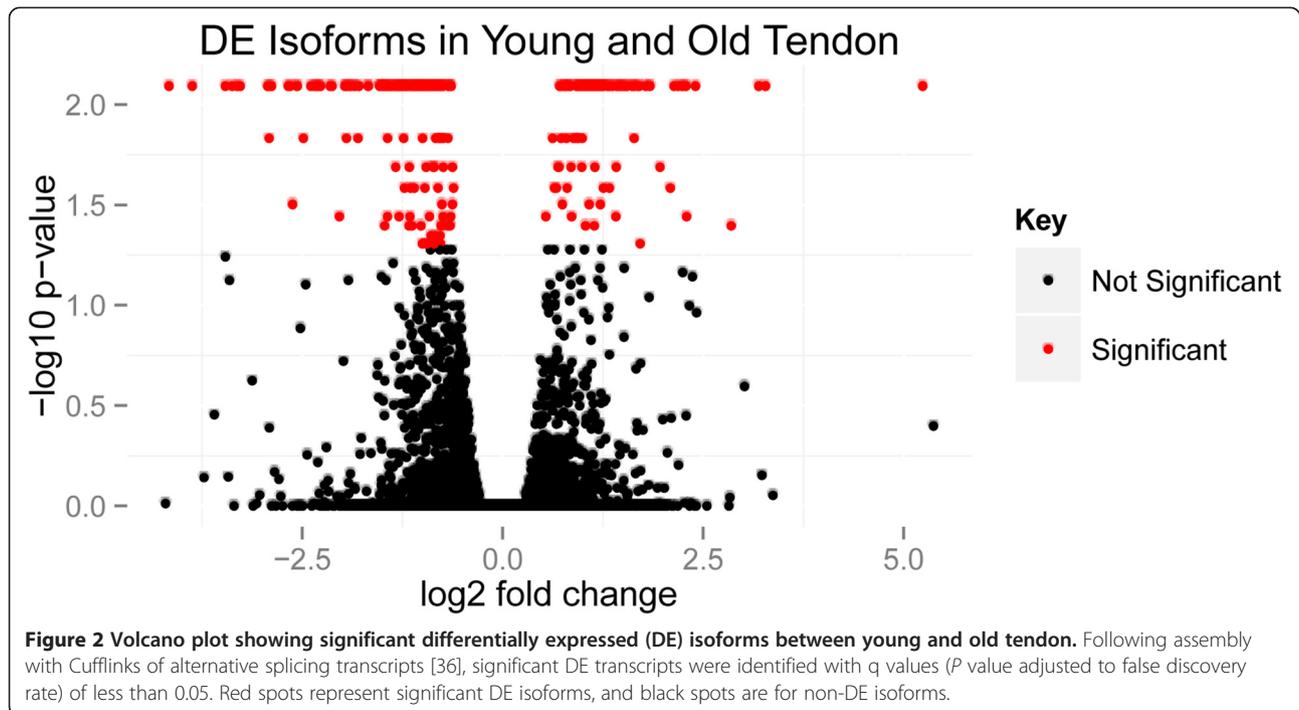
**Table 6 Non-coding RNAs with significantly reduced expression in old tendon**

Gene accession	Category	Type of transcript	Log <sub>2</sub> fold change	Description/Aliases	Function
LINC00957	LncRNA	Novel lncRNA	-1.73	Long intergenic non-coding RNA 957	Unknown
ZNF667-AS1	LncRNA	Novel lncRNA	-1.82	ZNF667 antisense RNA 1 (head to head)	Unknown
RP11-308 N19.1	LncRNA	Novel lncRNA	-2.43		Unknown
AF131215.9	LncRNA	Novel sense intronic	-2.78		Unknown
CTB-113P19.1	LncRNA	Novel antisense	-2.93	lnc-G3BP1-2	Unknown
AF131215.2	LncRNA	Novel sense intronic	-3.24		Unknown
MAFG-AS1	LncRNA	Known antisense	-3.45	MAFG antisense RNA 1 (head to head)	Unknown
AC020571.3	LncRNA	Novel antisense	-3.50	lnc-CCDC150-1	Unknown
AC012613.2	LncRNA	Novel antisense	-3.86	lnc-IL17B-3	Unknown
RP11-923I11.1	LncRNA	Novel antisense	-3.91	lnc-SCN8A-1	Unknown
RP11-270 M14.5	LncRNA	Novel lncRNA	-4.01	lnc-TLL5-1	Unknown
CTD-2540B15.9	LncRNA	Novel lncRNA	-4.50	Novel lincRNA	Unknown
CTD-3049 M7.1	LncRNA	Novel lncRNA	-4.53	lnc-RGMA7	Unknown
RP5-1198O20.4	LncRNA	Novel lncRNA	-5.25	lnc-KLF17-1	Unknown
RP11-300E4.2	LncRNA	Novel antisense	-6.24	lnc-JPH1_5	Unknown
RP11-445 L6.3	LncRNA	Novel lncRNA	-6.31	lnc-TNC-2	Unknown
MIR1245A	MiRNA	Known miRNA	-2.15		Unknown
MXRA5P1	Pseudogene	Pseudogene	-3.52	MXRA5P1	Unknown
RNY3P2	Pseudogene	Pseudogene	-5.30	RNA Ro-associated Y3 pseudogene 2	Unknown
RP11-494 K3.2	Pseudogene	Pseudogene	-6.36		Unknown
CTC-260E6.10	Pseudogene	Pseudogene	-7.23		Unknown
Y_RNA	SnoRNA	Novel misc RNA	-5.53		DNA replication and cell proliferation

Terms are derived from Ensemble [46] and Vega [47]. 'Antisense' overlaps the genomic span of a protein-coding locus on the opposite strand. 'Known' indicates identical to known cDNA or proteins from the same species and has an entry in a model database. 'Novel' indicates identical or homologous to cDNAs from the same species or proteins from all species. 'Processed transcript' does not contain open reading frame and cannot be placed in any other category. 'Pseudogene' indicates homology to protein but from a disrupted coding sequence and an active homologous gene can be found at another locus. 'Sense intronic' has a long non-coding transcript in introns of a coding gene that does not overlap any exons. LncRNA, long non-coding RNA (which can be further classified as LINC RNA, which is a long intergenic non-coding RNA locus of more than 200 base pairs); miRNA, microRNA; SnoRNA, small non-coding RNA.

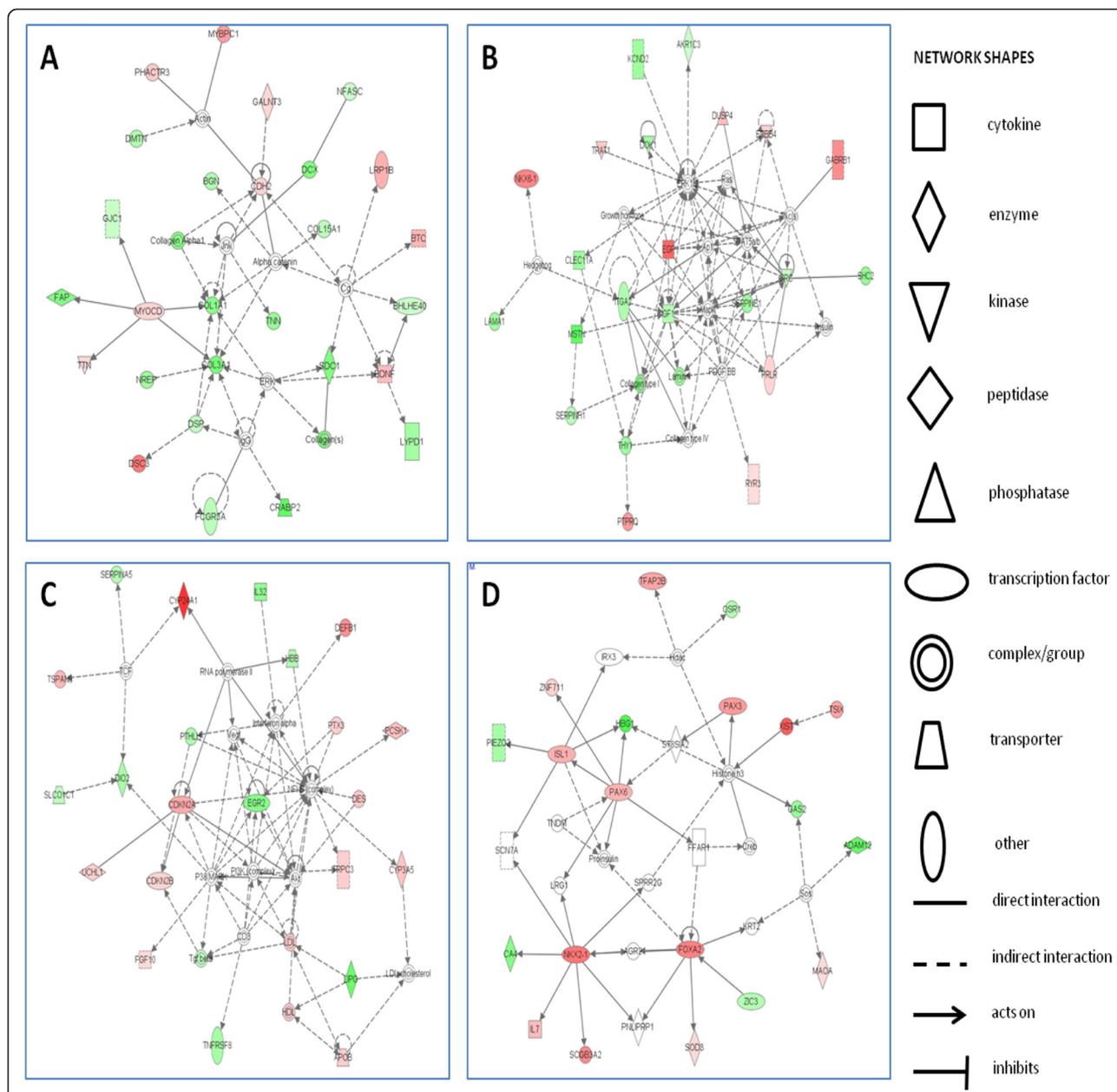
keeping with the hypothesis of increased matrix turnover, with an imbalance favoring catabolism. For example, various studies have demonstrated increased expression of collagen 1 alpha 1 (COL1A1) [5,48,49] and proteins more typical of cartilage COL2A1, aggrecan, and SOX9 [5,52]. Tendinopathic samples show an upregulation of various MMPs, including MMP23 [5,51], a disintegrin and metalloproteinase 12 (ADAM12) [5,50,51], and downregulation of MMP3. The results of our study are in stark contrast to this with very low expression levels for COL2A1, aggrecan, SOX9, most MMPs (except MMP3), and a significant downregulation of ADAM12 in the old group. Therefore, the results suggest that degeneration is not an inevitable consequence of ageing and that ageing and disease-associated degeneration are distinct processes.

In this study, we identified DEG gene sets with ageing related to a dysregulation of cellular function and maintenance, cellular growth and proliferation, cellular cycling, and cellular development. Therefore, these changes suggest that the cellular component of tendon may lose the ability to respond appropriately to mechanical and chemical signals. Other studies have linked cellular senescence, a state of irreversible growth arrest, in a small subset of cells (progenitor cells) in tendon with tendon ageing [60]. A senescence phenotype has been described, although no marker of senescence identified thus far is entirely specific to the senescent state [9]. Most senescent cells express p16(INK)4a, which is not commonly expressed by quiescent or terminally differentiated cells [9]. In this study, p16(INK)4a was expressed at higher levels in the old group, although transcript levels overall

**Table 7** List of top 10 up-/down-regulated isoforms in ageing tendon

Comparison	Gene	Locus	Log <sub>2</sub> (fold change)	q value	
Higher in old	PM20D2	6:89855768-89875284	0.54	0.04	
	ILF3-AS1	19:10762537-10764520	0.62	0.01	
	PERP	6:138411922-138428648	0.65	0.03	
	ODC1	2:10580093-10588630	0.67	0.03	
	PRKAA2	1:57110994-57181008	0.68	0.03	
	PNRC1	6:89790469-89794879	0.69	0.02	
	SYT11	1:155829299-155854990	0.70	0.02	
	F13A1	6:6144317-6321246	0.71	0.01	
	CITED2	6:139693392-139695757	0.72	0.01	
	LDB1	10:103867316-103880210	0.73	0.01	
	Lower in old	HLA-DRB5	6:32485119-32498064	-4.72	0.01
		MYH1	17:10286448-10527201	-4.52	0.01
		CRABP2	1:156669397-156675608	-4.44	0.03
CPXM1		20:2774714-2781283	-4.29	0.01	
COL3A1		2:189839045-189877472	-4.16	0.01	
COL3A1		2:189839045-189877472	-3.87	0.01	
ADAM12		10:127700949-128077024	-3.84	0.01	
CAPN6	X:110488330-110513751	-3.72	0.03		
COL1A1	17:48260649-48278993	-3.46	0.01		
FAP	2:163018279-163101661	-3.37	0.01		

The differentially expressed genes (DEGs) in young versus old tendon were determined by Cuffdiff. The fold change is the ratio of fragments per kilobase of exon per million fragments mapped (FPKM) of those genes in young to FPKM of those genes in old tendon. The significant DEGs (q values of less than 0.05) calculated with a Benjamini-Hochberg correction were ranked on their log<sub>2</sub> fold change, and 10 genes with the highest or lowest fold changes are shown in the table.



were low, which may indicate that a small subpopulation of cells is responsible for the difference. Senescent cells have been shown to contribute to an inflammatory profile, and the term ‘inflamm-aging’ has been coined [13]. Studies have shown upregulation of inflammatory mediators such as cytochrome oxidase 2 (COX2), interleukin 6 (IL-6), and prostaglandin E2 (PGE2) and downregulation of

the lipoxin A4 (LXA4) receptor FPR2 (formyl peptide receptor 2)/ALX in human or equine tendinopathic tissue [14,49]. Inflammatory pathways, however, were not recognised in our GO mapping of DEGs between young and old groups in this study.

An interesting finding in this study was the differential expression of isoforms and those with reduced expression

**Table 8 A number of ingenuity pathway analysis canonical pathways were significantly affected in ageing tendon**

Name of canonical pathway	P value	Ratio
Hepatic fibrosis/Hepatic stellate cell activation	$5.69 \times 10^{-4}$	8/142 (0.056)
Oestrogen biosynthesis	$1.61 \times 10^{-3}$	4/38 (0.105)
Transcriptional regulatory network in embryonic stem cells	$1.95 \times 10^{-3}$	4/40 (0.1)
Glioblastoma signalling	$3.66 \times 10^{-3}$	7/159 (0.044)
Bupropion degradation	$4.89 \times 10^{-3}$	3/26 (0.115)

The significance of the association between the data set and the canonical pathway was measured by using a ratio of the number of molecules from the data set that mapped to the pathway divided by the total number of molecules that map to the canonical pathway. Fisher's exact test was used to calculate *P* values.

in the older tendons mapping to ECM, degradative proteases, cytokines, and growth factors. AS is a significant regulatory mechanism in gene expression as it enables versatility at the post-transcriptional level accounting for proteome complexity and may affect up to 92% of human

**Table 9 Significant gene ontology terms annotated according to DAVID's SP-PIR-Keywords**

Term	Count	Percentage	P value
GO terms identified in downregulated isoforms in old tendon			
Signal	88	45.1	$6.30 \times 10^{-18}$
Secreted	58	29.7	$7.04 \times 10^{-15}$
Extracellular matrix	24	12.3	$3.25 \times 10^{-14}$
Glycoprotein	85	43.6	$2.72 \times 10^{-9}$
Collagen	12	6.2	$1.62 \times 10^{-7}$
Trimer	7	3.6	$8.71 \times 10^{-6}$
Hydroxylation	9	4.6	$2.22 \times 10^{-5}$
Triple helix	7	3.6	$2.00 \times 10^{-5}$
Hydroxylysine	7	3.6	$2.00 \times 10^{-5}$
Disulfide bond	57	29.2	$1.85 \times 10^{-5}$
Cell adhesion	18	9.2	$3.57 \times 10^{-5}$
Hydroxyproline	7	3.6	$4.38 \times 10^{-5}$
Ehlers-Danlos syndrome	5	2.6	$7.34 \times 10^{-5}$
Angiogenesis	6	3.1	0.011506
Metalloprotease	8	4.1	0.016954
Pyroglutamic acid	5	2.6	0.021871
GO terms identified in upregulated isoforms in old tendon			
Secreted	30	24.2	$1.48 \times 10^{-4}$
Plasma	7	5.6	0.003767
Signal	39	31.5	0.006761
Acute phase	4	3.2	0.052997

Gene ontology (GO) terms were revealed in upregulated and downregulated differentially expressed gene isoforms of tendon ageing. *P* value represents Benjamini-Hochberg corrected *P* value. DAVID, Database for Annotation, Visualization and Integrated Discovery.

genes [61]. Differences between isoforms of the same protein extend from a complete loss of function, acquiring a new function to subtle modulations, the latter observed in the majority of cases [62]. Few AS events have been reported in tendon to date. Those that have include versican, in which AS may contribute to changes in ECM structure and function in tendinopathies [48]; lubricin, which is location-dependent [63]; and insulin-like growth factor 1, which is mechanical stress-dependent [64]. The isoforms showing the greatest difference between young and old tendon groups in our study (for example, COL1A1, COL3A1, and ADAM12) are recognised as some of the most important proteins for tendon function and the relevance of these isoforms requires further investigation.

The results of our study have yielded new information relating to tendon cell phenotype and to the ageing process, identifying transcripts that are not generally recognised as being important in tendon. For example, the gene most highly expressed, disregarding ribosomal proteins, was angiopoetin-like 7 (ANGPTL7). This protein has previously been identified as highly expressed in microarray analysis of human tendinopathic tissue from various tendons [50]. Angiopoetins are involved in angiogenesis [65], inflammation, and glucose [66] and lipid [67] metabolism. In the cornea, ANGPTL7 may function as negative regulator of angiogenesis, contributing to the avascular properties of the tissue [68], whereas in the human ocular trabecular meshwork cells, it has a role in the organization of the ECM [69]. Thus, we suggest that ANGPTL7 may have a role in maintaining the relatively avascular nature of tendon tissue and in the organisation of the ECM. This represents an area for further investigation.

One of the limitations of this study is that the samples were taken from patients with malignant disease. We consider that it is very unlikely that this has influenced the results as samples are taken only when the tumour is at a site distant to the tendon and the tendon is macroscopically normal; however, we cannot rule out the possibility that some of the genes showing high expression, such as metastasis associated lung adenocarcinoma transcript 1 (MALAT1), ANGPTL7, and S100A6, are related to the disease state.

Another point of interest was the expression of genes associated with muscle cells. For example, we observed a reduced DEG of myosin heavy chain 1 and an increase DEG of myogenic factor 5 (MYF5) and MYOCD. Our previous studies in ageing cartilage also identified DEG of muscle-related genes: myosin heavy chain 2, myosin 3A, and myosin 1B, which were all reduced in cartilage ageing [24]. Samples of Achilles tendon were taken at a region far removed from muscle insertion, and the identification of muscle genes in cartilage and tendon is unlikely to be due to inadvertent inclusion of muscle tissue.

**Table 10 Real-time polymerase chain reaction analysis of 14 selected genes reveals good correlation with RNA-Seq results**

Gene name	RNA-Seq results			RT-PCR results		
				Age		P value
	Differential expression	Significant log <sub>2</sub> fold change	q value	Young	Old	
<i>EGF</i>	Higher in old	6.50	$2.06 \times 10^{-8}$	$0.003 \pm 0.002$	$0.01 \pm 0.007$	0.04
<i>POU3F4</i>		8.20	$7.79 \times 10^{-16}$	$0.021 \pm 0.015$	$0.245 \pm 0.147$	0.02
<i>MYF5</i>		6.30	0.003	$0.061 \pm 0.057$	$0.25 \pm 0.07$	0.01
<i>TGFB3</i>	Lower in old	-1.60	0.035	$10.717 \pm 6.022$	$0.897 \pm 0.34$	0.05
<i>MMP16</i>		-2.30	0.022	$0.107 \pm 0.087$	$0.014 \pm 0.02$	0.07
<i>COL3A1</i>		-3.70	0.008	$13.696 \pm 17.535$	$1.05 \pm 0.594$	0.27
<i>MYH1</i>		-4.60	0.003	$0.288 \pm 0.155$	$0.14 \pm 0.101$	0.23
<i>IGF1</i>		-2.10	0.004	$2.094 \pm 1.234$	$0.483 \pm 0.278$	0.04
<i>COL1A1</i>		-3.30	0.009	$47.968 \pm 32.676$	$8.515 \pm 9.154$	0.42
<i>MMP3</i>	No change	Not significant	1.000	$51.181 \pm 362.11$	$99.323 \pm 66.10$	0.15
<i>ACAN</i>			1.000	$1.439 \pm 1.43$	$0.848 \pm 0.99$	0.66
<i>XIST</i>	Higher in old	6.90	0.000	$0.827 \pm 0.712$	$7.41 \pm 3.465$	0.03
<i>LINC00957</i>	Lower in old	-1.70	0.036	$1.044 \pm 0.584$	$0.363 \pm 0.174$	0.08
<i>RP11.308 N19.1</i>		-2.430	0.040	$0.02 \pm 0.065$	$0.005 \pm 0.001$	0.03

Values for real-time polymerase chain reaction (RT-PCR) are the mean  $\pm$  standard deviation of relative expression levels normalised to expression of RSP16 (to two decimal places). Statistical significance was tested by using Student's *t* test. RT-PCR results are expressed as  $2^{-\Delta\Delta CT}$ . ACAN, aggrecan; COL1A1, collagen type 1 alpha 1; COL3A1, collagen type 3 alpha 1; EGF, epidermal growth factor; IGF1, insulin growth factor 1; LINC00957 long intergenic non-protein coding RNA 957; MMP3, matrix metalloproteinase; MMP16, matrix metalloproteinase 16; MYF5, myogenic factor 5; MYH1, myosin heavy chain 1; POU3F4, POU class 3 homeobox 4; RP11.308 N19.1, lnc-ZNF462-2; TGFB3, transforming growth factor  $\beta$ .

Interestingly, there was an increase in the expression of a large set of transcription factors in old compared with young tendon. In *Caenorhabditis elegans* [70] and a number of tissues, including heart [71] and brain [72], transcription factors have been implicated in ageing. Interestingly, deacetylates Nk2 homeobox 1 (NKX2-1), a transcription factor showing upregulation in old tendon in our study, is involved in neuronal activation in dorso-medial and lateral hypothalamic nuclei, a function thought to contribute to a more 'youthful' physiology during ageing [73]. Conversely, an isoform of scleraxis (SCX), a critical transcription factor in tendon development [74], was reduced in old tendon. In addition, the reduced expression in old tendon of EGR2 and AS EGR1, both of which are required for tendon differentiation [75], may affect tendon repair [76].

We identified eight pseudogenes showing upregulation in the old group of tendons. Pseudogenes have similar sequences to their counterpart coding genes, but owing to mutation/deletion or insertion of nucleotides they cannot be transcribed. It is hypothesised that pseudogenes act as post-transcriptional regulators of the corresponding parental gene [74]. In other studies, pseudogenes have been identified as increasing with age, such as pseudogene cyclin D2 in the ovary [76], and recent work has indicated that they may have a role in inflammation [77]. This provides an exciting new frontier to explore in ageing research, and further work is required to determine whether

any of the pseudogenes identified in this study have functional significance.

Our study is the first to profile lncRNAs in tendon. lncRNAs are a large and functionally heterogeneous class of RNAs with a length of more than 200 nucleotides. They have been shown to regulate mRNA transcription, splicing, stability, translation, and epigenetic modification, providing a complex spectrum of gene regulatory functions [78], and a number of studies have identified roles for lncRNAs in ageing [22,79]. In this study, lncRNAs were shown to be DEGs in ageing tendon, and 34 showed upregulation in old tendon. In musculoskeletal disease, relatively little work interrogates the role of lncRNAs in tissue physiology and disease except for a few studies in cartilage/OA [80-82] and muscle (reviewed [83]) and an osteosarcoma study [84]. The lncRNA transcriptome signatures in ageing tendon provide an interesting set of genes for further studies to determine their role in tendon ageing and disease.

## Conclusions

Our study is the first to interrogate tendon by using RNA-Seq. We demonstrate dynamic alterations in RNA with age, at numerous genomic levels, which indicate changes in the regulation of transcriptional networks. Further extensive follow-up analysis of modulator genes, splice variants, and non-coding RNAs found in this study may be useful in understanding tendon ageing.

## Additional files

**Additional file 1: Transcript expression in all samples.** A complete list of significantly expressed genes and DAVID (Database for Annotation, Visualization and Integrated Discovery) analysis of the expression patterns is shown in a complete table of all transcript expression as an estimate of overall abundance of genes in ageing tendon. Reads were used to estimate transcript expression of all nine samples by using fragments per kilobase of exon per million fragments mapped (FPKM) in order to identify the most abundant genes in tendon. Based on the gene annotation file, a gene contains certain features. The total feature length of a gene is known (GFL). The count numbers obtained from read alignment for a gene depend on the real expression level of the gene and its GLF value. Therefore, counts cannot be used for comparing the expression level among genes. The sequence depth also influences the count values for given gene expression level. FPKM corrects for these two effects. FC, fold change; FDR, false discovery rate; PV, P-value; GFL, Gene feature length SD, standard deviation.

**Additional file 2: The entire set of significantly differentially expressed gene transcripts.** Significant differentially expressed genes have a  $\log_2$  fold change (FC)  $\pm 1.4$  and false discovery rate (FDR)-adjusted  $P < 0.05$ . Highest in old (gene names marked in red) are genes more highly expressed in old tendon compared with young tendon, and lowest in old (marked in blue) are genes less expressed in old tendon compared with young tendon.

**Additional file 3: The entire list of significantly differentially expressed gene isoforms.** A complete table of significant differentially expressed isoforms with a  $\log_2$  fold change  $\pm 1.4$  and false discovery rate-adjusted  $P < 0.05$  is shown. FPKM, fragments per kilobase of exon per million fragments mapped.

**Additional file 4: All identified networks and their respective molecules.** Network-eligible molecules were overlaid onto molecular networks. All identified networks and their respective molecules are tabulated.

**Additional file 5: Associated molecules of the top canonical pathways.** Molecules in the tendon RNA-Seq data set associated with the top five canonical pathways are shown.

**Additional file 6: Skeletal and muscular disorders were the top diseases associated with the gene set.** Disease annotation and associated  $P$  value, activation  $z$ -score, and molecules associated with skeletal and muscular disorders are shown.

**Additional file 7: Quantitative real-time polymerase chain reaction (qRT-PCR) analysis of 12 selected genes reveals good correlation with RNA-Seq results in an independent cohort.** Values for qRT-PCR are the mean  $\pm$  standard deviation of relative expression levels normalised to expression of RPS16. Statistical significance was tested by using Student's  $t$  test. ACAN, aggrecan; COL1A1, collagen type I alpha 1; COL3A1, collagen type III, alpha 1; EGF, epidermal growth factor; FC, fold change; IGF1, insulin growth factor 1; LINC00957, long intergenic non-protein coding RNA 957; MMP-3, matrix metalloproteinase 3; MMP-16, matrix metalloproteinase 16; MYF5, myogenic factor 5; MYH1, myosin heavy chain 1; POU3F4, POU class 3 homeobox 4; TGF $\beta$ 3, transforming growth factor  $\beta$ 3; XIST, X inactive specific transcript.

## Abbreviations

ADAM12: a disintegrin and metallopeptidase domain 12; ALX: ALX homeobox; ANGPTL7: angiopoietin-like 7; AS: alternative splicing; bp: base pair; COL: collagen; DAVID: Database for Annotation, Visualization and Integrated Discovery; DEG: differentially expressed gene; ECM: extracellular matrix; EGR2: early growth response protein; FDR: false discovery rate; FPKM: fragments per kilobase of exon per million fragments mapped; GO: gene ontology; IL: interleukin; IPA: Ingenuity Pathway Analysis; lncRNA: long non-coding RNA; miRNA: microRNA; MMP: matrix metalloproteinase; MYOCD: myocardin; NCBI GEO: National Center for Biotechnology Information Gene Expression Omnibus; NKX2: thyroid transcription factor 1; PAX: paired box transcription factor; PCA: principal component analysis; PCR: polymerase chain reaction; qRT-PCR: quantitative real-time polymerase chain reaction; rRNA: ribosomal RNA; RSP: ribosomal protein; snoRNA: small non-coding RNA.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

MJP conceived the study, carried out most of the laboratory work and bioinformatics analysis, and drafted the manuscript. YF carried out post-processing of data, including resolution of indexes, and helped to draft the manuscript. KC undertook AS analysis and helped with drafting the manuscript. TKJW undertook data acquisition and processing and helped draft the manuscript. PDC participated in the design and coordination of the study and helped to draft the manuscript. HLB participated in the design and coordination of the study, harvested the tissue and carried out the macroscopic examination, and helped to draft the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

MJP was supported by a Wellcome Trust Veterinary Integrated Research Fellowship (WT088557MA). The study was funded by a Biotechnology and Biological Sciences Research Council (BBSRC) grant (BB/K007785/1) and supported by the Medical Research Council (MRC) and Arthritis Research UK as part of the MRC – Arthritis Research UK Centre for Integrated research into Musculoskeletal Ageing (CIMA).

## Author details

<sup>1</sup>Comparative Musculoskeletal Biology, Institute of Ageing and Chronic Disease, University of Liverpool, Leahurst, Chester High Road, Neston, Wirral CH64 7TE, UK. <sup>2</sup>Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK. <sup>3</sup>Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. <sup>4</sup>School of Life Sciences and Chemical Technology, Ngee Ann Polytechnic, 535 Clementi Road, Singapore 599489, Singapore. <sup>5</sup>Institute of Orthopaedics and Musculoskeletal Science, University College London, Stanmore Campus, Royal National Orthopaedic Hospital, Brockley Hill, Stanmore HA7 4LP, UK.

Received: 24 September 2014 Accepted: 26 January 2015

Published online: 18 February 2015

## References

1. Beard JR, Biggs S, Bloom BD, Fried LP, Hogan L, Kalache A, et al. Global Population Ageing: Peril or Promise. Geneva: World Economic Forum; 2011.
2. Clayton RA, Court-Brown CM. The epidemiology of musculoskeletal tendinosis and ligamentous injuries. *Injury*. 2008;39:1338–44.
3. Hess GW. Achilles tendon rupture: a review of etiology, population, anatomy, risk factors, and injury prevention. *Foot Ankle Spec*. 2010;3:29–32.
4. Movin T, Gad A, Reinholt FP, Rolf C. Tendon pathology in long-standing achillobodynia. Biopsy findings in 40 patients. *Acta Orthop Scand*. 1997;68:170–5.
5. Corps AN, Robinson AH, Harrall RL, Avery NC, Curry VA, Hazleman BL, et al. Changes in matrix protein biochemistry and the expression of mRNA encoding matrix proteins and metalloproteinases in posterior tibialis tendinopathy. *Ann Rheum Dis*. 2012;71:746–52.
6. Riley GP, Goddard MJ, Hazleman BL. Histopathological assessment and pathological significance of matrix degeneration in supraspinatus tendons. *Rheumatology (Oxford)*. 2001;40:229–30.
7. Birch HL, Bailey AJ, Goodship AE. Macroscopic 'degeneration' of equine superficial digital flexor tendon is accompanied by a change in extracellular matrix composition. *Equine Vet J*. 1998;30:534–9.
8. Tavernarakis N. Ageing and the regulation of protein synthesis: a balancing act? *Trends Cell Biol*. 2008;18:228–35.
9. Rodier F, Campisi J. Four faces of cellular senescence. *J Cell Biol*. 2011;192:547–56.
10. Riley G. Tendinopathy—from basic science to treatment. *Nat Clin Pract Rheumatol*. 2008;4:82–9.
11. Thorpe CT, Streeter I, Pinchbeck GL, Goodship AE, Clegg PD, Birch HL. Aspartic acid racemization and collagen degradation markers reveal an accumulation of damage in tendon collagen that is enhanced with aging. *J Biol Chem*. 2010;285:15674–81.
12. Dudhia J, Scott CM, Draper ER, Heinegard D, Pitsillides AA, Smith RK. Aging enhances a mechanically-induced reduction in tendon strength by an active process involving matrix metalloproteinase activity. *Aging Cell*. 2007;6:547–56.

13. Franceschi C, Bonafe M, Valensin S, Olivieri F, De Luca M, Ottaviani E, et al. Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann N Y Acad Sci*. 2000;908:244–54.
14. Dakin SG, Dudhia J, Werling NJ, Werling D, Abayasekara DR, Smith RK. Inflamm-aging and arachadonic acid metabolite differences with stage of tendon disease. *PLoS One*. 2012;7:e48978.
15. Kirkwood TB. A systematic look at an old problem. *Nature*. 2008;451:644–7.
16. Molloy TJ, Wang Y, Horner A, Skerry TM, Murrell GA. Microarray analysis of healing rat Achilles tendon: evidence for glutamate signaling mechanisms and embryonic gene expression in healing tendon tissue. *J Orthop Res*. 2006;24:842–55.
17. Molloy TJ, Kemp MW, Wang Y, Murrell GA. Microarray analysis of the tendinopathic rat supraspinatus tendon: glutamate signaling and its potential role in tendon degeneration. *J Appl Physiol*. 2006;101:1702–9.
18. Maeda E, Fleischmann C, Mein CA, Shelton JC, Bader DL, Lee DA. Functional analysis of tenocytes gene expression in tendon fascicles subjected to cyclic tensile strain. *Connect Tissue Res*. 2010;51:434–44.
19. Eliasson P, Andersson T, Aspenberg P. Influence of a single loading episode on gene expression in healing rat Achilles tendons. *J Appl Physiol* (1985). 2012;112:279–88.
20. Eliasson P, Andersson T, Hammerman M, Aspenberg P. Primary gene response to mechanical loading in healing rat Achilles tendons. *J Appl Physiol* (1985). 2013;114:1519–26.
21. Troy A, Sharpless NE. Genetic 'Inc'-age of noncoding RNAs to human disease. *J Clin Invest*. 2012;122:3837–40.
22. Abdelmohsen K, Panda A, Kang MJ, Xu J, Selimyan R, Yoon JH, et al. Senescence-associated lncRNAs: senescence-associated long noncoding RNAs. *Aging Cell*. 2013;12:890–900.
23. Kato M, Slack FJ. Ageing and the small, non-coding RNA world. *Ageing Res Rev*. 2013;12:429–35.
24. Peffer MJ, Liu X, Clegg PD. Transcriptomic signatures in cartilage ageing. *Arthritis Res Ther*. 2013;15:R98.
25. Peffer MJ, Beynon RJ, Clegg PD. Absolute quantification of selected proteins in the human osteoarthritic secretome. *Int J Mol Sci*. 2013;14:20658–81.
26. Martin M. Cutadapt removes adapter sequences from high through put sequencing reads. *EMB Net J*. 2011;17:10–2.
27. Human reference genome sequence. [ftp://ftp.ensembl.org/pub/release-73/fasta/homo\\_sapiens/dna/](ftp://ftp.ensembl.org/pub/release-73/fasta/homo_sapiens/dna/). 2014.
28. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
29. Human genome annotation. [ftp://ftp.ensembl.org/pub/release-73/gtf/homo\\_sapiens/](ftp://ftp.ensembl.org/pub/release-73/gtf/homo_sapiens/). 2014.
30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met*. 1995;57:289–300.
32. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
34. Picard, a set of Java command line tools for manipulating high-throughput sequencing. <http://broadinstitute.github.io/picard/>. 2014.
35. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
36. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–78.
37. Wickham H, editor. ggplot2: elegant graphics for data analysis. New York: Springer; 2009.
38. Ingenuity systems. Ingenuity pathway analysis. <http://www.ingenuity.com>. 2014.
39. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44–57.
40. Martin I, Jakob M, Schafer D, Dick W, Spagnoli G, Heberer M. Quantitative analysis of gene expression in human articular cartilage from normal and osteoarthritic joints. *Osteoarthritis Cartilage*. 2001;9:112–8.
41. Noh EM, Kim JS, Hur H, Park BH, Song EK, Han MK, et al. Cordycepin inhibits IL-1beta-induced MMP-1 and MMP-3 expression in rheumatoid arthritis synovial fibroblasts. *Rheumatology*. 2009;48:45–8.
42. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002;3:RESEARCH0034.
43. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods*. 2001;25:402–8.
44. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998;8:175–85.
45. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. 2008;18:1509–17.
46. A software system which produces and maintains automatic annotation on selected eukaryotic genomes. <http://www.ensembl.org>. 2014.
47. A repository for high-quality gene models produced by the manual annotation of vertebrate genomes. <http://vega.sanger.ac.uk/>.
48. Corps AN, Robinson AH, Movin T, Costa ML, Ireland DC, Hazleman BL, et al. Versican splice variant messenger RNA expression in normal human Achilles tendon and tendinopathies. *Rheumatology*. 2004;43:969–72.
49. Legerlotz K, Jones ER, Screen HR, Riley GP. Increased expression of IL-6 family members in tendon pathology. *Rheumatology (Oxford)*. 2012;51:1161–5.
50. Jelinsky SA, Rodeo SA, Li J, Gulotta LV, Archambault JM, Seeherman HJ. Regulation of gene expression in human tendinopathy. *BMC Musculoskelet Disord*. 2011;12:86.
51. Jones GC, Corps AN, Pennington CJ, Clark IM, Edwards DR, Bradley MM, et al. Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human achilles tendon. *Arthritis Rheum*. 2006;54:832–42.
52. Archambault JM, Jelinsky SA, Lake SP, Hill AA, Glaser DL, Soslowsky LJ. Rat supraspinatus tendon expresses cartilage markers with overuse. *J Orthop Res*. 2007;25:617–24.
53. Birch HL, Bailey JV, Bailey AJ, Goodship AE. Age-related changes to the molecular and cellular components of equine flexor tendons. *Equine Vet J*. 1999;31:391–6.
54. Thorpe CT, Birch HL, Clegg PD, Screen HR. The role of the non-collagenous matrix in tendon function. *Int J Exp Pathol*. 2013;94:248–59.
55. Peffer MJ, Thorpe CT, Collins JA, Eong R, Wei TK, Screen HR, et al. Proteomic analysis reveals Age-related changes in tendon matrix composition, with Age- and injury-specific matrix fragmentation. *J Biol Chem*. 2014;289:25867–78.
56. Faria A, Gabriel R, Abrantes J, Bras R, Moreira H. Biomechanical properties of the triceps surae muscle-tendon unit in young and postmenopausal women. *Clin Biomech (Bristol, Avon)*. 2011;26:523–28.
57. Thermann H, Frerichs O, Biewener A, Krettek C, Schandelmaier P. Biomechanical studies of human Achilles tendon rupture. *Unfallchirurg*. 1995;98:570–5.
58. Kubo K, Kanehisa H, Miyatani M, Tachi M, Fukunaga T. Effect of low-load resistance training on the tendon properties in middle-aged and elderly women. *Acta Physiol Scand*. 2003;178:25–32.
59. Magnusson SP, Narici MV, Maganaris CN, Kjaer M. Human tendon behaviour and adaptation, in vivo. *J Physiol*. 2008;586:71–81.
60. Kohler J, Popov C, Klotz B, Alberton P, Prall WC, Haasters F, et al. Uncovering the cellular and molecular changes in tendon stem/progenitor cells attributed to tendon aging and degeneration. *Aging Cell*. 2013;12:988–99.
61. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;456:470–6.
62. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, et al. Function of alternative splicing. *Gene*. 2005;344:1–20.
63. Sun Y, Berger EJ, Zhao C, Jay GD, An KN, Amadio PC. Expression and mapping of lubricin in canine flexor tendon. *J Orthop Res*. 2006;24:1861–8.
64. Heinemeier KM, Olesen JL, Schjerling P, Haddad F, Langberg H, Baldwin KM, et al. Short-term strength training and the expression of myostatin and IGF-I isoforms in rat muscle and tendon: differential effects of specific contraction types. *J Appl Physiol* (1985). 2007;102:573–81.
65. Hato T, Tabata M, Oike Y. The role of angiotensin-like proteins in angiogenesis and metabolism. *Trends Cardiovasc Med*. 2008;18:6–14.
66. Xu A, Lam MC, Chan KW, Wang Y, Zhang J, Hoo RL, et al. Angiotensin-like protein 4 decreases blood glucose and improves glucose tolerance but

- induces hyperlipidemia and hepatic steatosis in mice. *Proc Natl Acad Sci U S A*. 2005;102:6086–91.
67. Ono M, Shimizugawa T, Shimamura M, Yoshida K, Noji-Sakikawa C, Ando Y, et al. Protein region important for regulation of lipid metabolism in angiotensin-like 3 (ANGPTL3): ANGPTL3 is cleaved and activated in vivo. *J Biol Chem*. 2003;278:41804–9.
  68. Peek R, Kammerer RA, Frank S, Otte-Holler I, Westphal JR. The angiotensin-like factor cornea-derived transcript 6 is a putative morphogen for human cornea. *J Biol Chem*. 2002;277:686–93.
  69. Comes N, Buie LK, Borrás T. Evidence for a role of angiotensin-like 7 (ANGPTL7) in extracellular matrix formation of the human trabecular meshwork: implications for glaucoma. *Genes Cells*. 2011;16:243–59.
  70. Pincus Z, Slack FJ. Transcriptional (dys)regulation and aging in *Caenorhabditis elegans*. *Genome Biol*. 2008;9:233.
  71. Bodyak N, Kang PM, Hirumura M, Suljoadikusumo I, Horikoshi N, Khrapko K, et al. Gene expression profiling of the aging mouse cardiac myocytes. *Nucleic Acids Res*. 2002;30:3788–94.
  72. Lu T, Aron L, Zullo J, Pan Y, Kim H, Chen Y, et al. REST and stress resistance in ageing and Alzheimer's disease. *Nature*. 2014;507:448–54.
  73. Satoh A, Imai S. Hypothalamic Sirt1 in aging. *Aging (Albany NY)*. 2014;6:1–2.
  74. Brent AE, Schweitzer R, Tabin CJ. A somitic compartment of tendon progenitors. *Cell*. 2003;113:235–48.
  75. Guerquin MJ, Charvet B, Nourissat G, Havis E, Ronsin O, Bonnin MA, et al. Transcription factor EGR1 directs tendon differentiation and promotes tendon repair. *J Clin Invest*. 2013;123:3564–76.
  76. Gross G, Hoffmann A. Therapeutic strategies for tendon healing based on novel biomaterials, factors and cells. *Pathobiology*. 2013;80:203–10.
  77. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife*. 2013;2:e00762.
  78. Baker M. Long noncoding RNAs: the search for function. *Nat Methods*. 2011;8:379–83.
  79. Lukiw WJ, Handley P, Wong L, Crapper McLachlan DR. BC200 RNA in normal human neocortex, non-Alzheimer dementia (NAD), and senile dementia of the Alzheimer type (AD). *Neurochem Res*. 1992;17:591–7.
  80. 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 Rheum*. 2013;66:969–78.
  81. Steck E, Boeuf S, Gabler J, Werth N, Schnatzer P, Diederichs S, et al. Regulation of H19 and its encoded microRNA-675 in osteoarthritis and under anabolic and catabolic in vitro conditions. *J Mol Med (Berl)*. 2012;90:1185–95.
  82. Barter MJ, Hyatt S, Xu Y, Young DA. Long non-coding RNAs in osteoarthritis and chondrogenesis. *Osteoarthritis Cartilage*. 2014;22(Supplement):S222.
  83. Neguembor MV, Jothi M, Gabellini D. Long noncoding RNAs, emerging players in muscle differentiation and disease. *Skelet Muscle*. 2014;4:8.
  84. Li JP, Liu LH, Li J, Chen Y, Jiang XW, Ouyang YR, et al. Microarray expression profile of long noncoding RNAs in human osteosarcoma. *Biochem Biophys Res Commun*. 2013;433:200–6.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

