

AGE-RELATED CHANGES IN MESENCHYMAL STEM CELLS IDENTIFIED USING A MULTI-OMICS APPROACH

M.J. Peffers^{1*}, J. Collins¹, Y. Fang², K. Goljanek-Whysall¹, M.Rushton³, J. Loughlin³, C. Proctor³ and P.D. Clegg¹

¹Institute of Ageing and Chronic Disease, University of Liverpool, Leahurst, Chester High Road, Neston, Wirral, CH64 7TE, UK

²Centre for Genomic Research, Institute of Integrative Biology, Biosciences Building, Crown Street, University of Liverpool, Liverpool L69 7ZB, UK

³Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

⁴Newcastle University Institute for Ageing, Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, NE4 5PL, UK

Abstract

Mesenchymal stem cells (MSC) are capable of multipotent differentiation into connective tissues and as such are an attractive source for autologous cell-based treatments for many clinical diseases and injuries. Ageing is associated with various altered cellular phenotypes coupled with a variety of transcriptional, epigenetic and translational changes. Furthermore, the regeneration potential of MSCs is reduced with increasing age and is correlated with changes in cellular functions. This study used a systems biology approach to investigate the transcriptomic (RNAseq), epigenetic (miRNAseq and DNA methylation) and protein alterations in ageing MSCs in order to understand the age-related functional and biological variations, which may affect their applications to regenerative medicine. We identified no change in expression of the cellular senescence markers. Alterations were evident at both the transcriptional and post-transcriptional level in a number of transcription factors. There was enrichment in genes involved in developmental disorders at mRNA and differential methylated loci (DML) level. Alterations in energy metabolism were apparent at the DML and protein level. The microRNA miR-199b-5p, whose expression was reduced in old MSCs, had predicted gene targets involved in energy metabolism and cell survival. Additionally, enrichment of DML and proteins in cell survival was evident. Enrichment in metabolic processes was revealed at the protein level and in genes identified as undergoing alternate splicing. Overall, an altered phenotype in MSC ageing at a number of levels implicated roles for inflammation and mitochondrial ageing. Identified changes represent novel insights into the ageing process, with implications for stem cell therapies in older patients.

Keywords: Mesenchymal stem cells, RNAseq, epigenetics, proteomics, ageing.

*Address for correspondence:

Mandy Peffers BSc (Hons) MPHIL PhD BVet Med MRCVS,

Wellcome Trust Clinical Intermediate Research Fellow,

Department of Musculoskeletal Biology,

Institute of Ageing and Chronic Disease,

The University of Liverpool, Leahurst, Neston,

CH64 7TE, UK,

Telephone number: +44 (0)78 726 21 02,

Email: peffs@liv.ac.uk

Introduction

Mesenchymal stem cells (MSC) are post-embryonic cells, capable of self-renewal and multipotent differentiation into connective tissues including bone, tendon and cartilage. They are an attractive source for autologous cell-based treatments for a range of clinical diseases and injuries. Furthermore, these cells may be considered useful tools for understanding the physiopathological understanding of specific diseases, and to develop disease-specific biomarkers (Valenti *et al.*, 2015).

MSCs possess an age-related loss of cellular functions, including differentiation potential and capacity for proliferation (Huang *et al.*, 2010; Kretlow *et al.*, 2008; Yu *et al.*, 2011). This has implications for stem cell therapies in older patients. Moreover, the reduction in differentiation potential could contribute to ageing and age-related disease (Zaim *et al.*, 2012). It has become increasingly apparent that the therapeutic efficacy mediated by MSCs is through the production of soluble factors (cytokines and growth factors) that control diverse disease-associated processes, including activation of tissue resident stem cells, apoptosis and inhibition of inflammation (Kolf *et al.*, 2007).

Ageing is a complex process with effects at both the cellular and organism level. Ageing has been defined as “the sum of the primary restrictions in regenerative mechanisms of multicellular organisms” (Sethe *et al.*, 2006). It presents huge challenges for society, but a key issue is that whilst life span increases, quality of life faced by individuals in old age is often poor. As MSCs provide regenerative capacity it is rational that functional deficiencies may contribute to the ageing process and age-related diseases such as osteoarthritis, osteoporosis and tendinopathy.

Natural ageing is coupled with a progressive reduction in the regulation of cellular, tissue and organ interactions, resulting in senescence (the irreversibly arrest of cellular division). It can affect the decline in regenerative potential of tissue and cellular functions in numerous organs. Animal studies and clinical trials have demonstrated that the regeneration potential of bone (Stenderup *et al.*, 2004) and other tissues declines with age due to a reduced number or frequency of stem cells present in adult organs. These factors could contribute to human ageing and age-related disease (Rando, 2006). Thus, understanding the age-related functional and biological alterations in MSCs will be decisive to the success of therapeutic application of MSCs. Furthermore, it is proposed that long-term culture and

ageing may be regulated by similar mechanisms (Wagner *et al.*, 2009).

A number of studies have investigated biological ageing on MSCs (reviewed by: Oh *et al.*, 2014; Sethe *et al.*, 2006; Stolzing *et al.*, 2008). The cell number (Kasper *et al.*, 2009; Sethe *et al.*, 2006), potential to proliferate and differentiate (Kretlow *et al.*, 2008; Zaim *et al.*, 2012), frequency of colony forming units (Baxter *et al.*, 2004) and senescence-associated beta-galactosidase (Zhou *et al.*, 2008) are age-related. Additionally, MSC ageing is affected by oxidative stress (Stolzing *et al.*, 2006), DNA damage (Beausejour, 2007), telomere shortening (Beausejour, 2007) and apoptosis (Stolzing *et al.*, 2008). Using microarrays differentially expressed genes engaged in cell cycle. DNA replication and repair have been identified as age-related (Hacia *et al.*, 2008; Yu *et al.*, 2011), whilst a number of miRs are differentially expressed in MSC ageing (Pandey *et al.*, 2011; Wagner *et al.*, 2008; Yu *et al.*, 2011). Global proteomic studies of biological MSC ageing have been scarce, although studies have investigated *in vitro* ageing (Madeira *et al.*, 2012) and MSCs in old osteoarthritis patients (Camafeita *et al.*, 2014; Rollin *et al.*, 2008). However, one study in the rat identified age-affected susceptibility towards senescence, and molecular functions associated with cytoskeleton organisation and antioxidant defence (Kasper *et al.*, 2009). Previous studies have shown that genes involved in cell signalling, apoptosis and cell death undergo differential methylation in late passage MSCs when compared to early passage MSCs (Choi *et al.*, 2012; Redaelli *et al.*, 2012). In addition, analysis of methylation in biological ageing identified differentially methylated homeobox genes and genes pertinent to cell differentiation (Bork *et al.*, 2010).

Important bio-molecules in MSC ageing consist of the genes functioning as the information source defining limits of cell behaviour. Some of this information is processed through transcription, forming mRNA. This can be regulated by methylation and miRNA epigenetic control. mRNA transcription enables the production of proteins, the effector molecules that control the cellular state and behaviour. The purpose of this study was to investigate the global epigenetic (DNA-methylation and miRNA), RNA and protein changes in biological ageing of MSCs in a bid to understand the age-related functional and biological changes impacting upon their therapeutic applications in regenerative medicine and use as disease-specific biomarkers. The identification of individual transcripts, functionally coherent gene sets, or proteins that are under- or over-expressed with ageing may provide key insights into the mechanisms of ageing processes and age-related disease.

Materials and Methods

Mesenchymal stem cell culture

Human MSCs were purchased from young; $n = 4$ (21.8 y \pm 2.4 SD) (Stem Cell Technologies, Grenoble, France) and old; $n = 4$ (65.5 y \pm 8.3 SD) (Promocell, Heidelberg, Germany) donors and RNA, DNA and protein extracted following 8 population doublings. MSCs were

incubated to passage 4 in 5 % oxygen, as previously described (Peffers *et al.*, 2010). Briefly, they were grown as monolayers in Dulbecco's modified eagles medium (DMEM) (Invitrogen, Paisley, UK), supplemented with 10 % foetal calf serum, 100 units/mL penicillin, 100 mg/mL streptomycin (all from Invitrogen, Paisley, UK) and 500 ng/mL amphotericin B (BioWhittaker, Lonza, USA).

RNA isolation, library preparation for RNASeq and small RNASeq and sequencing

Total RNA was isolated from approximately two million cells using TRIzol (Invitrogen™ Life Technologies, Carlsbad, USA) (Chomczynski and Sacchi, 1987). RNA was purified using RNeasy spin columns with on-column DNase treatment (Qiagen, Crawley, UK) to remove residual gDNA, according to the manufacturer's instructions. RNA-Seq and small RNASeq library preparations and sequencing was undertaken by the Centre for Genomic Research, University of Liverpool. Sequencing used the Illumina HiSeq 2000 (Illumina, San Diego, USA) at 2 × 100-base pair (bp) paired-end sequencing with v3 chemistry for RNASeq. Multiplexed size selected small RNA library pools were sequenced on one lane of the Illumina HiSeq 2500 (Illumina, San Diego, USA) at 1 × 50 bp sequencing.

Total RNA integrity was confirmed, using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Ribosomal RNA (rRNA) was depleted from 8 total RNA samples by using the Ribo-Zero™ rRNA Removal Kit (Epicentre, Madison, USA) in accordance with manufacturer's instructions. For RNASeq cDNA libraries were prepared with the ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre, Madison, USA) using 50 ng rRNA depleted RNA as starting material in accordance with manufacturer's protocols, as previously described (Peffers *et al.*, 2013a). For both RNASeq and small RNASeq (smallSeq), the quantity and quality of the final pools were assessed using Qubit (Invitrogen™ Life Technologies, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). The final pooled library was diluted to 8 pmol before hybridisation. Sequencing was undertaken with the Illumina HiSeq 2000 (Illumina, San Diego, USA) at 2 × 100-base pair (bp) paired-end sequencing with v3 chemistry. 1 µg of total RNA were used to prepare small RNA libraries using the NEBNextMultiplex Small RNA Library Preparation Set for Illumina (New England BioLabs, Ipswich, USA), according to manufacturer's instructions. Following 12 cycles of amplification, the PCR amplified cDNA constructs were purified using the Qiagen QIAQuick PCR Purification kit (Qiagen, Crawley, UK). Multiplexed size selected small RNA library pools were sequenced on one lane of the Illumina HiSeq 2500 (Illumina, San Diego, USA) at 1 x 50 bp sequencing with rapid-run mode chemistry.

RNA data processing

The RNASeq data was processed as previously described (Peffers *et al.*, 2013a). Briefly, RNASeq sequence libraries 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 and quality trimming performed. The

trimmed R1-R2 read pairs, for each sample, were aligned to reference sequence (Web ref. 1) by using TopHat2 version 2.0.10 (Kim *et al.*, 2013) with default settings, except for the option `-g 1`. Read counts were obtained from the mapping results by using HTSeq-count and genome annotation (Web ref.2). The differential gene expression analysis was performed on R platform by using the edgeR package (Robinson *et al.*, 2010) 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 (log2 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 (Benjamini and Hochberg, 1995). In addition, FPKM (fragments *per* kilobase of exon *per* million fragments mapped) values were converted from count values for comparing expression levels among genes.

For smallSeq base-calling and de-multiplexing of indexed reads was performed by CASAVA version 1.8.2 (Illumina, San Diego, USA) to produce data, in fastq format. The raw fastq files were trimmed to remove Illumina adapter sequences using Cutadapt version 1.2.1 (Martin, 2011). The option `"-O 3"` was set, so the 3' end of any reads which matched the adapter sequence over at least 3 bp was trimmed. The reads were trimmed to remove low quality bases, using Sickle version 1.2 with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed.

Trimmed R1 reads were mapped to mature human microRNA (miRNA) and other human small RNAs included in Affymetrix miRNA 4.0 arrays. Reads were mapped to the reference sequences using Bowtie2 version 2.1.0 (Langmead and Salzberg, 2012). Single-end mapping was carried out using the "very-sensitive-local" set of options, which increases the sensitivity and accuracy of the alignment relative to the default parameters. In addition, FPKM (fragments *per* kilobase of exon *per* million fragments mapped) values were converted from count values for comparing expression levels among small RNAs.

For smallSeq DEGs and transcripts were extracted by applying the threshold FDR of less than 0.05 *p* values, which were generated by using Benjamini and Hochberg approach (Benjamini and Hochberg, 1995) and a 1.4 log₂ fold change (Log₂FC). All sequence data produced in this study have been submitted to National Centre for Biotechnology Information Gene Expression Omnibus (NCBI GEO) under Array Express accession number E-MTAB-3427.

Genomic DNA isolation, bisulphite treatment and methylation profiling

Genomic DNA was extracted from 2 million cells using the SureSelect gDNA Extraction Kit (Agilent, Santa Clara, USA) according to manufacturer's instructions, 500 ng of genomic DNA was then bisulphite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, USA). DNA methylation profiling of the samples was carried out

by Cambridge Genomic Services (Cambridge, UK), using the Illumina Infinium HumanMethylation450 Beadchip array (Illumina, Inc., San Diego, USA).

Methylation data processing

GenomeStudio (Illumina Inc., San Diego, USA) was used to extract the raw data. GenomeStudio provides the methylation data as β values: $\beta = M/(M + U)$ (*M* represents the fluorescent signal of the methylation probe; *U* represents the signal of the unmethylated probe). β values range from 0 (no methylation) to 1 (100 % methylation). The raw methylation data was processed using R (version 3.0.1) and the Watermelon package (version 2.12) as has been previously described (Pidsley *et al.*, 2013). Probes with a detection *p* value > 0.01 were removed. Age-related differential methylation was defined as Benjamini-Hochberg corrected *p* value (Benjamini and Hochberg, 1995) < 0.01 and a mean methylation difference ($\Delta\beta$ score) ≥ 0.10 (10 %), as has been previously reported (Whitaker *et al.*, 2013).

Protein extraction, sample preparation and mass spectrometry

Cell pellets containing 2 million cells were lysed using a probe sonicator in the presence of protease inhibitors (Complete protease Inhibitors EDTA-free, Roche, Lewes, UK). For each sample 100 μ g protein was reduced, alkylated and trypsin digested (Peffers *et al.*, 2013b). Liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed using a NanoAcquity™ ultraperformance LC (Waters, Manchester, UK) on-line to an LTQ-Orbitrap Velos mass spectrometer (Thermo-Fisher Scientific, Hemel Hempstead, UK) (Peffers *et al.*, 2013b), as previously described. The proteomics data were deposited to the ProteomeXchange Consortium (Vizcaino *et al.*, 2014) via the PRIDE partner repository with the dataset identifier PXD001952.

LC-MS/MS analysis was performed using nanoAcquity™ ultraperformance LC (Waters, Manchester, UK) on-line to an LTQ-Orbitrap Velos mass spectrometer (Thermo-Fisher Scientific, Hemel Hempstead, UK), using an electrospray ionisation ion source containing a 10 μ m coated Pico-tip emitter (Presearch Ltd., Basingstoke, UK). Aliquots of tryptic peptides equivalent to 300 ng of tendon fascicle protein were loaded onto a 180 μ m \times 20 mm C18 trap column (Waters) at 5 μ L/min in 99 % solvent A (water plus 0.1 % formic acid) and 1 % solvent B (acetonitrile plus 1 % formic acid) for 5 min and subsequently back-flushed onto a C18 pre-equilibrated analytical column (75 μ m \times 15 mm Waters) using a flow rate of 0.3 μ L/min. Xcalibur 2.0 software (Thermo-Electron, Hemel Hempstead, UK) was used to operate the LTQ-Orbitrap Velos mass spectrometer in data-dependant acquisition mode. The survey scan was acquired in the Orbitrap with a resolving power set to 30,000 (at 400 *m/z*). MS/MS spectra were concurrently acquired on the 20 most intense ions from the high-resolution survey scan in the LTQ mass spectrometer. Charge state filtering > 1 was used where unassigned precursor ions were not selected for fragmentation. Fragmentation parameters in the LTQ mass spectrometer were: normalised collision energy, 30;

Table 1. qRT-PCR primer sequences.

Gene	Ensemble Gene ID	Primer Sequences
ACAN ^a	ENSG00000157766	F: TCGAGGACAGCGAGGCC R: TCGAGGGTGTAGCGTGTAGAGA
ALX1	ENSG00000180318	F: TTTTCACTGACTCTCTTCTACTGG R: ATATTGGCGGTGTGCTCCTT
COL3A1 ^b	ENSG00000168542	F: TTCTGAAATAGTCAAATACGAAATTAGA R: CTGGGACTCATAGAATAACAATCTG
EGF ^b	ENSG00000138798	F: ACAGGAGGCTTCGGAGTT R: AATCAGGCAATTTACTTACAATCTT
HOXB6	ENSG00000108511	F: CGCCTCCCCTCCCAATG R: CGCATAGCCCCGACGAATAG
HOXB7	ENSG00000260027	F: GTTGCTGCCCTCCTGAG R: GCGAGTGGTAGGTTTTGGG
MAB21L2	ENSG00000181541	F: CCGCTCAACAACCTACCACAT R: TCGTCCCAGTCCGTTTCTC
MMP16 ^b	ENSG00000156103	F: ACCCGTGTAACCCTTTGAAGA R: AACCTGAACTTCTTGAACCTGTG
RPS8	ENSG00000142937	F: CAAGAAATACCGTGCCTGAG R: CGATGATCCTTGTTTTACGAGTAC
PITX2	ENSG00000164093	F: AACTCCGCCCTTGAAAGACT R: GATTCAAGTGGTTTCTTTTCTCTATA
TGFA	ENSG00000163235	F: GCCACCTCCCCACAACAG R: ATTAACAGTTTCCCCTCCTTCAC
TGFB1	ENSG00000105329	F: CACTCCCACTCCCTCTCTC R: GTCCCTGTGCCTTGATG
RBL1	ENSG00000080839	F: CAGATTGTGGGTAAGCAGAAGAGTC R: AACTGGCACTGAGGAAAAGGC
p21	ENSG00000124762	F: GGCAGACCAGCATGACAGATTTC R: CGGATTAGGGCTTCTCTTGG
p85	ENSG00000145675	F: TTGCGAGGGAAGCGAGATGGC R: TGCTGCACAAGGGAGGTGTGT
p53	ENSG00000141510	F: CTCCGGGTGAGCAGTAGG R: TGGGGTTATAGGGAGGTCAA

^a(Martin *et al.*, 2001), ^b(Peffers *et al.*, 2015) denote primer pairs published previously. F, forward; R, reverse. ACAN; aggrecan, ALX1; ALX homeobox 1, COL3A1; collagen III alpha 1, EGF; epidermal growth factor, HOXB6; homeobox B6, HOXB7; homeobox B7, MAB21L2; mab-21-like 1, MMP16; matrix metalloproteinase 16, KMT2A; lysine (K)-specific methyltransferase 2A, RPS8; ribosomal protein 8, PITX2; paired-like homeodomain 2, TGFA; transforming growth factor alpha, TGFB; transforming growth factor beta, RBL1; retinoblastoma-like 1, p21; cyclin-dependent kinase inhibitor 1A, p85; phosphoinositide-3-kinase, regulatory subunit 1, p53; tumour protein p53.

activation, 0.250; activation time, 10 ms; minimum signal threshold, 500 counts with isolation width 2 *m/z*.

Label-free quantification

For label-free quantification ageing MSCs the Thermo raw files of the acquired spectra from in-solution tryptic digests were analysed by the ProgenesisQI™ software (Version 1, Waters, Manchester, UK), as previously described (Peffers *et al.*, 2013b). Briefly, the top five spectra for each feature were exported from ProgenesisQI™ and utilised for peptide identification in PEAKS® 7 (Bioinformatics Solutions Inc., Ontario, Canada) using the Unihuman database. Search parameters used were: 10 ppm peptide mass tolerance and 0.6 Da fragment mass tolerance; one missed cleavage allowed; fixed modification, carbamidomethylation; variable modifications, methionine oxidation. Proteins were identified with a false discovery rate (FDR) of 1 % and a minimum of two peptides *per* protein. The resulting peptide-spectrum matches were imported into ProgenesisQI™ for label-free relative quantification. Differentially expressed proteins were defined with FDR $p < 0.05$ and ± 2 -fold regulation.

Functional analysis of transcriptomic, proteomic and methylation data

To determine gene ontology, functional analyses, networks, canonical pathways and protein-protein interactions of

age-related differentially expressed genes, methylated genes and proteins we performed the analyses using Panther Classification System (Mi *et al.*, 2013), the functional analysis and clustering tool from the Database for Annotation, Visualisation, and Integrated Discovery (DAVID bioinformatics resources 6.7) (Huang *et al.*, 2009), Ingenuity Pathway Analysis (IPA) (Web ref. 3) and the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) tool Version 10.0 (Franceschini *et al.*, 2013), as previously described (Peffers *et al.*, 2013a; Peffers *et al.*, 2015). Targetscan v6.2 (Web ref. 4) was used to identify potential targets of miR-199.

Relative gene expression using real-time polymerase chain reaction (qRT-PCR)

qRT-PCR was undertaken on similarly sourced MSCs at P4 from an independent cohort from those used for the RNA-Seq analysis young; $n = 3$ (22.2 y \pm 2.3 SD) and old; $n = 3$ (64.8 y \pm 6.6 SD) using methods previously described (Peffers *et al.*, 2015). Primers were either validated in previous publications (Bracken *et al.*, 2003; Conklin *et al.*, 2012; Martin, 2011; Peffers *et al.*, 2015) and supplied by Eurogentec (Seraing, Belgium) or were designed and validated commercially (Primer Design, Southampton, UK). Steady-state transcript abundance of potential endogenous control genes was measured in the RNA-Seq data. Assays for four genes – Gluteraldehyde

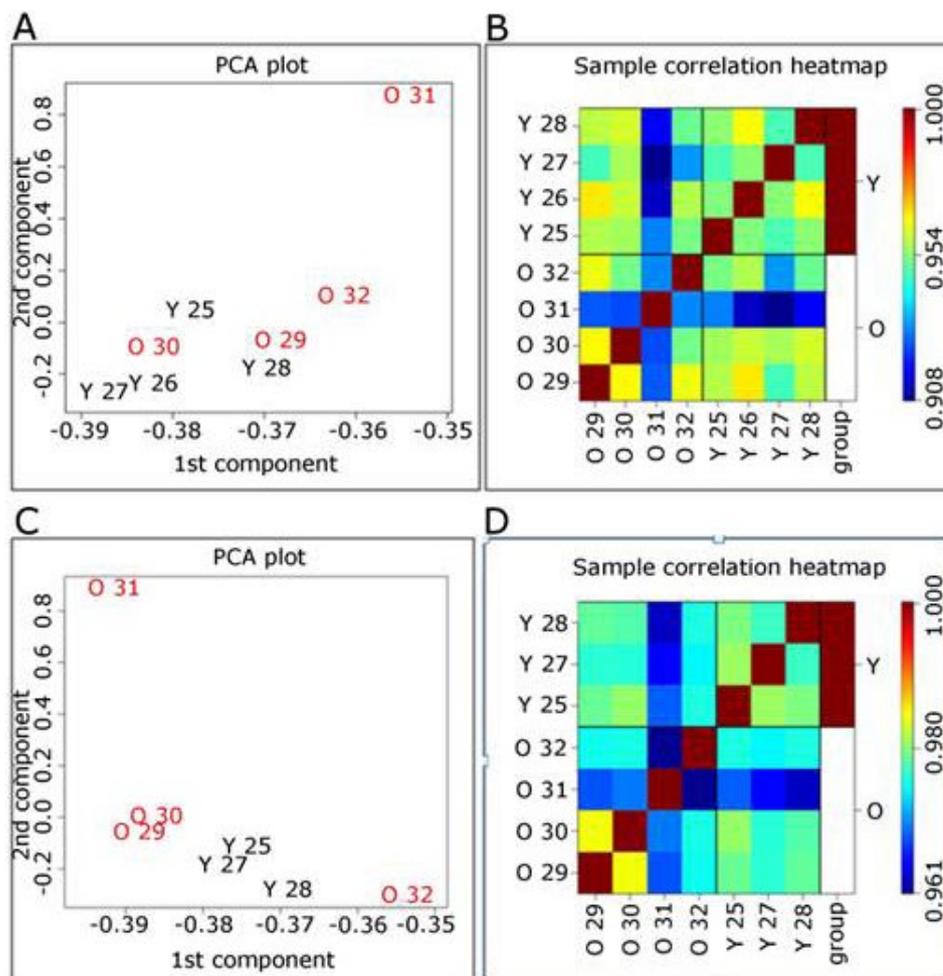
Table 2. Summary of RNA sequence alignment to the human genome.

Sample	Total reads	Mapped Reads	% reads mapped	mapped read pairs	Singly mapped reads	Genes aligned	% genes aligned
msc_young3	66,306,012	60,634,324	91.45	28,916,012	2,802,300	27,930	44.23
msc_young5	69,842,190	64,120,013	91.81	30,648,269	2,823,475	28,101	44.50
msc_young6	58,743,174	51,880,907	88.32	24,653,613	2,573,681	25,548	40.45
msc_young8	58,194,714	53,513,610	91.96	25,501,727	2,510,156	28,506	45.14
msc_old3	51,067,276	46,692,381	91.43	22,140,647	2,411,087	26,381	41.77
msc_old4	57,931,724	53,115,861	91.69	25,462,506	2,190,849	27,168	43.02
msc_old5	59,896,936	53,901,889	89.99	25,288,270	3,325,349	27,929	44.22
msc_old6	58,865,142	53,602,928	91.06	25,325,008	2,952,912	28,339	44.87

Table 3. Summary of small RNA sequence alignment to the human genome.

Sample	Total reads	Mapped Reads	% reads mapped	% mapped to rRNA	Genes aligned	% genes aligned
msc_young3	9,790,558	7,875,474	80.44	5.8	1443	34.30
msc_young6	11,161,583	8,685,493	77.82	7.6	1568	37.27
msc_young8	10,438,112	8,199,841	78.56	7.8	1424	33.85
msc_old3	13,025,963	10,102,460	77.56	7.2	1542	36.65
msc_old4	10,804,811	8,544,026	79.08	5.8	1492	35.46
msc_old5	13,316,974	10,048,663	75.46	7	1643	39.05
msc_old6	8,281,321	5,926,098	71.56	11.3	1420	33.75

Fig. 1. Variation in the RNASeq data demonstrated by principal component analysis (PCA) and hierarchical clustering. (A) The differentially abundant transcripts of the eight samples in their 2D plane spanned by their first two principle components. PCA plot of \log_2 gene expression data showed the effect of age was weak. (B) To complement the PCA, we undertook a hierarchical clustering of the samples to determine if the samples cluster into young and old groups. Samples from the same age group were closely correlated. (C) PCA plot of \log_2 small RNA gene expression showed the effect of age was weak. (D) Hierarchical clustering of the samples to determined clustering into young and old groups.



dehydrogenase (GAPDH), ribosomal protein 13 (RPS8), ribosomal protein 13 (RPS13), 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

(Vandesompele *et al.*, 2002). RPS8 was selected as the most stable endogenous control gene. For miR-199b cDNA synthesis was performed using 100 ng RNA and miRscript RT kit II (Qiagen, Crawley, UK) according to the manufacturer’s protocol. qRT-PCR analysis was performed

using miRScript SybrGreen Mastermix (Qiagen, Crawley, UK) using Rnu-6u as the endogenous control. Relative expression levels were calculated by using the $2^{-\Delta Ct}$ method (Livak and Schmittgen, 2001). Primers pairs used in this study are listed (Table 1).

Western blot validation

Validation studies were undertaken with the same independent cohort as for the relative gene expression studies using methods previously described using 20 μ g protein (Peffers *et al.*, 2013b). Membranes were probed with primary antibodies against the following; rabbit polyclonal to SOD2 (1:1000, Enzo Life Sciences, Exeter, UK), and anti GAPDH-horseradish peroxidase (HRP) conjugate (1:10,000; Sigma, Dorset, UK). A HRP conjugated goat anti-rabbit secondary antibody was used at 1:2000 (Sigma, Dorset, UK).

Statistical analysis

Statistically significant differences in relative gene expression were identified using a Mann-Whitney U test. Differences in relative band intensity in immunoblotting identified with a Student's *t*-test. Statistical analyses were undertaken using S-Plus (version 7.0; Tibco Software Inc., Palo Alto, USA) and Excel (version 2007; Microsoft, Redmond, USA).

Results

Overview of RNASeq and smallSeq data

For the RNASeq data an average of 54.7 million pairs of 100 bp paired-end reads *per* sample was generated that aligned to the human reference sequence. Mapping results are summarised in Table 2. Of the 63,152 human genes, between 40.5 % and 45.1 % had at least one read aligned. 13,595 genes had no reads aligned from any of the samples. This is similar to the output of other RNA-Seq studies (Peffers *et al.*, 2013a; Peffers *et al.*, 2015).

In the smallSeq data, an average of 8.2 million 50 bp single-end reads was generated. Based on data variation assessment, one young sample was classed as an outlier and so removed from subsequent data analysis. This represented 75.1 % of reads mapped. 2,267 human small RNAs (reference sequences) had at least one read aligned (between 1,041 and 1,586 in 7 different samples). 1,939 of the small RNAs had no reads aligned from any of the 7 samples. Mapping results are summarised in Table 3 and are similar to other small RNASeq studies (Buitrago *et al.*, 2015; Yang *et al.*, 2014). Reads were used to estimate small RNA transcript expression of all samples using FPKM in order to identify the most abundant microRNAs (miRs) and small nucleolar RNAs (snoRNAs). Table 4 highlights the top 10 highly expressed small RNAs genes within each class.

Identification of differentially expressed genes and differentially spliced genes using RNASeq

For RNASeq a principal component analysis (PCA) plot of \log_2 gene expression data identified biological variation within both sample groups indicating that the age-effect is

weak. The variability between old donors was greatest (Fig. 1A). However, hierarchical clustering using a sample-to-sample distance matrix identified clustering by age group (Fig. 1B).

There were 46 differentially expressed transcribed elements between young and old, principally protein coding genes, but also three long non-coding RNAs (lncs) (*per* $\pm 1.4 \log_2$ fold change, FDR < 0.05). Of these 31 were higher in the older MSCs and 15 were lower (Table 5).

Table 4. The top 10 most abundant small RNAs (CDBox, HAcaBox, scaRNAs, snoRNAs and miRs) within each class determined using FPKM.

Identification	Type of small RNA	Mean All	Standard Deviation All
U63	CDBox	52627	16653
U74	CDBox	40952	15610
U50	CDBox	33773	13785
U26	CDBox	29755	8271
U81	CDBox	29553	12486
mgh28S-2409	CDBox	26763	8334
U104	CDBox	25581	6348
U58B	CDBox	25280	13114
U44	CDBox	24741	12804
U30	CDBox	24549	10752
ACA18	HAcaBox	241	124
E3	HAcaBox	153	51
HBI-100	HAcaBox	152	54
ACA55	HAcaBox	146	53
ACA61	HAcaBox	116	32
U64	HAcaBox	79	25
ACA7	HAcaBox	77	28
ACA3	HAcaBox	70	21
U19	HAcaBox	70	24
E2	HAcaBox	69	30
ENSG00000222345	snoRNA	2363	1238
ENSG00000202252	snoRNA	2004	498
ENSG00000265706	snoRNA	1449	603
ENSG00000207118	snoRNA	872	327
ENSG00000222185	snoRNA	697	302
ENSG00000252787	snoRNA	157	73
ENSG00000222937	snoRNA	130	48
ENSG00000200879	snoRNA	120	34
ENSG00000199262	snoRNA	114	67
ENSG00000253067	snoRNA	82	85
U90	scaRna	78	26
U93	scaRna	69	20
ACA66	scaRna	53	9
U88	scaRna	39	21
ACA26	scaRna	30	10
U85	scaRna	21	10
U100	scaRna	7	3
U92	scaRna	7	3
ACA35	scaRna	5	3
mgU2-19-30	scaRna	2	1
hsa-miR-21-5p	miRNA	15433276	2935100
hsa-miR-100-5p	miRNA	8777127	1623490
hsa-let-7i-5p	miRNA	2009764	522883
hsa-miR-125b-5p	miRNA	1441052	479950
hsa-miR-26a-5p	miRNA	1305126	384004
hsa-miR-143-3p	miRNA	1197489	679965
hsa-let-7a-5p	miRNA	998801	200421
hsa-miR-221-3p	miRNA	881823	230714
hsa-let-7f-5p	miRNA	766655	101225
hsa-miR-222-3p	miRNA	670149	199778

Table 5. The DEG between young and old MSCs.

Gene	Gene Name	Log ₂ fold change	FDR-adjusted <i>p</i> value
ALX1	ALX homeobox 1	6.77	0.000
CST2	cystatin SA	4.96	0.042
PODXL	podocalyxin-like	4.80	0.021
MKRN3	makorin ring finger protein 3	4.64	0.032
PITX2	paired-like homeodomain 2	4.32	0.000
CCL20	chemokine (C-C motif) ligand 20	4.25	0.025
HOXB6	homeobox B6	4.12	0.005
TGFA	transforming growth factor, alpha	4.08	0.032
HOXB4	homeobox B4	3.85	0.009
HOXB7	homeobox B7	3.68	0.000
IRF6	interferon regulatory factor 6	3.62	0.042
PLAT	plasminogen activator, tissue	3.25	0.001
MTND4P12	MT ND4 pseudogene 12	2.87	0.035
OSBPL6	oxysterol binding protein-like 6	1.79	0.048
EML1	echinoderm microtubule associated protein like 1	-1.43	0.009
TBX18	T-box 18	-1.98	0.009
MITF	microphthalmia-associated transcription factor	-2.02	0.007
FAM134B	family with sequence similarity 134, member B	-2.34	0.043
MAF	v-maf avian musculoaponeurotic fibrosarcoma	-2.65	0.002
PDE1A	phosphodiesterase 1A, calmodulin-dependent	-2.69	0.036
FOXP2	forkhead box P2	-3.26	0.007
DPT	dermatopontin	-3.55	0.036
NOS1	nitric oxide synthase 1 (neuronal)	-3.66	0.010
FMO3	flavin containing monooxygenase 3	-3.89	0.011
OSR2	odd-skipped related transcription factor 2	-3.93	0.010
CYP19A1	cytochrome P450, family 19, subfamily A, polypeptide 1	-3.94	0.032
CD36	CD36 molecule (thrombospondin receptor)	-4.08	0.007
MAB21L1	mab-21-like 1 (<i>C. elegans</i>)	-4.42	0.005
DSG2	desmoglein 2	-4.70	0.004
FAM132B	family with sequence similarity 132, member B	-4.75	0.019
GMNC	geminin coiled-coil domain containing	-4.87	0.018
DIO2	deiodinase, iodothyronine, type II	-4.89	0.036
IGSF9B	immunoglobulin superfamily, member 9B	-4.90	0.033
SEMA3E	semaphorin 3E	-5.03	0.004
EPHA7	EPH receptor A7	-5.04	0.002
TNXB	tenascin XB	-5.21	0.018
HAND2	heart and neural crest derivatives expressed 2	-5.30	0.023
MAB21L2	mab-21-like 2 (<i>C. elegans</i>)	-5.49	0.000
MCF2L	MCF.2 cell line derived transforming sequence-like	-5.52	0.003
KLHL41	kelch-like family member 41	-5.99	0.007
RP11-670E13.5	RP11-670E13.5	-6.73	0.048
AC004870.4	AC004870.4	-7.57	0.042
OGN	osteoglycin	-8.13	0.007
ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	-9.25	0.001
GPC5	glypican 5	-9.61	0.036

Positive log₂ fold change values represent lower expression in old MSCs, negative log₂ fold change values represent higher expression in old MSCs. FDR; false discovery rate.

In total, 79,387 isoforms were detected in young and 77,189 isoforms were detected in old MSCs. No isoforms were differentially expressed. However, using Cuffdiff to calculate changes in the relative splice abundances by quantifying the square root of the Jensen-Shannon divergence on primary transcripts with at least two isoforms, identified 23 with differential splicing events (alternative splicing (AS)) (Table 6).

For the smallSeq PCA of log₂ gene expression data indicated the age-effect was weak. The variability between old donors was greatest (Fig. 1C). Hierarchical clustering clustered into young and old groups (Fig. 1D). The only

differentially expressed small RNA was miR199b (FDR; *p* = 0.002, log₂FC; -4.3) which was increased in older donors and confirmed using qRT-PCR (*p* = 0.0085) (Fig. 2).

Similar to the reports from our (Peffers *et al.*, 2013a; Peffers *et al.*, 2015) and other groups (Marioni *et al.*, 2008), quantification of transcripts by deep sequencing correlates with the results derived from qRT-PCR analyses reflecting the accuracy and reliability of deep sequencing analyses. There was no change in expression of the cellular senescence markers RBL1, p21, p16, p53 and p85 (Table 7).

Table 6. Table contains the genes in which there was a significant difference in splicing events in ageing MSCs.

Gene ID	Gene	Gene Name	Type	Locus	Square root Jensen-Shannon divergence	q-value
XLOC_026797	CENPA	centromere protein A	protein coding	2:26987151-27023935	0.83	0.00
XLOC_033690	CNOT10	CCR4-NOT transcription complex, subunit 10	protein coding	3:32726636-32815367	0.83	0.00
XLOC_014317	DHRS4L2	dehydrogenase/reductase (SDR family) member 4 like 2	protein coding	14:24407939-24475617	0.83	0.00
XLOC_002851	EMC1	ER membrane protein complex subunit 1	protein coding	1:19536994-19586622	0.83	0.00
XLOC_034298	GTF2E1	general transcription factor IIE, polypeptide 1, alpha 56 kDa	protein coding	3:120405527-120501916	0.83	0.00
XLOC_011608	ING4	inhibitor of growth family, member 4	protein coding	12:6759445-6772314	0.83	0.00
XLOC_036966	MRPL1	mitochondrial ribosomal protein L1	protein coding	4:78783673-78873944	0.83	0.00
XLOC_045977	NT5C3A	5'-nucleotidase, cytosolic IIIA	protein coding	7:32535037-33102409	0.83	0.00
XLOC_045911	OSBPL3	oxysterol binding protein-like 3	protein coding	7:24836157-25021253	0.43	0.02
XLOC_003042	PEF1	penta-EF-hand domain containing 1	protein coding	1:32083286-32169920	0.83	0.00
XLOC_010138	PUS3	pseudouridylylase synthase 3	protein coding	11:125753508-125773116	0.83	0.00
XLOC_023041	RAB31	RAB31, member RAS oncogene family	protein coding	18:9708161-9862548	0.83	0.00
XLOC_041114	RAPGEF6	Rap guanine nucleotide exchange factor (GEF) 6	protein coding	5:130759613-131132710	0.83	0.00
XLOC_017355	SLC12A6	solute carrier family 12 (potassium/chloride transporter), member 6	protein coding	15:34525459-34630261	0.83	0.00
XLOC_021864	SREBF1	sterol regulatory element binding transcription factor 1	protein coding	17:17584786-17740325	0.83	0.00
XLOC_017808	ULK3	unc-51 like kinase 3	protein coding	15:75128456-75135687	0.83	0.00
XLOC_007014	XPNPEP1	X-prolyl aminopeptidase (aminopeptidase P) 1, soluble	protein coding	10:111624523-111683311	0.83	0.00
XLOC_040967	GJA1P1		pseudogene	5:108083522-108532542	0.83	0.00
XLOC_040175	HMGB3P22		pseudogene	5:179105583-179157926	0.83	0.00
XLOC_038137	RP11-123J14.2		pseudogene	4:77870855-77961537	0.83	0.00
XLOC_026337	CTC-301O7.4		LNC	19:49834873-49891338	0.83	0.00
XLOC_026537	CTD-2537I9.12		novel antisense	19:56165511-56186081	0.83	0.00
XLOC_041176	VTRNA2-1	vault RNA 2-1	premiR and vault RNA	5:135416159-135416286	0.83	0.00

Gene ontology (GO) and IPA analysis of DEGs and AS genes

All DEGs (adjusted $p < 0.05$ and 1.4 \log_2 fold change) were analysed in DAVID. Significant annotations included terms 'developmental protein', 'cell proliferation' and skeletal system morphogenesis. The DEGs were input into IPA (Peffers *et al.*, 2015). Interesting age-related features were determined from the gene networks inferred. The top-scoring network was 'embryonic development, organismal development, cellular function and maintenance' (Fig. 3). The top diseases and biological functions are shown in Table 8. Transforming growth factor β (TGF β) was identified as a potential upstream regulator of some of the pathway effects identified. The genes podocalyxin (PODXL), homeobox B6 (HOXB6) and TGF α were identified as biological targets with expression pairing consistent with miR199 (Table 9).

GO analyses indicated enrichment in genes associated with metabolic processes in 12 out of the 23 genes identified as undergoing AS (Fig. 4). Additionally 20 % of AS protein-coding genes were identified in PANTHER as transcription factors.

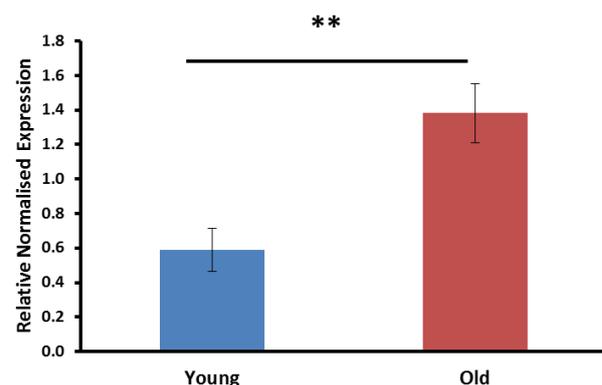


Fig. 2. Validation of age-related differential expression of miR199b with smallSeq using qRT-PCR. Gene expression is relative to Rnu-6u. Data are represented as fold change. Histograms represent means \pm SEM * $p = 0.0085$. Data were evaluated using Student's t -test.

Comparison of the DNA methylome in ageing MSCs

Unsupervised hierarchical clustering revealed that young and old samples are largely distinguished by their DNA methylome (Fig. 5). In total, 187 differentially methylated loci (DML) were identified, including 44

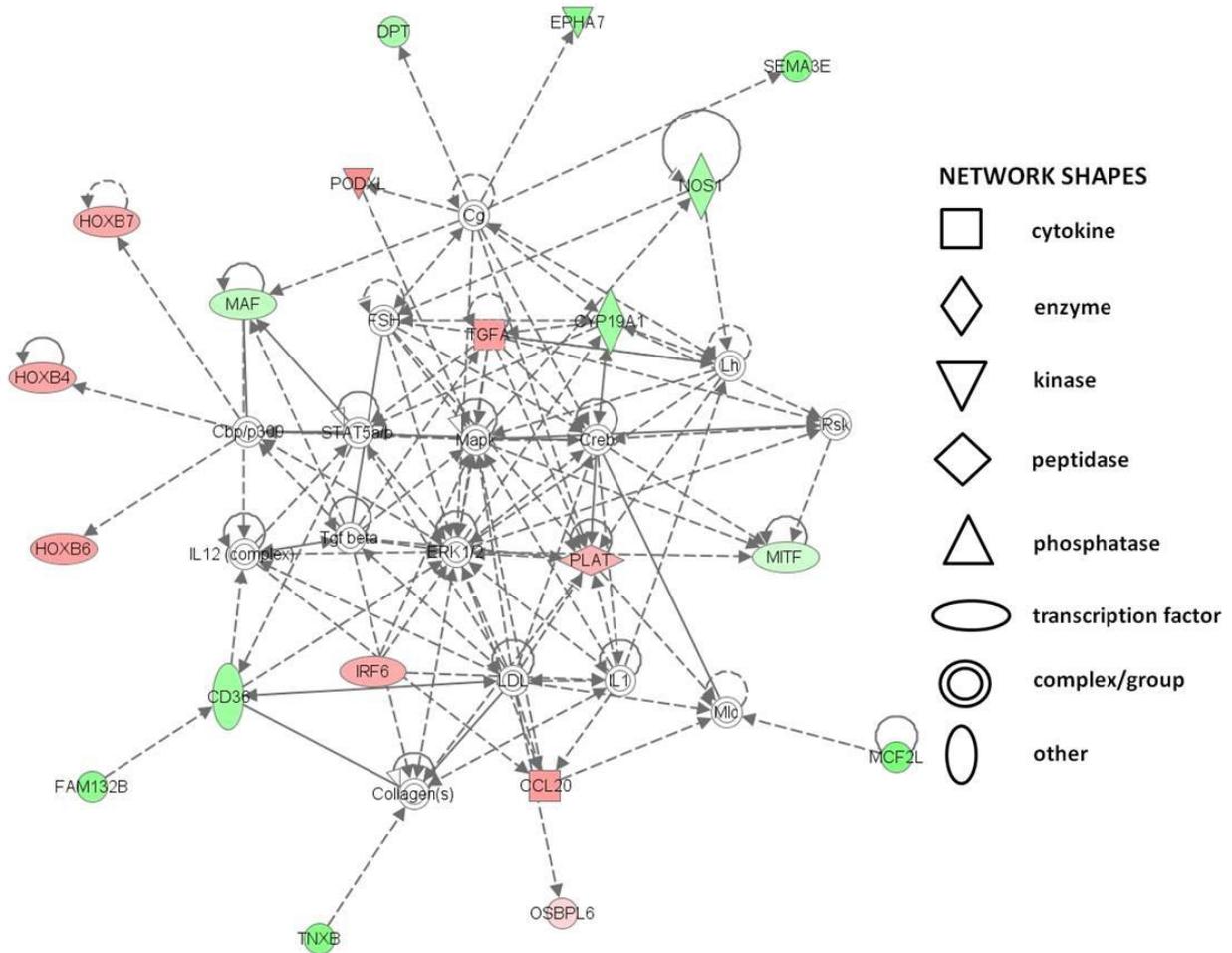


Fig. 3. Pathway analysis of DEGs. Top-scoring network derived from the 43 genes differentially expressed in ageing. Ingenuity pathway analysis (IPA) identified connective tissue development and function, embryonic development, organ development as the principal associated network functions with a score of 40. This is a graphical representations between molecules identified in our data in their respective networks. Green nodes; upregulated gene expression in older MSCs, red nodes; downregulated gene expression in older MSCs. Intensity of colour is related to higher fold-change. Key to the main features in the networks is shown.

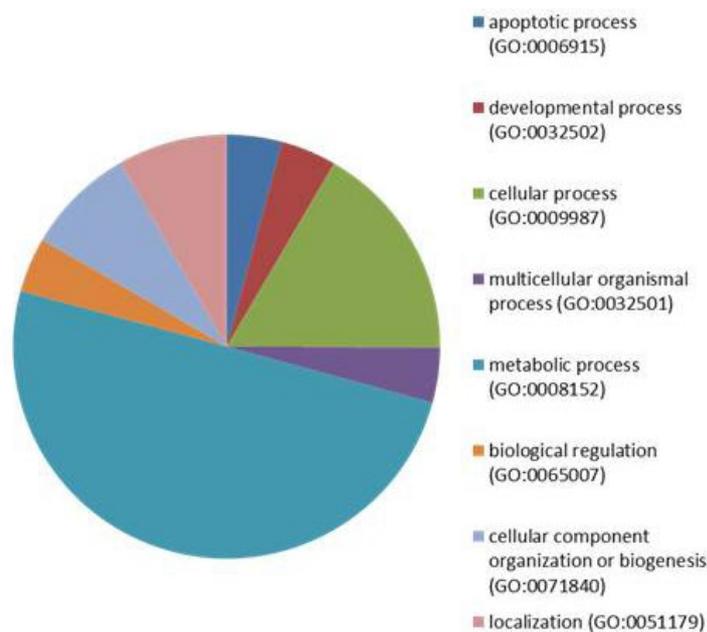


Fig. 4. Pie chart representing the occurrence of major ontological annotations associated with AS in ageing MSCs in biological process categories of the Panther Classification System.

Table 7. Real-time polymerase chain reaction analysis of selected genes reveals good correlation with RNA-Seq results.

Gene name	RNA-Seq Results			RT-PCR Results		
	Differential expression	Significant Log ₂ FC	q-value	Age		
				p-value Young	Old	
ALX1	lower in old	6.77	0.00	0.55 ± 0.14	0.04 ± 0.03	0.01
PITX2		4.32	0.00	0.23 ± 0.12	0.05 ± 0.03	0.32
HOXB6		4.12	0.00	327.47 ± 62.3	0.85 ± 0.52	0.00
TGFA		4.08	0.03	1.18 ± 0.93	0.14 ± 0.02	0.04
HOXB7		3.68	0.00	6.68 ± 64.2	0.56 ± 0.43	0.04
MAB21L2	higher in old	-5.49	0.00	0.25 ± 0.13	0.11 ± 0.07	0.05
ACAN	no change	no change	0.84	0.17 ± 0.13	0.09 ± 0.07	NS
COL3A1			NS	0.11 ± 0.02	0.27 ± 0.12	NS
EGF			NS	0.21 ± 0.18	0.17 ± 0.08	NS
MMP16			NS	11.38 ± 7.17	0.38 ± 0.35	NS
TGFB			NS	0.27 ± 0.28	0.78 ± 0.46	0.02
RBL1			NS	2.8 ± 0.49	0.22 ± 0.46	NS
p21			NS	1.10 ± 0.46	1.31 ± 0.23	NS
p85			NS	0.48 ± 0.04	0.61 ± 0.13	NS
p53			NS	0.88 ± 0.22	0.83 ± 0.35	NS
p16			NS	0.00 ± 0.00	1.19 ± 1.18	NS

Values for quantitative real-time polymerase chain reaction (qRT-PCR) are the mean ± standard error of relative expression levels normalised to expression of RPS8 (to two decimal places). Statistical significance was tested by using Mann-Whitney U Test. qRT-PCR results are expressed as $2^{-\Delta CT}$, NS; not significant. ACAN; aggrecan, ALX1; ALX homeobox 1, COL3A1; collagen III alpha 1, EGF; epidermal growth factor, HOXB6; homeobox B6, HOXB7; homeobox B7, MAB21L2; mab-21-like 1, MMP16; matrix metalloproteinase 16, RPS8; ribosomal protein 8, PITX2; paired-like homeodomain 2, TGFA; transforming growth factor alpha, TGFB; transforming growth factor beta, RBL1; retinoblastoma-like 1, p21; cyclin-dependent kinase inhibitor 1A, p85; phosphoinositide-3-kinase, regulatory subunit 1, p53; tumour protein p53, p16; tumour protein p16.

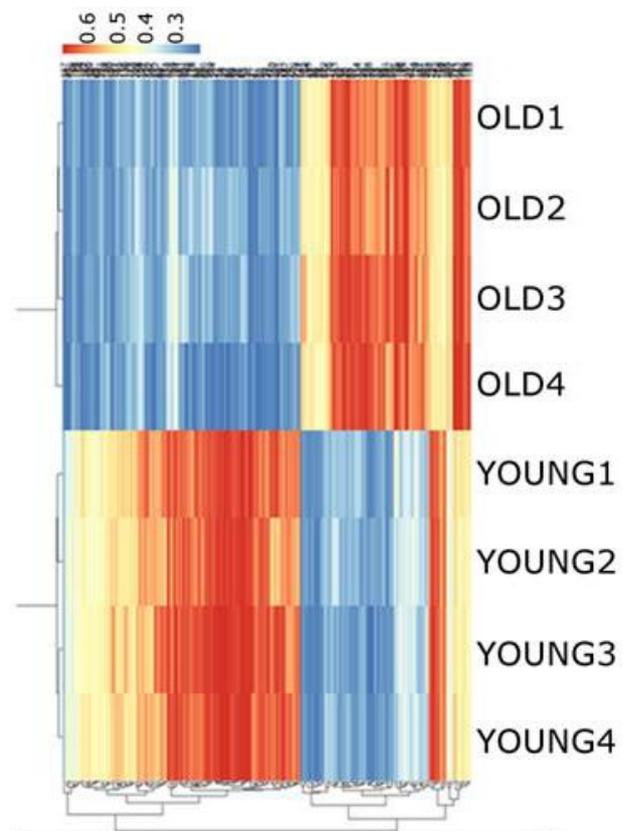


Fig. 5. Heatmap showing the unsupervised clustering of the 187 DMLs between the young and old samples. DMLs were defined as at least a 10 % difference in methylation between the two groups, and an FDR-corrected p value of < 0.05 . The dendrogram at the top shows the clustering of the loci and the dendrogram to the left show clustering of the samples. The methylation scale is shown at the top of the heatmap (1 = 100 % methylation, 0 = no methylation).

Table 8. Table of the top diseases and biological functions as determined by IPA for DEGs from RNASeq analysis, DML following methylation analysis and proteomics analysis demonstrates many similarities in the types on processes effected in ageing MSCs at the transcriptome, epigenetic and protein level.

Category	Classification	RNASeq		Methylation		Proteomics	
		p-value	No. molecules	p-value	No. molecules	p-value	No. molecules
Diseases and disorders	Developmental Disorder	5.36E-08 - 8.41E-3	21	3.82E-07 - 1.18E-2	18	1.40E-08 - 5.79E-3	23
	Gastrointestinal Disease	1.03E-06 - 8.41E-3	14				
	Connective Tissue Disorders	2.95E-06 - 8.41E-3	11	4.48E-06 - 1.18E-2	10	6.32E-07 - 5.79E-3	13
	Skeletal and Muscular Disorders	2.95E-06 - 8.41E-3	12	3.82E-07 - 1.18E-2	16		
	Cancer	3.12E-05 - 8.41E-3	33	9.76E-06 - 9.87E-3	11		
	Neurological Disease			1.06E-04 - 1.18E-2	9		
	Metabolic Disease					3.17E-11 - 5.79E-3	41
	Hereditary Disorder					1.40E-08 - 5.79E-3	37
Molecular and cellular functions	Dermatological Diseases and Conditions					1.49E-07 - 5.79E-3	24
	Cell Morphology	3.12E-06 - 8.41E-3	12				
	Cellular Assembly and Organization	3.12E-06 - 8.41E-3	10				
	Cellular Function and Maintenance	1.71E-05 - 8.41E-3	13				
	Cellular Development	8.34E-05 - 8.41E-3	22	1.16E-05 - 1.10E-2	20		
	Gene Expression	1.66E-04 - 6.31E-3	16				
	Cellular Growth and Proliferation			1.28E-05 - 1.18E-2	20	9.31E-08 - 5.79E-3	54
	Gene Expression			6.98E-05 - 1.18E-2	14		
	Cell Death and Survival			3.12E-04 - 9.87E-3	19	3.11E-07 - 5.79E-3	51
	Cellular Assembly and Organization			1.12E-03 - 5.94E-3	4		
	Nucleic Acid Metabolism					1.51E-06 - 5.79E-3	19
	Small Molecule Biochemistry					1.51E-06 - 5.79E-3	40
Physiological system, development and function	Energy Production					2.93E-06 - 6.72E-5	16
	Embryonic Development	6.74E-13 - 8.41E-3	25	2.68E-12 - 1.23E-2	21		
	Organismal Development	6.74E-13 - 8.41E-3	26	2.68E-12 - 1.23E-2	24		
	Organ Development	4.75E-10 - 8.41E-3	21	3.19E-11 - 1.23E-2	19		
	Tissue Development	4.75E-10 - 8.41E-3	30			4.44E-05 - 5.79E-3	31
	Skeletal and Muscular Development, Function	8.66E-08 - 8.05E-3	17	9.02E-14 - 1.23E-2	20	9.59E-06 - 5.79E-3	22
	Connective Tissue Development and Function			3.19E-11 - 1.18E-2	17		
	Organ Morphology					9.59E-06 - 5.79E-3	23
	Tissue Morphology					4.32E-05 - 5.79E-3	17
	Haematological System Development and Function					9.10E-05 - 5.79E-3	21

Table 9. Using Targetscan 6.2 for miR199 targets were identified in which the corresponding genes were also DEG in ageing MSCs.

Gene	Gene name	Representative transcript	RNASeq		qRT-PCR	miR and mRNA pairing
			RNASeq log2FC	Higher in		
PODXL	podocalyxin-like	NM_001018111	4.8	young	no difference	yes
MAB21L1	mab-21-like 1 (<i>C. elegans</i>)	NM_005584	-4.4	old	no difference	no
EPHA7	EPH receptor A7	NM_004440	-5.0	old	no difference	no
HOXB6	homeobox B6	NM_018952	4.1	young	higher in young	yes
TGFA	transforming growth factor, alpha	NM_001099691	4.1	young	higher in young	yes

Representative miRs were all hsa-miR-199a-5p. All were sites in the untranslated regions. Expression pairing relates to when the expression of miR199 and the corresponding target gene in the corresponding direction.

Table 10. Significant gene ontology terms from DAVID using DML as input data. Significant primary term results are with a Benjamini-Hochberg adjusted *p* value <0.05.

Term	Count	Fold Enrichment	Benjamini
GO:0048598~embryonic morphogenesis	12	16.52	0.00
GO:0048704~embryonic skeletal system morphogenesis	8	59.33	0.00
GO:0048705~skeletal system morphogenesis	9	33.97	0.00
GO:0007389~pattern specification process	11	17.42	0.00
GO:0048706~embryonic skeletal system development	8	43.92	0.00
GO:0009952~anterior/posterior pattern formation	9	27.18	0.00
GO:0001501~skeletal system development	11	14.58	0.00
GO:0043009~chordate embryonic development	11	14.05	0.00
GO:0009792~embryonic development ending in birth or egg hatching	11	13.92	0.00
GO:0003002~regionalisation	9	19.31	0.00
GO:0048562~embryonic organ morphogenesis	8	25.43	0.00
GO:0048568~embryonic organ development	8	19.66	0.00
GO:0002009~morphogenesis of an epithelium	6	25.11	0.00
GO:0048729~tissue morphogenesis	6	14.09	0.00
GO:0006355~regulation of transcription, DNA-dependent	14	3.34	0.00
GO:0051252~regulation of RNA metabolic process	14	3.26	0.00
GO:0060429~epithelium development	6	11.17	0.00
GO:0045449~regulation of transcription	16	2.60	0.01
GO:0051216~cartilage development	4	22.85	0.02
GO:0035113~embryonic appendage morphogenesis	4	19.44	0.02
GO:0030326~embryonic limb morphogenesis	4	19.44	0.02
GO:0035108~limb morphogenesis	4	17.08	0.03
GO:0035107~appendage morphogenesis	4	17.08	0.03
GO:0060173~limb development	4	16.42	0.04
GO:0048736~appendage development	4	16.42	0.04

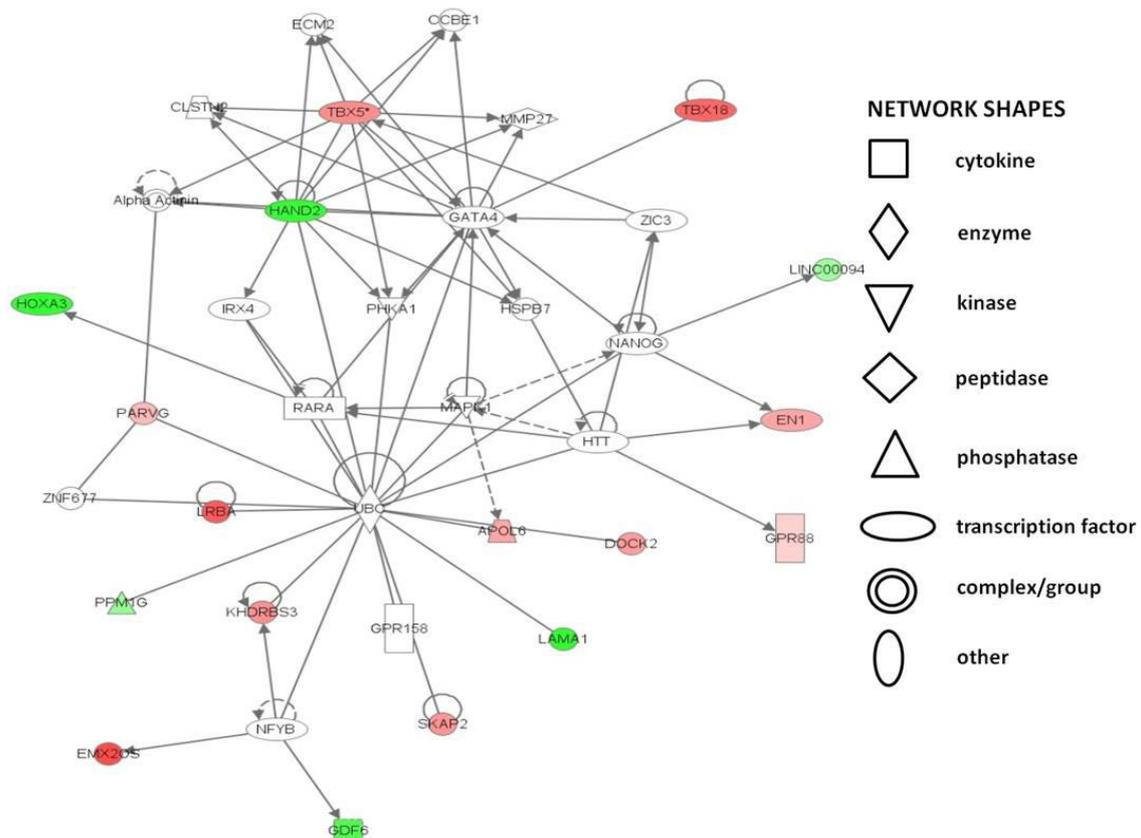


Fig. 6. Top-scoring network derived from the 187 DMLs in ageing. Ingenuity pathway analysis (IPA) identified ‘connective tissue development and function, embryonic development, organ development as the principal associated network functions with a score of 40. This is a graphical representation between molecules identified in our data in their respective networks. Green nodes; hypermethylated genes in older MSCs, red nodes; hypomethylated genes in older MSCs. Intensity of colour is related to higher fold-change. Key to the main features in the networks is shown.

Table 11. Summary of genes and DML correlating relationships.

Gene ID	DML (B value)	Methylation status in old	Log ₂ fold change (RNASeq)	Gene expression status in old	Location of methylation	Promoter or body	Data correlation
HAND2	-0.66	hypermethylated	-5.3	higher old	3'UTR	body	yes
HOXB4	0.62	hypomethylated	3.85	lower old	TSS1500	promoter	no
HOXB6	0.5	hypomethylated	4.12	lower old	TSS1500	promoter	no
HOXB7	0.71	hypomethylated	3.68	lower old	Body	body	yes
MKRN3	-0.33	hypermethylated	4.64	lower old	5'UTR	promoter	yes
TBX18	0.67	hypomethylated	-1.98	higher old	Body	body	no

The 3'UTR is encompassed in the gene body. The promoter is classified as the 5'UTR up to 1500 bp upstream of the start codon.

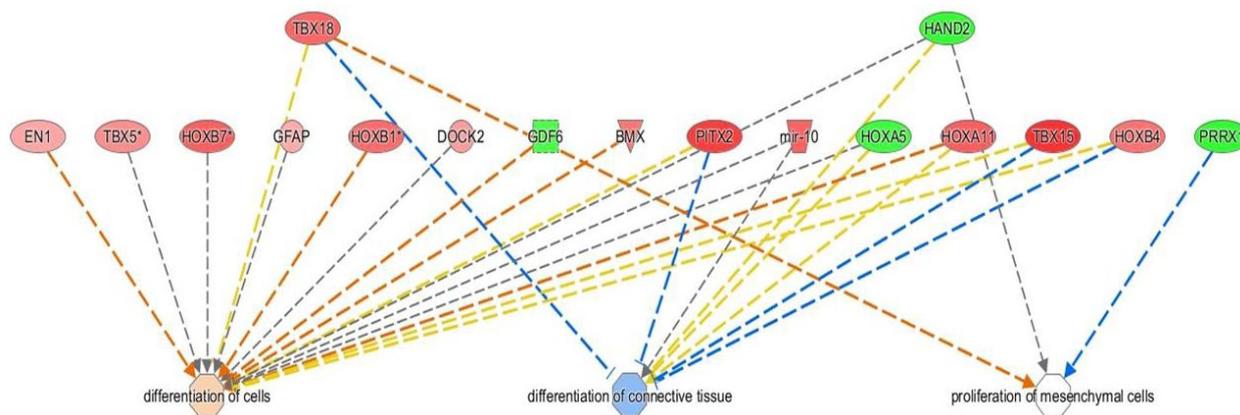


Fig. 7. Differentiation of cells and connective tissue and MSC proliferation were identified as significantly enriched at the DML level. Green nodes; hypermethylated genes in older MSCs, red nodes; hypomethylated genes in older MSCs. Intensity of colour is related to higher fold-change. Key to the main features in the networks is shown.

within CpG islands and 3 within promoters. 117 DMLs were hypomethylated and 70 hypermethylated in old. GO analysis of genes containing DMLs indicated enrichment in embryonic skeletal tissue morphogenesis pathways and the terms homeobox, developmental proteins, transcriptional regulation and tissue morphogenesis (Table 10). In IPA, the top scoring network was 'connective tissue development and function, embryonic development, organ development' (Fig. 6A). The top diseases and biological functions were similar to those identified in the RNASeq analysis (Table 8). There was also a significant enrichment for DML of genes involved in differentiation of cells and connective tissue and the proliferation of MSCs (Fig. 7). The top two networks were connective tissue development and function, embryonic development, organ development (score 40), skeletal and muscular system development and function, embryonic development, organismal development (score 37).

The methylation of gene promoters and/or enhancers is known to correlate with decreased gene expression contrastingly methylation within non-enhancer regions of the gene body correlates with increased gene expression (Jjing *et al.*, 2012). Therefore, we compared DEGs from RNASeq with location of DMLs. The DEGs HAND2, HOXB7 and TBX18 contained DMLs within the gene body. HOXB7 demonstrated decreased methylation in old samples with a decrease in gene expression in old; HAND2 demonstrated an increase in methylation and concurrent increase in gene expression. For the genes HOXB4,

HOXB6 and MKRN3 that contained DMLs within promoter regions between young and old samples only MKRN3 demonstrated the expected inverse relationship with gene expression (Table 11).

Proteomics

Protein Identification, label-free quantification

2,347 proteins were identified in combined samples from MSCs. Of these 777 were quantifiable using at least two unique peptides in Progenesis Q1 and 118 differentially expressed; 116 higher in old and 2 lower in old (Table 12A and B). Peptides clustered according to age with a principal component of 80%. Interestingly the highest differentially expressed protein was superoxide dismutase 2 (SOD2) at 20.8 higher in old MSCs (Fig. 8A). This was validated with western blotting.

Gene ontology (GO) and IPA analysis of differentially expressed proteins

GO analyses indicated enrichment in genes associated with mitochondria, metabolic processes and transcriptional regulation (Fig. 8B). IPA identified the top-scoring network was 'developmental disorders and metabolic disease'. The top diseases and biological functions included energy production, increased cell survival and metabolic diseases (Table 7). Age-related changes in oxidative state were identified by the canonical pathways mitochondrial dysfunction and oxidative phosphorylation. Interleukins-1 and -6 were identified as upstream regulators of apoptosis,

Table 12A. Differentially expressed proteins identified following database search using at least two unique peptides.

Accession	Anova (<i>p</i>)	q-Value	Max fold change	Description
F5GYZ5	0.08	0.04	20.80	Superoxide dismutase 2, mitochondrial
F5H2U8	0.03	0.03	5.50	High mobility group AT-hook 2
O14558	0.00	0.00	4.86	Heat shock protein beta-6
B4DJ51	0.01	0.03	4.51	Calmodulin 1 , isoform CRA_a
Q9BS40	0.01	0.03	4.51	Latexin
P27487	0.02	0.03	3.65	Dipeptidyl peptidase 4
F5H157	0.01	0.03	3.21	RAB35, member RAS oncogene family
Q6IBC4	0.04	0.03	3.19	NADH dehydrogenase Fe-S protein 6, 13kDa , isoform CRA_a
Q9Y383	0.03	0.03	3.13	Putative RNA-binding protein Luc7-like 2
P61604	0.02	0.03	3.09	10 kDa heat shock protein
Q16647	0.02	0.03	3.06	Prostacyclin synthase
E9PH29	0.05	0.03	2.99	Peroxiredoxin 3
Q6FHF5	0.00	0.02	2.98	Proliferating cell nuclear antigen
Q12765	0.01	0.02	2.94	Secernin-1
D3DUZ3	0.02	0.03	2.93	Interferon, gamma-inducible protein 16, isoform CRA_a
Q02952	0.02	0.03	2.88	A-kinase anchor protein 12
F5H801	0.03	0.03	2.87	Oxoglutarate (alpha-ketoglutarate) dehydrogenase
A2A274	0.01	0.02	2.86	Aconitase 2
B3KQK4	0.00	0.01	2.84	highly similar to Cathepsin L
Q59GA0	0.03	0.03	2.84	Thy-1 cell surface antigen variant
P08559	0.08	0.04	2.83	Pyruvate dehydrogenase E1 component subunit alpha, somatic form
A8K1Z2	0.05	0.03	2.78	highly similar to glycerol-3-phosphate dehydrogenase 2
B4DT77	0.00	0.02	2.78	Annexin
Q5TBU5	0.00	0.02	2.77	Adipose specific 2
Q9Y499	0.03	0.03	2.75	clone 23742
A8K5W7	0.01	0.02	2.73	highly similar to mitochondrial isoleucine tRNA synthetase
Q0QEY7	0.01	0.02	2.69	Succinate dehydrogenase complex subunit B
P00367	0.00	0.02	2.64	Glutamate dehydrogenase 1
P42224	0.02	0.03	2.63	Signal transducer and activator of transcription 1-alpha/beta
B4DE89	0.00	0.02	2.63	highly similar to Tripeptidyl-peptidase 1
P10809	0.01	0.02	2.62	60 kDa heat shock protein
B2R5M8	0.05	0.04	2.60	Isocitrate dehydrogenase
Q59EK1	0.05	0.03	2.58	Adducin 3 isoform a variant
B4DUL5	0.04	0.03	2.58	highly similar to Ubiquinol-cytochrome-c reductase complex coreprotein I
P30084	0.01	0.02	2.57	Enoyl-CoA hydratase,
B4DDM6	0.02	0.03	2.54	Budding uninhibited by benzimidazoles 3 homolog
B4DPJ2	0.03	0.03	2.54	Annexin
P24752	0.01	0.02	2.52	Acetyl-CoA acetyltransferase
Q9UQ80	0.00	0.02	2.52	Proliferation-associated protein 2G4
Q9Y5X1	0.00	0.02	2.52	Sorting nexin-9
P28066	0.01	0.03	2.50	Proteasome subunit alpha type-5
F8VY04	0.04	0.03	2.49	Adenylate kinase 2
Q2TNI1	0.02	0.03	2.48	Caveolin
E5KNY5	0.07	0.04	2.48	Leucine-rich PPR-motif containing
O95425	0.03	0.03	2.46	Supervillin
F5GZ16	0.04	0.03	2.44	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 2
P06703	0.05	0.03	2.44	Protein S100-A6
B4DUC1	0.06	0.04	2.43	Uncharacterized protein
Q969G5	0.02	0.03	2.42	Protein kinase C delta-binding protein
O60437	0.01	0.02	2.41	Periplakin
C9JGC2	0.02	0.03	2.40	Uncharacterized protein
Q63ZY3	0.00	0.01	2.40	KN motif and ankyrin repeat domain-containing protein 2
O00479	0.02	0.03	2.39	High mobility group nucleosome-binding domain-containing protein 4
E9PFT9	0.07	0.04	2.38	Transmembrane protein 119
F5H4I0	0.00	0.02	2.37	Dihydropolipoyl dehydrogenase
B4E2W0	0.04	0.03	2.37	Hydroxyacyl-CoA dehydrogenase B
E9PQW4	0.00	0.02	2.36	Mitogen-activated protein kinase 3
Q9Y6N5	0.03	0.03	2.36	Sulfide:quinone oxidoreductase,
P10606	0.03	0.03	2.35	Cytochrome c oxidase subunit 5B,
P62942	0.01	0.03	2.34	Peptidyl-prolyl cis-trans isomerase
Q6NZI2	0.01	0.02	2.34	Polymerase I and transcript release factor
P45880	0.03	0.03	2.33	Voltage-dependent anion-selective channel protein 2

Table 12B. Differentially expressed proteins identified following database search using at least two unique peptides.

Accession	Anova (p)	q-Value	Max fold change	Description
Q53HQ8	0.05	0.04	2.32	Granulin variant
Q86TY5	0.03	0.03	2.32	Full-length cDNA clone CS0DI041YE05 of Placenta
F8VU90	0.04	0.03	2.32	Peptidyl-prolyl cis-trans isomerase
P02545	0.01	0.03	2.29	Prelamin-A/C
P14854	0.00	0.02	2.28	Cytochrome c oxidase subunit 6B1
P61106	0.04	0.03	2.27	Ras-related protein Rab-14
B4DZ88	0.02	0.03	2.25	KTN1 protein
Q92973	0.03	0.03	2.24	Transportin-1
Q99436	0.00	0.02	2.22	Proteasome subunit beta type-7
B4E2K4	0.06	0.04	2.22	cDNA FLJ54576, highly similar to Aspartyl/asparaginyl beta-hydroxylase
P40939	0.00	0.02	2.22	Trifunctional enzyme subunit alpha
P13073	0.07	0.04	2.19	Cytochrome c oxidase subunit 4 isoform 1
Q53R19	0.08	0.04	2.19	Actin related protein 2/3 complex, subunit 2, 34kDa, isoform CRA_a
P15374	0.04	0.03	2.19	Ubiquitin carboxyl-terminal hydrolase isozyme L3
Q549N0	0.04	0.03	2.17	Cofilin 2 (Muscle), isoform CRA_a
H0YLU7	0.03	0.03	2.17	Electron-transfer-flavoprotein, alpha polypeptide
E9PFH8	0.01	0.03	2.17	Sjogren syndrome antigen B
E9PRJ8	0.05	0.03	2.16	CD81 molecule
Q96C19	0.00	0.02	2.16	EF-hand domain-containing protein D2
E7EW69	0.01	0.02	2.15	Septin 10
Q9Y2W1	0.05	0.04	2.15	Thyroid hormone receptor-associated protein 3
B4DRB7	0.09	0.04	2.14	highly similar to Proactivator polypeptide
Q03252	0.02	0.03	2.14	Lamin-B2
Q6FHN3	0.08	0.04	2.14	Nucleoside diphosphate kinase
B3KTN4	0.02	0.03	2.13	Citrate synthase
Q6FHP5	0.00	0.02	2.13	PHB protein
A6XMH5	0.06	0.04	2.12	Beta-2-microglobulin
P31949	0.08	0.04	2.12	Protein S100-A11
Q9H3P7	0.06	0.04	2.12	Golgi resident protein GCP60
Q76LA1	0.06	0.04	2.11	CSTB protein
Q07812	0.04	0.03	2.11	Apoptosis regulator BAX
Q15075	0.01	0.03	2.10	Early endosome antigen 1
Q9Y230	0.02	0.03	2.10	RuvB-like 2
E9PKT8	0.02	0.03	2.09	Nucleosome assembly protein 1-like
P04083	0.02	0.03	2.07	Annexin A1
P35908	0.00	0.02	2.07	Keratin, type II cytoskeletal 2 epidermal
Q5TB52	0.03	0.03	2.07	3'-phosphoadenosine 5'-phosphosulfate synthase 2
Q9NX63	0.06	0.04	2.07	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3,
A8K7A4	0.01	0.03	2.07	highly similar to methionine adenosyltransferase II, beta, transcript variant 2
B2R984	0.04	0.03	2.06	highly similar to histone 1
P25398	0.06	0.04	2.06	40S ribosomal protein S12
Q6IBA2	0.00	0.02	2.06	PC4 protein
P04632	0.08	0.04	2.06	Calpain small subunit 1
P06576	0.00	0.02	2.05	ATP synthase subunit beta
F8W118	0.03	0.03	2.04	Nucleosome assembly protein 1-like
Q53HG3	0.00	0.02	2.04	Peroxisomal D3,D2-enoyl-CoA isomerase isoform 1 variant
Q9Y6M1	0.06	0.04	2.04	Insulin-like growth factor 2 mRNA-binding protein 2
P08670	0.00	0.02	2.03	Vimentin
F5GY50	0.08	0.04	2.03	Prostaglandin reductase 1
Q9Y6C9	0.07	0.04	2.03	Mitochondrial carrier homolog 2
Q549C5	0.03	0.03	2.02	HCG2010808, isoform CRA_a
B4DFP5	0.05	0.03	2.01	highly similar to Hepatocyte growth factor-regulated tyrosine kinase substrate
A8K766	0.04	0.03	2.01	highly similar to electron-transfer-flavoprotein, beta polypeptide
G3V279	0.01	0.02	2.01	Enhancer of rudimentary homolog
Q9UBG0	0.06	0.04	-2.00	C-type mannose receptor 2
P52943	0.03	0.03	-2.06	Cysteine-rich protein 2

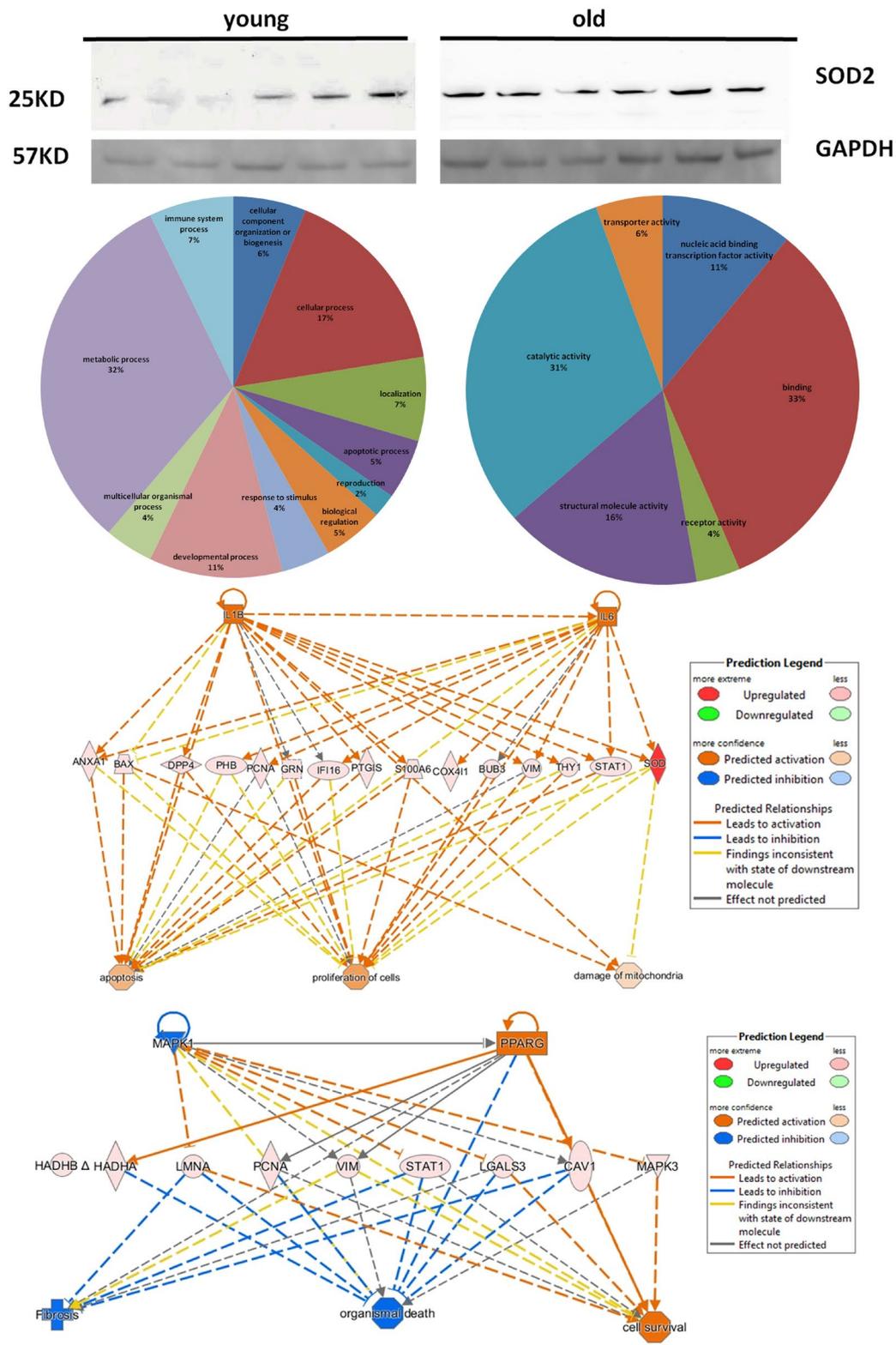
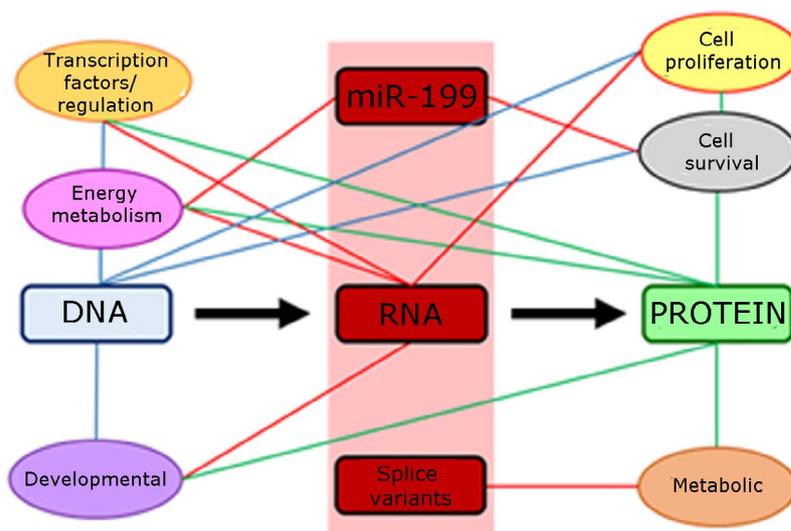


Fig. 8. Proteomic changes associated with MSC ageing; (A) Increased SOD2 protein expression as determined by western blotting of young and old MSCs. Western blot analysis of SOD2 relative to GAPDH demonstrated an increase in SOD2 with age confirming label-free quantification results. (B) Pie chart representing the occurrence of major ontological annotations associated with differentially expressed proteins in ageing MSCs for i.) biological processes and ii) molecular function categories of the Panther Classification System. (C) Interleukins-1 and -6 were identified as upstream regulators of a set of differentially expressed proteins leading to activation of apoptosis, cell proliferation and mitochondrial damage. (D) One of the top regulator effects networks, which explains how predicted upstream regulators cause the observed changes in protein expression explaining phenotypic or functional outcomes downstream. These causal hypotheses take the form of directionally coherent networks formed from the merger of upstream regulator networks with downstream effects networks. Mitogen activated protein kinase 1 (MAPK1) and peroxisome proliferator-activated receptor gamma (PPARG) were identified as upstream regulators of downstream inhibition of fibrosis and organismal death but activation of cell survival through a set of nine differentially expressed proteins.

Fig. 9. Summary diagram of MSC ageing using multi-level analysis; DNA (DNA methylation), RNA (RNASeq), MicroRNA 199 (miR199) and protein level. Ontology terms in the bubbles are significantly enriched terms derived following bioinformatics using IPA/DAVID/PANTHER. Connections between each level and general ontology term are red for RNA, blue for DML and green for protein.



cell proliferation and mitochondrial damage activation (Fig. 8C). Top regulatory effects networks included cell survival and reduced fibrosis in ageing (Fig. 8D).

Fig. 9 summaries major findings from the study.

Discussion

Major sources of cells for musculoskeletal tissue regeneration (both *in vivo* and *in vitro*) are adult MSCs (Natsu *et al.*, 2004; Shirley *et al.*, 2005). Recent studies have revealed that MSC cell biology is influenced by biological ageing (Campagnoli *et al.*, 2001) with MSCs from older donors having a shorter life-span, and a reduced proliferative and differentiation potential (Sethe *et al.*, 2006). Thus age-related changes to their functionality may contribute to musculoskeletal ageing; a major factor in many musculoskeletal diseases. Our study aimed to increase our understanding of age-related changes in the systems of biological components of MSCs at the RNA, epigenetic and protein level. This has identified common networks across scales enabling a ‘big picture view’ of MSC ageing.

Previous investigations on ageing MSCs have been undertaken on a variety of subsets of stem cells with variable potentials and characteristics. The combinatorial effects of genomic, epigenetic and proteomic changes lead to a decline in cellular function, which in turn contributes to tissue dysfunction and ageing. Here, we used young and old MSCs that had been tested for ‘stemness’ at a number of levels, including analysis of a comprehensive panel of markers and differentiation assays, performed for each lot prior to inclusion. Furthermore, as the site of extraction can affect cell behaviour (Bork *et al.*, 2010; Matsubara *et al.*, 2004), we used MSCs derived from alveolar bone. As low oxygen tension improves MSC vitality and metabolic state in culture (Estrada *et al.*, 2012; Knuth *et al.*, 2013), all cell culture were undertaken in 5 % oxygen tension. Finally, as challenges arise in distinguishing ageing *in vivo* and *in vitro*, we used the same passage number for

all donors from identical seeding densities and confluence that can influence cellular ageing (Wagner *et al.*, 2010). It is worth noting that whilst one study found no difference between passage 2 and 10 in proliferation capacity and differentiation potential in MSCs (Bruder *et al.*, 1997), others demonstrated a reduction in colony forming units from passage 4 (Peter *et al.*, 2012).

There is a detrimental effect on MSC clonogenicity, proliferation capacity and differentiation potential when expansion is undertaken. Our studies were undertaken on ageing MSCs in *in vitro* conditions, which will produce different results from MSCs within an *in vivo* physiological environment. However, in order to obtain enough cells to undertake the systems biology approach used in our study, this was required. It would be beneficial to repeat the study on freshly isolated MSCs, in order to predict their ageing before expansion. As insufficient cell numbers would be problematic, this could be undertaken in a more targeted workflow in the future, but was beyond the scope of this study. Furthermore, although we have identified molecular changes at a number of levels, a validation of the system is required. As ageing can affect MSC differentiation potential in further work we have undertaken transcriptomic, proteomic and epigenetic analysis of changes in chondrogenic, osteogenic and tenogenic constructs produced from the young and old MSCs utilised in this study.

One of the main challenges of this study was the integration of the different types of data. We used network analysis and annotation of the pathways identified. However, we were unable to integrate key environmental factors such as nutrition and other lifestyle factors; affecting the pathways due to limited information on the donors. These factors can change molecular measurements. Other potential limitations include small sample size and significance threshold filtering, which can affect the subsequent pathway/network analysis. Additional parameters, such as histone modification, would benefit the study but this was beyond the financial scope of the study. Dynamic histone methylation can contribute to the ageing process through influencing DNA repair and transcriptional

regulation of ageing processes (reviewed by: McCauley and Dang, 2014).

Our study found weak age-related effects on expression at the miR level. Despite this, an added benefit of using smallSeq was that we identified for the first time, in an unbiased manner, a catalogue of miRs and small nucleolar RNAs expressed in MSCs and their relative expression levels which will be a useful tool in future studies. Results contrast to a recent study in monkeys that identified 40 differentially expressed miRs in ageing by microarray (Yu *et al.*, 2011). This could be due to species, platform and workflow differences. The differential expression of miR-199 has been previously reported in this study and human MSC ageing (Alt *et al.*, 2012). Here we demonstrate the age-related increase in miR-199b-5p expression. Analysis of miR predicted target genes and networks demonstrates that miR-199b-5p may be involved in regulating SIRT1, TGF α and PODXL. The expression of other predicted targets: MAB21L and EPHA7, was negatively correlated with that of miR-199b-5p during ageing, suggesting that these may indeed be targets. Interestingly, TGF α has been shown to mediate secretion of VEGFA by MSCs, positively contributing to processes like wound healing and injury response (Wang *et al.*, 2008). Moreover, SIRT1 appears to regulate the differentiation of MSCs (Simic *et al.*, 2013) and their long-term growth (Yuan *et al.*, 2012). Together, these data support the hypothesis that age-related upregulation of miR-199b-5p expression in MSCs may contribute to age-related deterioration of MSC function through regulating key genes.

In transcriptomic studies, it is difficult to justify each individual change at the gene level. Generally, it is more informative to look at the changes as a whole. Therefore, we used gene ontology and network analysis tools. However, there are a few interesting findings for the some individual genes. We investigated the levels of gene expression previously reported to be associated with cellular senescence: p16, p21, Rb, p53 and p85. Previous studies have found increased levels of p21 and p53 in ageing MSCs (Stolzing *et al.*, 2008; Zhou *et al.*, 2008), although none of these were significantly changed in ageing in this study. This may be due to differences in individual studies or the variability in the expression between individuals, especially apparent in MSCs from older individuals. ALX1 has the most reduced expression change in old. This gene is important in skeletal development and has previously demonstrated to have increased expression in old tendon (Peffers *et al.*, 2015) and in bone-marrow-derived MSCs compared to those derived from umbilical cord (Hsieh *et al.*, 2010). The most highly DEGs in the old were CPC5 (involved in cell division and growth regulation), ADH1B (an enzyme able to metabolise many substrates) and OGN (an osteoblastic differentiation gene). We found age-related changes in gene profiles included differences in cell proliferation, signalling, function and maintenance. These changes suggest an age-related loss in MSCs ability to respond to biological cues. In addition, there was a dysregulation of transcription factors at the transcript and AS level in ageing with a reduction in expression of ALX1, PITX2, HOXB6, HOXB7, IRF6 and an increased expression of TBX18 and FOXP2 in ageing. Transcription

factors, including HOX genes have been implicated in ageing of tissues (Stelnicki *et al.*, 1998) including tendon (Peffers *et al.*, 2015), and heart (Bodyak *et al.*, 2002). Interestingly, HOX genes are an intrinsic property of MSCs and play a role in cellular identity, are required for tissue appropriate regeneration (Ackema and Charite, 2008; Sagi *et al.*, 2012) and may be involved in the timing of ageing (Bork *et al.*, 2010). Loss of cellular identity, through alterations in HOX gene expression (Trivedi *et al.*, 2011), results in aberrant cellular homeostasis and age-related functional deterioration. Conversely, HOX-mediated transcriptional memory limits stem cell-mediated tissue regeneration (Chang, 2009). Potentially, the altered expression in ageing may affect the differentiation potential of older MSCs particularly in relation to osteogenesis (Shi *et al.*, 1999). This has particular relevance to both tissue engineering and musculoskeletal repair in ageing.

There is a disruption of splicing with ageing (Harries *et al.*, 2011). We have previously observed an age-associated disruption to the balance of alternatively expressed isoforms for selected genes in tendon ageing (Peffers *et al.*, 2015). Alterations in the splicing patterns of the mRNA transcripts contribute to some age-related diseases, including Hutchison Gilford Progeria syndrome (Eriksson *et al.*, 2003) and Alzheimer's (Chen *et al.*, 2010). Thus, whilst most MSCs may be expressing genes at comparable levels in MSC ageing, there are also differences in the relative balance of splice products or increase in the occurrence of atypical transcripts. Interestingly, ontology analysis identified AS genes altered in ageing primarily affected metabolic processes similar to that in ageing brain (Tollervey *et al.*, 2011).

We identified two differentially regulated LNCs in MSC ageing; AC004870.4 and RP11-670E13.5. LNCs (Long Noncoding RNA) are a diverse group of non-protein coding RNA molecules longer than 200 nucleotides whose expression is developmentally regulated and both cell and tissue specific. The function of the majority of LNCs is unknown. However, they have been implicated in many diseases (Ponting *et al.*, 2009). Studies have revealed that LNCs have significant roles in biological processes, including, transcription, imprinting, chromatin modification (Kung *et al.*, 2013). Furthermore, they are important in controlling tissue differentiation (Hu *et al.*, 2012). Previously, a study has identified differentially expressed LNCs in bone marrow *versus* periodontal ligament derived MSCs from young donors (Dong *et al.*, 2014). The expression of RP11-670E13.5 has been previously identified in bone marrow (Uhlen *et al.*, 2015). However, there is no expression data for RP11-670E13.5, also known as lnc-DGKE-2, in MSCs. There may be a significant role for LNCs in phenotypic regulators of MSC differentiation and they could aid the establishment of lineage specification through independent mechanisms (Tye *et al.*, 2015). LNCs represent potential biomarkers of disease and novel therapeutic targets for ageing, but much work is needed to elucidate their roles in MSC ageing.

Epigenetic processes have been implicated in ageing and age-related musculoskeletal diseases such as osteoarthritis (Rushton *et al.*, 2014). Here, we analysed DNA methylation; the most characterised epigenetic

mechanism, which is primarily involved in gene regulation. Alterations in DML within ageing tissues (Calvanese *et al.*, 2009; Zykovich *et al.*, 2014) and cells (Fernandez *et al.*, 2012), as well as aged MSCs (Bork *et al.*, 2010), and following successive *in vitro* culture (reviewed (Bentivegna *et al.*, 2013)). Our study provides the most comprehensive DNA methylation analysis of ageing MSCs, to date, using the HumanMethylation 450 Beadchip array. This has advantages over the 27K array used previously (Bork *et al.*, 2010) as the 27K array contains probes within promoters only, whereas the 450K array includes probes within multiple genomic features as well as less dense CpG regions. Most methylation changes occurring in disease occur in the less dense CpG regions thus additional probes on the 450K array is advantageous. However, there was significant concordance with the results from this study. There was a global hypomethylation in ageing, as described previously (Calvanese *et al.*, 2009). Methylation changes in the study were associated with HOX genes, similar to other studies (Bork *et al.*, 2010), the expression of which is tightly regulated temporally during vertebrate development. An association between HOX genes and longevity has been previously proposed (Venkataraman and Futerman, 2002). HOXB4, HOXB6 and HOXB7 were also DEGs in the RNASeq study. However, only for HOXB7 did DEG correlate (reduced in old) with DNA hypomethylation within the gene body. Despite this, the dysregulation of HOX genes at the mRNA and epigenetic level consolidate their role in MSC ageing.

We identified many DMLs in genes involved in differentiation of cells and connective tissue as well as the proliferation of MSCs. The latter was also identified at the transcriptomic level. Methylation changes at specific CpG sites can affect MSC differentiation (Noer *et al.*, 2007; Zimmermann *et al.*, 2008), whilst alterations in MSC potential have been previously noted (Khan and Marsh, 2009; Sethe *et al.*, 2006).

We did not observe a general correlation of DNA methylation with differentially expressed genes. This could be due to the differing sensitivities of the two platforms. It is commonly accepted that DNA methylation regulates gene expression. However, the effect of DNA methylation on gene expression is dependent upon the site of methylation. DNA methylation within the gene body or 3' untranslated region (UTR) of a gene generally positively correlates with gene expression, whereas methylation in the promoter is inversely correlated with gene expression. Others have hypothesised that DMLs at specific CpG sites does not inevitably correlate with gene expression changes. Significant gene expression fold-changes can only be anticipated if the CpG sites are principally methylated and that selective modification of certain promoter regions activate or inactivate binding of transcription regulators (Bork *et al.*, 2010). Further suggestions for the lack of correlation could be that a minimal number of altered DMLs are required in order to correlate with gene expression changes (Zykovich *et al.*, 2014) or that methylation changes may reflect historical gene expression changes.

Interestingly, a recent report on reprogramming of old MSCs into induced pluripotent stem cells (iPS) and

their subsequent re-differentiation into MSCs, indicated a reduction in some age-related MSCs features. Epigenetic recapitulation was maintained, with the age-affected DNA methylation accumulation remaining reset in iPS-MSCs (Frobel *et al.*, 2014).

The maintenance of the MSC proteome is imperative for assuring stem cell identity. Label-free quantification provided the most comprehensive proteomic analysis of ageing MSCs phenotype to date. Whilst studies imply that alterations in mRNA levels are a significant factor in the determination of protein changes, gene expression – as reflected by mRNA – is correlated with protein 40 % of the time (Web ref. 5). Protein expression is subject to a diversity of regulations of transcription and translation, including mRNA and protein half-life, translation rate, post-translational modifications and subcellular localisation. Furthermore, differences in sensitivity between platforms can also account for differences. In this study, there was minimal overlap at the RNA and protein at the individual molecule level but there were similarities at the functional level for many changes; metabolism, transcriptional regulation, energy metabolism, development, cell proliferation and cell survival in MSC ageing.

There was an age-related increase in proteins involved in anti-oxidant changes, energy, nucleic acid and fatty-acid metabolism, musculoskeletal abnormalities, increased cell survival and reduced cell death. A number of inflammatory cytokines were upregulated in ageing (TGF α at the gene level) or identified as upstream regulators at the gene (IL6, TNF) and protein level (IL4, IL6, IL1 β). Together, these findings, indicate that MSCs undergo inflamm-ageing (reviewed by: Lepperdinger, 2011). Furthermore, similar to others, mitochondrial ageing (Kasper *et al.*, 2009) was demonstrated.

Mitochondrial dysfunction is a hallmark of the ageing phenotype, and alterations in redox balance may, in part, explain the observed reduction in cellular function associated with age (Chan, 2006). The differences in mitochondrial protein expression could be indicative of potentially dysfunctional mitochondrial alterations in ageing. A key finding of this study was a large increase in mitochondrial superoxide dismutase (SOD2) protein expression in ageing, which is in contrast to others (Duscher *et al.*, 2014; Stolzing *et al.*, 2008). The mitochondria are a primary source of reactive oxygen species (ROS) and SOD2 represents the principle antioxidant molecule that scavenges mitochondrial superoxide, thus of key importance in maintaining redox balance (Miao and St Clair, 2009). An increase in SOD2 protein was associated with an increase in mitochondrial peroxiredoxin 3, a key scavenger of H₂O₂, which suggests that in this study, the antioxidant capacity of ageing cells increases similar to rat ageing cartilage (Fu *et al.*, 2013), but in contrast to other studies (Duscher *et al.*, 2014; Stolzing *et al.*, 2008). This observation may be a compensatory response by the cell to restore oxidative balance, given that our data also demonstrate increased protein expression of many mitochondrial proteins. This includes members of the electron transport chain; crucial for mitochondrial homeostasis. Whether or not the increase in mitochondrial

antioxidant proteins correlates with increased activity is beyond the scope of this study, but represents a future avenue of research for our group. Alternatively, the dramatic increase in SOD2 levels may be evidence of mitochondrial dysfunction, potentially exposing the cell to large amounts of H₂O₂, which could enhance oxidative stress and cause aberrant H₂O₂-mediated redox signalling (Veal *et al.*, 2007).

There was a reduction in proteasome subunits PSMB7, PSMA5 and PSMD2 in ageing MSCs. This has previously been associated with reduced proteasome activity in ageing fibroblasts (Hwang *et al.*, 2007). Furthermore, PSMB7 silencing induces autophagy in cardiomyocytes (Kyrychenko *et al.*, 2014). Cellular damage and dysfunction can be caused by defects in proteostasis, resulting in the accumulation of damaged proteins (Bucciantini *et al.*, 2002). The age-dependant deficits evident here in the proteasome subunits may affect both protein quality and autophagy, contributing to MSC ageing.

Conclusion

Our data suggest that age-related changes in MSCs result in impaired therapeutic potential of aged progenitor cells. These results have significant implications for therapeutic cell source decisions (autologous or allogeneic) revealing the necessity of approaches to improve functionality of ageing MSCs.

Acknowledgements

MJP was supported by a Wellcome Trust Veterinary Integrated Research Fellowship (WT088557MA) 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). We thank Sophie Sutton for her assistance with qRT-PCR.

References

Ackema KB, Charite J (2008) Mesenchymal stem cells from different organs are characterized by distinct topographic Hox codes. *Stem Cells Dev* **17**: 979-991.

Alt EU, Senst C, Murthy SN, Slakey DP, Dupin CL, Chaffin AE, Kadowitz PJ, Izadpanah R (2012) Aging alters tissue resident mesenchymal stem cell properties. *Stem Cell Res* **8**: 215-225.

Baxter MA, Wynn RF, Jowitt SN, Wraith JE, Fairbairn LJ, Bellantuono I (2004) Study of telomere length reveals rapid aging of human marrow stromal cells following *in vitro* expansion. *Stem Cells* **22**: 675-682.

Beausejour C (2007) Bone marrow-derived cells: the influence of aging and cellular senescence. *Handb Exp Pharmacol* **180**: 67-88.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Methodology* **57**: 289-300.

Bentivegna A, Miloso M, Riva G, Foudah D, Butta V, Dalpra L, Tredici G (2013) DNA methylation changes during *in vitro* propagation of human mesenchymal stem cells: implications for their genomic stability? *Stem Cells Int* **2013**: 192425.

Bodyak N, Kang PM, Hiromura M, Suljoadikusumo I, Horikoshi N, Khrapko K, Usheva A (2002) Gene expression profiling of the aging mouse cardiac myocytes. *Nucleic Acids Res* **30**: 3788-3794.

Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, Wagner W (2010) DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell* **9**: 54-63.

Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K (2003) EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *The EMBO J* **22**: 5323-5335.

Bruder SP, Jaiswal N, Haynesworth SE (1997) Growth kinetics, self-renewal, and the osteogenic potential of purified human mesenchymal stem cells during extensive subcultivation and following cryopreservation. *J Cell Biochem* **64**: 278-294.

Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, Stefani M (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**: 507-511.

Buitrago DH, Patnaik SK, Kadota K, Kannisto E, Jones DR, Adusumilli PS (2015) Small RNA sequencing for profiling microRNAs in long-term preserved formalin-fixed and paraffin-embedded non-small cell lung cancer tumor specimens. *PLoS One* **10**: e0121521.

Calvanese V, Lara E, Kahn A, Fraga MF (2009) The role of epigenetics in aging and age-related diseases. *Ageing Res Rev* **8**: 268-276.

Camafeita E, Lamas JR, Calvo E, Tornero-Esteban P, Lopez JA, Fernandez-Gutierrez B (2014) Selected reaction monitoring assays in mesenchymal stem cells from osteoarthritis patients. *Clin Proteomics* **11**: 33.

Campagnoli C, Roberts IA, Kumar S, Bennett PR, Bellantuono I, Fisk NM (2001) Identification of mesenchymal stem/progenitor cells in human first-trimester fetal blood, liver, and bone marrow. *Blood* **98**: 2396-2402.

Chan DC (2006) Mitochondria: dynamic organelles in disease, aging, and development. *Cell* **125**: 1241-1252.

Chang HY (2009) Anatomic demarcation of cells: genes to patterns. *Science* **326**: 1206-1207.

Chen S, Townsend K, Goldberg TE, Davies P, Conejero-Goldberg C (2010) MAPT isoforms: differential transcriptional profiles related to 3R and 4R splice variants. *J Alzheimers Dis* **22**: 1313-1329.

Choi MR, In YH, Park J, Park T, Jung KH, Chai JC, Chung MK, Lee YS, Chai YG (2012) Genome-scale DNA methylation pattern profiling of human bone marrow mesenchymal stem cells in long-term culture. *Exp Mol Med* **44**: 503-512.

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**: 156-159.

- Conklin JF, Baker J, Sage J (2012) The RB family is required for the self-renewal and survival of human embryonic stem cells. *Nat Commun* **3**: 1244.
- Dong R, Du J, Wang L, Wang J, Ding G, Wang S, Fan Z (2014) Comparison of long noncoding RNA and mRNA expression profiles in mesenchymal stem cells derived from human periodontal ligament and bone marrow. *Biomed Res Int* **2014**: 317853.
- Duscher D, Rennert RC, Januszyk M, Anghel E, Maan ZN, Whittam AJ, Perez MG, Kosaraju R, Hu MS, Walmsley GG, Atashroo D, Khong S, Butte AJ, Gurtner GC (2014) Aging disrupts cell subpopulation dynamics and diminishes the function of mesenchymal stem cells. *Sci Rep* **4**: 7144.
- Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, Dutra A, Pak E, Durkin S, Csoka AB, Boehnke M, Glover TW, Collins FS (2003) Recurrent *de novo* point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* **423**: 293-298.
- Estrada JC, Albo C, Benguria A, Dopazo A, Lopez-Romero P, Carrera-Quintanar L, Roche E, Clemente EP, Enriquez JA, Bernad A, Samper E (2012) Culture of human mesenchymal stem cells at low oxygen tension improves growth and genetic stability by activating glycolysis. *Cell death and differentiation* **19**: 743-755.
- Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, Yamamoto H, Hidalgo M, Tan AC, Galm O, Ferrer I, Sanchez-Cespedes M, Villanueva A, Carmona J, Sanchez-Mut JV, Berdasco M, Moreno V, Capella G, Monk D, Ballestar E, Ropero S, Martinez R, Sanchez-Carbayo M, Prosper F, Agirre X, Fraga MF, Grana O, Perez-Jurado L, Mora J, Puig S, Prat J, Badimon L, Puca AA, Meltzer SJ, Lengauer T, Bridgewater J, Bock C, Esteller M (2012) A DNA methylation fingerprint of 1628 human samples. *Genome Res* **22**: 407-419.
- Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**: 808-815.
- Frobel J, Hemeda H, Lenz M, Abagnale G, Jousen S, Denecke B, Saric T, Zenke M, Wagner W (2014) Epigenetic rejuvenation of mesenchymal stromal cells derived from induced pluripotent stem cells. *Stem Cell Reports* **3**: 414-422.
- Fu Y, Kinter M, Griffin T (2013) Increase in mitochondrial antioxidant Sod2 in response to aging-associated oxidative stress and osteoarthritis in rat articular cartilage. *Proceedings of the ORS Annual Meeting*. Abstract 61.
- Hacia JG, Lee CC, Jimenez DF, Karaman MW, Ho VV, Siegmund KD, Tarantal AF (2008) Age-related gene expression profiles of rhesus monkey bone marrow-derived mesenchymal stem cells. *J Cell Biochem* **103**: 1198-1210.
- Harries LW, Hernandez D, Henley W, Wood AR, Holly AC, Bradley-Smith RM, Yaghootkar H, Dutta A, Murray A, Frayling TM, Guralnik JM, Bandinelli S, Singleton A, Ferrucci L, Melzer D (2011) Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**: 868-878.
- Hsieh JY, Fu YS, Chang SJ, Tsuang YH, Wang HW (2010) Functional module analysis reveals differential osteogenic and stemness potentials in human mesenchymal stem cells from bone marrow and Wharton's jelly of umbilical cord. *Stem Cells Dev* **19**: 1895-1910.
- Hu Y, Cai K, Luo Z, Zhang Y, Li L, Lai M, Hou Y, Huang Y, Li J, Ding X, Zhang B, Sung KL (2012) Regulation of the differentiation of mesenchymal stem cells *in vitro* and osteogenesis *in vivo* by microenvironmental modification of titanium alloy surfaces. *Biomaterials* **33**: 3515-3528.
- Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.
- Huang SC, Wu TC, Yu HC, Chen MR, Liu CM, Chiang WS, Lin KM (2010) Mechanical strain modulates age-related changes in the proliferation and differentiation of mouse adipose-derived stromal cells. *BMC Cell Biol* **11**: 18.
- Hwang JS, Chang I, Kim S (2007) Age-associated decrease in proteasome content and activities in human dermal fibroblasts: restoration of normal level of proteasome subunits reduces aging markers in fibroblasts from elderly persons. *J Gerontol A Biol Sci Med Sci* **62**: 490-499.
- Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK (2012) On the presence and role of human gene-body DNA methylation. *Oncotarget* **3**: 462-474.
- Kasper G, Mao L, Geissler S, Draycheva A, Trippens J, Kuhnisch J, Tschirschmann M, Kaspar K, Perka C, Duda GN, Klose J (2009) Insights into mesenchymal stem cell aging: involvement of antioxidant defense and actin cytoskeleton. *Stem Cells* **27**: 1288-1297.
- Khan WS, Marsh DR (2009) The effects of ageing on the isolation, proliferation and differentiation of mesenchymal stem cells. *Current Research Journal of Biological Sciences* **1**: 1-6.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: 36.
- Knuth CA, Clark ME, Meeson AP, Khan SK, Downen DJ, Deehan DJ, Oldershaw RA (2013) Low oxygen tension is critical for the culture of human mesenchymal stem cells with strong osteogenic potential from haemarthrosis fluid. *Stem Cell Rev* **9**: 599-608.
- Kolf CM, Cho E, Tuan RS (2007) Mesenchymal stromal cells. *Biology of adult mesenchymal stem cells: regulation of niche, self-renewal and differentiation*. *Arthritis Res Ther* **9**: 204.
- Kretlow JD, Jin YQ, Liu W, Zhang WJ, Hong TH, Zhou G, Baggett LS, Mikos AG, Cao Y (2008) Donor age and cell passage affects differentiation potential of murine bone marrow-derived stem cells. *BMC Cell Biol* **9**: 60.
- Kung JT, Colognori D, Lee JT (2013) Long noncoding RNAs: past, present, and future. *Genetics* **193**: 651-669.
- Kyrychenko VO, Nagibin VS, Tumanovska LV, Pashevin DO, Gurianova VL, Moibenko AA, Dosenko VE, Klionsky DJ (2014) Knockdown of PSMB7 induces autophagy in cardiomyocyte cultures: possible role in endoplasmic reticulum stress. *Pathobiology* **81**: 8-14.

- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.
- Lepperdinger G (2011) Inflammation and mesenchymal stem cell aging. *Curr Opin Immunol* **23**: 518-524.
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.
- Madeira A, da Silva CL, dos Santos F, Camafeita E, Cabral JM, Sa-Correia I (2012) Human mesenchymal stem cell expression program upon extended *ex-vivo* cultivation, as revealed by 2-DE-based quantitative proteomics. *PLoS One* **7**: e43523.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**: 1509-1517.
- Martin I, Jakob M, Schäfer D, Dick W, Spagnoli G, Heberer M (2001) Quantitative analysis of gene expression in human articular cartilage from normal and osteoarthritic joints. *Osteoarthritis Cartilage* **9**: 112-118.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10-12.
- Matsubara T, Tsutsumi S, Pan H, Hiraoka H, Oda R, Nishimura M, Kawaguchi H, Nakamura K, Kato Y (2004) A new technique to expand human mesenchymal stem cells using basement membrane extracellular matrix. *Biochem Biophys Res Commun* **313**: 503-508.
- McCauley BS, Dang W (2014) Histone methylation and aging: lessons learned from model systems. *Biochim Biophys Acta* **1839**: 1454-1462.
- Mi H, Muruganujan A, Thomas PD (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41**: 377-386.
- Miao L, St Clair DK (2009) Regulation of superoxide dismutase genes: implications in disease. *Free Radic Biol Med* **47**: 344-356.
- Natsu K, Ochi M, Mochizuki Y, Hachisuka H, Yanada S, Yasunaga Y (2004) Allogeneic bone marrow-derived mesenchymal stromal cells promote the regeneration of injured skeletal muscle without differentiation into myofibers. *Tissue Eng* **10**: 1093-1112.
- Noer A, Boquest AC, Collas P (2007) Dynamics of adipogenic promoter DNA methylation during clonal culture of human adipose stem cells to senescence. *BMC Cell Biol* **8**: 18.
- Oh J, Lee YD, Wagers AJ (2014) Stem cell aging: mechanisms, regulators and therapeutic opportunities. *Nat Med* **20**: 870-880.
- Pandey AC, Semon JA, Kaushal D, O'Sullivan RP, Glowacki J, Gimble JM, Bunnell BA (2011) MicroRNA profiling reveals age dependent differential expression of nuclear factor kB and mitogen activated protein kinase in adipose and bone marrow-derived human mesenchymal stem cells. *Stem Cell Res Ther* **2**: 49.
- Peffers M, Liu X, Clegg P (2013a) Transcriptomic signatures in cartilage ageing. *Arthritis Res Ther* **15**: 98.
- Peffers MJ, Beynon RJ, Clegg PD (2013b) Absolute quantification of selected proteins in the human osteoarthritic secretome. *Int J Mol Sci* **14**: 20658-20681.
- Peffers MJ, Fang Y, Cheung K, Wei TK, Clegg PD, Birch HL (2015) Transcriptome analysis of ageing in uninjured human Achilles tendon. *Arthritis Res Ther* **17**: 33.
- Peffers MJ, Milner PI, Tew SR, Clegg PD (2010) Regulation of SOX9 in normal and osteoarthritic equine articular chondrocytes by hyperosmotic loading. *Osteoarthritis Cartilage* **18**: 1502-1508.
- Peter S, Evans C, Ow SY, Scutt AM, Wright PC, Biggs CA (2012) Proteomic analysis of the impact of static culturing on the expansion of rat bone marrow mesenchymal stem cells. *Biotechnol Lett* **34**: 1589-1596.
- Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**: 293.
- Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* **136**: 629-641.
- Rando TA (2006) Stem cells, ageing and the quest for immortality. *Nature* **441**: 1080-1086.
- Redaelli S, Bentivegna A, Foudah D, Miloso M, Redondo J, Riva G, Baronchelli S, Dalpra L, Tredici G (2012) From cytogenomic to epigenomic profiles: monitoring the biologic behavior of *in vitro* cultured human bone marrow mesenchymal stem cells. *Stem Cell Res Ther* **3**: 47.
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
- Rollin R, Marco F, Camafeita E, Calvo E, Lopez-Duran L, Jover JA, Lopez JA, Fernandez-Gutierrez B (2008) Differential proteome of bone marrow mesenchymal stem cells from osteoarthritis patients. *Osteoarthritis Cartilage* **16**: 929-935.
- Rushton MD, Reynard LN, Barter MJ, Refaie R, Rankin KS, Young DA, Loughlin J (2014) Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis Rheumatol* **66**: 2450-2460.
- Sagi B, Maraghechi P, Urban VS, Hegyi B, Szigeti A, Fajka-Boja R, Kudlik G, Nemet K, Monostori E, Gocza E, Uher F (2012) Positional identity of murine mesenchymal stem cells resident in different organs is determined in the postsegmentation mesoderm. *Stem Cells Dev* **21**: 814-828.
- Sethe S, Scutt A, Stolzing A (2006) Aging of mesenchymal stem cells. *Ageing Res Rev* **5**: 91-116.
- Shi X, Yang X, Chen D, Chang Z, Cao X (1999) Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J Biol Chem* **274**: 13711-13717.
- Shirley D, Marsh D, Jordan G, McQuaid S, Li G (2005) Systemic recruitment of osteoblastic cells in fracture healing. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society* **23**: 1013-1021.
- Simic P, Zainabadi K, Bell E, Sykes DB, Saez B, Lotinun S, Baron R, Scadden D, Schipani E, Guarente L (2013) SIRT1 regulates differentiation of mesenchymal stem cells by deacetylating beta-catenin. *EMBO Mol Med* **5**: 430-440.

- Stelnicki EJ, Komuves LG, Kwong AO, Holmes D, Klein P, Rozenfeld S, Lawrence HJ, Adzick NS, Harrison M, Largman C (1998) HOX homeobox genes exhibit spatial and temporal changes in expression during human skin development. *J Invest Dermatol* **110**: 110-115.
- Stenderup K, Rosada C, Justesen J, Al-Soubky T, Dagnaes-Hansen F, Kassem M (2004) Aged human bone marrow stromal cells maintaining bone forming capacity *in vivo* evaluated using an improved method of visualization. *Biogerontology* **5**: 107-118.
- Stolzing A, Jones E, McGonagle D, Scutt A (2008) Age-related changes in human bone marrow-derived mesenchymal stem cells: consequences for cell therapies. *Mech Ageing Dev* **129**: 163-173.
- Stolzing A, Sethe S, Scutt AM (2006) Stressed stem cells: Temperature response in aged mesenchymal stem cells. *Stem Cells Dev* **15**: 478-487.
- Tollervey JR, Wang Z, Hortobagyi T, Witten JT, Zarnack K, Kayikci M, Clark TA, Schweitzer AC, Rot G, Curk T, Zupan B, Rogelj B, Shaw CE, Ule J (2011) Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res* **21**: 1572-1582.
- Trivedi CM, Cappola TP, Margulies KB, Epstein JA (2011) Homeodomain only protein x is down-regulated in human heart failure. *J Mol Cell Cardiol* **50**: 1056-1058.
- Tye CE, Gordon JA, Martin-Buley LA, Stein JL, Lian JB, Stein GS (2015) Could lncRNAs be the missing links in control of mesenchymal stem cell differentiation? *J Cell Physiol* **230**: 526-534.
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigartyo CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**: 1260419.
- Valenti MT, Mori A, Malerba G, Dalle Carbonare L (2015) Mesenchymal stem cells: A new diagnostic tool? *World J Stem Cells* **7**: 789-792.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**: RESEARCH0034.
- Veal EA, Day AM, Morgan BA (2007) Hydrogen peroxide sensing and signaling. *Mol Cell* **26**: 1-14.
- Venkataraman K, Futerman AH (2002) Do longevity assurance genes containing Hox domains regulate cell development *via* ceramide synthesis? *FEBS Lett* **528**: 3-4.
- Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dianes JA, Sun Z, Farrar T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolome S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**: 223-226.
- Wagner W, Bork S, Horn P, Kronic D, Walenda T, Diehlmann A, Benes V, Blake J, Huber FX, Eckstein V, Boukamp P, Ho AD (2009) Aging and replicative senescence have related effects on human stem and progenitor cells. *PLoS One* **4**: e5846.
- Wagner W, Ho AD, Zenke M (2010) Different facets of aging in human mesenchymal stem cells. *Tissue Eng Part B Rev* **16**: 445-453.
- Wagner W, Horn P, Castoldi M, Diehlmann A, Bork S, Saffrich R, Benes V, Blake J, Pfister S, Eckstein V, Ho AD (2008) Replicative senescence of mesenchymal stem cells: a continuous and organized process. *PLoS One* **3**: e2213.
- Wang Y, Crisostomo PR, Wang M, Markel TA, Novotny NM, Meldrum DR (2008) TGF- α increases human mesenchymal stem cell-secreted VEGF by MEK- and PI3-K- but not JNK- or ERK-dependent mechanisms. *Am J Physiol Regul Integr Comp Physiol* **295**: 1115-1123.
- Whitaker JW, Shoemaker R, Boyle DL, Hillman J, Anderson D, Wang W, Firestein GS (2013) An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med* **5**: 40.
- Yang KC, Yamada KA, Patel AY, Topkara VK, George I, Cheema FH, Ewald GA, Mann DL, Nerbonne JM (2014) Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support. *Circulation* **129**: 1009-1021.
- Yu JM, Wu X, Gimble JM, Guan X, Freitas MA, Bunnell BA (2011) Age-related changes in mesenchymal stem cells derived from rhesus macaque bone marrow. *Aging Cell* **10**: 66-79.
- Yuan HF, Zhai C, Yan XL, Zhao DD, Wang JX, Zeng Q, Chen L, Nan X, He LJ, Li ST, Yue W, Pei XT (2012) SIRT1 is required for long-term growth of human mesenchymal stem cells. *J Mol Med* **90**: 389-400.
- Zaim M, Karaman S, Cetin G, Isik S (2012) Donor age and long-term culture affect differentiation and proliferation of human bone marrow mesenchymal stem cells. *Ann Hematol* **91**: 1175-1186.
- Zhou S, Greenberger JS, Epperly MW, Goff JP, Adler C, Leboff MS, Glowacki J (2008) Age-related intrinsic changes in human bone-marrow-derived mesenchymal stem cells and their differentiation to osteoblasts. *Aging Cell* **7**: 335-343.
- Zimmermann P, Boeuf S, Dickhut A, Boehmer S, Olek S, Richter W (2008) Correlation of COL10A1 induction during chondrogenesis of mesenchymal stem cells with demethylation of two CpG sites in the COL10A1 promoter. *Arthritis Rheum* **58**: 2743-2753.
- Zykovich A, Hubbard A, Flynn JM, Tarnopolsky M, Fraga MF, Kerksick C, Ogborn D, MacNeil L, Mooney SD, Melov S (2014) Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. *Aging Cell* **13**: 360-366.

Web references

Web ref. 1 Human reference genome sequence ftp://ftp.ensembl.org/pub/release-73/fasta/homo_sapiens/dna/01012014

Web ref. 2 www.ebi.ac.uk/arrayexpress/experiments [01012014]

Web ref. 3 Ingenuity systems. Ingenuity pathway analysis [<http://www.ingenuity.com/>] [01012015]

Web ref. 4 Target scan, <http://www.targetscan.org/> [01012015]

Web ref. 5 A gene's mRNA level does not usually predict its protein level, Kendrick, N [01012014]

Discussion with Reviewers

Reviewer I: Do you think this work could offer the base for future studies to predict the age of freshly isolated MSC?

Authors: This study was undertaken to determine the difference at a systems level (transcripts, epigenetic and protein) of chronological ageing on bone marrow derived MSCs. This is of course not the same as looking at *in vitro* ageing. Age-related functional and biological alterations in MSCs were identified. The ages used in the study were in two distinct groups twenties and sixties, due to limitations in resources, but provides some interesting features not previously interrogated, namely chronological MSC ageing. It would be interesting in future studies to look at age groups between these extremes and also beyond, to perhaps eighties, to get a more refined idea of transcript, protein and epigenetic changes through age and their consequences. From the study, one could develop (following additional recruitment of donors through a wider age range and with further validation) a panel of molecules that may be able to predict the phenotypic age of the MSCs from a particular donor. Given the age-related changes in MSCs found in this study with potential for impaired therapeutic capacity of aged progenitor cells, these data could aid in the choice of MSCs (autologous or allogeneic) in particular patients that are undergoing cartilage or tendon repair.

Reviewer I: Which functional assay would you use to validate your results?

Authors: This is a very good question. The implications of the study are two-fold. Firstly, the results suggest that age-related dysregulation at all molecular levels tested in MSCs may have consequences in the ability of a damaged musculoskeletal tissue to repair and the mechanical properties of that tissue and thus functionality. Secondly, there will be a consequential effect on the content, viability and functionality of tissue engineered constructs made from ageing MSCs that are used in cartilage or tendon repair therapies. Functional assay, on a practical level, should determine the effects of the deterioration in regenerative potential on ageing. This question has partially been addressed in further studies, which are not alluded to

in this study and are not yet unpublished. Chondrogenic pellets, osteogenic 2D constructs and tendon 3D fibrin constructs from the young and old MSCs were produced and then subjected to the same platforms the MSCs were subject to (RNASeq, miRNASeq, methylation array and proteomics). Similar and distinct age-related pathways were identified. Further studies, which have not been undertaken but could be performed, would then look at the mechanical competence of the tissue-engineered constructs to establish the age-related effect and thus translational impact. These further studies did not look at differentiation potential, as there was only a single end-stage time point. However, other functional tests could ascertain if there was a difference in differentiation potential by comparing chondrogenic, osteogenic and tenogenic markers (the latter is difficult, as these are not well defined in tendon).

Reviewer II: Why were epigenetic histone marks and other ncRNAs (such as lncRNAs) besides miRNAs not investigated?

Authors: We tried to include as many 'molecular' platforms as possible in our small study, given financial constraints. Therefore, histone marks were not assessed in this study but would be beneficial. With regards to long non-coding RNAs analysis, we did in fact identify two differentially expressed lncRNAs (LNCs) in this study; AC004870.4 and RP11-670E13.5. The size selection in the library preparations for RNASeq enables LNCs to be identified and the data analysed. LNCs are a diverse group of non-protein coding RNA molecules longer than 200 nucleotides, whose expression is developmentally regulated and both cell and tissue specific. The function of the majority of LNCs is unknown. However, they have been implicated in many diseases. Studies have revealed that LNCs have significant roles in biological processes, including, transcription, imprinting, chromatin modification. Furthermore, they are important in controlling tissue differentiation. Previously a group has identified differentially expressed LNCs in bone marrow *versus* periodontal ligament derived MSCs from young donors. Here, RP11-670E13.5 was identified to be differentially expressed in bone marrow. However, there is no expression data for RP11-670E13.5 also known as lnc-DGKE-2 in MSCs. There may be a significant role for LNCs in phenotypic regulators of MSC differentiation and they could aid the establishment of lineage specification through independent mechanisms. LNCs represent potential biomarkers of disease and novel therapeutic targets for ageing, but much work is needed to elucidate their roles in MSC ageing.

Editor's Note: Scientific Editor in charge of the paper: Christine Hartmann.