

UNIVERSIDAD DE CANTABRIA

PROGRAMA DE DOCTORADO EN MEDICINA Y
CIENCIAS DE LA SALUD




TESIS DOCTORAL

Análisis genómico y funcional de las células madre mesenquimales de médula ósea de pacientes osteoporóticos

PHD THESIS

Genome-Wide Methylation and Transcriptome Analyses of Bone Marrow Mesenchymal Stem Cells from Osteoporotic Patients

A decorative molecular structure composed of various colored spheres (blue, yellow, orange, green, pink, grey) connected by thin lines, resembling a complex protein or DNA structure, positioned diagonally across the lower half of the page.

Realizada por: Álvaro del Real Bolt
Dirigida por: José Antonio Riancho Moral

Escuela de Doctorado de la Universidad de Cantabria
Santander 2019

UNIVERSIDAD DE CANTABRIA

PROGRAMA DE DOCTORADO EN MEDICINA Y
CIENCIAS DE LA SALUD



TESIS DOCTORAL

**Análisis genómico y funcional de las células
madre mesenquimales de médula ósea de
pacientes osteoporóticos**

PHD THESIS

**“Genome-Wide Methylation and Transcriptome
Analyses of Bone Marrow Mesenchymal Stem
Cells from Osteoporotic Patients”**

Realizada por: Álvaro del Real Bolt

Dirigida por: José Antonio Riancho Moral

Escuela de Doctorado de la Universidad de Cantabria

Santander 2019



Facultad de Medicina

Departamento de Medicina y Psiquiatría

D. José A. Riancho Moral, Doctor en Medicina y Catedrático del Departamento de Medicina y Psiquiatría de la Universidad de Cantabria, CERTIFICO:

Que D. Álvaro del Real Bolt ha realizado, bajo mi dirección, el trabajo titulado: “Análisis genómico y funcional de las células madre mesenquimales de médula ósea de pacientes osteoporóticos”.

Considero que este trabajo tiene originalidad y calidad científica suficientes para ser presentado como Tesis para obtener el grado de Doctor.

Facultad de Medicina-U.C. Laboratorio de Investigación Médica Traslacional (LIMT). Avda. Cardenal Herrera Oria s/n. 39011 Santander

Research carried out in this thesis was mainly developed in the Department of Medicine and Psychiatry of the Faculty of Medicine, University of Cantabria/ Mineral and lipid metabolism group of IDIVAL.

The research was funded with grants from the Instituto de Salud Carlos III (*PI12/615* and *PI16/915*). I have been funded by a predoctoral fellowship from the University of Cantabria and the Research Institute of Marques de Valdecilla Hospital (IDIVAL) (*CVE-2016-11669*).

‘Chaos is merely order waiting to be deciphered’

-José Saramago-

Acknowledgements

This thesis could not have been possible without the help of various people, who have influenced on my work directly and also in my mood, day to day.

At first, I would like to thank Dr. Riancho for being such an absolute director, for helping and guiding me in my investigation, every week, with intense work but always with enthusiasm. He has taught me to think and grow as a good investigator with his dedication. I feel very fortunate to have had the opportunity to work with him. I hope that this thesis is not the end of our relationship and our work together.

I would also like to thank my laboratory partner and better said, my friend Carolina Sañudo. We have spent much time together and the lab has given to us lots of moments. Thank you to make my work easier.

Thanks to my lab meeting partners, Laura and Pablo, that have endured me, with the good and the bad of my endless data analysis.

To IDIVAL and the University of Cantabria for providing me the necessary means and economic support to work in my thesis.

I appreciate Leandro Castellano and the Imperial College (London), for letting me use its facilities and for the attention I have received.

For sure, thanks to my parents and sister. They have taught me the values of life, hard work and ethics. We have been far away this years, but you have always been there, listening and supporting me.

And certainly, last but not least, I must thank to my wife Elena Mora, for all her love and support. To you I dedicate this thesis.

TABLE OF CONTENTS

ABBREVIATIONS.....	13
RESUMEN	21
SUMMARY.....	25
INTRODUCTION.....	29
Bone tissue	29
The osteoblastic lineage	31
Bone resorption.....	34
Bone remodelling	36
Mesenchymal stem cells.....	38
Aging and prevalent bone disorders: Osteoporosis and osteoarthritis	41
Epigenetics: definition and mechanisms	44
Chromatin structure, post-translational histone modifications and bone tissue	47
DNA methylation and bone	49
Expression of non-coding RNAs in bone	53
OBJECTIVES.....	73
METHODS	77
Bone samples and bone marrow mesenchymal stem cell (BMSC) isolation and culture	77
Osteogenic differentiation of BMSCs.....	78
Proliferation analysis	78
Cell lines cultures.....	79
DNA isolation.....	79
RNA extraction and purification	80
Genome-wide DNA methylation analysis.....	81
DNA methylation age predictor.....	84
Relative telomere length analysis	84
DNA methylation analysis by pyrosequencing	85
Transcriptome analysis	87
Real Time PCR analysis.....	91
Transfection analysis with siRNA and expression vectors.....	93
Plasmid cloning, purification and transfection experiments.	94
Bioinformatic tools and statistics	96
RESULTS.....	103
Isolation and characterization of BMSCs	103
Proliferative capacity of BMSCs	103

DNA methylation profiling	103
Epigenetic aging.....	109
Gene Expression Profiling	111
DNA methylation and gene expression replication.....	116
Osteogenic capacity of BMSCs	120
Functional experiments with ID2 inhibition	122
Expression of long non-coding RNAs.....	123
Replication of RNA sequencing.....	124
Replication of transcription analysis and signature of osteogenic differentiation	126
Long Non-coding RNA replication and functional studies.....	130
DISCUSSION	137
Osteoporosis as a prevalent complex disorder	137
MSCs and aging.....	137
MSCs and osteoporosis.....	138
Epigenetic marks of MSCs, aging and osteoporosis	141
DNA methylation of MSCs, aging and osteoporosis	141
MSCs, lncRNAs and osteoporosis.....	148
Study limitations	152
What this thesis adds	153
Future perspectives	153
CONCLUSIONS	157
REFERENCES	161
APPENDIX	185
Appendix 1: List of differentially expressed genes	185
Appendix 2: List of genes from the intersections	191
Appendix 3: List of publications	193

ABBREVIATIONS

ALP	Alkaline phosphatase
ALPL	Alkaline Phosphatase, Biomineralization Associated
ANRIL	Antisense RNA in the INK4 Locus
APC	Allophycocyanin
ASCs	Adult stem cells
BER	Base excision repair
BGLAP	Bone Gamma-Carboxyglutamate Protein
bHLH	Basic helix-loop-helix
BMD	Bone mineral density
BMI	Body mass index
BMP4	Bone morphogenetic protein 4
BMPs	Bone morphogenetic proteins
BMSCs	Bone marrow stem cells
BMU	Basic multicellular unit
CCD	Cleidocranial dysplasia
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A
CEBPA	CCAAT Enhancer Binding Protein Alpha
ceRNAs	Competing endogenous RNAs
CMV	Cytomegalovirus
COL1A1	Collagen Type I Alpha 1 Chain
CpG	Dinucleotide. Cytosine that precedes a Guanine
Csf1r	Colony stimulating factor 1 R
Ct	Threshold cycles
CTs	Chromosome territories
DANCR	Differentiation Antagonizing Non-Protein Coding RNA
DICER	dsRNA-specific RNase III family ribonuclease
DKK1	Dickkopf WNT Signaling Pathway Inhibitor 1
DLX2	Distal-Less Homeobox 2
DLX5	Distal less homeobox 5
DMEM	Dulbecco's Modified Eagle Medium
DMRs	Differentially methylated regions
DNMTs	DNA methyltransferases
DXA	Dual-energy x-ray absorptiometry
ENCODE	The Encyclopedia of DNA Elements
eRNAs	Enhancer-derived RNAs
ESCs	Embryonic stem cells
EZH2	Enhancer of Zeste Homolog 2
FAL1	Eukaryotic Translation Initiation Factor 4A3 (EIF4A-III)
FBS	Fetal bovine serum
FDR	False discovery rate
FENDRR	FOXF1 Adjacent Non-Coding Developmental Regulatory RNA
FITC	Fluorescein isothiocyanate
FOXC1	Forkhead Box C1
FOXP2	Forkhead Box P2

FPKM	Fragments per kilobase per million
FRX	Osteoporotic fracture
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GR	Glucocorticoid Receptor
GREAT	Genomic Regions Enrichment of Annotations Tool
GWAS	Genome-wide association studies
hASCs	Human adipose-derived mesenchymal stem cells
HCELL	Hematopoietic Cell E-/L-selectin Ligand
HDAC	Histone deacetylase family
HDAC5	Histone deacetylase 5
Hes	Hairy And Enhancer Of Split
Hey	Hairy Ears, Y-Linked
hnRNPI	Heterogeneous nuclear ribonucleoprotein I
hnRNPK	Heterogeneous nuclear ribonucleoprotein K
hnRNP-K	Heterogeneous Nuclear Ribonucleoprotein K
hnRNPL	Heterogeneous Nuclear Ribonucleoprotein L
HOTAIR	HOX Transcript Antisense RNA
HOTTIP	HOXA Distal Transcript Antisense RNA
HOXA-AS3	HOXA Cluster Antisense RNA 3
HULC	Hepatocellular Carcinoma Up-Regulated Long Non-Coding RNA
IBSP	Integrin Binding Sialoprotein
ID2	Inhibitor Of DNA Binding 2
IGF2	Insulin Like Growth Factor 2
IGFBP4	Insulin Like Growth Factor Binding Protein 4
Ihh	Indian hedgehog
IOF	International Foundation of Osteoporosis
ISCT	International Society for cellular Therapy
KCNQ1OT1	KCNQ1 Opposite Strand/Antisense Transcript 1
KREMEN2	Kringle Containing Transmembrane Protein 2
LADs	Lamina-associated domains
LAMC1	Laminin Subunit Gamma 1
LASP2	LIM And SH3 Protein 2
LB	Luria-Bertani media
LINC00963	Long Intergenic Non-Protein Coding RNA 963
LincRNAs	Long intergenic non-coding RNAs
lncRNA-OG	LncRNA Osteogenesis Associated
lncRNAs	Long non-coding RNAs
LOXL2	Lysyl Oxidase Like 2
LRP	Low-density lipoprotein receptor-related proteins
LSD1	Lys-specific demethylase 1
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
M-CSF	Macrophage colony-stimulating factor
MDS	Multidimensional scaling
MeCP2	Methyl-CpG binding-protein 2
MEG3	Maternally Expressed 3
MET	Mesenchymal-to-epithelial transition

MIR31HG	MIR31 Host Gene
miRNAs	Micro RNAs
MITF	Microphthalmia-associated transcription factor
MMPs	Matrix metalloproteinases
MODR	Osteogenesis differentiation-related lncRNA of membrane stem cells
mRNA	Messenger RNA
MSCs	Mesenchymal stem cells
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
ncRNAs	Non-coding RNAs
NFATc1	Calcineurin/nuclear factor of activated T cell c1
NRIP1e	Nuclear Receptor Interacting Protein 1e
OA	Osteoarthritis
OP	Osteoporosis
OPG	Osteoprotegerin
OSCAR	Osteoclast-associated receptor
OSX	Osterix
PACER	p50-associated COX-2 extragenic RNA
PACT	Protein activator of PKR
PANDA	p21-associated ncRNA DNA damage-activated
PAPAS	Proline-Serine-Threonine Phosphatase Interacting
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PcG	Polycomb group
PE	Phycoerythrin
PINT	Long Intergenic Non-Protein Coding RNA, P53 Induced Transcript
piRNAs	Piwi RNAs
Pol II	Polymerase II
PPAR	Peroxisome Proliferator Activated Receptor
PPARG	Peroxisome Proliferator Activated Receptor Gamma
PRC	Polycomb repressive complexes
PRC1	Protein Regulator Of Cytokinesis 1
PRC2	Protein Regulator Of Cytokinesis 2
pri-miRNA	Primary micro RNA
PTCH1	Protein Patched Homolog 1
PTENP1	Phosphatase And Tensin Homolog Pseudogene 1
PTGS2	Prostaglandin-Endoperoxide Synthase 2
PTMs	Post-translational histone modifications
PU.1	Hematopoietic Transcription Factor PU.1
RANK	Receptor activator of nuclear factor (NF)-kB
RANKL	RANK ligand
RB	Retinoblastoma protein
RBP	RNA binding proteins
RdRP	RNA-dependent RNA polymerase
RIN	RNA Integrity Number
RISC	RNA-induced silencing complex
RNAseq	RNA sequencing

RoR	Long Intergenic Non-Protein Coding RNA, Regulator Of Reprogramming
RPK	Reads per Kilobase
RPKM	Reads per kilobase per million
RPL13A	Ribosomal Protein L13a
rRNA	Ribosomal RNA
RT-qPCR	Real time quantitative PCR
RUNX2	Runt-Related Transcription Factor 2
SAM	Donor S-adenosyl methionine
sFRP	Secreted frizzled related protein
siRNAs	Small interfering RNAs
SIRT1	Sirtuin 1
SLC5A3	Solute Carrier Family 5 Member 3
SLPI	Secretory Leukocyte Peptidase Inhibitor
SNORAs	H/ACA box snoRNAs
SNORDs	C/D box snoRNAs
snoRNAs	Small nucleolar RNAs
SOCS3	Suppressor Of Cytokine Signaling 3
SOST	Sclerostin
SP7	Sp7 Transcription Factor
SPARC	Secreted Protein Acidic And Cysteine Rich
SPP1	Secreted Phosphoprotein 1
SRA	Steroid Receptor RNA Activator
SSCs	Skeletal stem cells
SWAN	Subset-quantile Within Array Normalization method
TADs	Topological-associated domains
TARID	TCF21 Antisense RNA Inducing Promoter Demethylation
TBP	TATA-Box Binding Protein
TDG	Thymine DNA glycosylase
TERC	Telomerase RNA component
TERRA	LncRNA telomeric repeat containing RNA
TERT	Telomerase reverse transcriptase
TET	Ten-eleven translocation
TFG-B	Transforming growth factor beta
THRIL	TNF And HNRNPL Related Immunoregulatory Long Non-Coding RNA
TMM	Trimmed mean of M values
TNF	Tumour necrosis factor
TNFRSF11B	TNF Receptor Superfamily Member 11b
TPM	Transcript per million
TRAFs	Receptor-activating factors
TRAP	Tartrate-resistant acid phosphatase
TRBP	Tar RNA binding protein
tRNA	Transfer RNA
TSS	Transcription start site
UNC5B	Unc-5 Netrin Receptor B
WNT	Wingless type family
XIST	X Inactive Specific Transcript

YWHA2	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Eta
-------	---

1. RESUMEN

RESUMEN

El hueso es un tejido activo que está en continua renovación en un proceso equilibrado, llevada a cabo por los osteoblastos, encargados de formar nuevo hueso, y por los osteoclastos, encargados de la resorción ósea. La diferenciación de las células madre mesenquimales (MSCs), precursores de los osteoblastos, es esencial para el mantenimiento de la masa ósea. En la osteoporosis hay un desequilibrio en el proceso de remodelado óseo, con una predominancia de la actividad osteoclástica sobre la osteoblástica. Esto conduce a una pérdida de la densidad mineral ósea y una mayor susceptibilidad a padecer fracturas. Los mecanismos epigenéticos son esenciales para la regulación celular y, por tanto, también son claves en el desarrollo de diversas enfermedades. Entre estos mecanismos se encuentran la metilación de ADN y la expresión de ARN largos no codificantes (lncRNAs). La capacidad funcional de las MSCs puede verse comprometida en personas de edad avanzada, en relación con los cambios epigenéticos asociados con el envejecimiento. Sin embargo, el papel de las MSCs en la patogenia de la osteoporosis no está bien definido. Por lo tanto, nuestro objetivo fue caracterizar las marcas de metilación, el patrón de expresión génica (codificante y no codificante de proteína) y la capacidad de diferenciación de las MSC de médula ósea (BMSCs) de pacientes con fracturas de cadera osteoporóticas.

Obtuvimos BMSCs de las cabezas femorales de mujeres que se sometieron a un reemplazo de cadera debido a fracturas de cadera y de controles con artrosis de cadera. La metilación del ADN se exploró con el microchip Infinium 450K (Illumina). El análisis del transcriptoma se realizó mediante secuenciación de ARN.

Las BMSCs de pacientes con fracturas mostraron una mayor proliferación y expresión de los genes reguladores osteogénicos RUNX2/OSX. Cuando se cultivaron en medio osteogénico, las BMSCs de pacientes con fracturas mostraron una capacidad de diferenciación alterada, con una actividad de fosfatasa alcalina reducida y una acumulación deficiente de una matriz mineralizada. Además, mostraron algunos signos de envejecimiento acelerado de la metilación.

Los análisis de metilación de ADN revelaron que la mayoría de los sitios diferencialmente metilados se dan en regiones genómicas con actividad reguladora (“enhancer”), a distancia de los promotores de genes. Estas regiones se asociaron, a su vez, con genes expresados diferencialmente, que estaban sobre-representados en vías relacionadas con el crecimiento de BMSCs y la diferenciación osteogénica.

Cuando nos centramos en la expresión de la parte no codificante del genoma, vimos que la mayoría de los transcritos eran de tipo antisentido. Los genes codificantes de proteínas en posición *cis* de estos lncRNAs antisentido, que también se expresaban diferencialmente en fracturas y artrosis, estaban altamente representados en vías relacionadas con la formación ósea.

En general, nuestros resultados sugieren que los mecanismos epigenéticos, y específicamente el estado de metilación de las regiones reguladoras y los lncRNAs juegan un papel importante en la determinación del patrón de expresión génica de BMSCs de pacientes con osteoporosis. Un mejor conocimiento de estas vías no sólo contribuirá a comprender mejor los mecanismos patogénicos de la osteoporosis, sino que puede llevar a identificar nuevas dianas para el diseño de fármacos potenciadores de la formación ósea.

2. SUMMARY

SUMMARY

Bone is an active tissue, continuously renewed in a balanced process carried out by the osteoblasts, bone forming cells, and by the osteoclasts, responsible for bone resorption. Mesenchymal stem cells (MSCs), precursors of osteoblasts, are essential for the maintenance of bone mass. In osteoporosis, exists an imbalance in the bone remodeling process, with a predominance of osteoclast activity over osteoblastic bone formation. This leads to low bone mineral density and an increased susceptibility to fractures. Epigenetic mechanisms are essential for cell differentiation and activity and, therefore, are also important for the pathogenesis of different diseases. These mechanisms include DNA methylation and the expression of long non-coding RNAs (lncRNAs), among others. In theory, the functional capacity of MSCs may be compromised in elderly people, in relation to the epigenetic changes associated with aging. However, the role of these cells in the pathogenesis of osteoporosis is not well established. Therefore, the aim of this thesis was to analyze DNA methylation marks, gene expression (protein coding and non-protein coding) and the differentiation capacity of bone marrow MSCs (BMSCs) of patients with osteoporotic hip fractures.

We obtained BMSCs from the femoral heads of women undergoing hip replacement surgery due to hip fractures and from controls with hip osteoarthritis. DNA methylation was explored with the Infinium 450K array (Illumina). Transcriptome analysis was performed by RNA sequencing.

BMSCs of patients with fractures showed greater cell proliferation and gene expression of the master regulator genes RUNX2/OSX. When cultured in osteogenic medium, BMSCs of patients with fractures showed an altered differentiation capacity, with reduced alkaline phosphatase activity and a deficient accumulation of a mineralized matrix. In addition, they showed some signs of accelerated epigenetic aging, as assessed by the methylation of some CpGs.

Genome-wide methylation analysis showed that most sites differentially methylated in BMSCs from patients with osteoporotic fractures, in comparison with controls with osteoarthritis, are located in genomic regions with enhancer activity. In turn, those enhancer regions were associated with differentially expressed genes, and these genes were enriched in bone related pathways, such as osteogenic differentiation.

Regarding the non-protein coding specific gene expression, we saw that most transcripts are antisense type. The protein coding genes in cis position associated with those antisense lncRNAs, which are also differentially expressed, are highly represented in pathways related to bone formation.

In general, our results suggest that both epigenetic mechanisms, DNA methylation marks of enhancer regions and lncRNAs, play an important role in the regulation of gene expression of BMSCs derived from patients with osteoporosis. A better knowledge of these pathways will not only improve our understanding of the pathogenesis of osteoporosis, but may also help to identify new targets for anabolic bone therapy.

3. INTRODUCTION

INTRODUCTION

Bone tissue

Bone is a mineralized connective tissue whose particular structure and composition allow to succeed in its important functions in a vertebrate organism. Bones have specific properties that make them capable to have mechanical, metabolic and hematopoietic functions: a) they support and protect internal soft organs, like heart, brain or lungs; b) they are the niche for the bone marrow cells and haematopoiesis; c) they also are the most important calcium and phosphorus reservoir and contribute to mineral homeostasis; d) they serve as energy reservoir, represented by the marrow fat; e) finally, bones also permit the body movement by anchoring muscles, ligaments and tendons.

Bone tissue originates from the mesodermal layer in the gastrulation during embryogenesis. The mesoderm is the intermediate layer which forms the skeleton (with exception of the craniofacial region) and muscles. Distinguishing the mesoderm in different parts, depending on the distance to the centre of the embryo, we differentiate the lateral plate of the mesoderm, which forms the appendicular skeleton (limbs), and the paraxial mesoderm, where the sclerotome that will give rise to the axial skeleton is situated. Craniofacial bones are formed from the cells of the neural crest. There are two forms of bone formation or ossification, intramembranous ossification and endochondral ossification. Intramembranous ossification is characterized by bone formed directly from undifferentiated connective tissue; this is, mesenchymal progenitors differentiate into osteoblasts to form the membranous bone. On the other hand, in the endochondral ossification cartilage is formed by chondrocytes before bone, then bone replaces the hyaline cartilage. Cartilage is not transformed into bone, but it is a template to be replaced by bone afterwards.

Bone composition is distinct from that of any other tissue because it has the only extracellular matrix that is mineralized. Bone matrix is mainly composed of mineral, collagen, noncollagenous proteins, water, and depending on the site, a small proportion of lipids. The mineral phase consists basically of hydroxyapatite $[\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2]$, with some substitutions like carbonate, magnesium and acid phosphate. This mineral strengthens the collagen structure, and also represents

a readily available source of ions, such as calcium, magnesium and phosphate, that helps maintaining the serum levels of minerals.

In addition to the matrix, there are three main types bone cells: the bone forming osteoblasts; osteocytes, which are mature bone cells immersed in mineralized bone; and the bone-destroying osteoclasts.

Bones have two types of tissue, compact or cortical bone, and cancellous or trabecular bone, which differ in how tightly the tissue is packed, so they have a different tissue density. Compact bone has closely packed osteons, forming a solid, compacted mass. Trabecular bone consists of a highly porous structure of rod-shaped trabeculae, adjacent to irregular cavities that contain red bone marrow. Obviously, trabecular bone is less dense (this is, has a smaller amount of mineralized tissue per volume) and more flexible than compact bone.

Bones in the skeleton can be classified in five different types according to their shape: long, short, irregular, sesamoid and flat bones. Long bones are characterized by having a central elongation filled with bone marrow, called diaphysis and at both ends a wider region known as epiphysis. Between the epiphysis and the diaphysis is the metaphysis, which is formed by trabecular bone and, during the growth period, a cartilaginous disc that allows the lengthening of the bone. The outside of the bone has a layer of connective tissue called the periosteum, which contains blood vessels that supply bone, as well as the nerve endings. There is also a membranous layer that covers the inner surface of the bone that is in contact with the marrow, known as the endosteum. Some examples of long bones are the femur, tibia and humerus.

Short bones are about as wide as long, formed by a thin layer of cortical bone filled with trabecular bone. They provide support with less movement and include the bones of the hands or feet for example. Sesamoid bones are bones buried in a tendon to protect it, like the patella (knee cap). Flat bones are thin and curved. They are composed of two thin layers of compact bone and an inner layer of spongy bone. Within this group are the bones of the skull. And finally, irregular bones are those bones that cannot be classified within the previous groups, such as the vertebrae or the bones of the face.

Bone is a highly dynamic tissue. Adult bone is continuously remodelled by a regulated process coordinated by bone forming osteoblasts and bone resorbing osteoclasts. This physiological process is called bone remodelling. After bone

lesions, the process leads to bone repair or regeneration. Any imbalance in the bone remodelling process leads to bone diseases, such as osteoporosis, among others ¹.

The osteoblastic lineage

The lineage of bone forming cells include mesenchymal stem cells (MSCs), osteoblasts and osteocytes. MSCs differentiate into osteoblasts (bone forming cells), which secrete most components of the non-mineralized bone matrix (osteoid), and finally they become embedded as osteocytes in the matrix that becomes progressively mineralized ^{2,3}.

Osteoblasts are cuboid cells that are responsible for bone formation. This type of cell is the result of the differentiation of MSCs, a pluripotent cell that can be found in a variety of tissues, including bone marrow, muscles, and fat. MSCs can differentiate into different cell components of the mesoderm, including bone, cartilage, muscle, fat, ligaments, and tendons. Thus, the stages of the osteoblastic lineage include different cell types, such as mesenchymal progenitors, osteo/chondroprogenitors, pre-osteoblasts, immature osteoblasts, osteoblasts, bone lining cells and osteocytes. The differentiation of MSCs into osteoblasts is controlled by specific cytokines and transcription factors, at each stage of differentiation. Among the humoral endocrine and paracrine factors involved in osteoblast differentiation are the Hedgehogs, BMPs, TGF- β , PTH and WNTs. These developmental signals regulate the expression of specific transcription factors, such as, RUNX2, OSX, NFAT and SOX9 ⁴.

Transcription factors in bone differentiation

RUNX2 is indispensable for osteoblast differentiation. This protein acts as a transcription factor that regulates the activation or inhibition of its target genes. In 1997 studies of several groups elucidated the role of RUNX2 in osteoblast differentiation and consequently bone formation ^{5,6}. They demonstrated that mice with a homozygous mutation in RUNX2 died just after birth and presented an absence of ossification. They also showed specific skeletal deformities, which are characteristic in cleidocranial dysplasia (CCD), a human heritable skeletal disorder. Furthermore, this gene regulates the expression of several osteoblast

marker genes: osteocalcin, bone sialoprotein, osteopontin and type 1 collagen ⁷. During the differentiation process, distinct transcription factors are expressed at different developmental stages and with different functions. SOX9 is necessary for chondrogenesis, and it seems to be also important in the first stage of differentiation of the mesenchymal progenitors to osteoblasts. In fact, deletion of SOX9 in the limb bud mesenchyme causes a lack of chondrocytes and osteoblasts. So, it is thought that SOX9 is important for the formation of osteochondroprogenitor cells, but it is not further expressed in the osteoblastic lineage⁸.

Osterix (OSX) is a zinc finger transcription factor expressed in osteoblasts and is also required for bone formation, as RUNX2 factor. OSX works specifically in osteoblasts and acts downstream of RUNX2. It means that OSX expression depends on the expression of RUNX2, but not vice versa ⁹. First, mesenchymal progenitors express RUNX2 and differentiate into preosteoblasts that do not express osteoblast marker genes yet. Then, RUNX2- and OSX-expressing preosteoblasts in both endochondral and intramembranous ossification differentiate into mature osteoblasts, this is, cells that express osteoblast specific marker genes ¹⁰.

Other signalling factors

Hedgehog (Hh) signalling proteins bind to the receptor PTCH1, which regulates the expression of the GLI family of transcriptions factors. There are three Hh proteins in mammals (Sonic, Indian, and Desert hedgehog) that play critical roles in organ development. Indian hedgehog (Ihh) is essential for the development of the osteoblast lineage in the endochondral skeleton by promoting osteoblast differentiation and inhibiting the alternative chondrocyte pathway ¹¹.

Bone morphogenetic proteins (BMPs) belong to the transforming growth factor beta (TGF- β) superfamily and are essential components of various biological processes including bone formation. BMP2, BMP6, BMP7 and BMP9 stimulate osteoblastic differentiation. They increase alkaline phosphatase activity in pre-osteoblasts, and also other early markers of osteoblast differentiation. Not all members of this family are bone inducers; BMP3 and BMP13 inhibit bone formation ¹².

Notch proteins mediate communication between cells. They are transmembrane proteins that undergo proteolytic cleavage by presenilin. The resultant Notch intracellular domain translocates into the nucleus and regulates the expression of transcription factors like *Hes* or *Hey* family of transcription factors. MSC differentiation into osteoblasts is inhibited via *Hey1*. In fact, Runx2 transcriptional activity is physically antagonized by the protein encoded by Notch target gene *Hey1* ⁴.

WNTs are glycoproteins with important roles in regulating the osteoblastic lineage. WNT ligands are transduced by a family of seven-pass transmembrane receptor of the frizzled family, and the co-receptors low-density lipoprotein receptor-related proteins (LRP4-6). The binding of a WNT ligand to a frizzled receptor and LRP activates an intracellular signalling cascade, which can include the canonical β -catenin-dependent pathway or the noncanonical β -catenin-independent pathways. Some rare monogenic disorders underscore the important role of Wnt pathways in skeletal homeostasis. Thus, inactivating mutations of the WNT co-receptor LRP5 result in osteoporosis pseudoglioma syndrome. On the other hand, gain-of-function mutations cause a high bone mass phenotype. In addition, mutations in the sclerostin gene (*SOST*), an inhibitor of this pathway, cause sclerosteosis or Van Buchem disease (also high bone mass disorders) depending on whether the mutation is in coding or in regulatory regions of the gene, respectively ^{13,14}.

At the end of the bone formation phase, destiny of osteoblasts can be one of the three following stages (Figure 1).

1. Become embedded in bone as osteocytes. Osteocytes are the most abundant cells in bone (90-95% of the total number of cells), as well as the most long lived, with a life span of up to 25 years. However, in high bone turnover states, the life of osteocytes may be shorter. Osteocytes are spider-shaped cells embedded in the mineralized bone matrix, with several long and branched cell processes located inside canaliculi, which permit to establish cell-cell communication between different osteocytes or between osteocytes and cells located on the bone surface. Osteocytes are mechanosensing cells involved in the skeletal responses to mechanical stimuli, by translating the physical signals such as mechanical loads into biochemical signals. Osteocytes regulate bone remodelling, via

paracrine signals that control the differentiation of osteoblasts and osteoclasts ^{2,15}.

2. Become quiescent osteoblasts in the bone surface as “bone lining cells”. The majority of cancellous and endocortical bone surfaces are covered by flat bone lining cells. These abundant cells may play an important role in matrix metabolism during bone remodelling process ¹⁶. Also, there is some evidence that bone lining cells, under certain stimuli, can be activated and transformed into active osteoblasts ¹⁷.
3. Finally, some osteoblasts experience apoptosis and die after they finish forming bone matrix.

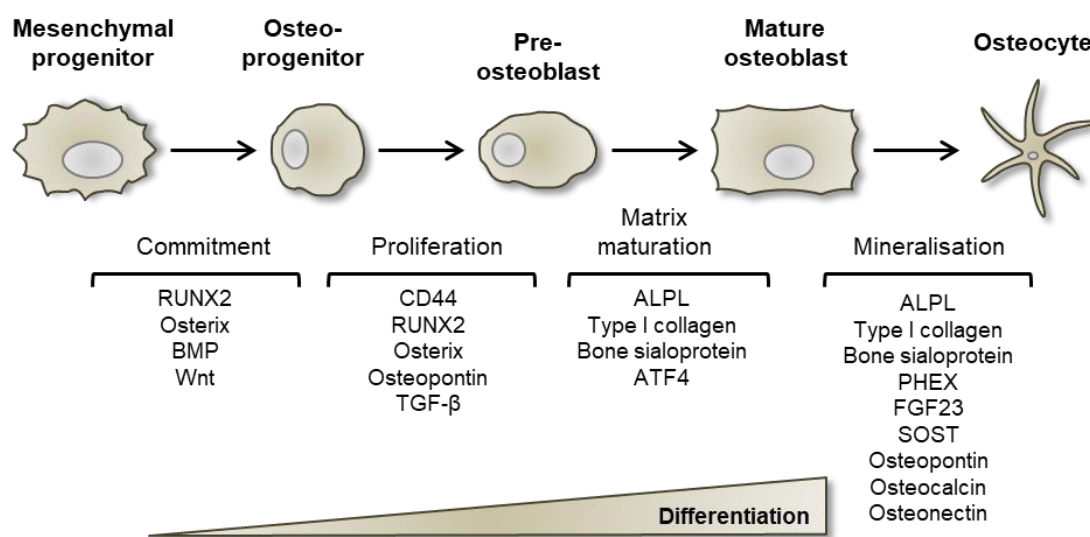


Figure 1. Diagram of the morphological changes associated with osteogenic differentiation, and markers that usually characterize each of the different stages through which a mesenchymal stem cell commits to the osteoblastic lineage. Adapted from ¹⁸.

Bone resorption

Bone resorption is the process of degradation of bone tissue as a necessary step for tissue renovation. Osteoclasts are terminally differentiated multinucleated cells, responsible for resorbing bone, and are related to the monocyte/macrophage family. Due to its terminal stage, they are unable of self-replication. Osteoclasts degrade bone tissue by secreting H^+ , Cl^- , cathepsin K and matrix metalloproteinases (MMPs) to dissolve the mineral and organic components of the bone matrix ¹⁹. Consequently, severe dysfunction of osteoclasts causes osteopetrosis, a disorder characterized by high bone mass and reduced marrow space due to impaired bone resorption. Several

investigations of human and mice models of osteopetrosis have identified macrophage colony-stimulating factor (M-CSF), receptor activator of nuclear factor (NF)- κ B (RANK), RANK ligand (RANKL) and tumour necrosis factor- α (TNF- α) as essential factors for osteoclastogenesis and osteoclast survival ²⁰.

Osteoclasts derive from hematopoietic precursors of the myeloid lineage, specifically of the monocytic line. The first stages of differentiation require critical genes, such as M-CSF, colony stimulating factor 1 R (Csf1r) and the transcription factor PU.1. PU.1 binds to the promoter Csf1r to upregulate its transcription.

M-CSF induces the proliferation of osteoclast progenitors and upregulates the expression of RANK, which is a very important receptor for the complete differentiation of osteoclasts. M-CSF also activates Microphthalmia-associated transcription factor (MITF), that regulates the anti-apoptotic protein BCL-2 in the osteoclast and thus promotes cell survival ²¹.

The next step of osteoclast maturation involves RANKL and RANK, ligand and receptor, respectively, that are members of the TNF superfamily. RANKL binds to its receptor (RANK), expressed by osteoclast precursors. Activated RANK, like other TNF family receptors, recruits TNF receptor-activating factors (TRAFs), which activate an intracellular cascade of protein-kinases and transcription factors. Specifically, TRAF6 signalling activates NF- κ B and the calcineurin/nuclear factor of activated T cell c1 (NFATc1) signalling to induce osteoclast formation ²⁰. In summary, RANKL enhances NFATc1 transcription, which regulates several osteoclast-specific genes, including tartrate-resistant acid phosphatase (TRAP), cathepsin K and osteoclast-associated receptor (OSCAR) (Figure 2).

RANKL/RANK is a regulated system, in which osteoprotegerin (OPG) has an important inhibitory role. OPG is a decoy receptor for RANKL, so that it prevents the binding of RANKL to its receptor RANK and consequently inhibits osteoclast formation. The important role of OPG is revealed by the fact that transgenic mice that overexpress OPG exhibit osteopetrosis. On the contrary, OPG-deficient mice develop early osteoporosis related to enhanced osteoclast differentiation ²².

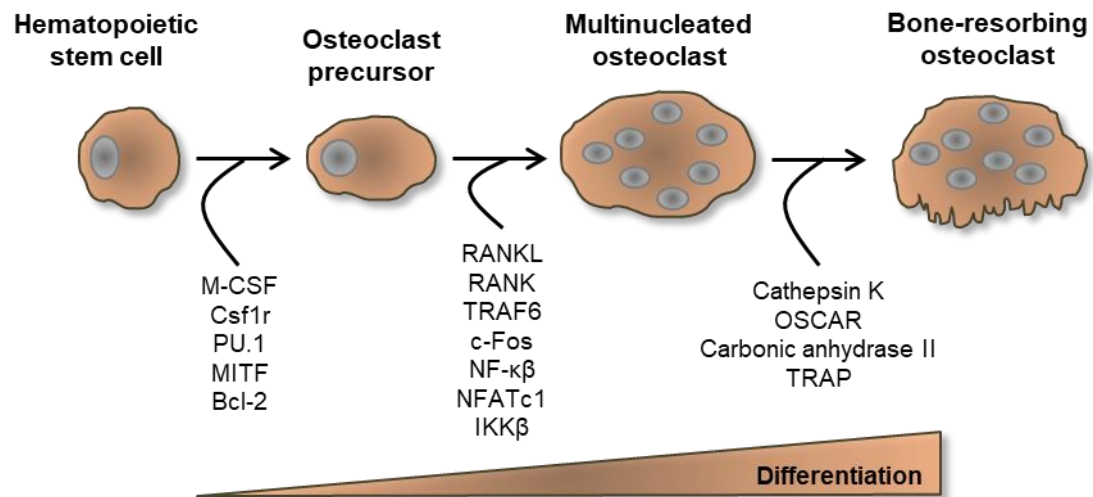


Figure 2. Diagram of osteoclastic lineage associated with the stage-specific key molecules for osteoclast differentiation and function. Adapted from ²³.

Bone remodelling

Bone remodelling is the process of bone tissue renovation to preserve bone quality and maintain skeletal homeostatic equilibrium. Bone is an active tissue that undergoes continuous remodelling by the activity of osteoblasts and osteoclasts. Bone resorption and formation are balanced in a homeostatic equilibrium so that old bone is continuously replaced by new tissue and, thus, microfractures are repaired. This contributes to the maintenance of the mineral composition, bone density and quality throughout adult life.

Osteoblasts and osteoclasts cooperate in the remodelling process in what is called basic multicellular unit (BMU). The spatial organization of BMUs is somewhat different in cortical and trabecular bone. Trabecular bone is more actively remodelled than cortical bone due to a larger ratio of surface/volume, and the BMUs are located on the surface, where osteoclasts are resorbing bone. After the resorption, the cells of the environment prepare the surface for bone formation, by providing signals for osteoblast differentiation and function. Finally, the surface is covered with flattened lining cells. In the cortical bone, during a remodelling cycle, several osteoclasts dig a circular tunnel that move through bone, followed by numerous osteoblasts that fill the tunnel, in conjunction with vessels and nerves that are also filling the space ^{24,25}.

Bone remodelling process needs to select the regions to be remodelled. Osteocytes are responsible for this initiating event in the BMU, which forms under

a canopy of cells, probably including lining cells, and with nearby capillaries. Those capillaries and nearby bone marrow provide the hematopoietic precursors for osteoclast formation, and contribute to provide molecular signals needed for osteoclast differentiation. Other cells, including osteocytes, osteoblasts and vascular cells likely contribute to regulate the differentiation of osteoclast precursors ²⁵.

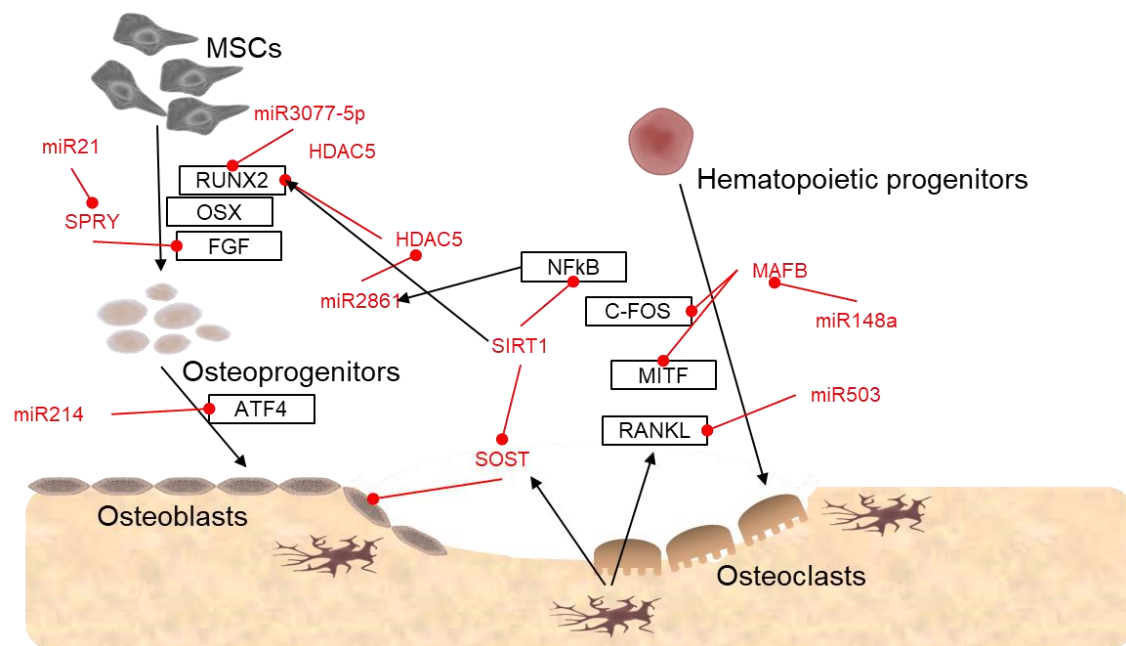


Figure 3. Bone remodelling scheme with various key molecules, known to regulate bone formation and bone resorption. Surrounded in black rectangles are important genes and transcription factors. Coloured in red, some coding and non-coding regulators, with dot-ended lines pointing inhibition over genes or cell processes.

Osteocytes are the most abundant cells in bone. They are responsible for the adaptation of bone to mechanical forces, and the response to several hormones and they also regulate bone remodelling through different molecular mechanisms. Mature osteocytes embedded in the bone matrix secrete sclerostin, the protein encoded by SOST gene, which is a strong inhibitor of bone formation, through the inhibition of the canonical Wnt signalling. So, osteocytes regulate bone formation through the secretion of sclerostin. On the other side, they also regulate bone resorption through RANKL signalling. Osteocyte apoptosis induce osteoclastogenesis by stimulating other stromal or osteoblastic cells to secrete RANKL. Moreover, osteocytes directly secrete RANKL to regulate osteoclast differentiation ²⁶. However, RANKL and sclerostin are not the only factors

implicated in the regulation of bone remodelling. In fact, there is a complex biochemical network regulating bone formation and bone resorption. The genes involved include protein coding genes, non-protein coding genes, such as microRNAs, and proteins that are regulating other epigenetic marks, including HDAC5 or SIRT1 (Figure 3).

Mesenchymal stem cells

Stem cells are unspecialized cells with the potential to divide in exact copies, or to differentiate into specialized cells with a specific function. Stem cells are essential for development, maintenance and repair of our organs. That is the reason why stem cells have become the focus of cell-based therapy for regenerative medicine. Stem cells can be classified based on their potency or their sources. Potency refers to the range of differentiation options into diverse cell types. Regarding this classification, there are four main types, totipotent, pluripotent, multipotent and unipotent cells:

- a. Totipotent cells have the ability to differentiate into all the possible cell types, including the extraembryonic cells. These cells are the first division of the zygote.
- b. Pluripotent stem cells can differentiate into all body cells, including embryonic stem cells (ESCs) and all the cells derived from the three germ layers (mesoderm, endoderm and ectoderm).
- c. Multipotent stem cells can differentiate into more than one cell type from the same germ layer.
- d. Unipotent stem cells may differentiate into a single cell type.

However, the easiest way to classify stem cells is in two types: embryonic and non-embryonic or adult stem cells. Embryonic stem cells (ESCs) are from the inner cell mass of the blastocyst, they are pluripotent and may differentiate into all cell types of the three germ layers. Adult stem cells (ASCs) are obtained from postnatal tissues; they are necessary to maintain tissue and organ mass ²⁷.

One specific type of ASCs, denominated mesenchymal stem cells (MSCs), have raised a huge interest in the field of regenerative medicine. In the 60s, Friedenstein and co-workers demonstrated the osteogenic potential of

heterotopic transplantation of a specific colony of bone marrow cells, that were called osteogenic stem cells. They observed that these cells have a rapid adherence to tissue culture vessels and, when placed into a diffusion chamber, reticular or bone tissue is formed instead of haemopoietic elements ²⁸. Later, the name MSCs was first used in the early 90's by Caplan ²⁹, who demonstrated that these cells are capable to differentiate into different lineages of the mesenchyme, such as bone and cartilage.

MSCs are characterized in vitro by their adherent properties and fibroblastic morphology. They express specific surface markers (CD90⁺, CD73⁺, CD105⁺, CD34⁻, CD45⁻ and CD14⁻, among others) and they may differentiate into different cells of the mesenchymal lineage, including osteoblasts, adipocytes or chondrocytes. MSCs have immunomodulatory properties, so they have been related to the immune dis-regulation of some diseases ³⁰. They lack immunogenicity because they express low levels of the major histocompatibility complex-I and show absence of expression of the major histocompatibility complex-II. They also interact with T and B lymphocytes to modulate its proliferation or activation ³¹.

In addition, MSCs have the capacity to migrate to target tissues, by transmigration across the endothelium, a process involving integrins and selectins ³². Another distinctive property of MSCs is the secretion of bioactive molecules, such as cytokines, chemokines and growth factors that affect local and systemic homeostasis. The secretome of MSCs may modulate immune responses, apoptosis, wound healing, tissue repair and angiogenesis ³³. Thus, MSCs have been the focus of many investigations of regenerative medicine and cellular therapy, both autologous and allogeneic ³⁴.

For many years, bone marrow has been considered the main source of MSCs, but more recent investigations have found the presence of MSCs in other tissues, such as adipose tissue, peripheral blood, synovial membrane, dental pulp, pancreatic islets, umbilical cord and placenta ³⁵. The adipose tissue seems to be one of the most interesting sources due to the large amount of cells obtained with relatively easy and safe collecting procedures ³⁶.

Nowadays, the scientific community has realized that MSCs isolated from distinct tissues show differences in their functions and developmental potential, and therefore in their therapeutic uses. Accordingly, it is recommended to revise the

name MSCs ³⁷. Caplan suggested to change the name of MSCs to “Medicinal signalling cells” to precisely point the fact that these cells migrate into sites of injury and secrete molecules with specific properties to heal and regenerate the tissue ³⁸. Another prominent study has pointed that MSCs, from different sources of origin, have dissimilar differentiation properties ³⁹. They conclude that “MSCs” terminology should be abandoned and replaced by a new and more explicit nomenclature. Furthermore, Bianco and Robey also called to abandon the term “MSCs”. Mesenchymal cells from non-skeletal tissues may be moved to osteogenesis by signalling molecules that induce bone differentiation, such as bone morphogenetic protein (BMP). However, they are not skeletal progenitor themselves, and differ from the mesenchymal precursors in the bone marrow, which naturally express the osteogenic transcription factor, runt-related transcription factor 2 (Runx2), and do not require the induction of exogenous factors. Therefore, some investigators suggested to replace the general term MSCs by skeletal stem cells (SSCs) ⁴⁰. Thus, SSCs are multipotent cells located in the bone marrow, which could differentiate into cartilage, bone, stromal cells and marrow adipocytes. These cells will likely become very important in the skeletal development and regenerative medicine approaches, because they home to skeletal injury sites and obviously have bone formation properties ⁴¹.

From now on, whenever possible, MSCs will be referred to in this thesis according to the source of origin, usually the bone marrow, as bone marrow stem cells (BMSCs).

As already explained, BMSCs may differentiate into different tissues of the mesoderm lineage, including adipose, cartilage and bone tissue. During the differentiation process towards a specific cell type, there are several factors with both, stimulation or inhibition roles ⁴². In the first steps of osteoblastic differentiation, the transcription factors Runx2 and Osx play critical roles (Figure 1) ⁸. The importance of mesenchymal stem cells in tissue engineering and regenerative medicine is emphasized by the fact that there are currently 273 completed trials including “mesenchymal stem cells” in the clinicaltrials.gov database, including 55 in the United States, 71 in Europe and 59 in East Asia.

Aging and prevalent bone disorders: Osteoporosis and osteoarthritis

The aging process is characterized by a progressive decline in the physiological integrity, which causes a loss of normal tissue function and greater susceptibility to disease. However, molecular mechanisms involved in this aging process are not well understood yet. Structural changes associated with aging are visible, both macro and microscopically, and are associated with a loss of normal tissue function. Lopez-Otín and colleagues pointed at nine hallmarks of cellular and molecular aging (genomic instability, altered intercellular communication, stem cell exhaustion, cellular senescence, mitochondrial dysfunction, deregulated nutrient-sensing, loss of proteostasis, epigenetic alterations, and telomere attrition). Each hallmark contributes to the aging process itself, and the exacerbation of the hallmark accelerate tissue aging. They are interconnected and related each other ⁴³. Aging impairs the functionality of stem cells. For example, some authors reported that BMSCs suffer a drift in their functionality and show a decreased proliferation capacity with preferential differentiation towards adipogenesis instead of osteogenesis. This could translate into decreased bone formation, thus increasing the risk of fractures ⁴⁴.

Due to the progressive aging of society, the prevalence of aging-related diseases has increased significantly in recent years. Regarding the skeleton, the most prevalent diseases are osteoporosis and osteoarthritis. Both are highly prevalent diseases, and they are considered as “complex diseases”. Complex diseases are disorders caused by a combined effect of different factors, including genetic and environmental factors, that interact in complex ways finally resulting in disease.

Osteoporosis

Osteoporosis is the most prevalent bone disorder, characterized by a progressive decrease in bone mineral density (BMD), a more porous bone, and consequently an increased susceptibility to fractures. Patients with osteoporosis may suffer bone fractures after low-energy trauma (or sometimes even during normal activity in the absence of trauma). The International Foundation of Osteoporosis (IOF) estimates that 1 in 3 women and 1 in 5 men over 50 years of age are at risk of an osteoporotic fracture, worldwide. One third of all postmenopausal women have

osteoporosis in the United States and in Europe. The most common fractures associated with osteoporotic fragility occur at the hip, wrist and the spine. Hip fractures are of concern due to their impact on life expectancy and quality of life. These fractures usually require surgery and represent an important proportion of the activity of Traumatology departments.

Osteoporosis can be classified as primary and secondary. Primary osteoporosis is related to changes occurring with aging, such as the decrease in sex hormones, renal function, intestinal absorption of calcium and muscle mass, as well as, perhaps, age-related changes in the intrinsic properties of bone cells responsible for remodeling ⁴⁵.

There is a long list of diseases and other factors associated with secondary osteoporosis. They include endocrine, nutritional, haematological, renal and autoimmune disorders; drugs, such as glucocorticoids and blockers of sex hormone synthesis, among others ⁴⁶.

In current practice, osteoporosis is diagnosed by the occurrence of fragility fractures (i.e. fractures in the absence of high-energy trauma) or a decreased bone mass, as measured by DXA or other suitable procedure. In this regard, osteoporosis is usually diagnosed when the patient has a t-score below -2.5. However, it must be emphasized that, even though there is an exponential inverse relationship between DXA scores and fracture risk, most fragility fractures occur in individuals with “osteopenia” (T-scores between -1 and -2.5) or even with normal DXA scores.

Osteoporosis is caused by a combined action of genetic and environmental factors. Environmental factors act on a genetically susceptible individual to cause disease. Among acquired factors influencing bone mass and fracture risk, there are both external and internal ones, including nutrition, exercise, toxic consumption, diseases and drugs. Genome-wide association studies (GWAS) are being important tools to confirm the role of different genes and pathways in osteoporosis and other complex disorders. Thus, several individual GWAS and the subsequent combined meta-analyses have confirmed the skeletal role of known pathways such as estrogen, Wnt, RANK/RANKL and other novel genes not previously associated with bone biology ^{47,48}. The mechanisms mediating the interaction between the environmental factors and the genetic factors underlying

individual susceptibility to disease are largely unknown. Nevertheless, they likely involve epigenetic mechanisms that are described below ^{49,50}.

Osteoarthritis

Osteoarthritis (OA) is the most common joint disease. It is characterized by the progressive degeneration of articular cartilage. The prevalence is very high, affecting about 10% men and 18% women over 60 years of age. Due to the high prevalence, OA is a major cause of disability and socioeconomic costs in developed countries ⁵¹. Risk factors for OA include some influencing the overall skeleton and others that influence just a single joint or a group of joints (including injury and abnormal loading, among the most representative ones) ⁵². Specific person-level risk factors include age, gender, obesity and genetics. Age is a multidimensional risk factor, for aging is associated with multiple changes involved in the disease, such as sarcopenia, thinning of cartilage and oxidative stress. OA is more prevalent in females than in males, but the frequency of OA increases with age in both sexes. Obesity is strongly associated with knee OA. In addition, OA has a strong genetic predisposition. Joint injury is another risk factor for developing OA. It is thought that traumatic lesions damage the cartilage. Subsequently, an abnormal response of chondrocytes, which form new cartilage, cause the release of enzymes that promote cartilage degradation.

Several candidate gene studies and GWAS have shown multiple genes involved in OA susceptibility ⁵³.

Although cartilage alterations are considered a crucial characteristic of OA, other joint structures, including bone and synovial membrane, are also involved. Bone changes include thickening of the subchondral bone and the appearance of osteophytes (immature bone protrusions), already in early stages of cartilage degeneration ⁵⁴. Those changes characteristic of OA go in the opposite direction of those observed in osteoporosis, characterized by bone loss. Several studies have also shown an association between OA and a trend for higher bone mass over the whole skeleton, however this remains a controversial issue ^{55,56}.

Epigenetics: definition and mechanisms

Epigenetics literally means ‘above the genetics’, but this definition is incomplete and controversial. Nowadays, epigenetics is conceived as the study of heritable changes in gene expression (activation or deactivation of gene expression) that do not modify the DNA sequence. Briefly, epigenetics is responsible for activating genes that cells have to use and turning down genes that are not needed. This modifies the phenotype without altering the genotype. The term “epigenetics” was first adopted by Conrad Waddington in 1942. He defined it as the “whole complex of developmental processes that lie between genotype and phenotype” and proposed the concept of epigenetic landscape. It was illustrated by a ball on top of an inclined surface; when rolling down there are different ways the ball could take (Figure 4). Similarly, according to Conrad’s “epigenetic landscape” a cell can develop and differentiate following a number of defined pathways ⁵⁷. However, recent studies have shown that cell biology is less deterministic than previously suggested, and there are numerous examples of de-differentiation and trans-differentiation of cells ⁵⁸.

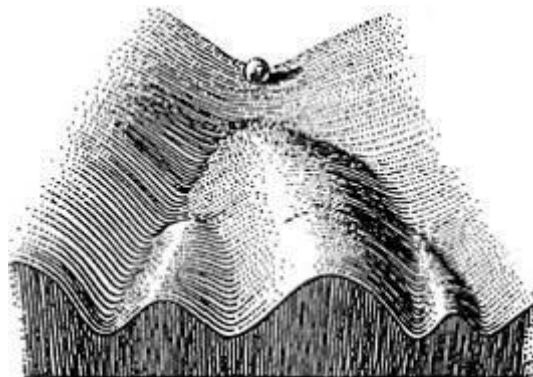


Figure 4. The epigenetic landscape proposed by C. H. Waddington (1940)

Epigenetic changes switch genes “on” or “off” and determine which proteins are transcribed. Thus, epigenetic mechanisms and changes of epigenetic marks are elements of normal cell biology. They are influenced by internal and external factors, like age, lifestyle, environment and disease. On the other hand, they participate in the pathogenesis of numerous neoplastic and non-neoplastic conditions. Epigenetic marks are reversible, so, unlike the genotype, the epigenome changes with age and environmental factors.

The best known and most studied epigenetic mechanisms are DNA methylation, post-translational histone modifications (PTMs, such as acetylation, methylation,

etc.) and non-coding RNA (ncRNA) expression. They act at different levels of the genetic code; at the chromatin level, as marks that either facilitate or inhibit transcription; or at the post-transcriptional level, modulating protein translation. DNA methylation and histone modifications mainly regulate chromatin assembly and gene transcription. ncRNAs act at the transcriptional or post-transcriptional level ⁴⁹.

Epigenetic mechanisms do not work alone, they are all related with each other, and if one of them fails, the regulation fails. Genomic imprinting represents one of the first-known and strongest evidences for the role of epigenetics in human biology and pathophysiology. Genomic imprinting consists in an epigenetic regulation characterized by the restriction of gene expression to one of the two parental chromosomes, this is, the one inherited from the mother or the one inherited from the father. The discovery of uniparental disomy leads to uncover disorders of genomic imprinting, such as Prader-Willy syndrome, Angelman syndrome, pseudohypoparathyroidism or Beckwith-Wiedemann syndrome ⁵⁹.

Normal and abnormal phenotypic differences among individuals are due to genotypic differences, environmental factors, and epigenetic differences. In this line, Fraga and collaborators found, in a large cohort of twins, that different phenotypes can originate from the same genotype by altering epigenetic marks and thus modulating gene expression ⁶⁰.

One of the most striking properties of genomes is their capacity to create a range of different cell types in a highly ordered manner. Understanding this process has helped to encourage new advances in epigenetics. When fertilization occurs, gametes fuse to form the first cell of the embryo, called zygote. Fused gametes undergo epigenetic reorganisation and restore cell totipotency. In mouse embryos, transcriptional activity begins in the first cell cycle when paternal and maternal chromatin are still in separate nuclear entities in the same cytoplasm. Paternal or maternal gametes do not show hyperacetylated histone H4, but immediately after fertilization, paternal chromatin shows hyperacetylation in H4 (a modification associated with active gene transcription). When transcriptional differences in S/G2 phase occur, male and female pronuclei have similar levels of H4 hyperacetylation ⁶¹. Furthermore, DNA methylation levels are different in the paternal and maternal pronucleus of the zygote ^{62,63}. Nowadays, it is not well established when are the first specified cells appearing but probably between the

8-16-cell morula, that give rise to the embryo and extraembryonic lineages ⁶⁴. Additionally, epigenetic mechanisms are regulating all embryo differentiation states, such as the formation of the distinct layers occurring in the inner cell mass during fetal development.

Environmental factors impact the epigenome not only during postnatal life, but also in utero. Several epidemiological studies have disclosed the influence of maternal nutrition on the epigenome later in life. One of the most significant studies was carried out in individuals who were prenatally exposed to hunger during the Dutch Hunger Winter in 1944-45. Investigators showed that these individuals, studied more than 6 decades later, presented less DNA methylation of the imprinted IGF2 gene and had an increased prevalence of high body mass index (BMI), insulin resistance or hypercholesterolemia. These data were among the first to provide empirical support for the relationship between early-life environmental conditions and epigenetic changes in humans, and their persistence throughout life ⁶⁵. On the other hand, more recently maternal obesity and gestational weight gain have also been associated with epigenetic changes in offspring. Various studies suggest that an important component of metabolic disease risk has a prenatal developmental basis and is associated with later obesity ^{66,67}. Besides, several studies support that protein restriction is associated with impaired fetal growth and development of diabetes and hypertension in childhood. For example, fetal protein restriction changed DNA methylation and gene expression levels of the AGTR1B gene, implicated in hypertension. Low protein maternal diet also changes the methylation and expression of glucocorticoid receptor (GR) and PPAR genes, which are important for the regulation of blood pressure and lipids in adults ⁶⁶.

Bone development alterations during intrauterine life is a factor for osteoporosis. Maternal calcium and vitamin D intake are important in the building of fetal bones. The mechanisms involved are unclear, but may include epigenetic changes, such as modifications of DNA methylation of genes near to vitamin D response elements, and genes encoding placental calcium transporters ⁶⁸.

Nutrition, and other lifestyle factors are also very important in postnatal individuals. For example, vitamin B12 and folate deficiency increases homocysteine levels and reduces SAM (the methyl group donor), thus causing a global DNA hypomethylation ⁶⁹. Smoking has been found to broadly impact DNA

methylation levels at many regions of the genome, and these changes are persistent many years after stopping smoking ⁷⁰. Childhood abuse is associated with adult disease risk, and has been associated with some DNA methylation profiles, suggesting that abuse in childhood has long-lasting epigenetic influences ⁷¹. Another stressful situation in human minds are suicidal thoughts or suicidal depression. Gene expression and DNA methylation changes have been reported in the brains of suicide completers, which may help to explain behavioural changes increasing suicidal thoughts ⁷².

Stressful experiences, and other life factors, induce changes in the methylation and expression of genes that are also implicated in the regulation of bone. Thus, epigenetic changes are a link between these factors and the skeletal condition.

Chromatin structure, post-translational histone modifications and bone tissue

Chromatin is the state of DNA within the cell nucleus. It is composed by DNA and proteins that very condensed, organised and wrapped, form the chromosomes. Nucleosomes, the smallest structural unit of the chromatin, are octamers of a core formed by four histones (H3, H4, H2A and H2B) wrapped with about 147 base pairs of double stranded DNA. Histones are basic proteins and their positive charges allow them to associate with DNA. Histones are chemically modified in their terminal N and C tails by different types of enzymes, and these modifications regulate chromatin's state. There are more than 60 different aminoacids on histone tails where these modifications occur. Most common modifications are acetylation and methylation of lysines and arginines and phosphorylation of serines and threonines. However, there are also other types of modifications, including ribosylation, ubiquitynation and sumoylation of lysines. Moreover, methylation modifications are particularly complex because they may have three different forms: mono-, di- or trimethyl for lysines and mono- or di-methyl for arginines ⁷³. Some marks promote gene expression (H3K27ac and H3K4me1), whereas other marks (H3K9me3 and H3K27me3) repress gene expression.

Lysine residues have a positive charge that interacts with the negatively charged phosphate of nucleotides. Covalent modifications, above explained, lead to alterations in chromatin structure and gene accessibility. These modifications are

reversible and have capacity of folding (heterochromatin) and unfolding (euchromatin) chromosomes ⁷⁴.

Moreover, several histone post-translational modifications can generate an anchor site for nuclear proteins that modulate chromatin. These adaptors are usually part of large protein complexes implicated in the chromatin regulation.

PcG (Polycomb Group) family is responsible for both di- and trimethylation of H3K27, controlling gene expression. Most PcG proteins form two major polycomb repressive complexes, PRC1 and PRC2. The first one, PRC1, mediates the monoubiquitylation of histone H2A, impairs transcriptional elongation and repress gene expression. By contrast, PRC2, is responsible for the above described di- and tri-methylation of lysine 27 in histone H3. PcG proteins are epigenetic regulators of transcription with important roles in stem cell regulation, development and differentiation ^{75,76}.

Spatial organization of the chromatin also plays a role in the nucleus by modulating gene expression. Chromosome conformation capture (3C) techniques can assess spatial organization of entire genomes. They have revealed that chromatin is not occupying a random space in the nucleus but is placed in distinct “chromosome territories” (CTs). Gene rich chromosome regions are situated preferentially in the centre of the nucleus, whereas gene-poor regions tend to be near the nuclear lamina. Another method that studies the three-dimensional architecture of whole genomes by coupling proximity-based ligation with massively parallel sequencing, Hi-C, has revealed the existence of different regions within chromosomes that interact with each other. There are transcriptionally active regions with a higher GC content, enriched in open chromatin, genes and active chromatin marks. Other regions are less transcriptionally active and are enriched in the repressive histone mark H3K9me3. Repressed regions are positioned in lamina-associated domains (LADs), so they are localized, preferentially, at the nuclear periphery ⁷⁷. In addition, the genome is divided into topological associated domains (TADs), which are megabase scale chromatin regions that interact more frequently with themselves. TADs are cell specific and are conserved across cell types and species. Enhancers usually contact genes located within TADs. Distant enhancers and their specific gene promoters are in contact one with the other thanks to DNA-loops formed by CTCF and cohesins ⁷⁸.

PTMs and chromatin structure play an important role in bone tissue regulation, controlling osteoblast and osteoclast differentiation. They may also participate in the pathogenesis of osteoporosis ⁷⁹. For instance, the deletion of a TAD boundary has been proposed to cause Liebenberg syndrome, an autosomal dominant upper-limb malformation that shows arms with morphological characteristics of a leg ⁸⁰.

Furthermore, histone modifying enzymes can influence skeletal development, bone mineral density, ossification and bone resorption. As an example, osteocytes secrete paracrine factors that regulate the balance between bone formation and destruction. Among these molecules, sclerostin inhibits the formation of bone matrix by osteoblasts. Wein and collaborators demonstrated that histone deacetylase 5 (HDAC5) downregulates sclerostin levels in osteocytes. Inhibiting HDAC5 with shRNA increases SOST, while the overexpression of HDAC5 decreases the expression of SOST. HDAC5 knockout mice show increased levels of SOST mRNA and sclerostin-positive osteocytes, reduced Wnt activity, trabecular bone density and osteoblast bone formation ⁸¹. This thesis focuses on DNA methylation and long non-coding RNA. Therefore, both epigenetic marks are further explained below.

DNA methylation and bone

DNA methylation is an epigenetic mechanism that consists in the addition of a methyl group onto the C5 position of the cytosine ring to form 5-methylcytosine (5mC). This covalent modification does not take place in any DNA cytosine, but occurs mostly in cytosines that precede a guanine, which is known as "CpG" dinucleotide. In humans, 5mC is found in 80% of all CpG dinucleotides, and the proportion of 5mC is approximately 1% of all DNA bases, which are distributed in the human genome asymmetrically. 5mC is a relatively stable epigenetic mark whose patterns are inheritable during cell division. There are regions with a high density of these dinucleotides, called CpG islands. CpG islands are segments approximately 1000 base pairs long that have a higher CpG density than the rest of the genome and are located in the promoter region of approximately one half of the genes. CpG rich promoter regions are normally demethylated, while the rest of CpGs are methylated. DNA hypermethylation of promoter regions has

been associated with the repression of gene expression. In addition, alterations of these DNA methylation patterns are associated with the appearance of numerous diseases, such as cancer; in which a global hypomethylation takes place, together with a hypermethylation of CpG islands of gene promoters, resulting in the inhibition of tumour-repressor genes ⁸².

DNA methylation is catalysed by a family of enzymes called DNA methyltransferases (DNMTs), which transfer the methyl group from the donor S-adenosyl methionine (SAM) to the fifth carbon of cytosine. There are two types of DNMTs: those involved in de novo methylation (DNMT3A and DNMT3B), and DNMT1, which is responsible for maintaining the methylation pattern during cell division. There is a global passive demethylation of the existing 5mC over time. Furthermore, active demethylation of the 5mC takes place by two mechanisms, deamination and direct demethylation. Deamination by activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex (AID/APOBEC) transform 5mC into thymine and this cause a G/T mismatch, which is corrected by the base excision repair (BER) pathway. 5mC can also be demethylated directly by the action of the ten-eleven translocation (TET) enzymes. TET enzymes oxidize 5mC to form 5-hydroxymethyl-cytosine (5hmC), then another two oxidation steps proceed to oxidize 5hmC to 5-formyl-cytosine (5fC) and then to 5-carboxy-cytosine (5caC). 5hmC could be deaminated by AID/APOBEC, as 5mC, to form 5-hydroxymethyl-uracil. BER pathway uses thymine DNA glycosylase (TDG) to replace 5-hydroxymethyl-uracil, 5fC and 5caC to a non-modified cytosine ⁸³ (Figure 5).

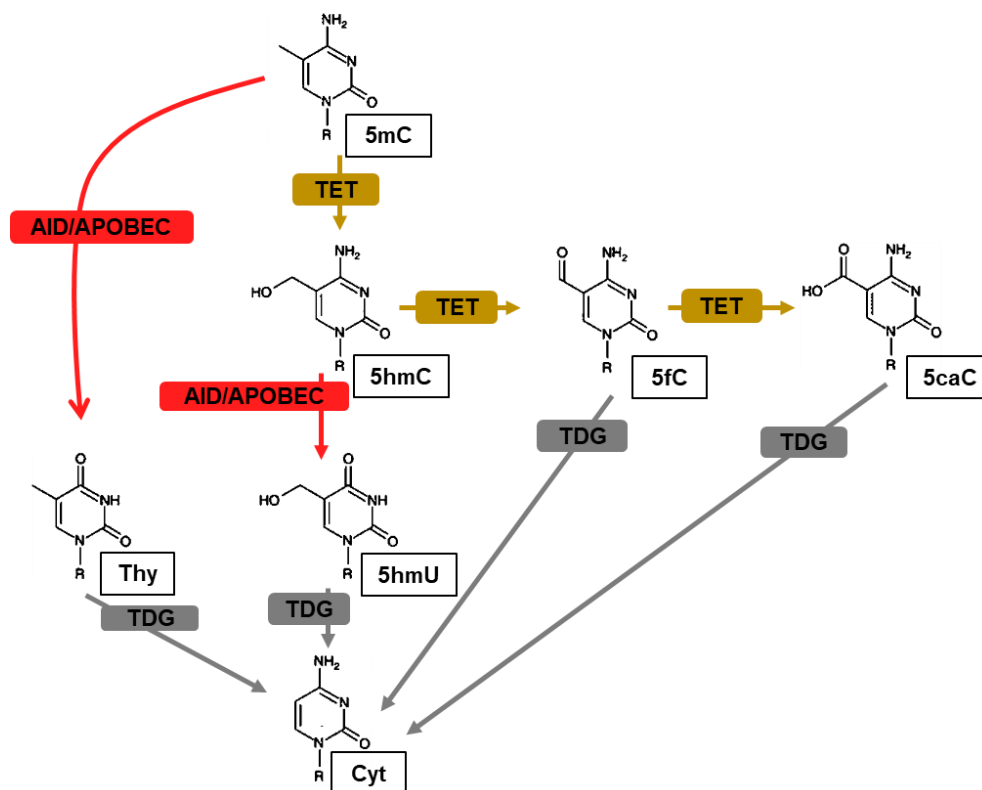


Figure 5. Cytosine cycle. 5mC could take two ways: through oxidation catalyzed by the TET family of oxygenases forming 5hmC, or deamination by members of the AID/APOBEC family and the resulting Thymine (Thy) is excised by thymine-DNA glycosylase (TDG). Then base excision repair process (BER), replace for an unmethylated cytosine (Cyt). 5hmC could be further oxidized into 5fC and 5caC, respectively. 5hmC, 5fC and 5caC may be specially recognized and excised by TDG and subsequent BER activity ⁸³.

DNA methylation is essential for regulating tissue-specific gene expression. But this regulation may be different depending on the localization of these methylated genomic regions. For example, when promoter CpG islands are hypomethylated, gene expression is activated. However, DNA methylation of the gene body or enhancer regions is, in some cases, associated with a higher level of gene expression in dividing cells ⁸⁴.

Aging and age-related diseases have also been related with specific changes in DNA methylation marks. Since 2008 studies have shown that methylation at specific loci in the human genome change with age ⁸⁵. There are some “epigenetic clocks”, but the most widely used is the epigenetic clock by Steve Horvath, which has been developed and can be applied to tissues other than blood ⁸⁶. Horvath developed a multi-tissue predictor of age that allows to estimate the DNA methylation age of several tissues. His software uses a set of 353 CpGs

of the 27k and 450k arrays of Illumina, to characterize the “epigenetic age” and correlate it with the chronologic age. The comparison between both ages permit to establish an accelerated or decelerated age of the tissue studied.

In non-terminally differentiated cells, DNA methylation marks increase progressively in specific genes, thus leading to the loss of the developmental potential and adoption of a more differentiated state. Due to the critical role of DNA methylation in gene expression and cell differentiation, it is obvious that errors in methylation marks can disrupt gene regulation and consequently lead to various diseases. Thus, DNA methylation has a relevant function in bone tissue, skeletal disorders and age-related bone diseases, among others. In bone tissue, DNA methylation has been shown to play an important role in the regulation of genes such as alkaline phosphatase, sclerostin, osterix, estrogen receptor 1, osteopontin, RANKL, osteoprotegerin and leptin, among others ⁵⁰.

A few studies have explored DNA methylation in prevalent skeletal disorders, such as osteoporosis and osteoarthritis. In a previous study of our group, we determined bone DNA methylation profiles (by Illumina 27K array) in 26 patients with hip osteoarthritis (OA) and 27 patients with osteoporotic hip fractures (OP), as well as the transcriptome in pooled samples. The study revealed an inverse relationship between methylation and gene expression in both groups of patients. The comparison of OP and OA bones revealed 241 CpG sites, located in 228 genes, with significant differences in methylation. Of these, 217 were less methylated and 24 more methylated in OP than in OA. The pathways analysis revealed an enrichment in the genes involved in the metabolism of glycoproteins or in cell differentiation, and in the Homeobox superfamily of transcription factors. Likewise, differentially methylated regions were enriched in genes related to cell differentiation and skeletal embryogenesis, which suggests the existence of a developmental component in the predisposition to this disorder ⁸⁷.

Reppe and collaborators explored the epigenetic regulation of SOST expression in bone of postmenopausal women. They measured serum sclerostin and markers of bone remodelling in two groups of post-menopausal women: healthy (T-score of BMD > -1) and established OP (T-score of BMD < -2.5, with at least a typical fracture). They found that the promoter region of the SOST gene showed an increased methylation in OP patients. Free serum sclerostin levels and sclerostin mRNA in bone were lower in patients with osteoporosis. In addition,

after adjusting for age and body mass index, they found an inverse association of sclerostin levels with total bone mineral density. They concluded that the genetic and epigenetic variations of SOST influence gene expression and sclerostin serum levels in postmenopausal women. They also suggested that the increase in DNA methylation of the SOST promoter was a compensatory mechanism, which reduces serum concentrations of sclerostin and reduces the inhibition of Wnt signalling in an attempt to promote bone formation ⁸⁸.

More recently, Cheishvili et al. determined the methylation pattern in blood samples from 22 post-menopausal women with osteoporosis and compared them with 22 postmenopausal women without osteoporosis. The analysis identified 77 differentially methylated CpG sites. They also used the 13 most significant genes to build a weighted score of the DNA methylation of these genes suitable to predict osteoporosis at an early stage with high sensitivity and specificity and correlated it with bone mineral density ⁸⁹. Although those results look quite promising, replication studies in other populations are badly needed to establish their actual relevance and the potential role in clinical practice.

Age-related bone loss is related to an imbalance between bone formation and bone resorption, where the latter is predominant. Although mesenchymal stem cells from the bone marrow (BMSCs) are quite appealing as the precursors of the bone forming cells, there are very few studies of DNA methylation marks in these precursors. Roforth and collaborators analysed RNA and DNA methylation profiles in BMSCs from young people and elderly women ⁹⁰. They identified various cellular pathways associated with aging, some that are known and other that are completely new, thus pointing at new age-related pathways in BMSCs, and osteoblast formation and function.

Expression of non-coding RNAs in bone

The Genetic code is defined as the set of rules by which information encoded in the DNA sequence is translated into 20 aminoacids, which are basic building blocks of proteins. The central dogma of genetic biology had recognized the importance of RNA as an intermediate molecule in this flow of information. At the beginning, transcription and translation processes distinguished between three classes of RNA molecules: messenger RNA (mRNA), transfer RNA (tRNA), and

ribosomal RNA (rRNA). However, the vast majority of DNA does not code for proteins and was previously considered as “junk” DNA, without a known function. The Encyclopedia of DNA Elements (ENCODE) project delineated all functional elements encoded in the human genome and established new functions for the so-called “junk” DNA. ENCODE project concluded that most of those DNA regions, at least 80%, have a biochemical function and many of them are transcribed into RNA. Recent studies estimated that 93% of the human genome is actively transcribed. Among these transcripts, 1% represent protein-coding exons and 39% entire genes including from promoters to poly(A) signal. The rest (54%) are non-coding transcripts ^{91,92}.

RNA molecules that are not being translated, are known as non-coding RNA (ncRNA). They are classified in two main groups. Long non-coding RNAs (lncRNA), which exceed 200 nucleotides in length, and small ncRNAs that are those with less than 200 nucleotides. Small ncRNAs can be further classified into micro RNAs (miRNAs), small interfering RNAs (siRNAs), piwi RNAs (piRNAs), and small nucleolar RNAs (snoRNAs) ⁹³.

Small noncoding RNAs

snoRNAs

SnoRNAs are small noncoding RNAs of 60 to 200 nucleotides, allocated in the nucleolus, associated with a set of proteins forming small nucleolar RNPs (snoRNPs). They are indispensable for the nucleolytic processing of rRNAs and also in the post-transcriptional modification in rRNAs, small nuclear RNAs (snRNAs) and some mRNAs. snoRNAs are not transcribed independently, but are usually processed from other transcribed fragments, as introns from pre-mRNA. There are two major classes of snoRNAs, which have different evolutionarily conserved sequences. C/D box snoRNAs (SNORDs) and H/ACA box snoRNAs (SNORAs), related to the methylation and pseudouridylation of ribosomal RNAs, respectively ⁹⁴. They act as guides for binding RNAs to form a complex where the target is modified by RNA binding proteins (RBP), which act as enzymes. snoRNAs are important for a variety of cell functions and might become useful as disease biomarkers. In fact, distinct changes in snoRNAs have been described in some human cancers. Several studies have explored the detection of snoRNAs in body fluids, during the development of diseases,

including oncological processes, viral diseases, neurodegenerative diseases, and others ⁹⁵. The recent discovery of stable snoRNAs in serum, have increased the interest in their study as biomarkers in cancer. In addition, some studies showed specific snoRNAs associated with aging and osteoarthritic joints. Zhang et. al. examined serum noncoding RNAs as biomarkers for cartilage deterioration after an anterior cruciate ligament (ACL) injury. They analysed serum of 80 patients a year after surgery and 60 serum of control individuals. They found that snoRNAs U38 and U48 were significantly increased in serum of patients developing cartilage damage. Hence, they suggested that serum levels of snoRNA U38 might be useful to aid in the diagnosis of patients with cartilage deterioration after ACL injury ⁹⁶.

piRNAs

Another class of small ncRNAs are PIWI-interacting RNAs (piRNAs). This class of animal specific small ncRNA guide PIWI-clade Argonautes (PIWI proteins). They are 21-35 nucleotides in length and contribute to silence transposable elements, fight viral infection and regulate gene expression. Most argonaute (AGO) proteins associate with miRNAs and siRNAs, but there is a subclass of AGO proteins termed PIWI which are required for germ cell development. piRNAs in mammals have been shown to silence transposons in male germ cells. However, many piRNAs non related with transposon sequences have been also found in mammalian testis, and they are thought to regulate expression of their host mRNA ⁹⁷. PIWI-piRNAs complexes directly target and degrade retrotransposons in the cytoplasm. In addition, they also guide genomic silencing at specific loci through the regulation of DNA methylation marks and specific repressive histone modifications, such as, H3K9me3. In piRNA-deficient mice there exists an increased retrotranscription. PIWI-piRNA is necessary for male fertility ⁹⁸. Apart from the well-known mechanism of piRNAs in spermatogenesis, Rajasethupathy et. al. found that PIWI-piRNA complex facilitates stable changes in neurons for memory tenacity. Specifically, they concluded that this complex ease serotonin-dependent methylation of CREB2 promoter, which inhibits the memory mechanism. This leads to enhanced synaptic processes in memory pathways ⁹⁹.

siRNA and miRNA

There is also a small interfering RNA class (siRNA), characterized by a length of 20-30 nucleotides, which are always associated with the family of Argonaute proteins. siRNA derive from exogenous double stranded RNA (dsRNA) that is cleaved into double stranded siRNA by a dsRNA-specific RNase III family ribonuclease (Dicer). Dicer cleave a double strand sequence of 20-25 nucleotides, each strand is carrying two free bases at the 3' hydroxyl ends and bearing 5' phosphate end. A strand is the guide and directs silencing, whereas the other strand is the passenger and is destroyed. Target silencing by siRNAs is mediated by RNA-induced silencing complex (RISC), which is a family of heterogeneous complexes. This complex has to be at least an Argonaute protein bound to a small RNA, and this is sufficient for target RNA silencing. However, Argonaute proteins may have diverse binding proteins with their own function within complex. After recognition, mRNA is sliced or degraded, breaking the reading sequence of the protein. Therefore, a gene could be silenced by siRNAs-RISC mechanism, preventing protein translation through mRNA degradation ¹⁰⁰. Double-stranded siRNAs are endogenously synthesized by RNA-dependent RNA polymerase (RdRP), but this molecule has been identified in plants and nematodes, but not in mammals. However, synthetic siRNAs have been extensively used for gene silencing in human cells ¹⁰¹.

Nevertheless, the most studied class of small ncRNAs are miRNAs which are very important regulators of gene expression and contribute controlling numerous physiological cell processes. miRNAs are single stranded RNA molecules with about 20 to 23 nucleotides. They also bind to Argonaute proteins to form RISC complex, necessary to post-transcriptional silencing, as occurs with siRNAs, but they differ in their origin and biogenesis. miRNAs are processed in the nucleus and first of all, they need the specific transcription of a double stranded primary miRNA (pri-miRNA) by RNA polymerases II or III. Both polymerases recognize specific promoters differently, promoting a wide variety of miRNA species. Then, pri-miRNA are modified by adenosine deaminases acting on RNAs (ADARs), which transform adenosine into inosine, and this permit to change the sequence and base pairing capacities. Pri-miRNA is transformed into pre-miRNA through the cleavage action of the complex formed by RNase III enzyme Drosha and the enzyme DGCR8 (Drosha-DGCR8). Resulting pre-miRNA is then exported to the

cytoplasm by exportin-5. In the cytoplasm, pre-miRNA is cleaved by Dicer in association with the action of Tar RNA binding protein (TRBP) and protein activator of PKR (PACT). This complex recruits Ago2 to constitute RISC loading complex. The first endonucleolytic action is done by Ago2 activity that cleaves the passenger strand in the middle, and this facilitates Dicer action to form the mature miRNA-Ago2 complex. Finally, RISC mediates mRNA degradation or deadenylation and translational inhibition (Figure 6) ¹⁰². These steps described for miRNA biogenesis are not universal. Some steps can be replaced or modified in various ways. For example, some miRNAs are localized in the cell nucleus, evidencing that these miRNAs are also regulating gene expression within the nucleus.

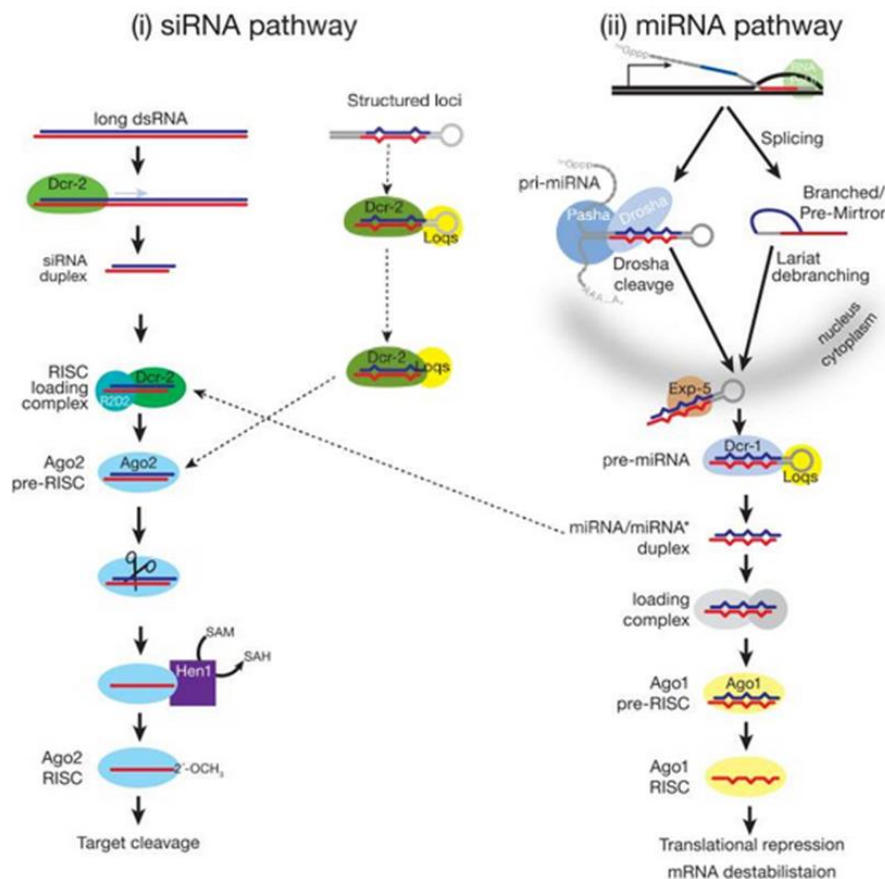


Figure 6. Biogenesis of siRNAs and miRNAs pointing out the differences between both pathways. Adapted from ¹⁰³.

More than 2000 miRNAs have been reported in humans. A half of the human expressed miRNAs are transcribed from non-protein-coding genes and the rest from introns of coding genes. Overall, miRNAs, are estimated to regulate approximately one third of the human genome. Consequently, they have many

functions and participate in the regulation of development, differentiation and metabolism of cells. Thus, deficiencies or excesses of miRNA expression have been associated with different human pathologies including cardiovascular disorders, autoimmune diseases, cancer, skin, muscle and neurologic disorders^{102,104}. In addition, there are diverse studies relating miRNA expression patterns with skeletal diseases, such as osteoporosis. Several miRNAs target molecules implicated in the β -catenin pathway. For example, Let-7f reduces Axin-2 levels, which tends to increase β -catenin and promote osteoblast differentiation¹⁰⁵. Some miRNAs also target different inhibitors of the Wnt signal, including Dkkopf (DKK1), Kremen2, sclerostin (SOST) or secreted frizzled related protein (sFRP). In addition, numerous miRNA inhibit osteogenesis through BMP receptors Smad1 and 5, distal less homeobox 5 (DLX5), the Homeobox family, Nuclear Factor Kappa B (NF- κ B) pathway and the master transcription factors RUNX2 or OSX, among others^{50,106,107}. As for osteoblasts, several miRNAs are involved in osteoclast differentiation affecting RANK/RANKL and Macrophage colony-stimulating factor (M-CSF) pathways¹⁰⁶.

Garmilla-Ezquerro et. al. explored the expression of 760 miRNAs in patients with osteoporotic hip fractures and non-fracture controls with hip osteoarthritis by real-time PCR. After the validation stage, they confirmed statistically significant differences between osteoporosis and controls for two miRNAs (miR-187 and miR-518f)¹⁰⁸.

Therapeutic properties of miRNA have been also studied. Kazuki et. al. identified miRNA-182 as a regulator of osteoclastogenesis and bone homeostasis. They showed how the deletion of this miRNA in ovariectomized mice protects the mouse from excessive bone resorption. In addition, this target also serves to find new pathways involved in bone pathology. In this case they identify PKR protein kinase, which is an inhibitor of osteoclastogenesis via interferon beta regulation, as the target of miRNA-182¹⁰⁹.

Another study about miRNAs and osteoporosis identified specific miRNAs in osteoporotic patients compared to controls with osteoarthritis. Total RNA was extracted from serum. The levels of 179 serum miRNAs were analyzed by real-time PCR, and 12 passed the false discovery rate test for multiple comparisons. After the replication stage, they identified 3 miRNAs (miR-122-5p, miR-125b-5p, and miR-21-5p) as potential biomarkers. Specifically, miRNA-21-5p, showed a

difference between both groups independently of age, and with high discriminative capacity. Additionally, their levels were correlated with those of the bone resorption marker CTx.

miRNAs are of great interest as candidate disease biomarkers, as they are stable in fluids such as serum, can be detected in samples easy to obtain, such as peripheral blood, and in some cases are quite specific for tissue and disease ¹¹⁰.

Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are transcribed RNAs that are not translated into proteins because they do not have a protein coding sequence. They are longer than 200 nucleotides in length. In 1990s first non-coding gene was discovered (H19), but it was not classified as a ncRNA. Brannan and collaborators looked for the protein of the H19 transcript, but they found diverse translation stop signals and 35 small open reading frames. They cloned H19 gene to sequence it and revealed no conserved open reading frame. H19 was not associated with the translational machinery. They concluded that because H19 is transcribed by RNA polymerase II and is polyadenylated, it may be a non-classical mRNA ¹¹¹. Furthermore, in the same decade, early 1990s, X chromosome inactivation in females was associated with X Inactive Specific Transcript (XIST) activity, which is exclusively expressed in the inactive X chromosome. Human XIST include several conserved tandem repeats. It does not contain significant conserved open read frames and does not appear to encode a protein. Therefore, researchers concluded that XIST works as a structural RNA in the nucleus, and demonstrated that XIST is localized in the X inactivation-associated Barr body ¹¹². ncRNA research has revealed that these species are involved in a huge number of cellular processes, regulating gene expression at the chromatin level, transcriptional level, RNA splicing, mRNA degradation, and translational control ^{113,114}.

LncRNAs are less conserved than protein-coding genes as shown in a study looking at the non-coding transcriptome of 185 samples from 11 species (human, chimpanzee, bonobo, gorilla, orangutan, macaque, mouse, opossum, platypus, chicken and frog), which raised questions about their biological relevance ¹¹⁵. Recently, another study showed, by comparing 16 vertebrate species, that more than 70% of lncRNAs do not have homologs in species that diverged more than

50 million years ago. However, thousands of human lncRNAs have homologs in other species. These homologs are conserved in short sequences of the 5' position while the rest of the gene architecture has been restructured mostly by transposable element exonization. They suggest that lncRNA functions require only short conserved fragments of specific sequences and may tolerate changes in the rest of gene architecture ¹¹⁶. Certainly, the most conserved lncRNA type are from long intergenic subfamily, involved in developmental functions. lncRNA transcripts are proposed to arise through several mechanisms, including DNA-based duplication of existing sequence, metamorphosis of protein-coding genes to pseudogenes, transposable element exaptation or exaptation of noncoding DNA ¹¹⁶.

Pseudogenes are an example of protein coding genes metamorphosis that arise from the accumulation of disruptions restraining its potential to be translated into protein. Pseudogene transcription might produce lncRNAs, which is the case of PTENP1, which has been found to be biologically active by regulating coding gene expression, and has also been found to be lost in human cancer ¹¹⁶.

Transposable elements are located throughout the genome and they are often enclosed in noncoding RNAs, more specifically they appear in approximately a 30% of lncRNA sequences. This is the reason why they are considered a major cause of lncRNA development in vertebrates. Transposable elements contribute to signaling for lncRNA biogenesis, including transcription initiation, polyadenylation or splicing ¹¹⁶.

In addition, another mechanism is exaptation of non-coding sequences into lncRNA through recruitment of regulatory elements that permit expression in a region which was previously silent. The testis specific multi-exonic lncRNA Poldi is an example of these de novo origin of lncRNAs, but its transformation mechanism from noncoding DNA into functional lncRNAs is not exactly understood yet ¹¹⁷.

The vast majority of genomic regions used for RNA production remain under-explored. Annotation of the universe of lncRNAs is still at its beginning, and there are different classifications based on their length, genomic locations, properties and functions.

For the purposes of this thesis, we focused on the classification according to lncRNA location with respect to protein coding genes, which is the most widely

used in bioinformatic databases, including the GENCODE/Ensembl annotations (Figure 7). Transcripts were first classified as intergenic or intragenic. Long intergenic non-coding RNAs (LincRNAs) do not overlap with protein coding genes or other ncRNA genes. They are transcribed independently by RNA polymerase II, 5' capped, 3' polyadenylated and spliced. As explained before, there are several LincRNAs with conserved sequences (although rarely the whole sequence), commonly in the 5' position, nested in exons ¹¹⁸. Many of them have been found in the nucleus. That is the case, for example, of LincRNA-p21, which regulates p53-dependent transcriptional responses ¹¹⁹.

On the other hand, intragenic lncRNAs overlap with protein coding genes. They can be further classified as antisense, intronic and sense overlapping lncRNAs. Intragenic lncRNAs can be transcribed in the opposite direction of a protein coding gene promoter, head to head, originating bidirectional lncRNAs. They can originate from the antisense strand of a protein coding gene (Antisense lncRNA), or transcribed from the sense strand of a protein coding gene, thus overlapping introns (intronic lncRNAs) ¹²⁰.

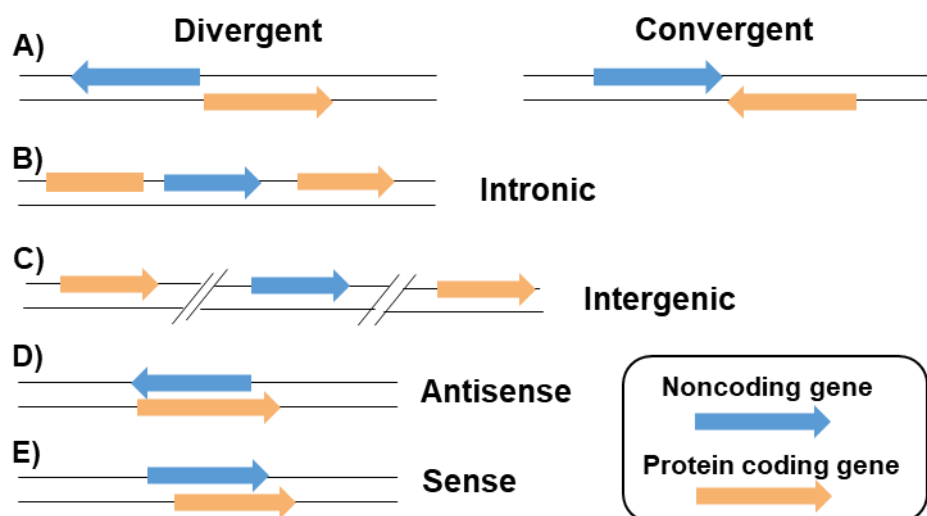


Figure 7. lncRNAs classification: (A) Divergent lncRNA originate from the opposite strand of the same promoter region of the protein coding gene as the adjacent and Convergent lncRNA, encoded on the opposite strand of the protein coding gene and facing each other. (B) Intronic lncRNAs are transcribed in intronic regions of other protein coding genes. (C) Intergenic lncRNAs are located between other two genes, usually >10 kilobases. (D) Overlapping antisense lncRNAs are overlapping on the opposite strand of other protein coding genes and transcribed in the opposite way (complementary sequence). (E) Overlapping sense are overlapping on the opposite strand of other protein coding genes and transcribed in the same direction. Adapted from ¹²¹.

Another common studied lncRNA type are enhancer derived lncRNAs. Numerous transcripts classified as lncRNAs have an effect at the DNA level as enhancers, independently of the lncRNA type according to their genomic location. Enhancer-derived RNAs (eRNAs) are a group of RNAs transcribed by RNA polymerase II and constitute a major type of cis-regulatory elements in the genome ^{122,123}. However, eRNAs function is still poorly understood, and only the function of a few eRNAs, such as FOXC1e and NRIP1e, has been elucidated ¹²⁴.

Some studies have revised lncRNA stability in different species. In mouse, Clark MB and collaborators determined the half-life of approximately 800 lncRNAs and 12000 mRNAs, by inhibiting ongoing transcription with actinomycin D. They found that a minority of lncRNAs are unstable. Their half-lives vary over a wide range, with some lncRNAs having half-life of less than 30 minutes and others showing extreme stability with a half-life over 48 hours. The median lncRNA half-life was 3.5 h, whereas the median half-life for protein-coding transcripts was 5.1 h. So, lncRNAs are not particularly unstable, although their half-lives are shorter than those of protein-coding RNAs. When investigators analyzed different types of lncRNAs, they realized that intergenic and bidirectional lncRNAs are more stable than intronic lncRNAs. Regarding cellular localization, nuclear lncRNAs are more unstable ¹²⁵. Similar results were obtained by Ayupe and collaborators in human cells. They used a microarray to investigate biogenesis, processing, stability, conservation and cellular localization of approximately 6000 intronic lncRNAs and 10000 antisense lncRNAs. Antisense lncRNAs (median $t_{1/2}$ = 3.9 h) were more stable than mRNAs (median $t_{1/2}$ = 3.2 h), whereas intronic lncRNAs (median $t_{1/2}$ = 2.1 h) comprised a more heterogeneous class with both stable ($t_{1/2}$ > 3 h) and unstable ($t_{1/2}$ < 1 h) transcripts ¹²⁶.

lncRNA expression patterns are extremely specific of cell or tissue type and state, so they have a spatio-temporal specificity, usually higher than protein coding genes. Few lncRNAs are expressed in a wide variety of tissues, many are present in a few tissues, and some are expressed only in a single tissue type. On the other hand, many protein coding genes are much more frequently expressed in a variety of different tissues. Not only lncRNAs exhibit cell or tissue specificity, but their expression is state and time dependent, displaying a highly dynamic expression. Ubiquitous lncRNAs, this is, lncRNAs that are expressed in various tissues, are often more abundant, whereas tissue-specific lncRNAs tend to have

low expression levels ¹²⁷. Brain and testis are complex tissues with a high amount of specific lncRNAs. Particularly, brain is the richest source of lncRNAs in the body. In general, the high lncRNAs gene expression specificity suggests regulatory roles of these molecules in various biological processes, including normal or pathological development ¹²⁸.

lncRNAs: mechanisms of action

The evolution of lncRNAs from transcriptional noise to functional regulators of gene expression has opened a whole new field of study. Despite many studies demonstrating the important functions of some lncRNAs, the actual biological role of many lncRNAs is unknown yet.

The mechanisms involved in the regulatory pathways of lncRNAs are multiple. They regulate chromatin structure through histone modifications. From the first documentation of lncRNA on mammalian X chromosome inactivation, as described above, XIST has been associated with Polycomb Repressive Complex 2 (PRC2) recruitment, which mediates methylation of lysine 27 in histone 3 ¹²⁹. However, numerous lncRNAs have been related with PRC2 or PRC1 actions or other histone modifying enzymes, including lncRNAs HOTAIR, lincRNA Pint, FENDRR, SRA, HOTTIP, FAL1 and ANRIL among others ¹³⁰. Some of these lncRNAs have been shown to bind more than one histone-modifying complex. For example, HOTAIR physically associates with PRC2 and with Lys-specific demethylase 1 (LSD1). However, the detailed mechanism of how lncRNAs target specific DNA regions, remains unclear.

Apart from modulating chromatin, lncRNAs regulate transcription through multiple mechanisms. They can interact with the transcriptional machinery directly or can regulate the activity of transcription factors. For instance, lncRNA Evf2 can act either as an activator or repressor, depending on whether it recruits the transcriptional activator DLX2 or the transcriptional repressor MeCP2 (methyl-CpG binding-protein 2) to specific DNA regulatory elements ¹³¹.

Antisense lncRNA, as the name suggests, are transcribed from the opposite strand (antisense strand). The host gene in the 'sense' strand may be a protein coding gene or a noncoding one. However, most commonly in the mammalian genome is a non-protein-coding antisense gene with a protein-coding gene in the opposite strand. It is estimated that more than 20% genes in eukaryotes are

antisense transcripts. Antisense lncRNAs can act in a cis manner, when they are complementary to their sense partner, or trans-manner, when they interact with transcripts from a different, distant, locus. Interestingly, cis-acting antisense transcripts regulate the expression of their sense partners, either in a discordant or concordant manner. Thus, depending on the particular coding-noncoding pair, antisense knockdown results in sense-transcript increase or decrease ¹³².

The ability of lncRNAs to bind to proteins provide them multiple regulatory capacities. They may act as decoys that preclude the access of regulatory proteins to DNA. For example, after DNA damage five lncRNAs are promoted from the CDKN1A gene promoter, and one of them is induced in a p53-dependent manner (lncRNA PANDA). PANDA interacts with the transcription factor NF-YA to inhibit the expression of pro-apoptotic genes. Additionally, if PANDA is depleted, human fibroblasts are sensitized to apoptosis by doxorubicin ¹³³.

Moreover, lncRNAs can function as scaffolds to bring together different elements of protein complexes. The telomerase RNA (TERC) is an example of an RNA working as a scaffold that assembles the telomerase complex ¹³⁴. Another example of lncRNA scaffold is, as described before, HOTAIR, which binds both PRC2 and LSD1, and these interactions induce H3K27 methylation and H3K4me2 demethylation, causing gene silencing. An additional example is ANRIL, which combines interactions with PRC2 and PRC1 ¹³⁵.

Finally, another lncRNA-protein binding function is as guides. Many lncRNAs are required for the proper localization of specific protein complexes. LncRNAs involved in imprinting, such as XIST, serve as guides to target gene silencing activity. LincRNA-p21 is directly induced by p53 upon DNA damage, and it physically associates with nuclear factor hnRNP-K to drive this protein to specific promoters ¹³⁶.

Numerous lncRNAs are exported to the cytoplasm, where they regulate mRNA translation and stability. An important regulation function of lncRNA in the cytoplasm, is as miRNA sponges. They act as competing endogenous RNAs (ceRNAs) with miRNA response elements, so that this lncRNA titrate competing miRNAs (Figure 8) ¹³⁷.

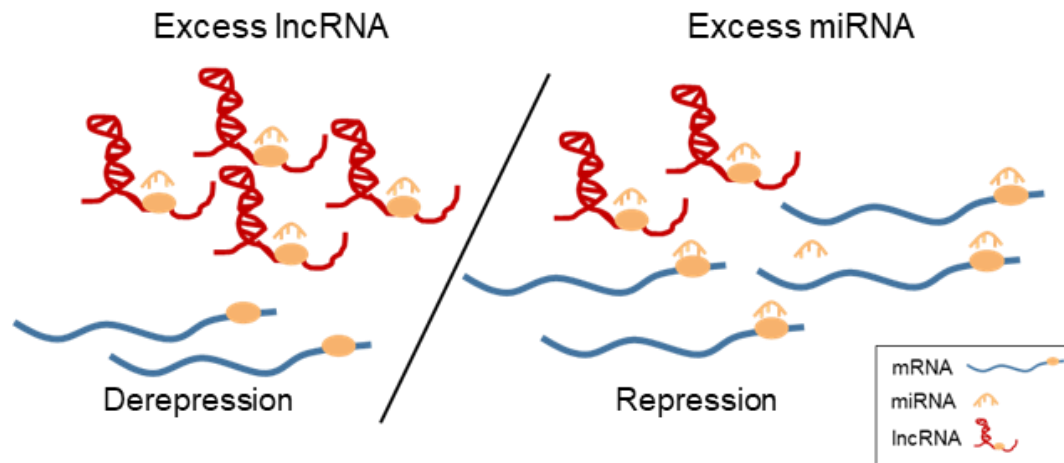


Figure 8. Networks of mRNA/miRNA/lncRNA, competing endogenous RNA (ceRNA), showing the regulation of lncRNAs as sponges. Excess of lncRNAs in comparison with miRNAs levels result in derepression of mRNA translation, whereas, excess of miRNAs compared to lncRNAs leads to mRNA translation and protein synthesis. Adapted from ¹³⁷.

Enhancer-derived lncRNAs (eRNAs) participate in the interactions between enhancers and promoters through chromosome looping ¹³⁸. Actually, there are three current views on the functional implications of enhancer transcription:

- a. Enhancer transcription represents noise; it is due to the Pol II action over open and accessible chromatin;
- b. Enhancer transcription effects are independent of RNA synthesis. Rather, they are related to chromatin remodelling induced by RNA polymerases and the associated proteins that provoke histone tail modifications and other changes.
- c. eRNAs are functional molecules involved in transcriptional control either locally or at distant locations ¹³⁹.

lncRNAs in bone tissue and aging

Several investigators have explored the role of lncRNAs in aging and related phenotypes ^{140,141}. This association of lncRNAs and aging is illustrated by their regulatory effects over aging-related genes or pathways. In this line, lncRNAs have been related to almost all the molecular hallmarks of aging ^{43,141}, including stem cell state, telomere stability, DNA damage, proteostasis, intercellular communication, cellular senescence and epigenetic alterations.

Telomeres are DNA-protein complexes that protect the ends of chromosomes against damage and instability; their length is inversely related with DNA

replication and cell senescence. This means that telomeric shortening is associated with a high rate of proliferation eventually causing cellular senescence. The length of telomeres is regulated by the association of telomerase RNA component (TERC) and telomerase reverse transcriptase (TERT), as well as, the lncRNA telomeric repeat containing RNA (TERRA). lncRNA TERC plays a role in the maintenance of telomere length, working as a template and scaffold between telomeric repeats and the proteins that form the telomerase complex ¹⁴². On the other hand, lncRNA TERRA is a competitive inhibitor for telomeric DNA and suppresses telomere elongation. It contains 5'-UUAGGG-3' repeats which are complementary to telomeric sequences to suppress elongation ¹⁴³.

As already discussed, epigenetic mechanisms control gene transcription and chromatin structure, and are implicated in a large variety of processes concerning aging and disease. DNA methylation changes with aging and cellular senescence. Several lncRNAs regulate DNA methylation (XIST, H19, PAPAS, Airn or TARID among others) ¹⁴¹. However, it is important to keep in mind that the information flow is dual, this means that DNA methylation can also regulate the expression of various lncRNAs implicated in aging and/or disease. Similarly, some specific post-translational histone modifications have also been associated with aging such as lncRNA PINT, which activates genes of the p53 pathway by directly interacting with PRC2 for repression ¹⁴⁴. Thus, other lncRNAs regulate DNA damage response. lncRNA RoR is a p53 repressor, which inhibits p53 translation to protein due to its interaction with the heterogeneous nuclear ribonucleoprotein I (hnRNPI) ¹⁴⁵. Additionally, antisense non-coding RNA in the INK4 locus (ANRIL) is involved in the repression of INK4B-ARF-INK4A locus, tumor suppressor genes that trigger the anti-proliferative functions of p53 and Retinoblastoma protein (RB) ¹⁴⁶.

Human aging is characterized by a low-grade chronic inflammatory state, sometimes referred to as 'inflammaging', which may be an important factor contributing to the pathogenesis of age-related diseases. Therefore, it is important to explore pathways controlling the inflammation process in order to benefit elderly people ¹⁴⁷. In this regard, some lncRNAs have been related to the inflammation process and its regulation. TNF α - and hnRNPL-related immunoregulatory lincRNA (THRIL) interacts with hnRNPL to regulate the

expression of many immune response genes like specific cytokines and transcriptional regulators of TNF α expression. Another lncRNA, p50-associated COX-2 extragenic RNA (PACER), has been characterized as an inflammation regulator. PACER interacts with the repressive subunit of NF- κ B p50, activating p65/p50 dimers, and this block NF- κ B from the COX2 promoter^{148,149}.

Several lncRNAs have a significant role in stemness regulation and aging. They regulate directly the level of transcription factors or participate within the network of reprogramming processes. This is confirmed with the association of pluripotency transcription factors (Oct4, Sox2 or Nanog) to specific lncRNAs promoters, suggesting their direct relationship in stemness preservation^{150,151}. Likewise, Mesenchymal-to-epithelial transition (MET) is an important cell event in cells undergoing reprogramming. E-cadherin expression and Zeb2 downregulation augments MET efficiency. lncRNA Zeb2-AS, which is a natural antisense transcript of Zeb2, controls its expression and consequently improves reprogramming processes¹⁵².

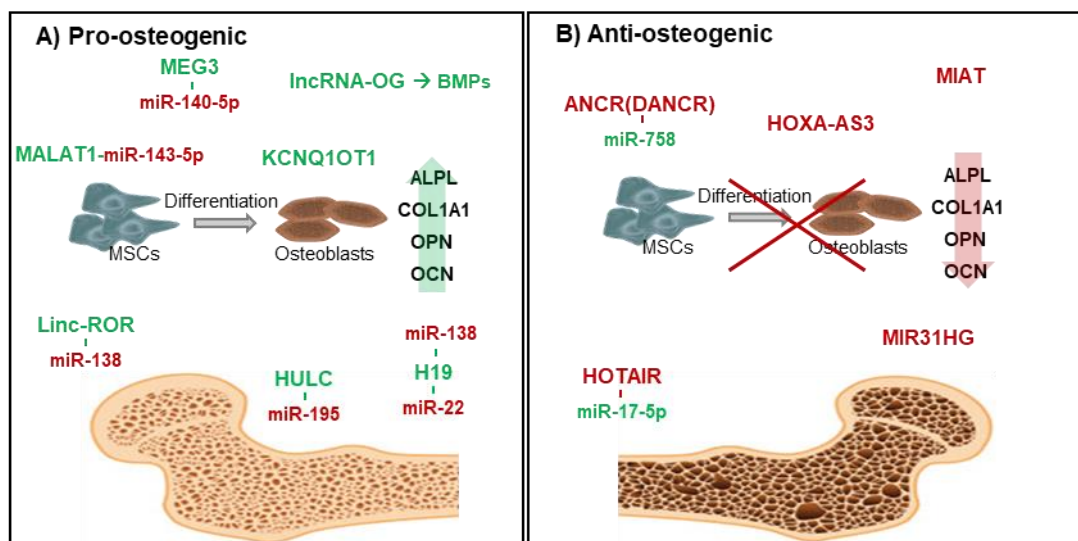


Figure 9. lncRNAs classified by their osteogenic properties. They could be pro-osteogenic (A) or anti-osteogenic (B), this is, they can promote or inhibit osteogenesis, respectively. lncRNAs are key transcriptional and translational regulators. They usually work with other regulators, such as miRNAs. The pro-osteogenic molecules are coloured in green and anti-osteogenic molecules in red. Adapted from¹⁵³.

Rapid advances in genome sequencing and transcriptome analysis have settled lncRNAs as significant intermediaries in gene regulation. In recent years, a number of studies have pointed at lncRNAs as molecules playing a role in the regulation of bone metabolism and remodeling, and consequently in the

pathogenesis of skeletal disorders. Also this opens a window to the discovery of new drug targets. Different lncRNAs may have a positive or negative impact on osteogenic differentiation processes, as reviewed by Silva AM and collaborators¹⁵³ (Figure 9).

Some anti-osteogenic lncRNAs are DANCR, HOTAIR, MIR31HG and HOXA-AS3, among others. Anti-differentiation noncoding RNA (ANCR), also denominated differentiation antagonizing non-protein coding RNA (DANCR), have been shown to promote osteogenesis when downregulated. DANCR is associated with Enhancer of Zeste Homolog 2 (EZH2), which results in the inhibition of RUNX2 expression inhibition, thus blocking osteoblast differentiation¹⁵⁴. In addition, DANCR regulates osteogenic differentiation of human BMSCs via the p38 MAPK pathway¹⁵⁵. However, DANCR downregulation in dental tissue-derived stem cells has been shown to promote not only the osteogenic differentiation, but also adipogenic and neurogenic differentiation. This indicate that DANCR is a very important regulatory factor in different stem cell differentiation processes, not exclusively in the differentiation towards bone cell lineage¹⁵⁶. Likewise, HOXA-AS3 expression has been shown to play a role in the switch between adipo- and osteogenic differentiation¹⁵⁷. This lncRNA is upregulated in adipogenic differentiation, but unaltered in the osteogenic one. Moreover, silencing HOXA-AS3 in BMSCs suppresses adipogenesis and adipogenic markers like PPARG, CEBPA and ADIPOQ. On the contrary, this silencing effect in BMSCs enhanced osteogenesis and the expression of osteogenic markers such as RUNX2, SPP1, SP7 and COL1A1¹⁵⁷.

On the other hand, several lncRNAs promote osteogenesis, including lncRNA-OG, MALAT1, MODR, AK141205, KCNQ1OT1, HULC, LINC-ROR and MEG3¹⁵³.

Specifically, lncRNA MALAT1 is able to promote the expression of Osterix, by inhibiting miRNA-143 in BMSCs. When the expression of MALAT1 is inhibited in these cells, the relative expression of Osterix also decreases. However, when it is co-transfected with an inhibitor of miRNA143, the relative expression of Osterix increases again. Osterix expression is linked to the expression of ALP (alkaline phosphatase) and other osteogenic markers¹⁵⁸.

A more frequently studied lncRNA gene is maternally expressed gene 3 (MEG3), which is paternally imprinted, and acts controversially as a pro-osteogenic

lncRNA. MEG3 knockdown in MSCs from multiple myeloma patients reduces the expression of RUNX2, osterix and osteocalcin, whereas promoting MEG3 expression enhanced the expression of these osteogenic markers. Mechanistically, MEG3 dissociates the transcription factor SOX2, which is a repressor, from the Bone morphogenetic protein 4 (BMP4) promoter and this permits to activate its expression. Those results suggest that MEG3 plays an essential role in the osteogenic differentiation of BMSCs, from patients with multiple myeloma, by indirectly activating BMP4 transcription ¹⁵⁹. In line with the positive effect of MEG3 on osteogenesis, another study showed that MEG3 is downregulated during adipogenic differentiation, but upregulated with osteogenic differentiation of human adipose-derived mesenchymal stem cells (hASCs). Apparently, MEG3 acts via miR-140-5p, which is upregulated during adipogenesis and downregulated during osteogenesis, thus with an inverse correlation ¹⁶⁰. However, the actual effect of MEG3 on bone is controversial. The microRNA miR-133a-3p has been related with an abnormal osteogenic differentiation. In a study of human BMSCs, there was a positive correlation between MEG3 and miR-133a-3p expression; and both RNAs were down-regulated after osteogenic differentiation. Thus, these data suggested that MEG3 regulates miR-133a-3p expression, and inhibits osteogenesis of BMSCs from post-menopausal women ¹⁶¹. The reasons explaining those contradicting results are unclear, but may depend on the patients' characteristics, because lncRNA expression is rather tissue- and time-specific, as already discussed.

Tang et. al. found that a novel osteogenesis-associated lncRNA (lncRNA-OG) is upregulated during osteogenic differentiation of BMSCs, and functionally lncRNA-OG promotes this process. This lncRNA interacts with heterogeneous nuclear ribonucleoprotein K (hnRNPK) protein to regulate bone morphogenetic protein signaling pathway ¹⁶².

Although the role of lncRNAs in osteoporosis remains largely unexplored, several studies suggested that lncRNAs are implicated in other skeletal disorders, such as rheumatoid arthritis or osteoarthritis ¹⁶³. Recently, a small study suggested that some lncRNAs are differentially expressed in osteoporosis ¹⁶⁴. They performed a RNA sequencing to obtain the expression profile in blood samples of postmenopausal osteoporotic patients (diagnosed as defined by the World Health Organization criteria) and controls without osteoporosis. They found 5

differentially expressed lncRNAs, including LINC00963, LOC105376834, LOC101929866, LOC105374771 and LOC100506113, which may be involved in the pathogenesis of postmenopausal osteoporosis, by regulating the expression of other five co-expressed, differentially expressed, mRNAs (ALPL, SOCS3, ADM, SLPI and CD177) ¹⁶⁴.

4. OBJECTIVES

HYPOTHESIS

Osteoporosis is a bone disease characterized by low bone mass and a higher risk of fracture. Bone remodelling involves a regulated activity of osteoclasts and osteoblasts, responsible of bone resorption and bone formation, respectively. Bone marrow stem cells (BMSCs) are the precursors of osteoblasts and, consequently, are needed to maintain an adequate bone formation. Epigenetic mechanisms play an important role in cell physiology, the specification of the gene expression pattern and cell differentiation. Among epigenetic mechanisms, DNA methylation and certain long non-coding RNAs (lncRNAs) have been postulated as important regulators of the differentiation and activity of BMSCs. Therefore, the general hypothesis underlying this project is that DNA methylation patterns and the specific expression of certain lncRNAs by BMSCs are involved in the pathogenesis of osteoporosis, by influencing the ability of BMSCs to differentiate into osteoblasts and therefore to maintain bone mass.

OBJECTIVES

The main objective is to study differential epigenetic marks (specifically DNA methylation and expression of lncRNAs) by the BMSCs of patients with two prevalent skeletal disorders, namely osteoporosis and osteoarthritis, which tend to present with opposing changes in bone mass. This general objective translates into the following specific objectives.

- 1- To analyse the methylation pattern of BMSCs of patients with hip fractures and controls with osteoarthritis, and their relationship with the expression of genes involved in bone metabolism.
- 2- To explore the epigenetic aging of BMSCs of patients and controls.
- 3- To determine the gene expression signature of BMSCs from patients with osteoporotic fractures in comparison with controls
- 4- To characterize the lncRNA signature of BMSCs from patients and controls, as well as its relationship with the DNA methylation patterns
- 5- To explore the changes of the lncRNAs signature of BMSCs after their differentiation towards the osteoblastic phenotype in vitro.

5. METHODS

METHODS

Bone samples and bone marrow mesenchymal stem cell (BMSC) isolation and culture

Bone tissue samples were obtained from the femoral head of patients undergoing hip replacement surgery due to osteoporotic hip fractures or hip osteoarthritis. The study was approved by the institutional review board (Comité de Ética en Investigación Clínica de Cantabria) and all donors gave informed consent. Patients with secondary osteoporosis, high-impact fractures or secondary osteoarthritis were excluded. Cylinders of trabecular bone were extracted with a trephine, after removing the subchondral and subfracture edges. They were washed with 50 mL of phosphate-buffered saline (PBS) to obtain a cell suspension. Cells were subjected to a Ficoll density gradient. Then, cells at the interface, between the aqueous phase and the Ficoll-Paque layer, were cultured on polystyrene culture flasks in Mesencult™ MSC Basal media completed with 10% of Mesenchymal Stem Cell Stimulatory supplements (Stem Cell Technologies®, Vancouver, Canada)¹⁶⁵. In some experiments, cells were maintained in Dulbecco's Modified Eagle Medium (DMEM®, Merck KGaA, Darmstadt, Germany) with high glucose 4.5g/L, phenol red and L-glutamine. Media were supplemented with 10% of Fetal bovine serum (FBS) (Merck KGaA, Darmstadt, Germany) and 1% Penicillin-Streptomycin (Merck KGaA, Darmstadt, Germany).

When cultures attained 80% confluence, cell differentiation experiments and nucleic acid isolation were carried out. Only cells of the two first passages were used.

In addition, the trabecular bone were cut into approximately 1 cm fragments, snap-frozen into liquid nitrogen, and stored at -80°C until nucleic acid isolation from bone cells⁸⁷.

Representative samples of BMSCs were characterized by staining for surface markers of BMSCs in a FACSCanto II flow cytometer (Becton Dickinson, New Jersey, USA). Results were analyzed using the FACSDiva 6.1.3 software. The antibodies used were: CD45 labelled with peridinin chlorophyll protein complex (PerCP), CD34 labelled with fluorescein isothiocyanate (FITC), CD73 labelled

with allophycocyanin (APC), CD90 labelled with phycoerythrin (PE) and CD105 labelled with violet blue. All the antibodies were purchased from Miltenyi Biotec (Bergisch Gladbach, Germany).

Osteogenic differentiation of BMSCs

BMSCs were plated at a density of 100,000 cells/well in a 6 well plate and kept in culture for 2 days to reach confluence. Then, osteogenic induction medium was added. This medium was composed of DMEM supplemented with 100 nM of dexamethasone, 50 μ M of ascorbic acid and 10 mM of glycerol 2-phosphate (all from Sigma-Aldrich, which now is Merck KGaA, Darmstadt, Germany). The medium was changed twice a week for up to 3 weeks. At several time points, cells were used for DNA extraction, RNA extraction or staining with alizarin red (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Alizarin red staining procedure was used to measure the calcium deposition into cell matrix. Cell cultures were fixed with ethanol 70% for 1 hour, washed with distilled water three times and then exposed to alizarin red 2% solution (pH 4.2) for 10 minutes. Finally, stained cells were washed with distilled water three times and dried at room temperature. The staining was semiquantitatively evaluated by 2 independent observers who were blind of the culture origin ¹⁶⁵.

Proliferation analysis

The proliferation status of BMSCs was assessed by immunocytochemistry using an anti Ki-67 antibody (Thermo Fisher Scientific, Waltham, Massachusetts, USA), a nuclear protein associated with cell proliferation. Briefly, cells were permeabilized with 0.5% Triton X-100 in PBS for 15 minutes. After this step, cells were thoroughly washed with PBS three times. A final wash with PBS-Tween 20 0.05% was performed for 5 minutes. Cells were incubated in the same buffer in the presence of a Ki67 antibody at 4° for 12 hours. After the incubation with the antibody, cells were washed as previously. Secondary rabbit antibody labelled with fluorescein isothiocyanate (FITC) was added to the cells and incubated for 45 minutes at room temperature. After incubation with the secondary antibody, cells were washed again three times with PBS and once with 0.5% Triton X-100

in PBS for 5 minutes. After mounting, the proportion of Ki67-positive cells was determined under a fluorescence microscope.

The results were confirmed by a cell proliferation colorimetric assay based on the reduction of the tetrazolium dye MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) to insoluble formazan. Cells were plated in triplicate on a 96-well plate (3×10^3 cells per well) with their normal growth medium. On days 3, 5, 7 and 9, cells were treated with 1 mg/mL of MTT at 37° for 5 hours. Next, the dye was extracted with 100 μ L Isopropanol to solubilize the formazan for 10 minutes at 37°C. Optical density was measured at a wavelength of 550 nm.

Cell lines cultures

For some parallel experiments and replications several cell lines were used. They included human osteoblastic cell lines, such as HOS, SAOS-2 and MG-63 (osteosarcoma derived cells), and non-osteoblastic cell lines, such as HEK-293T (derived from human embryonic kidney cells). All cell lines, excepting HOS, were cultured with DMEM (Merck KGaA, Darmstadt, Germany) with high glucose 4.5g/L, phenol red and L-glutamine, supplemented with 10% FBS (Merck KGaA, Darmstadt, Germany) and 1% Penicillin-Streptomycin to control bacterial contamination (Merck KGaA, Darmstadt, Germany). Human osteoblastic cell line HOS was maintained in culture with Eagle minimum essential medium (MEM, Merck KGaA, Darmstadt, Germany), supplemented with 10% FBS and 1% antibiotics. All cells were kept and grown in a humidified incubator, with 5% of CO₂ and 37°C of temperature ¹⁶⁶.

DNA isolation

Genomic DNA from cell cultures was obtained after cell lysis with Proteinase K (0.2 mg/mL) and a lysis buffer (Tris-HCl 2 M, EDTA 0.5 M, sodium acetate 0.3 M and 10% of SDS). The mix was incubated for 1 hour in a water bath at 56°C. DNA isolation was performed with Phenol:Chloroform:Isoamyl alcohol (Thermo Fisher Scientific, Waltham, Massachusetts, USA), as previously reported ⁸⁷. Pellet of DNA was cleaned with 75% ethanol and resuspended with 50 μ L of distilled water. All centrifugation steps were done at 4°C to avoid DNA degradation.

On the other hand, to extract DNA from the frozen bone fragments, they were mechanically homogenized, with a Polytron homogenizer (IKA, Staufen,

Alemania), in 1mL sterile PBS over ice. After this step, the same procedure was used to isolate DNA.

To determine DNA concentration, we used spectrophotometric analysis (DeNovix, Wilmington, USA), which is based on the axiom that nucleic acids absorb ultraviolet light, and then Lambert Beer law is employed to calculate concentrations without standard curves. Nucleic acids absorb at a wavelength of 260 nm and the quantity of absorption is proportional to DNA/RNA quantity. For more accurate quantification, the Qubit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) technique was used. Qubit is an instrument that uses fluorescent dyes to determine DNA, RNA or protein quantification.

RNA extraction and purification

RNA extraction from cell cultures was performed with 1mL TRIzol™ Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, USA) for the cell lysis, following the instructions provided by the manufacturer. Summarising, after the lysis, the separation phase was done with 0.2 mL of chloroform and RNA precipitation from the aqueous phase with isopropyl alcohol. This precipitation may be improved with an incubation of an hour in the freezer. The RNA pellet was then cleaned with 1 mL cold 75% ethanol. At the end of the procedure, RNA pellet was dried and resuspended in RNase-free water by leaving it a few hours on ice. It is important not to let the pellet dry completely because that would prevent its proper dissolution. All centrifugation steps were done at 4°C as with DNA procedure.

A similar procedure was used to extract RNA from bone tissue samples. Frozen bone fragments were mechanically homogenized with a polytron as with the DNA procedure, but in this case into 2 mL of TRIzol™ and then the extraction procedure was followed as described above.

For some experiments, RNA samples were further cleaned and purified. RNA was purified with the RNA Clean & Concentrator kit (Zymo Research, Irvine, California, USA), which is a RNA clean up kit with an efficient procedure for the preparation of high-quality DNA-free RNA. The procedure uses a spin column technology for the recovery of total RNA, small RNAs and long RNAs. Over the column, the procedure consists of different steps, including loading with binding

buffer, ethanol wash and elution of pure RNA. DNase I (provided in the kit) treatment was carried out for complete elimination of DNA contamination.

To determine RNA concentration, we used spectrophotometric analysis (DeNovix, Wilmington, USA), and for more accurate quantification, the Qubit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) technique was used.

Genome-wide DNA methylation analysis

BMSCs at first passages were used for DNA methylation experiments. Thus, 22 osteoporotic fractures (FRX) and 17 osteoarthritis (OA) samples were used for the analyses of DNA methylation (patient age: 62 to 95 years). DNA was extracted as explained before with Phenol:Chloroform:Isoamyl alcohol and bisulfite-converted before genome-wide analysis of methylation with the Infinium Human Methylation450 BeadChip array (Illumina®, San Diego, CA, USA) at the Spanish “Centro Nacional de Genotipado” (CEGEN-ISCI).

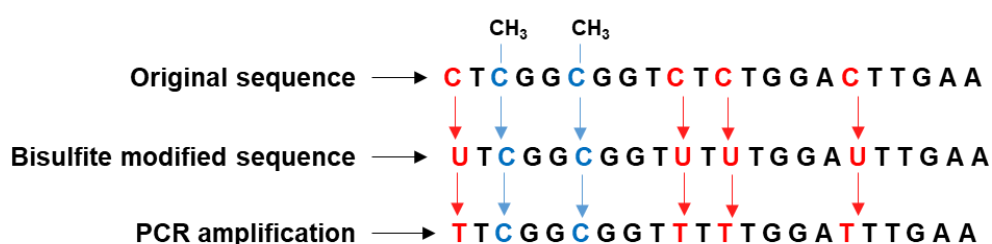


Figure 10. Bisulfite conversion scheme on a random sequence, showing how this conversion works. In the original sequence there are cytosines (coloured in red) and methyl cytosines (coloured in blue). After the bisulfite conversion, the methylated cytosines (blue) still remain as cytosines (blue), whereas the unmethylated cytosines (red), are transformed into uracils, which are read as thymine after the PCR reaction.

Bisulfite conversion was performed with EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, California, USA) with 1 µg of DNA per sample. Sodium bisulfite is able to deaminate cytosines into uracil, but this does not occur in methylated cytosines. Hence, after DNA deamination with bisulfite, unmethylated cytosines are converted into uracil that after a polymerase chain reaction (PCR) are read as thymine, but 5-methylcytosines are still cytosines, so depending on the thymine/cytosine ratio we can calculate the DNA methylation levels in the original sample (Figure 10). EZ DNA Methylation-Gold Kit protocol was provided by the

manufacturer and is briefly explained below. In brief, 1 µg of DNA is diluted in 20 µL of water and mixed with 'CT Conversion Reagent' and then incubated 10 minutes at 98°C and 2.5 hours at 64°C. The samples are loaded in a spin column and are consequently washed, desulphonated, washed and finally eluted into a new clean tube.

Infinium Human Methylation450k BeadChip array targets more than 450,000 CpG methylation sites with a high genome coverage and can analyse twelve samples in parallel. This array uses two different types of chemical assays (Infinium types I and II assays), both are designed to distinguish Cytosine and Thymine, the two species generated after DNA bisulfite conversion. Infinium I uses two types of probes, one hybridizes the methylated allele (Cytosine) and the other the unmethylated allele (Thymine), next base extension is the same for both alleles. Infinium II assay uses only a single probe for both alleles, and in this case, base extension depends on the methylation state with different two colours, so it could generate A-T signal (colour 1) or C-G signal (colour 2) ¹⁶⁷ (Figure 11).

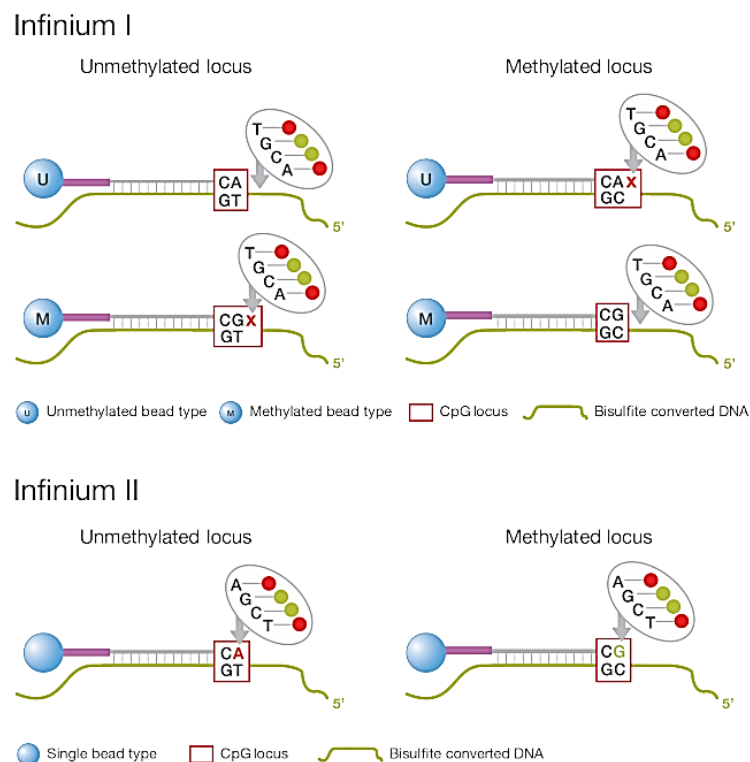


Figure 11. Infinium types I and II assays. Infinium I uses two types of probes (U and M), one hybridizes the methylated allele (M) and the other the unmethylated allele (U). Next base extension is the same for both alleles. Infinium II assay uses only a single probe for both alleles. base extension could be unmethylated A-T signal (red) or methylated C-G signal (green) ¹⁶⁷.

Human Methylation 450k array included 485,577 hybridization sites, from which 482,421 were CpG sites, 3091 non-CpG sites and 65 random SNPs. This coverage is distributed along the genome and covered almost all protein coding genes and the known CpG islands. Specifically, the array covers 21,231 RefSeq genes (99%), 26,658 CpG islands (96%), 26,249 CpG island shores (92%), 24,018 CpG island shelves (86%), 62,600 Hidden Markov Model-defined CpG islands, 9426 high CpG content FANTOM 4 promoters, 2328 low CpG content FANTOM 4 promoters, 80,538 in silico-identified enhancers, and other regions of potential interest, like DNase hypersensitive sites, Ensemble regulatory features and almost all HumanMethylation27k loci, ¹⁶⁸ (previous designed methylation array).

Due to the two probe types of this array, there are some type II bias during data processing. Therefore, several softwares have been developed for array data normalisation. Likewise, a number of analysis pipelines have been suggested. Most software packages are freely available and written in R language (methyumi, minfi, watermelon, ChAMP, RnBeads, and others) ^{169–172}. Relevant steps of the 450k analysis pipeline are: import raw data files (IDAT files), probe filtering, background correction, adjustment for type II bias, batch effect analysis and correction, and identification of differentially methylated sites and regions ¹⁷³. DNA methylation levels are commonly expressed as β value, which ranges between 0 (no methylation) to 1 (full methylation). We analysed our data using R/Bioconductor package RnBeads ¹⁶⁹. This package permits to analyse DNA methylation data from human 450k array comprehensively, including import data, filter data, batch correction and identification of differentially methylated sites and/or regions. Differentially methylated sites were considered as significant with a false discovery rate (FDR) value < 0.05 and an absolute difference in methylation higher than 10% ($\Delta\beta > 0.10$). We also used an analysis module from RnBeads that identifies differentially methylated regions (DMRs) when comparing groups of CpGs. So, we did not only look at single CpGs for differential analysis, but also for pre-defined genomic regions, such as CpG islands, promoters, genes, and enhancers. DMRs analysis increases the statistical power to detect differential DNA methylation, and it also facilitates the interpretation of identified DMRs.

Genomic Regions Enrichment of Annotations Tool (GREAT) was used to analyse the functional significance of differentially methylated enhancer regions by association with both, proximal and distal putative target genes. GREAT also uses gene annotations from numerous ontologies to calculate statistical enrichments in gene ontology terms or metabolic pathways for such associations¹⁷⁴. Further details of the bioinformatic analyses are given below.

DNA methylation age predictor

DNA methylation age, also denominated as epigenetic age, was calculated with Steve Horvath's multi-tissue predictor of age, which permits to estimate the epigenetic age of several tissues and cell types. Epigenetic age was calculated from the methylation level of a set of 353 CpGs, which has been shown to change with aging in a wide variety of tissues¹⁷⁵. Epigenetic age allows to compare ages of different tissues from the same subject and it may identify tissues with evidence of an altered epigenetic age due to disease. We used our DNA methylation data from BMSCs (n=39, from which FRX=22 and OA=17) above described, and also bone DNA methylation results (n=39, from which FRX=20 and OA=19), previously obtained in our lab⁸⁷. BMSCs and bone DNA samples were not paired because they were obtained at different times and projects.

Horvath's software was implemented with R function as described in his tutorial (additional file 20, from¹⁷⁵). The relationship of the epigenetic age and the chronological age was explored by linear regression analysis.

Relative telomere length analysis

The relative telomere length was analyzed using a quantitative PCR procedure employing β -globin as a control gene and the telomere primer sequences proposed by Cawthon¹⁷⁶:

Primer Name	Sequence
Telomere forward	CGGTTTGTTTGGGTTTGGGTTTGGGTT
Telomere reverse	GGCTTGCCTTACCCTTACCCTTACCCT
β-globin forward	GCTTCTGACACAACGTGTGTTCACTAGC
β-globin reverse	GGCTTGCCTTACCCTTACCCTTACCCT

Table 1. Primers for Telomere and β -globin polymerase chain reaction.

DNA from BMSCs was quantified with a fluorometric assay (Qubit®). Telomere and β -globin amplification reactions were performed in quadruplicate in 96-well plates in a BioRad iCycler iQ™ Real-Time PCR detection system. The reaction mix contained Sybrgreen master mix (Applied Biosystems, Foster City, California, USA), 6 ng of DNA and the corresponding primers (100 nM forward and 900 nM reverse, for telomeres; 300 nM forward and 700 nM reverse for β -globin). The amplification protocol for telomeres included an initial denaturation step at 95°C for 10 minutes, 35 cycles of amplification at 95°C for 15 seconds and a hybridization step of 54°C for 2 minutes. The β -globin amplification protocol included an initial denaturation step at 95°C for 10 minutes, followed by 35 cycles of amplification at 95°C for 15 seconds and 58°C for 1 minute¹⁷⁶. In each experiment, 3 control samples were run to normalize the results and to compensate for inter-assay variability. The relative telomere length of the samples was then estimated as $2^{\Delta\Delta Ct}$. ΔCt for each unknown sample and control was estimated as (Ct β -globin - Ct telomere). Average ΔCt of the 3 control samples was calculated. Then $\Delta\Delta Ct$ was calculated as (ΔCt unknown sample - ΔCt controls).

DNA methylation analysis by pyrosequencing

The DNA methylation level of selected CpGs identified as differentially methylated by the methylation array was replicated by pyrosequencing with the PyromarkQ24 Advanced System (Qiagen N.V., Hilden, Germany). Pyrosequencing is a high resolution quantitative method that allows to determine the C/T ratio in the CpGs analysed (Figure 12)¹⁷⁷.

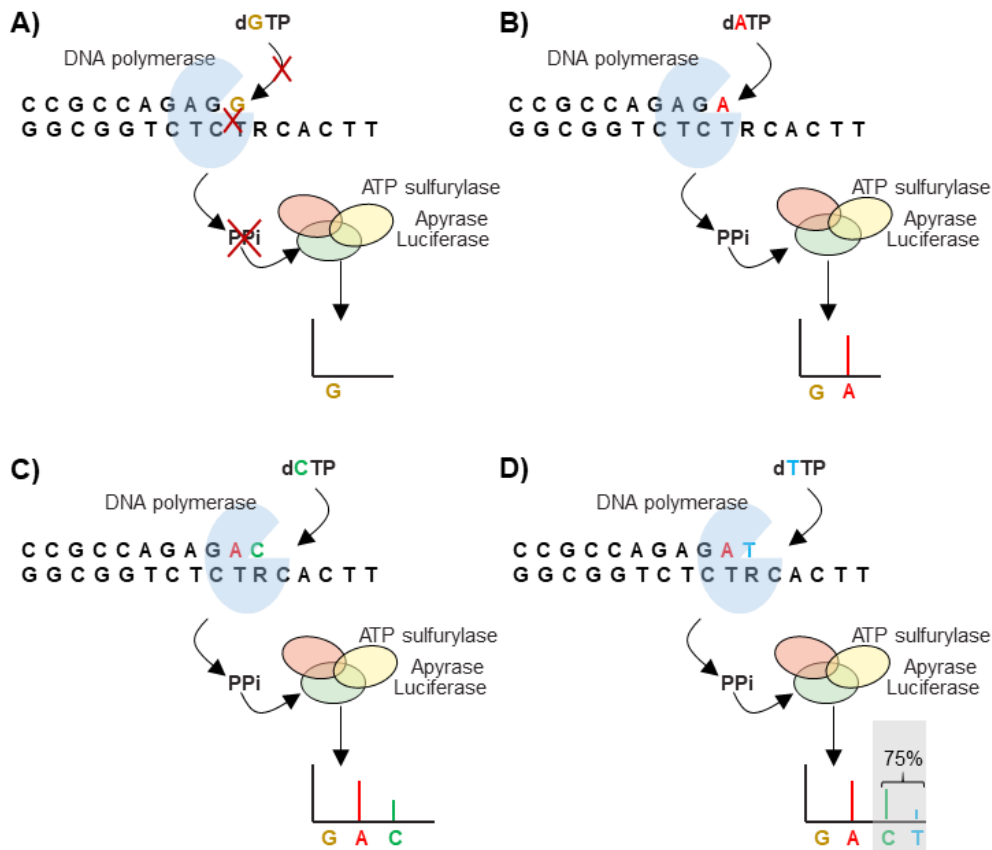


Figure 12. Enzymatic reactions of pyrosequencing in the example of a random bisulfite-treated sequence. (A) dGTP is added as nucleotide for extension, but it is not complementary to the sequence, therefore no PPi is released and the nucleotide is degraded by the apyrase. (B) dATP is added and it is complementary to the template so that PPi is released and used by the ATP sulfurylase and Luciferase enzymes generating a signal which level is proportional to the available PPi. (C) The 'R' in the template sequence corresponds to a CpG with variable methylation levels. Thus, after bisulfite treatment and PCR, either cytosine or thymine can be incorporated. First, dCTP is added and a signal is released. (D) dTTP is added and another light signal is released; in this case three times lower than with cytosine, so that the DNA methylation level in this CpG correspond to 75%.

Primers used for PCR amplification and sequencing were designed with the PyroMark assay design program (Qiagen N.V., Hilden, Germany), Table 2. Assay design software ensures that primers hybridize in a methylation-independent manner. Sodium bisulphite modification of 1 µg genomic DNA was performed as described above. Bisulphite-converted DNA was eluted in 20 µL, using 1 µL for each PCR. Primers sequences were designed to hybridize with CpG-free sites to ensure methylation-independent amplification.

PCR was performed with a biotinylated primer, which permit its purification in a single-stranded DNA template, using the PyromarkQ24 Vacuum Workstation (Qiagen N.V., Hilden, Germany) (according to manufacturer's instructions). Finally, pyrosequencing reactions and methylation quantification were performed in a PyroMark Q24 Advance System ¹⁶⁶. The statistical significance of the differences between osteoporotic and osteoarthritic patients was tested by t-test when normal distribution exists. Instead, when normal distribution does not exist, non-parametric Mann-Whitney U tests was used. All reported p values are two-tailed and with a significance threshold of 0.05.

Name	Sequence	(Tm)
ID2-Fw	5'-GTTTTTAGGTTTGGGGGAGAA	60°
ID2-Rev	5'-[Btn]CTCAACCCCTTCCTCTC	
ID2-Seq	5'-AAAGTTTAGGAGGGAG	
OPG-Fw	5'-TAGTGGTTTTGTTTTTTGATTAAATTTTG	62°
OPG-Rev	5'-[Btn]CAATCACACACATTATTTTCCTAACAA	
OPG-Seq	5'-TTGTTTTTTTGATTAAATTTTGA	
UNC5B-Fw	5'-GTGAAGGGTTTTTGTTTAATAGTTAGAT	60°
UNC5B-Rev	5'-[Btn]AAAAAAAACCTTAAACACC	
UNC5B-Seq	5'-GAGATTTATAATAGATTGAGGTA	
LOXL2-Fw	5'-GATTAGTATTTAGAATAATTGGGTAAGTGT	59°
LOXL2-Rev	5'-[Btn]AAATCATCTAATTTTTTTCCCCTACA	
LOXL2-Seq	5'-TTGAATTTAATTTTTTTTAGTTGTG	

Table 2. Primer sequences of four gene candidates with differential DNA methylation (ID2, UNC5B, OPG and LOXL2) for pyrosequencing analysis. The biotinylated primer is marked with [Btn]. The third column shows the melting temperature of the PCR.

Transcriptome analysis

As with genome-wide DNA methylation analysis, in the case of the transcriptome, BMSCs at first passages were used for RNA isolation. We used two sets of samples that were subjected to high-throughput RNA sequencing RNA (RNA-Seq). The samples were collected and analysed at different times both from practical reasons and to ensure biological and technical replication.

Thus, the first RNA-Seq experimental set consisted in BMSCs isolated from 10 osteoporotic fractures (FRX) and 10 BMSCs from patients with osteoarthritis

(OA). After extraction, RNA samples were extracted as described before, and then quantified and studied for their integrity with an Agilent RNA Screen Tape analysis. RNA degradation information is given with an RNA Integrity Number (RINe) from 0 to 10, and RINe is recommended to be higher than 7 for sample quality control in sequencing workflows. Samples were then prepared using the NEBNext Ultra Directional RNA Library Protocol (Illumina, San Diego, California, USA) and were sequenced on an Illumina Hi-Seq 2000 sequencer (NTX-Dx, Gent, Belgium). Library preparation was with ribosomal RNA depletion, paired end, stranded and in 100 base pairs fragments. Total paired end reads were between 10 and 30 million, excepting a sample with 46 million, all of them with a mapping percentage higher than 80 percent.

The second set of samples used for RNA-seq included 17 predifferentiated BMSCs samples (9 FRX and 8 OA). Additionally, we analysed 6 differentiated BMSCs samples, after three weeks of osteogenic differentiation in vitro (Table 3). In this case, library preparation was done with the kit TRuseq stranded total RNA Ribo-Zero (Illumina, San Diego, California, USA) for ribosomal RNA depletion, 100 base pairs fragments, paired end and stranded. Sequencing instrumental was the new Novaseq System (Illumina, San Diego, California, USA), with at least 30 million reads per sample (Macrogen, Seoul, South Korea).

Lab code	Biopsy	Disease	Sex	Differentiation State	Age	RNAseq
JAR1	JAR1	FRX	woman	PREDIFF	79	1 st
JAR2	JAR2	FRX	woman	PREDIFF	82	1 st
JAR3	JAR3	FRX	woman	PREDIFF	82	1 st
JAR4	JAR4	FRX	woman	PREDIFF	86	1 st
JAR5	JAR5	FRX	woman	PREDIFF	80	1 st
JAR6	JAR6	FRX	woman	PREDIFF	86	1 st
JAR13	JAR13	FRX	woman	PREDIFF	74	1 st
JAR14	JAR14	FRX	woman	PREDIFF	73	1 st
JAR15	JAR15	FRX	woman	PREDIFF	92	1 st
JAR19	JAR19	FRX	woman	PREDIFF	87	1 st
JAR7	JAR7	OA	woman	PREDIFF	72	1 st
JAR12	JAR12	OA	woman	PREDIFF	62	1 st
JAR17	JAR17	OA	woman	PREDIFF	67	1 st
JAR20	JAR20	OA	woman	PREDIFF	67	1 st
JAR22	JAR22	OA	woman	PREDIFF	73	1 st

Lab code	Biopsy	Disease	Sex	Differentiation State	Age	RNAseq
JAR23	JAR23	OA	woman	PREDIFF	73	1 st
JAR24	JAR24	OA	woman	PREDIFF	82	1 st
JAR16	JAR16	OA	woman	PREDIFF	66	1 st
JAR8	JAR8	OA	woman	PREDIFF	80	1 st
JAR9	JAR9	OA	woman	PREDIFF	67	1 st
JARM5	2	FRX	man	PREDIFF	79	2 nd
JARM9	5	FRX	woman	PREDIFF	77	2 nd
JARM12	7	FRX	woman	PREDIFF	91	2 nd
JARM21	9	FRX	woman	PREDIFF	84	2 nd
JARM24	10	FRX	man	PREDIFF	67	2 nd
JARM32	15	FRX	woman	PREDIFF	68	2 nd
JARM38	18	FRX	woman	PREDIFF	76	2 nd
JARM46	20	FRX	man	PREDIFF	95	2 nd
JARM6	3	OA	man	PREDIFF	87	2 nd
JARM7	4	OA	woman	PREDIFF	63	2 nd
JARM10	6	OA	woman	PREDIFF	72	2 nd
JARM25	11	OA	woman	PREDIFF	71	2 nd
JARM26	12	OA	woman	PREDIFF	72	2 nd
JARM29	13	OA	man	PREDIFF	65	2 nd
JARM50	21	OA	woman	PREDIFF	73	2 nd
JARM51	22	OA	man	PREDIFF	68	2 nd
JARM1	1	FRX	woman	PREDIFF	88	2 nd
JARM2	1	FRX	woman	21	88	2 nd
JARM11	6	OA	woman	21	72	2 nd
JARM43	19	OA	woman	21	75	2 nd
JARM54	23	FRX	woman	14	89	2 nd
JARM53	7	FRX	woman	14	91	2 nd
JARM14	7	FRX	woman	21	91	2 nd

Table 3. Biopsies information of the RNA samples sent to the first and second RNAseq. Column 6th, 'differentiation state', shows the status of BMSCs in terms of osteogenic induction in vitro. 'PREDIFF' refers to the basal state of BMSCs and the numbers '14' and '21' refer to the days with osteogenic differentiation media.

In both experimental sets, the analysis started with a general quality control test of the raw sample reads using FastQC software from Babraham Institute. FastQC looks at the quality of all reads from a sample and gives information for the next pre-processing steps. Pre-processing step was used to eliminate low quality

bases, reads or other artifacts, such as, adapters or library derived sequences. Adapter dimers, corresponding to the first 15 nucleotides, were trimmed with Fastx trimmer tool (FASTQ/A Trimmer). This tool is used for shortening reads in FASTQ files to remove barcodes or noise (http://hannonlab.cshl.edu/fastx_toolkit/). These trimmed samples were then used to align reads to mammalian-sized genomes.

Sample reads were mapped to the human consortium reference genome build 37 (GRCh37), also known as hg19, by using TopHat tool. TopHat is “a fast splice junction mapper for RNA sequenced reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra-high-throughput short read aligner Bowtie, and then analyses the mapping results to identify splice junctions between exons”¹⁷⁸. TopHat uses Bowtie alignment method that permits to analyse and detect alternative splicing in the align sequences. After mapping with TopHat the following pipeline was followed, as described by Cole Trapnell and collaborators¹⁷⁸, and summarized below.

Mapped reads were assembled to transcriptome using two different ways for the final differential expression analysis. One way uses Cufflinks suite of tools that assemble transcripts, estimates abundances and tests for differential in all samples. Cufflinks suite is composed by several software tools, such as cufflinks, cuffcompare, cuffquant, cuffnorm and cuffdiff (<http://cole-trapnell-lab.github.io/cufflinks/manual/>). Cufflinks assembles reads into transcripts and quantifies their expression. Cuffcompare associates each assembled sample to known transcripts, employing a specific reference transcriptome. In this case gencode V19 gtf file was applied as a comprehensive gene annotation on the reference chromosomes (www.gencodegenes.org/human/release_19.html). Cufflinks and cuffcompare output files can be used with cuffdiff to compare expression levels of genes and/or transcripts between sample groups. Cuffdiff shows up- and down- regulated genes comparing our two conditions (FRX and OA), and also which genes have differentially spliced or differentially expressed isoforms. However, this step is very computationally expensive, therefore, it is recommended to use cuffquant, which save gene and transcripts profiles to files that are easier to analyse with cuffdiff and cuffnorm. Cuffnorm was employed to obtain the normalized expression levels of all our samples.

Secondly and processed in parallel, those aligned files with TopHat were sorted with Samtools sort option, and then HTSeq-count tool was used to quantify the expression of the samples in all genes ¹⁷⁹. Finally, HTSeq-count output files were examined for differential expression between our groups with two free R packages (DESeq2 and EdgeR) ^{180,181}.

The overrepresentation of genes with differential expression in different cell pathways (Wikipathways) was obtained from the output of Web-based gene set analysis toolkit (WebGestalt) software ¹⁸², which incorporates information from different public resources. The Gene Ontology enrichment analyses from the common terms between differentially methylated enhancers and differential expression were done with ArrayTrack software ¹⁸³. Further details of the bioinformatic analysis are given below.

Real Time PCR analysis

Gene expression was validated by Reverse Transcription followed by a real time quantitative polymerase chain reaction (RT-qPCR). Complementary DNA (cDNA) was synthesized with the TaKaRa kit PrimeScript RT (TaKaRa, Shiga, Japan). We used 1 µg of RNA as template, random hexamers and oligo-dT as primers, with manufacturers quantities and conditions, which are 2 µL of their buffer and random hexamers; and 0.5 µL of enzyme and oligo-dT; 1 µg of RNA in 5 µL of sterile and RNase free water. This reaction mixture per sample is incubated at 37° for 15 minutes (reverse transcription) and 85° for 5 seconds (enzyme inactivation). Then complementary DNA (cDNA) in 10 µL was diluted four times, up to a final volume of 40 µL. Transcript abundance of messenger RNAs was assessed by RT-qPCR using commercially available Taqman assays (Thermofisher Scientific, Waltham, MA, USA) on an Applied Biosystems 7300 Real-Time PCR System.

However, for long non-coding RNAs (lncRNAs) analysis, new primers had to be designed, by using the “RealTime qPCR Assay Entry” IDT online tool (<https://eu.idtdna.com/scitools/Applications/RealTimePCR/>). For intergenic lncRNA reverse transcription was performed with the same protocol as above. However, a different protocol was established for analysing the antisense type of lncRNAs. In that case, reverse transcription must be strand specific to distinguish between both strands. Strand specific reverse transcription was carried out with

the same kit without random hexamers and oligo-dT as primers, but with sense primers for the antisense strand selection, or vice versa. Specific primers for reverse transcription were used with a concentration of 2 μ M in the final mix and the reverse transcription reaction took place at 50°, instead of 37°, to improve the results by increasing the specificity.

Assays used for the RNA sequencing validation, other bone related genes and lncRNA genes are reflected in Table 4. Housekeeping genes GAPDH and RPL13A were used for normalization in BMSCs, and GAPDH and TBP for cell lines. Threshold cycles (Ct) are the amplification cycles at which the fluorescence threshold was reached, and they were estimated for the target and housekeeping genes. The average of the housekeeping genes was used as control for normalization. Thus, the relative expression of target genes was estimated by the $2^{\Delta C_t}$ method, where ΔC_t = average Ct housekeeping – average Ct target gene. For statistical results, the data collected from real time qPCR was subjected to Shapiro-Wilk normality test. Pearson correlation coefficient was used for the correlation analysis when the data were normally distributed. When the data were not normally distributed or the relationships were not linear, Spearman rank correlation method was employed.

Gene	Assay / Sequence
ALPL	Taqman Hs00758162_m1
BGLAP	Taqman Hs01587814_g1
COL1A	Taqman Hs00164004_m1
FOXP2	Taqman Hs00362818_m1
GAPDH	Taqman Hs99999905_m1
IBSP	Taqman Hs00173720_m1
ID2	Taqman Hs04187239_m1
IGFBP4	Taqman Hs01057900_m1
LAMC1	Taqman Hs00267056_m1
LASP1	Taqman Hs00196221_m1
LOXL2	Taqman Hs00158757_m1
OPG	Taqman Hs00900360_m1
OSX	Taqman Hs00541729_m1
PPARG	Taqman Hs01115513_m1
RPL13A	Taqman Hs04194366_g1

Gene	Assay / Sequence
RUNX2	Taqman Hs00231692_m1
SLC5A3	Taqman Hs00272857_m1
SPARC	Taqman Hs00234160_m1
SPP1	Taqman Hs00959010_m1
UNC5B	Taqman Hs00900710_m1
PAPPA-AS1 Fw	CCCACCCAACAACAACAATAAC
PAPPA-AS1 Rv	GCCTCAGTAGGTAGACACAAAC
CTB-51J22.1 Fw	AGCTTAGGGATGGTGGGAATTG
CTB-51J22.1 Rv	CCCTGCCCACTAAATGCTTAT
LINC00341 Fw	CAATACGCAGAGGGACCATATC
LINC00341 Rv	TCCAATACTGCTTGCCTTCC
LINC01279 Fw	GGAGGCGTGGTAAAGGTATATG
LINC01279 Rv	AATCCCACTGCCCTTATCTTG
LINC012008 Fw	GGTTCCATCCAGCCCAATAA
LINC012008 Rv	CTACAGGTCAACACTGCGATAG
CTD-2541J13.1 Fw	ACAGCGGCAATCCCTAAA
CTD-2541J13.1 Rv	GTTCTCCTTACTCATCCCTCAC
PACERR Fw	CCCTCTCCTCCCCGAGTTCC
PACERR Rv	CAGGGCCGCTCAGATTCCTG

Table 4. Primers and assays.

Transfection analysis with siRNA and expression vectors

For transfection analysis in both BMSCs and cell lines, 50,000 cells in 0.5 ml medium were seeded in each well of a 24 well plate, so that in the next day they were between 70-90% confluent. Then, we prepared the transfection mix with lipofectamine 3000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA), as recommended by the manufacturer. The mixture is prepared with 1.2 µl of lipofectamine 3000; the material to be transfected at the desired concentration (for DNA, 500 ng); 1 µl of P3000 Reagent and 50 µl of Optimem (Gibco®, Thermo Fisher Scientific, Waltham, Massachusetts, USA) reduced serum media, which is ideal for use during cationic lipid transfections. After 5 minutes of incubation, the mixture was added dropwise on the centre of the well. In the case of small interfering RNAs (siRNAs) of interest, the transfection mix included 50 µL of

Optimem with 2.75 μ L of lipofectamine and 50 nM of siRNA, according to the manufacturer's recommendations.

siRNAs were obtained from Guangzhou RiboBio Co., Ltd. (Guangzhou, China), lyophilized. For each gene, three tubes with different target sequences were obtained, and a negative control (scramble sequence). Each of the three tubes of siRNA for the gene to be silenced, were mixed at equal volumes, obtaining a final concentration of 20 μ M.

We acquired siRNA for ID2, with three different target sequences given in Table 5.

Gene	Sequence
genOFFTM st-h-ID2_001	CGATGAGCCTGCTATACAA
genOFFTM st-h-ID2_002	GGACTCGCATCCCCTATT
genOFFTM st-h-ID2_003	TCAGCATCCTGTCCTTGCA

Table 5. Total of three different sequences of the siRNA of ID2 obtained.

For each transfection, different controls were implemented, some wells without transfection mix, and some negative silencing controls with the scramble sequence. After cell transfection, RNA was extracted at different time points with Trizol as described above and gene expression was determined by RT-qPCR. For each transfection well, the cells were lysed with TRIzol™ Reagent for further RNA isolation and gene expression analysis.

Plasmid cloning, purification and transfection experiments.

Mammalian expression cloning vectors pcDNA3.1(+) were ordered with our sequences of interests under a cytomegalovirus (CMV) promoter (General Biosystems, Morrisville, USA). The expression vectors included the sequences of two human genes, PACERR (NR_125801.1) and LINC00341 gene (NR_026779.1), respectively. Both plasmids contained two antibiotic resistance genes, beta-lactamase gene that confers resistance to ampicillin, and other beta-lactams; and aminoglycoside phosphotransferase from Tn5 gene that confers resistance to kanamycin, neomycin and G418 (Geneticin). Figure 13.

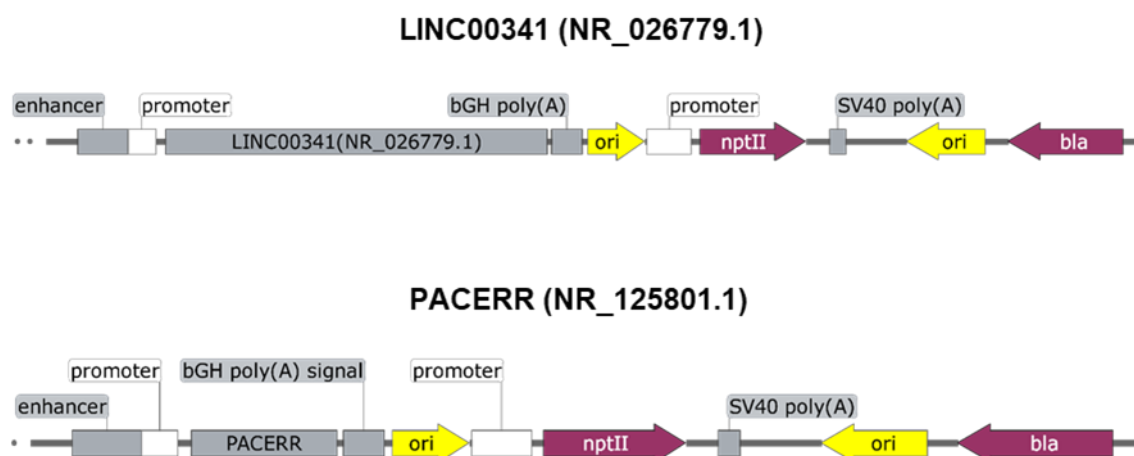


Figure 13. Mammalian expression vectors with the CMV promoter previous to the gene sequence (PACERR and LINC00341, independently). Both plasmids have beta lactamase gene that confers resistance to ampicillin and aminoglycoside phosphotransferase gene from Tn5 that confers resistance to neomycin and kanamycin.

Plasmids arrived lyophilized and were dissolved in nuclease free water at a final stock concentration of 100 ng/ μ L. After reconstitution, stock solution was stored at -20°C for long term storage. A working solution of 1 ng/ μ L was prepared, and 10 μ L of the working solution were used to transform the plasmids into 50 μ L of *E. coli* DH5 α competent cells (Thermo Fisher Scientific, Waltham, Massachusetts, USA). They were gently mixed by flicking the bottom of the tube with the fingers and incubated on ice for 30 minutes. Then each tube was placed in a water bath at 42°C for 20 seconds and then put back on ice for 2 minutes. Then, 950 μ L of Luria-Bertani (LB) Broth media (Thermo Fisher Scientific, Waltham, Massachusetts, USA) were added, without antibiotic, and maintained at 37°C in a shaking incubator for 1 hour. This step allows the bacteria to produce the antibiotic resistance proteins encoded by the plasmid so that they will be prepared to grow in an antibiotic containing agar plate. Two LB agar (Thermo Fisher Scientific, Waltham, Massachusetts, USA) plates with ampicillin were prepared, per plasmid, to plate different volumes of the transformed outgrowth in LB. Plates were incubated at 37°C overnight.

Only the cells that contained the plasmid will be able to grow in the LB agar with ampicillin and to form colonies. Single colonies were removed with a sterile pipette tip, deposited into 3 mL of liquid LB and maintained at 37°C for 8 hours in a shaking incubator. Next, large overnight culture was prepared from each starter culture, using 1:1000 dilution in a flask with 50 mL of LB Broth medium with

ampicillin and incubated at 37°C. Plasmids were purified from cultures with the Nucleobond Xtra Midi Plus kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany), using high copy plasmid purification protocol, as described by the manufacturer. Plasmid concentration was determined by UV spectrophotometry and its integrity by agarose gel electrophoresis.

Plasmids were used for transfection experiments on HEK293T, HOS, MG63 cell lines and BMSCs, in a 24 well plates. Lipofectamine 3000 was used for transfection as explained above. Lipofectamine 3000 and plasmid DNA proportions were 1 µL of Lipofectamine for 0.5 µg of DNA. Transfection efficacy was assessed by gene expression analysis by RT-qPCR.

Bioinformatic tools and statistics

For “wet-lab” experiments standard univariate statistical tests were used (Pearson and Spearman for correlation analysis; t-tests and Mann-Whitney tests for pair-wise comparisons. Therefore, this sub-section is divided into two parts, the first one about the methylation analysis and the second one about the transcriptome analysis. In general, the false discovery rate (FDR) method ¹⁸⁴ was used to minimize the inflation of type I error due to multiple comparisons.

Methylome data

DNA methylation results from Infinium DNA methylation 450k array, were analysed, as explained above, with RnBeads software, which consists in an R package for comprehensive analysis of DNA methylation data. RnBeads pipeline contain different analysis modules that can be run automatically by only giving some annotation characteristics, or they can also be run separately ¹⁶⁹. DNA methylation analysis modules used for this thesis comprehend:

- Data Import. In this case data is imported as typically Infinium formats (IDAT format files) and normalized by default with Subset-quantile Within Array Normalization (SWAN) method ¹⁸⁵.

- Quality control. RnBeads identifies technical and biological biases including assay failures and batch effects.

- Preprocessing. Quality control results determine which samples or CpG sites are susceptible to filter out due to low-quality with the objective of reducing risk of misleading results. RnBeads removes probes overlapping SNPs that are

interfering in DNA methylation levels, and additionally, discards samples and/or probes previously included in the normalization, but not in the final analysis because of too much missing values or without any variability.

-Tracks and Tables. Preprocessed data is exported in several formats to use for example in data visualization in genome browsers or to use in other software to analyse further. Moreover, a table is created summarising simple statistics such as the final number of probes (CpG sites), genomic regions involved and number of CpG sites per region type.

-Exploratory Analysis. This module explores sample subgroups, methylation profiles in terms of CpG sites and regions, and explores associations with sample annotations. It shows through several density plots the genomic regions covered in the dataset, their size distributions, number of sites per region and distribution of sites across different regions. It also implements two methods for dimension reduction, principal component analysis (PCA) and multidimensional scaling (MDS). Methylation value distributions are assessed based on sample groups on probe and region levels. And finally, this module also creates clustered samples on heatmaps using Euclidean distance metric per genomic region.

-Differential DNA methylation. DNA methylation changes can be analysed at the CpG sites level or across genomic regions. Combined estimations in larger genomic regions increase statistical power and may arise in more interpretable results. On the site level, p-values were calculated using limma's software method that uses hierarchical linear models with the DNA methylation M-values, which distribution is more appropriate for this statistical model than the beta values. β -values range from 0 to 1, generally used to measure DNA methylation percentage. Whereas, M-values are calculated as the log2 ratio of the intensities of methylated probe versus unmethylated probe and almost range from -6 to 6

185.

With this assumption the difference in mean methylation levels of both groups (FRX and OA) were compared and a t-test assessing if methylation values from this two groups originate from distinct distributions. In the case of the region level, differential methylation was assessed as the mean difference in means across all sites in a given region, as well as a combined p-value calculated from all site p-values in the region.

DNA methylation changes were considered as significant with FDR less than 0.05 and beta value differences between both groups greater than 0.1, which corresponds to a 10%.

Transcriptome data

The main objective of transcriptome analysis is to find differentially expressed genes. Those are genes that are expressed in significantly different quantities when comparing distinct groups of samples, for example, disease versus control individuals. This analysis is done in a univariate way, so referring on one gene at a time and stage. This fact is due to that there are tens of thousands of genes and the number of samples used is much smaller, so it is very difficult to fit a statistical model that considers all genes as a whole ¹⁸⁶. Reads that are independently sampled from a population would be expected to follow a Poisson distribution, and it has been used to test differential expression. However, Poisson distribution assumed that mean is equal to variance and when samples are taken from different biological individuals, variance is much higher. This means that Poisson distribution is too restrictive and does not control the probability of false discoveries. To manage this overdispersed distribution problem, it has been proposed a model of negative binomial distributions, which is used in the EdgeR and DESeq2 packages ¹⁸⁰.

Mapped reads count to one gene or isoform is proportional with the isoform or gene abundance, this means that the more abundant a gene is expressed the more reads are sequenced. The problem is that genes and/or isoforms are not equal, and this assumption would be so for expressed genes with same length. Moreover, sequencing depth differs between samples, which alters the comparison of read counts of a given gene in different samples. Therefore, all read counts must be normalized to be comparable across genes, gene isoforms and samples. Normalization is also used to make expression level distribution able to presumptions used in specific statistical methods. There are some standards adjustments used for this correction in RNA-seq. The easiest way is to correct for gene length, dividing the number of reads by the total number of bases in the sequence and multiplying by one thousand, resulting in Reads per Kilobase (RPK). RPK does not adjust expression levels across samples. For this, Reads or Fragments per kilobase per million (RPKM and FPKM) adjustment methods

are used. RPKM is made for single-end RNA-Seq, where every read is a single sequenced fragment, whereas FPKM is made for paired-end RNA-Seq, where two reads may correspond to a single fragment. Thus, FPKM considers that two reads can map to one fragment and does not count the fragment twice. There also exist other commonly used normalization methods, such as, Transcript per million (TPM) and Trimmed mean of M values (TMM) that are correcting for different compositions of RNA pools. However, the most common used method and likewise the one used in this work for expression from RNA-Seq is FPKM^{187,188}.

Differential expression analysis methods are constantly under development and there is not an established best method for this analysis. There are some controversies about best ways to normalize and statistically analyse gene expression. Therefore, in this thesis we explored our data by using different softwares. Cuffdiff is frequently regarded as the best choice for differentially expressed isoforms study, and this makes cuffdiff, in addition, more appropriate on the gene level study because gene level changes are closely related to isoform level changes. However, other R/Bioconductor packages were also used (EdgeR and DESeq2), because some authors have shown that their performance as equal as that of Cuffdiff and may have the advantage that they can work with more than a covariate using a generalized linear model, for example, using disease and experimental conditions as two independent covariates. This has been implemented for differentially expressed genes analysis using pre and post-differentiated samples as a covariate and disease as a second covariate¹⁸⁹. Cuffdiff assumes beta negative binomial distribution and uses t-test for differential expression evaluation. EdgeR and DESeq2 use negative binomial distribution and exact test for differential expression analysis.

Significant gene expression changes were considered with an FDR less than 0.05 and a fold change greater than 2.

Pathways and gene ontology (GO) analysis

Over representation methods with WebGestalt online software were implemented with Wikipathways database for pathways analysis and the gene ontology database for biological processes terms¹⁹⁰. Webgestalt uses the hypergeometric test methodology for over representation enrichment analysis.

Enhancer analysis was done with GREAT. GREAT (Genomic Regions Enrichment of Annotation Tools) software associates genomic regions to genes by using a single transcription start site (TSS) to specify the location of each gene. Two options have been used with this software for each gene list obtained in our results. The first one consists in the association of ‘two nearest genes’, which extends each region given to the nearest upstream and downstream TSS up to 1000 kilobases in each direction. The other option used was the ‘single nearest gene’ association rule, which extends each region to the nearest adjacent TSS, up to 1 Mb in each direction. The TSS used is that of the ‘canonical isoform’ of the gene as defined by the UCSC Known Genes track. For the enrichment analysis of a set of cis-regulatory regions (the regions given as input), GREAT implements both, the binomial test (regulatory domain bias) and the traditional hypergeometric gene-based test (gene specific enrichment), and it highlights ontology terms enriched by both tests separately ¹⁷⁴.

Finally, the ‘CommonPathway’ function in ArrayTrack recognizes the common genes between different sets of genes, and identifies which pathway is significantly altered for each combination of sets. The statistical significance of each pathway is estimated using Fisher’s exact test ¹⁸³.

6. RESULTS

RESULTS

Isolation and characterization of BMSCs

The procedure to isolate BMSCs from the femoral head is described in methods. We found a similar success rate in establishing a BMSC culture with samples from patients with fractures (FRX: 27 out of 41, 66%) or with OA (22 out of 27, 59%). Since long-term culture has been previously shown to introduce changes in the DNA methylation pattern ¹⁹¹, once isolated from the bone tissue, the BMSCs were allowed to proliferate for a short period until confluency and then RNA and DNA were collected at first passage. To confirm the stem cell nature of the cells used for the study, some cultures were used to check their phenotype by flow cytometry using a combination of five different markers (CD45-, CD34-, CD90+, CD73+, CD105+) that define the phenotype of BMSCs, as established by the International Society for cellular Therapy (ISCT) ¹⁹². Cytometry results showed that the starting cells for the subsequent transcriptome and methylome analysis were obtained from a highly pure population of BMSCs.

Proliferative capacity of BMSCs

Actively dividing cells express Ki-67 protein, therefore this is a commonly used marker for cell proliferation. We tested our cultures of BMSCs, and interestingly found that the proportion of Ki-67 positive cells was significantly higher in BMSCs isolated from patients with osteoporotic fractures (60%) than in cultures from osteoarthritic patients (40%, $p=0.0003$). These results were confirmed by assessing cell proliferation by MTT assay experiments.

DNA methylation profiling

We used 39 BMSCs samples for the DNA methylation study (22 were isolated from osteoporotic patients and 17 from patients with osteoarthritis). Human DNA methylation 450k array used for this analysis provided an idat format file per sample, where 485577 different probes were detected. The first step for microarray data preprocessing analysis is to remove bad quality probes. Thus, 4713 sites were filtered out because they were overlapping with SNPs. Technical variation in background fluorescence signal was corrected using the “noob” method from

methyumi package ¹⁹³, and the signal intensity values were normalized using the SWAN normalization method. Additionally, 3156 probes of distinct specific sequence contexts, which are probes dedicated to SNP detection, were also removed. Thus, after the filtering procedures, 7869 probes were removed and 477708 probes were retained for further analysis.

Besides the single CpG site output, the methylation analysis included CpGs grouped in different ways, such as 137536 tiling regions, 30794 gene regions, 30945 promoter regions and 26649 CpG islands. Single probe analysis revealed a bimodal distribution of methylation, as also observed in other studies. However, in CpG islands and promoter regions, the unmethylated sites were much more common, whereas in gene body regions DNA methylation values were intermediate, with a pyramidal distribution peaking at 50 percent of methylation levels (Figure 14).

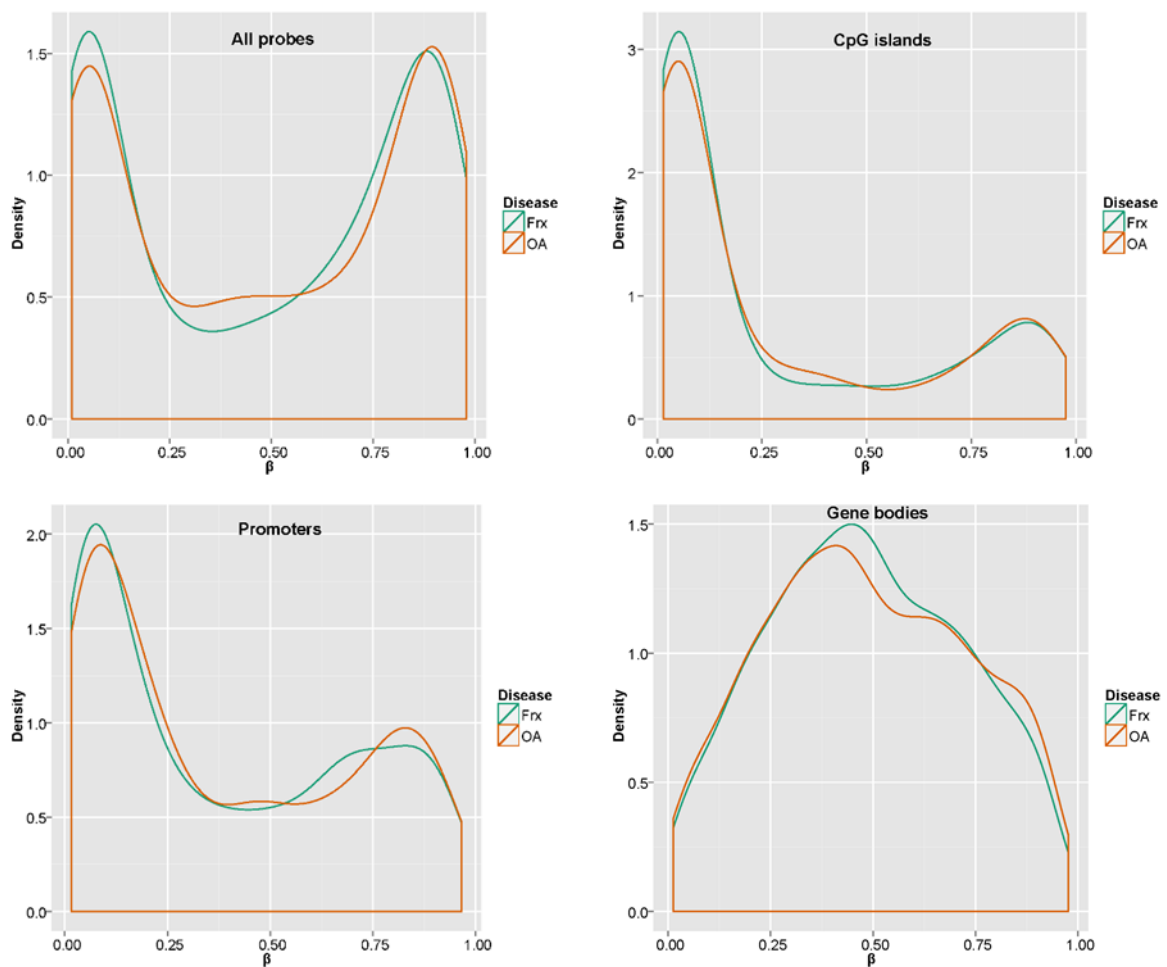


Figure 14. Density plots comparing methylation value distributions (FRX in green and OA in orange) according to the probe and region levels (CpG islands, promoters and gene bodies).

RnBeads suite uses two methods for dimensional reduction of DNA methylation data, principal component analysis (PCA) and multidimensional scaling (MDS). Using all CpG sites, independently of the location, we observed an overall trend for the separation of fracture and osteoarthritis groups by both procedures (Euclidean distance in MDS or the first and second principal components in PCA) (Figure 15).

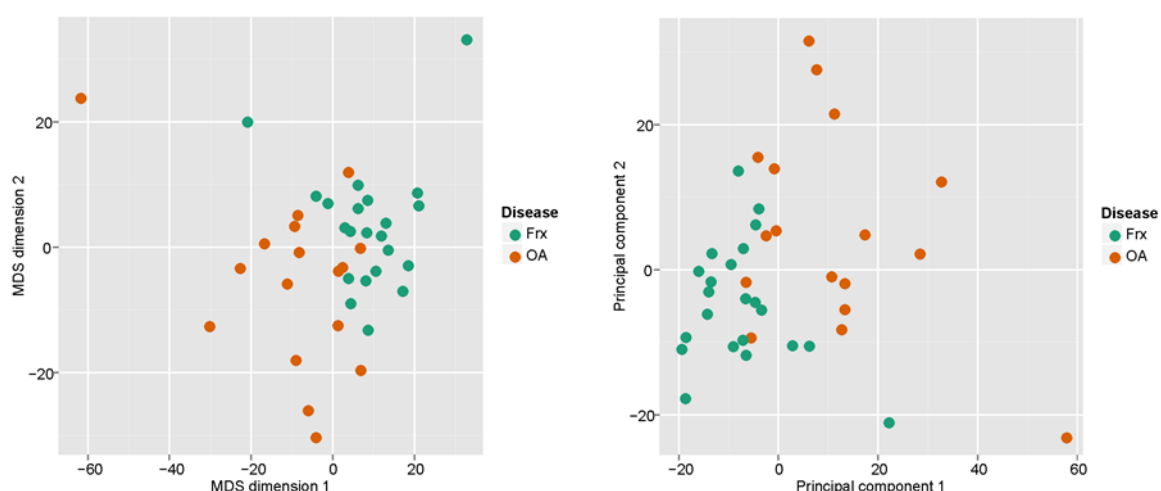


Figure 15. The scatter plot on the left shows all samples transformed into a two-dimensional space using multidimensional scaling approach. Similarly, on the right, a scatter plot with the values of the first and second principal components of all samples. In both cases sample groups are coloured (FRX in green and OA in orange).

Differential methylation analysis was conducted both at the single probe and at region level. Age was included as a covariable because age, as explained above, is an important factor for DNA methylation, and per se age may change DNA methylation levels in some regions, among other factors. Differential methylation at the single site level was analysed by using a t-test that compared the mean methylation levels of each CpG in both groups. The resulting p-values were corrected for multiple test comparisons, using the false discovery rate (FDR), which is a method used to diminish type I errors in null hypothesis testing.

The average methylation level was similar in both groups as shown in Figure 1. However, among all CpG sites analyzed (477708), we found 9038 sites differentially methylated, considering a combined $FDR < 0.05$ and β -differences of the group mean greater than 10 percent. Of these 9038 differentially methylated CpG sites, 4417 were more methylated (hypermethylated) and 4621 were less methylated (hypomethylated) in BMSCs obtained from patients with hip fractures

(FRX). These sites were distributed in different genomic regions, specifically 1586 sites in CpG islands, 1105 in shores, 353 in shelves and 5994 in open sea regions.

In parallel, the analysis at the region level revealed 217 differentially methylated gene promoters out of 30877, including 111 hypermethylated and 106 hypomethylated in FRX in comparison with OA. Furthermore, from these 217 regions, only 62 correspond to protein coding genes, whereas the majority (155) were non protein coding genes. Among CpG islands, 40 regions were differentially methylated, from which 16 were hypermethylated and 24 hypomethylated in FRX. Among gene bodies, 62 were found to be hypermethylated and 67 hypomethylated in FRX (Figure 16).

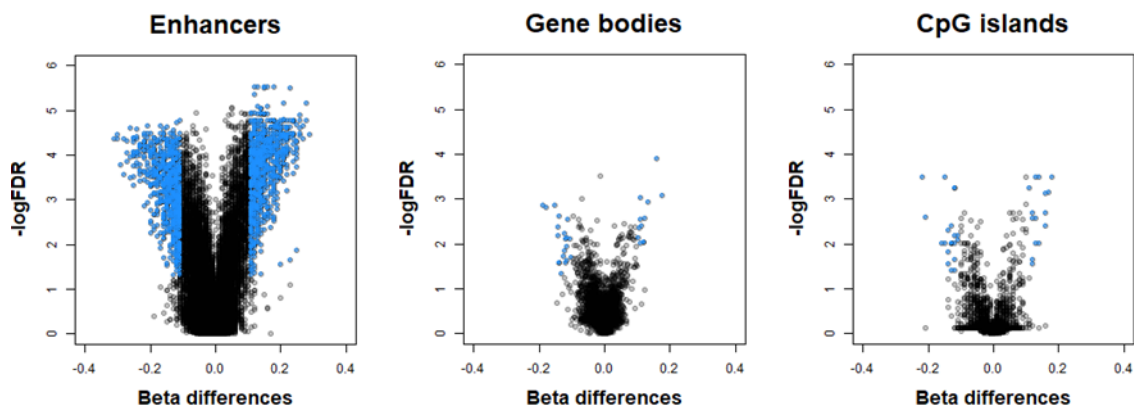


Figure 16. Volcano plots showing different genome regions (Enhancers, gene bodies and CpG islands) with significant DNA methylation differences in blue dots ($FDR < 0.05$ and absolute beta differences > 0.1) between BMSCs obtained from fractured patients (FRX) and OA patients.

In addition to this region level analysis, RnBeads permits to analyze other regions according to custom annotations. So, we used this option to retrieve annotation data for genomic enhancers specified by a chromatin state segmentation approach employed in the ENCODE project. The region level analysis at enhancers revealed 1684 differentially methylated regions (out of the 41280 regions explored), from which 870 were hypermethylated and 814 hypomethylated in FRX.

All significant regions have at least a CpG site differentially methylated but in some cases there were more sites in the same region, so we looked at the differentially methylated individual sites enriched in these regions and obtained

402; 1355; 124 and 2425 CpG sites in promoters, gene bodies, CpG islands and enhancer regions, respectively (Table 6).

Differential methylation on the region level analysis				
Genomic regions	Promoters	Gene bodies	CpG islands	Enhancers
Analyzed regions	30877	30725	26649	41280
Analyzed CpG sites per region	208588	383795	149149	62169
Differential CpG sites	402	1355	124	2425
Differential regions	217	129	40	1684
Hyper-methylated regions	111	62	16	870
Hypo-methylated regions	106	67	24	814

Table 6. Distribution of differentially methylated CpGs between BMSCs from fractured patients (FRX) and patients with osteoarthritis (OA) in various genomic regions (Promoters, gene bodies, CpG islands and enhancer regions). Each analyzed region could have from 1 to several CpGs. Significant hyper- or hypo-methylated regions were calculated with the mean difference in means across all sites in a region of the two groups being compared, as well as, a combined p-value calculated from all site p-values in the region (mean-mean difference > 0.1 and combined FDR < 0.05). Hyper-methylation refers to higher methylation in FRX than in OA. Some CpG sites could have been analyzed in different set of regions.

As previously described, we observed that differentially methylated CpG sites were enriched in enhancer regions, 2425 of 9038, which corresponds to more than a quarter of all of them.

We then focused on the differentially methylated enhancer regions and their nearest protein-coding gene, which is assumed to be the putative target gene. This enhancer-gene association was determined with the Genomic Regions Enrichment of Annotation Tool (GREAT), which predicts functions of cis-regulatory regions. GREAT associates proximal input regions with their target genes and uses annotations from several gene ontology databases to associate the genomic regions to significant annotation terms ¹⁷⁴. We used the significant enhancer regions as input in GREAT software and obtained associated genes, using the 'single nearest gene' option within 1000 kb of distance extended in both directions. Thus, 1684 differentially methylated enhancer regions were related with 1400 genes, 722 genes linked with hypermethylated enhancers and 678 with hypomethylated enhancers. Distances between each enhancer and its associated transcription start site (TSS) are variable and are shown in Figure 17.

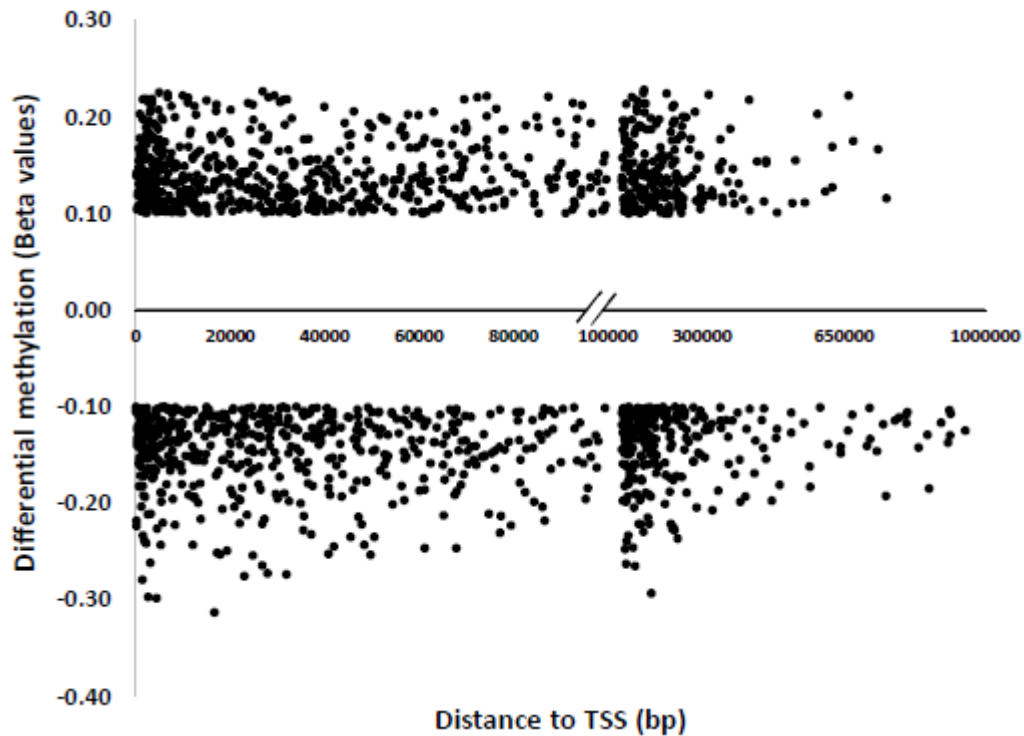


Figure 17. Distances of the differentially methylated enhancers to the transcription start sites (TSS) of the associated genes with GREAT software. The distances to TSS are represented versus the mean differences (Beta values) of the enhancer regions when comparing BMSCs from fractured patients (FRX) and patients with osteoarthritis (OA).

Genes with differentially methylated enhancers were enriched in stem cell development and bone-related pathways such as the Wnt receptor signaling pathway ($p = 4.5 \cdot 10^{-8}$; binomial test), regulation of osteoblast differentiation ($p = 9.1 \cdot 10^{-5}$), regulation of hMSCs proliferation ($p = 7.6 \cdot 10^{-6}$), and bone development ($p = 5.4 \cdot 10^{-5}$) pathways (Figure 18).

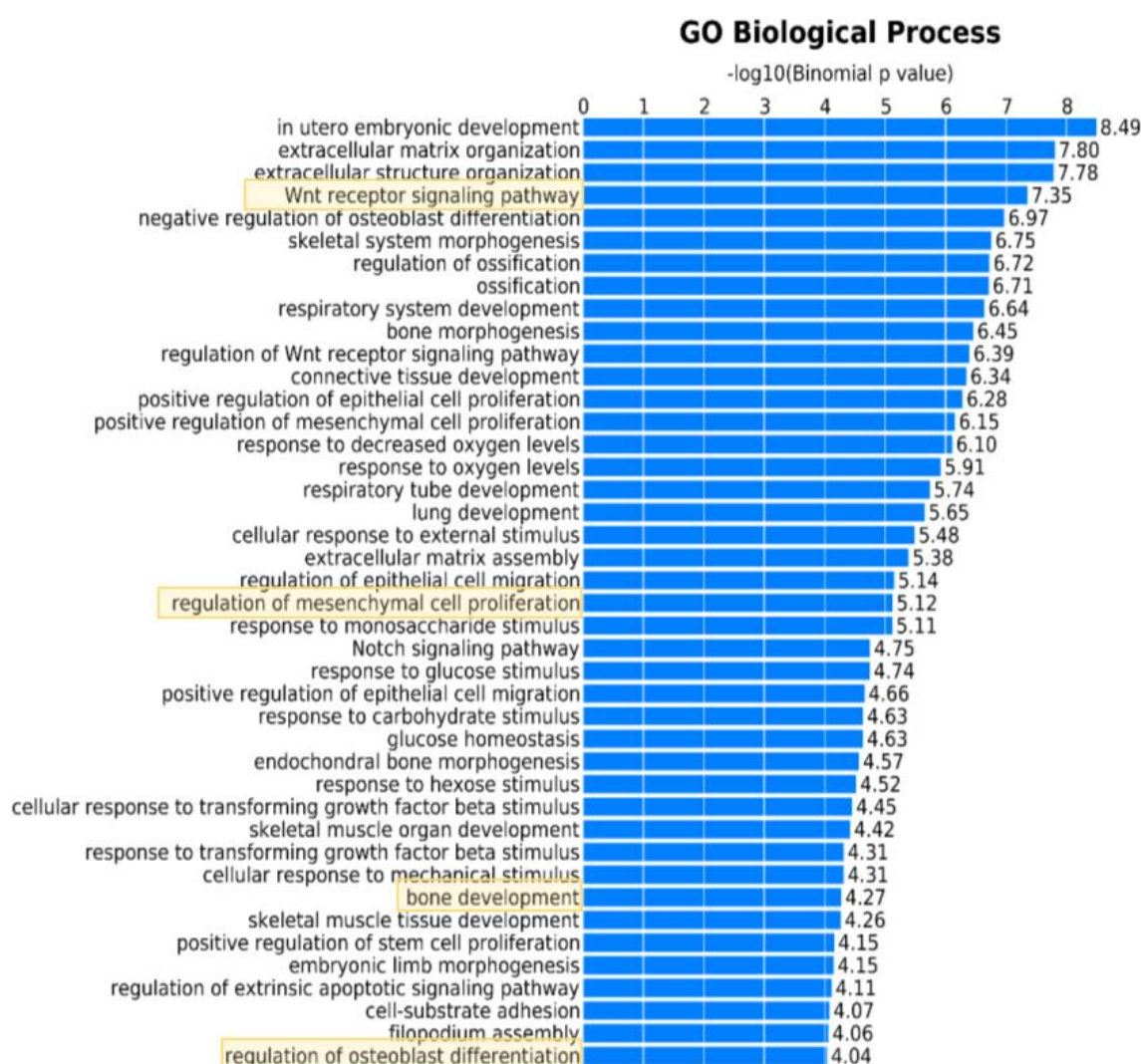


Figure 18. Bar chart of the gene ontology (GO) enrichment analysis, concerning biological processes terms, of those genes associated to differentially methylated enhancers between BMSCs derived from FRX and controls with OA. Surrounded in yellow, some interesting bone related processes.

Epigenetic aging

Studies of DNA methylation have shown regions whose methylation level tend to change quite consistently with advancing age. Likewise, telomere length is known to shorten with aging. Thus, we explored that “epigenetic aging” pattern in our BMSCs and bone samples from osteoporotic or osteoarthritic origin. The methylation patterns were analyzed with Horvath’s Epigenetic clock software ⁸⁶, based on Illumina DNA Infinium 450k data. As explained above, this software employs a set of CpG sites (353 sites) showing age-related changes in DNA methylation. Bone DNA of 20 patients with hip FRX and 19 patients with hip OA

methylation data were obtained from previous studies of our laboratory. We found a significant correlation between the predicted epigenetic age and the chronological age ($r = 0.64$, $p = 1.36 \cdot 10^{-5}$). Moreover, BMSCs from osteoporotic patients showed an accelerated aging when compared with BMSCs from patients with osteoarthritis ($p = 0.001$). However, there were not differences in DNA from bone tissue samples obtained from patients with FRX or OA ($p = 0.111$) (Figure 19A).

Further, we explored the relative telomere length in BMSCs and bone samples as for DNA methylation age. Telomeres shortening alters cell proliferation and divisions and is considered an aging hallmark, also related with aging-associated diseases. There were not significant differences in the relative telomere length of either bone samples or BMSCs obtained from patients with osteoporosis or with osteoarthritis (Figure 19B). However, the interindividual variation was very large, which might limit the power to find statistically significant differences.

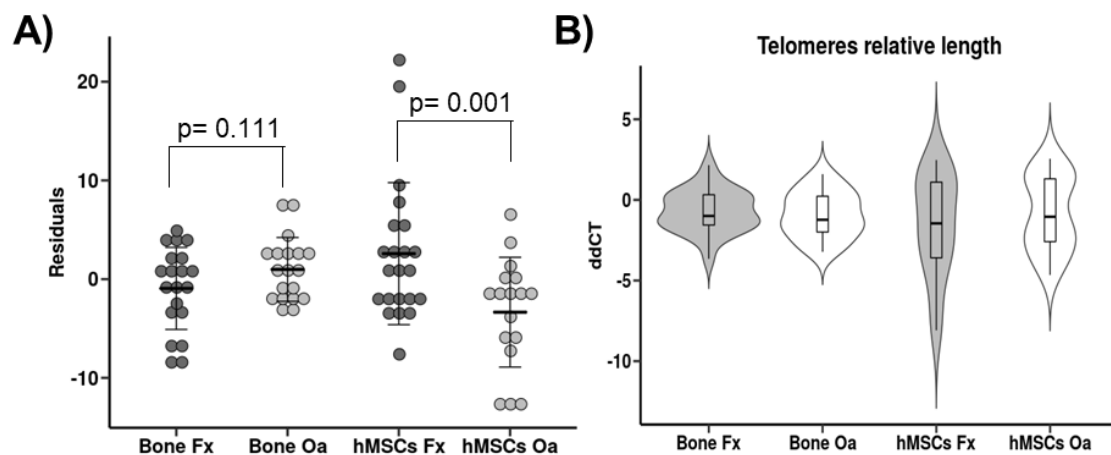


Figure 19. Epigenetic marks in bone tissue and BMSCs from patients with hip fractures and with hip osteoarthritis. (A) Epigenetic aging as revealed by age-related DNA methylation marks. Dot plot with the mean and SD of the residuals (deviation from the overall regression line of epigenetic age and chronological age) from bone and BMSCs isolated from osteoporotic fractures (Fx, dark gray dots) and osteoarthritis (Oa, gray dots) are shown. Between-group differences were compared by ANCOVA, with the chronological age as covariable. Data derived from previously reported results (Delgado-Calle et al. 2013; Vidal-Bralo et al. 2016; del Real et al. 2017). (B) Violin plots showing the distribution of the relative telomere length density from bone and BMSCs isolated from osteoporotic fractures (in gray) and osteoarthritis (in white).

Gene Expression Profiling

We used 20 RNA samples from BMSCs (10 FRX and 10 OA) for sequencing with the Hi-Seq technology (Illumina), using a paired end library per sample with at least 20 million of total reads. Reads were mapped using the star RNA aligner v2.3.0 with an efficiency of 81.4 ± 5.6 (Table 7).

Sample	Total Reads	Total Paired Reads	Mapping %	Mapped Paired Reads
1	26,703,234	13,351,617	82.54	11,020,425
12	33,815,406	16,907,703	72.23	12,212,434
2	33,351,538	16,675,769	82.99	13,839,221
3	44,932,312	22,466,156	80.65	18,118,955
4	36,214,556	18,107,278	81.18	14,699,489
5	32,308,526	16,154,263	82.23	13,283,651
6	92,627,492	46,313,746	82.54	38,227,366
7	32,164,458	16,082,229	84.33	13,562,144
8	34,046,988	17,023,494	82.39	14,025,657
9	31,728,064	15,864,032	82.46	13,081,481
22	50,920,756	25,460,378	62.05	15,798,165
23	41,267,446	20,633,723	79.62	16,428,571
24	56,045,870	28,022,935	86.38	24,206,212
13	42,720,916	21,360,458	85.89	18,346,498
14	33,283,220	16,641,610	86.04	14,318,442
15	28,453,606	14,226,803	81.93	11,656,020
16	34,283,956	17,141,978	84.6	14,502,114
17	35,710,624	17,855,312	77.65	13,864,650
19	33,464,362	16,732,181	86.21	14,424,814
20	29,907,096	14,953,548	83.44	12,477,241

Table 7. Summary of the reads alignment to the human genome (hg19) for each sample.

Among them, 11390 genes were expressed (defined, for the purpose of this study, as those having more than 10 reads per group) both in cells from FRX patients and in cells from OA patients, whereas 496 genes were expressed only in FRX and 1695 in OA. The average number of reads was similar in FRX and OA and both groups were comparable regarding the expression of 4 housekeeping genes frequently used as controls (GAPDH, TBP, RPL13A, YWHA2).

AltAnalyze analysis pipeline incorporates a program (lineage profiler) that calculates lineage correlation (Z-scores) of the samples subjected to transcriptome analysis in terms of different tissues¹⁹⁴. This tool showed that the expression signature of our samples was typical for hMSCs, as expected (Table 8).

	FRX	OA
Bone Marrow Mesenchymal Stem Cells	2.42	2.02
Osteoblasts	2.41	2.11
Adipocyte Progenitor	2.65	2.10
Bone Marrow Stem Cells	3.16	2.70
Astrocytes	2.00	1.54
Aorta Smooth Muscle Cells	2.23	1.64
Myoblast	1.19	0.93
Fibroblasts	1.35	1.00
Neural Crest Mesenchymal Stem Cells	1.20	0.71

Table 8. Lineage correlations (Z-scores) of the cells subjected to transcriptome analysis.

Differential gene expression analysis with both, AltAnalyze and EdgeR softwares^{181,194}, revealed 338 differentially expressed genes (defined as $FDR < 0.10$ and fold change > 2). Among them, 99 genes were upregulated, whereas 239 were downregulated in FRX (Figure 20 and Appendix 1).

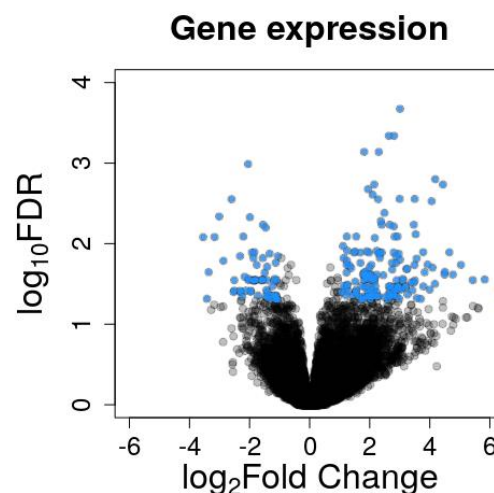


Figure 20. Volcano plot of differential gene expression analysis in BMSCs grown from patients with fractures (FRX) and controls with osteoarthritis (OA). Significant genes are those with an $FDR < 0.1$ and an absolute fold change > 2 (blue dots) between BMSCs obtained from FRX and OA patients.

These genes were subjected to Gene ontology and pathways analysis and we observed that upregulated genes in FRX were enriched in hMSCs differentiation and bone formation processes, whereas genes that were downregulated in FRX are enriched principally in immune related pathways (Table 9)¹⁶⁵.

Upregulated genes		Downregulated genes	
GO: Biological process	AdjP	GO: Biological process	AdjP
Anatomical structure morphogenesis	1.9E-03	Immune system process	1.7E-08
Angiogenesis	3.0E-03	Complement activation, classical pathway	5.0E-08
Skeletal system development	4.0E-03	Innate immune response	5.0E-08
Anatomical structure development	6.1E-03	Defense response	9.1E-08
C21-steroid hormone biosynthetic process	6.1E-03	Regulation of immune system process	9.9E-08
Organophosphate catabolic process	8.2E-03	Response to stimulus	1.5E-07
Bone remodeling	8.2E-03	Regulation of immune response	2.5E-07
Negative regulation of glucocorticoid biosynthetic process	8.2E-03	Complement activation	2.5E-07
Regulation of bone mineralization	8.2E-03	Humoral immune response mediated by circulating immunoglobulin	3.2E-07
Biomineral tissue development	8.2E-03	Leukocyte mediated immunity	3.3E-07

Table 9. Top ten pathways after enrichment analysis of differentially up- and down-regulated genes in FRX. Overrepresentation enrichment analysis was implemented with webgestalt software on wikipathways database.

We, then, compared those differentially expressed genes (338 genes) with the genes previously found associated with differentially methylated enhancers (1400 genes). Association between up- or down-expression and hyper- or hypomethylation was variable, but there was a trend for an inverse correlation between enhancer methylation and the associated gene expression (Appendix 2). In Figure 21 we can see that 18 upregulated genes in FRX had 8 hypermethylated enhancers and 10 had hypomethylated enhancers, whereas 54 downregulated genes had 39 hypermethylated and 15 hypomethylated enhancer regions, all in reference to FRX [odds ratio (OR): 0.3; 95% confidence interval: 0.12–0.99; $p = 0.05$].

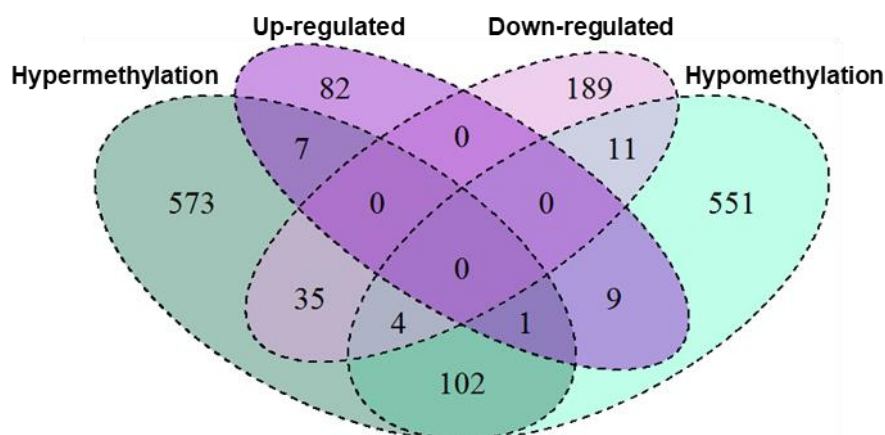


Figure 21. Relationship between DNA methylation and gene expression signatures. Venn diagram summarizing the association between differential DNA methylation and differential gene expression (comparisons of BMSCs from fractures over BMSCs from controls). As shown, enhancers for the same gene with differential DNA methylation may have both changes (Hypermethylation and hypomethylation).

ArrayTrack is a bioinformatic tool with multiple possibilities in microarray data analysis¹⁸³. This tool has the option to analyze common pathways between two sets of differentially expressed genes, or, in general, any two sets of genes. We hypothesized that pathways that are common to differentially expressed genes and differentially methylated enhancers are more likely to be true disease-related pathways than other pathways present in just one of those gene lists. By using ArrayTrack software, Gene Ontology Enrichment analysis revealed that genes with hypomethylated enhancers and upregulated gene expression in FRX are enriched in bone related pathways, such as positive regulation of mesenchymal cell proliferation, endochondral bone morphogenesis or regulation of bone mineralization, as well as some neuron-related pathways (Figure 22A). The three other combinations of up- or down-expressed and methylated genes did not highlight bone-related pathways, excepting the case of genes with hypermethylated enhancers and overexpressed genes in FRX (Figure 22), but those gene-pathways are in common with the first one.

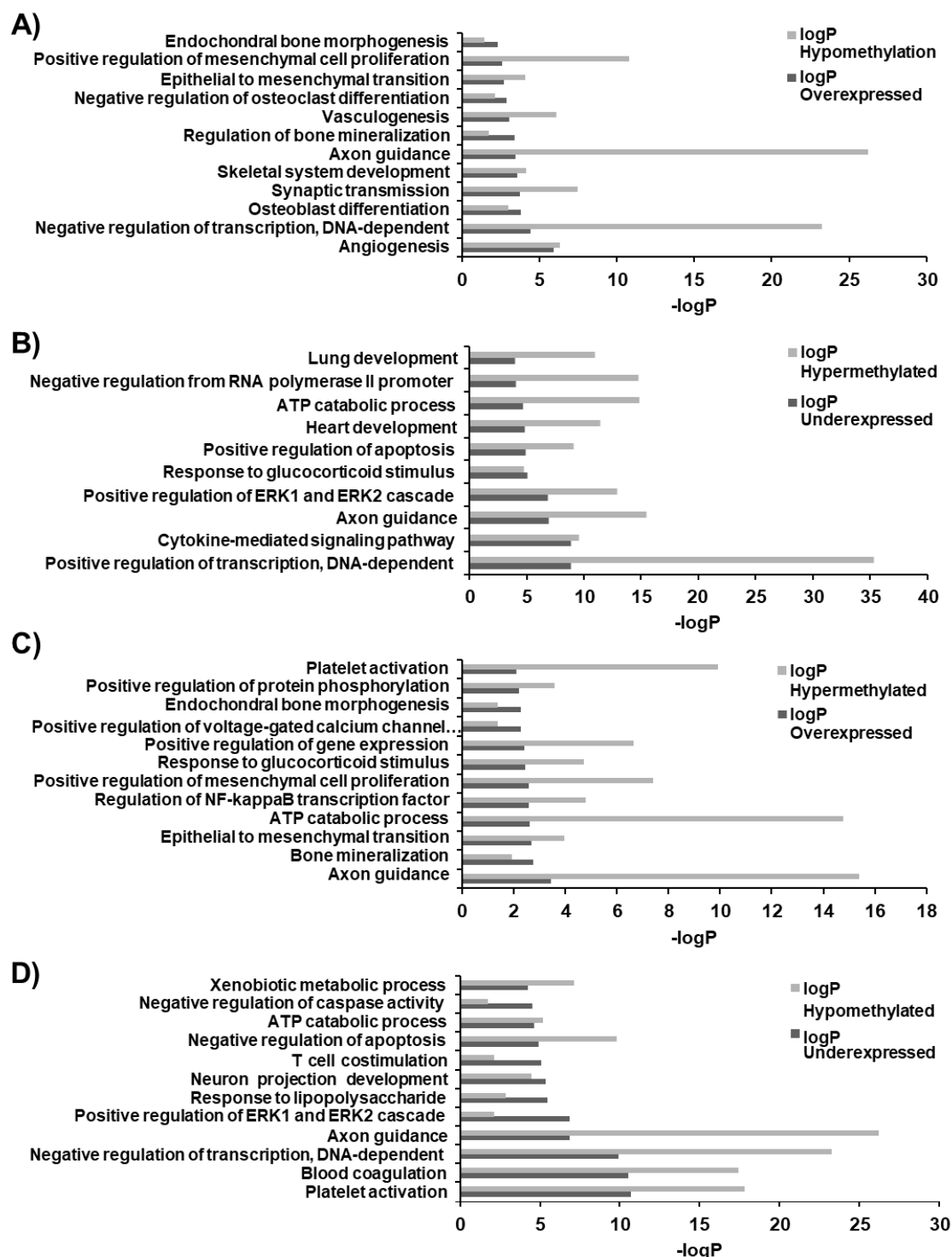


Figure 22. Relationship between DNA methylation and gene expression signatures. Pathways enrichment analysis of genes with differentially methylated enhancers and differentially expressed genes, in BMSCs from fractures over BMSCs from controls. All possible combinations are shown. (A) Hypomethylated enhancers with overexpressed genes. (B) Hypermethylated enhancers with underexpressed genes. (C) Hypermethylated enhancers with over expressed genes. (D) Hypomethylated enhancers with underexpressed genes.

DNA methylation and gene expression replication

After all the above bioinformatic data analysis, we chose some candidates to replicate the results of both DNA methylation and gene expression obtained with the methylation array and RNAseq procedures, by pyrosequencing and RT-qPCR, respectively.

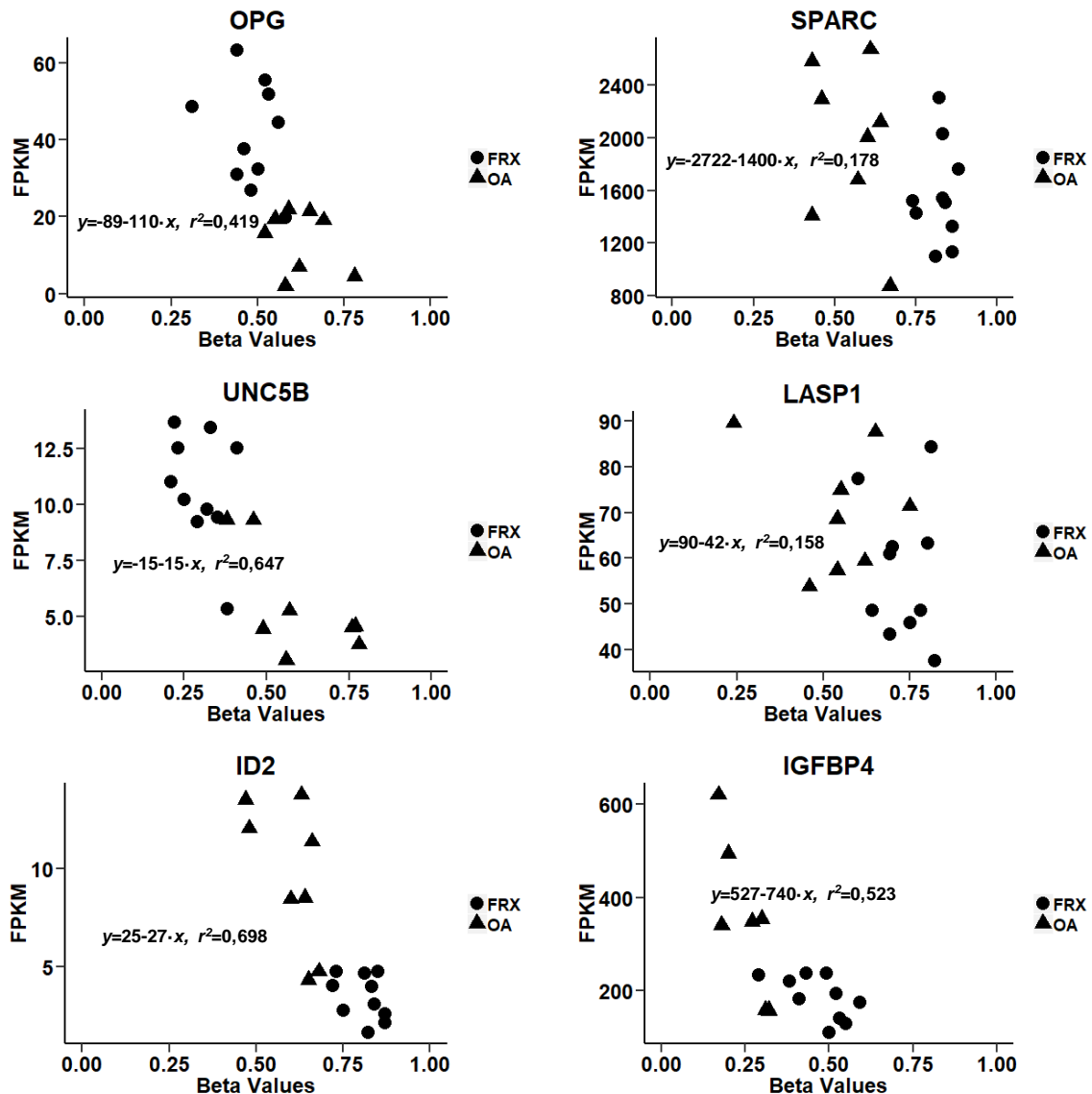


Figure 23. Scatter plots showing the correlation of DNA methylation (Beta values) and gene expression (FPKM) levels for some representative genes, which have been found to have differential expression and DNA methylation patterns when comparing BMSCs isolated from patients with FRX (circles) and patients with OA (triangles). Equations shown are calculated with a linear regression analysis.

For replication, we selected 10 differentially expressed genes with differentially methylated enhancers: SPARC, FOXP2, LOXL2, SLC5A3, LAMC1, TNFRSF11B

(OPG), ID2, IGFBP4, LASP2 and UNC5B. The inverse correlation between DNA methylation and the expression of several genes is shown in Figure 23.

For the validation of gene expression by RT-qPCR, we used 8 samples previously analyzed by RNAseq (4 FRX and 4 OA) and 19 additional samples (9 FRX and 10 OA), also obtained in our lab with the same procedures. Among the 10 genes studied, the results were replicated (i.e., differential expression) in 4, considering $p < 0.05$ between sample groups and differences in gene expression in the same direction as in RNAseq. Specifically, 3 genes (LOXL2, ID2 and OPG) were replicated in the technical validation (8 samples), 2 genes (ID2 and UNC5B) considering all 'new' samples, and 4 genes (LOXL2, ID2, OPG and UNC5B) when the results of both sample groups were analyzed.

We then designed primers to replicate the differential methylation of enhancers associated with those 4 genes by pyrosequencing. The following CpG sites were explored: CpG 22489510 (ID2); CpG 26711508 (OPG); CpG 02675344 (UNC5B) and CpG 24911388 (LOXL2). We used a total of 24 samples (12 FRX and 12 OA) for technical replication, therefore all of them were previously used in our 450k array. The results showed similar average methylation (% values) with array and pyrosequencing in OPG (57.6 ± 11.2 and 47.6 ± 9.4 , respectively), ID2 (68.9 ± 15.9 and 75.1 ± 14.3 , respectively) and UNC5B (45.8 ± 17.6 and 48.1 ± 13.6 , respectively). However, for LOXL2 the values obtained with both techniques were rather different (80.6 ± 10.5 and 48.8 ± 11.9 , respectively) (Figure 24A). Nevertheless, the methylation differences between both groups (FRX and OA) previously found with the methylation array were replicated by pyrosequencing (Figure 24B), and methylation values by both techniques were highly correlated (Figure 24C).

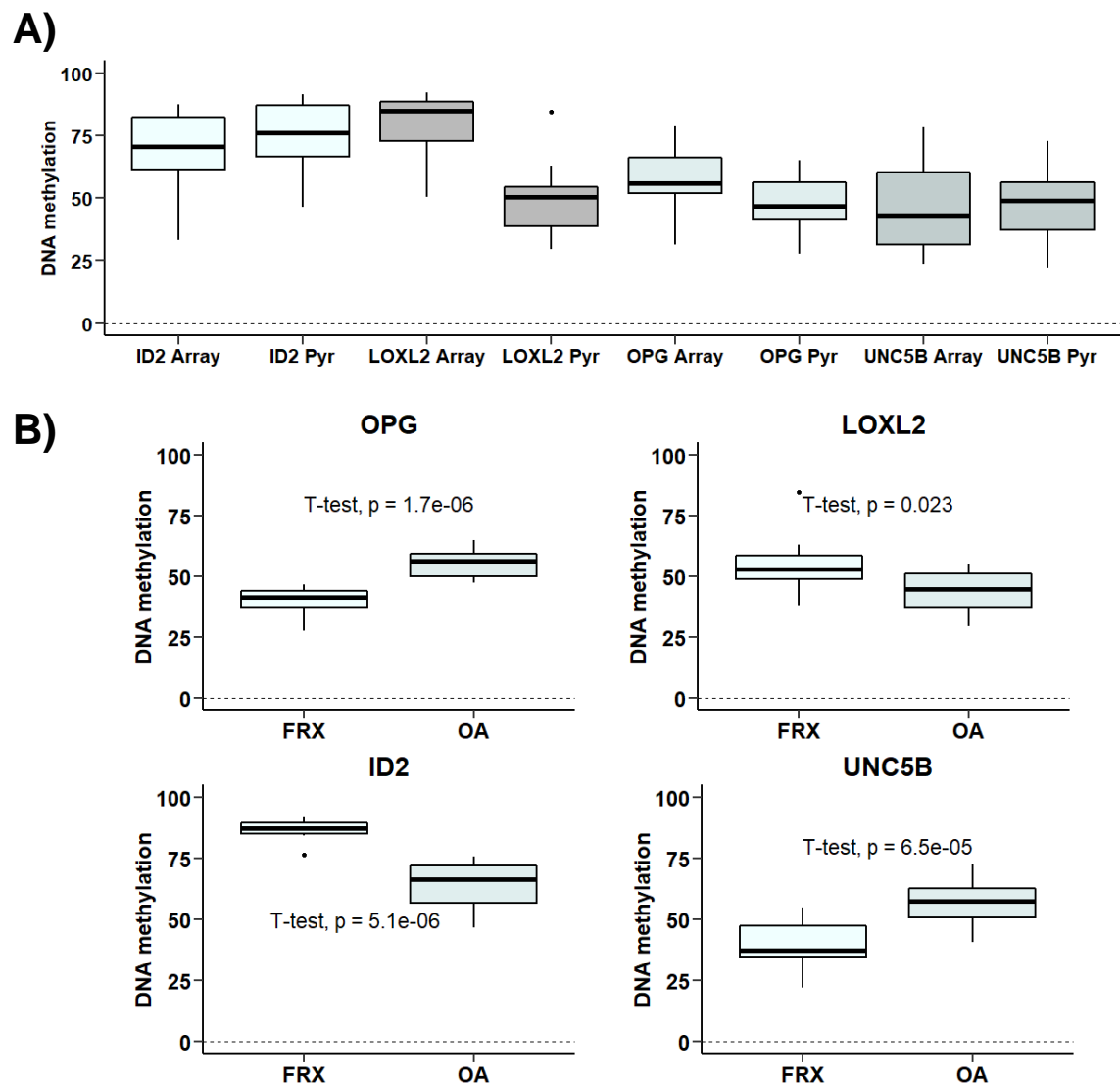


Figure 24. A) Box plot with the DNA methylation levels from the array and pyrosequencing (Pyr) of the replicated genes (ID2, LOXL2, UNC5B and OPG). B) Box plot comparing DNA methylation levels measured by pyrosequencing in BMSCs derived from fractures (FRX) and controls with osteoarthritis (OA). The differences between both groups are significant, as with our array results.

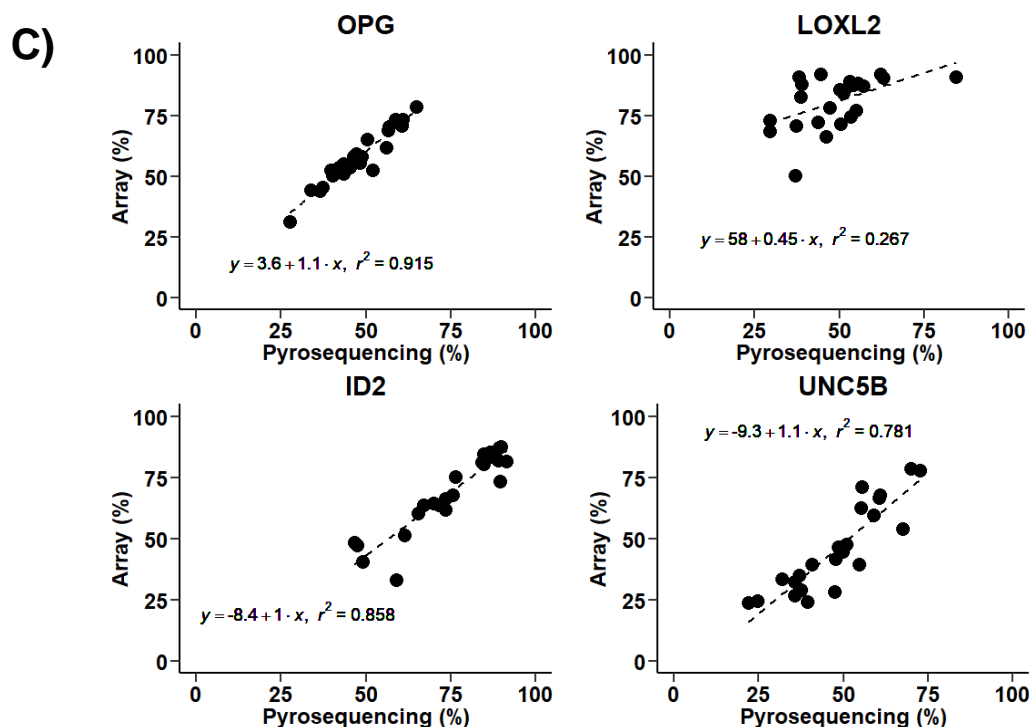


Figure 24. C) Scatter plots showing the DNA methylation levels correlation between array values and pyrosequencing values. Correlation is tested with generalized linear model (linear regression analysis).

Once we confirmed that those 4 genes were differentially methylated and expressed in BMSCs obtained from patients with FRX and OA, we explored their relevance for osteogenic differentiation.

For these experiments, we used BMSCs grown from 17 patients (8 FRX and 9 OA), which gene expression patterns were analyzed at baseline and after 21 days of osteogenic differentiation in vitro. Pairwise comparisons revealed statistically significant differences between pre-differentiated samples and post-differentiated samples in ID2, UNC5B and LOXL2 genes ($p=0.0346$, 0.0063 and 0.0033 , respectively), but not in OPG ($p=0.4196$) (Figure 25).

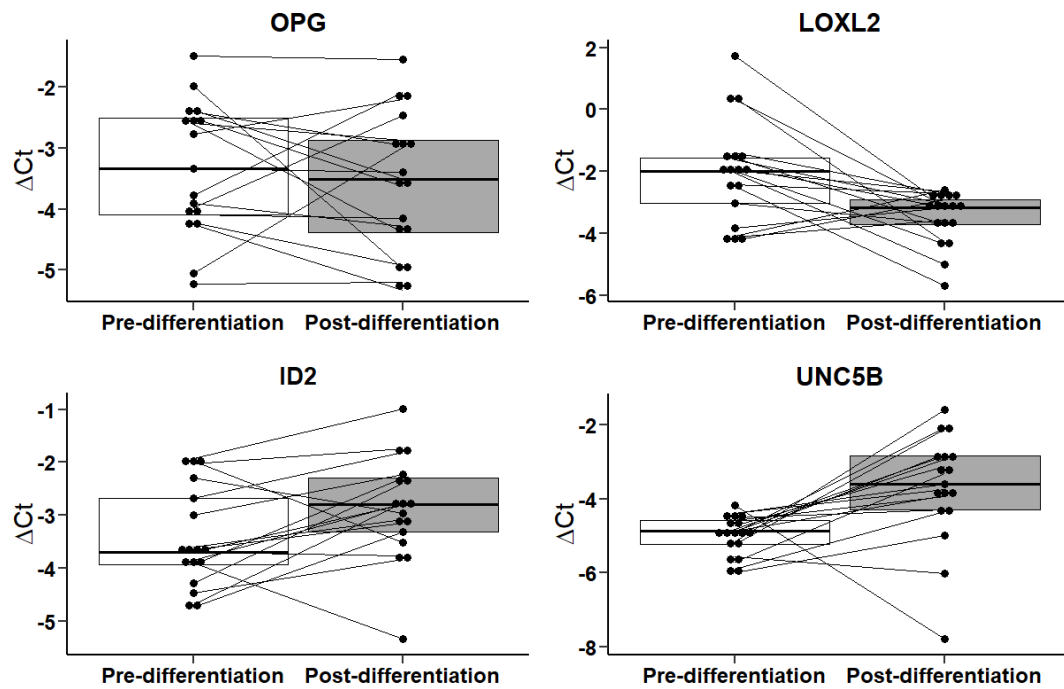


Figure 25. Box plots showing gene expression levels (deltaCt) of BMSCs in the basal state (white boxes) and after 21 days of osteogenic differentiation in vitro (gray boxes). Each point corresponds to a sample, which is linked by the line with its pair after differentiation.

Osteogenic capacity of BMSCs

In parallel experiments, we used RT-qPCR to study the expression of some genes typical of the osteoblastic lineage, including RUNX2, OSX, ALPL, SPP1, BGLAP, COL1A1 and IBSP. These experiments showed that OSX and BGLAP were significantly more expressed in FRX than in OA ($p=0.024$ and 0.002 , respectively). Whereas ALPL and SPP1 were significantly more expressed in OA than in FRX ($p=0.046$ and 0.016 , respectively). Collagen expression was similar in both groups and in the case of adipogenic markers, PPARG was significantly more expressed in OA than in FRX BMSCs ($p=0.01$) (Figure 26).

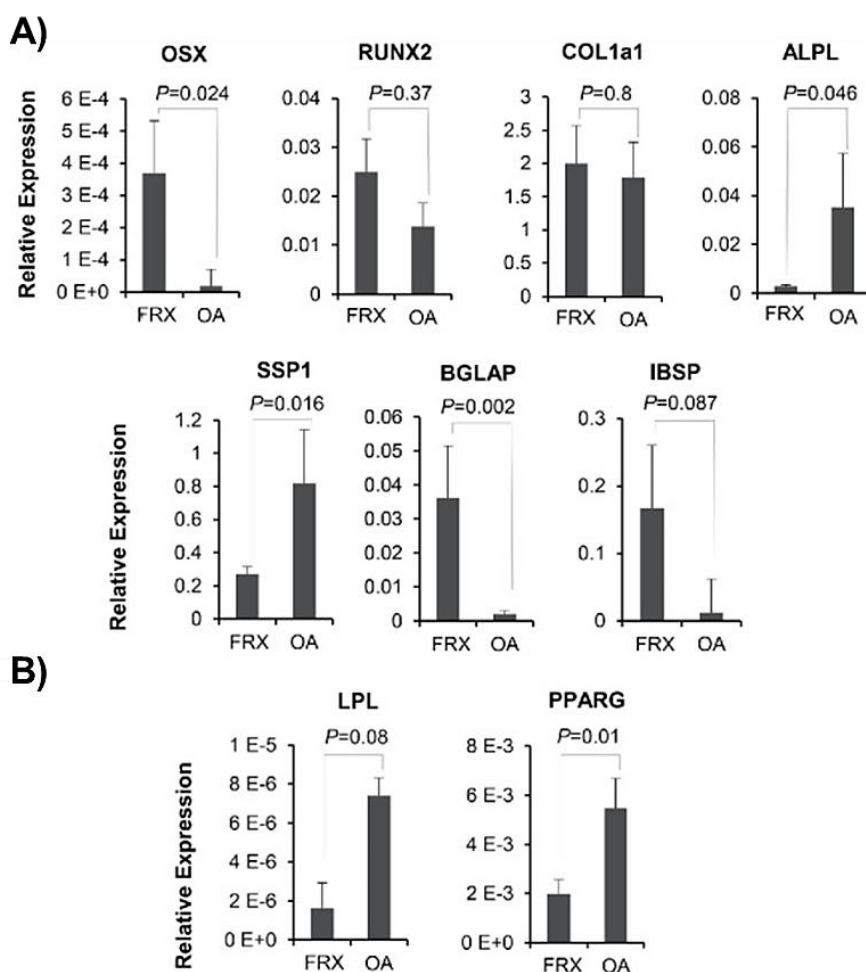


Figure 26. A) Bar plots showing the relative expression of some specific osteogenic and adipogenic markers by BMSCs from patients with fractures (FRX) and osteoarthritis (OA). B) LPL and PPARG genes in both groups of BMSCs (FRX and OA). T test is used for the statistical differences.

BMSCs from FRX and OA were differentiated into osteoblasts in vitro, and both had the capacity to respond to the osteogenic induction. However, there were some significant differences, because the capacity to form a mineralized matrix was markedly decreased in BMSCs from FRX ($p=0.00015$), as shown by alizarin red staining (Figure 27A). Additionally, alkaline phosphatase activity was also lower in BMSCs from FRX than in those from OA (Figure 27B), and there was a positive correlation between matrix mineralization and alkaline phosphatase activity (Spearman's $r=0.84$, $p<0.001$). Surprisingly, OSX and RUNX2 expression tended to be higher in BMSCs isolated from patients with osteoporosis ($p=0.34$ and 0.043 , respectively).

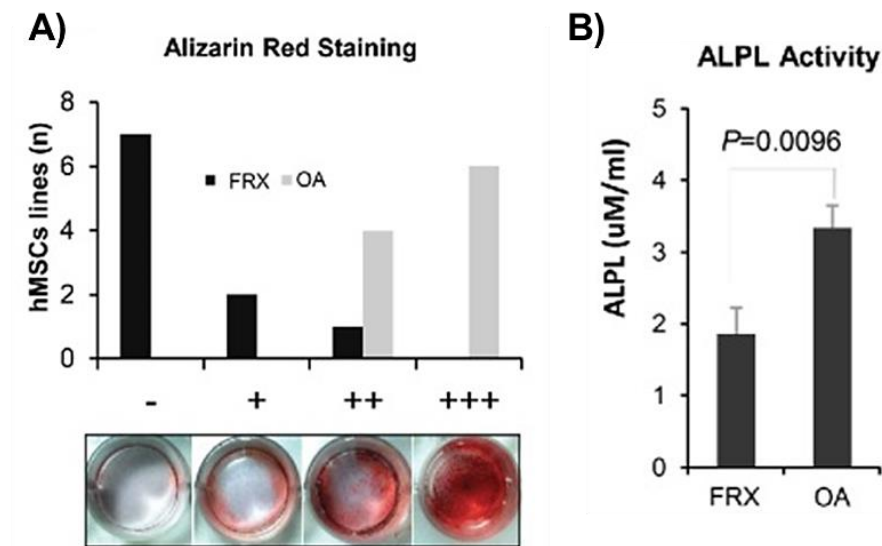


Figure 27. Differentiation capacity of BMSCs. (A) Osteogenic differentiation evaluation after alizarin red staining of BMSCs from patients with fractures (FRX) and with osteoarthritis (OA), semiquantitative analysis by double blind test. (B) Alkaline phosphatase activity in these BMSC maintained in osteogenic medium.

Functional experiments with ID2 inhibition

From our results and also some data in the literature, ID2 seemed to be associated with the osteogenic differentiation of BMSCs¹⁹⁵. Hence, we used some siRNA sequences to inhibit ID2 expression in our BMSCs.

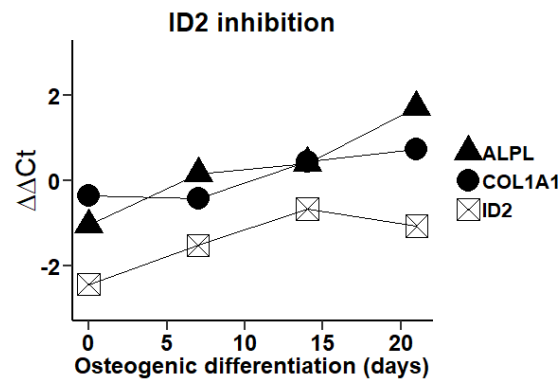


Figure 28. Relative gene expression levels of genes ID2 (squares), ALPL (triangles) and COL1A1 (circles) in BMSCs with osteogenic differentiation media. Levels are measured in the days 1, 8, 15 and 22 after inhibition by siRNA ID2 transfection. They correspond to the basal day and the days 7, 14 and 21 after osteogenic differentiation. $\Delta\Delta Ct$ is calculated from the differences between ΔCt of the samples inhibited with siRNA ID2 and with those transfections with a scramble sequence (negative controls). Each ΔCt corresponds to the normalization of the expression with the housekeeping genes GAPDH and RPL13A.

We found significant difficulties in transfecting BMSCs, nevertheless we obtained some inhibition of ID2 expression. We did not find significant effects on the expression of the osteogenic markers COL1A1 and ALPL, neither at baseline nor during osteogenic differentiation in vitro. Moreover, we did not observe differences in ID2 gene expression before and after osteogenic differentiation of BMSCs in vitro (Paired t test; $n=17$; $p=0.0968$) (data not shown).

Expression of long non-coding RNAs

Since ncRNAs are increasingly recognized as important factors in the regulation of cell differentiation, we next focused our attention on the non-coding regions of the genome. Specifically, the aim of this analysis was to determine the expression of lncRNAs in MSCs from patients with FRX or OA and its relation with DNA methylation marks and the expression of protein-coding genes.

DNA methylation analysis showed that the frequency of differentially methylated CpG sites ($FDR < 0.05$ and Beta differences > 0.1) was similar in non-protein coding transcribed regions (946 out of 53084; 1.8%) and in other genomic regions (8092 out of 424625; 1.9%).

Transcriptome analysis using different softwares (EdgeR and DESeq2) showed 118 non-protein coding genes from 234 total genes differential expressed ($FDR < 0.05$; $FC > 2$) with EdgeR software; And 59 non-protein coding genes of a total of 140 genes differentially expressed with DESeq2 software. The implemented softwares are different as those used in the previous analyses (EdgeR and AltAnalyze) because they are using distinct normalization statistical methods and they usually have diverse results. This is important to elucidate a greater number of possible target genes.

Among differentially expressed ncRNAs, most of them belonged to the antisense type (72%), followed by lincRNAs and sense overlapping lncRNAs. Antisense lncRNAs tend to regulate neighboring protein-coding genes (cis regulation). Our data are in line with this concept. In fact, approximately 50% of the protein-coding genes in cis position of differentially expressed antisense lncRNAs were also differentially expressed. Furthermore, enrichment analysis of these associated lncRNAs-protein coding genes pairs showed significant overrepresentation in bone related pathways including “regulation of ossification” and “osteoblast differentiation”.

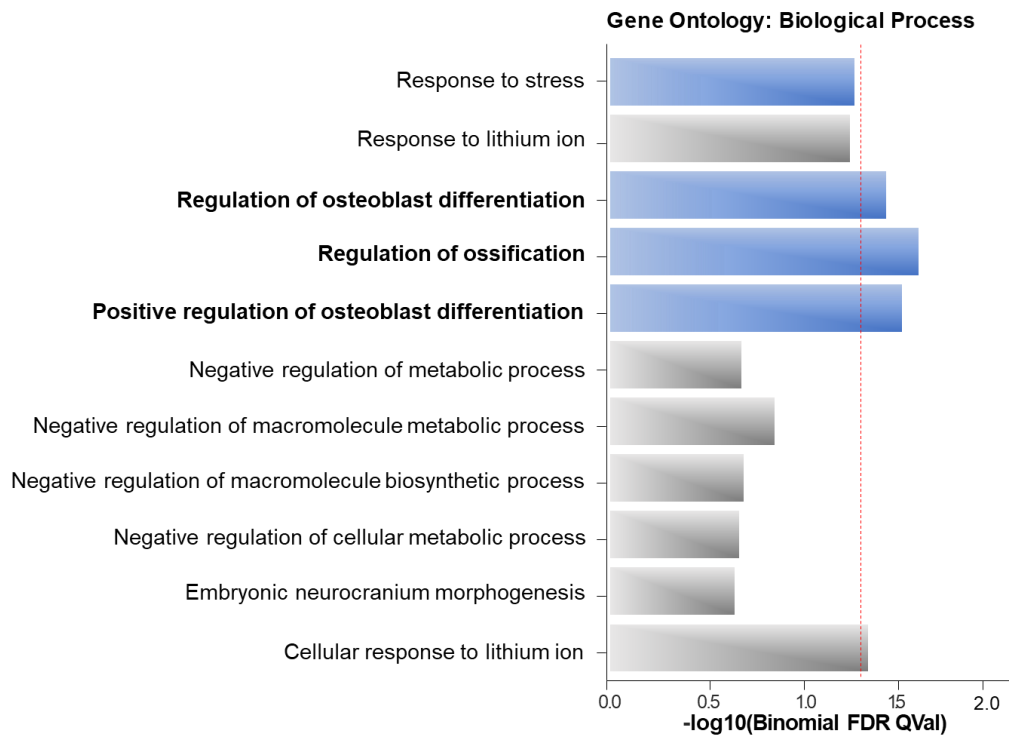


Figure 29. Gene ontology (GO) enrichment analysis (biological processes), of those protein coding genes linked to differentially expressed antisense lncRNAs of BMSCs derived from FRX and controls with OA. Coloured in blue, various interesting bone related processes. FDR=0.05 is marked by the red dashed line.

Replication of RNA sequencing

Given the importance of the noncoding gene expression, we aimed to increase our sample size and performed another RNA sequencing experiment including a new independent set of BMSC RNAs. Specifically, we analyzed 16 pre-differentiated samples (8 FRX, 8 OA), as well as 3 post-differentiated samples (paired with predifferentiated ones), maintained in osteogenic culture medium for three weeks.

Since RNA sequencing had been performed by two different external companies, we re-analyzed the raw data in house to be able to apply equivalent analysis procedures to both datasets.

Before differential gene expression analysis, we performed several quality control procedures. Fastq files were checked for their quality on raw sequences data with FastQC tool, and all of them passed. Sequences were aligned using the Human GRCh37/Hg19 as the reference genome with TopHat2/Bowtie software, and then

cufflinks programs were implemented to assemble all transcriptomes and quantify their expression.

Differential gene expression analysis was performed, similarly to the previous analysis, with DESeq2, EdgeR and CuffDiff softwares. Before the final analysis, additional quality control tests were implemented to detect possible outliers. First, the square of the coefficient of variation (CV^2) was used to evaluate the quality of RNA-seq. The differences in CV^2 may result in a low number of differentially expressed genes due to a high degree of variability among the FPKMs of the estimated replicates. Quite reassuringly, we did not find significant differences (Figure 30A).

Also, we built a dendrogram using all genes to identify samples that were separated from the group of origin. We found two samples (1 FRX and 1 OA) grouped apart (Figure 30B, marked in red). Next, we implemented dimensional reduction, which is a method that serves to cluster samples and explore the relationship between conditions. It can be used to identify variability in the data. We applied multidimensional scaling (MDS) strategy and found the same two samples in apparently different clusters (surrounded in red Figure 30C).

RNAseq data sometimes contain very large counts that are not related to the experimental or study design, and that can be considered outliers. They may result from technical or experimental artifacts, mapping problems, and rare biological events. We made a box plot of Cook's distances to see if some samples were consistently higher than others, and found that the same samples considered as potential outliers in the dendrogram and multidimensional scaling had a higher cook's distance (Figure 30D). Hence, we discarded those two samples for the differential expression analysis and compared the results with and without them. Nevertheless, the results were very similar, suggesting that the results of those two samples was not introducing a significant bias in the results.

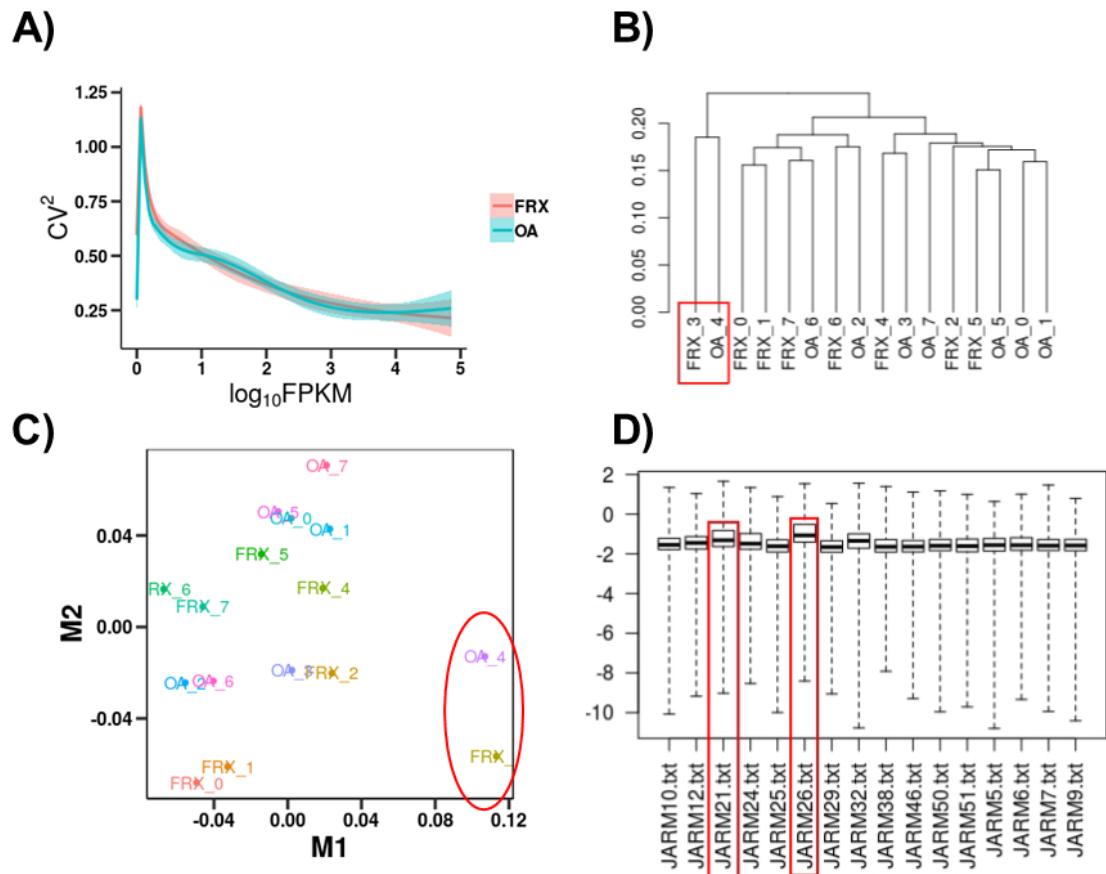


Figure 30. Various quality tests of the sequenced and aligned samples. (A) The square of the coefficient of variation (CV^2) of both BMSCs groups (FRX in red and OA in blue). (B) Dendrogram using all genes, grouping the samples by hierarchical clustering. Two samples are separated from the principal group, surrounded in red. (C) Multidimensional scaling (MDS) test for all samples with the same two samples in apparently different clusters (surrounded in red). (D) Box plot showing Cook's distances and surrounded in red, both samples surrounded in (B) and (C), with consistently higher distances than others.

Replication of transcription analysis and signature of osteogenic differentiation

We found 85 differentially expressed genes when comparing pre-differentiated samples, from which 33 were lncRNA type, and relate them with those previously obtained in the prior RNAseq. Concerning the three paired samples, pre- and post-differentiated, there were 163 genes differentially expressed, from which 99 were lncRNA type.

We used the intersected genes ($n=53$) for further analysis, from which 21 were lncRNAs and 32 were protein-coding genes. The intersection group included common genes being present in the group of genes differentially expressed in

predifferentiated BMSCs from FRX and OA, as well as the group of genes differentially expressed between pre- and post-differentiated samples. With this intersected gene list we built a Spearman correlation matrix between lncRNAs and protein-coding genes (Figure 31).

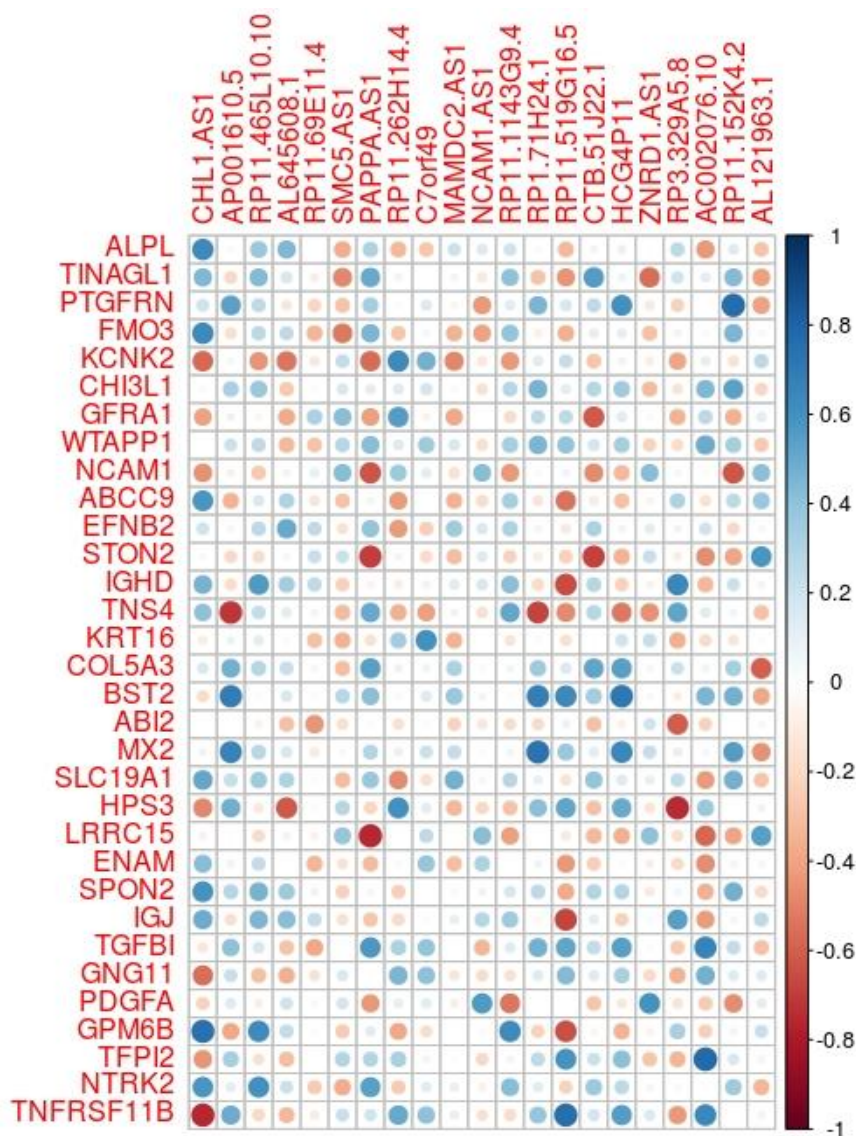


Figure 31. Correlation plot with lncRNA gene type in the columns and protein coding genes in the rows. The size of the circles is proportional to the correlation value (the larger, the higher is the value). Coloured in red are the negative correlation coefficients and in blue the positive correlation coefficients.

In the following table we show the 21 differentially expressed lncRNAs from the intersection analysis explained above, and the cis-associated protein-coding genes (Table 10).

GENE	logFC	FDR	Cis.genes
AL121963.1	-2.88	4.48E-13	COL10A1
AL645608.1	2.80	3.59E-02	SAMD11
AP001610.5	-4.50	3.77E-25	MX1
C7orf49	-1.78	3.30E-03	TMEM140
CHL1-AS1	2.87	2.37E-04	CHL1
CTB-51J22.1	-3.70	4.98E-07	ELN
HCG4P11	-1.50	3.59E-02	HLA-F
MAMDC2-AS1	-2.74	1.98E-03	MAMDC2
NCAM1-AS1	-3.53	6.53E-06	NCAM1
PAPPA-AS1	1.09	7.61E-02	PAPPA
RP1-71H24.1	-4.27	1.59E-19	OAS1;OAS3
RP11-1143G9.4	5.10	1.26E-26	LYZ
RP11-152K4.2	2.36	4.31E-02	CDH6
RP11-262H14.4	5.83	8.99E-03	Intergenic
RP11-465L10.10	1.76	8.16E-03	SLC12A5;MMP9
RP11-519G16.5	-3.29	4.45E-09	C15ORF48
RP11-69E11.4	-2.84	2.96E-03	BMP8A
RP3-329A5.8	4.56	4.39E-02	SCUBE3
SMC5-AS1	-3.05	2.75E-03	MAMDC2
WTAPP1	-4.68	2.71E-12	MMP1&3
ZNRD1-AS1	-1.27	2.01E-02	ZNRD1

Table 10. List of differentially expressed lncRNAs which are common in all performed analysis. Predifferentiated samples analysis with FRX and OA, and an analysis with paired pre- and post-differentiated samples. The fourth column shows the nearest protein coding gene associated to each lncRNA.

We used those results to try to select candidates for further replication and functional studies. The selection was based upon the following criteria:

- Candidates show gene expression differences between FRX and OA.
- Candidates show expression differences between pre- and post-differentiated samples.
- The associated protein coding genes (ie, genes in the nearest –cis-position; or highly correlated in the correlation matrix), are preferentially bone-related.
- There is previously published evidence for their role in skeletal biology.

According to those criteria, we chose two antisense lncRNAs as candidates, PAPPAS1 and CTB-51J22.1:

- PAPPAS1 was more expressed in OA than in FRX, but there was considerable individual variability. In both groups It was expressed more abundantly in pre-differentiated MSCs than in post-differentiated cells, thus, it is downregulated when MSCs undergo osteogenic differentiation (Figure 32A). The PAPPAS gene (its antisense protein-encoding gene) is involved in the development of the skeleton. KO mice not expressing PAPPAS show smaller skeleton and delayed ossification.
- CTB-51J22.1 was less expressed in FRX than in OA, and was also less abundant in differentiated than in pre-differentiated MSCs, although the difference was very small in the FRX group (Figure 32B). Elastin (ELN) is its antisense protein-coding gene. KO mice for the ELN gene have abnormal epiphysis, but the most common affections are those of the cardiovascular system.

Given the relatively small differences in gene expression between FRX and OA and the large variability, we decided to select other additional lncRNAs that showed marked pre-/ post-differentiation differences. Thus, we chose four more lncRNAs to replicate: LINC00341, LINC02008, LINC01279 and PTGS2-AS.

- LINC00341 is a non-coding long intergenic RNA in the 3'UTR position of the SYNE3 gene and 5'UTR of the calmin gene (CLMN). It was similarly expressed (very low expression) in FRX and OA, but markedly increased in differentiated cells (Figure 32C). In addition, with the BLAT tool for alignment on the genome, certain regions are found in common with certain coding genes, such as LTBP2, that is involved in bone development.
- LINC02008 is an intergenic long non-coding RNA found in the 5'UTR position of GBE1 gene. It was not expressed in the pre-differentiated samples, but increased after differentiation (Figure 32D).
- LINC01279 is a long non-coding intergenic RNA that is located between the coding genes CCDC80 and SLC35A5. It is more expressed in pre-differentiated than in differentiated cells, so it is downregulated after

osteogenic differentiation. There were no differences between fractures and osteoarthritis (Figure 32E).

PTGS2-AS is also known as PTGS2 Antisense NFKB1 Complex-Mediated Expression Regulator RNA, or P50-Associated COX-2 Extragenic RNA (PACERR). It was downregulated after osteogenic differentiation and there were not significant differences between FRX and OA (Figure 32F).

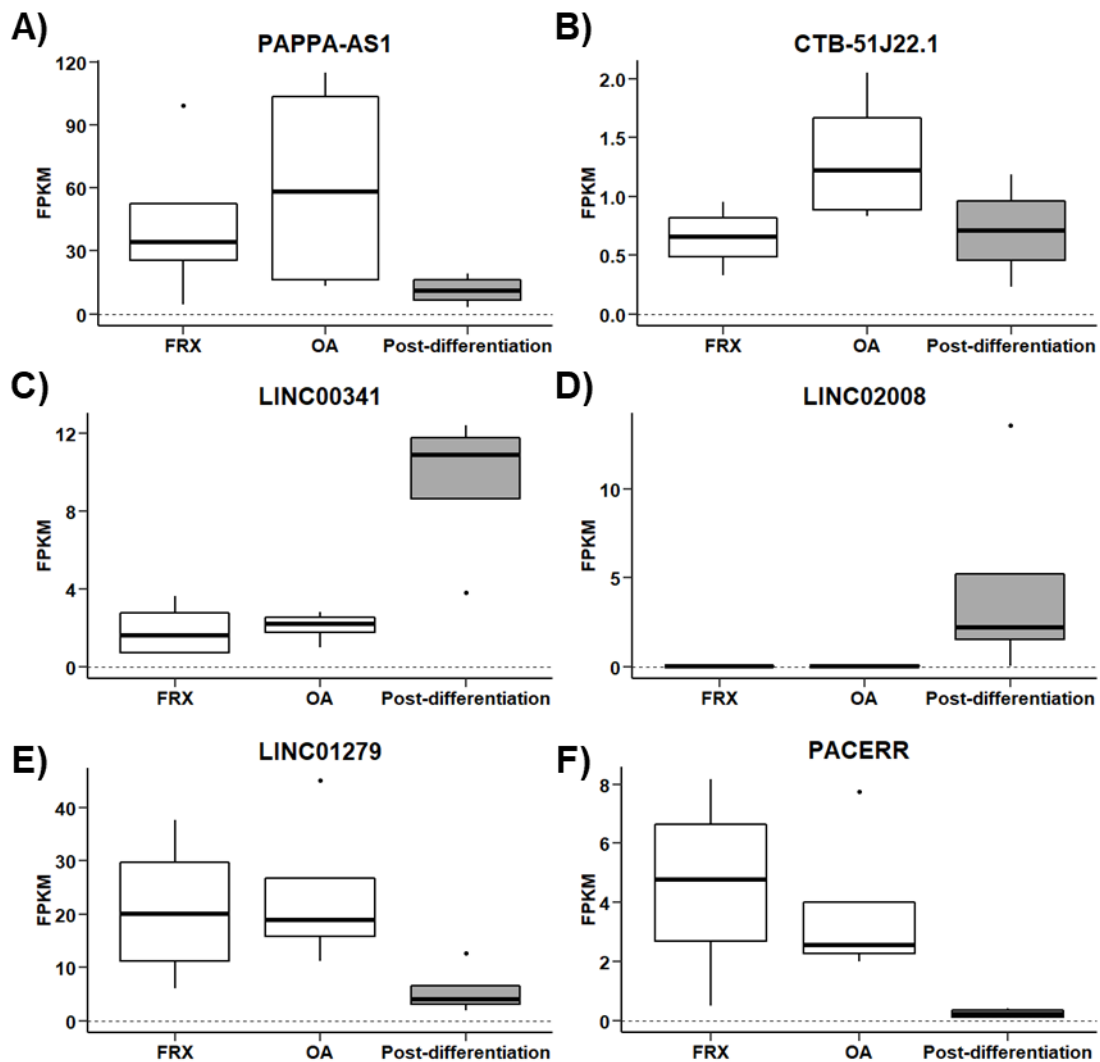


Figure 32. Gene expression levels of the selected lncRNA from the RNA sequencing normalized data (FPKM), distinguishing between FRX and OA in the basal state (white boxes) and differentiated samples after 21 days of osteogenic induction (gray boxes).

Long Non-coding RNA replication and functional studies.

We replicated the gene expression of the chosen lncRNAs by real time qPCR with sybr green and primers, previously designed (Table 4). In all cases dissociation curves after amplification confirmed that our PCR was gene specific.

The expression was validated for two lncRNAs, LINC00341 and PACERR (Figure 33).

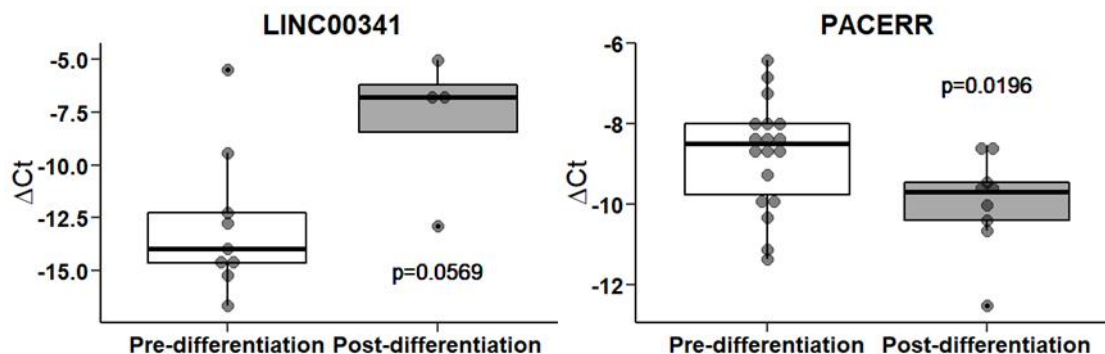


Figure 33. Replication of the gene expression of LINC00341 (on the left) and PACERR (on the right) by real time PCR. Comparing BMSCs in the basal state (white boxes) with differentiated BMSCs (gray boxes). For the statistical analysis, T test has been used. ΔC_t are calculated from the difference between C_t of the lncRNAs genes and C_t from the reference genes (GAPDH and TBP)

Both lncRNAs are outside protein coding genes, PACERR is upstream head to head with the protein coding gene and LINC00341 is intergenic. However, the rest of experimented lncRNAs were not replicated by qPCR. As we saw before the major part of differentially expressed lncRNA are antisense type, which are within the protein coding sequence, and this is the case of PAPPA-AS and CTB-51J22.1. For distinguishing their transcription from those of the associated protein coding gene, we had to use different reverse transcription strategies. In theory, if we used random hexamers and oligodT as primers to anneal RNA, we obtained cDNA of the complete transcriptome, and in this case it does not permit to study the coding and the non-coding fragments independently. Therefore, we had to use strand specific reverse transcription with the corresponding primer for each strand. Besides, we also used a reverse transcription without primers (only enzyme) as a negative control. But, somewhat unexpectedly in all these reverse transcriptions we obtained sequence amplicons by qPCR. This fact meant that we could not discriminate between both strands in the reverse transcription process, and the results were not strand specific (Figure 34).

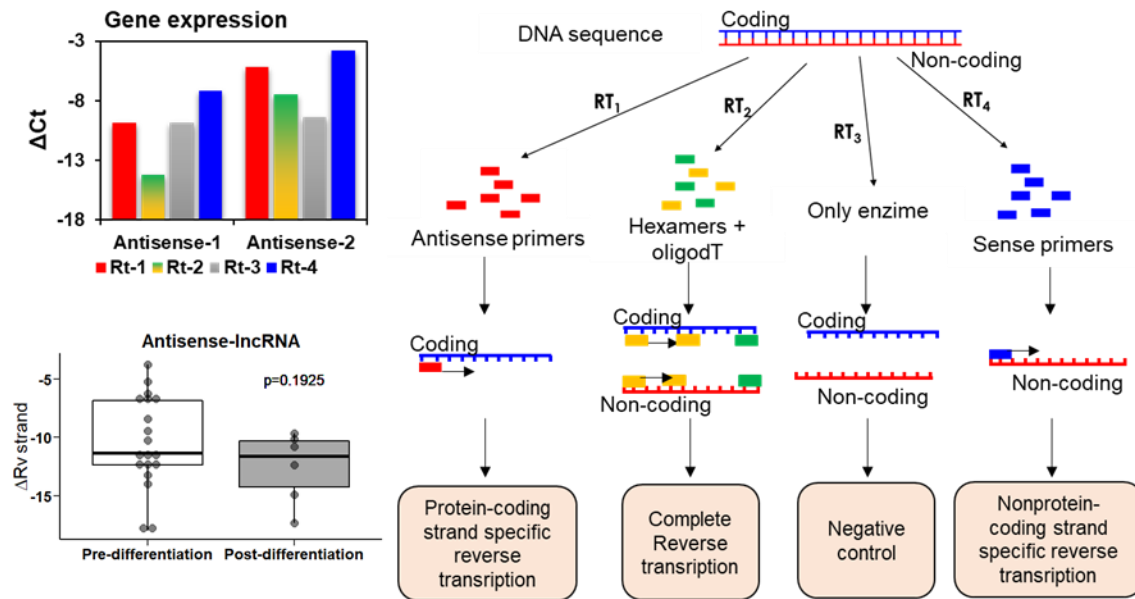


Figure 34. Study of the antisense lncRNA region by real time PCR, using four different reverse transcription reactions (graphical scheme on the right). On the left, expression levels after reverse transcription theoretically distinguishing both strands of two different antisense lncRNAs. ΔC_t are calculated from the difference between C_t of the antisense genes and C_t from the reference genes (GAPDH and TBP).

Finally, we focused on and the two replicated lncRNAs (LINC00341 and PACERR) by real time qPCR experiments to confirm the biological role of those genes in bone formation. The strategy used was to transfect expression vectors into human cell lines (HOS and HEK-293T) and BMSCs, and subsequently studying expression changes in genes centrally involved in bone formation (COL1A1 and ALPL). LINC00341 overexpression in HOS cell line induced an increase in the bone-related markers (Figure 35A). This effect appeared to be specific for cells of the osteoblastic lineage, because in HEK-293T cells these changes were not clear (Figure 35C). On the other hand, when PACERR was overexpressed, we did not find consistent changes in either COL1A1 or ALPL, neither in HOS nor in HEK-293T (Figure 35B and D).

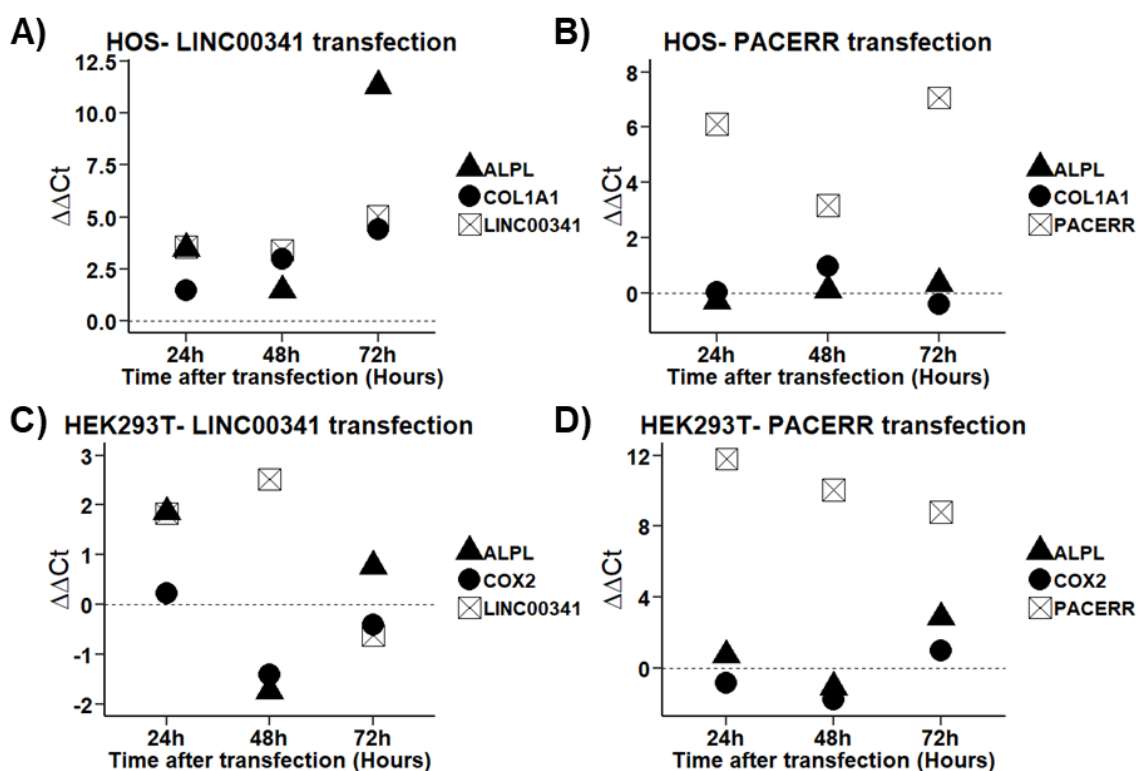


Figure 35. Gene expression changes after transfection analysis in different cell lines (HOS and HEK-293T) with LINC00341 and PACERR expression vectors, at different times (24, 48 and 72 hours). ΔCt are calculated from the difference between the Ct of the lncRNAs and the reference genes (GAPDH and RPL13A). $\Delta\Delta Ct$ refers to the comparison with empty vector transfection (ΔCt lncRNA vector – ΔCt empty vector). ALPL relative expression is marked with a black triangle. COL1A1 (in HOS cells) and COX2 (in HEK-293T cells) are shown with a black circle. Respective LINC00341 or PACERR overexpression is reflected with a box with a 'X' in inside. (A) Transfection of LINC00341 expression vector into HOS cells. (B) Transfection of PACERR expression vector into HOS cells. (C) Transfection of LINC00341 expression vector into HEK-293T cells. (D) Transfection of PACERR expression vector into HEK-293T cells.

We also tried to transfect BMSCs with both expression vectors (LINC00341 and PACERR), independently. Indeed, we tried different methods for transfection, including several lipid-based reagents and electroporation. However, the transfection efficiency was very low and consequently we could not confirm or refute those results in BMSCs (Figure 36).

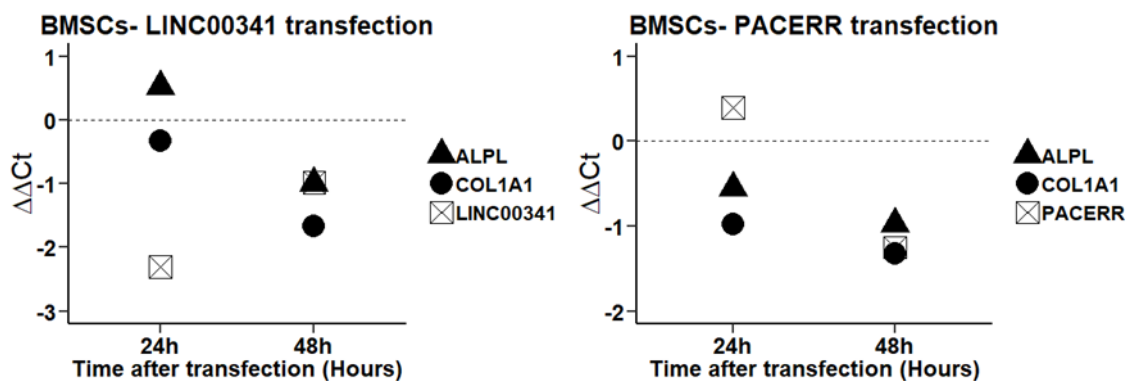


Figure 36. Gene expression levels of genes COL1A1 (circles), ALPL (triangles) and the experimentally overexpressed gene, PACERR or LINC00341 (squares) in BMSCs. Levels are measured after overexpression of LINC00341 (left) or PACERR (right) vector transfection. Relative gene expression compares those transfected wells with our genes of interest with those transfections with an empty vector (negative controls). ΔC_t are calculated from the difference between the C_t of the lncRNAs and the reference genes (GAPDH and RPL13A). $\Delta\Delta C_t$ refers to the comparison with the transfected wells with an empty vector (ΔC_t lncRNA vector – ΔC_t empty vector).

7. DISCUSSION

DISCUSSION

Osteoporosis as a prevalent complex disorder

Osteoporosis is a prevalent bone disorder characterized by low bone mineral density due to an imbalance in the bone remodelling process. This imbalance consists in a predominance in the osteoclastic activity over the osteoblastic one, which leads to bone loss. Genetic studies have confirmed that osteoporosis is a complex disease with a polygenic pattern of inheritance in which several genes are involved, as well as environmental factors, such as nutrition, physical activity, aging, hormone functions and exposure to various substances, like alcohol, tobacco or some medicines ^{49,196}. Osteoporosis is very common in elderly people, with a higher frequency in postmenopausal women than in premenopausal women or in men. Bone remodelling regulation mechanisms are remarkably complex and include a series of specific gene expression changes that lead to cell differentiation along specific pathways, thus allowing specialized cell activities.

Osteoporosis is characterized by an insufficient bone formation, either in absolute terms (low turnover osteoporosis), or in relation to the degree of bone resorption (high turnover osteoporosis). Bone Marrow Mesenchymal stem cells (BMSCs) are pre-osteogenic cells, and have the capacity to differentiate into osteoblasts, the bone forming cells. This thesis focuses on the transcriptional and epigenetic signatures of BMSCs from osteoporotic patients.

MSCs and aging

MSCs can differentiate into a variety of cell types, including adipocytes, osteoblasts, chondroblasts and myocytes. Recently, it has been identified a subpopulation of self-renewing and multipotent MSC that generates progenitors of osteoblasts and chondrocytes and stroma, but not fat, and has been termed “skeletal stem cells” ⁴¹. Osteoblasts are responsible for bone formation. Hence, bone formation deficiency in aging individuals may be caused by age-related changes in the migration, proliferation and differentiation capacity of MSCs.

Indeed, several studies suggest that the pool size and function of MSCs diminishes with aging ¹⁹⁷. BMSCs represent only a minor fraction of the marrow cell population (between 1/10,000 and 1/100,000) ¹⁹⁸, and it has been reported

to be even smaller in aged subjects ¹⁹⁹. MSCs from old individuals appear to have reduced proliferative capacity, with longer doubling times, increased senescence and apoptosis and reduced number of duplication cycles ²⁰⁰. Also, an aging-associated reduced capacity of BMSCs to differentiate into osteoblasts, along a preferential differentiation towards adipocytes has been reported ^{201–203}. For example, Mueller et al found a decline in the differentiation potential of BMSCs from individuals over 60 years of age, in comparison with those below 50. The mechanisms involved may be multiple, including DNA damage, telomere shortening, epigenetic abnormalities, oxidative stress, etc. ⁴³. However, this is by no means a fully elucidated question ²⁰⁴. For example, the group of Stenderup et al found that MSCs from old individuals have an accelerated aging when cultured in vitro, but they maintain a normal capacity of proliferation and bone formation in vivo ^{205–207}. This suggests that age-associated changes may depend not only on cell intrinsic changes, but on changes in the environment where MSCs develop, the so-called “stem cell niche”, and a variety of local and systemic humoral factors ^{204,208,209}. In this line, the role of cell senescence in age-related bone disorders is receiving great attention in recent years. Although the absolute number of skeletal senescent cells seems to be rather small, even in old individuals, they may negatively impact bone homeostasis through the so-called senescence-associated phenotype (SASP), which includes a number of secreted factors impairing the function of neighbor cells ²¹⁰. Moreover, MSCs from different sources may vary regarding age-dependent changes. Thus, Beane et al. reported that BMSCs have impaired proliferation, senescence, and chondrogenic response in association with aging, whereas muscle-derived stem cells and adipose-derived stem cells exhibited no negative effects ²¹¹. While age reduced overall cell yield and adipogenic potential of all MSC populations, osteogenesis and clonogenicity remained unchanged.

MSCs and osteoporosis

Several investigators have explored MSC function in osteoporosis. Thus, it has been reported that BMSCs from postmenopausal women with osteoporosis have a reduced capacity to proliferate and differentiate into osteoblasts, with less calcified nodule formation and expression of collagen and alkaline phosphatase

after osteogenic induction in vitro ^{212–214}. However, such a deficiency has not been confirmed in other studies ²⁰⁵.

In some studies, the impaired differentiation towards an osteoblastic phenotype has been accompanied by a parallel increase in the differentiation towards an adipocytic phenotype, which might be related to the increase in marrow fat frequently observed with aging and osteoporosis ²¹⁵.

It is to note that MSCs are present in the bone marrow (BMSCs), but they may also be present in the circulation, at least in some circumstances. The number of osteogenic circulating precursors has been reported to decrease with aging, particularly among frailty subjects ²¹⁶, but a number of methodological issues limit the validity of those studies. In fact, some investigators could not demonstrate the presence of bona-fide circulating MSCs in normal conditions, although they did in patients with hip fractures or extensive tissue injury and multiple fractures ^{217,218}.

Mesenchymal stem cell migration is an essential step for endochondral or intramembranous ossification during skeletal development, as well as for fracture healing ²¹⁹. The role of circulating MSCs in diffuse bone disorders such as osteoporosis is less clear, but homing of these cells to remodelled areas is of course critical for proper bone formation. Likewise, homing of MSCs to damaged bone areas is a crucial step when systemic infusion of MSCs is planned for regenerative purposes. Thus, new strategies are being applied to enhance MSCs migration to bone tissue (osteotropism), so that they can be infused systematically to heal generalized bone diseases, such as osteoporosis, as well as localized bone defects. In this regard, it is particularly interesting the approach by Sackstein and collaborators, who modified the glycosylation pattern of CD44 ligand to convert it into the so-called hematopoietic cell E-/L-selectin ligand (HCELL), which can bind to expressed E-selectin in bone marrow endothelial cells, thus increasing osteotropism ^{220,221}.

Cell differentiation is regulated by biological, physical and environmental chemical factors. The biological factors include transcription factors, their signalling pathways and miRNAs. Physical or chemical factors include mechanical stimulation, radiation or diet among others ^{222–226}. Transcription factors CEBP and PPARG appears to be critical for adipocyte differentiation ²²⁷, whereas RUNX2 and OSX signals are the master regulators of osteogenesis ²²⁸.

In this regard, it is interesting a recent study by Rauch et al showing that BMSCs are relatively predisposed to differentiate towards the osteoblastic phenotype, so that osteogenesis involves activation of pre-established enhancer, whereas adipogenesis is driven by more profound changes of the chromatin and de novo activation of enhancers ²²⁹. This study is in line with previous observations about the tendency of MSCs cultured in vitro to spontaneously differentiate towards an osteoblastic phenotype ²³⁰.

We analysed BMSCs derived from the femoral head of patients with osteoporotic fractures and controls with osteoarthritis. BMSCs of patients with fractures showed higher proliferation rate, which in theory is an attractive feature if these cells are to be used for regenerative therapy. This was a somewhat unexpected finding. Since BMSCs were obtained from fracture patients, it could be speculated that it was a phenomenon related to the response to bone injury. We avoided the fractured bone edges, yet a regional or even systemic effect of fracture cannot be avoided. In fact, studies suggest that BMSCs are activated and mobilized following tissue injury ^{218,230}.

Despite their higher proliferation rate, BMSCs from patients with fractures presented a diminished ability to express some bone makers, such as alkaline phosphatase, and to deposit a mineralized extracellular matrix. The mechanisms involved are unclear but might be related to an upregulation of the RUNX2 pathway. In fact, we quantified the expression of several genes by qPCR, and observed that the osteogenic transcription master regulators RUNX2 and OSX were upregulated in BMSCs from fracture patients. These two transcription master regulators target other genes, usually related with osteoblastic phenotype, including BGLAP and IBSP and indeed both were also upregulated in fractures. However, late-stage genes of the osteogenic process, such as, ALPL and SPP1, were downregulated in BMSCs from patients with osteoporosis. This fact could be connected with their reduced capacity to differentiate in vitro and form mineralized matrix, and maybe due to the tenacious expression of RUNX2. This is a critical factor for BMSCs commitment towards osteoblastogenesis, but persistent activation may impair the terminal differentiation of osteoblasts ²³¹.

Thus, in the presence of high proliferation rate but reduced terminal differentiation in vitro, it is unclear whether the overall consequence is a preserved bone formation activity of BMSCs in patients with fractures or not. In vivo studies

currently ongoing in our laboratory may help solving this issue. It is to note that, as expected by their different epidemiology, there was an age imbalance between patients with fractures and controls with osteoarthritis. However, it is unlikely that it played a role in the observed differences, as we did not find significant age-related changes in the behaviour of BMSCs, within the age range of the patients included in the present study.

Epigenetic marks of MSCs, aging and osteoporosis

Epigenetic mechanisms are key elements for the gene expression regulation and consequently in the regulation of particular cell physiology and cell differentiation processes.

Several studies have explored the role of miRNAs on osteoblastogenesis⁷⁹. Most miRNAs are negative regulators of osteogenesis and their inhibition is a potential therapy to repress their function in bone diseases, including osteoporosis²³². Some of this molecules are miRNA-204 and miRNA-211, which bind to RUNX2 3'-UTR, thereby inhibiting osteogenesis²³³. miRNA-103a also inhibits bone formation through RUNX2 repression²³⁴. Many more miRNAs have been found to regulate bone remodelling, including miRNA-455-3p, -23a, -30c, -34c, -133a, -135, -137, -204, -205, -217 and -338-3p²³⁵.

DNA methylation of MSCs, aging and osteoporosis

Regarding DNA methylation, another important mechanism, Pasumarthy and collaborators, observed that BMSCs suffer age-related DNA methylation changes associated with gene expression changes. They used BMSCs from young donors (20 to 24 years, n=5) and from aged donors (62 to 82 years, n=5) and concluded that those changes were frequently found within enhancer regions determined by the H3K4me1 histone mark²³⁶. Another study explored the DNA methylation profile of BMSCs obtained from individuals from a wide age range (2 to 92 years old), and they identified 18,735 hypermethylated CpG sites and 45,407 hypomethylated sites, associated with aging. The hypermethylated sites were enriched in repressive chromatin labels. In addition, hypomethylated CpG sites were strongly enriched in the active chromatin brand H3K4me1, related to the enhancer activity²³⁷. Roforth et al. examined the gene expression and DNA

methylation patterns of BMSCs from young (mean age, 29 years) versus old (mean age, 73 years) women. They found significant differences in methylation between the young and old subjects surrounding the promoters of 1528 target genes that also exhibited significant differences in gene expression ⁹⁰.

In our analysis, we have not found any DNA methylation significant differences across individuals of different ages. However, as already mentioned, the age range was rather limited. Specifically, all subjects were above 60 years of age. Therefore, we cannot establish if there are differences in BMSC responses between elderly patients and young or middle age subjects.

DNA methylation levels of specific CpG sites depend on genetic factors, environmental factors, and lineage-specific cues, apart from certain random variation. Methylation levels also tend to change with aging, and specifically some age-related changes are quite universal at some CpG sites. This has led to several groups to propose the analysis of those CpG sites as an index of the “epigenetic aging” of the individual. There are three main tools for analysing blood DNA methylation age. The first one, developed by Hannum and collaborators, was built with 450k array data from whole blood of 656 human individuals, ranging between 19 and 101 years’ old ²³⁸. Another one that has used only blood samples is from Weidner and collaborators. They have found three specific age-related CpGs and correlated them with chronologic age with an absolute deviation of at most 5 years ²³⁹. Levine and colleagues have created a software to use DNA methylation levels as an epigenetic biomarker of aging phenotypes. The current version, called “DNAm PhenoAge” includes a variety of aging-related outcomes, all-cause mortality, familiar longevity, socioeconomic status factors, among others ²⁴⁰.

However, the most extensively applied is the method proposed by Steve Horvath. Horvath described a set of 353 CpG sites showing age-related changes in a wide variety of cells and tissues, which makes this software particularly useful for studies of many tissues, not only blood projects ¹⁷⁵. There have been many studies in recent years, with Horvath’s software. For example, it has been used in lymphoma ²⁴¹, alcohol dependence ²⁴², breast cancer ²⁴³, brain tissue in Huntington’s disease ²⁴⁴ and cerebellum ²⁴⁵.

Using Horvath’s software, we have seen that BMSCs isolated from fracture patients have an accelerated epigenetic aging when compared with BMSCs from

controls with osteoarthritis. In theory, we could speculate that such accelerated aging could somewhat impair the activity of BMSCs, thus contributing to a reduced bone formation. However, in bone tissue there were no differences between both groups. This could sound paradoxical, but it is important to note that most cells in bone tissue samples are cells of the osteoblastic lineage, and specifically osteocytes, not BMSCs. Osteocytes are non-dividing cells with a very long lifespan (estimated as 10-20 years). Since changes in methylation may occur more rapidly during DNA replication, the methylation of cells in bone tissue samples may not reflect the current status of BMSCs and other actively proliferating cells. Additionally, osteocytes lifespan is dependent on bone turnover. If bone turnover is lower in older patients with osteoporosis, as suggested in some studies, osteocytes in patients with fractures may be older than those from controls with osteoarthritis of the same age. In other words, when current osteocytes were born, patients with fractures were younger than patients with osteoarthritis. Nevertheless, for the time being this remains a speculative explanation.

Whatever the real explanation might be, it is important to note that, different from genetic features, epigenetic marks vary across tissues. In this line, in a previous study in which we collaborated, an accelerated epigenetic aging was observed in cartilage cells from patients with osteoarthritis, but not in blood cells ²⁴⁶. Additionally, Fernandez-Rebollo and colleagues did not find differences in the epigenetic age of blood samples from patients with OP and controls without OP ²⁴⁷. Similarly, Morris and collaborators that did not find a correlation between DNA methylation age in blood and bone mineral density levels ²⁴⁸. Overall, those results emphasize that epigenetic marks are not only disease-specific, but also tissue-specific.

DNA methylation is the most studied epigenetic mark, and a huge number of studies have demonstrated its important role as a regulator of gene expression patterns in multiple tissues. Hence, studying DNA methylation in bone cells in normal and pathological conditions may permit to identify new candidate genes regulated by these marks, which eventually could be used as new therapeutic targets. At the same time, some marks might be helpful for diagnosis and/or treatment monitorization.

Other studies have investigated differences in DNA methylation patterns in blood samples from patients with osteoporosis versus people without osteoporosis ²⁴⁸. However, DNA methylation is a tissue specific mark and diverse studies have not found a significant association between blood DNA methylation levels and BMD ^{247,248}. Thus, blood seems not to be the right tissue to study osteoporosis. However, DNA methylation levels in blood of patients with osteoporosis remains a controversial issue. Rather surprisingly, in a small study, Cheisvili and collaborators found that women at risk of developing osteoporosis can be diagnosed by using whole blood DNA methylation analysis ⁸⁹.

In a previous study from our lab, we explored genome-wide DNA methylation in bone tissue samples using a 27k array. This work used bone samples from osteoporotic patients and from osteoarthritic patients as controls, and the comparison identified several differentially methylated regions, which were enriched in genes associated with skeletal embryogenesis, like homeobox genes, indicating the possible existence of a developmental component in osteoporosis ⁸⁷.

In the present study, we looked for DNA methylation differences between BMSCs isolated from patients with osteoporosis and controls with osteoarthritis. To our knowledge, this is the first epigenome-wide and transcriptome-wide study in BMSCs isolated from patients with osteoporosis. There are a few epigenome-wide studies in osteoporosis in other cell types, such as, bone tissue ^{87,249} or circulating leukocytes ²⁴⁸.

We explored more than 450000 CpG sites with the Illumina Infinium 450K Human Methylation array. We demonstrated that DNA methylation patterns of BMSCs isolated from patients with osteoporotic hip fractures and controls showed considerable differences at 9038 CpG sites. For the statistical comparison of the groups the values considered were an FDR< 0.05 and differences in methylation values greater than 10%, as threshold of significance. Those are the standard values used for this type of studies. These differentially methylated sites were distributed throughout the genome, involving promoter regions (217 regions), CpG islands (40 regions), within various gene bodies (129 regions) and enhancer regions (1684 regions). Thus, they were enriched in enhancer regions, which are commonly associated with the regulation of their genes in cis position. Enhancers are present thousands of bases away from the gene and function as amplifiers of

gene expression; As previously described, DNA methylation marks have been shown to vary with aging in enhancer regions. We looked at genes associated to the differentially methylated enhancers (approximately 1400 genes) and observed that they were enriched among stem cell and bone-related pathways. Enhancers regulate the transcription of particular genes, by the action of binding proteins (transcription factors) which in turn regulate the transcription process through diverse mechanisms. Regarding DNA methylation levels, the most common view is that high levels of DNA methylation tend to diminish gene expression and vice versa. Nevertheless, we have seen that this relationship is certainly variable in enhancer regions, where all four possible combinations of DNA methylation and gene expression were observed. Additionally, these data suggest that the methylation of regulatory regions, distant from gene bodies, may be more important than the methylation of gene bodies in determining cell phenotype.

Recently, a study from Rauch examined the transcriptional and epigenomic programming during adipocyte and osteoblast lineage determination of MSCs. They observed that during the differentiation of MSCs (independently of the tissue of origin), adipogenesis needs changes of a greater number of genes than those needed for osteogenesis. They also proved that osteoblast differentiation of BMSCs is induced by the concerted action of a subgroup of transcription factors (TEAD1, TEAD4, NKX3-1, FLI1, MEF2A, HIF1A and SNAI2), which are already active in undifferentiated BMSCs. Further activation of these factors drives osteogenesis and blocks adipogenesis. Furthermore, they correlated this gene expression changes with enhancer regions defined by DNase I hypersensitive sites. Hence, they showed that BMSCs promote osteogenesis by the activation of pre-established enhancers, whereas adipogenesis needs the activation of a new establishment of enhancers ²²⁹. In our dataset, we found some DNA methylation changes in 6 of those 8 transcription factors (TEAD1, HIF1A, SNAI2, MEF2A, SMAD3 and FLI1). All of them with DNA methylation changes in enhancer regions too. However, we did not see gene expression differences in these factors (data not shown), so that, both groups (osteoporosis and osteoarthritis) seem to have a similar transcriptional network of transcription factors.

In a hypothesis-free genome-wide approach, using the Infinium methylation array and next generation sequencing of the transcriptome, the comparison of BMSCs from osteoporotic patients and controls revealed differences in the methylation and expression of genes enriched in several pathways related to cell differentiation and osteogenesis, thus suggesting that indeed differences in methylation marks contributed to differences in BMSC function. Among those genes, we chose 10 to replicate by qPCR; 4 of those genes were significantly replicated in terms of gene expression (ID2, OPG, SPARC and UNC5B). These four genes were also replicated by pyrosequencing, to validate DNA methylation results from array. Both techniques, pyrosequencing and qPCR, showed a high capacity to replicate DNA methylation values from array and gene expression levels from RNAseq, respectively. Some characteristics of these four validated genes are explained below.

ID2: DNA-binding protein inhibitor ID2 according to UniProt database ²⁵⁰ is a “transcriptional regulator (lacking a basic DNA binding domain) which negatively regulates the basic helix-loop-helix (bHLH) transcription factors by forming heterodimers and inhibiting their DNA binding and transcriptional activity. It is implicated in regulating a variety of cellular processes, including cellular growth, senescence, differentiation, apoptosis, angiogenesis, and neoplastic transformation. Its role in bone biology is unknown. A study observed that ID2 is upregulated by IL-27 in osteoclast precursors, through EGR-2, and this upregulation represses RANKL-mediated osteoclastogenesis ²⁵¹. In our BMSCs, ID2 was downregulated in patients with fractures in comparison with those with osteoarthritis. Moreover, DNA methylation of the enhancer region associated with ID2 was inversely correlated with gene expression, so that the higher the levels of DNA methylation, the lower the values of gene transcription. In fact, methylation levels of the enhancer region of ID2 were higher in fractures. Additionally, we investigated if there were some ID2 gene expression changes after culturing BMSCs in osteogenic medium, and we found that ID2 was significantly upregulated. Thus, our results and those of the literature suggest that ID2 could exert a beneficial effect on bone mass, associated with anti-osteoclastogenic and pro-osteoblastogenic actions. Unfortunately, our attempts to block ID2 in cultured BMSCs were unsuccessful. Therefore, elucidating

whether the increased methylation/reduced expression in BMSCs from patients with fractures plays a role in the pathogenesis of bone fragility will need further functional studies.

OPG: TNF Receptor Superfamily Member 11b, also known as osteoprotegerin (OPG), is a receptor inhibitor normally secreted by osteoblasts and other cells. OPG acts as decoy receptor for TNFSF11/RANKL and thereby neutralizes its function in osteoclastogenesis, inhibits the activation of osteoclasts and promotes osteoclast apoptosis in vitro. Thus, bone homeostasis seems to depend on the local ratio between TNFSF11 and TNFRSF11B. Because of its important function in the bone remodelling process, OPG is a well-studied protein in osteoporosis. In a previous study from our lab, Delgado-Calle and collaborators observed that the DNA methylation has a repressive influence on the expression of OPG and RANKL ¹⁶⁶. There is no doubt that the balance OPG/RANKL is important to regulate osteoclastogenesis. In fact, in many experimental models, blocking RANKL by genetic methods or by the infusion of OPG markedly inhibits osteoclastogenesis ²². Also, in this line, denosumab, an anti-RANKL neutralizing antibody, is commonly used as an antiosteoporotic drug ²⁵². Several cells in the bone microenvironment produce RANKL and OPG. It has been suggested that BMSC-produced OPG inhibits osteoclastogenesis in vitro ²⁵³. However, its actual role in vivo is unknown. In our study, OPG expression was increased in BMSCs from fracture patients, which might seem counterintuitive. Additionally, we did not see differences between undifferentiated BMSCs and BMSCs differentiated in osteogenic medium.

UNC5B: Unc-5 Netrin Receptor B gene is a member of the netrin family of receptors. According to RefSeq database ²⁵⁴ “this particular protein mediates the repulsive effect of netrin-1 and is a vascular netrin receptor. This encoded protein is also in a group of proteins called dependence receptors (DpRs) which are involved in pro- and anti-apoptotic processes. Many DpRs are involved in embryogenesis and in cancer progression”. Despite its importance in vascular functions, netrin-1 and UNC5B axis have been demonstrated to play an important role in both, osteoclast and osteoblast biology. In osteoclasts, this axis results in an impairment in actin polymerization and fusion, preventing bone erosion ²⁵⁵.

UNC5B is expressed in osteoblasts, and, if it is knocked down in combination with other netrin receptors, alkaline phosphatase expression is diminished ²⁵⁶. In the case of our BMSCs, UNC5B was upregulated in FRX and DNA methylation of its associated enhancer region is significantly hypomethylated in FRX. Besides, UNC5B is also significantly upregulated after osteogenic differentiation in vitro, which is concordant with literature. Overall, these results could be consistent with the concept that UNC5B expression is involved in the response of BMSCs to promote osteogenesis after fracture.

LOXL2: Lysyl Oxidase Like 2 gene, is a protein coding gene that according to RefSeq “is essential to the biogenesis of connective tissue, encoding an extracellular copper-dependent amine oxidase that catalyses the first step in the formation of crosslinks in collagens and elastin. A highly conserved amino acid sequence at the C-terminus end appears to be sufficient for amine oxidase activity, suggesting that each family member may retain this function. The N-terminus is poorly conserved and may impart additional roles in developmental regulation, senescence, tumor suppression, cell growth control, and chemotaxis to each member of the family”. LOXL2 is very important for the formation of collagen fibers, which are the main component of the osteoblast-secreted bone matrix. This gene has been studied in primary mouse calvaria cells during osteogenic differentiation, and it is suggested that besides its role in the formation of mature collagen fibers, LOXL2 may be an important regulator of osteogenesis ²⁵⁷. Our transcriptome results showed hypermethylation and decreased expression of LOXL2 in BMSCs from fracture patients. Interestingly, LOXL2 was downregulated after osteogenic differentiation, which might suggest that this gene is necessary for early stages rather than later stages of osteogenic differentiation.

MSCs, lncRNAs and osteoporosis

Accumulating experimental evidence highlights the role of long non-coding RNAs (lncRNAs) in bone biology, including osteogenesis. Indeed, in the last decade studies in this field have raised exponentially, highlighting the importance of lncRNAs in gene expression regulation in many cells and tissues, including MSCs

and bone. A few studies have explored the role of lncRNAs in osteoblast differentiation, but their actual roles remain largely unknown. Some lncRNAs promote osteogenesis (H19, MEG3, MODR and MALAT1), whereas others tend to inhibit osteogenesis (HOTAIR, DANCR and MIAT)^{258,259}.

In our analysis of BMSCs from fracture patients and controls, differentially expressed lncRNAs were enriched in the antisense type. Antisense lncRNAs are transcribed from the opposite strand (antisense strand), to regulate sense transcription. An antisense transcript may act by controlling neighbour protein-coding genes in the sense strand (cis regulation) or may act on distant genes (trans regulation). Recent studies of the tri-dimensional organization of the chromatin have shown that DNA looping and organization in topologically associating domains (TADs) permit interactions between distant genomic regions and their transcripts^{79,260}. Antisense lncRNAs might also silence the expression of its juxtaposed gene, through DNA methylation of the associated CpG island^{261,262}.

We associated the differentially expressed antisense lncRNAs to their cis protein-coding genes and observed that the majority of them were also significantly dysregulated. Moreover, these associations were highly enriched in pathways related to bone metabolism. Additionally, we found that differentially methylated CpG sites, were similarly distributed between regions corresponding to coding and non-coding RNAs, thus suggesting a complex interplay between both transcription regulatory mechanisms, DNA methylation and lncRNAs.

In parallel, we compared pre-differentiated and post-differentiated BMSCs after an osteogenic induction, *in vitro*, for three weeks. This analysis was performed with paired pre-post samples. In order to identify candidate lncRNAs potentially important in BMSC differentiation and osteoporosis, we compared lncRNA signatures in osteoporosis and controls, as well as in undifferentiated and differentiated BMSCs. Then, we chose as potential candidates lncRNAs that a) were differentially expressed in fractures and controls; b) were differentially expressed following BMSC differentiation; c) were associated to bone-related protein coding genes in our own data, in the literature or in knock-out mice and other bioinformatic databases²⁶³. Finally, we chose 6 lncRNAs as candidates to be taken to further study: 2 antisense type lncRNAs (PAPPA-AS1 and CTB-51J22.1), 3 lincRNAs (LINC00341, LINC02008 and LINC01279) and a divergent

lncRNA (PACERR). Since we could not reliably separate the expression of sense and anti-sense transcripts by using strand-specific RT-PCR, we finally replicated the expression of two lncRNAs (LINC00341 and PACERR).

Some characteristics of these 2 replicated lncRNAs genes are explained below.

LINC00341: long intergenic non-protein coding RNA 341. It is located in the 3'UTR position of the SYNE3 gene and 5'UTR of the calmin gene (CLMN). Both protein-coding genes are related to the actin binding gene ontology annotation. There are just a few studies about LINC00341 in different cell types, including endothelial cells ²⁶⁴, vascular smooth muscle cells ²⁶⁵, breast cancer metastasis ²⁶⁶, bronchial epithelial cells ²⁶⁷ and chondrocytes ²⁶⁸. The functions of this lncRNAs are still unclear. Nevertheless, in line with a possible role in bone biology, an interesting study found that LINC00341 was coexpressed with the transcription factor MEF2C (involved in SOST gene expression) ²⁶⁹.

In our samples, LINC00341 was similarly expressed (very low expression) in undifferentiated BMSCs from fracture patients and controls, but it markedly increased in differentiated cells. Overexpressing this gene in human cell lines induced a slight increase in the expression of the osteogenic markers ALPL and COL1A1. However, to confirm its anabolic role in bone, further studies are needed that can demonstrate how knocking down LINC00341 in BMSCs impairs osteogenic differentiation.

PTGS2-AS: is also known as PTGS2 Antisense NFkB1 Complex-Mediated Expression Regulator RNA, or P50-Associated COX-2 Extragenic RNA (PACERR). It consists in a head-to-head antisense (divergent lncRNA), which interacts with NFkB1 transcriptional p50 subunit to promote the expression of COX-2 in U937 pro-monocytic, human myeloid leukaemia cell line and in U937-derived macrophages ²⁷⁰. Furthermore, COX-2 may function as an oncogene in osteosarcoma. Hence, PACERR promotes the proliferation and metastasis of osteosarcoma cells through COX-2 activation ²⁷¹. Numerous studies suggest that COX-2 may have the potential to accelerate the osteogenic differentiation of MSCs ^{272–274}.

However, in our samples, PACERR was downregulated after osteogenic differentiation and there were not significant differences between fractures and

controls. After overexpressing this lncRNA in human cell lines (HEK-293T and HOS), we did not see any significant change in the expression of COX-2, ALPL or COL1A1. Nevertheless, the value of these experiments is limited by the very low level of COX-2 expression in the cell lines tested.

Since we were unable to consistently transfect either expression vectors or siRNAs in our primary BMSCs, we could not get definitive evidence for the functional roles of these lncRNAs.

Several studies in the literature have explored lncRNAs expression in relation with bone biology and pathophysiology. For example, Wang and collaborators studied BMSCs in their basal state and at different days of osteogenic differentiation (days 7 and 14) and found that lncRNA KCNQ1OT1 is up-regulated during the process of osteogenesis, whereas miRNA-214 is down-regulated. In vitro data suggested that KCNQ1OT1 positively regulates osteogenesis by regulating BMP2 expression through sponging miRNA-214 ²⁷⁵. Another study suggested that lncRNA MALAT1 regulates OSX expression through sponging miRNA-143 during BMSCs osteogenic differentiation ¹⁵⁸. LncRNA MEG3 has also been related to osteogenesis, but with controversial results ^{159,161,276}. An interesting study by Li et al ²⁷⁷ suggested that lncRNAs Bmncr influences MSCs, enhancing commitment towards the osteoblastic lineage and reversing the age-associated switch between osteoblast and adipocyte differentiation.

We have not found any significant changes in the expression of these lncRNAs, nor in other candidate lncRNAs suggested in the literature. Hence, a lack of replication in the field is evident. The explanations may be multiple, including differences in clinical characteristics of patients, protocols for BMSC isolation and culture, other experimental differences, etc. It is to note that single-cell analyses show that MSCs are a heterogeneous population showing substantial differences in gene transcription ²⁷⁸. The small sample size of most studies is another important limitation.

Study limitations

Our study has some limitations. Due to practical reasons, we isolated BMSCs from patients with fractures and with hip osteoarthritis. So, we used osteoarthritis rather than “normal” BMSCs controls. This is certainly a limitation, as in theory BMSCs from patients with osteoarthritis may have some disease-related features. However, results of studies about gene expression by BMSCs in osteoarthritis have been controversial. Normal osteogenesis has been found in some studies ^{279,280}, whereas abnormal proliferation and gene expression and chondrogenic differentiation have also been reported ^{281–284}. Nevertheless, osteoarthritis is mainly a cartilage disease and osteoporosis is a bone disease, and several studies have pointed out that osteoarthritis is associated with increased bone mineral density (BMD), which is the opposite of osteoporosis ^{285,286}. It has been suggested that subchondral bone in OA is changing inversely to cartilage loss, so to avoid this singularity, we removed the subchondral regions before BMSCs isolation. Thus, we thought that osteoarthritis could be regarded as feasible, convenient controls for comparison with osteoporotic fractures.

In addition, as a source of “osteoporotic BMSCs”, we analysed BMSCs from patients with hip fractures, also for ethical and feasibility reasons. We tried to extract cylinders in the central part of the femoral head, thus avoiding fractured areas. However, as discussed above, we cannot exclude that local and systemic responses after fracture could exert some influences on BMSCs. In other words, any abnormal behaviour of BMSCs in our study could represent changes either prior to fracture (ie, osteoporosis-related) or after fracture (ie, related to fracture-induced responses). Nevertheless, cells were grown in culture for several weeks before the experiments, which may have tempered any fracture-related influences.

On the other hand, culture itself induces a number of changes in cell characteristics, including senescence and epigenetic changes ^{191,230,287,288}. To diminish this bias, we used cells at two first passages. The small proportion of BMSCs among bone cells prevented from using freshly isolated, uncultured BMSCs, which could be otherwise ideal.

Aging is associated with changes in epigenetics marks like DNA methylation, which has been shown to gradually diminish with age ²³⁷. In our case, patients with osteoporotic fractures were about 10 years older than donors with

osteoarthritis. We adjusted the DNA methylation values with age as a covariate to limit age-related bias.

Another considerable limitation is that BMSCs proliferate very slowly and are very hard to transfect with lipid-based and other conventional procedures. Due to practical reasons we could not use lentiviral-based methods, which may be more efficient. This precluded us from obtaining meaningful results in loss-of-function and gain-of-function experiments.

In addition, BMSCs cultures are heterogeneous. This means that they are a mixed population with somewhat different phenotypes and/or functions ²⁸⁹. This variability may limit power to find significant differences in the gene expression and methylation patterns, particularly when the sample size is not very large. This may be important for the study of lncRNAs, which are frequently expressed at low levels.

Additionally, we found singular difficulties in confirming antisense lncRNAs expression by using commonly postulated protocols for strand-specific RT-PCR. These drawbacks made us rather cautious when interpreting other published results of antisense-type RNAs.

What this thesis adds

This thesis contributes to the current understanding of how epigenetic mechanisms, including DNA methylation and lncRNAs, regulate BMSCs in osteoporosis. It is the first DNA methylation and gene expression study, using BMSCs obtained from patients with osteoporosis. It included transcriptome analysis of both mRNAs and lncRNAs. The analysis of these gene expression signatures, epigenetic marks and regulatory networks in osteoporosis suggested new protein coding genes involved in the disease. The study points to some new signalling pathways that should be studied for a better understanding of the disease and to find new potential biomarkers and drug targets.

Future perspectives

Previous data and experimental data from this thesis suggest that epigenetic mechanisms have an important influence on the osteogenic differentiation. However, the exact role of these mechanisms in osteoporosis is not entirely clear.

In this sense, the comparison of global DNA methylation and gene expression patterns in subjects with bone diseases can lead to the identification of differentially methylated regions and differentially expressed genes involved in the pathogenesis of the disease. New sequencing technologies, including bisulfite-sequencing may help to further elucidate the epigenome of bones in skeletal diseases.

DNA methylation gives a global view on epigenome-wide changes and highlight some target regions and gene candidates, which can be useful for diagnosis and prevention. RNA transcripts are very state specific, which makes them very prone to study as biomarkers. However, sample heterogeneity may complicate the study with less precise results. In line with this, single cell RNA sequencing can distinguish cell populations, uncover regulatory relationships between genes and compare distinct cell lineages ²⁹⁰. An interesting route for future research is to conduct more longitudinal studies with paired differentiated samples, at different times along BMSCs osteogenic differentiation. This could reveal certain protein or non-protein coding genes involved in this process and in osteoporosis. Also, analysis of BMSCs from osteoporotic patients without recent fracture would help to distinguish pathogenetic from adaptative changes. Likewise, in vivo studies will clarify the true potential of BMSCs from osteoporotic patients as an autologous source of osteogenic cells and as elements for bone regeneration therapy.

8. CONCLUSIONS

CONCLUSIONS

- 1- BMSCs of patients with osteoporotic fractures appear to maintain good proliferation capacity, but diminished capacity for terminal osteoblastic differentiation in vitro, in comparison with BMSCs from control patients with osteoarthritis.
- 2- Combined transcriptome and epigenome analyses, revealed signature differences between both patient groups, including 9038 differentially methylated CpG sites. DNA methylation differences were enriched in enhancer regions, thus suggesting that epigenetic mechanisms may particularly target enhancer regions to modulate gene transcription.
- 3- Enhancer regions showing differential methylation were associated with differentially expressed genes enriched in pathways related to BMSCs growth, osteoblast differentiation and bone formation.
- 4- Among differentially methylated and expressed genes we identified LOXL2, ID2, OPG and UNC5B.
- 5- The epigenetic age of BMSCs, from patients with osteoporotic hip fractures, was accelerated in comparison with cells from patients with osteoarthritis. The mechanisms involved and the actual consequences for bone formation in vivo remain to be elucidated.
- 6- Almost 50% of differentially expressed genes in BMSCs derived from patients with osteoporotic hip fractures, belonged to the lncRNA class of genes. Among them, 72% were of the antisense type.
- 7- Approximately half of the protein-coding genes in cis position of differentially expressed antisense lncRNAs, were also differentially expressed. These lncRNAs-protein coding genes pairs showed significant overrepresentation in bone related pathways.
- 8- Experimental difficulties for transfecting BMSCs and for reliable strand-specific RT-PCR impeded the functional validation of candidate transcripts.

- 9- Overall, our results suggest that epigenetic mechanisms, and specifically the methylation status of enhancer regions and lncRNAs play significant roles in determining the gene expression pattern of BMSCs and may represent additional targets to enhance bone anabolism in patients with osteoporosis, as well as to increase the efficacy of cell regenerative therapy.

9. BIBLIOGRAPHY

REFERENCES

1. Paiva, K. B. S. & Granjeiro, J. M. Bone tissue remodeling and development: Focus on matrix metalloproteinase functions. *Arch. Biochem. Biophys.* **561**, 74–87 (2014).
2. Franz-Odenaal, T. A., Hall, B. K. & Witten, P. E. Buried alive: how osteoblasts become osteocytes. *Dev. Dyn.* **235**, 176–90 (2006).
3. Dallas, S. L. & Bonewald, L. F. Dynamics of the transition from osteoblast to osteocyte. *Ann. N. Y. Acad. Sci.* **1192**, 437–43 (2010).
4. Long, F. Building strong bones: molecular regulation of the osteoblast lineage. *Nat. Rev. Mol. Cell Biol.* **13**, 27–38 (2011).
5. Komori, T. *et al.* Targeted Disruption of *Cbfa1* Results in a Complete Lack of Bone Formation owing to Maturation Arrest of Osteoblasts. *Cell* **89**, 755–764 (1997).
6. Otto, F. *et al.* *Cbfa1*, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**, 765–71 (1997).
7. Bruderer, M., Richards, R., Alini, M. & Stoddart, M. Role and regulation of RUNX2 in osteogenesis. *Eur. Cells Mater.* **28**, 269–286 (2014).
8. Rutkovskiy, A., Stensl kken, K.-O. & Vaage, I. J. Osteoblast Differentiation at a Glance. *Med. Sci. Monit. Basic Res.* **22**, 95–106 (2016).
9. Nishio, Y. *et al.* Runx2-mediated regulation of the zinc finger Osterix/Sp7 gene. *Gene* **372**, 62–70 (2006).
10. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108**, 17–29 (2002).
11. Long, F. *et al.* *Ihh* signaling is directly required for the osteoblast lineage in the endochondral skeleton. *Development* **131**, 1309–18 (2004).
12. Shahi, M., Peymani, A. & Sahmani, M. Regulation of Bone Metabolism. *Reports Biochem. Mol. Biol.* **5**, 73–82 (2017).
13. Balemans, W. *et al.* Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J. Med. Genet.* **39**, 91–7 (2002).
14. Brunkow, M. E. *et al.* Bone dysplasia sclerosteosis results from loss of the

- SOST gene product, a novel cystine knot-containing protein. *Am. J. Hum. Genet.* **68**, 577–89 (2001).
15. Capulli, M., Paone, R. & Rucci, N. Osteoblast and osteocyte: Games without frontiers. *Arch. Biochem. Biophys.* **561**, 3–12 (2014).
16. Wein, M. N. Bone Lining Cells: Normal Physiology and Role in Response to Anabolic Osteoporosis Treatments. *Curr. Mol. Biol. Reports* **3**, 79–84 (2017).
17. Raggatt, L. J. & Partridge, N. C. Cellular and molecular mechanisms of bone remodeling. *J. Biol. Chem.* **285**, 25103–8 (2010).
18. Collins, F. L., Rios-Arce, N. D., Schepper, J. D., Parameswaran, N. & McCabe, L. R. The Potential of Probiotics as a Therapy for Osteoporosis. *Microbiol. Spectr.* **5**, (2017).
19. Ono, T. & Nakashima, T. Recent advances in osteoclast biology. *Histochem. Cell Biol.* **149**, 325–341 (2018).
20. Boyce, B. F. Advances in the Regulation of Osteoclasts and Osteoclast Functions. *J. Dent. Res.* **92**, 860–867 (2013).
21. Asagiri, M. & Takayanagi, H. The molecular understanding of osteoclast differentiation. *Bone* **40**, 251–64 (2007).
22. Liu, W. & Zhang, X. Receptor activator of nuclear factor- κ B ligand (RANKL)/RANK/osteoprotegerin system in bone and other tissues. *Mol. Med. Rep.* **11**, 3212–3218 (2015).
23. Asagiri, M. & Takayanagi, H. The molecular understanding of osteoclast differentiation. *Bone* **40**, 251–264 (2007).
24. Brunetti, G., Di Benedetto, A. & Mori, G. Bone remodeling. *Imaging Prosthet. Joints A Comb. Radiol. Clin. Perspect.* **1092**, 27–37 (2014).
25. Sims, N. A. & Martin, T. J. Coupling the activities of bone formation and resorption: a multitude of signals within the basic multicellular unit. *Bonekey Rep.* **3**, 481 (2014).
26. Bellido, T. Osteocyte-Driven Bone Remodeling. *Calcif. Tissue Int.* **94**, 25–34 (2014).
27. Kalra, K. & Tomar, P. C. Stem Cell: Basics, Classification and Applications. *Am. J. Phytomedicine Clin. Ther.* **2**, 919–930 (2014).
28. Friedenstein, A. J., Piatetzky-Shapiro, I. I. & Petrakova, K. V. Osteogenesis in transplants of bone marrow cells. *J. Embryol. Exp. Morphol.* **16**, 381–90

- (1966).
29. Caplan, A. I. Mesenchymal stem cells. *J. Orthop. Res.* **9**, 641–650 (1991).
 30. Uccelli, A., Moretta, L. & Pistoia, V. Immunoregulatory function of mesenchymal stem cells. *Eur. J. Immunol.* **36**, 2566–73 (2006).
 31. Gao, F. *et al.* Mesenchymal stem cells and immunomodulation: current status and future prospects. *Cell Death Dis.* **7**, e2062 (2016).
 32. De Becker, A. & Riet, I. Van. Homing and migration of mesenchymal stromal cells: How to improve the efficacy of cell therapy? *World J. Stem Cells* **8**, 73–87 (2016).
 33. Vizoso, F. J., Eiro, N., Cid, S., Schneider, J. & Perez-Fernandez, R. Mesenchymal Stem Cell Secretome: Toward Cell-Free Therapeutic Strategies in Regenerative Medicine. *Int. J. Mol. Sci.* **18**, 1852 (2017).
 34. Griffin, M. D. *et al.* Anti-donor immune responses elicited by allogeneic mesenchymal stem cells: what have we learned so far? *Immunol. Cell Biol.* **91**, 40–51 (2013).
 35. Pereira, A. R., Trivanović, D. & Herrmann, M. Approaches to mimic the complexity of the skeletal mesenchymal stem/stromal cell niche in vitro. *Eur. Cell. Mater.* **37**, 88–112 (2019).
 36. Zuk, P. A. *et al.* Human adipose tissue is a source of multipotent stem cells. *Mol. Biol. Cell* **13**, 4279–95 (2002).
 37. Sipp, D., Robey, P. G. & Turner, L. Clear up this stem-cell mess. *Nature* **561**, 455–457 (2018).
 38. Caplan, A. I. Mesenchymal Stem Cells: Time to Change the Name! *Stem Cells Transl. Med.* **6**, 1445–1451 (2017).
 39. Sacchetti, B. *et al.* No Identical ‘Mesenchymal Stem Cells’ at Different Times and Sites: Human Committed Progenitors of Distinct Origin and Differentiation Potential Are Incorporated as Adventitial Cells in Microvessels. *Stem cell reports* **6**, 897–913 (2016).
 40. Bianco, P. & Robey, P. G. Skeletal stem cells. *Development* **142**, 1023–1027 (2015).
 41. Chan, C. K. F. *et al.* Identification of the Human Skeletal Stem Cell. *Cell* **175**, 43-56.e21 (2018).
 42. Almalki, S. G. & Agrawal, D. K. Key transcription factors in the differentiation of mesenchymal stem cells. *Differentiation*. **92**, 41–51

- (2016).
43. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–217 (2013).
 44. Infante, A. & Rodríguez, C. I. Osteogenesis and aging: lessons from mesenchymal stem cells. *Stem Cell Res. Ther.* **9**, 244 (2018).
 45. Bouxsein, M. L. *The Nature of Osteoporosis. Osteoporosis* (Academic Press, 2013). doi:10.1016/B978-0-12-415853-5.00002-9
 46. Mirza, F. & Canalis, E. Management of endocrine disease: Secondary osteoporosis: pathophysiology and management. *Eur. J. Endocrinol.* **173**, R131–R151 (2015).
 47. Meng, X.-H. *et al.* Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions. *Bone* **113**, 41–48 (2018).
 48. Sabik, O. L. & Farber, C. R. Using GWAS to identify novel therapeutic targets for osteoporosis. *Transl. Res.* **181**, 15–26 (2017).
 49. Del Real, A., Riancho-Zarrabeitia, L., López-Delgado, L. & Riancho, J. A. Epigenetics of Skeletal Diseases. *Curr. Osteoporos. Rep.* **16**, 246–255 (2018).
 50. Vrtačnik, P., Marc, J. & Ostanek, B. Epigenetic mechanisms in bone. *Clin. Chem. Lab. Med.* **52**, 589–608 (2014).
 51. Glyn-Jones, S. *et al.* Osteoarthritis. *Lancet (London, England)* **386**, 376–87 (2015).
 52. Palazzo, C., Nguyen, C., Lefevre-Colau, M.-M., Rannou, F. & Poiraudreau, S. Risk factors and burden of osteoarthritis. *Ann. Phys. Rehabil. Med.* **59**, 134–138 (2016).
 53. Jeffries, M. A. Osteoarthritis year in review 2018: genetics and epigenetics. *Osteoarthr. Cartil.* **27**, 371–377 (2019).
 54. Aho, O.-M., Finnilä, M., Thevenot, J., Saarakkala, S. & Lehenkari, P. Subchondral bone histology and grading in osteoarthritis. *PLoS One* **12**, e0173726 (2017).
 55. Vina, E. R. & Kwok, C. K. Epidemiology of Osteoarthritis: Literature Update. *Curr. Opin. Rheumatol.* **30**, 160 (2018).
 56. Gregson, C. L. *et al.* High Bone Mass is associated with bone-forming features of osteoarthritis in non-weight bearing joints independent of body

- mass index. *Bone* **97**, 306–313 (2017).
57. Deichmann, U. Epigenetics: The origins and evolution of a fashionable topic. *Dev. Biol.* **416**, 249–254 (2016).
 58. Qin, H., Zhao, A. & Fu, X. Small molecules for reprogramming and transdifferentiation. *Cell. Mol. Life Sci.* **74**, 3553–3575 (2017).
 59. Zoghbi, H. Y. & Beaudet, A. L. Epigenetics and Human Disease. *Cold Spring Harb. Perspect. Biol.* **8**, a019497 (2016).
 60. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10604–9 (2005).
 61. Adenot, P. G., Mercier, Y., Renard, J. P. & Thompson, E. M. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development* **124**, 4615–25 (1997).
 62. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–44 (2012).
 63. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Dev. Biol.* **241**, 172–182 (2002).
 64. Cantone, I. & Fisher, A. G. Epigenetic programming and reprogramming during development. *Nat. Struct. Mol. Biol.* **20**, 282–289 (2013).
 65. Heijmans, B. T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17046–9 (2008).
 66. Lee, H.-S. Impact of Maternal Diet on the Epigenome during In Utero Life and the Developmental Programming of Diseases in Childhood and Adulthood. *Nutrients* **7**, 9492–507 (2015).
 67. Godfrey, K. M. *et al.* Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes* **60**, 1528–34 (2011).
 68. Bocheva, G. & Boyadjieva, N. Epigenetic regulation of fetal bone development and placental transfer of nutrients: progress for osteoporosis. *Interdiscip. Toxicol.* **4**, 167–72 (2011).
 69. Crider, K. S., Yang, T. P., Berry, R. J. & Bailey, L. B. Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv. Nutr.* **3**, 21–38 (2012).

70. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **9**, 436–447 (2016).
71. Suderman, M. *et al.* Childhood abuse is associated with methylation of multiple loci in adult DNA. *BMC Med. Genomics* **7**, 13 (2014).
72. Labonté, B. *et al.* Genome-Wide Methylation Changes in the Brains of Suicide Completers. *Am. J. Psychiatry* **170**, 511–520 (2013).
73. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
74. Chrun, E. S., Modolo, F. & Daniel, F. I. Histone modifications: A review about the presence of this epigenetic phenomenon in carcinogenesis. *Pathol. Res. Pract.* **213**, 1329–1339 (2017).
75. Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525–34 (2013).
76. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20**, 1147–1155 (2013).
77. Dubé, J. C., Wang, X. Q. D. & Dostie, J. Spatial Organization of Epigenomes. *Curr. Mol. Biol. reports* **2**, 1–9 (2016).
78. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–80 (2012).
79. van Meurs, J. B., Boer, C. G., Lopez-Delgado, L. & Riancho, J. A. Role of Epigenomics in Bone and Cartilage Disease. *J. Bone Miner. Res.* **34**, 215–230 (2019).
80. Spielmann, M. *et al.* Homeotic Arm-to-Leg Transformation Associated with Genomic Rearrangements at the PITX1 Locus. *Am. J. Hum. Genet.* **91**, 629 (2012).
81. Wein, M. N. *et al.* HDAC5 Controls MEF2C-Driven Sclerostin Expression in Osteocytes. *J. Bone Miner. Res.* **30**, 400–411 (2015).
82. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
83. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
84. Jin, B., Li, Y. & Robertson, K. D. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* **2**, 607–17 (2011).
85. Unnikrishnan, A. *et al.* The role of DNA methylation in epigenetics of aging.

- Pharmacol. Ther.* **195**, 172–185 (2018).
86. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
 87. Delgado-Calle, J. *et al.* Genome-wide profiling of bone reveals differentially methylated regions in osteoporosis and osteoarthritis. *Arthritis Rheum.* **65**, 197–205 (2013).
 88. Reppe, S. *et al.* Methylation of bone SOST, its mRNA, and serum sclerostin levels correlate strongly with fracture risk in postmenopausal women. *J. Bone Miner. Res.* **30**, 249–56 (2015).
 89. Cheishvili, D. *et al.* Identification of an Epigenetic Signature of Osteoporosis in Blood DNA of Postmenopausal Women. *J. Bone Miner. Res.* **33**, 1980–1989 (2018).
 90. Roforth, M. M. *et al.* Global transcriptional profiling using RNA sequencing and DNA methylation patterns in highly enriched mesenchymal cells from young versus elderly women. *Bone* **76**, 49–57 (2015).
 91. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
 92. Consortium, T. E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
 93. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437 (2014).
 94. Kufel, J. & Grzechnik, P. Small Nucleolar RNAs Tell a Different Tale. *Trends Genet.* **35**, 104–117 (2019).
 95. Stepanov, G. A. *et al.* Regulatory role of small nucleolar RNAs in human diseases. *Biomed Res. Int.* **2015**, 206849 (2015).
 96. Zhang, L. *et al.* Serum non-coding RNAs as biomarkers for osteoarthritis progression after ACL injury. *Osteoarthr. Cartil.* **20**, 1631–1637 (2012).
 97. Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
 98. Reznik, B. *et al.* Heterogeneity of transposon expression and activation of the repressive network in human fetal germ cells. *Development* **146**, dev171157 (2019).

99. Rajasethupathy, P. *et al.* A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**, 693–707 (2012).
100. Bhattacharjee, S., Roche, B. & Martienssen, R. A. RNA-induced initiation of transcriptional silencing (RITS) complex structure and function. *RNA Biol.* **18**, 1–14 (2019).
101. Maida, Y., Kyo, S., Lassmann, T., Hayashizaki, Y. & Masutomi, K. Off-target effect of endogenous siRNA derived from RMRP in human cells. *Int. J. Mol. Sci.* **14**, 9305–18 (2013).
102. Iswariya, G. T., Paital, B., Padma, P. R. & Nirmaladevi, R. microRNAs: Epigenetic players in cancer and aging. *Front. Biosci.* **11**, 29–55 (2019).
103. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).
104. Cheng, V. K.-F., Au, P. C.-M., Tan, K. C. & Cheung, C.-L. MicroRNA and Human Bone Health. *JBMR Plus* **3**, 2–13 (2019).
105. Egea, V. *et al.* Tissue inhibitor of metalloproteinase-1 (TIMP-1) regulates mesenchymal stem cells through let-7f microRNA and Wnt/ -catenin signaling. *Proc. Natl. Acad. Sci.* **109**, E309–E316 (2012).
106. Bellavia, D. *et al.* Deregulated miRNAs in bone health: Epigenetic roles in osteoporosis. *Bone* **122**, 52–75 (2019).
107. Marini, F., Cianferotti, L. & Brandi, M. L. Epigenetic Mechanisms in Bone Biology and Osteoporosis: Can They Drive Therapeutic Choices? *Int. J. Mol. Sci.* **17**, 1329 (2016).
108. Garmilla-Ezquerro, P. *et al.* Analysis of the Bone MicroRNome in Osteoporotic Fractures. *Calcif. Tissue Int.* **96**, 30–37 (2015).
109. Inoue, K. *et al.* Bone protection by inhibition of microRNA-182. *Nat. Commun.* **9**, 4108 (2018).
110. Panach, L., Mifsut, D., Tarín, J. J., Cano, A. & García-Pérez, M. Á. Serum Circulating MicroRNAs as Biomarkers of Osteoporotic Fracture. *Calcif. Tissue Int.* **97**, 495–505 (2015).
111. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* **10**, 28–36 (1990).
112. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–42 (1992).

113. Li, J. & Liu, C. Coding or Noncoding, the Converging Concepts of RNAs. *Front. Genet.* **10**, 496 (2019).
114. Yao, R.-W., Wang, Y. & Chen, L.-L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **21**, 542–551 (2019).
115. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
116. Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110–22 (2015).
117. Marques, A. C. & Ponting, C. P. Intergenic lncRNAs and the evolution of gene expression. *Curr. Opin. Genet. Dev.* **27**, 48–53 (2014).
118. Jarroux, J., Morillon, A. & Pinskaya, M. History, Discovery, and Classification of lncRNAs. in *Advances in experimental medicine and biology* **1008**, 1–46 (2017).
119. Chen, S. *et al.* lincRNA-p21: function and mechanism in cancer. *Med. Oncol.* **34**, 98 (2017).
120. Balas, M. M. & Johnson, A. M. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-coding RNA Res.* **3**, 108–117 (2018).
121. Wu, T. & Du, Y. LncRNAs: From Basic Research to Medical Application. *Int. J. Biol. Sci.* **13**, 295–307 (2017).
122. Kim, T.-K., Hemberg, M. & Gray, J. M. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harb. Perspect. Biol.* **7**, a018622 (2015).
123. Liu, F. Enhancer-derived RNA: A Primer. *Genomics. Proteomics Bioinformatics* **15**, 196–200 (2017).
124. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–20 (2013).
125. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–98 (2012).
126. Ayupe, A. C. *et al.* Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol.* **12**, 877 (2015).
127. Gloss, B. S. & Dinger, M. E. The specificity of long noncoding RNA expression. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 16–22

- (2016).
128. Ward, M., McEwan, C., Mills, J. D. & Janitz, M. Conservation and tissue-specific transcription patterns of long noncoding RNAs. *J. Hum. Transcr.* **1**, 2–9 (2015).
 129. Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C.-Y. & Lee, J. T. Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol. Cell* **74**, 101–117 (2019).
 130. Akhade, V. S., Pal, D. & Kanduri, C. Long Noncoding RNA: Genome Organization and Mechanism of Action. in *Advances in experimental medicine and biology* **1008**, 47–74 (2017).
 131. Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712 (2013).
 132. Wahlestedt, C. Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug Discov. Today* **11**, 503–508 (2006).
 133. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–9 (2011).
 134. Zapulla, D. C. & Cech, T. R. RNA as a Flexible Scaffold for Proteins: Yeast Telomerase and Beyond. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 217–224 (2006).
 135. Spitale, R. C., Tsai, M.-C. & Chang, H. Y. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* **6**, 539–43 (2011).
 136. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–66 (2012).
 137. Chan, J. J. & Tay, Y. Noncoding RNA:RNA Regulatory Networks in Cancer. *Int. J. Mol. Sci.* **19**, (2018).
 138. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
 139. Natoli, G. & Andrau, J.-C. Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
 140. Abdelmohsen, K. *et al.* Senescence-associated lncRNAs: senescence-associated long noncoding RNAs. *Aging Cell* **12**, 890–900 (2013).

141. Grammatikakis, I., Panda, A. C., Abdelmohsen, K. & Gorospe, M. Long noncoding RNAs(lncRNAs) and the molecular hallmarks of aging. *Aging (Albany. NY)*. **6**, 992–1009 (2014).
142. Collins, K. Physiological Assembly and Activity of Human Telomerase Complexes. *Mech. Ageing Dev.* **129**, 91 (2008).
143. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38**, 5797 (2010).
144. Marín-Béjar, O. *et al.* Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* **14**, R104 (2013).
145. Zhang, A. *et al.* The human long non-coding RNA-RoR is a p53 repressor in response to DNA damage. *Cell Res.* **23**, 340 (2013).
146. Wan, G. *et al.* Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway. *Cell. Signal.* **25**, 1086–95 (2013).
147. Franceschi, C. & Campisi, J. Chronic Inflammation (Inflammaging) and Its Potential Contribution to Age-Associated Diseases. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **69**, S4–S9 (2014).
148. Costa, M. C., Leitão, A. L. & Enguita, F. J. Noncoding Transcriptional Landscape in Human Aging. in *Long Non-coding RNAs in Human Disease* 177–202 (Springer, Cham, 2015).
149. Krawczyk, M. & Emerson, B. M. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-κB complexes. *Elife* **3**, e01776 (2014).
150. Sousa-Franco, A., Rebelo, K., da Rocha, S. T. & Bernardes de Jesus, B. LncRNAs regulating stemness in aging. *Aging Cell* **18**, e12870 (2019).
151. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295 (2011).
152. Bernardes de Jesus, B. *et al.* Silencing of the lncRNA Zeb2-NAT facilitates reprogramming of aged fibroblasts and safeguards stem cell pluripotency. *Nat. Commun.* **9**, 94 (2018).
153. Silva, A. M. *et al.* Long noncoding RNAs: a missing link in osteoporosis. *Bone Res.* **7**, 10 (2019).
154. Zhu, L. & Xu, P.-C. Downregulated LncRNA-ANCR promotes osteoblast

- p differentiation by targeting EZH2 and regulating Runx2 expression.
- Biochem. Biophys. Res. Commun.*
- 432**
- , 612–617 (2013).
155. Zhang, J., Tao, Z. & Wang, Y. Long non-coding RNA DANCR regulates the proliferation and osteogenic differentiation of human bone-derived marrow mesenchymal stem cells via the p38 MAPK pathway. *Int. J. Mol. Med.* **41**, 213–219 (2018).
 156. Jia, Q. *et al.* The Regulatory Effects of Long Noncoding RNA-ANCR on Dental Tissue-Derived Stem Cells. *Stem Cells Int.* **2016**, 3146805 (2016).
 157. Zhu, X.-X. *et al.* Long non-coding RNA HoxA-AS3 interacts with EZH2 to regulate lineage commitment of mesenchymal stem cells. *Oncotarget* **7**, 63561–63570 (2016).
 158. Gao, Y. *et al.* Long noncoding RNA MALAT1 promotes osterix expression to regulate osteogenic differentiation by targeting miRNA-143 in human bone marrow-derived mesenchymal stem cells. *J. Cell. Biochem.* **119**, 6986–6996 (2018).
 159. Zhuang, W. *et al.* Upregulation of lncRNA MEG3 Promotes Osteogenic Differentiation of Mesenchymal Stem Cells From Multiple Myeloma Patients By Targeting BMP4 Transcription. *Stem Cells* **33**, 1985–1997 (2015).
 160. Li, Z. *et al.* Long non-coding RNA MEG3 inhibits adipogenesis and promotes osteogenesis of human adipose-derived mesenchymal stem cells via miR-140-5p. *Mol. Cell. Biochem.* **433**, 51–60 (2017).
 161. Wang, Q. *et al.* LncRNA MEG3 inhibited osteogenic differentiation of bone marrow mesenchymal stem cells from postmenopausal osteoporosis by targeting miR-133a-3p. *Biomed. Pharmacother.* **89**, 1178–1186 (2017).
 162. Tang, S. *et al.* LncRNA-OG Promotes the Osteogenic Differentiation of Bone Marrow-Derived Mesenchymal Stem Cells Under the Regulation of hnRNPK. *Stem Cells* **37**, 270–283 (2019).
 163. Pearson, M. J. & Jones, S. W. Review: Long Noncoding RNAs in the Regulation of Inflammatory Pathways in Rheumatoid Arthritis and Osteoarthritis. *Arthritis Rheumatol. (Hoboken, N.j.)* **68**, 2575 (2016).
 164. Fei, Q. *et al.* Identification of aberrantly expressed long non-coding RNAs in postmenopausal osteoporosis. *Int. J. Mol. Med.* **41**, 3537–3550 (2018).
 165. del Real, A. *et al.* Differential analysis of genome-wide methylation and

- gene expression in mesenchymal stem cells of patients with fractures and osteoarthritis. *Epigenetics* **12**, 113–122 (2017).
166. Delgado-Calle, J. *et al.* Role of DNA methylation in the regulation of the RANKL-OPG system in human bone. *Epigenetics* **7**, 83–91 (2012).
167. Dedeurwaerder, S. *et al.* A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief. Bioinform.* **15**, 929–41 (2014).
168. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (2011).
169. Assenov, Y. *et al.* Comprehensive Analysis of DNA Methylation Data with RnBeads. *Nat. Methods* **11**, 1138 (2014).
170. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430 (2014).
171. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
172. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
173. Morris, T. J. & Beck, S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* **72**, 3–8 (2015).
174. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495 (2010).
175. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
176. Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **30**, e47 (2002).
177. Tost, J. & Gut, I. G. DNA methylation analysis by pyrosequencing. *Nat. Protoc.* **2**, 2265–2275 (2007).
178. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
179. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
180. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

181. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).
182. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
183. Xu, J., Kelly, R., Fang, H. & Tong, W. ArrayTrack: a free FDA bioinformatics tool to support emerging biomedical research--an update. *Hum. Genomics* **4**, 428–34 (2010).
184. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
185. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* **13**, R44 (2012).
186. Korpelainen, E. *et al.* *RNA-seq Data Analysis*. (Chapman and Hall/CRC, 2014). doi:10.1201/b17457
187. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–5 (2010).
188. Evans, C., Hardin, J. & Stoebel, D. M. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.* **19**, 776–792 (2018).
189. Seyednasrollah, F., Laiho, A. & Elo, L. L. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* **16**, 59–70 (2015).
190. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–9 (2000).
191. Bork, S. *et al.* DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell* **9**, 54–63 (2010).
192. Dominici, M. *et al.* Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**, 315–317 (2006).

193. Triche, T. J. *et al.* Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).
194. Emig, D. *et al.* AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res.* **38**, W755-62 (2010).
195. Siddappa, R. *et al.* cAMP/PKA pathway activation in human mesenchymal stem cells in vitro results in robust bone formation in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7281–6 (2008).
196. Sabik, O. L. & Farber, C. R. Using GWAS to identify novel therapeutic targets for osteoporosis. *Transl. Res.* **181**, 15–26 (2017).
197. Paspaliaris, V. & Kolios, G. Stem cells in Osteoporosis: From Biology to New Therapeutic Approaches. *Stem Cells Int.* **2019**, 1730978 (2019).
198. Antebi, B., Pelled, G. & Gazit, D. Stem cell therapy for osteoporosis. *Curr. Osteoporos. Rep.* **12**, 41–7 (2014).
199. Katsara, O. *et al.* Effects of Donor Age, Gender, and In Vitro Cellular Aging on the Phenotypic, Functional, and Molecular Characteristics of Mouse Bone Marrow-Derived Mesenchymal Stem Cells. *Stem Cells Dev.* **20**, 1549–1561 (2011).
200. Zhou, S. *et al.* Age-related intrinsic changes in human bone-marrow-derived mesenchymal stem cells and their differentiation to osteoblasts. *Aging Cell* **7**, 335–43 (2008).
201. Mueller, S. M. & Glowacki, J. Age-related decline in the osteogenic potential of human bone marrow cells cultured in three-dimensional collagen sponges. *J. Cell. Biochem.* **82**, 583–90 (2001).
202. Kim, M. *et al.* Age-related alterations in mesenchymal stem cells related to shift in differentiation from osteogenic to adipogenic potential: Implication to age-associated bone diseases and defects. *Mech. Ageing Dev.* **133**, 215–225 (2012).
203. Moerman, E. J., Teng, K., Lipschitz, D. A. & Lecka-Czernik, B. Aging activates adipogenic and suppresses osteogenic programs in mesenchymal marrow stroma/stem cells: the role of PPAR- γ 2 transcription factor and TGF- β /BMP signaling pathways. *Aging Cell* **3**, 379–389 (2004).
204. Ganguly, P. *et al.* Age-related Changes in Bone Marrow Mesenchymal Stromal Cells. *Cell Transplant.* **26**, 1520–1529 (2017).
205. Stenderup, K., Justesen, J., Eriksen, E. F., Rattan, S. I. & Kassem, M.

- Number and proliferative capacity of osteogenic stem cells are maintained during aging and in patients with osteoporosis. *J. Bone Miner. Res.* **16**, 1120–9 (2001).
206. Stenderup, K., Justesen, J., Clausen, C. & Kassem, M. Aging is associated with decreased maximal life span and accelerated senescence of bone marrow stromal cells. *Bone* **33**, 919–26 (2003).
 207. Stenderup, K. *et al.* Aged Human Bone Marrow Stromal Cells Maintaining Bone Forming Capacity in vivo Evaluated Using an Improved Method of Visualization. *Biogerontology* **5**, 107–118 (2004).
 208. Zheng, C.-X. *et al.* Adipose mesenchymal stem cells from osteoporotic donors preserve functionality and modulate systemic inflammatory microenvironment in osteoporotic cytotherapy. *Sci. Rep.* **8**, 5215 (2018).
 209. Sui, B. D., Hu, C. H., Zheng, C. X. & Jin, Y. Microenvironmental Views on Mesenchymal Stem Cell Differentiation in Aging. *J. Dent. Res.* **95**, 1333–1340 (2016).
 210. Farr, J. N. & Khosla, S. Cellular senescence in bone. *Bone* **121**, 121–133 (2019).
 211. Beane, O. S., Fonseca, V. C., Cooper, L. L., Koren, G. & Darling, E. M. Impact of Aging on the Regenerative Properties of Bone Marrow-, Muscle-, and Adipose-Derived Mesenchymal Stem/Stromal Cells. *PLoS One* **9**, e115963 (2014).
 212. Zhang, Z.-M., Jiang, L.-S., Jiang, S.-D. & Dai, L.-Y. Osteogenic potential and responsiveness to leptin of mesenchymal stem cells between postmenopausal women with osteoarthritis and osteoporosis. *J. Orthop. Res.* **27**, 1067–1073 (2009).
 213. Rodríguez, J. P., Montecinos, L., Ríos, S., Reyes, P. & Martínez, J. Mesenchymal stem cells from osteoporotic patients produce a type I collagen-deficient extracellular matrix favoring adipogenic differentiation. *J. Cell. Biochem.* **79**, 557–65 (2000).
 214. Jin, W.-J., Jiang, S.-D., Jiang, L.-S. & Dai, L.-Y. Differential responsiveness to 17 β -estradiol of mesenchymal stem cells from postmenopausal women between osteoporosis and osteoarthritis. *Osteoporos. Int.* **23**, 2469–2478 (2012).
 215. Pino, A. M., Rosen, C. J. & Rodríguez, J. P. In osteoporosis, differentiation

- of mesenchymal stem cells (MSCs) improves bone marrow adipogenesis. *Biol. Res.* **45**, 279–87 (2012).
216. Gunawardene, P. *et al.* Association Between Circulating Osteogenic Progenitor Cells and Disability and Frailty in Older Persons: The Nepean Osteoporosis and Frailty Study. *J. Gerontol. A. Biol. Sci. Med. Sci.* **71**, 1124–30 (2016).
 217. Hoogduijn, M. J. *et al.* No Evidence for Circulating Mesenchymal Stem Cells in Patients with Organ Injury. *Stem Cells Dev.* **23**, 2328–2335 (2014).
 218. Alm, J. J. *et al.* Circulating plastic adherent mesenchymal stem cells in aged hip fracture patients. *J. Orthop. Res.* **28**, 1634–42 (2010).
 219. Su, P. *et al.* Mesenchymal Stem Cell Migration during Bone Formation and Bone Diseases Therapy. *Int. J. Mol. Sci.* **19**, 2343 (2018).
 220. López-Lucas, M. D. *et al.* Production via good manufacturing practice of exofucosylated human mesenchymal stromal cells for clinical applications. *Cytotherapy* **20**, 1110–1123 (2018).
 221. Sackstein, R. *et al.* Ex vivo glycan engineering of CD44 programs human multipotent mesenchymal stromal cell trafficking to bone. *Nat. Med.* **14**, 181–187 (2008).
 222. Yuan, Z. *et al.* PPAR γ and Wnt Signaling in Adipogenic and Osteogenic Differentiation of Mesenchymal Stem Cells. *Curr. Stem Cell Res. Ther.* **11**, 216–25 (2016).
 223. Choi, Y. J. *et al.* Transcriptional profiling of human femoral mesenchymal stem cells in osteoporosis and its association with adipogenesis. *Gene* **632**, 7–15 (2017).
 224. Kim, J. & Ko, J. A novel PPAR γ 2 modulator sLZIP controls the balance between adipogenesis and osteogenesis during mesenchymal stem cell differentiation. *Cell Death Differ.* **21**, 1642–1655 (2014).
 225. Hu, L. *et al.* Mesenchymal Stem Cells: Cell Fate Decision to Osteoblast or Adipocyte and Application in Osteoporosis Treatment. *Int. J. Mol. Sci.* **19**, 360 (2018).
 226. Wang, C. *et al.* Differentiation of Bone Marrow Mesenchymal Stem Cells in Osteoblasts and Adipocytes and its Role in Treatment of Osteoporosis. *Med. Sci. Monit.* **22**, 226–33 (2016).
 227. Li, Y. *et al.* PPAR- γ and Wnt Regulate the Differentiation of MSCs into

- Adipocytes and Osteoblasts Respectively. *Curr. Stem Cell Res. Ther.* **13**, 185–192 (2018).
228. Augello, A. & De Bari, C. The Regulation of Differentiation in Mesenchymal Stem Cells. *Hum. Gene Ther.* **21**, 1226–1238 (2010).
 229. Rauch, A. *et al.* Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nat. Genet.* **51**, 716–727 (2019).
 230. Li, Z. *et al.* Epigenetic dysregulation in mesenchymal stem cell aging and spontaneous differentiation. *PLoS One* **6**, e20526 (2011).
 231. Bruderer, M., Richards, R. G., Alini, M. & Stoddart, M. J. Role and regulation of RUNX2 in osteogenesis. *Eur. Cell. Mater.* **28**, 269–86 (2014).
 232. Li, B. MicroRNA Regulation in Osteogenic and Adipogenic Differentiation of Bone Mesenchymal Stem Cells and its Application in Bone Regeneration. *Curr. Stem Cell Res. Ther.* **13**, 26–30 (2017).
 233. Huang, J., Zhao, L., Xing, L. & Chen, D. MicroRNA-204 Regulates Runx2 Protein Expression and Mesenchymal Progenitor Cell Differentiation. *Stem Cells* **28**, 357–364 (2010).
 234. Zuo, B. *et al.* microRNA-103a Functions as a Mechanosensitive microRNA to Inhibit Bone Formation Through Targeting Runx2. *J. Bone Miner. Res.* **30**, 330–345 (2015).
 235. Zhao, W. *et al.* Runx2 and microRNA regulation in bone and cartilage diseases. *Ann. N. Y. Acad. Sci.* **1383**, 80–87 (2016).
 236. Pasumathy, K. K. *et al.* Methylome Analysis of Human Bone Marrow MSCs Reveals Extensive Age- and Culture-Induced Changes at Distal Regulatory Elements. *Stem cell reports* **9**, 999–1015 (2017).
 237. Fernández, A. F. *et al.* H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Res.* **25**, 27–40 (2015).
 238. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **49**, 359–367 (2013).
 239. Weidner, C. *et al.* Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **15**, R24 (2014).
 240. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany, NY)*. **10**, 573–591 (2018).

241. Wang, J. *et al.* DNA methylation patterns of adult survivors of adolescent/young adult Hodgkin lymphoma compared to their unaffected monozygotic twin. *Leuk. Lymphoma* **60**, 1429–1437 (2019).
242. Rosen, A. D. *et al.* DNA methylation age is accelerated in alcohol dependence. *Transl. Psychiatry* **8**, 182 (2018).
243. Hofstadter, E. W. *et al.* Increased epigenetic age in normal breast tissue from luminal breast cancer patients. *Clin. Epigenetics* **10**, 112 (2018).
244. Horvath, S. *et al.* Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany. NY)*. **8**, 1485–1512 (2016).
245. Horvath, S. *et al.* The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany. NY)*. **7**, 294–306 (2015).
246. Vidal-Bralo, L. *et al.* Specific premature epigenetic aging of cartilage in osteoarthritis. *Aging (Albany. NY)*. **8**, 2222–2231 (2016).
247. Fernandez-Rebollo, E. *et al.* Primary Osteoporosis Is Not Reflected by Disease-Specific DNA Methylation or Accelerated Epigenetic Age in Blood. *J. Bone Miner. Res.* **33**, 356–361 (2018).
248. Morris, J. A. *et al.* Epigenome-wide Association of DNA Methylation in Whole Blood With Bone Mineral Density. *J. Bone Miner. Res.* **32**, 1644–1650 (2017).
249. Reppe, S. *et al.* Distinct DNA methylation profiles in bone and blood of osteoporotic and healthy postmenopausal women. *Epigenetics* **12**, 674–687 (2017).
250. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115-9 (2004).
251. Shukla, P. *et al.* Interleukin 27 (IL-27) Alleviates Bone Loss in Estrogen-deficient Conditions by Induction of Early Growth Response-2 Gene. *J. Biol. Chem.* **292**, 4686–4699 (2017).
252. Kendler, D. L. *et al.* The risk of subsequent osteoporotic fractures is decreased in subjects experiencing fracture while on denosumab: results from the FREEDOM and FREEDOM Extension studies. *Osteoporos. Int.* **30**, 71–78 (2019).
253. Abe, T. *et al.* The effect of mesenchymal stem cells on osteoclast precursor cell differentiation. *J. Oral Sci.* **61**, 30–35 (2019).

254. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61-5 (2007).
255. Maruyama, K. *et al.* Bone-protective Functions of Netrin 1 Protein. *J. Biol. Chem.* **291**, 23854–23868 (2016).
256. Sato, T. *et al.* Functional Roles of Netrin-1 in Osteoblast Differentiation. *In Vivo (Brooklyn)*. **31**, 321–328 (2017).
257. Sharma-Bhandari, A., Park, S.-H., Kim, J.-Y., Oh, J. & Kim, Y. Lysyl oxidase modulates the osteoblast differentiation of primary mouse calvaria cells. *Int. J. Mol. Med.* **36**, 1664–1670 (2015).
258. Peng, S. *et al.* An Overview of Long Noncoding RNAs Involved in Bone Regeneration from Mesenchymal Stem Cells. *Stem Cells Int.* **2018**, 8273648 (2018).
259. Zhang, J., Hao, X., Yin, M., Xu, T. & Guo, F. Long non-coding RNA in osteogenesis: A new world to be explored. *Bone Joint Res.* **8**, 73–80 (2019).
260. Barman, P., Reddy, D. & Bhaumik, S. R. Mechanisms of Antisense Transcription Initiation with Implications in Gene Expression, Genomic Integrity and Disease Pathogenesis. *Non-coding RNA* **5**, E11 (2019).
261. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202–6 (2008).
262. Tufarelli, C. *et al.* Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* **34**, 157–165 (2003).
263. The Jackson Laboratory. Mouse Genome Database (MGD) at the Mouse Genome Informatics website. Available at: <http://www.informatics.jax.org/>.
264. Huang, T.-S. *et al.* LINC00341 exerts an anti-inflammatory effect on endothelial cells by repressing VCAM1. *Physiol. Genomics* **49**, 339–345 (2017).
265. Liu, X., Ma, B.-D., Liu, S., Liu, J. & Ma, B.-X. Long noncoding RNA LINC00341 promotes the vascular smooth muscle cells proliferation and migration via miR-214/FOXO4 feedback loop. *Am. J. Transl. Res.* **11**, 1835–1842 (2019).
266. Liao, M. *et al.* Relationship between LINC00341 expression and cancer

- prognosis. *Oncotarget* **8**, 15283–15293 (2017).
267. Xu, Y. *et al.* LncRNA LINC00341 mediates PM 2.5 -induced cell cycle arrest in human bronchial epithelial cells. *Toxicol. Lett.* **276**, 1–10 (2017).
268. Yang, Q. *et al.* A LINC00341-mediated regulatory pathway supports chondrocyte survival and may prevent osteoarthritis progression. *J. Cell. Biochem.* **120**, 10812–10820 (2019).
269. Penrod, N. M., Greene, C. S. & Moore, J. H. Predicting targeted drug combinations based on Pareto optimal patterns of coexpression network connectivity. *Genome Med.* **6**, 33 (2014).
270. Krawczyk, M. & Emerson, B. M. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-κB complexes. *Elife* **3**, e01776 (2014).
271. Qian, M. *et al.* P50-associated COX-2 extragenic RNA (PACER) overexpression promotes proliferation and metastasis of osteosarcoma cells by activating COX-2 gene. *Tumor Biol.* **37**, 3879–3886 (2016).
272. Liu, H. *et al.* Isolation, culture and induced differentiation of rabbit mesenchymal stem cells into osteoblasts. *Exp. Ther. Med.* **15**, 3715–3724 (2018).
273. Huang, C. *et al.* The Spatiotemporal Role of COX-2 in Osteogenic and Chondrogenic Differentiation of Periosteum-Derived Mesenchymal Progenitors in Fracture Repair. *PLoS One* **9**, e100079 (2014).
274. Pountos, I. *et al.* NSAIDS inhibit in vitro MSC chondrogenesis but not osteogenesis: implications for mechanism of bone formation inhibition in man. *J. Cell. Mol. Med.* **15**, 525–534 (2011).
275. Wang, C.-G. *et al.* LncRNA KCNQ1OT1 promoted BMP2 expression to regulate osteogenic differentiation by sponging miRNA-214. *Exp. Mol. Pathol.* **107**, 77–84 (2019).
276. Li, Z. *et al.* Long non-coding RNA MEG3 inhibits adipogenesis and promotes osteogenesis of human adipose-derived mesenchymal stem cells via miR-140-5p. *Mol. Cell. Biochem.* **433**, 51–60 (2017).
277. Li, C.-J. *et al.* Long noncoding RNA Bmncr regulates mesenchymal stem cell fate during skeletal aging. *J. Clin. Invest.* **128**, 5251–5266 (2018).
278. Cote, A. J. *et al.* Single-cell differences in matrix gene expression do not predict matrix deposition. *Nat. Commun.* **7**, 10865 (2016).

279. Stiehler, M. *et al.* In vitro characterization of bone marrow stromal cells from osteoarthritic donors. *Stem Cell Res.* **16**, 782–9 (2016).
280. Jones, E. *et al.* Large-scale extraction and characterisation of CD271 + multipotential stromal cells (MSCs) from trabecular bone in health and osteoarthritis: Implications for bone regeneration strategies based on minimally-cultured MSCs. *Arthritis Rheum.* **62**, 1944–54 (2010).
281. Campbell, T. M. *et al.* Mesenchymal Stem Cell Alterations in Bone Marrow Lesions in Patients With Hip Osteoarthritis. *Arthritis Rheumatol.* **68**, 1648–1659 (2016).
282. Jiang, S. S. *et al.* Gene expression profiling suggests a pathological role of human bone marrow-derived mesenchymal stem cells in aging-related skeletal diseases. *Aging (Albany. NY).* **3**, 672–684 (2011).
283. Murphy, J. M. *et al.* Reduced chondrogenic and adipogenic activity of mesenchymal stem cells from patients with advanced osteoarthritis. *Arthritis Rheum.* **46**, 704–713 (2002).
284. Tornero-Esteban, P. *et al.* Altered Expression of Wnt Signaling Pathway Components in Osteogenesis of Mesenchymal Stem Cells in Osteoarthritis Patients. *PLoS One* **10**, e0137170 (2015).
285. Jiang, L.-S., Zhang, Z.-M., Jiang, S.-D., Chen, W.-H. & Dai, L.-Y. Differential Bone Metabolism Between Postmenopausal Women With Osteoarthritis and Osteoporosis. *J. Bone Miner. Res.* **23**, 475–483 (2007).
286. Im, G.-I. & Kim, M.-K. The relationship between osteoarthritis and osteoporosis. *J. Bone Miner. Metab.* **32**, 101–109 (2014).
287. Bentivegna, A. *et al.* DNA Methylation Changes during In Vitro Propagation of Human Mesenchymal Stem Cells: Implications for Their Genomic Stability? *Stem Cells Int.* **2013**, 192425 (2013).
288. Nestor, C. E. *et al.* Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. *Genome Biol.* **16**, 11 (2015).
289. Mo, M., Wang, S., Zhou, Y., Li, H. & Wu, Y. Mesenchymal stem cell subpopulations: phenotype, property and therapeutic potential. *Cell. Mol. Life Sci.* **73**, 3311–3321 (2016).
290. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).

10.APPENDIX

APPENDIX

Appendix 1: List of differentially expressed genes

List of differentially expressed genes in BMSCs grown from patients with fractures (FRX) and controls with osteoarthritis (OA). Significant genes are considered with an FDR<0.1 and an absolute fold change > 2 between BMSCs obtained from FRX and OA.

Gene Expression upregulated in Osteoporosis			Gene Expression downregulated in Osteoporosis		
Gene_name	logFC	FDR	Gene_name	logFC	FDR
IBSP	-2,60	2,81E-03	ELN	2,64	4,59E-04
GLB1L3	-3,01	4,60E-03	LAYN	1,82	7,24E-04
SLC29A1	-1,46	6,32E-03	LYZ	4,18	1,59E-03
KLRC2	-3,16	8,29E-03	SYNPO2	4,44	1,84E-03
LGR6	-3,54	8,29E-03	EDIL3	2,15	1,84E-03
AMOT	-1,92	1,29E-02	OSBPL1A	1,94	2,11E-03
SH3GL1P3	-1,33	1,33E-02	ID4	3,50	2,79E-03
ACAD8	-2,32	1,43E-02	ZNF518A	2,37	5,24E-03
NCAM1	-1,87	1,51E-02	SLC19A1	2,40	5,83E-03
PKD1P6	-2,87	1,63E-02	LPAL2	3,47	5,83E-03
FBXL13	-1,76	1,83E-02	PPP1R14A	2,67	5,83E-03
WNT5B	-1,42	1,93E-02	TGFB1	2,80	6,09E-03
BGLAP	-3,36	2,26E-02	CD27-AS1	3,53	7,63E-03
EPHA2	-1,50	2,30E-02	HPS3	2,93	8,17E-03
UNC5B	-1,16	2,74E-02	LRRC16A	2,83	8,17E-03
STMN3	-2,03	2,74E-02	KLF9	1,55	8,17E-03
ENPP1	-1,56	2,80E-02	DMD-AS1	2,23	1,22E-02
KIAA1324L	-1,10	2,84E-02	HCG4P11	1,90	1,25E-02
PCDH10	-1,80	2,84E-02	IGHG3	4,65	1,28E-02
CXorf57	-1,61	2,84E-02	AMPD3	1,69	1,29E-02
ABCA3	-1,84	2,84E-02	ACSL5	2,34	1,29E-02
INSC	-2,51	2,84E-02	FMO3	2,80	1,29E-02
COL8A2	-2,04	2,84E-02	DGAT2	1,98	1,29E-02
TNFRSF11B	-1,87	2,84E-02	VGLL3	2,97	1,33E-02
SGCD	-1,42	2,84E-02	HLA-F-AS1	2,30	1,33E-02
MYO1D	-1,19	2,89E-02	SCRN1	2,88	1,43E-02
GALNT3	-1,86	2,89E-02	FAM110B	2,34	1,51E-02
LSP1	-1,24	3,39E-02	ENTPD1	3,01	1,58E-02
SORCS2	-2,51	3,83E-02	BASP1	1,24	1,63E-02
TRIM7	-1,96	3,84E-02	TXNRD1	2,87	1,73E-02
SLC12A7	-2,32	3,86E-02	PAPPA-AS1	1,69	1,73E-02
RTKL1	-2,53	3,99E-02	SLC24A3	3,93	1,82E-02
SCN9A	-2,29	3,99E-02	IGKV3-20	5,04	1,83E-02

Gene Expression upregulated in Osteoporosis			Gene Expression downregulated in Osteoporosis		
Gene_name	logFC	FDR	Gene_name	logFC	FDR
SHROOM2	-1,40	3,99E-02	LAMA4	2,05	1,83E-02
DKK3	-1,12	4,40E-02	FAM212B	1,23	1,83E-02
STARD5	-1,20	4,40E-02	IGHA1	4,09	1,99E-02
LETM2	-1,30	4,40E-02	CTSL1	1,99	1,99E-02
WDR90	-1,46	4,53E-02	FNBP1L	1,38	2,01E-02
LSP1P2	-2,25	4,56E-02	MUC15	3,25	2,04E-02
FMN1	-1,40	4,74E-02	CHL1-AS2	3,22	2,07E-02
GNG2	-1,09	4,86E-02	CHL1	4,50	2,24E-02
BMPER	-1,29	4,86E-02	IGJ	3,83	2,27E-02
CUBN	-1,13	4,95E-02	BPMS	1,23	2,29E-02
PDPK2	-1,06	4,97E-02	ADAMTS2	1,81	2,30E-02
BMP2	-2,27	5,10E-02	LOXL2	2,22	2,42E-02
P2RX6	-1,62	5,10E-02	COL4A1	2,02	2,42E-02
CADM1	-1,35	5,10E-02	EFEMP1	1,93	2,42E-02
SOX9	-1,48	5,40E-02	HLA-B	1,78	2,42E-02
SRPX	-1,14	5,44E-02	CELF2	1,89	2,47E-02
DNLZ	-1,04	5,46E-02	ANXA5	2,45	2,47E-02
TYMP	-1,10	5,74E-02	CCR7	1,85	2,47E-02
GNGT1	-3,18	5,74E-02	RNF43	2,05	2,58E-02
STXBP6	-1,86	5,80E-02	VCAN	1,95	2,58E-02
DNAJC22	-1,65	5,86E-02	RNF141	2,14	2,65E-02
HOMER2	-1,17	5,88E-02	LAMC1	2,17	2,70E-02
ARVCF	-1,34	5,99E-02	BST2	2,09	2,74E-02
GALNTL4	-1,16	5,99E-02	IGHG4	6,97	2,74E-02
RANBP3L	-2,88	6,04E-02	HERC5	1,93	2,74E-02
SERPINF1	-1,54	6,09E-02	MAN1C1	2,14	2,74E-02
CKM	-1,15	6,24E-02	SLC22A3	3,10	2,74E-02
CCDC158	-1,27	6,27E-02	CD4	1,88	2,74E-02
LRRC46	-1,08	6,42E-02	IGHD	5,83	2,79E-02
DNM1	-2,22	6,42E-02	MRVI1-AS1	2,08	2,82E-02
DLX6-AS2	-1,12	6,50E-02	SERPINB9	1,81	2,84E-02
CMKLR1	-1,22	6,51E-02	IGHGP	5,43	2,84E-02
KIAA1217	-1,06	6,54E-02	MAOB	2,90	2,84E-02
TDG	-1,30	6,64E-02	CRISPLD2	3,50	2,86E-02
C1QTNF9B	-1,45	6,69E-02	PDK1	2,15	2,95E-02
SYT12	-1,74	6,86E-02	PTGFRN	1,52	2,96E-02
LNK1	-1,03	6,86E-02	CP	3,40	3,01E-02
GALNTL1	-1,90	6,86E-02	TES	1,26	3,22E-02
EMBP1	-1,81	6,86E-02	GUCY1B3	1,25	3,32E-02
TNNC1	-2,03	7,03E-02	NME7	1,13	3,36E-02
FOXP2	-1,38	7,06E-02	ARHGAP20	3,20	3,37E-02
LRP4-AS1	-1,07	7,42E-02	NUAK2	1,73	3,37E-02
ENOX1	-1,11	7,75E-02	IGHG1	3,79	3,39E-02

Gene Expression upregulated in Osteoporosis			Gene Expression downregulated in Osteoporosis		
Gene_name	logFC	FDR	Gene_name	logFC	FDR
MEOX2	-1,83	7,97E-02	LARP4B	2,95	3,39E-02
RNF112	-1,87	7,97E-02	TM4SF19-AS1	2,64	3,39E-02
GRIA1	-2,02	8,06E-02	IGHMBP2	2,19	3,39E-02
TRIM59	-2,21	8,15E-02	CNTNAP2	2,77	3,45E-02
PRTG	-1,30	8,24E-02	RPS6KA2	2,95	3,45E-02
OSBP2	-2,32	8,24E-02	EFNB2	2,03	3,48E-02
LRRC17	-2,01	8,54E-02	ADCY2	3,19	3,48E-02
RAET1G	-1,06	8,54E-02	MED16	1,84	3,49E-02
LRRC15	-1,71	8,84E-02	PMP22	3,05	3,58E-02
BTBD11	-1,95	9,24E-02	CDKN1C	3,05	3,70E-02
SERINC2	-1,33	9,24E-02	CHRD1	4,00	3,72E-02
PNPLA3	-1,09	9,24E-02	PPP1R3C	2,67	3,76E-02
ARHGAP22	-1,41	9,29E-02	MMP9	2,86	3,77E-02
C1GALT1	-1,05	9,45E-02	SPARC	2,03	3,77E-02
ARFRP1	-1,95	9,50E-02	MAF	1,55	3,77E-02
MRPS14	-1,29	9,54E-02	IGHG2	3,42	3,80E-02
ARHGEF19	-1,15	9,61E-02	B2M	2,12	3,87E-02
CACNA1G	-1,97	9,72E-02	C1orf183	1,29	3,91E-02
TCEAL7	-1,82	9,72E-02	EFR3B	1,53	3,99E-02
IL16	-1,21	9,72E-02	ADA	1,06	3,99E-02
GLRB	-1,24	9,79E-02	CHI3L1	1,95	3,99E-02
LOXL4	-1,80	9,79E-02	PAPPA	2,23	4,02E-02
ETV1	-2,25	9,79E-02	ANO3	2,80	4,06E-02
			ATXN1L	3,26	4,29E-02
			IGFBP4	1,44	4,40E-02
			SHC2	2,69	4,40E-02
			TNS1	2,89	4,53E-02
			CYP1B1	2,29	4,53E-02
			XYLT1	2,23	4,53E-02
			STK3	2,23	4,53E-02
			FTL	2,06	4,53E-02
			MPZL1	1,98	4,53E-02
			PTX3	1,78	4,53E-02
			LGMN	2,69	4,56E-02
			LASP1	1,90	4,59E-02
			MED11	2,50	4,63E-02
			IL32	3,57	4,74E-02
			TFRC	2,98	4,83E-02
			DERL3	3,70	4,86E-02
			EYA1	1,75	4,86E-02
			FSTL1	1,79	4,86E-02
			GALNTL2	2,03	4,94E-02
			MAP6	1,72	4,97E-02

Gene Expression downregulated in
Osteoporosis

Gene_name	logFC	FDR
WASF2	2,61	4,97E-02
GPM6B	2,74	5,10E-02
KCNE4	1,48	5,10E-02
FBN2	1,51	5,15E-02
IGKV1-5	4,44	5,18E-02
COL14A1	2,94	5,31E-02
LMOD1	1,46	5,33E-02
TINAGL1	2,23	5,37E-02
PPARG	1,35	5,40E-02
PTMA	2,05	5,42E-02
STK38L	2,59	5,60E-02
TMTC3	3,38	5,76E-02
NDE1	3,38	5,77E-02
EDNRA	1,47	5,80E-02
GLIPR2	1,04	5,80E-02
COL3A1	1,73	5,81E-02
PEAR1	1,95	5,86E-02
NTRK2	2,85	5,86E-02
ID2	1,21	5,86E-02
IGKV2-28	5,44	5,96E-02
ITGA4	1,13	5,96E-02
IFI27	2,62	5,96E-02
GRK5	1,43	5,96E-02
RP11-145A3.1	1,37	5,99E-02
PREX2	2,27	6,12E-02
SPON2	2,12	6,12E-02
SAMD11	1,77	6,12E-02
IGHJ4	5,60	6,24E-02
TGFB3	1,22	6,24E-02
APBA2	1,88	6,24E-02
IGKV1-6	4,44	6,24E-02
MRVI1	2,28	6,25E-02
RP11-428C6.1	1,00	6,41E-02
VAMP8	1,73	6,42E-02
PTGES	1,14	6,42E-02
IL21R	1,13	6,42E-02
ID1	1,70	6,44E-02
KIT	2,43	6,50E-02
PRDM1	2,02	6,51E-02
DMD	2,60	6,51E-02
CPA3	3,96	6,51E-02
TCF7L1	1,07	6,51E-02
TAP1	1,25	6,58E-02

Gene Expression downregulated in
Osteoporosis

Gene_name	logFC	FDR
RND3	2,68	6,64E-02
CALCOCO2	2,54	6,64E-02
EMP1	2,28	6,80E-02
OSR2	2,37	6,85E-02
BLNK	2,52	6,85E-02
CCDC81	1,80	6,85E-02
MYOM1	1,27	6,85E-02
SLC22A4	1,20	6,86E-02
SLC5A3	3,28	6,87E-02
CDC5L	2,94	7,00E-02
PTBP3	2,62	7,02E-02
CALD1	1,83	7,02E-02
LYN	1,14	7,03E-02
ACTG2	3,60	7,11E-02
FAM89A	1,45	7,11E-02
DUSP1	1,05	7,31E-02
PPP2R5C	2,15	7,42E-02
JAG1	1,85	7,42E-02
ARHGAP23	1,48	7,42E-02
KREMEN1	2,97	7,69E-02
FAM65C	2,34	7,72E-02
APBB1IP	2,50	7,79E-02
AC017048.3	1,26	7,79E-02
LRRC28	2,74	8,03E-02
TNS3	1,94	8,06E-02
IGKV4-1	4,73	8,06E-02
ICAM1	1,61	8,09E-02
MORF4L1P1	2,09	8,18E-02
SLC2A5	2,36	8,19E-02
ARRB1	1,27	8,24E-02
IGKV3D-20	3,21	8,24E-02
ISLR	1,05	8,24E-02
CSF2RB	1,51	8,24E-02
C2	1,43	8,24E-02
CD74	1,16	8,27E-02
LEPR	2,14	8,45E-02
ANXA1	2,09	8,45E-02
GPRC5A	1,62	8,45E-02
STX12	3,27	8,50E-02
RBM3	2,31	8,50E-02
COL5A3	2,50	8,55E-02
ST5	1,31	8,56E-02
IGLJ3	3,65	8,56E-02

Gene Expression downregulated in
Osteoporosis

Gene_name	logFC	FDR
ARL6IP5	2,72	8,65E-02
TMED2	2,12	8,65E-02
PFDN5	2,14	8,82E-02
AKAP12	2,38	9,08E-02
IFI30	1,25	9,09E-02
IRF4	2,96	9,11E-02
MGST1	1,04	9,24E-02
CPNE5	2,67	9,24E-02
TMEM176A	3,68	9,35E-02
ADAM10	2,47	9,35E-02
CNTN1	2,13	9,49E-02
HSD17B6	1,69	9,49E-02
CANX	1,96	9,50E-02
CCDC41	1,43	9,50E-02
ANK3	2,82	9,54E-02
ADAM9	1,82	9,54E-02
STC2	1,91	9,56E-02
IGLC3	3,89	9,67E-02
GM2A	2,67	9,67E-02
KLHL13	1,24	9,72E-02
CD38	3,03	9,72E-02
CSTA	1,15	9,72E-02
NUB1	2,53	9,75E-02
ZNF423	2,15	9,75E-02
SMOC1	1,93	9,75E-02
MDM2	1,91	9,75E-02
HIST1H2AC	3,26	9,89E-02
SLC2A3	2,57	9,89E-02
TTC3P1	2,29	9,90E-02
INHBA	2,14	9,90E-02
ANP32A	1,88	9,90E-02
IGLC2	4,43	9,92E-02
LDHA	1,89	9,93E-02
MYL9	1,22	9,95E-02

Appendix 2: List of genes from the intersections

List of genes from the intersections between differentially methylated enhancers and differentially expressed genes in BMSCs isolated from FRX and controls with OA. Some genes may have both hypomethylated and hypermethylated regions.

Hypermethylated and hypomethylated	KCNMA1; TNK1; FOXP4; EHF; GRAMD1B; CGNL1; MSC; EDIL3; CBLN1; SLC25A37; EFNA5; NXN; MAX; TMEM174; SRP14; KIAA0513; PRL; TSKU; FAM65B; TLE3; FGGY; RHOU; DNAJC15; CCNY; SNX9; GMPR; LYVE1; SELPLG; FAM76B; PDZRN3; TBL1XR1; ADAMTS9; TMEM167A; CPEB4; PTPRG; TBX3; NRP2; MSI2; PLEC; ZNF423; DCP2; CBR4; ZBTB20; SLC41A2; DYNC1I1; ZEB2; GRHL2; EIF4E3; FBXL7; NOS1AP; DYRK2; ABCA4; MAP2K6; CYP26B1; TSPAN18; SP3; AZIN1; SOX9; GLI2; ZFPM2; OXNAD1; TGFB2; ANKRD46; CLDN20; SH3GL3; TCF7L2; FGFR2; PTPRJ; FOXP1; SPRY4; BOC; XYLT1; IRX1; TPM1; KITLG; WBSCR17; UBE2V2; TNFRSF19; SPECC1; PTPN14; TNS1; EPAS1; SPATS2L; THSD4; TJP1; ARHGAP26; ERFF1; BCAT1; CDH11; FAM20C; CHST2; SLC45A1; CPN2; AUTS2; GAB1; PFN4; CALM1; COL13A1; ANGPT1; IRF2BPL; SMAD6; RPH3A; NR2F2; TMTC2; ZNF608; CUL1; LBR
Hypermethylated and overexpressed	LRRC17; NCAM1; SOX9; RNF112; GRIA1; C1GALT1; BMPER; ARHGAP22
Hypermethylated and underexpressed	VAMP8; PTMA; EDIL3; SYNPO2; LAYN; TNS3; INHBA; LOXL2; ISLR; NME7; SERPINB9; AKAP12; PAPPAS1; ZNF423; CHI3L1; ST5; ID2; RBPMS; EYA1; LMOD1; GM2A; MPZL1; NUAK2; DUSP1; ANK3; ARRB1; SPARC; ANO3; LAMA4; RND3; TNS1; IGFBP4; SLC19A1; TFRC; SLC5A3; CRISPLD2; LASP1; LAMC1; PRDM1
Hypomethylated and overexpressed	FOXP2; UNC5B; ENOX1; TNFRSF11B; TNNC1; SLC29A1; SGCD; FMN1; SOX9; LRRC15
Hypomethylated and underexpressed	EDIL3; FBNP1L; CPNE5; KREMEN1; CNTN1; ZNF423; CNTNAP2; CELF2; APBA2; PDK1; SCRNI; TNS1; APBB1IP; ARHGAP23; XYLT1
Hyper and hypomethylated and overexpressed	SOX9
Hyper and hypomethylated and underexpressed	EDIL3; ZNF423; TNS1; TSPAN18

Appendix 3: List of publications

[1] del Real A, et al. Differential analysis of genome-wide methylation and gene expression in mesenchymal stem cells of patients with fractures and osteoarthritis, *Epigenetics*. 12 (2017) 113–122.

[2] del Real A, Riancho-Zarrabeitia L, López-Delgado L, Riancho JA, *Epigenetics of Skeletal Diseases*, *Curr. Osteoporos. Rep.* 16 (2018) 245–255.

[3] del Real A, Riancho-Zarrabeitia L, Riancho JA, *Epigenetic Aging in Osteoporosis*, *J. Bone Miner. Res.* 33 (2018) 1902–1903.

Differential analysis of genome-wide methylation and gene expression in mesenchymal stem cells of patients with fractures and osteoarthritis

Alvaro del Real, Flor M. Pérez-Campo, Agustín F. Fernández, Carolina Sañudo, Carmen G. Ibarbia, María I. Pérez-Nuñez, Wim Van Criekinge, Maarten Braspenning, María A. Alonso, Mario F. Fraga, and Jose A. Riancho

Epigenetics 2017, Vol. 12, No. 2, 113–122

<http://dx.doi.org/10.1080/15592294.2016.1271854>

Epigenetics of Skeletal Diseases

Alvaro del Real, Leyre Riancho-Zarrabeitia, Laura López-Delgado, José A. Riancho

Current Osteoporosis Reports (2018) 16:246–255

<https://doi.org/10.1007/s11914-018-0435-y>

Epigenetic Aging in Osteoporosis

Alvaro del Real,¹ Leyre Riancho-Zarrabeitia,² and Jose A Riancho¹

¹Department of Internal Medicine, Hospital U.M. Valdecilla IDIVAL, University of Cantabria, Santander, Spain

²Service of Rheumatology, Hospital Sierrallana, Torrelavega, Spain

To the Editor:

We have read with great interest the recent article in the *Journal of Bone and Mineral Research* "Primary Osteoporosis Is Not Reflected by Disease-Specific DNA Methylation or Accelerated Epigenetic Age in Blood."⁽¹⁾ The authors analyzed age-related DNA methylation profiles in peripheral blood cells and did not find significant differences between osteoporotic patients and non-osteoporotic controls.

DNA methylation is an epigenetic mechanism that regulates gene expression and cell differentiation and activity. Horvath described a set of specific CpGs that show age-related changes in methylation in a variety of cell types and tissues and therefore represent an index of the epigenetic aging of those tissues.⁽²⁾

The results by Fernandez-Rebollo and colleagues suggest that there are not differentially methylated age-related CpG sites in blood from osteoporotic patients when compared with blood from non-osteoporotic subjects. These results are in line with a previous report by Morris and colleagues, who did not find significant associations between DNA methylation in blood cells and bone mineral density.⁽³⁾

However, it is important to emphasize that epigenetic marks, including DNA methylation, are tissue specific, so in disorders like osteoporosis, methylation marks in leukocytes may not necessarily represent methylation signatures in other cells and particularly in those cells in bone tissue. In fact, we have previously shown that there are differences in the DNA methylation signature of bone tissue samples and mesenchymal stem cells (MSCs) between patients with fragility fractures and controls with osteoarthritis.^(4,5) Quite interesting, when we analyzed age-related methylation marks in patients with fractures and with osteoarthritis, we found an accelerated epigenetic aging in cartilage cells in the latter group, without significant differences in bone cells (Fig. 1A). These results would suggest that accelerated epigenetic aging is disease- and tissue-specific.⁽⁶⁾

Despite the lack of differences in age-related epigenetic marks in DNA extracted from bone tissue, an accelerated epigenetic aging in bone cells cannot be completely excluded. Bone is a complex and heterogeneous tissue, and analyzing whole bone samples may miss changes taking place in some cells. In fact, when we studied MSCs, an accelerated aging was evident in those grown from patients with osteoporotic hip fractures, in comparison with cells grown from patients with osteoarthritis (Fig. 1A).⁽⁵⁾ These results would be consistent with the concept that the potential role of epigenetic aging in

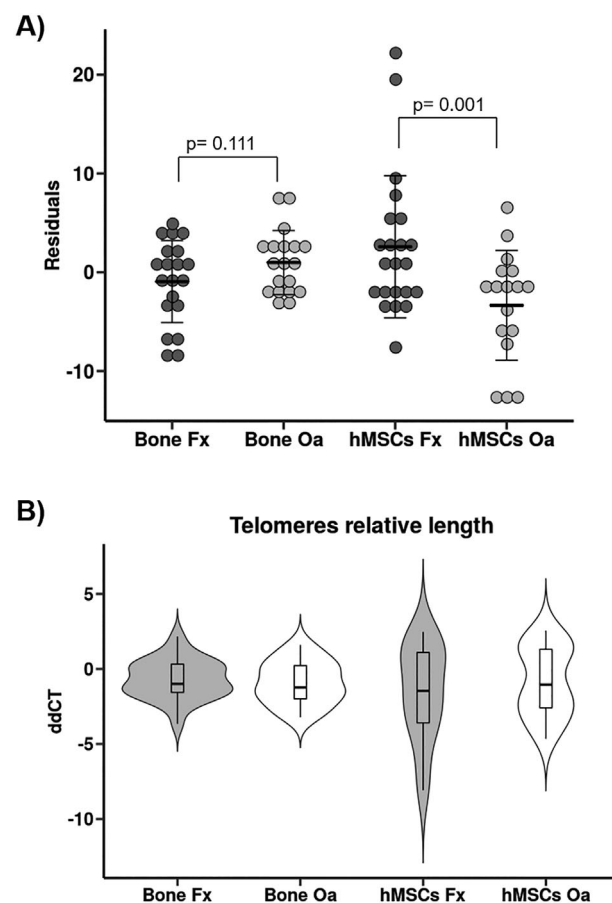


Fig 1. Epigenetic marks in bone tissue and MSCs from patients with hip fractures and with hip osteoarthritis. (A) Epigenetic aging as revealed by age-related DNA methylation marks. Dot plot with the mean and SD of the residuals (deviation from the overall regression line) from bone and human mesenchymal stem cells (hMSCs) isolated from osteoporotic fractures (Fx, dark gray dots) and osteoarthritis (Oa, gray dots) are shown. Between-group differences were compared by ANCOVA, with the chronological age as covariable. Data derived from previously reported results (Delgado-Calle et al. 2013; Vidal-Bralo et al. 2016; del Real et al. 2017). (B) Violin plots showing the distribution of the relative telomere length density from bone and MSCs isolated from osteoporotic fractures (in gray) and osteoarthritis (in white).

osteoporosis is more important at the precursor cell stage than in differentiated bone-forming cells. In theory, it could limit the pool of bone-forming cells and therefore the overall bone formation capacity, despite normal functioning of individual mature osteoblasts.

Telomeres shorten with cell divisions and may result in cell senescence when a length threshold is reached.⁽⁷⁾ Therefore, telomere length is also considered as a biomarker of aging and age-related diseases. So, we explored the relative telomere length in bone and in MSCs. We analyzed telomere length by real-time qPCR using a single copy gene (β -globin) as comparator as suggested by Cawthon.⁽⁸⁾ The results in each run were normalized in comparison with a set of three control DNAs. However, we did not find significant differences in the relative telomere length of either bone samples or MSCs of patients with fractures or with osteoarthritis (Fig. 1B).

Overall, the results of Fernandez-Rebollo and colleagues and those from our own group are consistent with the concept of an accelerated epigenetic aging, and specifically DNA methylation marks, in MSCs from osteoporotic patients that does not translate into blood leukocytes, which do not derive from MSCs, nor even in mature bone cells, at least when whole bone tissue is analyzed. This emphasizes the complexity of epigenome association studies, related to the cell and tissue specificity of epigenetic marks. Further studies in individual mature bone cells will be needed to clarify the role of DNA methylation and other aging-related epigenetic changes in the pathogenesis of bone disorders.

Acknowledgments

Supported in part by grants from the Instituto de Salud Carlos III (PI12/00615 and PI16/00915), which can be co-funded by FEDER funds from the European Union.

References

1. Fernandez-Rebollo E, Eipel M, Seefried L, et al. Primary osteoporosis is not reflected by disease-specific DNA methylation or accelerated epigenetic age in blood. *J Bone Miner Res*. 2018;33(2):356–61.
2. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
3. Morris JA, Tsai P-C, Joeheanes R, et al. Epigenome-wide association of DNA methylation in whole blood with bone mineral density. *J Bone Miner Res*. 2017;32(8):1644–50.
4. Delgado-Calle J, Fernández AF, Sainz J, et al. Genome-wide profiling of bone reveals differentially methylated regions in osteoporosis and osteoarthritis. *Arthritis Rheum*. 2013;65(1):197–205.
5. del Real A, Pérez-Campo FM, Fernández AF, et al. Differential analysis of genome-wide methylation and gene expression in mesenchymal stem cells of patients with fractures and osteoarthritis. *Epigenetics*. 2017;12(2):113–22.
6. Vidal-Bralo L, Lopez-Golan Y, Mera-Varela A, et al. Specific premature epigenetic aging of cartilage in osteoarthritis. *Aging (Albany NY)*. 2016;8(9):2222–31.
7. Wolkowitz OM, Mellon SH, Epel ES, et al. Leukocyte telomere length in major depression: correlations with chronicity, inflammation and oxidative stress—preliminary findings. *PLoS One*. 2011;6(3):e17837.
8. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res*. 2002;30(10):e47.