

DOCTORAL THESIS



**VNiVERSiDAD
D SALAMANCA**

**Experimental and bioinformatic
characterization of the
transcriptomic profile of human
Mesenchymal Stem Cells (MSCs)**

Elena Sánchez Luis

Ph.D. SUPERVISORS

Javier De Las Rivas Sanz, Ph.D.

Fermín Sánchez-Guijo Martín, M.D., Ph.D.

Sandra Muntión Olave, Ph.D.

Salamanca, Spain
2023

Dr. Javier De Las Rivas Sanz, con D.N.I. 15949000H, Investigador Científico del Consejo Superior de Investigaciones Científicas (CSIC), director del grupo de Bioinformática y Genómica Funcional en el Centro de Investigación del Cáncer (CiC-IBMCC), y profesor del Programa de Doctorado y del Máster de Biología y Clínica del Cáncer de dicho Centro y de la Universidad de Salamanca (USAL).

El **Dr. Fermín Sánchez-Guijo Martín**, con D.N.I 07869479Y, Profesor Titular de Medicina de la Universidad de Salamanca y Jefe del Servicio de Hematología del Hospital Universitario de Salamanca.

Y la **Dra. Sandra Muntión Olave** con D.N.I 16567353Q, Investigadora postdoctoral en el laboratorio de Terapia Celular del Servicio de Hematología del hospital Universitario de Salamanca.

CERTIFICAN

Que han dirigido la Tesis Doctoral titulada “**Experimental and bioinformatic characterization of the transcriptomic profile of human Mesenchymal Stem Cells (MSCs)**” realizada por Dña. **Elena Sánchez Luis**, dentro del Programa de Doctorado *Biociencias: Biología y Clínica del Cáncer y Medicina Traslacional* del Centro de Investigación del Cáncer (CiC-IBMCC, CSIC/USAL) y de la Universidad de Salamanca (USAL).


Y AUTORIZAN

La presentación de la misma, considerando que reúne las condiciones de originalidad y contenidos requeridos para optar al grado de Doctor por la Universidad de Salamanca.

En Salamanca, a 24 de Octubre de 2023



Dr. Javier De Las Rivas Sanz
Director



Dr. Fermín Sánchez-Guijo Martín
Codirector



Dra. Sandra Muntión Olave
Codirectora

Para la realización de esta Tesis Doctoral, la doctoranda **Elena Sánchez Luis** obtuvo en concurso público un **Contrato Predoctoral de Formación en Investigación en Salud (PFIS)** de cuatro años (2020-2023), financiada la **Acción Estratégica en Salud 2017-2020**, del Programa Estatal de Investigación Orientada a los Retos de la Sociedad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017- 2020, convocada por el **Instituto de Salud Carlos III (ISCIII)** (AES-19, Expediente FI19/00272).

Además, la investigación de esta Tesis Doctoral ha sido realizada gracias a los fondos proporcionados a varios Proyectos de Investigación competitivos concedidos al **Grupo del Dr. Javier De Las Rivas** (laboratorio 19) en el **Centro de Investigación del Cáncer (CiC-IBMCC, CSIC/USAL)**. En concreto, se pueden citar como proyectos en los que ha participado o colaborado la doctoranda: los Proyectos Nacionales de la AES del Instituto de Salud Carlos III (ISCiii) de referencia PI15/00328 y PI18/00591; el Proyecto Europeo Horizon 2020 *ArrestAD* de referencia 737390, iniciado el 1 de Enero de 2017 que terminó el 31 de Mayo de 2022 (<https://cordis.europa.eu/project/id/737390>); y el Proyecto Nacional IMPaCT-Data (Exp. IMP/00019).

Durante el tiempo de trabajo en esta Tesis Doctoral se realizó una **Estancia de Investigación** durante tres meses (de Enero a Abril de 2023) en un centro de investigación extranjero en Zúrich (Suiza), gracias a la concesión de una **Ayuda para la Movilidad de Personal Investigador Contratado** en el marco de la AES (**M-AES**) para trabajar en el laboratorio dirigido por el **Dr. Cesar Nombela Arrieta** (*Principal Investigator and Assistan Professor in UZH in Microenvironmental Regulation of Hematopoiesis Laboratory*, <https://www.nombelaarrieta.com/>) del **Departamento de Oncohematología** de la **Universidad y Hospital Universitario de Zúrich**. Como se indica, para la realización de dicha estancia se obtuvo de modo competitivo una Beca Internacional de la M-AES.

Finalmente, con los méritos anteriores y basado en el trabajo de investigación realizado en los últimos años, esta Tesis Doctoral opta a la **Mención de Doctorado Internacional** otorgado por parte de la **Universidad de Salamanca**, y por ello se presenta escrita en **inglés** en su totalidad, adjuntando también un **resumen en castellano**.

A mi familia

*Volverán las oscuras golondrinas
en tu balcón sus nidos a colgar,
y otra vez con el ala a sus cristales
jugando llamarán.*

*Pero aquellas que el vuelo refrenaban
tu hermosura y mi dicha a contemplar,
aquellas que aprendieron nuestros nombres...
ésas... ¡no volverán!*

*Volverán las tupidas madreselvas
de tu jardín las tapias a escalar
y otra vez a la tarde aún más hermosas
sus flores se abrirán.*

*Pero aquellas cuajadas de rocío
cuyas gotas mirábamos temblar
y caer como lágrimas del día...
ésas... ¡no volverán!*

*Volverán del amor en tus oídos
las palabras ardientes a sonar,
tu corazón de su profundo sueño
tal vez despertará.*

*Pero mudo y absorto y de rodillas
como se adora a Dios ante su altar,
como yo te he querido..., desengáñate,
así... ¡no te querrán!*

Gustavo Adolfo Bécquer

INDEX

LIST OF ABBREVIATIONS	1
INDEX OF FIGURES	5
INDEX OF TABLES	11
OBJECTIVES	13
INTRODUCTION	19
1. MESENCHYMAL STEM CELLS	21
a. General concepts about Mesenchymal Stem Cells	21
b. Mesenchymal Stem Cells obtained from different tissue sources	22
c. Mesenchymal Stem Cells in the bone marrow niche	23
d. Hematopoietic Stem Cells, hematopoiesis and immune response	25
e. Most abundant stromal cells: Fibroblasts	25
f. The gene regulatory system of Mesenchymal Stem Cells	26
g. Immunomodulation of Mesenchymal Stem Cells and Toll-Like receptors	28
2. CELLULAR GENE EXPRESSION AND TRANSCRIPTOMICS	33
a. Genes activation: the process of gene expression and transcription	33
b. Genome-wide expression measurements: Microarrays and RNA-seq techniques	34
3. CHARACTERIZATION OF MESENCHYMAL STEM CELLS: GENES AND CELLS	43
a. Novel markers of Mesenchymal Stem Cells	43
b. Cellular heterogeneity: different populations or states of Mesenchymal Stem Cells	44
 CHAPTER 1: Comparative analysis of expression profiles of human MSCs to identify novel and specific gene markers	47
1. MATERIAL AND METHODS	49
a. Identification of gene datasets	49
b. Gene markers selection	51
c. <i>Global Test</i> analysis	52
d. Confusion Matrices test	52
e. Human cell samples to test the markers	53
f. Cell processing	54
g. Quantitative gene expression measurements	55
2. RESULTS	57
a. Quality Control of the datasets	57
b. Expression signal of data sets in 151 genes	58
c. <i>Global Test</i> of the 151 MSCs candidate markers	60
d. <i>Global Test</i> of selected MSCs specific markers	65
e. Sensitivity, Specificity, FDR and Confusion Matrices of the selected MSCs markers	67

f. Experimental validation of the gene markers using RT-qPCR	70
3. DISCUSSION	73
4. FINAL SUMMARY of CHAPTER 1: <i>Specific gene markers for MSCs</i>	77
CHAPTER 2: Construction of gene regulatory networks and identification	
of master regulators of MSCs using transcriptomic expression data	79
1. MATERIAL AND METHODS	81
a. Integration of transcriptomic datasets obtained	
with high-density oligo Microarrays.....	81
b. Differential Expression Analysis	82
c. ARACNe algorithm to reconstruct gene regulatory networks.....	82
d. VIPER (Virtual Inference of Protein-activity	
by Enriched Regulon analysis) algorithm: selection of	
regulation genes and co-regulated networks.....	83
e. Enrichment Analysis to determine the function of the gene signature	84
f. Methylation Analysis	84
2. RESULTS.....	85
a. Transcriptomic profile of regulation genes in MSCs.....	85
b. Master Regulators of MSCs	86
c. Differential and Functional identification of regulation	
gene signature in MSCs	87
d. Methylation study of the MSCs Master Regulators	90
e. Gene expression co-regulation networks in Master Regulators	92
3. DISCUSSION	95
4. FINAL SUMMARY of CHAPTER 2: <i>Master regulators of MSCs</i>	97
CHAPTER 3: Analysis of the proinflammatory or immunosuppressive	
effect of MSCs through stimulation	99
1. MATERIAL AND METHODS	101
a. Characterization of MSCs	101
b. TLR4 and TLR3 stimulated RNA-seq datasets	101
c. <i>In vitro</i> Cell Cultures	102
d. TLR3 stimulation of MSCs with poly(I:C)	103
e. RNA-sequencing Technique.....	103
f. Differential Expression Analysis	103
g. Functional Enrichment Analysis	103
h. Differentiation assays of MSCs: Adipogenic differentiation.....	104
i. Differentiation assays of MSCs: Osteogenic differentiation.....	104
2. RESULTS.....	105
a. Quality control of stimulated MSCs with LPS and poly(I:C).....	105
b. Differential and Functional Analysis of LPS stimulated MSCs	106
c. Differential and Functional Analysis of poly(I:C) stimulated MSCs	109
d. Stimulation of MSCs with poly(I:C):	
short treatment 6h <i>versus</i> longer treatment 24h	111
e. Differentiation in control and stimulated MSCs.....	121
3. DISCUSSION	129
4. FINAL SUMMARY of CHAPTER 3: <i>Immunomodulatory effect of MSCs</i>	133

CHAPTER 4: Analysis of BM-MSCs isolated <i>in vitro</i> and <i>in vivo</i>	
using single-cell transcriptomics.....	135
1. MATERIAL AND METHODS	137
a. Stimulation of MSCs at 6h with poly(I:C).....	137
b. Single-cell RNA sequencing Technique.....	137
c. scRNA-seq data preprocessing including demultiplexing.....	138
d. Analysis of scRNA-seq data using algorithm <i>Seurat</i>	139
e. Functional Enrichment Analysis of the markers found for different cell populations	139
2. RESULTS.....	141
2.1 Single-Cell analysis of stimulated poly(I:C) and control MSCs	141
a. Quality control of Mesenchymal Stem Cells in scRNA-seq.....	141
b. Global Analysis of stimulated MSCs <i>versus</i> Control MSCs in scRNA-seq.....	142
c. Identification of subpopulations in control MSCs	146
d. Identification of subpopulations in stimulated poly(I:C) MSCs	150
e. New Identification of subpopulations in control and stimulated MSCs.....	155
2.2 Comparison of <i>in vivo</i> an <i>in vitro</i> Mesenchymal stem Cells	157
3. DISCUSSION.....	163
4. FINAL SUMMARY of CHAPTER 4: <i>Single-Cell transcriptomics of MSCs</i>	167
CONCLUSIONS	169
BIBLIOGRAPHY	173
LIST OF PUBLICATIONS	191
ACKNOWLEDGMENTS	193
ANNEX I: SUPPLEMENTARY TABLES	195
ANNEX II: RESUMEN EN CASTELLANO	199
ANNEX III: ARTICLE CHAPTER 2	209

LIST OF ABBREVIATIONS

Abbreviations	Meaning/Description
AD-MSC	Adipose-derived Mesenchymal Stem Cell
ADIP	Adipoblast
ALCAM	Activated-Leukocyte Cell Adhesion Molecule (CD166)
ARACNe	Algorithm for the Reconstruction of Accurate Cellular Networks
AT	Adipose Tissue
BCL	Binary Base Call
b-FGF	basic Fibroblast Growth Factor
BM-MSC	Bone Marrow Mesenchymal Stem Cell
BTN	Butyrophilin
CAR	CXCL12 Abundant Reticular
CFU-F	Colony-Forming Unit Fibroblast
CHON	Chondroblast
CLEC11A	osteogenic factor ostelectin
CMO	Cell Multiplexing Oligos
COL4A1	Collagen Type IV Alpha 1 Chain
COL4A2	Collagen Type IV Alpha 2 Chain
COX2	Cyclooxygenase-2
cpm	counts per million
CTL	Control
CTSC	Cathepsin
CXCL10	C-X-C motif Chemokine Ligand 10
CXCL12	C-X-C motif chemokine ligand 12
CXCL8	C-X-C motif Chemokine Ligand 8
dADIP	Adipoblasts derived from MSCs
DAMPS	Damage-Associated Molecular Patterns
DAVID	Database for Annotation, Visualization, and Integrated Discovery
dCHON	Chondroblasts derived from MSCs
Df	Degrees of freedom
diff	difference in means
DMEM	Dulbecco's Modified Eagle's medium-low glucose
dNTP	deoxynucleoside triphosphate nucleotides
dOST	MSC-derived Osteoblasts
dOSTB	Osteoblasts derived from MSCs
DPP4	Dipeptidyl Peptidase 4
EBF1	Early B cell Factor
ECM	Extracellular Matrix
EMT	Epithelial-Mesenchymal Transition
ENG	Endoglin (CD105)
FACS	flow-activated cell sorting
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration

FDR	False Discovery Rate
FGS	Fibroblast Growth Supplement
FIB	Fibroblast
FN	False Negative
FN1	Fibronectin
FNR	False Negative Rate
FOXO1	Forkhead Box O1
FOXP3	Forkhead Box P3
FP	False Positive
FPR	False Positive Rate
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GEM	Gel bead Emulsion
GEO	Gene Expression Omnibus
GO	Gene Ontology
GTF	General Transfer Format
GVHD	Graft-versus-host disease
H	Entropy of an arbitrary variable
HBSS	Hank's Balanced Salt Solution
HIF- α	Hypoxia-Inducible Factor-1 α
hMSCs	Human Mesenchymal Stem Cells
HPC	High Performance Computing
HSC	Hematopoyetic Stem Cell
HSPC	Hematopoietic Stem Progenitor Cells
HUS	University Hospital of Salamanca
HVG	Highly variable genes
I	Entropy
ID1	Inhibitor of Differentiation 1
IFIT	Interferon-Related Genes
Ig	Immunoglobulin
IGF-1	Insulin-like Growth Factor 1
IL1	Toll-Interleukin 1
IL7	interleukin-7
iPSC	induced Pluripotent Stem Cells
ISCT	International Society for Cellular Therapy
ITGA1	Integrin subunit alpha 1 (CD49a)
lncRNA	Non-coding Ribonucleic acid
log(FC)	Logarithm Fold Change
log ₂ (FC)	Log ₂ fold change
LPS	Lipopolysaccharide
LYM	Lymphocytes
MAPK	Mitogen-Activated Protein Kinase
MCP-1	Monocyte Chemoattractant Protein-1
MI	Mutual Information
MIF	Migration Inhibitory Factor
MLPH	Melanophilin

MME	Membrane Metalloendopeptidase
MMP1	Matrix Metallopeptidase 1
MNC	Mononuclear Cells
MR	Master Regulator
MS	multiple sclerosis
MSC	Mesenchymal Stem Cell
MyD88	Myeloid Differentiation primary response gene 88
myo	Myogenic cells
N	Negative
NCBI	National Center for Biotechnology Information
NES	Normalized Enrichment Score
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next Generation Sequencing
NK	Natural Killer
NT5E	5'-nucleotidase ecto (CD73)
OAS	Oligoadenylate Synthetase
OCT4	Octamer-binding Transcription
OSTB	Osteoblast
osteo	Osteogenic cells
P	Positive
PB	Peripheral Blood
PBMNCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PDGFab	Platelet-Derived Growth Factor ab
PI3K	Phosphoinositide 3-Kinases
PI3K/Akt	Protein Kinase B/AKT
PL-MSC	Placenta derived Mesenchymal Stem Cell
Poly(I:C)	Polyinosinic:polycytidylic acid
PPARG	Peroxisome Proliferation-Activated Receptor γ
PRDK1	Protein Kinase D1
QC	Quality Control
qPCR	quantitative Polymerase Chain Reaction
RMA	Robust Multi-Array Average expression measure
RNAseq	RNA sequencing
RT-qPCR	Real-time quantitative Polymerase Chain Reaction
RUNX2	Runt-related transcription factor 2
SC	Stromal Cell
scRNA-seq	single cell RNA sequencing
SCUBE3	Signal Peptide, CUB Domain & EGF Like Domain Containing 3
SD	Standard Deviation
SLE	Lupus erythematosus
SNP	Single Nucleotide Polymorphism
SOX2	SRY box-containing factor 2

SOX9	SRY-related high mobility group-box gene 9
SSEA-4	Stage-Specific Embryonic Antigen-4
stMSCs	Stimulated Mesenchymal Stem Cells
stOST	Stimulated Osteoblasts
Sum sq	Sum of squares
t-SNE	T-distributed Stochastic Neighbour Embedding
TAGLN	Transgelin
TF	Transcription Factor
TFBS	Transcription Factors Binding Sites
TGF- β	Transforming Growth Factor β
THY1	Thy-1 Cell Surface antigen (CD90)
TIR	Toll-Interleukin 1 Receptor
TIRAP	TIR- domain-containing adapter protein)
TLR	Toll-Like Receptor
TLR3	Toll-Like Receptor 3
TLR4	Toll-Like Receptor 4
TN	True Negative
TNF	Tumor Necrosis Factor
TNFAIP3	TNF Alpha Induced Protein 3
TNR	True Negative Rate
TP	True Positive
TPM	Tropomyosin
TPR	True Positive Rate
TRAM	TRIF-related Adaptor Molecule
Tregs	Regulatory T Cells
TRIF	TIR- domain-containing adapter protein
tRNA	transfer Ribonucleic acid
UC	Umbilical Cord
UCB	Umbilical Cord Blood
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier
VEGF	Vascular Endothelial Growth Factor
VIPER	Virtual Inference of Protein-activity by Enriched Regulon analysis
ZEB1	Zinc finger E-box Binding homeobox 1
ZEB2	Zinc finger E-box Binding homeobox 2
ZNF145	Zinc-finger protein 145

*All Human genes described in this Doctoral Thesis have been written in capital letters.

INDEX OF FIGURES

Figure 1. Multipotency of Bone Marrow MSCs. (From Uccelli et al., 2008).	22
Figure 2. Schematic view of MSCs in hematopoietic niches within the bone marrow. CAR cells are CXCL12-abundant reticular cells. (From Reagan & Rosen, 2016).	24
Figure 3. Human TLRs Signaling, including the pathways of TLR3 and TLR4. (From R & D systems, 2012 (https://www.rndsystems.com/)).	29
Figure 4. Diagram showing the principle of MSCs cell stimulation using poly(I:C) at short time of 4 hours and longer time of 24 hours. (From Petri et al., 2017).	31
Figure 5. Schematic representation of the cellular gene expression process (From National Human Genome Research Institute. (https://www.genome.gov/)).	33
Figure 6. Microarray technique. (From Matheny M et al. 2011).	35
Figure 7. Workflow of RNA sequencing technique. A) RNA is fragmented, converted to cDNA and the sequencing adapters are added to form the DNA libraries. B) The libraries are introduced in the flow cell for the amplification and cluster generation. C) dNTPs are introduced one by one, adhering to their complementary base, thus forming different complementary sequences. D) the raw data obtained is processed by bioinformatic tools to obtain the mapped reads. (From Dipartimento Rete Oncologica Piemonte e Valle d'Aosta, 2018 (http://www.reteoncologica.it/)).	36
Figure 8. Workflow bioinformatic analysis of RNA sequencing. (From Biocore RNAseq course, 2019. (https://github.com/biocorecrg/RNAseq_course_2019)).	38
Figure 9. Types of RNA sequencing alignment. (From Haas & Zody, 2010).	39
Figure 10. Droplet-based microfluidic technique. A) Cells are encapsulated in a droplet, cells are lysed and mRNA is captured by uniquely barcoded beads. B) mRNAs are captured by oligos (conjugated in beads), UMIs and other adapter sequences. In this process cDNA is synthesized and perform the library generation. Then, library modification (via template switching) is done to allows the library amplification. (From H. Li & Humphreys, 2021).	40
Figure 11. Schematic representation of a sc-RNAseq data analysis workflow from raw reads to cell clustering and gene marker identification (From Mary P, et all 2022).	41
Figure 12. Density and PCA 3D plots of three types of datasets. A) The 264 samples of Meta-Analysis from Microarrays data with 10 different cell types. B) The 15 samples of Exon arrays in 5 different cell types. C) the 29 samples of RNA-seq with 3 different cell types.	58
Figure 13. Boxplot of the 151 gene markers in the Exon array data set with 3 samples of BM- MSC, 3 samples of AD- MSC, 3 samples of PL- MSC, 3 of HSC and 3 of FIB. The scale of the gene expression was in logarithm 2.	59

Figure 14. The subjects plot of the cell types: MSCs, HSCs and FIBs. **A)** Plot of the dataset 1 corresponding with the Meta-analysis of the microarrays. **B)** Plot of the dataset 2 corresponding with the exon arrays. **C)** Plot of dataset 3 corresponding with the RNA-sequencing data. 63

Figure 15. The covariates plot of 151 genes associated to the cell types: MSCs, HSCs and FIBs. **A)** Plot of the dataset 1 corresponding with the Meta-analysis of the microarrays. **B)** Plot of the dataset 2 corresponding with the exon arrays. **C)** Plot of dataset 3 corresponding with the RNA-sequencing data. 64

Figure 16. Subjects plots of Global test. **A, B and C)** for the standard genes from ISCT: ENG, THY1 and NT5E. **D, E and F)** proposal genes of the study: TAGLN, SCUBE3, COL4A1 and COL4A2, all are done for Metanalysis of microarrays, exon arrays and RNA-seq data, respectively. 66

Figure 17. RT-qPCRs boxplots of genes: ALCAM(CD166), COL4A1, COL4A2, ENG (CD105), ITGA1 (CD49a), NT5E (CD73), TAGLN, THY1 (CD90) SCUBE3 from BM-MSC, AD-MSC, FIB, mononuclear cells (MNC) and two stromal cell lines (HTERT and HY5) cell types. The housekeeping used was GAPDH. 71

Figure 18. ARACNe analysis workflow. **A)** Mutual information (MI) Diagram. **B)** Elimination of false interactions on genes by the ARACNe algorithm. **C)** Steps developed in the ARACNe application for the obtaining a coexpression matrix. (From Margolin et al., 2006). 83

Figure 19. Heatmap of 188 genes from the differential expression analysis (joining top 30 genes of each pair-wise comparison: MSC-HSC, MSC-LYM, MSC-FIB, MSC-OSTB, MSC-stMSC, and stMSC-HSC). 86

Figure 20. TFs and associated regulons found using VIPER. **A)** Table of the number of TF up-down in each of the 6 comparisons, the results with VIPER, VIPER-bootstrap and VIPER-pleiotropy. **B)** Table with top 10 up and top 10 down regulated TFs in MSCs-HSCs including the parameters from VIPER: NES, p-value, FDR and pleiotropy. **C)** Table with top 10 up (in red) and top 10 down (in blue) regulated TFs in MSCs-HSCs, with the activity and the expression on the right, and the bars as the number of TFs that configure the regulons. . 88

Figure 21. DNA normalized methylation boxplots of 12 MR in 3 independent plots of BM- MSC and 3 independent plots of HSC. Genes EPAS1, NFE2L1, SATB2, SANI2, TEAD2 and TULP3 in upregulation of MSC and genes ERG, GATA2, GATA3, HLP, MYB and POU2F1 upregulated in HSC. **A and B)** 12 MR up and down regulated in GSE79695 vs. GSE63409. **C and D)** 12 MR up and down regulated in GSE129266 vs. GSE63409. **E and F)** 12 MR up and down regulated in GSE87797 vs. GSE63409. 91

Figure 22. Gene correlation network with top 10 upregulated MR in MSC in red and top 10 downregulated MR in MSC in blue. The circles are the gene regulons associated, which are in red are upregulated and in blue are downregulated. 93

Figure 23. Gene correlation network with top 10 upregulated MR in MSC indicated as red rectangles: SNAI2, TEAD1, STAB2, IRX3, EPAS1, HOXC6, TULP3, TWIS1, PRRX1 and NFE2L1 with their associated regulons (in red which are upregulated and in blue the downregulated

ones). The yellow ellipses are 6 promoter TFs E2F1, EP300, GADD45A, MAFK, TCF12, and TEAD4 enriched with the *iRegulon* tool..... 94

Figure 24. Schematic representation of the isolation and expansion of MSCs. 102

Figure 25. Quality Control analysis of 21 samples from 3 datasets: GSE1478, GSE7723, GSE109181. **A)** Density Plot. **B)** PCA plot in two dimensions..... 105

Figure 26. Venn Diagram of the interaction between MSCs stimulated with LPS 1 µg/ml vs. CTL and MSCs stimulated with LPS 10 ng/ml vs. CTL. The values associated correspond to the result with $\log(FC) \geq 1$ of the 4 comparisons of *Limma-Voom* Differential analysis.... 106

Figure 27. Heatmap of MSCs: CTLs of LPS, LPS 10ng/ml and 1 µg/ml in 140 common genes in LPS stimulation. For each sample there were 3 technical replicates. The results were scaled in logarithm (cpm+1). 108

Figure 28. Heatmap of MSCs: CTLs of Poly(I:C) vs. Poly(I:C) 10µg/ml in 207 common genes in poly(I:C) stimulation. For each of 2 sample there were 3 technical replicates. The results were scaled in logarithm (cpm+1). 110

Figure 29. Quality control analysis of the samples: s389-19, s592-21 and s98-19 in different conditions: CTL, Poly(I:C) at short time 6 hours and Poly(I:C) at longer time 24 hours. **A)** Boxplots representing the unnormalized samples and normalized samples. **B)** PCA plot of the different samples in different conditions. 111

Figure 30. *Limma-Voom* differential analysis. **A)** Comparison between different conditions of MSCs: poly(I:C) at time 6 hours vs. CTL, poly(I:C) at time 24 hours vs. CTL and poly(I:C) at time 24 hours vs. poly(I:C) at time 6 hours; with p-values of 0.01 and 0.05 and $\log(FC)$ of 1, 1.5 and 2 for all comparisons. **B)** Intersection between different conditions of MSCs: poly(I:C) at time 6 hours vs. CTL, poly(I:C) at time 24 hours vs. CTL and poly(I:C) at time 24 hours vs. poly(I:C) at time 6 hours in the same p-values and $\log(FC)$ as **Figure 30A**. **C)** Up and down genes of the genes of intersection in each comparison. **D)** Venn Diagram plots of the intersection between: poly(I:C) at time 6 hours vs. CTL, poly(I:C) at time 24 hours vs. CTL and poly(I:C) at time 24 hours vs. poly(I:C) at time 6 hours with p-value 0.05 and $\log(FC)$ 1.5.. 114

Figure 31. Volcano plots of the comparisons. **A)** stimulation of poly(I:C) at 6 hours vs. CTL **B)** stimulation of poly(I:C) at 24 hours vs. CTL. The blue spots are the differential down expressed genes, the red spots are the differential up expressed genes and the grey ones are the non-significant genes. 116

Figure 32. Heatmaps and expression pattern plots of the differential *Limma-Voom* analysis with p-value= 0.05 and $\log(FC)>1.5$. **A)** Differential analysis of the contrast stimulated MSCs at time 6h vs CTL. **B)** Differential analysis of the contrast stimulated MSCs at time 24h vs CTL. **C)** Differential analysis of the contrast stimulated MSCs at time 24h vs stimulated MSCs at time 6h..... 118

Figure 33. Network and dotplot of enrichment analysis. A) genes correlated with the functions associated with the contrast of stimulated MSCs with poly(I:C) at time 6h vs. CTL. B) genes correlated with the functions associated with the contrast of stimulated MSCs with poly(I:C) at time 24h vs. CTL.	120
Figure 34. Analysis of differentiation genes in MSCs CTL, stimulated MSCs with poly(I:C) at 6h and stimulated MSCs with poly(I:C) at 24h. A) Boxplots of general gene markers of differentiation and osteogenic gene markers. B) Boxplots of adipogenic gene markers. C) Table of the significant genes of the Figure 34A	122
Figure 35. Control MSCs, stimulated MSCs with poly(I:C) in time 6h and stimulated MSCs with poly(I:C) in time 24h of sample s98-19 after 21 days of adipocyte differentiation. The pictures were taken in Optical Microscope at resolution 10x and 20x.....	123
Figure 36. Statistical analysis of adipocyte differentiation of MSCs. A) Boxplot of the 4 samples in each condition in function of the 7, 14 or 21 days of adipocyte differentiation. B) ANOVA and post-Hoc statistical analysis of the samples.....	124
Figure 37. Control MSCs, stimulated MSCs with poly(I:C) in time 6h and stimulated MSCs with poly(I:C) in time 24h of sample 98-19 after 21 days with (osteoblasts) and without (MSCs) osteogenic differentiation medium. The pictures were taken in Optical Microscope at resolution 10x and 20x.	126
Figure 38. Statistical analysis of osteogenic differentiation of MSCs. A) Boxplot of the 4 samples in MSCs without and with osteogenic differentiation medium. B) ANOVA and post-Hoc statistical analysis of the samples.....	127
Figure 39. Workflow of single cell analysis downstream.....	138
Figure 40. UMAP clustering plots. A) UMAP plot corresponding with the two conditions: Control MSCs and stimulated poly(I:C) MSCs. B) UMAP plot of clusters formed using Seurat package.....	142
Figure 41. Heatmap of the top 50 differenced marker genes in each of 6 clusters of both MSC conditions.	143
Figure 42. Enrichment analysis of unstimulated and stimulated MSCs. A) Barplot of the main functions of each cluster of Figure 40B . B) Network of the main functions of the most differentiated genes of the top 50 differential genes.	145
Figure 43. UMAP plot clustering of the control MSCs. The plot was analyzed in two samples grouped by clustering: 0, 1, 2.	146
Figure 44. Heatmaps of the top 50 genes in the Control MSCs. A) Heatmap of top 50 genes in 3 clusters of control MSCs. B) Heatmap of top 50 genes represented in the stimulated and control MSCs.....	147

Figure 45. Violin plots associated with the most differential expressed genes in control MSCs. **A)** Most differentiated genes of Cluster 1 in 3 clusters of Control analysis. **B)** Most differentiated genes of Cluster 1 in 6 clusters of Global analysis. **C)** Most differentiated genes of Cluster 2 in 3 clusters of Control analysis. **D)** Most differentiated genes of Cluster 2 in 6 clusters of Global analysis. 148

Figure 46. Barplots of enrichment analysis associated with the three cluster of control MSCs differential analysis. 149

Figure 47. UMAP clustering of stimulated MSCs cells. The plot was analyzed in two samples grouped by clustering: 0, 1, 2, 3 and 4. 150

Figure 48. Heatmaps of the top 50 genes in the stimulated MSCs. **A)** Heatmap of top 50 genes in 35clusters of stimulated MSCs. **B)** Heatmap of top 50 genes represented in the stimulated and control MSCs. 151

Figure 49. Violin plots associated with the most differential expressed genes in stimulated MSCs. **A and B)** Most differentiated genes of Cluster 0 in control and Global analysis. **C and D)** Most differentiated genes of Cluster 1 in control and Global analysis. **E and F)** Most differentiated genes of Cluster 2 in control and Global analysis. **G and H)** Most differentiated genes of Cluster 3 in control and Global analysis. **I and J)** Most differentiated genes of Cluster 4 in control and Global analysis. 153

Figure 50. Barplots of enrichment analysis associated with clusters 0, 1, 2 and 4 of stimulated MSCs differential analysis. 154

Figure 51. UMAP plots of unstimulated and stimulated MSCs. **A)** UMAP plot of Seurat Clusters. **B)** UMAP plot in function of control and poly(I:C) stimulation of the two patients. **C)** UMAP plot of new subpopulations established. 156

Figure 52. UMAP plot of *in vivo* MSCs from healthy donors. **A)** UMAP plot of CAR cells (CAR), osteogenic cells (osteo) and myogenic cells (myo) from *in vivo* MSCs from healthy donors. **B)** UMAP plot of Seurat clusters in the three groups. 157

Figure 53. UMAP plots of *in vitro* and *in vivo* MSCs integration. **A)** UMAP plot of the different groups of MSCs: CAR1, CAR2, Myo, Osteo and MSC invitro. **B)** UMAP plot of Seurat clusters. 158

Figure 54. Enrichment analysis of Cluster 1 and 3. **A)** Barplot of the correlation of functions. **B)** Dot Plot of the significative functions. **C)** Network plot of the genes associated with the functions. 160

Figure 55. Enrichment analysis of Cluster 7. **A)** Barplot of the correlation of functions. **B)** Dot Plot of the significative functions. **C)** Network plot of the genes associated with the functions. 161

INDEX OF TABLES

Table 1. Meta-Analysis of 264 Microarrays from 10 cell types: HSCs, LYM, MSCs, stMSCs, ADIP, CHON, OSTB, dOST, stOST and FIB.....	50
Table 2. RNA-seq of 29 samples from 3 different cell types: MSC, HSC and FIB. The samples are provided from 8 different datasets as indicate the batch and the GSE. The samples have been sequencing with Illumina Hiseq in different versions, as indicate the column of the sequencing and have different types of sequencing, being: single- or paired-end.	51
Table 3. Characteristics of the human samples of different cell types that are used to test the expression of genes markers (the samples are coming from healthy donors, D1-D14, or from cell lines, L14 & L15).	54
Table 4. Human genes tested by RT-PCR to quantify expression in different human cells. The Catalog number and the Entrez NCBI ID are provided to check gene identity and sequence information.	55
Table 5. <i>Global test</i> statistic of the selected covariates (genes): TAGLN, COL4A2, COL4A1, SCUBE3, standard genes: NT5E, THY1, ENG and hematopoietic genes: CD34, PTPRC (CD45). The parameters employed were: the weighted test statistic, the mean value of the gene in all samples, de deviation standard (SD), the number of deviations standard from the mean for datasets of Microarrays, Exon Arrays and RNA-seq data.	62
Table 6. Confusion matrices of the proposed single genes (TAGLN, COL4A1, COL4A2 and SCUBE3) and standard single genes (ENG, THY1 and NT5E). For each one was calculated: miss rate, Fall-out, sensitivity, specificity, precision and FDR.	67
Table 7. Confusion matrices of the two best proposal and standard gene; and the second table represent the combination of all proposed genes (TAGLN, COL4A1, COL4A2 and SCUBE3) and standard genes (ENG, THY1 and NT5E): For each table was calculated: miss rate, Fall-out, sensitivity, specificity, precision and FDR.	69
Table 8. Number of samples of different human cell types obtained from the GEO datasets indicated above (GSE) that were collected, integrated and analyzed in this study.	81
Table 9. Differential Expression table of Top 20 MR by <i>VIPER</i> analysis in the 6 comparisons analysed in Figure 18 . Table represents the Fold Change logarithm of each comparison, the mean expression of the gene and the FDR in the contrast of MSC-HSC.....	89
Table 10. Functional enrichment analysis done with the top 10 master regulators up-regulated (UP) and their corresponding gene regulons and with the top 10 master regulators down-regulated (DOWN) and their corresponding gene regulons.....	90
Table 11. RNA-seq datasets with the samples associated, the labeled patients, the stimulation being with LPS, poly(I:C) or control (CTL), the platform where the data was analyzed and the type of sequencing: paired or single, and the corresponding articles. .	101

Table 12. Differential analysis of MSC stimulated with Poly(I:C) 10µg/ml vs. CLT of MSCs in two patients with 3 technical replicates each one. The table represent the differential genes in without log(FC), $\log(\text{FC}) \geq 0.5$; $\log(\text{FC}) \geq 1$ and $\log(\text{FC}) \geq 2$. Also are represented the up and down expressed values. 109

Table 13. Summary table of numeric matrix of expression values on the Logarithm scale of differential analysis with *Limma-Voom* in the contrast of stimulated MSCs samples of poly(I:C) in 6 hours vs. MSCs CTLs and stimulated MSCs samples of poly(I:C) in 24 hours vs. MSCs CTLs..... 112

Table 14. Distribution of samples (cells) in function of the two patients in raw matrix, and filtered matrix after quality control. 141

OBJECTIVES

This Doctoral Thesis entitled "**Experimental and bioinformatic characterization of the transcriptomic profile of human Mesenchymal Stem Cells (MSCs)**", focuses on the study and characterization of two fundamental aspects of the Mesenchymal Stem Cells isolated from adult human tissues: (i) **Transcriptomic characterization**: the first aspect corresponds to the gene expression, transcriptomic profiling and gene regulatory mechanisms of the MSCs focusing on the identification of distinct specific markers that characterize these cells; and (ii) **Immunomodulatory analysis**: the second aspect corresponds to the analysis of the immunomodulatory properties of the MSCs under different stimulation conditions (focusing on TLR3 and TLR4 stimulation), that provide different phenotypes and promote distinct effects within their cellular niche.

Mesenchymal Stem Cells are a multipotent adult stem cells found in various human tissues, including bone marrow, adipose tissue, and dental pulp, among others. MSCs are characterized by their ability to differentiate into a variety of cell types, including osteoblasts, chondrocytes, and adipocytes, making them very valuable for regenerative medicine and tissue repair applications. While the immunomodulatory effects of the MSCs have gained recent attention, many facets of their molecular properties, functions, and cellular heterogeneity remain unresolved. To address these issues, pioneer high throughput techniques such as full transcriptome RNA-sequencing and single-cell RNA-sequencing are employed to achieve more precise characterization, facilitating the identification of novel biomarkers and their associated functions. Following these ideas, this study aims to provide an in-depth and detailed transcriptomic analysis of the human MSCs that sheds light on their molecular and cellular characteristics.

Main objectives:

Within this thematic framework, this Doctoral Thesis uses several experimental and bioinformatic methods to fulfill the main objectives. The dissertation has been written and is presented into **four chapters** arranged to address **four main specific objectives** of our scientific work:

OBJECTIVE 1: Determination of the transcriptomic signature of human MSCs to identify precise novel gene markers that distinguish them from other related cell types.

- a. Integration of a large compendium of genome-wide expression data (transcriptomic profiles) of MSCs samples from different tissue origins, alongside primary fibroblasts and hematopoietic stem cells using different transcriptomic techniques: Microarrays, Exon Arrays and RNA-seq.
- b. Analysis of the collected transcriptomic data, employing advanced bioinformatic and computational statistical methods to identify the optimal marker genes that exhibit the most significant differences between cell types and selecting new MSC-specific gene markers. Machine learning methods will be used to assess the accuracy and precision of the novel markers compared to the markers commonly used to identify MSCs.
- c. Experimental validation using real-time quantitative PCR (RT-qPCR) of the new gene markers found on the different cell types tested.

This Objective, including Material and Methods, as well as Results and Discussion, is presented in **Chapter 1** of this dissertation.

OBJECTIVE 2: Integration of transcriptomic data from different hematopoietic and non-hematopoietic cells types and construction of gene regulatory networks to identify the MSC master regulators.

- a. Collection and integration of a compendium of transcriptomic data from different human cell types associated to the bone marrow, including hematopoietic stem/progenitor cells, lymphocytes (LYM), mesenchymal stem cells, fibroblasts, osteoblasts (OSTB), and stimulated MSCs (stMSC).
- b. Generation of the gene signatures derived from the analysis of differential expression of MSCs compared to the different cell types described, followed by functional enrichment analysis of these signatures.
- c. Use of bioinformatic algorithms to construct gene regulatory networks and identify the master regulators of MSCs (i.e., the key transcription factors, TFs, that regulate the expression profile of MSCs). Combined analysis of transcriptomic expression data and methylation data to corroborate the activity of the TFs.

This Objective, including Material and Methods, Results and Discussion, is presented in **Chapter 2** of this dissertation. Moreover, this work was published as a research article in the journal *Biomolecules* in 2020 ([doi: 10.3390/biom10040557](https://doi.org/10.3390/biom10040557)). The article is included at the end of this Thesis.

OBJECTIVE 3: Transcriptomic analysis of the immunomodulatory properties of MSCs stimulated with TLR3 and TLR4, along with exploration of the time-dependent effects of TLR3 stimulation.

- a. Collection of RNA-seq data of the stimulation of MSCs with Lipopolysaccharide, LPS, and with polyInosinic-polyCytidylic acid, poly(I:C), followed by differential expression analysis and functional enrichment using different bioinformatic methods.
- b. Stimulation of human bone marrow MSCs from healthy donors with poly(I:C) at different time points (i.e., with short and longer treatment) to profile gene expression patterns and functions using full transcriptomic RNA-sequencing technology.
- c. Study of the differences in the osteogenic and adipogenic differentiation of MSCs through experimental and statistical analysis.

This Objective, including Material and Methods, Results and Discussion, is presented in **Chapter 3** of this dissertation.

OBJECTIVE 4: Identification and analysis of *in vitro* naive and stimulated MSCs subpopulations and comparison between MSCs *in vitro* and *in vivo* using single-cell expression technology.

- a. Stimulation of human bone marrow MSCs from healthy donors treated with poly(I:C) for a short time (6 hours). Single-cell RNA-seq (scRNA-seq) technology is applied to the comparative study of the 2 conditions: control MSCs and poly(I:C) stimulated MSCs.
- b. Implementation of a bioinformatics pipeline for the analysis of single-cell MSC data, including clustering analysis of different cell subsets in the MSC samples under each condition. Identification of unique gene patterns and biological functions in the cell groups found.
- c. Comparison of cell populations and gene signatures obtained from scRNA-seq of bone marrow MSCs derived different healthy donors, obtained *in vivo* or *in vitro*. Identification of shared clusters between both conditions and establishment of the different gene-function associations within each subgroup.

This Objective, including Material and Methods, Results and Discussion, is presented in **Chapter 4** of this dissertation. The third point of this chapter represents collaborative research with the *Department of Oncohematology of the University Hospital of Zurich*, currently ongoing and not disclosed at the time of writing this Thesis, which requires confidentiality measures for certain results.

INTRODUCTION

1. MESENCHYMAL STEM CELLS

a. General concepts about Mesenchymal Stem Cells

Mesenchymal Stem Cells (MSCs) have been studied in biomedical research for the past several decades due to their value in many cell-based therapies. Currently, many studies focus on the use of these stem cells in cell therapy with multiple clinical applications in regenerative medicine due to their tissue repair and immunomodulatory capabilities. Thus, MSCs are being used as a tool to treat various inflammatory or autoimmune diseases (many of which cause degenerative changes in the joints), and to reconstruct bone and cartilage, or repair damaged musculoskeletal tissues (Petri et al., 2017). These uses allow their application in numerous clinical settings, such as: plastic surgery, hematopoietic stem cell transplantation, cardiovascular diseases, endocrine and nervous system diseases.

Friedenstein and colleagues were the first scientists to culture bone-forming cells in 1976 and described mesenchymal stem cells in the hematopoietic niche. These MSCs demonstrated a remarkable capacity for prolonged self-maintenance and differentiation into multiple mesenchymal cell lineages. Subsequently, in 1991, Caplan and colleagues published a paper reviewing the properties and characteristics of mesenchymal stem cells (Caplan, 1991). These cells exhibited unlimited mitotic division and the ability to differentiate into various cell types. This differentiation occurred/emerged in response to both paracrine regulation and autocrine signaling, considering that these cells have a pluripotent nature (Stefańska et al., 2020).

In 2006, the *International Society for Cellular Therapy* (ISCT, recently renamed the *International Society for Cell & Gene Therapy*, www.isctglobal.org) established a statement of minimal criteria to define MSCs, based on the following three conditions (Dominici et al., 2006):

1. MSC must be plastic adherent when maintained under standard culture conditions.
2. MSC must express the markers CD73 (NT5E), CD90 (THY1) and CD105 (ENG), and lack the expression of CD45 (PTPRC), CD34, CD14 or CD11B (ITGAM), CD79A or CD19 and HLA-DR. All the CDs are cell surface molecules.
3. MSC must have the ability to differentiate *in vitro* into these cell types: osteoblasts, adipocytes and chondroblasts.

Following these developments, numerous infusion procedures have been performed to apply MSCs due to their powerful stem capacity and good safety profile. Currently, more than 1233 MSC clinical trials are registered and listed in the database of the *Food and Drug Administration* (FDA, www.clinicaltrials.gov). The most common and longest used adult tissues origin for human MSCs are the bone marrow and the stromal vascular fraction of adipose tissue (Bianco et al., 2008).

These characteristics collectively define MSCs as prototypical adult stem cells with the capacity for self-renewal and differentiation into a wide range of tissues. MSCs have the capacity to differentiate into mesodermal and non-mesodermal lineages (see **Figure 1**). The endogenous role of MSCs is to maintain stem cell niches, particularly in the hematopoietic context, and they are also involved in organ homeostasis, wound healing, and successful aging (Williams & Hare, 2011).

However, it is important to note that while much of the literature on the clinical benefits of MSCs has been conducted with populations selected according to the ISCT criteria mentioned above, it has been recognized that these stem cell populations are not homogeneous, with individual cells or some subpopulations often exhibiting distinct behaviors. Therefore, it has been reported that some isolated *in vitro* MSCs comprise a heterogeneous population, making it difficult to identify their specific role in disease, especially when derived from different tissue sources or obtained with different production protocols (Wilson et al., 2019).

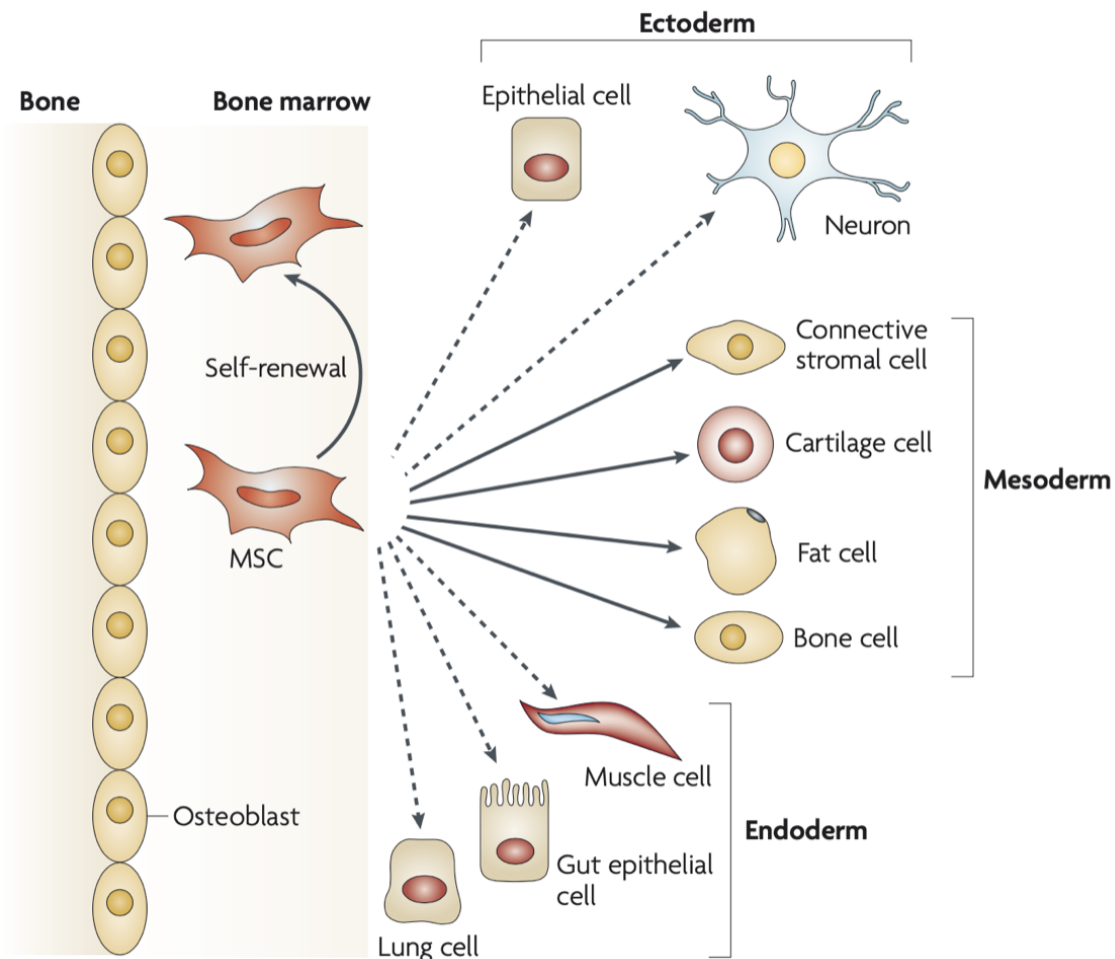


Figure 1. Multipotency of Bone Marrow MSCs. (From Uccelli et al., 2008)

b. Mesenchymal Stem Cells obtained from different tissue sources

While MSCs can hypothetically be derived from almost any tissue in the human body, the main sources of MSCs are bone marrow and adipose tissue. This is due to the challenges and invasiveness associated with the procurement process, compounded by the diverse biological characteristics of the source in the body. The recognized sources of MSCs discovered to date are (i) from adult tissues/organs: bone marrow, peripheral blood, adipose tissue, dental pulp, synovium and synovial fluid, endometrium, cervical tissue, lung, skin and muscle; and (ii) from fetal or perinatal tissues: amniotic fluid, placenta, umbilical cord and Wharton's jelly (Berebichez-Fridman & Montero-Olvera, 2018).

Bone marrow MSCs (BM-MSCs) were the first to be described as undifferentiated MSCs. They represent only between 0.01 and 0.001% of the nucleated bone marrow cells (proportion influenced by donor age). Obtaining BM-MSCs involves an invasive procedure to obtain bone marrow aspirates. Despite the difficulty in obtaining such biological samples, BM-MSCs are very useful in research due to the

existence of well standardized methods for BM samples, the ease of isolation of BM-MSCs by culturing on plastic and expansion, as well as their rapid growth (Pittenger, 2008). Regarding the use of BM-MSCs in cell therapies, several promising studies and clinical trials (especially on graft-versus-host disease, GVHD, a complication that occurs after allogeneic hematopoietic transplantation) highlight the key role of BM-MSCs in modulating inflammation. Additionally, BM-MSCs have also been used to treat Crohn's disease, systemic ischemic heart failure, lupus erythematosus, multiple sclerosis, autoimmune encephalomyelitis, liver cirrhosis, liver failure asthma, allergic rhinitis, COVID-19 or pulmonary fibrosis, among others (Berebichez-Fridman & Montero-Olvera, 2018; Javed et al., 2022; Li et al., 2006).

Adipose-derived MSCs (AD-MSCs) are abundant in our body but they are mostly obtained from therapeutic liposuction procedures, with over 98-100% of viability. The morphological, phenotypic and functional characteristics are similar to those of BM-MSCs, except for the major production of adipose stromal cells compared to the bone marrow fraction. In addition, AD-MSCs *in vitro* exhibit increased proliferative capacity, longer life-span, shorter doubling time and delayed senescence. Currently, AD-MSCs have several properties that are very useful for regenerative medicine, such as: their paracrine activity and enhanced promotion of wound healing. Furthermore, AD-MSCs have been used in various medical-therapeutic contexts, such as: wound healing, nerve regeneration, soft tissue reconstruction, as well as in craniofacial surgery, skin diseases and obesity treatment, as well as in the applications previously indicated for BM-MSCs. (Gentile et al., 2019; Semenova et al., 2018).

Placenta-derived MSCs (PL-MSCs) have gained importance in previous research performed in our laboratory. PL-MSCs show functional enrichment in mitosis, negative regulation of cell death and embryonic morphogenesis. In addition, PL-MSCs have higher growth rates in fetal cell culture and stronger associations with developmental processes compared to AD-MSCs and BM-MSCs. Furthermore, PL-MSCs express octamer-binding transcription factor-4 (OCT4, official gene symbol POU5F1) and stage-specific embryonic antigen-4 (SSEA-4), a glycosphingolipid, both of which are human markers characteristic of differentiated embryonic stem cells. This suggests that PL-MSCs are related to primitive embryonic stem cells as they retain greater differentiation potential than most adult stem cells (Roson-Burgo et al., 2014).

Despite the multiple sources of MSCs, the FDA has only approved treatments with stem cells derived from bone marrow aspirate concentrate. Therefore, while hematopoietic reconstitution with bone marrow or peripheral blood-derived MSCs is a known, safe and effective treatment; further research is needed on the safety and efficacy of other sources of MSCs (Berebichez-Fridman & Montero-Olvera, 2018).

c. Mesenchymal Stem Cells in the bone marrow niche

Mesenchymal Stem Cells have the ability to differentiate into mesodermal lineages, such as those that give rise to bone, fat, and cartilage tissues. Interestingly, they also possess the potential to differentiate into endodermal and neuroectodermal lineages, as illustrated in **Figure 1**.

Bone marrow-derived stromal cells represent a heterogeneous population, characterized by a complex transcriptome encoding a wide range of proteins involved in diverse developmental pathways and numerous biological processes. Despite substantial evidence indicating that MSCs can transdifferentiate into various cell types both *in vitro* and *in vivo*, their actual contribution to tissue repair, particularly in terms of significant engraftment and differentiation into biologically and functionally relevant tissue-specific cell types, remains uncertain.

In 1978, Schofield and colleagues published the first concept of the cell niche, outlining how microenvironments distinct from the spleen and bone marrow could induce hematopoiesis. The niche is defined as the microenvironment in which stem cells are present in an undifferentiated, self-renewing state and in which they interact with each other to maintain them or promote differentiation. The niche is a combination of quiescent states and proliferative activation of stem cells (Schofield, 1978). The niche is a fundamental unit of tissue physiology, orchestrating signals that mediate the balanced response of stem cells to the needs of the organism. Multiple factors are critical in regulating stem cell properties within the niche, including cell-cell interactions, oxygen tension, growth factors, adhesion molecules, extracellular matrix components, metabolites and ionic strength. Dysregulation between the proliferation and differentiation stages could lead to the development of different types of cancer (Scadden, 2006). This aspect is particularly important in the context of studying MSCs *in vitro*, considering the shifts between the *in vitro* and *in vivo* systems in which MSCs operate.

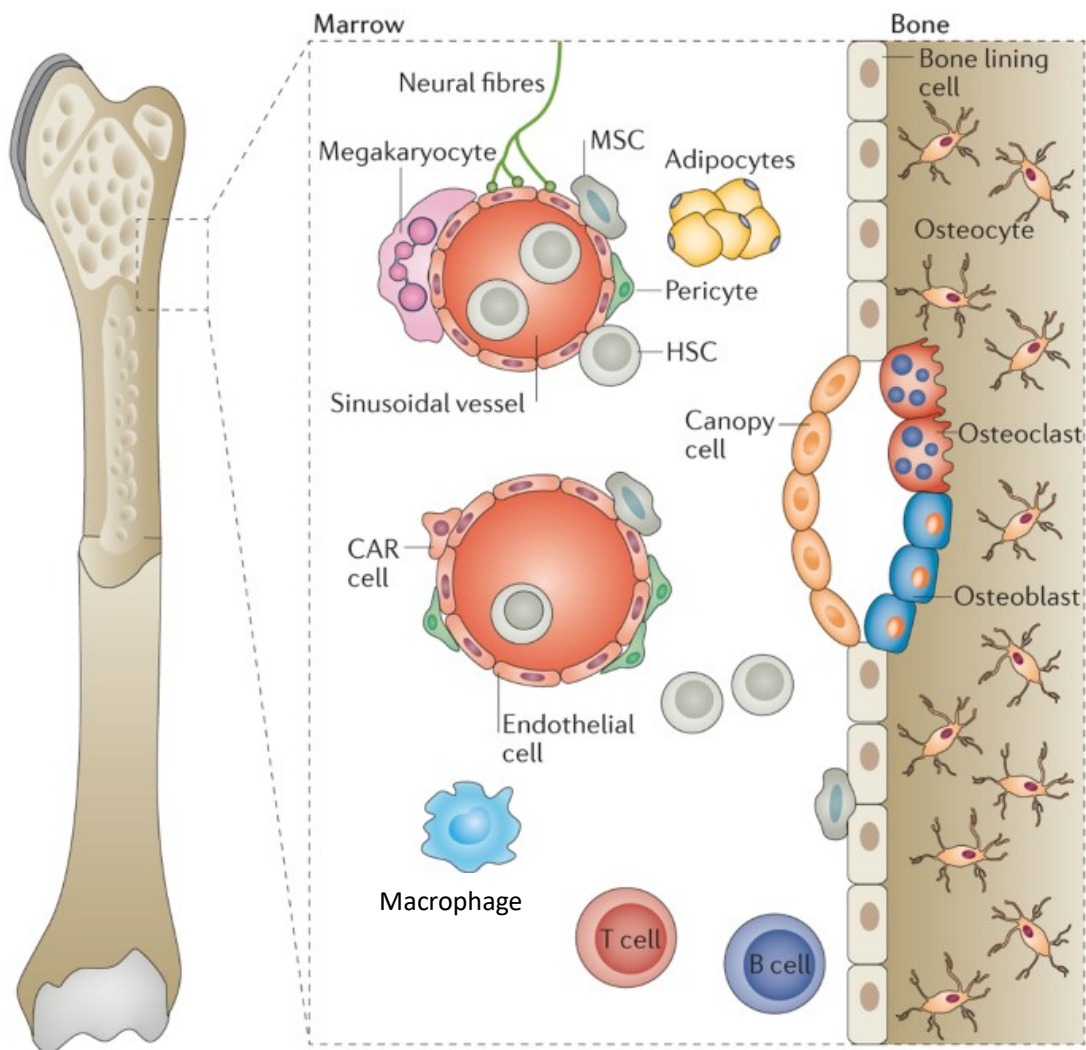


Figure 2. Schematic view of MSCs in hematopoietic niches within the bone marrow. CAR cells are CXCL12-abundant reticular cells. (From Reagan & Rosen, 2016)

The bone marrow niche is one of the most extensively studied, providing a specific atmosphere for proper hematopoiesis and consisting of hematopoietic and non-hematopoietic cells. **Figure 2** illustrates the MSCs within the bone marrow hematopoietic niche (Reagan & Rosen, 2016). The niche is perivascular and is formed between mesenchymal stromal cells and endothelial cells. In humans,

mesenchymal cells express CD146 and factors that promote hematopoietic stem cell (HSC) maintenance. Similarly, in mice, different types of MSCs can be identified based on the expression of transgenes such as LEPR-CRE64, PRX1-CRE62, CXCL12-GFP33 or NES-GFP that contribute to the HSC regulation. Perivascular stromal cells, including CXCL12-abundant reticular (CAR) cells, are associated with the bone formation. In addition, other cell types such as non-myelinating Schwann cells (which are also Nestin, NES, positive), macrophages and osteoclasts play a role in regulating the HSC niche. Extracellular matrix and calcium are additional regulators of HSCs. Endothelial cells also contribute to the perivascular niche of HSCs, and facilitate long-term reconstitution of HSC expansion in culture (Morrison & Scadden, 2014). Adipocytes are important for the production of stem cell factors, which are essential for HSCs. Adipogenesis is modulated by estrogens, which may lead to age differences in bone marrow homeostasis. Osteoblasts do not directly maintain HSCs, but facilitate the maintenance and the differentiation of certain lymphoid progenitors into B and T cells by releasing C-X-C motif chemokine ligand 12 (CXCL12), a chemokine essential for the maintenance of HSCs and their retention in the bone marrow. Additionally, the osteogenic factor osteolectin (CLEC11A), is crucial for osteogenesis and skeletal maintenance. These two differentiation roles of osteoblasts have an impact on the hematopoietic niche. Consequently, deregulation of BM niche cells can disrupt normal stem cell behavior, potentially leading to malignant transformation (Fröbel et al., 2021).

d. Hematopoietic Stem Cells, hematopoiesis and immune response

Hematopoietic Stem Cells are multipotent cells capable of differentiating into various blood cell types, including those of myeloid and lymphoid lineage. Like MSCs, HSCs have the capacity for self-renewal and differentiation to achieve the functional maturation. They are ubiquitously distributed in various organs, such as peripheral blood (PB), BM, and umbilical cord blood (UCB). HSCs give rise to different types of blood cells through the process of *hematopoiesis*, which can be divided into two main lineages based on the cells types they differentiate into: lymphoid or myeloid. Within the lymphoid lineage, lymphopoiesis generates subtypes such as T cells, B cells, and natural killer (NK) cells, which are essential components of the innate and adaptive immune systems. Conversely, the myeloid lineage of cells is generated in the myelopoiesis, which includes: erythrocytes from erythroblasts, platelets from megakaryocytes, monocytes from monoblasts (monocytes circulate in the bloodstream and can differentiate into macrophages when they invade tissues), and granulocytes (composed of neutrophils, involved in fighting bacterial infections, eosinophils, associated with allergic responses and defense against parasites, and basophils, involved in inflammation and allergic reactions). All of these cells significantly influence the BM niche.

The immune system orchestrates a dynamic interplay in which lymphoid lineage cells (T and B cells) facilitate the coordination of adaptive immune processes, while myeloid cells drive innate immune cell responses. This orchestration is important for enhancing their therapeutic effects in clinical applications (J. Y. Lee & Hong, 2020).

e. Most abundant stromal cells: Fibroblasts

Fibroblasts (FIB) are broadly characterized as connective-tissue-resident cells that are responsible for synthesizing the extracellular matrix (ECM) that supports and connects other tissues or organs in the body. These cells play several key roles such as the production of Collagen to form the ECM, tendons and bones. In wound healing, skin FIBs are important, while there are other Fibroblasts that play central role in inflammation in response to injury, infection or certain chronic diseases, modulating the immune response. FIBs are also involved in tissue development and aging, supporting heart tissue, tissue growth. In a negative action they can contribute to fibrosis and cancer (D'Urso & Kurniawan, 2020; Tracy et al., 2016; K. Wei et al., 2021).

Fibroblasts share a close relationship with MSCs, characterized by analogous morphological features. FIBs also express most of the ISCT-approved MSC markers, and show a variable percentage of multipotentiality and immunoregulatory capacity (Vaculik et al., 2012). Studies have delved into the gene expression profiles of Mesenchymal Stem Cells and Fibroblasts, using different sets of genes as gene markers. According to numerous investigations, MSCs and Fibroblasts also show similarities in their transcriptomic profiles since they belong to the same cell lineage; despite the differences in gene expression inherent to cell populations isolated by different laboratories, variations also appear when compared different sources of stromal or mesenchymal cells. Furthermore, some sets of FIBs have shown some differentiation capacity as MSCs, and similar immunomodulatory properties and epigenetic patterns, presenting serious challenges for correct identification and differentiation by researchers (Ugurlu & Karaoz, 2020).

f. The gene regulatory system of Mesenchymal Stem Cells

MSCs play a pivotal role in the regulation of stem cell niches in various tissues. Particularly, within the stem cell niche of the BM, MSCs are critical contributors, orchestrating homeostasis and regenerative processes through paracrine modulation of Hematopoietic Stem Cells and the interaction with other cells. Activation of MSCs occurs through engagement in a variety of signaling pathways. These stem cells also exhibit secretion of a number of growth factors and cytokines, in addition to their capacity to transfer organelles and extracellular vesicles to specific target cells.

In addition, MSCs are involved in the synthesis of the ECM, which has structural and signaling functions. Notably, the dynamic contribution of MSCs extends to the replenishment of cellular components within the niche. Thus, MSCs are able to differentiate into some niche components, while attracting relevant functional cells to the niche. In addition, MSCs are able to replenish the reservoir of the stem cell pool by conferring stemness properties to differentiated cells. Remarkably, MSCs have the ability to respond to metabolic, mechanical, and biological paracrine stimuli. This underscores the remarkable plasticity of MSCs within their given microenvironment.

Furthermore, MSCs exhibit the ability to act as precursors of individual niche components. For example, an imbalance between the osteogenic and adipogenic differentiation lineages of MSCs can culminate in a reduction within the osteoblast pool, which is critical for the maintenance of hematopoiesis. Recognized for their pro-angiogenic properties, these cells also produce abundant amounts of anti-angiogenic factors, such as pigment epithelium-derived factor and thrombospondins. Therefore, they could be modified by specific stimuli to limit the formation of new blood vessels or, instead, promote the stabilization of existing vascular structures (Sagaradze et al., 2020).

Cytokines, also play a key role in the regulation of MSCs, conferring them anti-inflammatory and immunomodulatory effects. The most important cytokines secreted by MSCs are IFN- γ , TNF- α and IL1 which can induce the secretion of anti-inflammatory immunosuppressive factors. Other cytokines and chemokines secreted by MSCs that modulate the regulatory system are: IL10, IL6, transforming growth factor β (TGF- β), CCL2/MCP1, CCL5, IDO1, vascular endothelial growth factor (VEGF), ICAM1 or PGE2 (Kyurkchiev, 2014).

The main directors that regulate the processes in MSCs are Transcription Factors (TFs). TFs include a large number of proteins, excluding RNA polymerase, that initiate and regulate the transcription of genes. In MSCs, many TFs are involved in their differentiation into different lineages. The major transcription factors that play key roles in the differentiation of MSCs into osteocytes are Runt-related transcription factor 2 (RUNX2), Osterix (gene SP7), and β -catenin (gene CTNNB1). The RUNX2 transcription factor is an essential regulator of bone formation and the osteogenic differentiation of MSCs, including MSCs to differentiate into pre-osteoblast and inhibiting adipogenic and chondrogenic differentiation. The

up-regulation of RUNX2 is due to many signaling pathways, including the Wnt, Bone Morphogenetic Protein (BMP), and Notch signaling pathways. The transcription factor, TWIST1, also acts as a downstream of hypoxia inducible factor-1 (heterodimeric protein consisting of two subunits alpha and beta, i.e. genes HIF1A and HIF1B) and suppresses the expression of RUNX2 in MSCs. Other transcription factors have also been studied and reported to play a functional role in the differentiation of MSCs into osteocytes including HOXB7, RUNX2, TNF- α (gene TNF), FOXC2, YAP1, HOXA2, BMP9 and β -catenin. TWIST1 and HIF1A have an inhibitory effect on MSC differentiation into osteocytes through their direct or indirect interaction with RUNX2 (Querques et al., 2019).

In adipogenic differentiation, peroxisome proliferation-activated receptor γ (gene PPAR γ) is the most important TF that regulates the expression of genes responsible for adipogenic differentiation (Zhuang et al., 2016). PPAR γ has been reported to be up-regulated during the adipogenic differentiation of MSCs, and the inhibition of this transcription factor suppresses adipogenesis. Moreover, early B-cell factor (EBF1) is a member of a cascade of transcripts that play important roles in cellular function and differentiation. Other transcription factors have also been reported to play a functional role in the differentiation of MSCs into adipocytes, including CEBPB, PRDM16, TWIST1, TWIST2, SOX2 and OCT4. However, GATA2, FOXA1 and HOXC8 have an inhibitory effect on the differentiation of MSCs into adipocytes.

Finally, in chondrogenic differentiation, the major transcription factor that plays a key role in the differentiation of MSCs into chondrocytes is the SRY-related high mobility group-box gene 9 (SOX9). SOX9 is an early transcription factor of chondrogenic differentiation and controls the expression of key genes in chondrogenesis. It controls the expression of collagen type IX (gene COL9A1) by binding to the promoter of this gene and forming trans-activating complexes with other proteins. In addition, Zinc-finger protein 145 (ZNF145) is a transcription factor that plays a role in the differentiation of MSCs into chondrocytes, its overexpression enhances SOX9 expression and chondrogenesis, according to the literature. Other relevant TFs involved in chondrogenesis that have been reported are: FOXO3A, HOXD9, HOXD10, HOXD11, HOXD13, STAT3 and WNT11 (Almalki & Agrawal, 2016).

Master Regulators (MR) are defined as the main transcription factors that differentially regulate groups of target genes (called *regulons*) so that they are expressed in a coordinated manner in a specific biological process or to achieve a specific cellular state. MRs regulate specific set of genes by activating or repressing their expression and, thereby activating or inhibiting the function of the corresponding proteins. In the case of MSC, not many genes with these defined characteristics have been described or are known. In embryonic stem cells, a small number of transcription factors, including OCT4, SRY box-containing factor 2 (SOX2), and NANOG, have been identified as master regulators. These MRs promote their self-renewal, hyperproliferation, suppressed apoptosis, evasion of immune surveillance, and plasticity (Hepburn et al., 2019). Furthermore, in adipogenic differentiation, PPARB and PPARD genes have been identified as MRs due to the maintenance of bone homeostasis by promoting Wnt signaling activity in osteoblasts and MSCs and their anti-inflammatory properties associated with the inhibition of pro-inflammatory cytokine production, Nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) signaling and expression of cell adhesion molecules (Djouad et al., 2017).

g. Immunomodulation of Mesenchymal Stem Cells and the Toll-Like receptors

The study of human MSCs (hMSCs) has been a prominent topic in biomedical research due to their biological significance in various cell-based therapies developed in recent years. Currently, numerous studies focus on the use of MSCs in cell therapy, with various clinical applications in the field of regenerative medicine due to their immunomodulatory capabilities. MSCs are used as a tool to treat various inflammatory or autoimmune diseases, such as degenerative changes in joints, bones or cartilage reconstruction, graft-versus-host disease, inflammatory bowel disease, etc. (Petri et al., 2017).

MSCs contain Toll-Like receptors (TLR), which are a family of type I transmembrane glycoproteins. These receptors possess an extracellular domain rich in leucine repeats that sense invading pathogens or signals of endogenous damage and initiate the innate and adaptive immune response. The interactions between TLRs and adapter proteins in TLR signaling are mediated by the Toll-Interleukin 1 Receptor (TIR) domains, which are conserved structural domains found in both TLRs and various adapter proteins. In humans, there are ten functional TLRs that are divided into two subgroups based on their localization in cellular compartments: (i) TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10 are expressed on the cell surface and primarily recognize microbial membrane and cell wall components; whereas (ii) TLR3, TLR7, TLR8 and TLR9 are expressed only in intracellular compartments, recognize viral nucleic acids and in endolysosomal compartments.

The signaling cascade of TLRs involves the recruitment of various Toll-Interleukin 1 (IL1) Receptor (TIR) domain-containing adaptor molecules. There are four cytosolic adaptor proteins associated with TLRs. Myeloid differentiation primary response gene 88 (MyD88), which is associated with all TLRs except TLR3. TIR domain-containing adapter protein (TIRAP), which is associated with TLR1, TLR2 and TLR4. TIR-domain containing adapter-inducing interferon- β (TRIF), corresponding to TLR3 and TLR4. And the TRIF-related adaptor molecule (TRAM) corresponding to TLR4 (Najar et al., 2017). These are all presented in **Figure 3**. MSCs express TLRs, especially TLR3 and TLR4, to modulate, polarize and activate their stem cell functions, to regulate migration or differentiation, and to modulate the immune system in a manner that depends on their activation state.

Toll-like receptor 4 (TLR4) is an immune sensor that detects invading microbes, and plays a central role in the induction of immune responses, essentially in innate and adaptive immunity. TLR4 activation involves two signaling pathways: MYD88-dependent and -independent pathways. These pathways activate the NF- κ B pathway, which is involved in the cellular response to stress stimuli, cytokines, ultraviolet radiation, oxidized LDL and bacterial or viral antigens. Additionally, the mitogen-activated protein kinase (MAPK), phosphoinositide 3-kinases (PI3K), and Wnt signaling pathways are involved in the regulation of cell cycle entry and proliferation. TLR4 signaling in MSCs influences their survival, differentiation, proliferation, migration and secretion of pro-inflammatory cytokines. TLR4 can be activated by endogenous compounds called damage-associated molecular patterns (DAMPs) released during tissue injury, by non-infectious conditions to promote tissue repair, and by lipopolysaccharides (LPS) from bacterial pathogens (He et al., 2016).

LPS are large molecules consisting of a lipid and a polysaccharide that are bacterial toxins found in the outer membrane of Gram-negative bacteria. LPS is a potent immune system activator and pyrogen. In severe cases, LPS can contribute to the development of deleterious effects in autoimmunity, obesity, depression, and cellular senescence (Moran et al., 1996). LPS stimulation of TLR4 leads to inflammatory modulation of MSCs toward a cellular phenotype commonly termed "MSC1". LPS binds TLR4 in collaboration with of LPS binding protein and CD14, involving the MD-2 protein, which is stably associated with the extracellular fragment of the receptor. The binding of an LPS molecule to the TLR4/MD-2 complex involves acyl chains and phosphate groups of lipid A, the conserved part of LPS and the main inducer of pro-inflammatory responses to LPS (Najar et al., 2017).

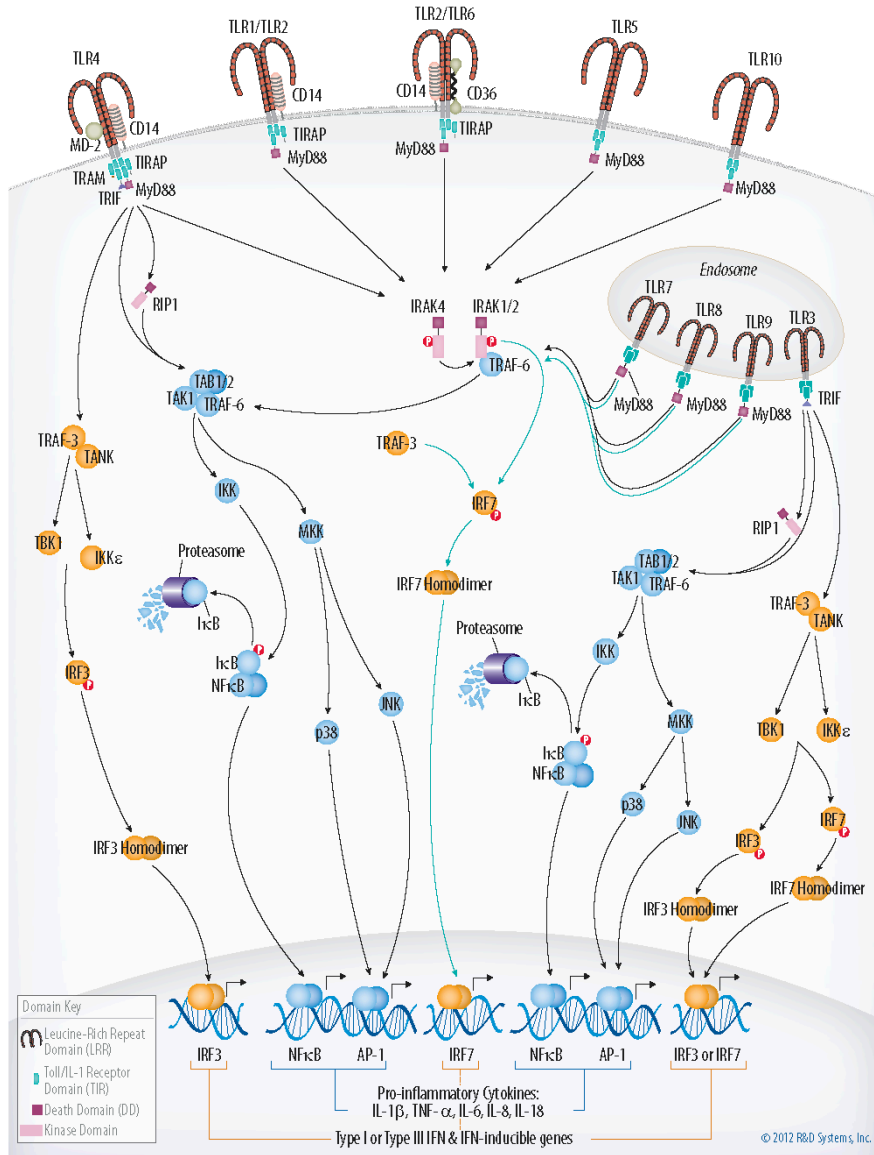


Figure 3. Human TLRs Signaling, including the pathways of TLR3 and TLR4. (From R & D systems, 2012; <https://www.rndsystems.com/>)

TLR4-primed MSCs protect them from oxidative stress-induced apoptosis and enhance their proliferation via PI3K/AKT signaling. Moreover, TLR4 activation in MSCs is associated with functions related to promoting osteogenic differentiation (Raicevic et al., 2012).

A study by Hwa Kim and colleagues explored the functional transcriptomic analysis of TLR4-primed in human MSCs to investigate differential gene expression associated with chemotaxis and inflammatory responses. They identified novel immune genes, as well as transcription factors involved in the NF- κ B and PI3K pathways, which orchestrate the expression of cytokines and chemokines (Hwa Kim et al., 2016). Genes such as CCL2, CXCL2, CXCL3, CXCL5, CXCL6, CXCL8 and CXCL10 were found to be upregulated in TLR4-primed hMSCs. These genes are related to anti-inflammatory and angiogenic functions in response to inflammatory stimulation (Cuesta-Gomez et al., 2021). The study also highlighted the upregulation of interleukins and interleukin-related genes, including IL6 and LIF, which act as potent activators of immune cells and recruit cells to sites of inflammation. Additionally, IL4 and IL10 have also been implicated in immunoregulatory effects on TLR4-primed hMSCs via the NF- κ B pathway (T. Lin et al., 2017).

Furthermore, interferon-related genes involved in antiviral immunity and signalling showed an increased expression in TLR4-primed MSCs, including genes such as: IFI44, IFI44L, IFIH1, IFIT2, IFIT5, IFITM1, IRG1, ISG15, ISG20, MX1, MX2, OAS2 and OAS3. Interferon signalling, particularly type-I IFN, is involved in the TRIF-dependent pathway. TLR3 signalling involves the recruitment of TRIF via the TRAM adaptor, whereas TLR4 initially recruits MAL and MYD88 to the plasma membrane, and subsequently delivers TRAM and TRIF. Downstream signalling activates IRF3, which forms complexes with IRF7 and is translocated to the nucleus, where IFN and IFN-inducible genes are transcribed. These results suggest that TRIF signalling involves multiple transcription factors.

Conversely, Toll-like receptor 3 (TLR3) recognizes polyinosinic:polycytidylic acid (poly(I:C)). Poly(I:C) is a mismatched double-stranded RNA (dsRNA) molecule in which one strand is a polymer of inosinic acid and the other a polymer of cytidylic acid, and is considered a synthetic analog of double-stranded RNA found in some viruses (Li et al., 2015). Poly(I:C) is an immunostimulant that binds TLR3, which is widely used to induce interferon response. TLR3 binding by dsRNA uses only the MyD88-independent pathway to activate IRF3 and, further, to induce the transcription of type I interferons. TLR3 activation enhances the therapeutic efficacy of MSCs in some inflammatory diseases, generating a subtype of MSCs called "MSC2". TLR3 can also homodimerize with TLR4 and TLR9, creating a cross-talk response against invading pathogens.

Stimulation of TLR3 in MSCs resulted an increase in the cytokines CCL5 (also known as RANTES), CXCL10 (also known as interferon gamma-induced protein 10, IFI10), and IL10. TLR3 stimulation triggers an indirect JAK/STAT signalling cascade through the induction of type I interferons, resulting in the activation of SOCS1 and SOCS3 (suppressor of cytokine signalling 1 and 3). Activation of these proteins leads to downregulation of the expression of the chemokine receptor, CXCR4, which influences the migratory behaviour of hMSCs dependent on the CXCR4-CXCR7 pathway.

TLR3-primed hMSCs exhibited higher levels of fibronectin, decreased expression of TGF- β 1 and β 3 and increased SMAD7, an inhibitor of the TGF- β signalling pathway. Consequently, this effect implies that TGF- β , an immunoregulatory factor, is associated with the TLR3-primed phenotype rather than the pro-inflammatory TLR4-primed phenotype. TGF- β cooperates in the TLR3-Jagged-1-Notch-1 pathway by increasing the expression of PGE2 to regulate the crucial transcription factor (FOXP3) which favours the development of regulatory T cells (Tregs) (Najar et al., 2017). In addition, TLR3-stimulated MSCs increase the expression ofIDO through the up-regulation of cyclooxygenase-2 (COX2). TLR3-primed hMSCs and unstimulated hMSCs suppressed T-lymphocyte activation, highlighting their immunosuppressive

properties. Specifically, TLR3-priming led to increased fibronectin deposition, regulation of immunoregulatory mediators, and maintenance of suppression of T-cell activation (Waterman et al., 2010).

In addition, TLR3-primed MSCs enhance their suppressive functions against NK cells. Robert Michael and colleagues described a time-dependent regulatory system of poly(I:C) stimulated hMSCs to model the dynamic interaction between NK cells and stimulated MSCs. At short time points (4 hours), activated MSCs secreted type I interferon, which enhanced the effector function of NK cells. However, at longer time points (24 hours), the functions of NK cells are restricted as a result of the influence of TGF- β and IL6, limiting NK cell effector function and inducing a senescent-like regulatory phenotype in NK cells, as illustrated in **Figure 4**.

In a short inflammatory phase, poly(I:C)-activated MSCs rapidly released IFN- α/β , IL6, and IL8 within 4–8 hours. Thereafter, from 8 to 24 hours, the levels of type I interferons decreased, while the levels of IL6 and IL8 continue to increase. Additionally, the supernatant was enriched in: IFN- γ inducible protein 10 (IFI10, also known as CXCL10), macrophage migration inhibitory factor (MIF), and TGF- β during this time. Notably, the induction of IL6, IL8, and IFI10 was triggered by poly(I:C), whereas MIF and TGF- β were not affected. Comparing the ratio of antiviral type I IFN to the regulatory factors (IL6, IL8, MIF, and TGF- β) at short (4 hours) and longer (24 hours) time points in the supernatant of poly(I:C)-stimulated MSC, a clear shift towards the regulatory factors becomes evident over time (Petri et al., 2017).

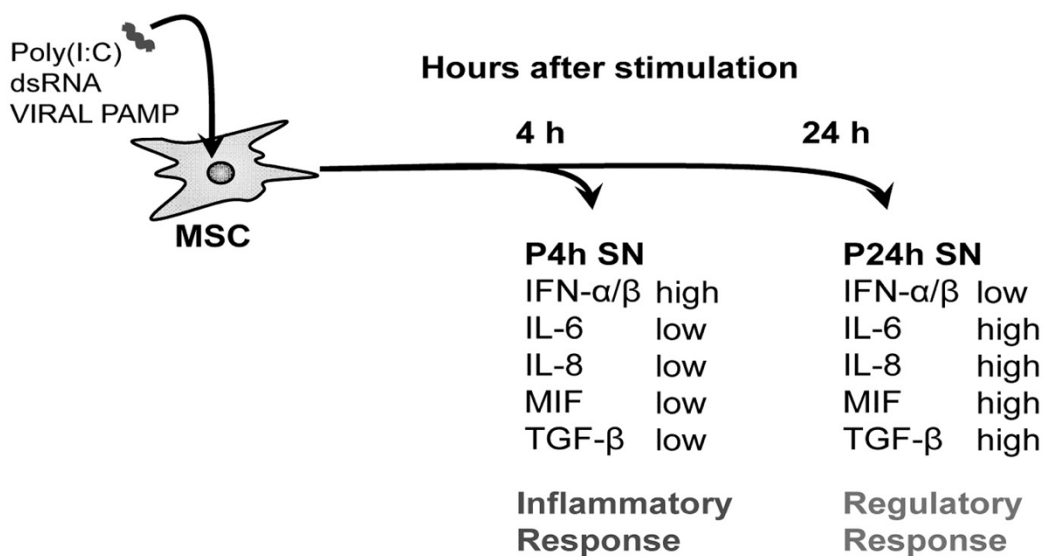


Figure 4. Diagram showing the principle of MSCs cell stimulation using poly(I:C) at short time of 4 hours and longer time of 24 hours. (From Petri et al., 2017.)

Finally, the activation of TLR3 in MSCs has been associated with the enhancement of the migratory potency in AD-MSCs. IGF1 facilitates migration by increasing the expression of chemokine receptors, including CCR5 (receptor of CCL5 on MSCs). It has also been shown that growth factors, including the basic fibroblast growth factor (FGF2, also known as bFGF) and the VEGF, released under hypoxemic stress, enhance migratory propensity by activating the PI3K/AKT pathway downstream of the bFGF receptor (BFGFR) on the MSCs. In addition, it has been shown that the synergistic effect of VEGF and platelet-derived growth factor ab (PDGFA and PDGFB) was shown to act as chemo-attractants to induce MSC migration.

Notably, in short times of poly(I:C) exposure (4 hours), expression of $\alpha 4$, $\alpha 5$, $\beta 1$ integrins, as well as, IL10 and TGF- β influence in the hMSCs migration. Conversely, the longer time of TLR3 stimulation (24 hours) in hMSCs leads to a decrease in migration ([Eskandari et al., 2023](#)).

2. CELLULAR GENE EXPRESSION AND TRANSCRIPTOMICS

a. Genes activation: the process of gene expression and transcription

Gene expression is the mechanism by which the information embedded in the sequence of a gene is used to synthesize a functional gene product. Transcription is the molecular process responsible for carrying out gene expression. During transcription, a specific segment of DNA is transcribed into messenger RNA (mRNA) by the enzymatic complex of the RNA polymerase. The process shown in **Figure 5** illustrates the sequential progression of information transfer: beginning at a gene locus within the DNA, progressing to the generation of immature mRNAs (the pre-mRNAs, which include both exon and intron sequences), followed by the pre-mRNAs splicing and processing to produce mature mRNAs, which then exit the nucleus. Following these transcription steps, the mRNA is used as a template for the production of protein copies, which is carried out by the translation machinery which involves the ribosomes and many molecular components responsible for reading the nucleotide sequence to translate it into the polypeptide sequence in the process of protein synthesis. These steps describe protein-coding transcriptomics, but there are many gene loci that produce transcripts that do not encode proteins.

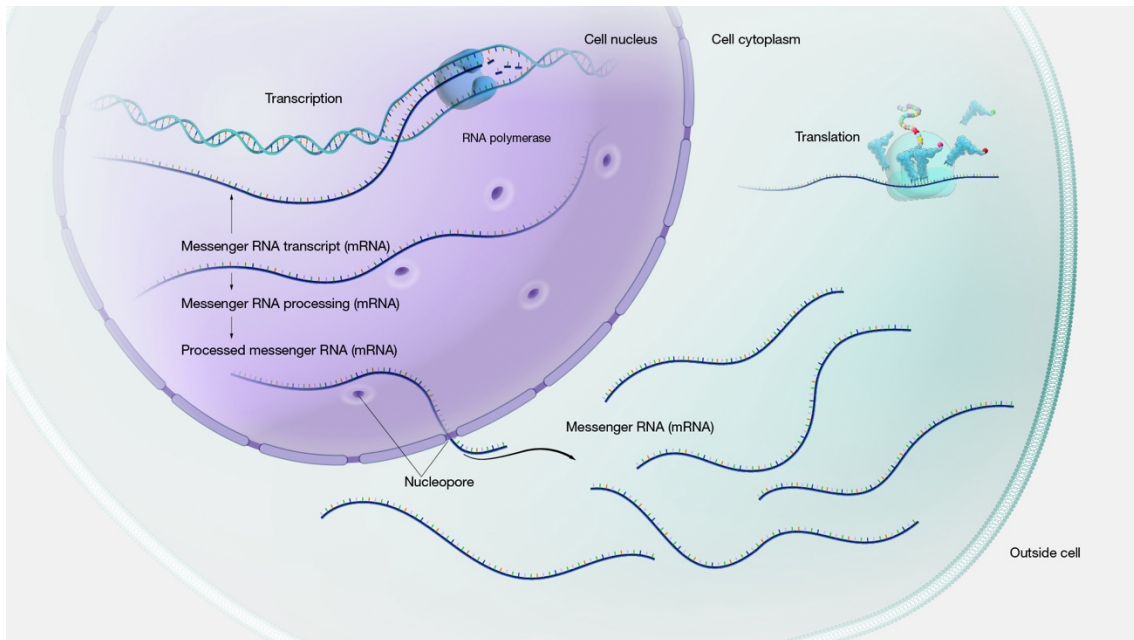


Figure 5. Schematic representation of the cellular gene expression process. (From [National Human Genome Research Institute, https://www.genome.gov/](https://www.genome.gov/))

Therefore, the global transcriptomics of a cell also includes transcripts derived from non-protein coding genes, such as ribosomal RNAs, transfer RNAs, long non-coding RNAs, microRNAs, small nuclear RNAs, and others. These RNAs, although they do not encode proteins, have relevant biological functions and assume multiple operational, structural, or regulatory roles, even during gene expression.

Most transcriptomic studies focus on the identification and quantification of all mRNAs produced in the cell, as this analysis is the vehicle to convert the genetic information encoded in the genome into the proteome. The amount of each gene transcribed into mRNA defines the level of expression of that gene presents in a cellular system. Proteo-transcriptomic studies in human samples have shown that mRNA production and concentration generally correlate well with protein production and concentration. However, it is clear that due to post-transcriptional changes and regulation, the quantity of many proteins can exhibit very specific changes that cause deviations with respect to the concentration of their

corresponding mRNA. However, the precise measurement of all mRNAs expressed in a cellular system is a very good resource to identify genes that are actively involved in a specific biological context or in a given condition (Buccitelli & Selbach, 2020). Therefore, global gene expression profiling (i.e., complete transcriptomic profiling) is a powerful method to understand how the inherited genomic information is used by the cell, and it is now the most comprehensive gateway to unravelling and understanding the phenotype of a biological system.

b. Genome-wide expression measurements: Microarrays and RNA-seq techniques

In recent years, the development of RNA-based global gene expression measurement has been widely extended to biomedical research and clinical applications in recent years. Gene expression profiling plays a crucial role in identifying gene signatures for disease diagnosis, predicting prognosis, and guiding the development of new cellular therapies. Currently, the most widely used techniques are: in small scale, real-time quantitative Polymerase Chain Reaction (RT-qPCR) (to measure up to a hundred genes); and in large scale, expression Microarrays, and RNA sequencing (RNA-seq) (to measure thousands to tens of thousands of genes, i.e., for a complete transcriptome profiling). More recently, these large-scale techniques, which measure expression in bulk samples (i.e., samples containing many cells), have been complemented by new emerging techniques, such as single-cell RNA sequencing (scRNA-seq), which is gaining importance in the biomedical field due to its ability to provide information on the expression profile of many individual cell types and unique cells.

In 1984, Kary Mullis introduced the Polymerase Chain Reaction (PCR), a technique that amplifies a specific segment of DNA, producing millions of copies in a matter of hours. In 1992, Higuchi and colleagues refined this method and developed qPCR to allow for accurate quantitative studies of gene expression (San Segundo-Val & Sanz-Lozano, 2016). Finally, real-time qPCR (RT-qPCR) was developed, which measures the amplification product as the reaction progresses, in real time, with quantification of the product after each cycle.

RT-qPCR is an hybridization technique the uses of oligonucleotide probes that hybridize to the target sequence. The TaqMan assay method uses the 5' endonuclease activity of Taq DNA polymerase to cleave an oligonucleotide probe during the PCR, producing a detectable signal. These probes are fluorescently labelled at the 5' end, and chemically modified at the 3' end to prevent extension. Specificity is ensured by the use of two PCR primers and the incorporated probe (Hoy, 2019). Quantification of mRNA by RT-qPCR can be either absolute or relative. Absolute quantification requires the construction of a calibration curve, whereas relative quantification expresses the target quantity for an experimental sample as an n-fold difference relative to a calibrator (Ho-Pun-Cheung et al., 2009). However, qPCR has limitations: normalization for small differences in gene expression and handling with a large number of genes.

In 1990, microarrays were conceptualized as a hybridization method. Microarrays consist of a collection of DNA probes that are usually attached to a solid surface, such as a glass slide, at defined positions. The probes are typically oligonucleotides that are applied in *ink-jet printed* onto slides (Agilent Microarrays) or synthesized *in situ* (Affymetrix Microarrays). Using high stringency conditions, labelled single-stranded DNA or antisense RNA fragments from a target sample of interest are hybridized to the DNA Microarray. The amount of hybridization detected for a given probe is proportional to the number of nucleic acid fragments present within the sample.

Figure 6 shows a schematic representation of a Microarray expression platform. The Microarray typically consists of a solid support material, such as a glass slide or silicon chip, which serves as the substrate for the attachment of oligonucleotide probes. Tens or hundreds of thousands of short, single-stranded DNA oligonucleotide probes are immobilized at a known location on the Microarray, each probe

designed to be complementary to a specific gene or transcript of interest. The preparation protocol for using a Microarray typically begins with the conversion of RNA isolated from a biological sample into more stable cDNA. The cDNA molecules are then fragmented using restriction endonucleases, and fluorescent labels are attached to these cDNA fragments. In the next step, the different fragmented and labelled cDNA sequences are introduced into the Microarray chip, where they bind to the affixed DNA probes (oligonucleotide probes) corresponding to the oligo sequences of the genes of a given species (for example, probes of all the human genes), which are located in micro or nanometric cells (arranged in well-labelled rows and columns) (**Figure 6**). Once the sample is introduced into the array, the different cDNA fragments hybridize with the specific cells where the complementary oligos are found located. In a final step, the specific fluorescent signal produced in each of the microarray cells is measured (using a high-sensitivity laser-based scanner) to obtain the raw data quantification, which is then integrated and summarized with a computational algorithm to allow calculation of the expression level of each gene.

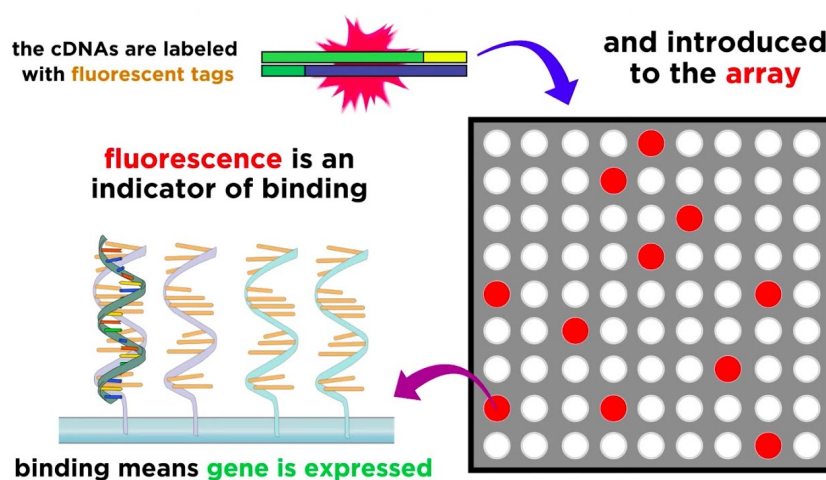


Figure 6. Microarray technique. (From [Matheny et al., 2011](#))

Microarray analysis is extremely useful in clinical research, providing insight into the molecular biology of disease. This technology provides opportunities for the molecular classification of diseases, the identification of novel molecular targets involved in disease development, and the analysis of therapeutic responses ([L. Wu et al., 2005](#)). However, despite the speed, sensitivity and specificity of this technique in identifying a large proportion of the mRNAs present in a biological sample, background hybridization limits the accuracy of expression measurements, particularly for transcripts present in low abundance. Furthermore, variations in hybridization characteristics between different probes and the fact that the arrays are limited to examining only the genes for which the probes are designed are further limitations ([Govindarajan et al., 2012](#)).

Over the past decade, high-throughput RNA sequencing (RNA-seq) has become a powerful and widely used technology for profiling the entire transcriptome of biological samples. RNA-seq is a sequencing technique that uses amplification-based next-generation sequencing (NGS) to reveal both the presence and amount of RNA in a biological sample. RNA-seq enables a comprehensive examination of the transcriptome, including coding RNAs (i.e., mRNAs) and non-coding RNAs (such as microRNAs, long non-coding RNAs), and provides information on isoforms and alternative splicing (as it detects all transcripts expressed at a gene locus), as well as the detection of sequence polymorphisms (i.e single nucleotide polymorphism, SNP).

A model of an RNA-seq workflow based on Illumina technology (<https://www.illumina.com/>) is depicted in **Figure 7**. The procedure begins with RNA extraction, followed by the RNA fragmentation. The RNA is then converted to cDNA by reverse transcription. The adapters are added to each end fragment of the cDNA and linked in the ligation process.

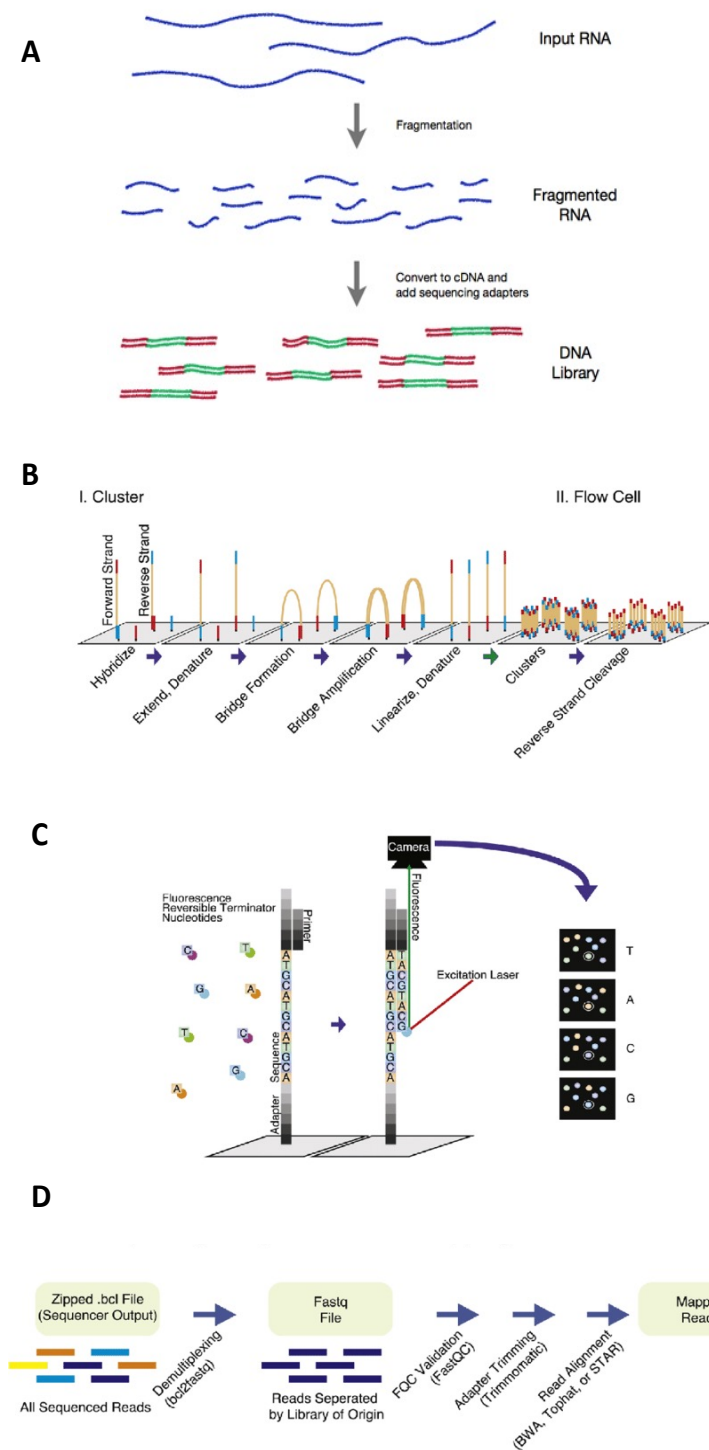


Figure 7. Workflow of RNA sequencing. **A)** RNA is fragmented, converted to cDNA and the sequencing adapters added to form the DNA libraries. **B)** The libraries are introduced in the flow cell for the amplification and cluster generation. **C)** dNTPs are introduced one by one, adhering to their complementary base, thus forming different complementary sequences. **D)** The raw data obtained is processed by bioinformatic tools to obtain the mapped reads. (Adapted from *Dpto. Rete Oncologica Piemonte e Valle d'Aosta, 2018* <http://www.reteoncologica.it/>)

The adapters contain elements that enable sequencing, including the amplification element, which facilitates clonal amplification of the fragments, and the primary sequencing priming site (**Figure 7A**). Sequences complementary to these adapters bind to the flow cell channel surfaces, allowing the binding of modified DNA molecules and providing a primer for DNA polymerase. DNA molecules from bridges, allowing bridge amplification and cluster generation (**Figure 7B**). Multiple rounds of amplification using deoxynucleoside triphosphate (dNTP) bases are performed. Each dNTP (dATP, dTTP, dCTP, and dGTP) is fluorescently labelled and acts as a terminator molecule due to the blocked ribose 3'-OH group, preventing elongation by polymerase. For each round of sequencing, a mixture containing all four labelled dNTP bases is present, and a single base is incorporated into each DNA molecule bound to the flow cell channels. The flow cell is then imaged to capture the dNTP base that was added at each specific cluster location. The fluorescent label and 3'-OH blocking group are then removed, followed by another round of sequencing. This results in the identification of the complete sequences of the DNA molecules bound to the flow cell channels (**Figure 7C**) (Anasagasti et al., 2012).

Traditionally, the sequencing depth has been around 10 to 50 million reads per sample on the most widely used sequencing platforms, which are those from Illumina (MiSeq, HiSeq, and NextSeq platforms). The final steps involve computational processing: alignment and/or assembly of sequencing reads into a transcriptome and quantification of reads that overlap specific transcripts and genes (**Figure 7D**).

The workflow for transcriptomic analysis constitutes a sequential process that starts with the acquisition of the *fastq* files, that are a common file format used in high-throughput sequencing technologies. The *fastq* files are a text-based format that stores both the sequence reads and their corresponding quality scores generated during the sequencing process. These files are the raw data output from the RNA-seq and are essential for downstream analysis of the sequencing data, including alignment, quality control, and quantification of gene expression. The *fastqc* (Fast Quality Control) software tool is then used to assess the quality of the raw high-throughput sequencing data.

Figure 8 illustrates all the steps in a workflow analysis of RNA-seq. This analysis includes several steps: (a) *Quality Control and Preprocessing*, which includes assessing the quality of the raw sequencing data using FastQC, followed by trimming to remove adapter sequences and filtering to remove very short reads or low quality nucleotides in the reads; (b) *Read Alignment or Mapping*, to align or map the processed reads to a reference genome or to a reference transcriptome using alignment tools (such as *STAR*, *HISAT2*, or *Bowtie2*) or pseudo-aligners (such as *Salmon* or *Kallisto*); (c) *Gene Expression Quantification and Normalization*, quantify gene or transcript expression by counting the number of reads (i.e., raw counts) mapped to each gene or transcript feature (using tools such as *featureCounts* or *HTSeq*), and normalizing read counts to account for library size and other variability factors (common methods include calculating TPM, Transcripts Per Million, or FPKM, Fragments Per Kilobase of transcript per Million mapped reads); (d) *Differential Expression (DE) Analysis*, comparing expression levels between different conditions (e.g., treatment versus control) to identify differentially expressed genes or transcripts (using tools such as *DESeq2*, *edgeR*, or *limma-voom*); (e) *Functional Enrichment Analysis (FEA)*, to understand the biological functions and processes associated with differentially expressed genes, performing functional enrichment analysis on various annotation databases such as Gene Ontology (GO), KEGG pathways, Reactome, etc.

The alignment of the sequenced reads can be performed against a reference genome or transcriptome, but if such references are not available, a *de novo* assembly of the transcripts should be performed, as shown in **Figure 9**. With the reference genome, two strategies or approaches can be followed for the gene mapping. The 'align-then-assemble' approach, which aligns short RNA-seq reads to the genome, taking into account possible splicing events, and then reconstructs transcripts from these

alignments. The 'assemble-then-align' approach, in which transcripts are first assembled *de novo* (directly from the RNA-seq reads) and then aligned to the genome to elucidate intron and exon structure and alternative splicing events (Haas & Zody, 2010). The align-then-assemble approach is more sensitive and is the one commonly used in RNA-seq analyses. Subsequently, read quantification assigns reads to a gene or transcript, and in the case of genes, this is done by mapping using a genome reference database and browser (such as ENSEMBL, <https://www.ensembl.org/>). Quasi-mappers or pseudo-aligners provide the counts in a fast way (with less effort to find the optimal alignment between the reads and the reference genome), but lose information, such as: variant detection, alternative splicing, position and orientation of the read on that transcript.

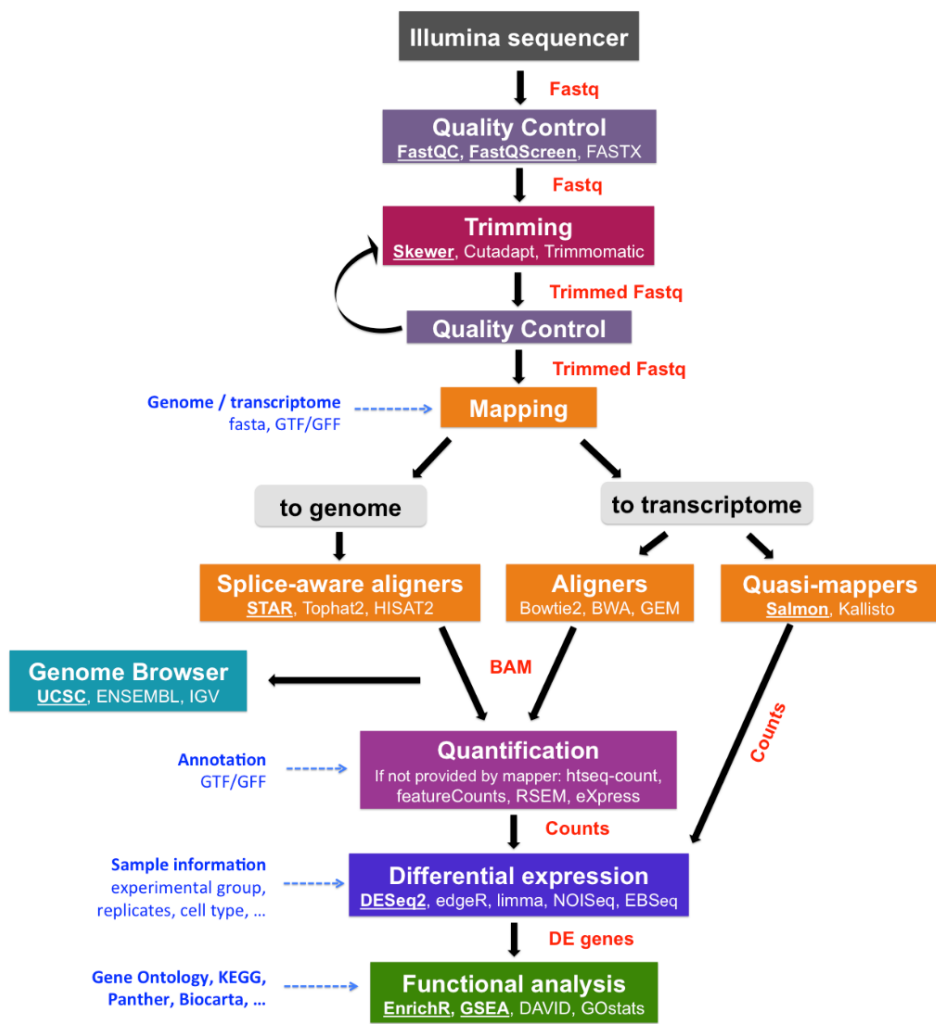


Figure 8. Workflow bioinformatic analysis of RNA sequencing.
(From [Biocore RNAseq course, 2019. https://github.com/biocorecrg/RNAseq_course_2019](https://github.com/biocorecrg/RNAseq_course_2019))

Once we have obtained an accurate quantification of gene expression from the RNA-seq data, we can proceed to multiple comparisons of the expression signal between different experimental conditions or different types of samples. This differential analysis can be done in many ways, especially if we have multiple samples and multiple conditions. As mentioned above, there are several bioinformatics tools to perform the differential expression analyses, and their statistical bases are not the same. Ultimately, researchers want to obtain a robust and stable identification of the genes or transcripts that show a significant change in a study condition. In general, the design of an RNA-seq data analysis workflow

is not a trivial matter and several studies have performed a comprehensive comparison of different analysis pipelines to reveal which are the most robust and accurate (Corchete et al., 2020).

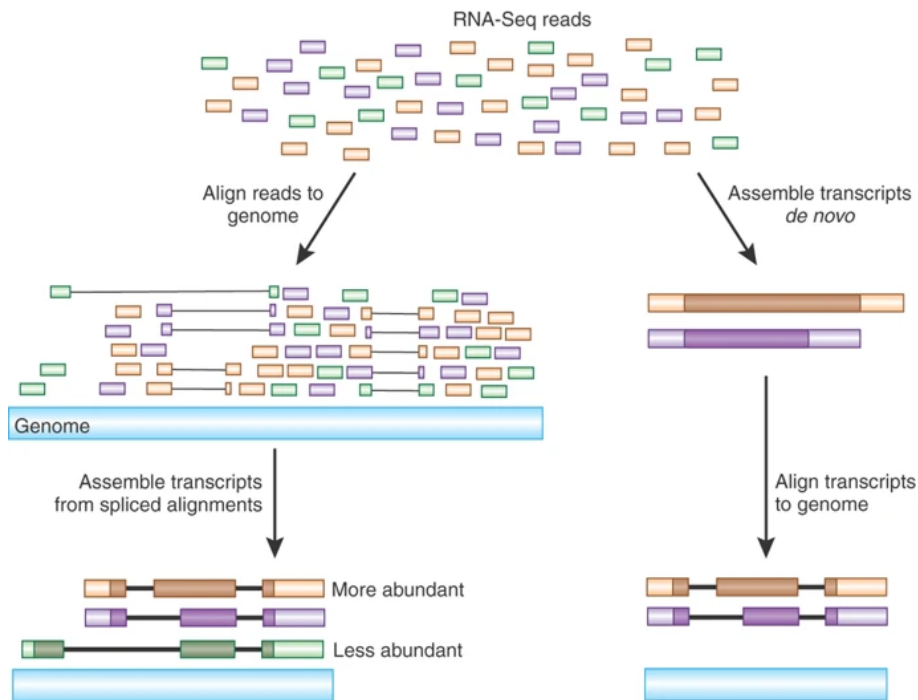


Figure 9. Types of RNA sequencing alignment. (From Haas & Zody, 2010)

RNA-seq offers numerous advantages as a comprehensive view of the transcriptome, with independence from prior sequence knowledge, RNA-seq has a number of strong capabilities, such as: high sensitivity, ability to detect polymorphisms, structural variations and alternative splicing, and accurate quantification of the absolute and relative abundance of transcripts. However, this technology, in contrast with Microarray expression data, involves more complex data analysis and a data storage challenge that requires a specialized computing infrastructure with qualified bioinformaticians; and considering this difficulty in data analysis, RNA-seq technology still has a higher cost than Microarrays technology today.

c. Cell-specific gene expression profiling: single-cell RNA sequencing

Recently, single-cell RNA sequencing (scRNA-seq) has attracted attention in clinical applications, particularly in oncology, immunology, and hematology, driven by the genomic and transcriptomic variability that occurs in each unique cell type and also by the changes that underlie cellular heterogeneity. Single-cell genomics is the application of NGS omic techniques to single cells. In 2009, the first article on scRNA-seq was published, and since then many developments and platforms have emerged that allow the precise analysis of thousands of cells in parallel (Tang et al., 2009).

Single cell isolation is the first step in the process. The traditional method for isolating single cells has been flow cytometry for many years. Within the context of this technique, two specific methods have been used to isolate cells: flow-activated cell sorting (FACS) is a commonly used strategy that requires large starting volumes (input numbers <10,000) and specific monoclonal antibodies to target the cells of interest. Alternatively, Magnetic activated cell sorting uses beads coated with antibodies that bind the antigen present on the cells of interest. After isolation, a strong magnetic field separates the cells that do not express the antigen flowing in the supernatant, while the cells with the antigen remain attached to the magnetic beads (Gross et al., 2015). The problem associated with flow cytometry is that the number

of molecular markers per cell is limited (currently to a maximum of about a hundred), and that the number of unique cells that can be phenotyped is limited to about a thousand or tens of thousands.

The scRNA-seq field has taken a different approach to flow cytometry, aiming to achieve single-cell profiling with large-scale omic techniques (i.e., identifying thousands of markers at once) and also aiming to isolate hundreds of thousands of distinct cells. The field of scRNA-seq field has evolved tremendously since the first paper was published back in 2009, because while the first methods analyzed only a handful of cells, the throughput and performance increased rapidly in a very short period of time. Indeed, it was only with the introduction of emulsion droplet methods, such as the well-known kits commercialized by 10x Genomics (<https://www.10xgenomics.com/>), that the robust and reproducible analysis of thousands of cells became feasible (Zheng et al., 2017).

Different capture techniques determine the number of measurable cells. The droplet microfluidics technology illustrated in **Figure 10** (corresponding to the 10x Genomics *Chromium* single cell platform) consists of encapsulating individual cells in aqueous microdroplets, each of which is tagged/labelled with a unique barcode for subsequent molecular indexing, as illustrated **Figure 10A**. Each individual cell is encapsulated in a droplet with a uniquely barcoded bead, the cells are lysed, and a specific cell barcode, Unique Molecular Identifier (UMI) and PCR adaptor are attached to the poly A chain of each mRNA. Finally, a cDNA is produced using reverse transcription, and template-switching reverse transcription is used to remove low-quality sequences prior to the amplification. This technology provides high throughput, high sensitivity in gene detection, and time efficiency (Li & Humphreys, 2021).

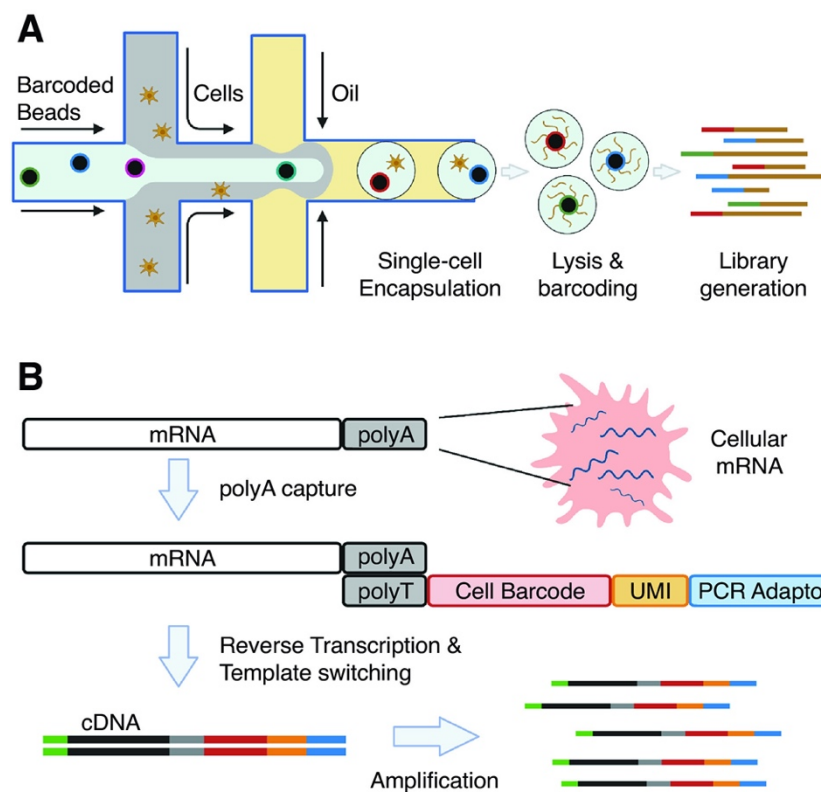


Figure 10. Droplet-based microfluidic technique. **A)** Cells are encapsulated in a droplet, cells are lysed and mRNA is captured by uniquely barcoded beads. **B)** mRNAs are captured by oligos (conjugated in beads), UMIs and other adapter sequences. In this process cDNA is synthesized and the library is generated. Then, library modification (via template switching) is done to allow the library amplification. (From Li & Humphreys, 2021)

Other platforms, other than 10x Genomics that produce single-cell libraries are based on the use of microfluidic plates. One of these is the Fluidigm platform (<https://www.standardbio.com/>), which allows automated isolation of 96 cells in a plate, cDNA synthesis, and amplification based on Smart-seq through microfluidics. However, this method only provides individual information of the cells but not data about the cell populations and can only separate hundreds or thousands of cells (Kashima et al., 2020).

After the construction of the single-cell libraries, RNA sequencing is performed by the standard method (mainly using Illumina platforms). The RNA sequencing is able to identify every transcript in every single cell thanks to the advancement of multi-Tag protocols. These protocols, combined with UMIs, allows multiplexing and improve quantification. Paired-end sequencing is used to eliminate limitations such as limited alignment, facilitating the detection of genomic rearrangements and novel transcripts. The high throughput of Tag-based approaches finds application in studies of gene expression levels, cell type discovery, and tissue composition. Alternatively, there are other sequencing methods, such as full-length based protocols, which provide coverage of different transcripts, and allow the discovery of alternative splicing events, allele-specific expression and single nucleotide polymorphisms. However, these protocols have a high complexity for correct data analysis (Wolfien et al., 2021). Indeed, adequate analysis of scRNA-seq data is a major challenge, as many bioinformatics tools are under development and good benchmarking is needed. **Figure 11** contains a schematic workflow for single-cell data analysis. Raw sequencing data are stored in *fastq* files and Binary Base Call (BCL) format. *Fastqc* can also be used to evaluate the quality of raw single-cell RNA-seq data.

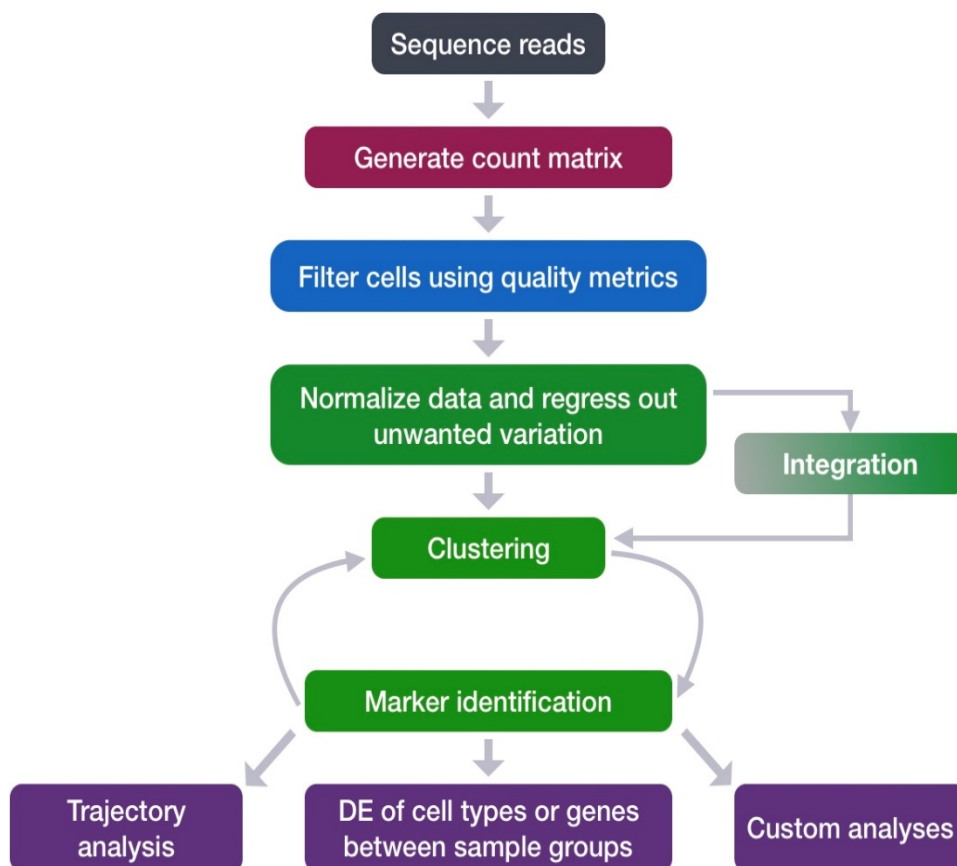


Figure 11. Schematic representation of a sc-RNAseq data analysis workflow from raw reads to cell clustering and gene marker identification. (From Mary et al., 2022)

High quality reads are aligned to the specific reference genome using an appropriate aligner (such as *STAR* or *TopHat2*), and cell counting is managed using Cell-Ranger tools, which perform alignment, filtering, UMI counting, and other analysis steps internally. In fact, Cell-Ranger (10x Genomics) is a set of analysis pipelines that process *Chromium* single cell data to align reads, generate feature barcode matrices, perform clustering, and other secondary analyses. Cell-Ranger uses *STAR* as its aligner, which performs a splice-aware alignment of reads to the genome and then uses a General Transfer Format (GTF) transcriptional annotation file to categorize these reads as exons, introns, and intergenic based on whether the reads are reliably aligned to the genome.

After generating the count matrix, a second quality control check is performed using tools such as: *Scran* (bioconductor.org/packages/scrans.html) released in 2016, *Scanpy* (github.com/theislab/Scanpy) released in 2018, and *Seurat* (satijalab.org/seurat/) released in 2018. The most commonly used package is *Seurat*, programmed under R, a free software for statistical computing and graphics. Quality control (QC) indicators allow to filter useful cells based on the number of genes, the number of UMIs (transcripts), the percentage of mitochondrial genes, and the percentage of ribosomal protein genes in each cell. Similar to RNA-seq samples, each cell is treated as an independent sample file when analyzing the scRNA-seq data. If the levels between cells are not comparable in the original raw data, normalization is required. Normalization is essential to counteract technical bias and ensure data comparability. Each scRNA-seq file is highly-dimensional, with tens of thousands of cells and thousands of genes expressed in each cell. Most genes are housekeeping or do not show significant changes in expression level, and their presence tends to interfere with true biological signals. Highly variable genes (HVGs) are critical to the analysis because they highlight biological signals while reducing computational complexity. High-quality HVGs include genes that can distinguish different cell types, and the quality of HVGs has an important impact on the clustering accuracy.

If the data originate from different sequencing platforms, integration of the data is necessary to avoid batch effects, as in RNA-seq. In this case, the *Seurat* package, with an integration tool can combine several datasets, including two at each step. Other packages, such as *Harmony* (github.com/immunogenomics/harmony) are suitable for integrating multiple single-cell datasets ([Korsunsky et al., 2019](#)).

In addition, dimensionality reduction techniques such as Principal Component Analysis (PCA) followed by t-distributed Stochastic Neighbour Embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP) are used to analyse and visualize of the single-cell data. UMAP is the most commonly used for visualization because it has faster run times, consistency and meaningful organization of cell clusters. After clustering, the final part of the analysis is the biological annotation of each group or cluster. Multiple analyses can be performed, such as the differential analysis of cell types or genes between groups, cell trajectories, functional enrichment, transcription factor identification, cell-cell interaction and cell marker analysis. In summary, scRNA-seq is a powerful tool to profile, identify, classify and discover novel or rare cell types and subtypes in different human organs and tissues. All this makes single-cell technology a very promising application that could be integrated into clinical diagnosis and in cell-specific characterization of diseases ([Jovic et al., 2022](#)).

3. CHARACTERIZATION OF MESENCHYMAL STEM CELLS: GENES AND CELLS

a. Novel markers of Mesenchymal Stem Cells

The identification of new MSC gene markers has been a major challenge in recent years. Since the ISCT statement in 2006, many researchers have been working in this field due to the interest in using these cells in many clinical trials. However, despite the consensus on minimum criteria for the definition of MSCs given by the ISCT (Dominici et al., 2006), the practical conditions for the identification and isolation of human MSCs remain the same and most laboratories continue to use the three positive gene markers: CD73 (NT5E), CD90 (THY1) and CD105 (ENG). At the same time, many recent studies conducted on hMSCs include relevant discoveries about the nature of these stem cells and claim a great need to identify novel and more specific gene markers.

The search for specific markers for MSCs, these can be categorized into two groups: singular markers and stemness markers. A singular marker, which allows positive MSC selection, replaces the conventional gene markers and being alone sufficient to identify or purify MSC-like cells from their *in vivo* environment. On the other hand, a stemness marker has the ability to identify a subset of MSCs with robust CFU-F potential (i.e., clear signal in the *colony-forming unit–fibroblast* assay to form adherent colonies under culture conditions) and the capacity to differentiate. These stemness markers can also, in certain cases, identify embryonic stem cell-like populations, facilitating the enrichment of subpopulations that exhibit superior CFU-Fs and multipotency. Considering the nature of these two distinct types of markers, the singular markers are usually highly expressed, while the stemness markers may be moderately detectable.

In recent years, several candidate surface markers for MSCs have been proposed as being closely related to the stemness capacity of these cells. These markers include: CD146 (official gene symbol MCAM), CD271 (official gene symbol NGFR), SSEA-4 (that is a surface glycosphingolipid), and Stro-1 (a gene detected in mice). The protein Nestin (gene NES) has also been proposed as an intracellular marker. In general, these markers have been reported to have a large difference in their expression between different MSCs isolated from different sources (Mabuchi et al., 2021).

CD146 (known as melanoma cell adhesion molecule (MCAM)) is a pivotal cell adhesion protein involved in vascular endothelial cell activity and angiogenesis. Notably, CD146 has emerged as a promising candidate for the identification of MSCs. CD146 has a higher CFU-F enrichment capacity than PROM1, THY1, or Stro-1 and also delineates MSCs with higher multipotency. In addition, CD146 also identifies MSCs with higher supportive capacity for haematopoiesis (Harkness et al., 2016). CD271 (also known as the low affinity nerve growth factor receptor (NGFR)) functions as a receptor for neutrophils, which promote the survival and differentiation of neuronal cells. CD271 has been used to select CFU-Fs from bone marrow mononuclear cells (BM-MNCs). CD271-positive BM-MSCs have demonstrated an enhanced ability to promote HSC engraftment. Furthermore, CD271 plays a significant role in maintaining the clonogenicity and functionality of MSCs.

SSEA-4 is a glycosphingolipid that has been identified as a good marker for embryonic stem cells. It has been documented that SSEA-4 can be used to effectively isolate MSCs from bone marrow. In addition to BM-MSCs, SSEA-4 expression has also been detected in MSCs derived from placenta, periodontal ligament, dental pulp, and synovial membrane. In contrast, MSCs derived from adipose tissue, umbilical cord (UC), or umbilical cord blood do not express SSEA-4. It has also been reported that SSEA-4 is not present in unsorted bone marrow samples. Stro-1 is a cell membrane single-pass type I protein, with a molecular mass of 75kD, that translocate from the endoplasmic reticulum to the cell membrane in response to the intracellular calcium depletion. Stro-1 has been defined as a protein marker of Mesenchymal Stem Cells; however, the presence of Stro-1 did not yield in significant CFU-Fs from human

endometrial stroma. Finally, Nestin (NES) is defined as a class VI intermediate filament protein. Some investigations have found that NES-positive MSCs in the bone marrow contain the full range spectrum of the mesenchymal progenitor activity (CFU-Fs) and have the capacity for self-renewal and trilineage differentiation. NES-positive MSCs were found to colocalize with HSCs, thereby supporting HSC maintenance and homing. However, challenges remain in the use of NES for the cell isolation, culture, and clinical therapy due to its intracellular expression (Xie et al., 2015).

Despite the discovery of these potential novel marker genes, there is no consensus opinion and the available reports supporting these marker genes are not consistent enough. The culture conditions in which MSCs are grown have the potential to influence their phenotype, contributing to the conflicting reports on the identification of specific expression markers. Furthermore, specific growth factors and cytokines, such as fibroblast growth factor (FGF) and interferon-gamma (IFN- γ), and disease conditions, such as inflammation, may also modulate the phenotype of MSCs.

b. Cellular heterogeneity: different populations or states of Mesenchymal Stem Cells

The therapeutic potential of MSCs has generated great interest in the search for specific gene markers that define these stem cells well. However, advances in genome-wide cellular studies and recent discoveries on stromal cells reveal the existence of heterogeneous cell populations within the MSCs, contributing to a still ambiguous characterization of specific markers to well identify them. Technological advances, such as scRNA-seq, have allowed substantial progress in this area, facilitating the identification of subtle variations and changes in the genome-wide expression profile of MSCs.

Zhoungyu Xie and colleagues found that the heterogeneity of MSCs can be deconstructed into three main groups: stemness, functional, and proliferative subpopulations. They found that the stemness subpopulation highly expressed the genes SOX4, GAS1 and CD26 (Xie et al., 2022). CD26 (gene DPP4) is a senescence marker that has been associated with reduced immunopotency of human multipotent stromal cells. The trajectory analysis in this study showed a clear segregation of a stemness subpopulation from the main functional and proliferative subpopulations. Furthermore, the study demonstrated that the functional subpopulation had stronger capabilities related to immune regulation and osteogenic differentiation, but lower potential for adipogenic differentiation and proliferation. The functional subpopulations presented increased expression of genes such as CCL2 and TGF- β ; as well as, the genes IGFBP2, PTX3, GREM1, and CTGF, which are related to various regulatory processes and multipotent differentiation capabilities of MSCs. The proliferative subpopulation presented numerous genes associated with G1/S and G2/M phases of cell division (Xie et al., 2022). A similar study also classified MSC populations, in this case derived from adipose tissue, bone marrow aspirates, foreskin samples and umbilical cords. They classified MSCs into three main groups: associated with extracellular matrix, associated with cell cycle regulation, and associated with cytokines and chemokines. The genes associated with the ECM subgroup had particularly immunomodulatory properties. These investigations also revealed a lack of intrinsic variability among the bone marrow healthy donors, enhancing the robustness of the results (Z. Wang et al., 2021). Additionally, an *ex vivo* study of MSCs also unveiled five subpopulations, including a stem-like APC subpopulation, a multipotent progenitor subpopulation, a specific adipocyte precursor subpopulation, a specific osteochondrocyte precursor subpopulation, and an immunoregulatory prechondrocyte subpopulation. The multipotent progenitor and immunoregulatory prechondrocyte subpopulations were those that showed the highest expression signal in BM-MSCs, with overexpression of CXCL12 or ANGTP1 in the *ex vivo* samples. However, the expression activity observed in the HSC niche was partially lost in the *in vitro* cultures, which presented high levels of expression in prechondrocytes associated with immunoregulatory effects (C. Zhang et al., 2022).

These studies reveal a certain diversity within the MSCs, providing different transcriptomic profiles that complicate the identification of specific and global MSC markers. Furthermore, these studies provide quite valuable information about the variability of the expression profiles of MSCs, and about the specific functions of different subpopulations. Collectively, all of these studies open up a novel way to understand the role and nature of MSCs at a deeper level, thereby enhancing the value of omic-wide and single-cell specific technologies for the field of Mesenchymal Stem Cell research.

CHAPTER 1:

Comparative analysis of expression profiles of human Mesenchymal Stem Cells to identify novel and specific gene markers

Short Title:

Specific gene markers for BM-MSCs

1. MATERIAL AND METHODS

a. Identification of gene datasets

In this first chapter, we have performed a comprehensive characterization of Mesenchymal Stem Cells with independent analysis of three different types of transcriptomic data: Microarrays, Exon Arrays and RNA-seq.

The data was processed using R statistic programming language (<https://www.r-project.org/about.html>) through the software *RStudio* (<https://www.rstudio.com/>). We used Bioconductor repository, an open platform (<https://www.bioconductor.org/>) that provides tools for analysis and high performance compression of genomic data, to download the required libraries (Huber et al., 2015).

The Microarrays analysis was a compendium of 264 samples from 18 data batches, comprising over 500 samples, previously designed previously in our laboratory. The platforms used for this analysis were: HG-U133A & HG-U133B and HG-U133 Plus 2.0 (Roson-Burgo, Sanchez-Guijo, Del Cañizo, et al., 2016). In **Table 1**, it was presented raw metadata, the groups included were: Hematopoietic Stem Cells, Hematopoietic Differentiated Cells (Lymphocytes) (LYM), Mesenchymal Stem Cells, Stimulated MSCs (stMSCs), MSC-derived Adipoblasts (ADIP), MSC-derived Chondroblasts (CHON), Osteoblasts (OSTB), MSC-derived Osteoblasts (dOST), Stimulated Osteoblasts (stOST) and Skin FIB. We then filtered a subset of 70 samples with 10 HSC, 49 MSCs and 11 FIB, all were from the i HG-U133 Plus 2.0 platform. All the Microarrays were obtained from Gene Expression Omnibus (GEO) platform, which is a public repository from the National Center for Biotechnology Information (NCBI) (Barrett et al., 2012).

The package used was *affy*, a specific method for analyzing Affymetrix Oligonucleotide Arrays, and we use Robust Multi-Array Average expression measure (RMA) function to normalize and reduce the batch effect of the datasets (Gautier et al., 2004). The result contained 16698 genes and 70 samples.

For the Exon Arrays analysis, data from the same previous work of our laboratory were also provided. The data are available in the GEO database under the identifier GSE72332. The dataset includes a total of 15 samples, including 3 samples of AD-MSCs, 3 samples of BM-MSC, 3 PL-MSC, 3 samples of HSC and 3 samples of dermal FIB; under platform Human Exon 1.0 array from Affymetrix. The data were analyzed following the same metadata analysis approach as previously described (Roson-Burgo, Sanchez-Guijo, Del Cañizo, et al., 2016). The resulting matrix obtained contains 39203 genes and 15 samples.

Table 1. Meta-Analysis of 264 Microarrays from 10 cell types: HSCs, LYM, MSCs, stMSCs, ADIP, CHON, OSTB, dOST, stOST and FIB.

Cell type	Annotation	N. Microarrays	BM-Microarrays	Filtered BM-Microarrays
Haematopoietic Stem/Progenitor Cells	HSCs	47		10
Haematopoietic Differentiated Cells (Lymphocytes)	LYM	9	56	
Mesenchymal Stem Cells	MSCs	116		49
Stimulated MSCs	stMSCs	27	143	
MSC-derived Adipoblasts	ADIP	3	3	
MSC-derived Chondroblasts	CHON	3	3	
Osteoblasts	OSTB	13		
MSC-derived Osteoblasts	dOST	12		
Stimulated Osteoblasts	stOST	23	48	
Skin Fibroblasts	FIB	11	11	11
Total		264	264	70

Finally, the RNA-seq analysis was performed on 29 samples, including 15 samples of HSC, 6 samples of BM-MSCs and 8 samples of FIB. To start the analysis, we extracted the raw *fastqc* of these samples from the GEO identifiers: GSE105145, GSE81478, GSE102881, GSE114922, GSE63569, GSE51518 and GSE119501. These files were obtained from Illumina HiSeq, with samples paired-end and single-end samples as it is indicated in **Table 2**. Next, we converted the *fastqc* to counts using the *Salmon* tool. *Salmon* is a tool compiled in Linux that is capable of estimating transcript abundance from RNA-seq data, as we commented in the Introduction. In our analysis, the *fastqc* were aligned to the reference genome of Homo sapiens GRCh38.79 (Patro et al., 2017). Normalization was performed by count per million (*cpm*) using the *edgeR* package and scaling to logarithm 2 with function $\log_2(\text{cpm}+1)$, yielding a matrix of 33290 genes and 29 samples (Robinson et al., 2010).

b. Gene markers selection

The selection of the 151 gene markers was derived from a combination of the bibliography and the results of the two previous projects designed in our laboratory (Roson-Burgo et al., 2014; Roson-Burgo, Sanchez-Guijo, Del Cañizo, et al., 2016). In addition, this compilation included all the known biomarker genes of MSCs, as well as, those associated with other cell types present in the microenvironment where the MSCs reside, including HSCs, lymphocytes, monocytes, endothelial cells and Fibroblasts. For further details and a comprehensive list of these genes along with their descriptions, we have included *Supplementary Information 1* in ANNEX I.

Table 2. RNA-seq of 29 samples from 3 different cell types: MSC, HSC and FIB. The samples are provided from 8 different datasets as indicate the batch and the GSE. The samples have been sequencing with Illumina Hiseq in different versions, as indicate the column of the sequencing and have different types of sequencing, being: single- or paired-end.

Cell type	Batch	Cell type Batch	GSE	GSM	Sequencing	Reads
MSC	1	BM.MSC.1	GSE105145	GSM2823171	Illumina Hiseq 2000	PAIRED
	1	BM.MSC.1.2	GSE105145	GSM2823172		
	1	BM.MSC.1.3	GSE105145	GSM2823173		
	2	BM.MSC.2	GSE81478	GSM2154690		
	2	BM.MSC.2.1	GSE81478	GSM2154691		
	2	BM.MSC.2.2	GSE81478	GSM2154692		
HSC	3	BM.HSC.3	GSE102881	GSM2747682	Illumina Hiseq 4000	PAIRED
	3	BM.HSC.3.1	GSE102881	GSM2747683		
	4	BM.HSC.4	GSE114922	GSM3167340		
	4	BM.HSC.4.1	GSE114922	GSM3167292		
	4	BM.HSC.4.2	GSE114922	GSM3167321		
	4	BM.HSC.4.3	GSE114922	GSM3167324		
	4	BM.HSC.4.4	GSE114922	GSM3167327		
	4	BM.HSC.4.5	GSE114922	GSM3167351		
	4	BM.HSC.4.6	GSE114922	GSM3167354		
	4	BM.HSC.4.7	GSE114922	GSM3167361		
	6	BM.HSC.6	GSE63569	GSM1552801		
	6	BM.HSC.6.1	GSE63569	GSM1552802		
	6	BM.HSC.6.2	GSE63569	GSM1552803		
	6	BM.HSC.6.3	GSE63569	GSM1552804		
	6	BM.HSC.6.4	GSE63569	GSM1552805		
	FIB	7	FIB.7	GSE51518		
7		FIB.7.1	GSE51518	GSM1246807	SINGLE	
7		FIB.7.2	GSE51518	GSM1246808	SINGLE	
7		FIB.7.3	GSE51518	GSM1246809	SINGLE	
7		FIB.7.4	GSE51518	GSM1246811	SINGLE	
8		FIB.8	GSE119501	GSM3375722	PAIRED	
8		FIB.8.1	GSE119501	GSM3375723	PAIRED	
8		FIB.8.2	GSE119501	GSM3375724	PAIRED	

c. **Global Test analysis**

Global test is a statistical package based in testing groups of features for significant association with a response variable (that is, test all features, that in our case are the genes, as candidate markers associated with a given subset of samples, that in our case are the different cell types). The algorithm works by employing a regression model to adjust by selecting the best features in the role of the proposed markers, based on significant associations or correlations with different classes of samples. The results obtained include a statistical test and the corresponding p-value, which indicates the strength of the association between the selected features (the selected genes) and the output state (the sample subtype or cell type). There are two different plots in the order of the results analyzed, being the outputs the correlation of genes or the samples (Goeman et al., 2004):

- The covariates plot: The covariates plot visualizes the influence of the covariates (genes) on the test result. The global test statistic is a weighted average of the global test statistics applied in each gene. The global test assigns relative weights to each covariate (gene) based on its contribution to the test result. These weights are proportional to the residual variance, which is represented by a value 'y' on the regression curve. A higher value of "y" indicates a stronger association with a particular cell type and such genes are further away from the regression line (corresponding to the mean of the gene weights). The weight ratio is scaled so that the maximum correlation is 1. Bars and stripes indicate the mean and standard deviation, respectively, under the null hypothesis, where the relative weight is equal to 0.

Genes are ordered in a hierarchical cluster based on absolute correlation distances, so if the covariates (genes) are highly correlated, will be ordered next to each other. In addition, the blond portion of the dendrogram indicates that genes have a significant multiplicity-correlated p-value, suggesting a confident association with the response (cell type) in this case.

- The subject plot: The subject plot visualizes the influence of the subjects (samples) on the test result. The methodology is the same as the covariates plot, but in this case the statistic and bars are plotted over the samples.

d. **Confusion Matrices test**

The confusion matrix, considered in statistics as the error matrix, is a specific matrix where each row represents the instances in a current class while each column represents the instances in a predicted class. There are two conditions: positive condition (P), which is the number of real positive cases in the data, and negative condition (N), which is the negative ones. In our case, the positives correspond to MSCs cell type and the negatives correspond to No MSCs, including fibroblasts and hematopoietic stem cells.

In the case of true positives (TP), the result is the presence of the condition and the true negative (TN) is the result of the absence of the condition. A false positive, also called Type I error (FP), is when the result falsely indicates the presence of the condition and a false negative, also called Type II error (FN), is when the result falsely indicates the absence of the condition (Stehman, 1997).

In the confusion matrices there are specific values to evaluate the results:

- The Sensibility or true positive rate (TPR): The sensibility describes the accuracy of the tests in indicating the presence of the condition. It is the probability of a positive test result. The function of this parameter is $TPR = TP / TP + FN$
- The Specificity or true negative rate (TNR): The specificity describes the accuracy of the tests that indicate the absence of the condition. It is the probability of a negative test result. The function of this parameter is $TNR = TN / TN + FP$
- The precision or positive predictive value (PPV): is the proportion of relevant instances among the retrieved instances, that apply to the data retrieved from all the samples. The function of this parameter is $PPV = TP / TP + FP$
- The fall out or false positive rate (FPR) is the expectation of the false positive rate. It is the probability of falsely rejecting the null hypothesis in the test. The function of this parameter is $FPR = FP / FP + TN$
- Miss rate or false negative rate (FNR) is the ratio of false negatives to the sum of false negatives and true positives. The ideal situation is when the model produces 0 false positives and 0 false negatives. The function of this parameter is $FNR = FN / FN + TP$

e. Human cell samples to test the markers

To validate the bioinformatic analysis performed in the study, an experimental study was performed in our laboratory. Samples from different human cell types were obtained from different sources, as described in **Table 3**. We obtained 3 samples of Mononuclear Cells (MNC) from peripheral blood; 4 samples of Mesenchymal Stem Cells isolated from BM; and 3 samples of Mesenchymal Stem Cells isolated from adipose tissue. All these were obtained from healthy donors. Furthermore, 3 samples of human primary Fibroblasts from dermis (from 3 different female donors), were purchased from Innoprot (<https://innoprot.com/>); and two Stromal Cell (SC) lines mentioned below. All samples from donors obtained at the University Hospital of Salamanca (MNCs, BM-MSCs at AD- MSCs) were collected after the appropriate informed consent and in accordance with the Ethical Standards and Good Clinical Practice established by the Ethics Committee of the University Hospital of Salamanca.

Table 3. Characteristics of the human samples of different cell types that are used to test the expression of genes markers (the samples are coming from healthy donors, D1-D14, or from cell lines, L14 & L15).

Donors/Lines	Cell Type	Cells Source	Age	Gender
D1	MNC	Peripheral Blood	18	Female
D2	MNC	Peripheral Blood	32	Male
D3	MNC	Peripheral Blood	20	Male
D4	FIB	Dermis	22	Female
D5	FIB	Dermis	38	Female
D6	FIB	Dermis	34	Female
D7	MSC	Bone Marrow	24	Female
D8	MSC	Bone Marrow	38	Male
D9	MSC	Bone Marrow	60	Male
D10	MSC	Bone Marrow	38	Male
d11	MSC	Adipose tissue	31	Female
D12	MSC	Adipose tissue	49	Female
D13	MSC	Adipose tissue	40	Female
L14 (HS-5)	SC	HS-5 cell Line	–	–
L15 (HTERT)	SC	hTERT cell line-AD	–	–

f. Cell processing

Mononuclear Cells were isolated from peripheral blood: 5ml of peripheral blood was collected in EDTA tubes from 3 healthy donors. Peripheral blood was lysed twice with 1X ammonium chloride for 20min and 5min respectively in cold, after incubation washed twice with Phosphate-Buffered Saline (PBS). The pellet was mixed with Trizol and RNA was extracted.

Fibroblasts: Cryopreserved fibroblast cells from 3 different adults (purchased from Innoprot) were thawed and seeded at the rate of 5000 cells/cm² in a culture flask with Fibroblast medium supplemented with 2% Fetal bovine Serum (FBS), 1% of Fibroblast Growth Supplement (FGS) and 1% of penicillin/streptomycin solution until confluence was reached. At confluence, the cells were trypsinized and subcultured to obtain the required number for further experiments.

Bone marrow harvest: Bone marrow harvest was performed under local anesthesia, by puncturing the posterior-superior iliac spine. Bone marrow Mononuclear cells were collected by density gradient with Ficoll-PaqueTM Plus (density:1.077k, GE Healthcare Bio-Sciences AB). They were centrifuged for 30 minutes at 500g according to the method previously described by Minguell et al. (Minguell et al., 2001). Finally, the interphase mononuclear cells were washed twice with HBSS medium (Hank's Balanced Salt Solution with Phenol Red, BioWhittaker Lonza Verviers, Belgium) for 10 minutes at 300g.

Adipose tissue procurement: Human adipose tissue was obtained by liposuction under general anesthesia. Briefly, 1 gram of adipose tissue from the liposuction was digested and incubated with collagenase at 37°C for one hour with shaking. The homogenized material was then passed through a 40µm filter and centrifuged. After cell button lysis with ACK 1X lysis Buffer (A10492 Gibco, Invitrogen) mononuclear cells were washed twice with PBS (Zuk et al., 2001).

Isolation and expansion of MSC from BM and Adipose tissue: The mononuclear cells obtained from BM and the cells from adipose tissue were seeded at a rate of 1x10⁶/cm² in Dulbecco's Modified Eagle's medium-low glucose (DMEM, Gibco Invitrogen Paisely, UK) supplemented with 10% Bovine Serum Fetal (SBF, BioWhittaker[®] Lonza, Verviers, Belgium) and 1% Penicillin/Estreptomycin (Gibco, Invitrogen,

Paisely, UK). The culture medium was changed every 2-3 days until 80-90% confluence was reached. At this time, the entire culture medium was removed from the flask, washed with sterile PBS (GIBCO Invitrogen Corporation, Paisely, UK) and incubated with 0.05% 1X trypsin (GIBCO Invitrogen Corporation, Paisely, UK) for 5 minutes at 37°C. Cells were then seeded at a concentration of 5,000 cells/cm² into larger surface culture flasks until they reached passage 3, which was used for this study (Muntión et al., 2012).

Stromal cell lines: We used two immortalized human stromal cell lines (hTERT and HS-5), both grown in mesenchymal medium (DMEM, 10%FBS and 1% P/S). The first cell line (hTERT) is an adipose human MSC line immortalized by expression of the telomerase reverse transcriptase gene. It was generously provided by Dr. D Campana (from the Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore); while the HS-5 cell line was derived from bone marrow stroma (HS-5, ATCC® CRL-11882™, <https://www.atcc.org/products/crl-11882>).

g. Quantitative gene expression measurements

In this study, RNA was extracted from all collected cells using the thiocyanate-phenol-chloroform method (Chomczynski & Sacchi, 2006). The Agilent 2100 Bioanalyzer system (Agilent, Palo Alto, CA) was used to measure RNA concentration and integrity. Reverse transcription was performed according to Van Dongen JJ (van Dongen et al., 1999), using the High-Capacity Kit (Applied Biosystems, Foster City, CA, USA). For qPCR, the Step One Plus Real-Time PCR System (Applied Biosystems, Foster City, CA) was used with commercial TaqMan® Gene Expression Assays (Applied Biosystems, Foster City, CA). The genes analyzed are listed in **Table 4**, along with the identification number of each gene tested. To quantify the expression of each target gene relative to the control gene (GADPH), we calculated the formula $2^{-\Delta Ct}$, where the Ct is the cycle threshold, where: $\Delta Ct = Ct \text{ study gene} - Ct \text{ GADPH}$. The differences between the expression signals obtained with the different cell types were analyzed to identify the significant differences (p-values <0.05) using a non-parametric Mann-Whitney U test between samples. This statistical test allowed for the detection of significant variations in gene expression levels between the different cell types studied.

Table 4. Human genes tested by RT-PCR to quantify expression in different human cells. The Catalog number and the Entrez NCBI ID are provided to check gene identity and sequence information.

CATALOG Nº	NCBI ID	GENE SYMBOL	GENE FULL NAME - DESCRIPTION
hs00977641_m1	214	ALCAM(CD166)	Activated-Leukocyte Cell Adhesion Molecule (CD166)
hs00266237_m1	1282	COL4A1	Collagen Type IV Alpha 1 Chain
hs05006309_m1	1284	COL4A2	Collagen Type IV Alpha 2 Chain
hs00923996_m1	2022	ENG (CD105)	Endoglin (CD105)
hs99999905_m1	2597	GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
hs00235006_m1	3672	ITGA1 (CD49a)	Integrin subunit alpha 1 (CD49a)
hs01573922_m1	4907	NT5E (CD73)	5'-nucleotidase ecto (CD73)
hs00162558_m1	6876	TAGLN	Transgelin
hs00174816_m1	7070	THY1 (CD90)	Thy-1 Cell Surface antigen (CD90)
hs00221277_m1	57758	SCUBE3	Signal Peptide, CUB Domain & EGF Like Domain Containing 3

2. RESULTS

a. Quality Control of the datasets

As is indicated in the Materials and Methods, we performed three independent experiments with different types of data: Meta-analysis of Microarrays, Exon Arrays and RNA-seq data. **Figure 12**, showed the quality control of the samples using density plots and PCA plots. The density plots of the **Figure 12A** and **Figure 12B** corresponded to the typical signal of the Affymetrix data whereas **Figure 12C** had a first peak with the high number of zeros and a second peak with the signal of the genes in the samples. This observed pattern in **Figure 12C** was common and expected as it is a characteristic of gene expression data in RNA-seq. Cell types or tissues express only a subset of the transcripts present in the genome. As a result, many genes are not expressed or have negligible expression levels, resulting in a large number of zeros in the data. The second peak corresponds to the expression levels of genes that are actively expressed in the samples.

In the PCAs shown in **Figure 12A** each plot represents the three main groups considered in this chapter: MSCs, HSCs and FIB. We could see a clear and distinct difference between MSC and HSC while the FIBs remained in the middle of the plot. The **Figure 12B** represented the 15 samples, including the AD-MSCs, BM-MSC, PL-MSCs and FIB. In this plot, there were not present differences in the distribution of the space while the FIB was more separated. Finally, **Figure 12C** referred to RNA-seq, where there were three groups: MSCs, HSCs and FIB clearly separated except for 3 samples of MSCs. These results showed that the PCA plots effectively represented and distributed the samples based on their cell types, allowing for accurate analysis and minimizing the risk of false positives. The separation between different cell types in the plots indicated that the gene expression data captured the differences between the groups, facilitating reliable downstream analysis and interpretation in the study.

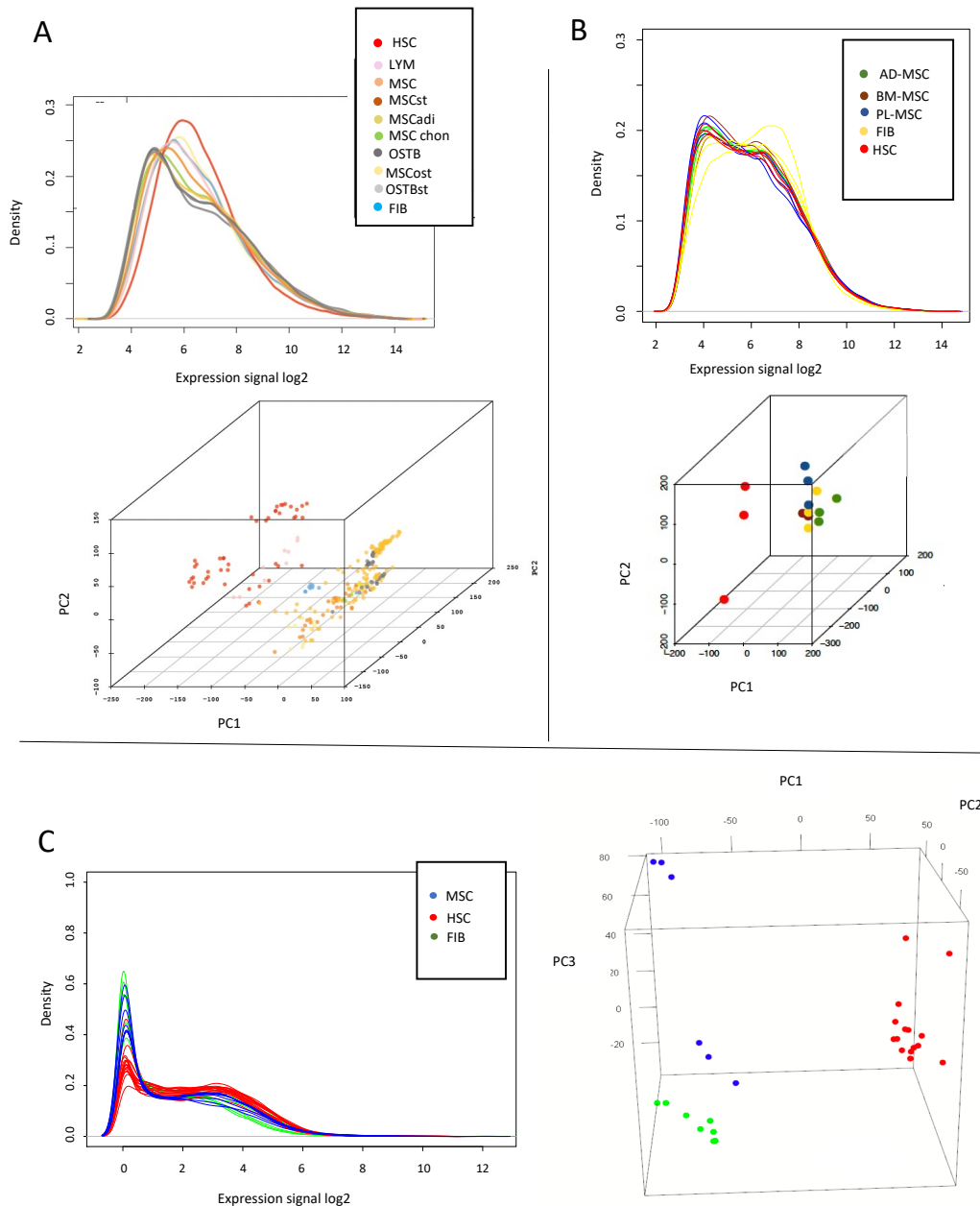


Figure 12. Density and PCA 3D plots of three types of datasets. **A)** The 264 samples of Meta-Analysis from Microarrays data with 10 different cell types. **B)** The 15 samples of Exon Arrays in 5 different cell types. **C)** The 29 samples of RNA-seq with 3 different cell types.

b. Expression signal of data sets in 151 genes

In this section of the results, we performed boxplots to analyze the distribution of our footprint of 151 genes. The Boxplots were done for the Meta-analysis of Microarrays, the Exon Arrays and the RNA-seq data. Due to the large number of samples in Microarrays (70 samples) and RNA-seq (29 samples), the plots were not conclusive and clear. However, although the results were similar across the different data types. The **Figure 13** showed the logarithm 2 of expression signal of our markers in Exon Arrays. In **Figure 13**, the first three genes detected were NT5E (CD73), THY1 (CD90) and ENG (CD105).

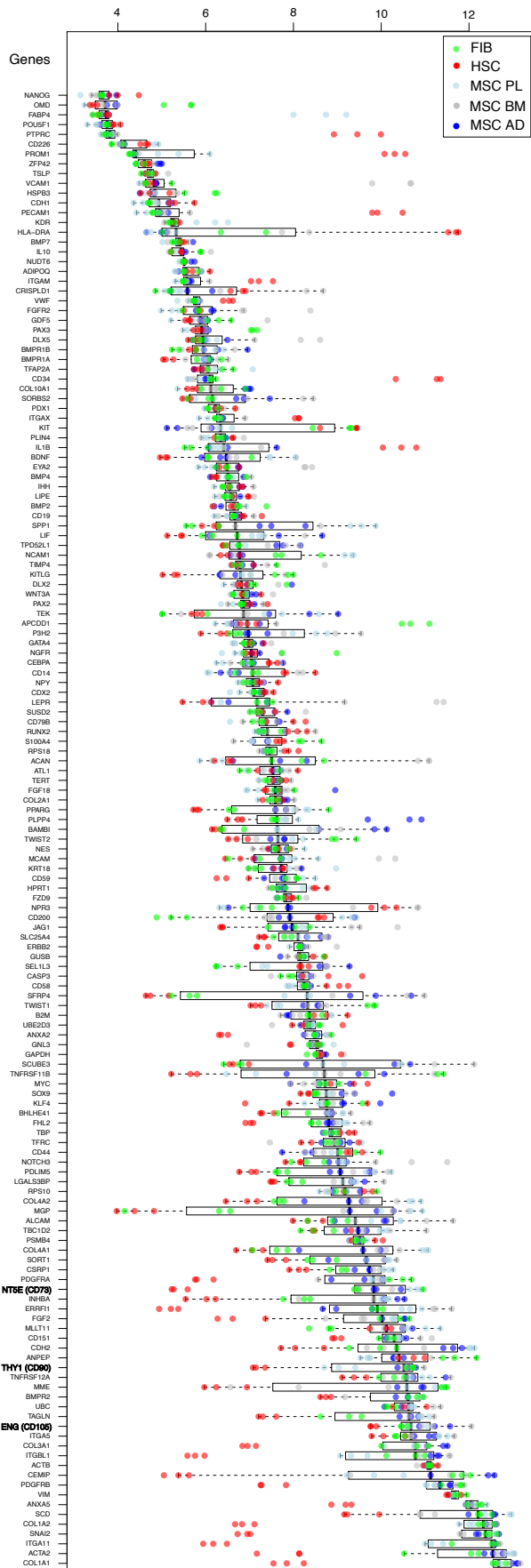


Figure 13. Boxplot of the 151 gene markers in the Exon Array data set with 3 samples of BM- MSC, 3 samples of AD-MSC, 3 samples of PL-MSC, 3 of HSC and 3 of FIB. The scale of the gene expression was in logarithm 2.

These three genes are the identifiers selected by the ISCT as Minimal criteria for defining multipotent Mesenchymal Stem Cells (Dominici et al., 2006). All of them were decreased in HSC and present a high expression in MSCs, even if they could not be separated from FIBs. Continuing with the ISCT criteria: HLA-DRA, CD34 and PTPRC (CD45) markers should be negative for MSCs, as it shows the **Figure 13**, these genes were clearly decreased in MSCs and FIBs while presenting a high expression signal in HSCs. The genes with the highest signal were: CEMIP, PDGFRB, VIM, ANXA5, SCD, COL1A2, SNAI2, ITGA11, ACTA2 and COL1A1; all of them with a decrease in HSC samples, but without variation between MSCs and FIBs. In contrast, genes such as: PROM1, KDR, CD226 and IL1B showed increased expression in HSCs compared to FIBs and MSCs. To find the best markers for the BM-MSCs, we identified 18 genes as the most promising candidates: EYA2, NGFR, ACAN, BAMBI, NES, MCAM, SFRP4, FHL2, PDLIM5, LGALS3BP, MGP, SORT1, CSRP1, SCUBE3, COL4A1, COL4A2, INHBA and TAGLN. These genes showed the best results in the analysis of BM-MSCs. For the FIB cell type, TWIST1 gene was clearly the best candidate, TNFRSF11B, MME, ANPEP, TWIST2 and OMD were also identified as relevant genes in the analysis.

c. *Global Test of the 151 MSCs candidate markers*

As is mentioned in Materials and Methods, the 151 gene markers were correlated with three different types of datasets: Meta-Analysis of Microarrays, Exon Arrays and RNA-seq data. Using the *Global Test* algorithm, we correlated the samples (subjects) based on the gene signature, which allowed for the correct classification of the samples according to their corresponding cell type. Furthermore, the algorithm was used in reverse to correlate the genes (covariates) based on the samples.

In the **Figure 14**, we see the *Global test* plots of the samples of the three different types of datasets. These subject plots were useful to identify the samples that have a significant influence on the result and to find samples that deviate from the main pattern. The main point of this figure was a good assignment of the samples to each of the three groups: MSCs, HSCs and FIB. In the case of Exon Arrays analysis and RNA-seq (**Figures 14B** and **14C**), the MSCs and FIB samples were associated with in the main branches of the dendrogram, which was expected due to their similarities. However, in the case of Meta-Analysis of Microarrays (**Figure 14A**) the FIBs were associated with the HSCs. Nevertheless, the ability to categorize the samples into the cell types was relevant, as indicated by the number of stripes corresponding to the number of standard deviations from the mean. It is important to note that the value of the weighted posterior effect in the case of Exon Arrays and RNA-seq is in logarithmic scale, which can lead to an apparent difference between the cell types in the distribution of the Meta-Analysis of the Microarrays. Finally, we can see in the Exon Array subjects plot, MSCs were distributed in three branches in the dendrogram, which correspond to the PL-MSCs, BM-MSCs and AD-MSCs.

Otherwise, the **Figure 15**, showed the covariates plot, which indicated the 151 gene markers used to identify the samples based on their cell types. In this case, each individual gene was associated with one of the cell types with a weighted test statistic value. **Figure 15A** and **15C**, corresponding to the Microarrays and RNA-seq datasets, respectively, showed more significant genes than **Figure 15B**, corresponding to the Exon Arrays. This is shown in the blond lines of the dendrogram, which select genes with a significant p-value (>0.05). In these plots, the size of the samples was taken into account to improve the correlation with the genes. This was the cause of the large increase of MSCs genes in the first two datasets and not in the third one, which contains 15 samples of HSCs. Despite this, the results presented a clear distribution of the genes associated with the cell type.

The **Figure 15A** had the highest cluster of MSCs: VCAM1, ERFF1, LGALS3BP, PDLIM5, COL4A1, COL4A2, INHBA, ALCAM, FHL2, ACTA2 and TAGLN, SCUBE3. TAGLN was one of the most significant gene with 16 orders of standard deviation from the mean and the most weighted value= 23.19, also were good indicators COL4A2 and a COL4A1 with weighted T-statistic of 18.56 and 20.77 and numbers of Standard

Deviation (SD) of 18 and 19, respectively. Moreover, SCUBE3 presents a weighted T-statistic of 9.15 and a number of SD of 13, as shown in **Table 5**. The standard MSCs genes presented by ISCT: NT5E (CD73), THY1 (CD90), and ENG (CD105) presented weighted T-statistics of 9.70, 8.74 and 8.10 and a number of SD of 10, 8 and 12, respectively, lower than our presented genes, indicating that the presented genes are better indicators of MSCs. For the genes associated with the HSCs, the cluster formed by: CD34, KDR, IL10, NANOG, IL1B, ITGAX, CDX2 and FGF18 was the most relevant, also considering the gene PTPRC (CD45) with a weighted T-statistic of 4.17 and number of SD of 11 that in combination of CD34 with a weighted T-statistic of 2.01 and number of SD of 8 were well labeled indicators of HSC. The association of genes with FIBs was less significant, but genes such as TWIST1 and S100A4 were identified as relevant indicators.

In the case of **Figure 15B** shows two different clusters associated with MSCs in Exon Arrays. The first cluster included genes such as: TAGLN, CSRP1, ACTA2, PDGFRA, COL1A1 and COL1A2, while the second cluster contained genes such as: FRP4, SLC25A4, SCUBE3, ENG, PDLIM5, SORT1, BHLHE41, MGP, COL4A2 and COL4A1. TAGLN remained as one of the most significant gene associated with 5 orders of standard deviation from the mean and weighted value= 30.19. COL4A2 and a COL4A1 with weighted T-statistic of 26.39 and 27.95, respectively an both with number of SD of 5. In case of HSCs, the weighted T-statistical power was lower, but as in **Figure 15A** appears KIT or MYC genes with high weighted T-statistical. However, the cluster associated had not significant value. The same occurs with FIB cell type, with two high statistical weight in genes: S100A4 and ANPEP but not significant p-value.

In **Figure 15C**, the genes associated with MSCs had lower statistical power. However, there was a cluster of genes: COL4A2, COL4A1, BDNF, NT5E, NOTCH3, ERRF1, SOX9, TAGLN and ACTA2, which had a high T-statistical weight and also significant p-value. TAGLN gene remained one of the most significant gene associated with 12 orders of standard deviation from the mean and weighted value= 31.83. Also presented high levels in the test COL4A2 and a COL4A1 with weighted T-statistic of 26.61 and 23.50 and numbers of SD of 12 and 11, respectively.

Moreover, the genes of the second cluster: TBC1D2, SORT1, BMP2, ALCAM, MCAM, PDLIM5, INHBA and SCUBE3 presented more than 5 stripes of standard deviation and weighted statistical test of 7.68, making them relevant in the analysis, as shown in **Table 5**. In the HSC cell type, the graph showed a large number of relevant genes, such as GAPDH, RPS18 and B2M with the most statistical power. Again, we saw the cluster of genes with: IL10, PTPRC, ITGAM, IL1B, ITGAX, MYC and KIT. Finally, in the FIB cell type, the cluster presented shows and increment number of genes and statistical power in: COL1A2, PDGFRA, TWIST1, SNAI2, ERBB2, PDGFRB, CD151, TNFRSF11B, ANXA5, CEMIP and ANXA2.

Based on the overlay of data analysis presented in **Figures 13, 14 and 15**, and considering the expression data and weighted test statistics from the *Global Test* in the Meta-Analysis of Microarrays, Exon Arrays, and RNA-seq data, the most relevant genes for differentiating BM-MSCs from HSCs and FIB were TAGLN, SCUBE3, COL4A1 and COL4A2. Furthermore, as we can see in **Table 5**, our proposed markers had a better weighted test statistic (as also shown in **Figures 15A, 15B and 15C**), which is reflected in these plots with a higher number of ticks (representing the deviations from the SD that is marked with a thicker tick for each gene, i.e. number of SD) compared to the standard markers: NT5E, THY1 and ENG (with the exception of the SCUBE3 gene in the RNA-seq dataset, which had only Number of SD = 5). In the **Table 5**, CD34 and PTPRC (CD45) were also presented because they are typical HSCs markers, to confirm our results.

Table 5. *Global test* statistic of the selected covariates (genes): TAGLN, COL4A2, COL4A1, SCUBE3, standard genes: NT5E, THY1, ENG and hematopoietic genes: CD34, PTPRC (CD45). The parameters employed were: the weighted test statistic, the mean value of the gene in all samples, de deviation standard (SD), the number of deviations standard from the mean for datasets of Microarrays, Exon Arrays and RNA-seq data.

Microarrays					
Genes	p-value	Weighted test statistic	Mean	SD	Number of SD
TAGLN	3.66E-12	23.1926	1.3168	1.3633	16
COL4A2	1.68E-11	18.5605	0.9226	0.9631	18
COL4A1	5.13E-12	20.7785	0.9849	1.0293	19
SCUBE3	1.74E-07	9.1558	0.6176	0.6450	13
NT5E	2.18E-08	9.7083	0.8120	0.8331	10
THY1	9.39E-06	8.7419	0.8639	0.8912	8
ENG	9.45E-08	8.1060	0.5581	0.5810	12
CD34	2.69E-07	4.1783	0.3257	0.3354	11
PTPRC (CD45)	6.72E-08	2.0112	0.2253	0.2184	8
Exon Arrays					
Genes	p-value	Weighted test statistic	Mean	SD	Number of SD
TAGLN	0.0002	30.1943	5.1138	4.9160	5
COL4A2	0.0003	26.3950	4.2427	4.1633	5
COL4A1	0.0004	27.9514	4.4327	4.3823	5
SCUBE3	0.0003	28.1321	4.4107	4.3338	5
NT5E	0.0017	19.0957	4.7754	4.4351	3
THY1	0.0265	15.9119	4.9521	4.7460	2
ENG	0.0005	30.8115	5.8109	5.4700	4
CD34	0.0000	12.5627	3.0145	2.7457	3
PTPRC (CD45)	0.0000	7.7998	1.9770	1.8009	3
RNASEQ					
Genes	p-value	Weighted test statistic	Mean	SD	Number of SD
TAGLN	4.53E-09	31.8339	2.2771	2.2832	12
COL4A2	3.93E-09	26.6082	1.9769	1.9697	12
COL4A1	2.26E-08	23.4991	1.9071	1.8878	11
SCUBE3	0.0003	7.6833	1.1733	1.1453	5
NT5E	5.11E-09	19.8728	1.4518	1.4491	12
THY1	6.45E-11	30.6339	1.8775	1.8886	15
ENG	7.88E-06	9.3085	1.0661	1.0393	7
CD34	2.48E-09	26.3858	1.7843	1.7969	13
PTPRC (CD45)	1.52E-10	29.6913	1.8414	1.8503	15

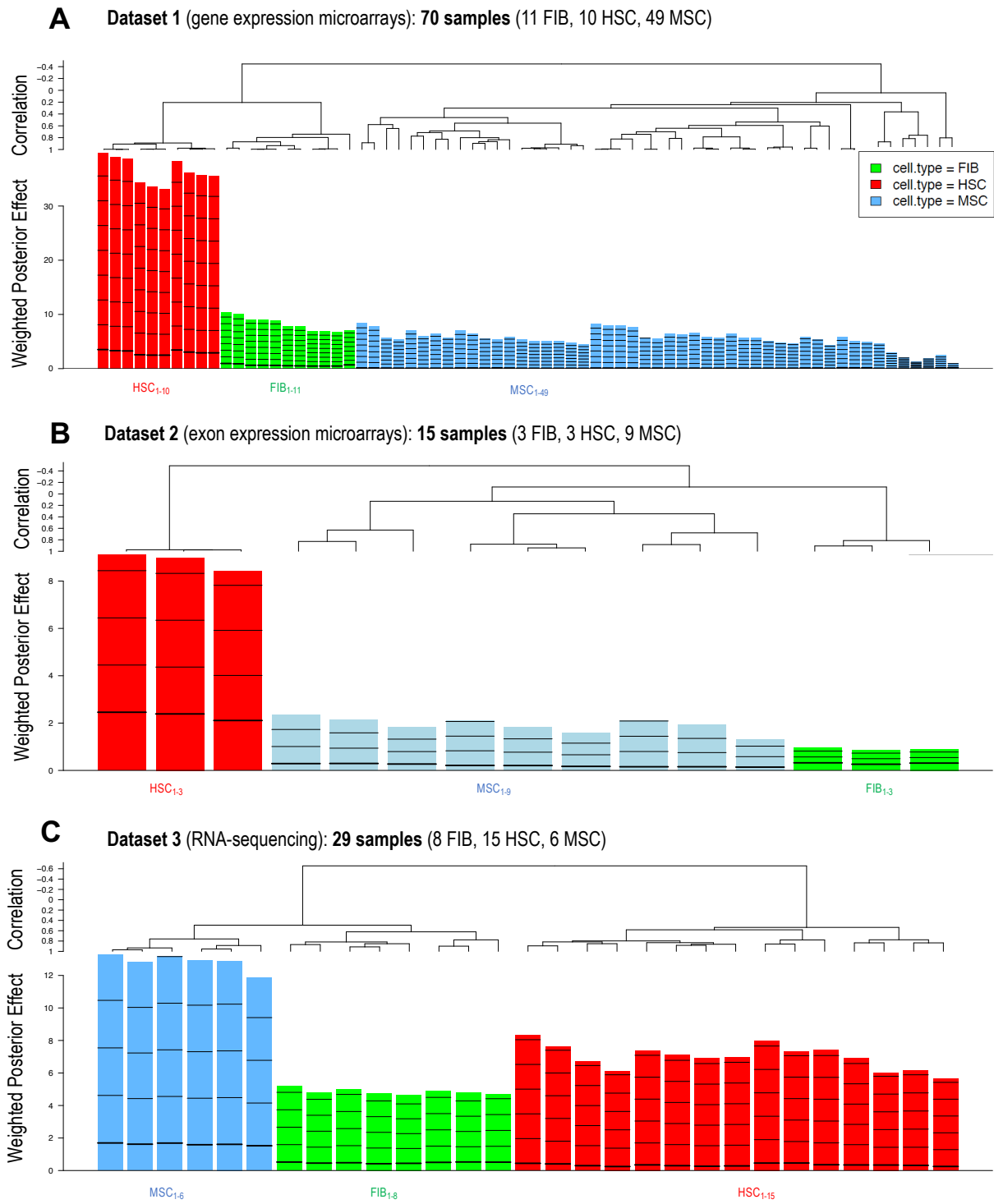


Figure 14. The subjects plot of the cell types: MSCs, HSCs and FIBs. **A)** Plot of the dataset 1 corresponding with the Meta-analysis of the microarrays. **B)** Plot of the dataset 2 corresponding with the exon arrays. **C)** Plot of dataset 3 corresponding with the RNA-sequencing data.

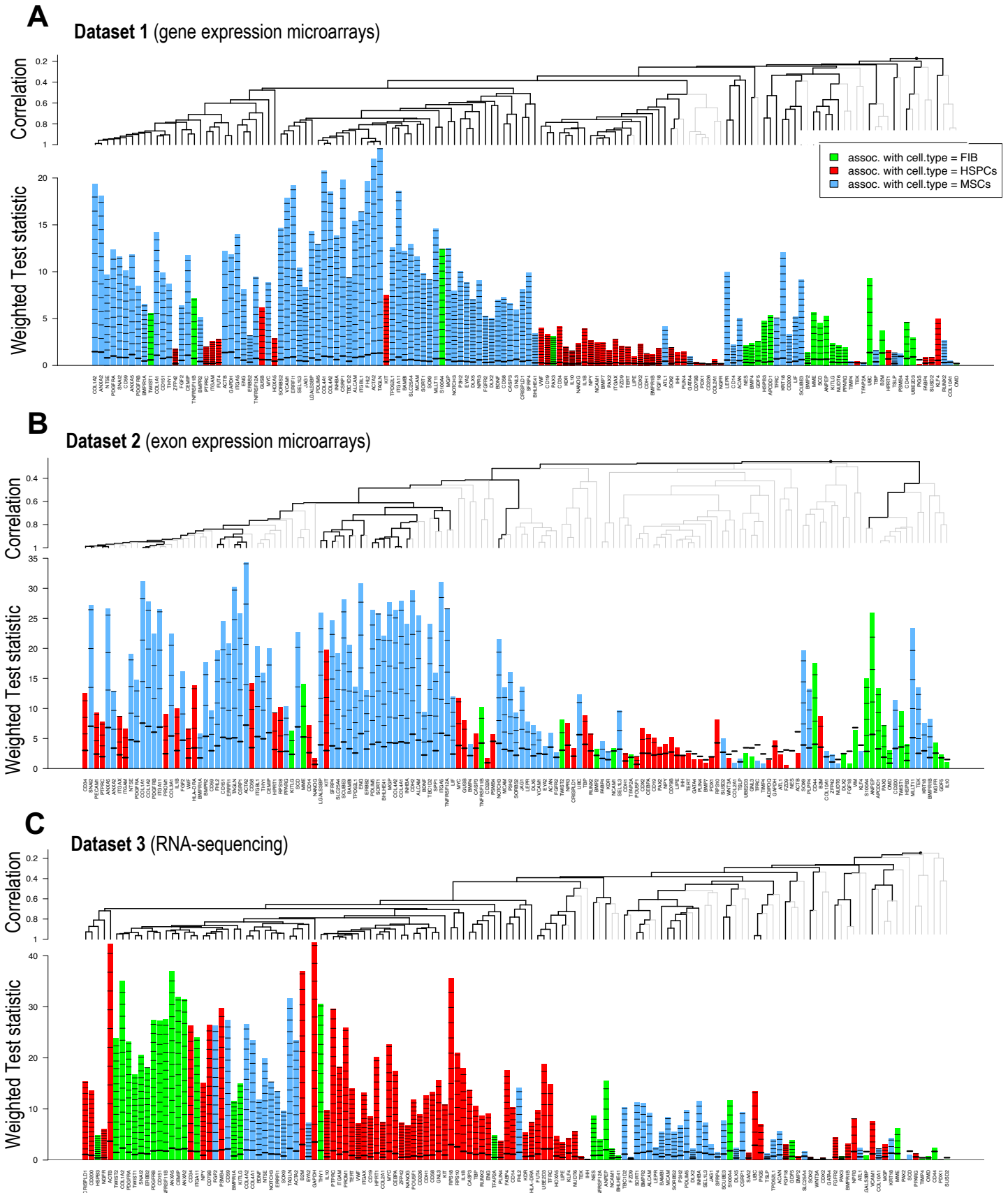


Figure 15. The covariates plot of 151 genes associated to the cell types: MSCs, HSCs and FIBs. **A)** Plot of the dataset 1 corresponding with the Meta-analysis of the microarrays. **B)** Plot of the dataset 2 corresponding with the exon arrays. **C)** Plot of dataset 3 corresponding with the RNA-sequencing data.

d. *Global Test* of selected MSCs specific markers

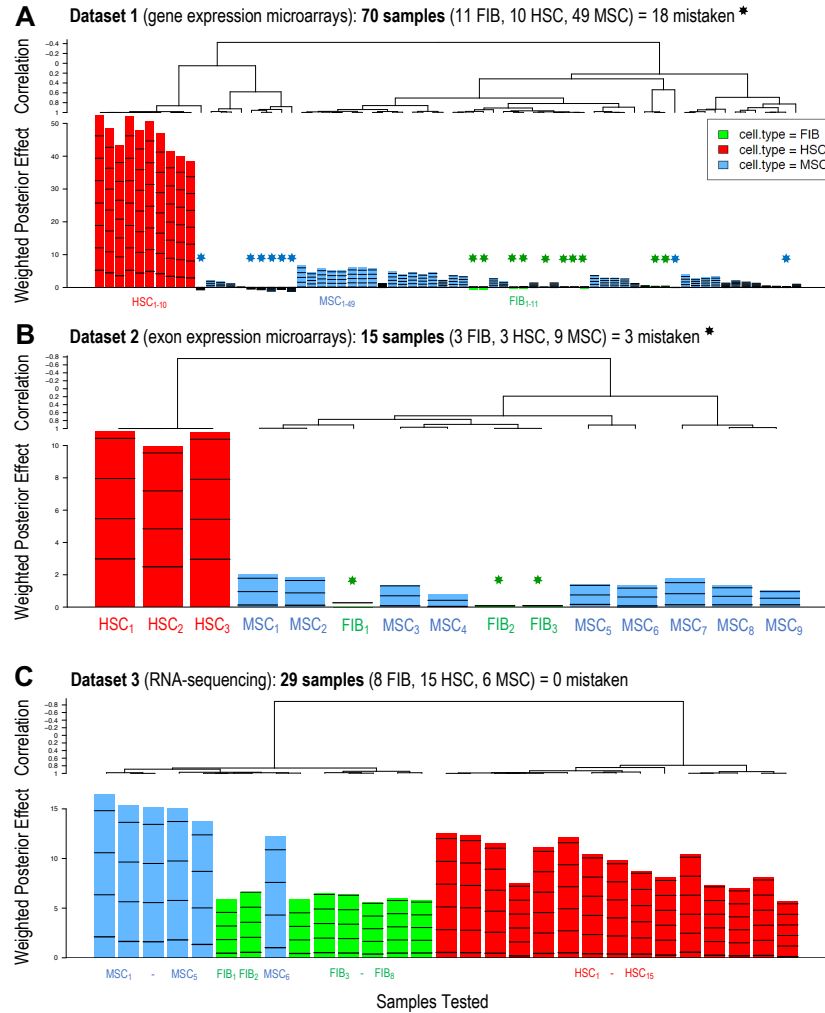
To gain insight into the results obtained in the previous section; we performed a new analysis with *Global Test* only with our genes of interest: TAGLN, SCUBE3, COL4A1 and COL4A2. This analysis aimed to confirm the power of these genes in separating and selecting the MSCs from the FIBs and HSCs. Moreover, we will correlate our proposed gene markers with those established by the ISCT: ENG (CD105), THY1 (CD90), NT5E (CD73).

In **Figure 16**, a new analysis was performed with the *Global Test* using only the genes of interest: TAGLN, SCUBE3, COL4A1, and COL4A2 and our standard genes: ENG (CD105), THY1 (CD90), NT5E (CD73). We performed 6 different subject plots.

In **Figure 16B**, using the Exon Arrays data, the standard genes were not able to categorize Fibroblasts well, often confusing them with MSCs samples. With our proposed gene markers (**Figure 16E**), there was only one mistake with an MSCs sample. In **Figure 16A**, samples from the Meta-Analysis of Microarrays (70 samples) were correlated based on the standard genes. There were 18 samples that were not well categorized, of which 8 were from MSCs samples and 10 were from fibroblasts. Using our proposed gene markers (**Figure 16D**), only one MSCs sample was not well categorized, indicating a better classification. The **Figure 16B** was done using the data from Exon Arrays, the standard genes were not able to categorize Fibroblasts well, often confusing them with MSCs samples (3 mistaken). However, with our proposed gene markers (**Figure 16E**), there was only one error with an MSCs sample. Finally, in **Figure 16C**, we used the RNA-seq data, and the standard genes labeled MSCs better (no mistakes) than our proposed genes, which made one mistake (**Figure 16F**). However, the standard genes could not distinguish between HSCs and FIBs, while our proposed genes could.

With this data, it seems that our proposed genes had a better prediction to classify the samples into the correct cell type: MSCs, HSCs or FIBs. However, to verify our results in the next section we performed the statistics of the analysis with confusion matrices.

Global test of MSC standard genes:ENG(CD105),
THY1(CD90) and NT5E(CD73)



Global test of MSC proposal genes:TAGLN,
SCUBE3, COL4A1 and COL4A2

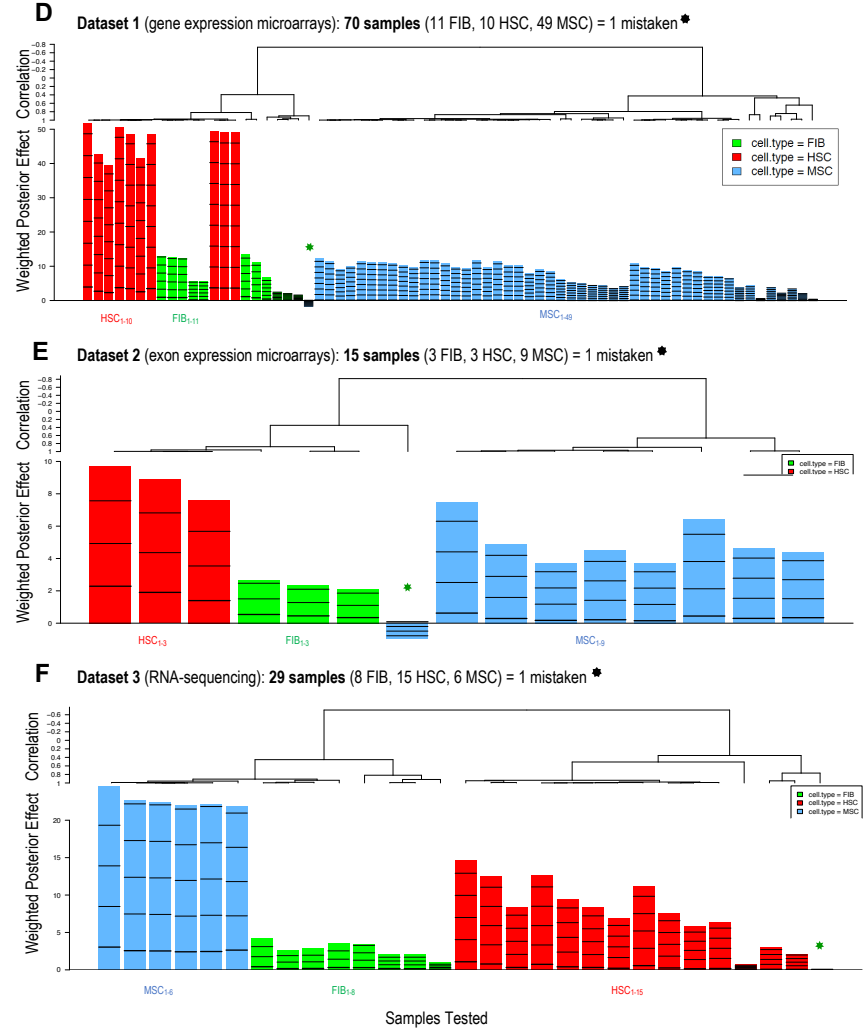


Figure 16. Subjects plots of Global test. **A, B** and **C**) for the standard genes from ISCT: ENG, THY1 and NT5E. **D, E** and **F**) proposal genes of the study: TAGLN, SCUBE3, COL4A1 and COL4A2, all were done for Meta-analysis of Microarrays, Exon Arrays and RNA-seq data, respectively.

e. Sensitivity, Specificity, FDR and Confusion Matrices of the selected MSCs markers

In the context of evaluating the error in the assignment of our MSCs from No MSCs (HSCs and FIBs), we realized a statistical analysis based in confusion matrices with the results of the *Global test* of **Figure 16**. We used the total of 114 samples that were the summatory of the three different expression platforms: Microarrays, Exon Arrays and RNA-seq. We based our best results on the sensitivity and specificity of the test.

Table 6 presents the confusion matrix for each gene, both the proposed gene markers and the standard genes. Among the proposed gene markers, the SCUBE3 was the worst predictor gene with a sensibility of 0.78 and a specificity of 0.64, the maximum error is in the first dataset (Microarrays) where had 19 false positives when categorizing the MSCs cell type. The best candidates were COL4A1 and TAGLN with the best sensitivity in COL4A1 (TPR= 0.98) and the best specificity in TAGLN (TNR= 1). Moreover, both had the best precision, with values of PPV= 0.92 and 1, respectively. The only corrected significant value was TAGLN with FDR=0.

Table 6. Confusion matrices of the proposed single genes (TAGLN, COL4A1, COL4A2 and SCUBE3) and standard single genes (ENG, THY1 and NT5E). For each one was calculated: miss rate, Fall-out, sensitivity, specificity, precision and FDR.

Confusion Matrices Using 1gene : PROPOSAL SINGLE GENES		TAGLN		COL4A1		COL4A2		SCUBE3	
3 DATASETS	MSC / No MSC	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)
DT1: 70 samples	49 MSC	49	0	46	3	45	4	30	19
Microarrays	21 No MSC	17	4	21	0	21	0	18	3
DT2: 15 samples	9 MSC	9	0	8	1	8	1	7	2
Exon Arrays	6 No MSC	5	1	6	0	6	0	6	0
DT3: 29 samples	6 MSC	6	0	6	0	6	0	6	0
RNA-seq	23 No MSC	23	0	22	1	22	1	14	9
TOTAL	64 MSC (Positive)	64	0	60	4	59	5	43	21
	50 No MSC (Negative)	45	5	49	1	49	1	38	12
SUM		109	5	109	5	108	6	81	33
TOTAL		114		114		114		114	
FNR(false negative rate)= FN/FN+TP	Miss rate	0,0725		0,0164		0,0167		0,2182	
FPR(false positive rate) = FP/FP+TN	Fall-out	0,0000		0,0755		0,0926		0,3559	
TPR(true positive rate) = TP/TP+FN	Sensitivity	0,9275		0,9836		0,9833		0,7818	
TNR(true negative rate)= TN/ TN+FP	Specificity	1,0000		0,9245		0,9074		0,6443	
PPV(positive predictive Value)= TP/TP+FP	Precision	1,0000		0,9375		0,9219		0,6719	
FDR= 1-PPV	FDR	0,0000000		0,0625000		0,0781250		0,3281250	

Confusion Matrices Using 1gene : STANDRAD SINGLE GENES		ENG (CD105)		THY1 (CD90)		NT5E (CD73)	
3 DATASETS	MSC / No MSC	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)
DT1: 70 samples	49 MSC	34	15	32	17	45	4
Microarrays	21 No MSC	19	2	10	11	10	11
DT2: 15 samples	9 MSC	7	2	0	9	9	0
Exon Arrays	6 No MSC	3	3	3	3	3	3
DT3: 29 samples	6 MSC	5	1	6	0	6	0
RNA-seq	23 No MSC	21	2	23	0	22	1
TOTAL	64 MSC (Positive)	46	18	38	26	60	4
	50 No MSC (Negative)	43	7	36	14	35	15
SUM		89	25	74	40	95	19
TOTAL		114		114		114	
FNR(false negative rate)= FN/FN+TP	Miss rate	0,1321		0,2692		0,2000	
FPR(false positive rate) = FP/FP+TN	Fall-out	0,2951		0,4194		0,1026	
TPR(true positive rate) = TP/TP+FN	Sensitivity	0,8679		0,7308		0,8000	
TNR(true negative rate)= TN/ TN+FP	Specificity	0,7049		0,5806		0,8974	
PPV(positive predictive Value)= TP/TP+FP	Precision	0,7188		0,5938		0,9375	
FDR= 1-PPV	FDR	0,2812500		0,4062500		0,0625000	

Otherwise, **Table 6** presented the confusion matrix of the standard genes. In general, the results of the standard genes were worse than our proposed markers. The best specificity and precision were in NT5E with a value of TNR= 0.89 and PPV= 0.93 while the best sensitivity was in gene ENG with a TPR= 0.86.

To further analyze the combination of gene markers, **Table 7** presented the confusion matrix for the two best proposed gene markers (TAGLN and COL4A1) and the two best standard gene markers (NT5E and ENG). In this **Table 7**, the combination of the two best proposed markers increased the sensitivity to the maximum, which was TPR = 1, whereas in the combination of the two standard genes created a medium between them of TPR= 0.83. The specificity of the combination of the two best proposed genes markers was TNR= 0.96 while in standard markers was TNR= 0.84. Finally, the precision of the two best candidate genes remained in the middle of the two precision values in both conditions, with PPV= 0.97 in proposal genes and PPV= 0.89 in standard genes.

In the case of the confusion matrix for all genes in each condition, the combination of all proposed gene markers (TAGLN, COL4A1, COL4A2, and SCUBE3) was significant. Although the specificity decreases a few tenths with values of TNR= 0.96 in gene proposal markers and TNR= 0.82, the values continuing to be more significant than individually gene TAGLN and SCUBE3. In case of the specificity, the combination of all proposed markers increased the values up to TPR= 0.98, as in the values of COL4A1 and COL4A2, which had the best values of sensitivity in the individual analysis. In the standard marker combination, the number was reduced a few tenths of the best individual gene NTE5E, but remained high with a value of TNR= 0.822, indicating the increase in the probability of correctly identifying the MSCs. In all cases except in the individual SCUBE3, the proposed markers: TAGLN, COL4A1, COL4A2 and SCUBE3 showed better results than the standard genes ENG, THY1 and NT5E in specificity, sensibility or precision. This indicated that the proposed gene markers had higher potential as specific gene markers for MSCs compared to the standard gene markers.

Table 7. Confusion matrices of the two best proposal and standard gene; and the second table represent the combination of all proposed genes (TAGLN, COL4A1, COL4A2 and SCUBE3) and standard genes (ENG, THY1 and NT5E): For each table was calculated: miss rate, Fall-out, sensitivity, specificity, precision and FDR.

Confusion Matrices Using 2genes : TWO BEST GENES		TAGLN, COL4A1		ENG (CD105), NT5E (CD73)	
3 DATASETS	MSC/ No MSC	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)
DT1: 70 samples	49 MSC	48	1	42	7
Microarrays	21 No MSC	21	0	13	8
DT2: 15 samples	9 MSC	8	1	9	0
Exon Arrays	6 No MSC	6	0	3	3
DT3: 29 samples	6 MSC	6	0	6	0
RNA-seq	23 No MSC	23	0	23	0
TOTAL	64 MSC (Positive)	62	2	57	7
	50 No MSC (Negaive)	50	0	39	11
SUM		112	2	96	18
TOTAL		114		114	
FNR(false negative rate)= FN/FN+TP	Miss rate		0.0000		0.1618
FPR(false positive rate) = FP/FP+TN	Fall-out		0.0385		0.1522
TPR(true positive rate) = TP/TP+FN	Sensitivity		1.0000		0.8382
TNR(true negative rate)= TN/ TN+FP	Specificity		0.9615		0.8478
PPV(positive predictive Value)= TP/TP+FP	Precision		0.9688		0.8906
FDR= 1-PPV	FDR		0.0312500		0.1093750

Confusion Matrices 4genes: ALL GENES		TAGLN, COL4A1, COL4A2, SCUBE3		ENG (CD105), THY1 (CD90), NT5E (CD73)	
3 DATASETS	MSC/ No MSC	TRUE	FALSE (FAIL)	TRUE	FALSE (FAIL)
DT1: 70 samples	49 MSC	48	1	41	8
Microarrays	21 No MSC	21	0	11	10
DT2: 15 samples	9 MSC	8	1	9	0
Exon Arrays	6 No MSC	6	0	3	3
DT3: 29 samples	6 MSC	6	0	6	0
RNA-seq	23 No MSC	22	1	23	0
TOTALES	64 MSC (Positive)	62	2	56	8
	50 No MSC (Negaive)	49	1	37	13
SUM		111	3	93	21
TOTAL		114		114	
FNR(false negative rate)= FN/FN+TP	Miss rate		0.0159		0.1884
FPR(false positive rate) = FP/FP+TN	Fall-out		0.0392		0.1778
TPR(true positive rate) = TP/TP+FN	Sensitivity		0.9841		0.8116
TNR(true negative rate)= TN/ TN+FP	Specificity		0.9608		0.8222
PPV(positive predictive Value)= TP/TP+FP	Precision		0.9688		0.8750
FDR= 1-PPV	FDR		0.0312500		0.1250000

f. Experimental validation of the gene markers using RT-qPCR

All these results were independently validated by RT-qPCR with samples of 6 human cell types at the Cellular Therapy Laboratory of the University Hospital of Salamanca. The results were presented in **Figure 17** and confirmed the expression level of 9 genes tested in MSC: ALCAM (CD166), COL4A1, COL4A2, ENG (CD105), ITGA1 (CD49a), NT5E (CD73), TAGLN, THY1 (CD90) SCUBE3 from different sources (BM-MSC and AD-MSC), FIB, MNC and two stromal cell lines (HTERT and HY5.5). The results indicated that COL4A1, COL4A2 and TAGLN are good and distinctive markers of MSC from bone marrow or adipose tissue, being highly expressed in these cells and with very low expression in the other cells tested. Moreover, these genes also showed a significant difference between each cell type in the comparison of MSC versus Fibroblasts, overall, in AD-MSCs. When comparing Fibroblasts vs BM-MSC the expression of NT5E (CD73) was also significantly higher in MSC from bone marrow but not so clear for AD-MSC.

However, for the remaining genes tested, there were differences between HSC and MSCs but not in the expression in Fibroblasts vs MSCs (either BM-MSC or AD-MSC). It may be relevant to note that in the panel of 9 genes tested in this study, the MNCs isolated from peripheral blood showed low expression levels and were not significantly different from the other cell types tested. These cells, MNC, although with small signal, were represented in ALCAM (CD166) and for ITGA1 (CD49a) genes, but these genes which have been used in some studies to characterize MSCs; were expressed in a similar level in FIB, indicating that they cannot be used as selective and distinct markers of MSCs.

In conclusion, our RT-qPCR experimental results confirmed that the genes COL4A1, COL4A2 and TAGLN, presented a high expression and specificity as gene markers of human Mesenchymal Stem Cells, showing clear differences with fibroblasts, especially TAGLN.

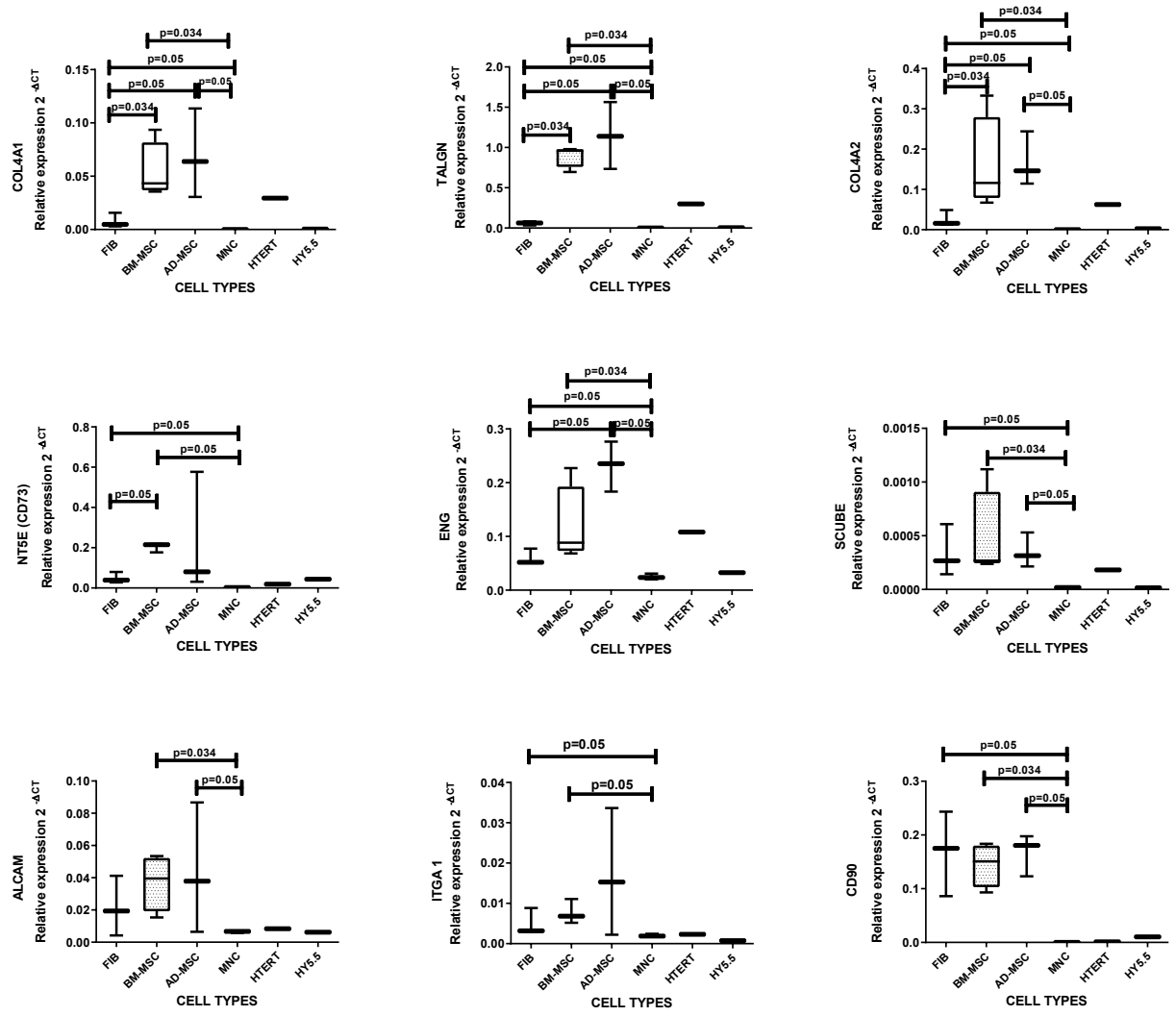


Figure 17. RT-qPCRs boxplots of genes: ALCAM(CD166), COL4A1, COL4A2, ENG (CD105), ITGA1 (CD49a), NT5E (CD73), TAGLN, THY1 (CD90) SCUBE3 from BM-MSC, AD-MSC, FIB, mononuclear cells (MNC) and two stromal cell lines (HTERT and HY5) cell types. The housekeeping used was GAPDH.

3. DISCUSSION

Mesenchymal Stem Cells have gained attention in regenerative medicine due to their potential to differentiate into various cell types and their immunomodulatory properties. Over the years, numerous studies have been conducted to accurately characterize and identify MSCs. Even if there are so many publications that try to incorporate new proposed MSCs gene markers, the ISCT still maintains in clinics the genes published in the articles written by M. Dominici et al. They established the three minimal criteria to consider MSCs: First, MSC must be plastic-adherent when maintained in standard culture conditions, second, MSC must express ENG (CD105), NTSE (CD73) and THY1 (CD90), and lack expression of PTPRC (CD45), CD34, CD14 or CD11b, CD79 α or CD19 and HLA-DR by flow cytometry and the last one, MSC must differentiate *in vitro* to osteoblasts, adipocytes and chondroblasts (Dominici et al., 2006).

However, as mentioned above, these canonical gene markers present a high similar level of expression as fibroblasts, which can cause confusion in a good characterization of BM-MSCs (Luo et al., 2020). Therefore, there are ongoing efforts to identify new and specific gene markers that can better distinguish MSCs from other cell types, especially fibroblasts, and improve the accuracy of MSC characterization. In murine experiments, a subpopulation of fibroblasts has been identified that are positive for CD105 and also exhibit multipotency (Lee et al., 2016) and CD73 is also present with high expression in cancer-associated fibroblasts (Yu et al., 2020). In the case of CD90, there is a specific article that explains the role and mechanisms of CD90 positive fibroblasts involved in pathophysiological processes, such as the inflammatory response, fibrotic process, and cell proliferation and differentiation (Zeng et al., 2023). This duality of the gene markers between MSCs and FIBs provides evidence for the need to find new possible MSCs-specific gene candidates.

The research presented in this study aimed to identify new gene candidates that can serve as potential markers for MSCs, providing a more robust and specific approach to their identification and differentiation from other cell types, especially than fibroblasts. The study employed transcriptomic and experimental analyses. The first screening of our 151 gene candidates presents the standard genes with high expression levels in MSCs, confirming the ISCT criteria. However, the expression levels of these genes were also observed to be high in fibroblasts, which can cause confusion in distinguishing between these two cell types.

Within this initial analysis, we found a cluster that showed potential in differentiating MSCs from fibroblasts. These genes included EYA2, NGFR, ACAN, BAMBI, NES, MCAM, SFRP4, FHL2, PDLIM5, LGALS3BP, MGP, SORT1, CSRP1, SCUBE3, COL4A1, COL4A2, INHBA and TAGLN. Some of them, such as MGP, are involved in the therapeutic mechanisms of MSCs for the treatment of Crohn's disease and contribute to their immunomodulatory functions (Feng et al., 2018).

Nestin is another gene that several articles have postulated as a potential marker to characterize a subset of bone marrow MSCs, participating in MSC angiogenesis and supporting HSC maintenance and homing (Karpenko et al., 2022; Lindsay & Barnett, 2017; Xie et al., 2015). Other common proposed marker MCAM glycoprotein gene (CD146), several articles postured it as the modulator of the migration of MSCs and possible new marker (Harkness et al., 2016; Roson-Burgo, Sanchez-Guijo, Del Cañizo, et al., 2016; Wangler et al., 2019), all of them commented in the Introduction. However, in correlation with the expression signal and the prediction algorithm *Global test* between the three independent datasets of different -omics methods: Microarrays, Exon Arrays and RNA-seq, the best candidates for MSCs were TAGLN, COL4A2, COL4A1 and SCUBE3.

These findings regarding Transgelin (TAGLN) are indeed interesting and demonstrate the complexity of gene expression patterns in different cell types. TAGLN is defined by the NCBI as a gene that

encoding a shape change and transformation sensitive actin-binding protein belonging to the calponin family. It is expressed in vascular and visceral smooth muscle. It has been implicated in early muscle differentiation [cited 04/06/2023; www.ncbi.nlm.nih.gov/gene/] and is a TGF β -inducible gene that regulates osteoblastic and adipogenic differentiation of human skeletal stem cells. In addition, the study by Elsafadi et al. proposes TAGLN as a marker of differentiation progression of hMSC based on its regulation of actin filament distribution and cytoskeletal organization, and presents in Western-Blot results how TAGLN is present in MSCs and not in Fibroblasts (Elsafadi et al., 2016).

However, TAGLN has been described in other previous articles as a marker of adult smooth muscle (Li et al., 1996). Moreover, another study of TAGLN gene expression levels in different cells from human bladder tissue showed that fibroblasts have lower expression than endothelial and smooth muscle cells (Tsui et al., 2019). This suggests that TAGLN expression levels can vary across different tissues and cell types, further highlighting the need for careful consideration and validation when using TAGLN as a marker for specific cell types.

The COL4A1 and COL4A2 genes, are key components of the basement membranes of blood vessels and various soft organs in mammals and belong to the family of collagen IV (Khoshnoodi et al., 2008). The mutations in these genes have been reported with a broader spectrum of cerebrovascular, renal, ophthalmologic, cardiac, and muscular abnormalities (Meuwissen et al., 2015). A study by C. Schinkle et. al all demonstrate a pattern of Mesenchymal Stem Cell expression genes as markers in the multiple myeloma, distinguishing these patterns from healthy donors. Despite this, in our results, COL4A1 and COL4A2 were up expressed in samples of human bone marrow MSCs healthy donors (Schinke et al., 2018). Moreover, other study characterizing the Fibroblasts in melasma, shows the down expression of COL4A1 in FIBs (Espósito et al., 2022). Nevertheless, other study compared skin Fibroblasts against MSCs from different sources and they found an increment of expression in Fibroblast in COL4A2 when fibroblasts were compared to MSCs from adipose tissue (Haydont et al., 2020).

SCUBE3 gene is a member of the signal peptide, complement subcomponent C1r/C1s, Uegf, bone morphogenetic protein-1 and epidermal growth factor-like domain-containing protein family. [provided by RefSeq, Dec 2014]. This gene is associated with loss of the epithelial marker E-cadherin and with increased expression of vimentin, in lung cancer tissue, and increased expression of SCUBE3 promotes the EMT in bone marrow MSCs (Yan et al., 2021). SCUBE3 is also expressed in osteoblasts and chondrocytes, where it interacts with BMPs to stabilize BMP receptors (Lin et al., 2021). In addition, other studies also published the presence of SCUBE3 gene in ectoderm, endoderm and endochondral tissues (Xavier et al., 2013).

All these information gain importance in the point of the combination of genes. We analyzed our genes independently by bioinformatic and experimental analyses to confirm our genes and even the best results were focus on the expression of TAGLN, the combination of our four proposal markers: TAGLN, COL4A1, COL4A2 and SCUBE3 proportionally high levels of specificity and sensitivity than the standard markers in our MSCs. Moreover, our results also demonstrate the ability of this combination of genes to distinguish the MSCs from the Hematopoietic Stem Cells and Fibroblasts. Given the heterogeneity of the MSCs, the combination of our four candidates may be interesting to consider the unknown subpopulations of MSCs. To corroborate this hypothesis, we can see as in our experimental results can identify TAGLN, COL4A2 and COL4A1 as the most expressed genes in the contrast of MSCs in 6 different human cell types. All these results could lead us to propose these proposed genes as a new MSCs gene marker combination.

For future studies, expanding the transcriptomic studies with a larger number of samples and including scRNA-seq data could indeed provide more comprehensive insights into the expression patterns

of these proposed gene markers and their potential to identify novel subpopulations of MSCs. scRNA-seq can help identify unique gene expression profiles in individual cells, shedding light on the heterogeneity and functional diversity within the MSC population. Furthermore, it could be interesting to analyze our proposed gene markers at the protein signal level using flow cytometry to validate their efficacy as MSC markers. Integration of multiple omics approaches, including transcriptomics, proteomics, and single-cell analyses, may lead to a better characterization and potential therapeutic applications. Taken together, our findings provide new promising candidate genes for the identification and characterization of MSCs.

4. FINAL SUMMARY of CHAPTER 1: *Specific gene markers for MSCs*

In recent years, MSCs have emerged as a key player in regenerative medicine due to their remarkable capacity to differentiate into various cell types and their ability to modulate the immune response. However, the accurate identification and characterization of MSCs remains a challenging task. The International Society for Cell and Gene Therapy (ISCT, <https://www.isctglobal.org/>) has established specific criteria to define human multipotent MSCs. These criteria include plastic adherence, and the expression of three positive CD markers: CD105, CD73, and CD90; and the absence of expression of CD45, CD34, CD14, CD11b, CD79 α , CD19, and HLA-DR markers. These criteria have been widely adopted in clinical practice. Despite the existence of these three commonly used gene markers (CDs 73, 90 and 105), it is recognized that they have similar expression levels in the Fibroblasts, making the proper distinction between MSCs and FIBs rather challenging. Consequently, there is continued effort and interest in identifying novel and more specific gene markers capable of reliably differentiating MSCs from all other cell types, particularly those of the stromal lineage, such as the fibroblasts.

This study presents a comprehensive exploration of multiple gene expression omics data to identify potential gene markers for MSCs through bioinformatic and experimental analyses. While many of the genes screened showed high levels of expression in MSCs isolated from different tissues, they also showed similar expression patterns in fibroblasts, highlighting the need for more discriminating markers. An initial selection of promising candidates that we found consisted of this set of 18 genes: ACAN, BAMBI, COL4A1, COL4A2, CSRP1, EYA2, FHL2, INHBA, LGALS3BP, MCAM (CD146), MGP, NES, NGFR (CD271), PDLIM5, SCUBE3, SFRP4, SORT1 and TAGLN.

After applying several machine learning algorithms to transcriptomic data from different platforms and different integrated datasets, we were able to select a subset of four genes that demonstrated a significant efficiency and accuracy in identifying MSCs. These genes were COL4A1, COL4A2, SCUBE3 and TAGLN. These genes demonstrated performance comparable to the three currently established standard markers (CDs mentioned above) in identifying MSCs compared to cells of the hematological lineage (e.g., HSCs), but also compared to cells of the stromal lineage, mainly FIBs. TAGLN, in particular, is presented as the most promising marker for MSCs. It plays a role in early muscle differentiation, TGF β signaling, and cytoskeletal organization. Additionally, it should be emphasized that TAGLN expression varies between different tissues and cell types, highlighting the importance of performing broader validations.

For further studies, we plan to test the proposed new markers in other large cohorts of independent samples coming from different adult tissues (including adipose tissue, muscle, tendon-ligament tissue, synovial tissue, dental pulp) as well as fetal tissues (such as placenta and umbilical cord). Furthermore, we aim to explore the expression patterns of these markers in multiple single-cell datasets. Additional analyses will explore the heterogeneity of MSC populations and how these gene markers change across such populations. In conclusion, our results and findings provide a better understanding of the molecular features that characterize MSCs and offer new potential markers to improve their precise identification and use in therapeutic applications.

CHAPTER 2:

Construction of gene regulatory networks and identification of master regulators of MSCs using transcriptomic expression data

Short Title:

Master regulators of MSCs

1. MATERIAL AND METHODS

This chapter 2 is based on the characterization of the regulation of Mesenchymal Stem Cells. To achieve this objective, we performed a bioinformatics study of MSCs and different cell types. In this chapter, we compiled a compendium of Microarrays data, performed differential expression analysis, functional enrichment analysis and analyzed the possible transcription factors and the “Master regulators” with different bioinformatic tools. Moreover, we further explored the interactions and relationships among these regulators by constructing co-regulatory networks. These results have been published in our article “**Deciphering Master Gene Regulators and Associated Networks of Human Mesenchymal Stromal Cells**”, attached in Annex III.

a. Integration of transcriptomic datasets obtained with high-density oligo Microarrays

The data were obtained from a meta-dataset of 264 samples obtained from 18 integrated datasets: GSE2666, GSE3823, GSE6029, GSE6460, GSE7637, GSE7888, GSE9451, GSE9520, GSE9593, GSE9764, GSE9894, GSE10311, GSE10315, GSE10438, GSE11418, GSE12264, GSE18043, and GSE46053; which are published data from the GEO platform of NCBI and subsequently analyzed in laboratory projects (Roson-Burgo, Sanchez-Guijo, del Cañizo, et al., 2016). Raw data were normalized using the *RMA* algorithm and the batch effect was corrected using the *frma* package (McCall et al., 2010).

We collected data for a set of 264 samples that were divided into several cell groups as indicated in **Table 8**, corresponding to: primary cells (including HSC, LYM, MSC, skin FIB; and OSTB), a set of differentiated cells derived from Mesenchymal Stem Cells (Adipoblasts derived from MSCs, dADIP; Chondroblasts derived from MSCs, dCHON, and dOSTB), and some stimulated cells (stMSC; and stOSTB). From this large set we considered a second subset that included 99 samples (**Table 8**), where we selected only the stem cells isolated from bone marrow, which were 10 samples of HSCs and 50 samples of MSCs; the primary cells (9 LYM, 11 FIB and 13 OSTB); and 6 samples of bone marrow stMSC.

Table 8. Number of samples of different human cell types obtained from the GEO datasets indicated above (GSE) that were collected, integrated and analyzed in this study.

Cell Type	Abbreviation	Type of cells (all isolated from healthy human donors)	Number of Samples (all)	Number of Samples (included in the comparisons)
Haematopoietic Stem/Progenitor Cells	HSC	primary cells	47	10 (bm HSC)
Lymphocytes (Haematopoietic Differentiated Cells)	LYM	primary cells	9	9
Mesenchymal Stromal/Stem Cells	MSC	primary cells	116	50 (bm MSC)
Stimulated MSCs	stMSC	stimulated	27	6 (bm stMSC)
Fibroblasts (derived from skin)	FIB	primary cells	11	11
MSC-derived Adipoblasts	dADIP	derived from SCs	3	–
MSC-derived Chondroblasts	dCHON	derived from SCs	3	–
MSC-derived Osteoblasts	dOSTB	derived from SCs	12	–
Osteoblasts	OSTB	primary cells	13	13
Stimulated Osteoblasts	stOSTB	stimulated	23	–
		Total =	264	99

b. Differential Expression Analysis

Differential Expression Analysis was performed using *Limma* package, with the following comparisons: MSC-HSC, MSC-LYM, MSC-FIB, MSC-OSTB, MSC-stMSC, and stMSC-HSC. *Limma* is an R package designed for the analysis of gene expression Microarray data, using linear models for the analysis and the evaluation of differential expression. Normalization and the data analysis are performed for two-color spotted Microarrays. To make the results more robust, *Limma* incorporates an analysis of variance (empirical Bayes method) to eliminate false positives, that is, genes with very small variations and that they do not really represent differences. eBayes modifies the previously obtained t-Student statistic depending on these variations, thus adjusting the differential expression of the genes to a common variance (Ritchie et al., 2015). The selection of significant differentially expressed genes was done using a 5% false discovery rate (FDR), and the top 30 genes with the most significant log fold change (log(FC)) for each group (up- and down-regulated).

c. ARACNe algorithm to reconstruct gene regulatory networks

The Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context (*ARACNe*) is an algorithm that calculates the correlation between the genes and their relationships through that their so-called “mutual information” between two genes expression profiles with the ultimate goal of creating regulatory networks. This method uses an information theoretic approach to eliminate most of the indirect interactions typically inferred by pairwise analysis (Margolin et al., 2006).

The *ARACNe* algorithm tries to find the mutual information (MI) from the gene expression profiles. Mutual Information analyzes the association and dependency of all the genes between them in a pair-wise manner and estimates the essential information they have in common. To do this, the MI is based on the entropy (I) values of the genes, as well as on the ability to infer the value of gene i from the information of a gene j , $I(g_i, g_j) \equiv I_{ij}$. We then filter the MI using an appropriate threshold, I_0 , computed for a given p-value, p_0 , in the null hypothesis that the two genes are independent. This MI value can be used to measure the relationship between the expression levels of two genes, with a high value indicating a higher level of interaction, as indicated in **Figure 18A**.

Figure 18B reflects the second step. *ARACNe* eliminates interactions between genes that are not direct, there may be genes where there is a relationship because they have a common connection with a third gene. In cases where there are three related genes, each triplet is analyzed regardless of whether its edges have been marked for removal, retaining the interactions that have a higher MI value and eliminating those relationships that may be indirect regardless of the order and thus false positives (Margolin et al., 2006). Mutual information then measures the degree of statistical dependence between two variables: x and y , and is defined as $I(x, y) = H(x) + H(y) - H(x, y)$, where $H(t)$ is the entropy of an arbitrary variable t , and t in this work is the value of the variable (logarithms).

As a result, *ARACNe* generates a matrix with three columns, in the first two there were genes that are related and the third corresponds to the MI value between genes. In this matrix, all the interactions between genes will be present (**Figure 18C**), but only those interactions in which there was at least one transcription factor will be of interest to us. For the transcription factors filter in the *ARACNe* algorithm, we used a list of 1544 Homo sapiens TFs (<http://bioinfo.life.hust.edu.cn/AnimalTFDB/index.shtml>). The algorithm then calculates a mean using the MI values as a cut-off point. The purpose of this is to eliminate the weakest genetic relationships and thus not include them in the next analysis (H.-M. Zhang et al., 2015).

The final result was a matrix with three columns containing all the information about the transcription factors found in the meta-dataset and their levels of association (MI) with the genes.

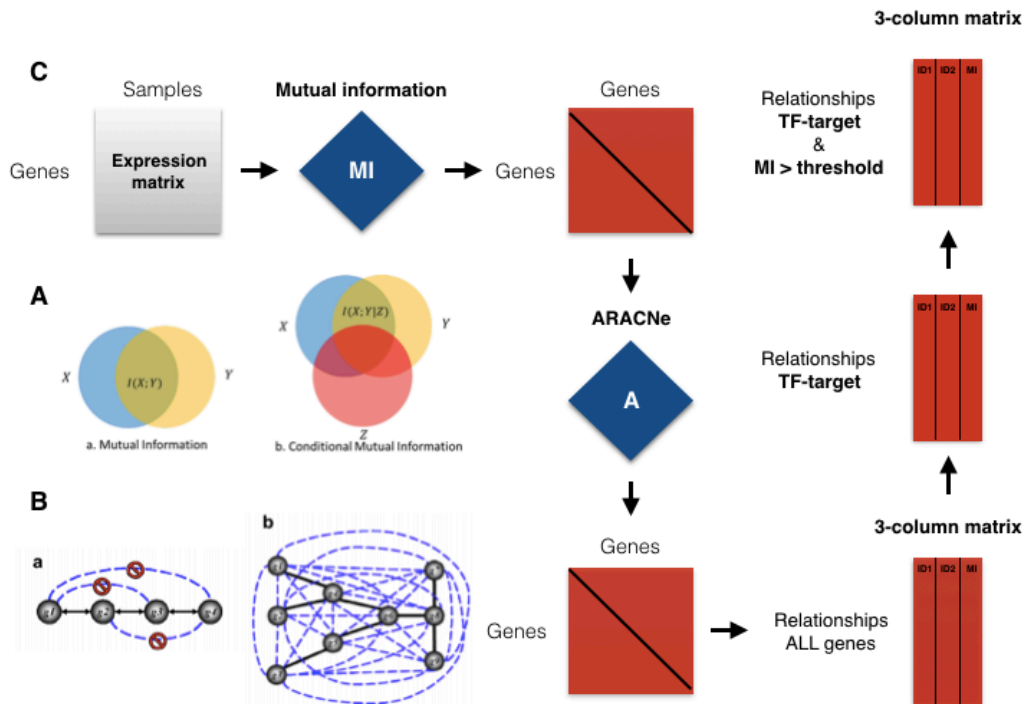


Figure 18. ARACNe analysis workflow. **A)** Mutual information (MI) Diagram. **B)** Elimination of false interactions on genes by the ARACNe algorithm. **C)** Steps developed in the ARACNe application for the obtaining a coexpression matrix (From Margolin et al., 2006).

d. VIPER (Virtual Inference of Protein-activity by Enriched Regulon analysis) algorithm: selection of regulation genes and co-regulated networks

The *VIPER* algorithm allows the computational inference of protein activity on an individual sample basis from gene expression profile data. This algorithm was computed in R programming and estimates the regulon activity derived from ARACNe analysis. In addition, while straight forward for TFs, *VIPER* assigns signal transduction proteins. For this purpose, we extended the concept of regulon to include the transcriptional targets that were directly affected by the activity of the protein, based on the maximization of information transfer total alternative paths. The method uses variable y as the value of MI returned by ARACNe, significant association TF-targets were filtered using as threshold = mean(y). Moreover, the *VIPER* algorithm includes bootstrapping that randomly selects 100 subsets of the samples to find and identify the most stable regulators and pleiotropy (Alvarez et al., 2016). The most significant TF regulators were selected with a p-value < 0.05. The comparative analysis was performed for the same groups of samples, using with the *limma* package for the differential expression. The pairwise comparisons were: MSC-HSC, stMSC-HSC, MSC-LYM, MSC-FIB, MSC-stMSC, and MSC-OSTB. Regulatory networks were visualized using *Cytoscape*.

Cytoscape is a software platform used to visualize molecular interaction networks and biological pathways and to integrate these networks into expression profiles (Shannon et al., 2003). The analysis was performed considering the top 10 most overexpressed TFs and the top 10 most repressed TFs found significant in the *VIPER* analysis, considering the 6 pairwise comparisons.

Moreover, we performed an enrichment identification of specific transcription factors binding sites (TFBS) within the list of genes studied by comparative analysis of the promoter sequences of a query list of genes against curated datasets using the *iRegulon* tool and associated the TFBS with our top 20 MR (Janky et al., 2014).

e. Enrichment Analysis to determine the function of the gene signature

To gain insight into the functions associated with the regulatory gene sets generated with TF regulator and the corresponding regulated genes as regulons, we performed a functional enrichment analysis. We used DAVID (Database for Annotation, Visualization, and Integrated Discovery) bioinformatics tool (<https://david.ncifcrf.gov/>) (Sherman et al., 2022) and GeneTerm-Linker bioinformatics tool (<http://gtlinker.cnbc.csic.es/>) which tools include a unified way: GO Biological Process, GO Molecular Function, GO Cellular Component, KEGG Pathways, and InterPro Motifs and Domains (Fontanillo et al., 2011).

f. Methylation Analysis

The analysis of DNA methylation profiles of the CpG island genes was performed with 3 datasets and 25 samples of human bone marrow MSCs from healthy donors: GSE79695, GSE129266 and GSE87797; and 1 dataset with 5 samples of human HSCs from healthy donors: GSE63409. The analysis was performed using the *minfi* package in R programming (Aryee et al., 2014). The DNA methylation of all these samples was measured using Illumina Infinium Human Methylation450 BeadChips (corresponding to platform GPL13534 in the GEO database). This technology allows the quantification of the global DNA methylation of the CpG islands across the genome based on the measurement of approximately 450000 methylation sites per sample at single nucleotide resolution.

All these samples were pre-processed and integrated together. The analysis is based on the β value, which consist of a ratio value of the proportion of methylated gene loci:

$$\beta = \frac{\text{methylated gene loci}}{(\text{unmethylated gene loci} + \text{methylated gene loci})}$$

The β value ranges from 0 to 1, with 1 indicating a complete methylated status and 0 indicating an unmethylated status. After the described global normalization of the samples and the calculation of the β values, we compared the top 20 gene TF regulators found to be overexpressed or repressed in the BM-MSCs with the methylation profiles of HSCs.

2. RESULTS

a. Transcriptomic profile of regulation genes in Mesenchymal Stem Cells

The transcriptomic characterization of MSCs was performed by the differential expression analysis of genome-wide expression signal. As indicated in Materials and Methods, we generated a compendium of different human hematopoietic primary cell types: HSCs and LYMs; FIBs, OSTBs, dADIPs, dCHONs, dOSTBs, and finally stMSCs. This compendium of 264 samples was reduced to 99 samples in order to discover the most differential genes. The subset was built with: MSC, HSC, LYM, FIB, OSTB, and stMSC. And the comparisons were: MSC-HSC, MSC-LYM, MSC-FIB, MSC-OSTB, MSC-stMSC, and stMSC-HSC.

The raw subset gene matrix consisted of 16698 genes and 99 samples. After applying the *Limma* algorithm to obtain the differential genes, we obtained a signature of 188 more differentiated genes by joining the top 30 most differentiated genes of each comparison. As we can see in the **Figure 19**, the genes were clustered in 4 groups. The names of these genes can be found in the *Supplementary 1* of our published article “**Deciphering Master Gene Regulators and Associated Networks of Human Mesenchymal Stromal Cells**” attached at the end of this Doctoral Thesis. Regarding the samples, the main differences were presented between hematopoietic and non-hematopoietic lineage. The LYM and the HSC were grouped together in the dendrogram, since they both came from the same hematopoietic lineage. On the other hand, the main differences within the non-hematopoietic lineage were shown between FIB and stMSCs, OSTB and MSCs.

The most different genes between MSCs and the hematopoietic lineage were in the Collagen family with genes such as: COL1A2, COL3A1, COL5A2, COL6A3, as well as the FN1 and CTGF. Analyzing the functional enrichment, these genes had the same characteristics in functions as extracellular matrix, focal adhesion, skeletal system development, adhesion, migration and differentiation lineage. Among the up expressed genes in MSCs versus hematopoietic lineage were ACTA2, MMP2, POSTN and especially SNAI2. Contrastingly, in hematopoietic lineage include genes such as CD69, FOSB, HLA-DRB1(ENSG00000206241), MYB, and SPINK2. CD69 and HLA-DRB1 are genes that play a central role in the immune system and are also expressed in T cells. In addition, MYB plays a key role in gene regulation throughout the hematopoietic hierarchy.

Within the differences between the hematopoietic lineages: LYM-HSC, we could see genes such as IGHA1, S100A8, S100A9, LTF, ARG1 and HCK were up expressed in LYM which had specific immune properties, or CXCR2 and C5AR1 genes related to the neutrophil chemotaxis.

In the other contrast, the MSCs vs FIB, genes such as CTSC, DPP, MMP1 and MME were associated in up expressed in FIB with functions such as proteolysis or protease functions.

In detail, the comparisons between MSCs and stMSCs revealed the up expression in A2M, APOD, COMP, CPE, DPT, PRELP, SERPINA3, SPARCL1, ITGA10, IGF2 and SPP1 related with functions of extracellular space and PI3K-AKT signaling. These results were to be expected because the stimulation with TGF- β in MSCs promotes the tissue regeneration, immune responses, EMT, differentiation and development. In the other hand, FLG, IL6, MEST, PLK2, TAGLN, COL4A1, COL11A1 and VCAM1 genes were down expressed in stMSCs vs MSCs, these genes are associated with the extracellular matrix.

The last comparison was MSCs versus OSTB, the up expressed genes in OSTB contains functions associated with collagen fibril organization and ECM, the most relevant genes were: COL15A1, COL15A1, GREM1, ITIH5, EGFR2, GREM1, CSF2RB, IGDCC4, NDNF and VCAM1.

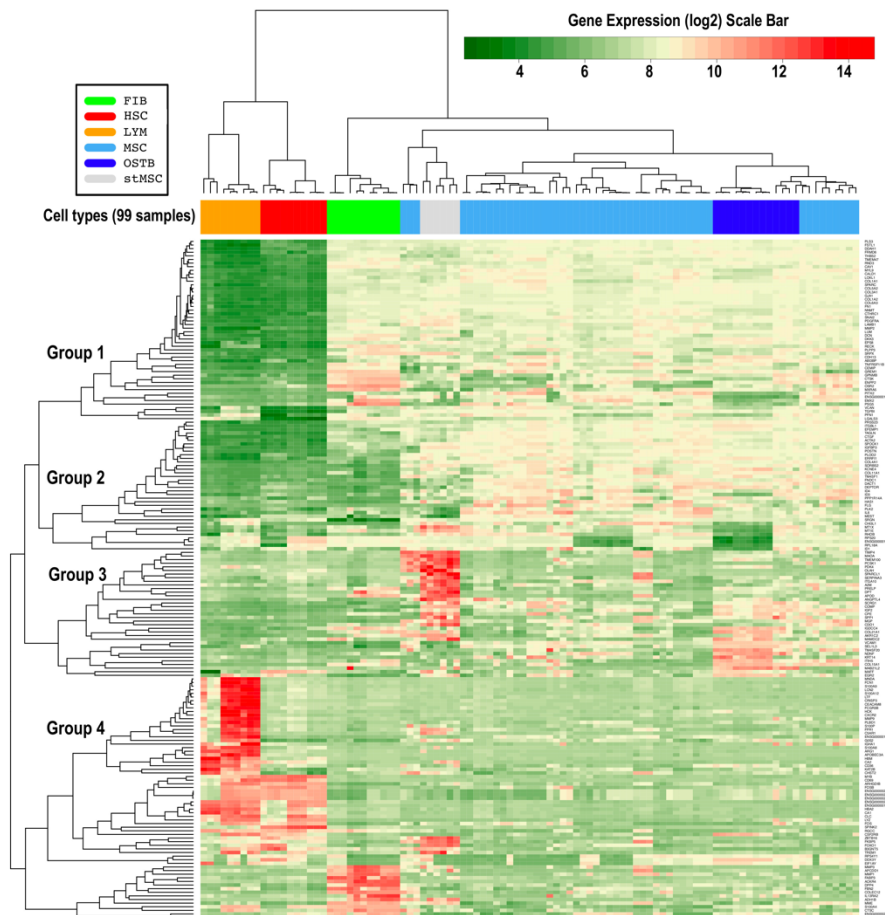


Figure 19. Heatmap of 188 genes from the differential expression analysis (joining top 30 genes of each pair-wise comparison: MSC-HSC, MSC-LYM, MSC-FIB, MSC-OSTB, MSC-stMSC, and stMSC-HSC).

b. Master Regulators of Mesenchymal Stem Cells

In this part of the results, we focused on identifying the role of Master Regulators in controlling the serial group of TF genes, which group of genes are called *regulons*. The bioinformatics algorithms used for this analysis were *ARACNe* and *VIPER*, as it was explained in the Materials and Methods section. The groups used in this analysis were the same as those used in the differential expression analysis: MSC-HSC, stMSC-HSC, MSC-LYM, MSC-FIB, MSC-stMSC and MSC-OSTB. The *ARACNe* algorithm was used to obtain the Mutual Information, which defines an interaction value between two genes in each pairwise comparison. Importantly, *ARACNe*, through MI, reflects the activity on the expression of the genes. These data were introduced into the *VIPER* algorithm, revealing the dissimilarity or distance between the cell types and all the associated regulons. The **Figure 20** showed the regulatory relationships between the MRs and their regulons in the different cell types mentioned above.

In **Figure 20A**, the differential analysis was presented, showing the most significant in the comparison of MSCs-HSCs, with 14 up-regulators and 22 down-regulators. The second most significant change occurred in the comparison of stMSC-HSC, with 13 up-regulators and 22 down-regulators, and

finally in MSC-LYM with 15 up-regulators and 13 down-regulators. The comparison of MSCs-FIB had relatively fewer changes, with 3 up-regulators and 10 down-regulators. Finally, the last two of MSC-stMSC and MSC-OSTB had fewer regulators; all the comparisons were found after the bootstrapping analysis. Bootstrapping analysis (a statistical technique for estimating quantities about a population by averaging estimates from multiple small data samples) provides a more accurate and robust result. The **Figure 20A** also showed pleiotropic effects due to the confluence of multiple genetic traits. This analysis greatly reduced the number of regulators, as it implies cooperativity.

The most relevant results were focused on the comparison between MSC-HSC, in **Figure 20B** we extracted from *VIPER* analysis common in the simple and bootstrap analysis; the top 10 upregulated genes in MSCs: *SNAI2*, *STAB2*, *IRX3*, *EPAS1*, *HOXC6*, *TWIST1*, *TULP3*, *PRRX1*, *TEAD1*, and *NFE2L1* and the top upregulated genes in HSC: *BCL11A*, *MYB*, *TFEC*, *HLF*, *GATA2*, *ERG*, *PLAGL2*, *DACH2*, *POU2F1*, and *GATA3*. The order of the genes was based on the normalized enrichment score (NES) value, which reflects the co-regulation of genes in their environment (Alvarez et al., 2016). On the other hand, the size of the *regulon* corresponded to the number of target genes found to be associated with each MR. In MSC the maximum size was found in *TEAD1* (111 genes), *PRRX1* (51 genes) and *IRX3* (56 genes), whereas in HSC, the maximum size of the *regulon* was found in *POU2F1* (76 genes), *GATA2* (70 genes) and *ERG* (49 genes). Finally, in the same figure was also represented which genes present pleiotropy, in the case of MSC are *SNAI1*, *EPAS1* and *HOXC6* while in HSC were *BCL11A*, *MYB*, *HLF*, *PLAGL2* and *GATA3*.

In the **Figure 20C**, the enrichment plot represented the same top 10 TF master regulators of MSCs and the same top 10 HSCs MR. In this plot, we wanted to represent the activity and the expression value of each MR in the whole dataset. It is crucial to clarify that the expression and the activity of the MR are not the same, it is understood as the activity of a transcription factor in a sample of cells, refers to the extent to which it exerts its regulatory potential, whereas the expression gene is the process by which the information encoded in a gene is translated into a function. The plot also shows the repressed targets (in blue) and activated targets (in red) for each MR, on the NES (x-axis), with the genes in the NES ranked from the most down-regulated to the on most up-regulated in the MSC vs HSC comparison. The plot highlights *SNAI2* and *TEAD1* as the most significant up-regulators in MSC, and the highest activity in up-regulators of HSC was observed for *BCL11A* and *MYB*.

c. Differential and Functional identification of regulation gene signature in MSCs

Consistent with the top 20 genes analyzed with *VIPER*, we characterized the top 20 up-down MR genes with the other 5 comparisons with a differential expression analysis using the *limma* package. **Table 9** represents the results of the differences between groups based on the \log_2 of the fold change ($\log_2(FC)$), the mean expression of the gene in the whole dataset and the FDR in the MSC-HSC comparison.

Among the top 20 up-regulated MR genes in the comparisons against the hematopoietic lineage, SNAI2 was represented as the most up-regulated gene in all the comparisons against the hematopoietic lineage, with values of log₂(FC) in MSC-HSCs: 5.93, stMSCs-HSC: 7.08 and MSC-LYM 6.49, but down-regulated in the contrasts of MSC-stMSCs and MSC-OST (log₂(FC)=-1.15 and log₂(FC)=-0.25, respectively) followed by TWIST1 with values of log₂(FC) in MSC-HSCs: 4.14; stMSCs-HSC: 5.15 and MSC-LYM 4.43 and down-regulated in the comparisons of MSC-FIB and MSC-stMSC with values of log₂(FC) of -1.22 and -1.01, respectively.

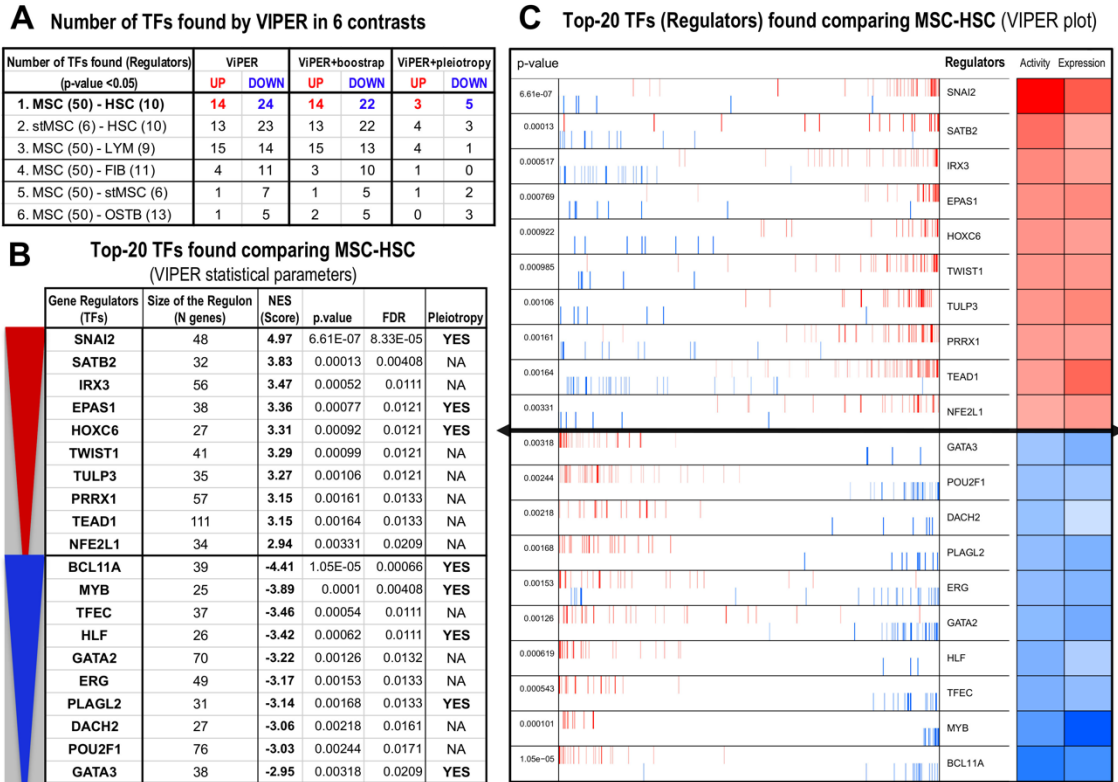


Figure 20. TFs and associated regulons found using *VIPER*. **A)** Table of the number of TF up-down in each of the 6 comparisons, the results with *VIPER*, *VIPER*-bootstrap and *VIPER*-pleiotropy. **B)** Table with top 10 up and top 10 down regulated TFs in MSCs-HSCs including the parameters from *VIPER*: NES, p-value, FDR and pleiotropy. **C)** Table with top 10 up (in red) and top 10 down (in blue) regulated TFs in MSCs-HSCs, with the activity and the expression on the right, and the bars as the number of TFs that configure the regulons.

In the top 10 down-regulated genes in MSC-HSC were determined by *VIPER*, MYB remained the most down-regulator in MSCs in the contrasts of MSC against the hematopoietic lineage. It showed significant down-regulation with log₂(FC) in MSC-HSCs: -6.03, stMSCs-HSC: -6.04 and MSC-LYM: -3.61, also was down-regulated in the contrast of MSC-FIB: -0.15. Similarly, GATA2 showed significant down-regulation in MSC-HSCs: -3.23; stMSCs-HSC: -3.24 and MSC-LYM: -1.29. GATA2 is a TF required for the generation survival and maintenance of HSC and its dysregulated expression is associated with human immunodeficiency syndromes and vascular integrity (de Pater et al., 2013).

The results of the **Table 9** showed small differences between the comparisons of MSC-FIB, MSC-stMSC, MSC-OSTB. These events were expected due to the similarity in stromal niche. However, a remarkable gene could be IRX2 in the contrast of MSC-FIB with a log₂(FC) of 2.13 which involves the regulation of cell fate determination in mesenchymal stem cells and it is involved in the development of MSCs (Narwidina et al., 2023).

Table 9. Differential Expression table of Top 20 MR by *VIPER* analysis in the 6 comparisons analysed in **Figure 18**. Table represents the Fold Change logarithm of each comparison, the mean expression of the gene and the FDR in the contrast of MSC-HSC.

N	Gene Regulators (TFs)	log2FC MSC-HSC	log2FC stMSC-HSC	log2FC MSC-LYM	log2FC MSC-FIB	log2FC MSC-stMSC	log2FC MSC-OSTB	Mean Expression	adjusted p-value MSC-HSC
1	EPAS1	3.5672	1.6307	3,5248	0.2669	1.9365	0.5094	8.4759	8.96E-39
2	HOXC6	3.6790	3.7755	4,5905	-0.1399	-0.0965	-0.0578	8.5355	1.39E-26
3	IRX3	3.9400	4.3672	5,2372	2.1315	-0.4272	-0.7877	8.0938	4.36E-44
4	NFE2L1	2.0413	1.8684	1,9363	0.7934	0.1729	-0.1964	9.5201	1.81E-42
5	PRRX1	4.2175	3.4723	4,9694	-0.2984	0.7452	0.5873	9.1210	4.04E-39
6	SATB2	2.7367	2.6208	3,5340	1.0307	0.1160	-0.5446	7.4116	2.05E-36
7	SNAI2	5.9277	7.0837	6,4919	0.3710	-1.1560	-0.2533	9.3545	3.89E-63
8	TEAD1	2.9846	1.8031	3,9882	-0.6595	1.1814	-0.4376	7.8122	2.03E-54
9	TULP3	1.9034	2.1224	1,8825	-0.1382	-0.2190	-0.1371	7.7463	3.59E-39
10	TWIST1	4.1399	5.1461	4,4375	-1.2236	-1.0061	0.2703	7.7358	3.42E-45
1	BCL11A	-2.7727	-2.9556	-2,3401	-0.5342	0.1830	0.0790	5.2395	1.17E-43
2	DACH2	-0.7719	-0.8682	-0,0860	-0.3389	0.0964	0.0332	4.0574	2.26E-18
3	ERG	-2.0832	-2.2306	-0,5000	0.0850	0.1474	0.1382	4.6198	2.19E-37
4	GATA2	-3.2355	-3.2464	-1,2964	-0.0883	0.0109	0.0436	6.3737	4.24E-37
5	GATA3	-1.9259	-1.9510	-1,5210	-0.3112	0.0251	0.0943	4.4646	4.91E-18
6	HLF	-3.1046	-3.0152	-0,3987	-0.3616	-0.0894	-0.1376	4.6562	7.52E-22
7	MYB	-6.0366	-6.0413	-3,6154	-0.1538	0.0047	0.2063	5.4332	1.21E-58
8	PLAGL2	-1.7381	-1.3196	-1,8485	-0.6596	-0.4186	-0.0203	5.7885	2.86E-32
9	POU2F1	-1.3250	-1.1853	-0,6829	-0.3178	-0.1397	-0.1775	5.7732	2.86E-25
10	TFEC	-2.0650	-2.1096	-2,3706	-0.0804	0.0446	-0.0046	4.1215	1.70E-22

The functional enrichment analysis shown in **Table 10** provides the top 10 up- and down-regulated MRs with their corresponding regulons. The top 10 up-regulated MRs and their regulons in MSC-HSC contrast were associated with typical functions of MSCs such as cell adhesion, cytoskeleton or cell differentiation. Moreover, the functional enrichment determined the second more important cluster with functions associated with organ morphogenesis and development indicated in the genes: SATB2, TULP3 EPAS1, TEAD1 and SNAI2.

On the other hand, the top 10 up-regulated MR and their regulons in HSC presented functions clearly associated with the immune system development and hematopoiesis. Notably, BCL11A and IKZF1 were among the top upregulated genes. We commented before the relevant characteristics of BCL11A, in the case of IKZF1 is an important TF in the hematopoietic stem cell niche in B-progenitor acute lymphoblastic leukemia, there are recent studies that reveals IKZF1 as a possible new gene marker to determine poor prognostic, although the heterogeneity of concomitant lesions in the patients avoid directly correlate with cell function and therapy resistance (Rogers et al., 2021).

Table 10. Functional enrichment analysis done with the top 10 master regulators (TFs) upregulated (UP) and their corresponding gene regulons and with the top 10 master regulators (TFs) downregulated (DOWN) and their corresponding gene regulons.

Enriched Functional Term	N Genes (in the Function)	N Genes (in the Query)	N in Function/N in Query (%)	Regulation (UP/DOWN)	p-value (adj. Benjamini)
generation of neurons	26	283	7.34	UP	0.015168
neurogenesis	30	283	8.47	UP	0.001735
nervous system development	42	283	11.86	UP	0.007296
cell-substrate adhesion	10	283	2.82	UP	0.014732
extracellular matrix	22	323	6.21	UP	0.001305
cytoskeleton	51	323	14.41	UP	0.003200
face morphogenesis	4	283	1.13	UP	0.047319
embryonic development	29	283	8.19	UP	0.001836
organ morphogenesis	31	283	8.76	UP	0.000405
organ development	57	283	16.10	UP	0.016957
cell differentiation	61	283	17.23	UP	0.000488
T cell activation	17	293	5.00	DOWN	0.000004
lymphocyte activation	21	293	6.18	DOWN	0.000005
hemopoiesis	22	293	6.47	DOWN	0.000010
hemopoietic or lymphoid organ development	22	293	6.47	DOWN	0.000043
immune system development	22	293	6.47	DOWN	0.000074
regulation of immune system process	23	293	6.76	DOWN	0.002043
calponin-like actin-binding	10	323	2.94	DOWN	0.002577
actin filament-based process	19	293	5.59	DOWN	0.000509
cytoskeleton organization	26	293	7.65	DOWN	0.000664
GTPase regulator activity	24	304	7.06	DOWN	0.001932

d. Methylation study of the MSCs Master Regulators

After *VIPER* analysis and corresponding differential and functional analysis, we decided to corroborate our top 20 MR with a methylation study profile. This study provided valuable insights into the epigenetic regulation of these genes and their correlation with their relative expression levels. The results indicated that certain MR genes were relatively hypomethylated in MSCs compared to HSCs, while others were hypermethylated in MSCs. The used *minfi* package was used in R programming as is indicated in the Materials and Methods section. We used three different DNA methylation profiles of BM-MSCs and compared them with HSC profiles. In the **Figure 21**, we used a differential analysis to see the differences in the methylation of the CpG island of our top 20 MR. The results showed that the genes EPAS1, NFE2L1, SATB2, SNAI2, TEAD1 and TULP3 were relatively hypomethylated in MSCs and the hypermethylation in MSCs genes of our top 20 MR were ERG, GATA2, GATA3, HLF, MYB and POU2F1.

The best results of hypomethylation in MSCs were in the gene SNAI2 with Beta-values of: 0.12, 0.14 and 0.1 in MSCs GSE identifiers, while the Beta-values of the HSC was 0.26, according the **Figure 21**. In addition, TEAD1 also had great Beta-values: 0.52, 0.55 and 0.57 in MSCs and Beta value of 0.63 in HSCs. In case of hypomethylation of the down-regulated MR, all genes presented higher differences except POU2F1. GATA2 and MYB were the most homogeneous with the lower variance between the samples. The results of Beta-value for GATA2 in MSCS were: 0.27; 0.24 and 0.25 and in HSC was: 0.13. In MYB gene the beta-values for MSCs were: 0.27, 0.28 and 0.28 while in HSC was 0.16.

The finding that SNAI2 was hypomethylated in BM-MSCs and was a Master Regulator was consistent with the functions in EMT of MSCs. All these results of MR in the epigenetics gain the idea of that these genes regulate the cell stemness, commitment and differentiation lineage of the MSCs, and add relevance and confidence to our top 20 MR.

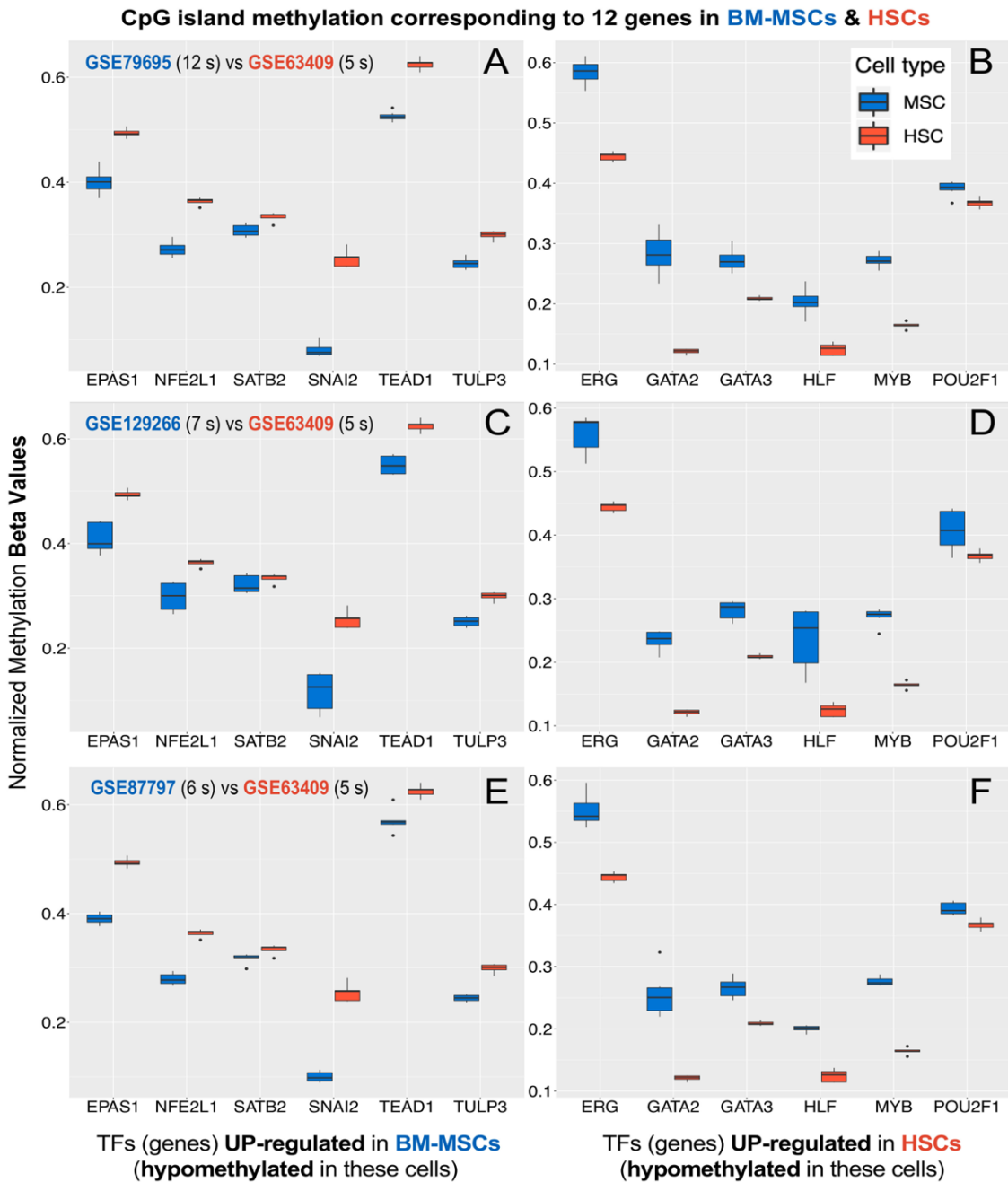


Figure 21. DNA normalized methylation boxplots of 12 MR in 3 independent plots of BM-MSC and 3 independent plots of HSC. Genes EPAS1, NFE2L1, SATB2, SNAI2, TEAD2 and TULP3 in upregulation of MSC and genes ERG, GATA2, GATA3, HLF, MYB and POU2F1 up-regulated in HSC. **A and B)** 12 MR up and down-regulated in GSE79695 vs. GSE63409. **C and D)** 12 MR up and down-regulated in GSE129266 vs. GSE63409. **E and F)** 12 MR up and down-regulated in GSE87797 vs. GSE63409.

e. Gene expression co-regulation networks in Master Regulators

In the last part of the results, we performed a regulatory network analysis with *Cytoscape* based in the results obtained with *VIPER*. These regulatory networks were represented as a bipartite graph containing two types of nodes: regulator TFs and *regulon* targets, and the directed links between TF and target. In this type of graph, the intensity of the colored nodes (genes) was proportional to the gene expression signal values, so that nodes were shown in red when they were up-regulated and in blue when they were down-regulated.

The **Figure 22** corresponded to the top up- and top 10 down-regulated regulators associated with their assigned gene *regulons*. Red nodes were the up-regulated MR in MSCs among which we can see again SNAI2 as one of the most intensity gene expression. This gene regulated genes such as CEBPB, EBF1, ERG, MYB, TGIF1, SERPINE2 and ZMAT1, genes associated with the regulation of the transcription and TF activity in progenitor states (data extracted from DAVID bioinformatics). Moreover, TEAD1 also presented a high intensity in the plot; its *regulons*: COPS6, KUF1BP, MYO6, NOTCH2, RAB31, PERP, OAT, LAMC1 and RGL2; and their functions were associated with phosphoproteins and cell membrane. TULP3 also had a high expression signal with genes such as CCND1, LAMB2, COL16A1, and THBS1, and regulates the PI3K/AKT signaling pathway, which plays a critical role in regulating various cellular functions, including metabolism, growth, proliferation, survival, transcription, and protein synthesis as well as in extracellular matrix organization, ECM–receptor interactions, cell adhesion, and integrin binding.

In contrast, the top 10 MR up expressed in HSC showed a different pattern of regulatory relationships. We found MYB as the strongest expression signal with important *regulons* such as AP2M1, ATP8A1, KDELR1, LAPTM5, TMED10, CDC42, TAGAP, BTK and UCP2. CDC42 is a gene present in the HSC that regulates the aging and correlates with a loss of polarity in aged HSCs (Florian et al., 2012), and all these genes are related to the vesicle transport and regulate EPS8, SYK, and YAP1, which are responsible for cell proliferation. IKZF1, GATA3 and HLF were less correlated and presented less expression and less group of *regulons*.

Figure 23 shows a more focused analysis on the top 10 up-regulated MR in MSCs and their associated *regulons*, as in the previous figure the up-regulated nodes in red and the down-regulated nodes in blue. Furthermore, in this **Figure 23**, we used the *iRegulon* tool, including the whole list of genes and TFs from *VIPER*, considering the TFBS of the gene network. We found 6 TFs that are more prominent in yellow ellipses. These transcription factors might play essential roles in regulating of the gene expression network associated with MSCs. We also marked in purple the MR corresponding to the hypomethylation profile analyzed in the previous section.

Considering EPAS1 also as a relevant gene due to its methylation, the associated *regulons* were MYL9, COL5A2, ACTN1, TMEM47, SERPINE2 and CDH11, where ACTN1 plays a relevant role in cell motility; notably, in the immune response it drives cells to sites of inflammation (Florian et al., 2012) and COL5A2 is presented as a TF that regulates the osteogenesis (Hou et al., 2021).

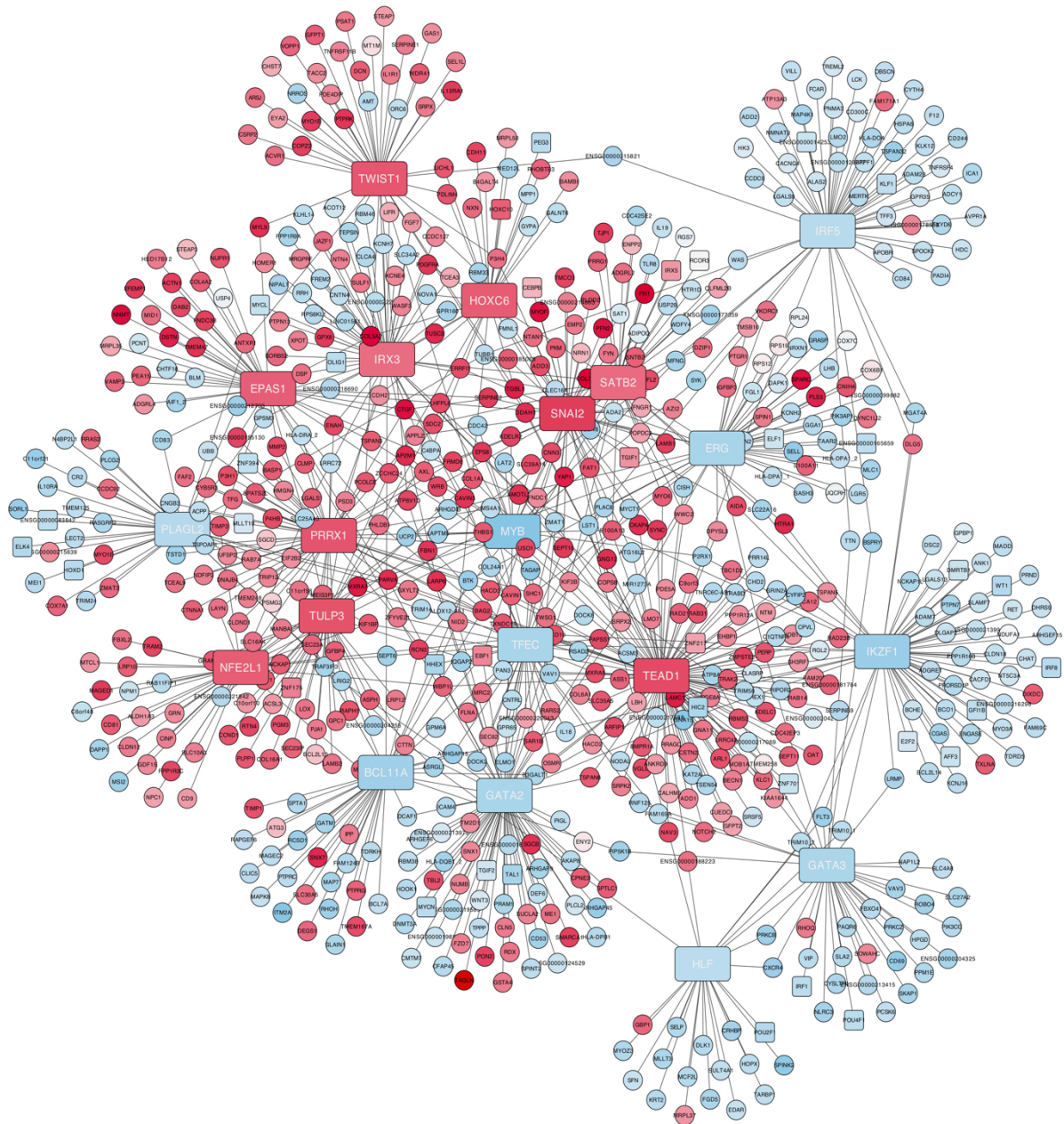


Figure 22. Gene correlation network with top 10 upregulated MR in MSC in red and top 10 downregulated MR in MSC in blue. The circles are the gene regulons associated, which are in red are upregulated and in blue are downregulated.

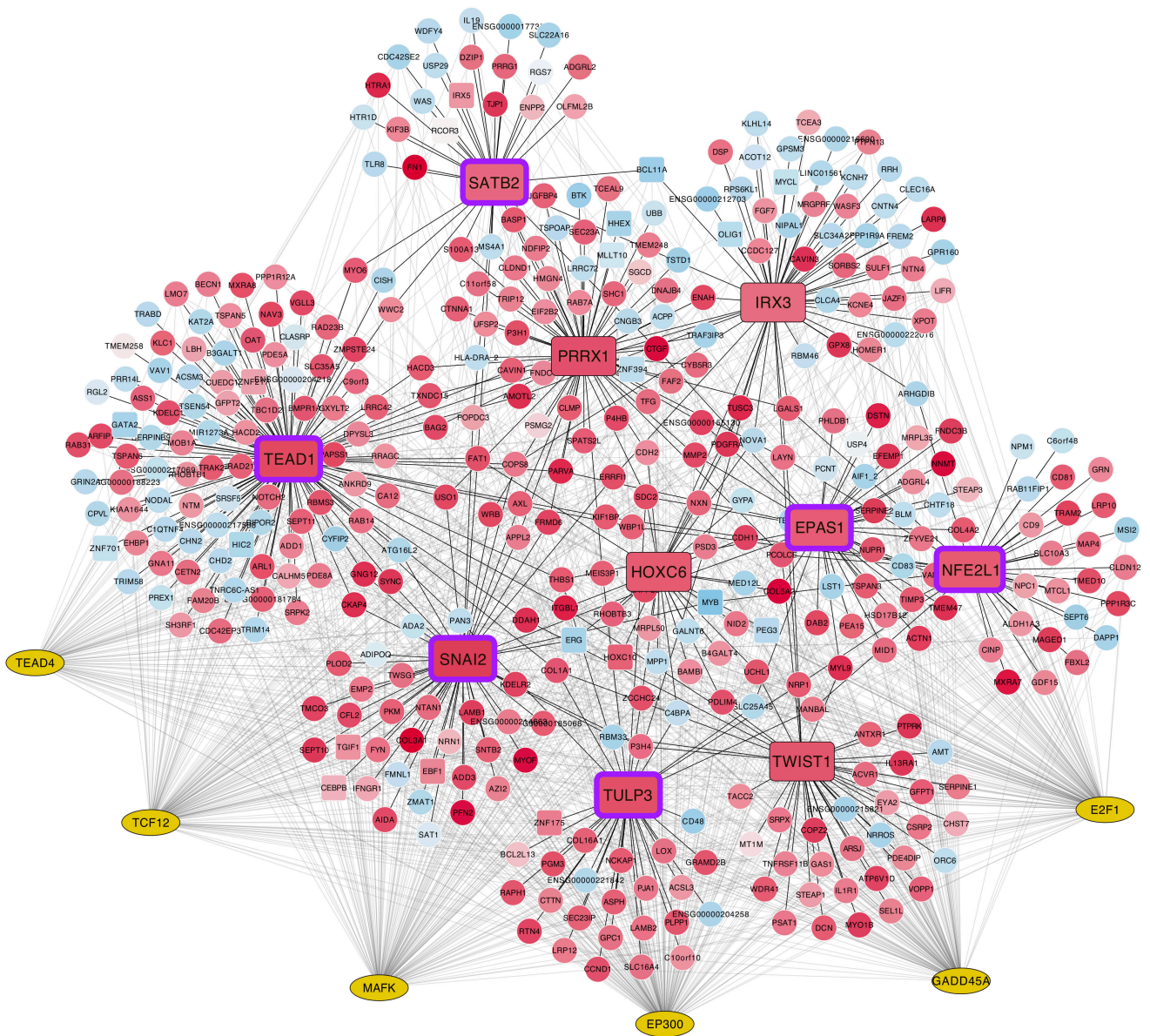


Figure 23. Gene correlation network with top 10 upregulated MR in MSC indicated as red rectangles: SNAI2, TEAD1, STAB2, IRX3, EPAS1, HOXC6, TULP3, TWIST1, PRRX1 and NFE2L1 with their associated regulons (in red which are upregulated and in blue the downregulated ones). The yellow ellipses are 6 promoter TFs E2F1, EP300, GADD45A, MAFK, TCF12, and TEAD4 enriched with the *iRegulon* tool.

3. DISCUSSION

The human bone marrow niche includes several non-hematopoietic cells, including Mesenchymal Stem Cells, osteoblasts, adipocytes, and endothelial and neural cells. These cells physically support Hematopoietic Stem Cells and regulate their homeostasis. Among them, MSCs play a key role in the bone marrow niche in the regulation and differentiation of the hematopoietic system (Crippa & Bernardo, 2018). Therefore, the study of genes and transcription factors is becoming increasingly important in the field of hematology.

In the first part of this chapter, the results obtained in the Meta-Analysis confirm the presence of known cellular lineages. Mesenchymal Stem cells, osteoblasts, adipocytes, chondroblasts and Fibroblasts show a strong correlation, suggesting similar gene expression profiles. Since osteoblasts, adipocytes and chondroblasts are derived from the differentiation of MSCs, it is logical that they exchange information with each other. Moreover, Fibroblasts are considered by many authors to be similar to MSCs due to their overlapping functions, especially as components of the extracellular matrix (Vaculik et al., 2012).

Interestingly, this stromal lineage is hardly related to the analyzed hematopoietic lineage (HSCs and lymphocytes). SNAI2 shows a significantly higher expression difference between the MSCs and the hematopoietic lineage, indicating the presence of EMT gene markers. This transition confers migratory and invasive properties to Mesenchymal Stem Cells; they become multipotent stromal cells that can differentiate into a variety of cell types (Lamouille et al., 2014). Collagen family genes also show differential expression in contrast to the hematopoietic cells, as shown in the Heatmap of Figure 19, especially COL1A2, COL3A1, COL5A2, COL6A3 were up expressed. Moreover, these genes are present in the PI3K-AKT pathway, which plays a pivotal role in cell metabolism, growth, proliferation and survival (Hemmings & Restuccia, 2012).

The results of the comparison between MSCs and stMSCs show that certain genes are down expressed in stMSCs, particularly VCAM1 and IL6, which are involved in inflammation signaling and regulate JAK/STAT and p53 pathways, respectively, to maintain the homeostasis (Ward & Hudson, 2014). These findings are important for the regulation of the stromal niche. stMSCs with TGF- β stimulation induce the production of M2 macrophages, which are responsible for the anti-inflammatory properties, inhibit the recruitment of CD68+ monocytes to the site of inflammation and exert anti-inflammatory effects through the activation of T cells Th1/Th17, NK cell activation and Treg induction (de Araújo Farias et al., 2018). Otherwise, in the MSCs versus OSTB comparison, up expressed genes in OSTB are associated with collagen fibril organization and ECM functions. The Collagen family genes are up expressed in OSTB: COL15A1, COL15A1. VCAM1 is also overexpressed, which are involved in skeletal development and bone homeostasis, and these genes are also activated during the response of bone cells to mechanical stress (Canalis et al., 2012; Zaman et al., 2012).

The subsequent focus shifts to the identification of Master Regulators for Mesenchymal Stem Cells. Among them, 10 up-regulated MR genes and 10 down-regulated MR genes are discovered, which contribute to a better characterization of this cell type. By studying the functionality of these genes in detail, it was found that the most of the 10 up-regulated MR in MSCs are related to the epithelial-mesenchymal transition, especially SNAI2 and TWIST1. This feature is essential for cell migration, as well as in metastasis, as it gives cells cancer cells the ability to migrate to other tissues (Mani et al., 2008). SNAI2 is also involved in wound healing and organogenesis (Ganesan et al., 2016), and previous studies in our laboratory were identified SNAI2 in the MSC lineage signature of transcription factor genes (Roson-Burgo, Sanchez-Guijo, Del Cañizo, et al., 2016). The methylation studies support the notion that DNA

methylation alters the drive of EMT, promoting it and contributing to resistance to various therapeutic agents in cancer (Galle et al., 2020), this relationship is also been demonstrated in induced pluripotent stem cells (iPSC) (Chen et al., 2013). Furthermore, the TWIST1 gene has also confirmed as a master regulator of EMT, which plays an essential role in tumor initiation, stemness, and angiogenesis in a variety of human cancers, especially in pancreatic cancer (Wang et al., 2020). In addition, several MRs, such as IRX3, SATB2, TULP3, and TWIST1, are associated with nervous system functions, including neuron generation, neurogenesis, and nervous system development, as MSC-derived growth factors and extracellular matrix components influence endogenous neuronal cell activity (Maltman et al., 2011). In the case of STATB2, we also found a correlation of hypomethylation of this gene with the osteogenic differentiation, facilitating bone formation and regeneration (Q. Han et al., 2015).

TEAD1 emerges as an important MR in our analysis, as there were no published articles on the relationship between the regulation of MSCs and TEAD1. However, other studies highlight the importance of TEAD1 as a regulator of cardiomyocyte function with potential therapeutic implications (Liu et al., 2017). The TEAD1 gene regulates the Hippo pathway, which is essential for the regulation of cell growth and proliferation (Li et al., 2022). The hypomethylation of TEAD1 gene, which has been reported to be a requirement for the ability of MSCs to differentiate lineage, gains importance as a MR. These studies also suggested that aberrant DNA hypermethylation of the loci of genes of the TEAD1 gene could alter the role of the development of BM-MSCs and could promote malignant as multiple myeloma (Garcia-Gomez et al., 2021). EPAS1, another hypomethylated gene, is also critical in the role of the cells from bone marrow and adipose MSCs, linking these data with the possible stem potential (Aranda et al., 2009). Additionally, EPAS1 has been implicated in promoting differentiation in BM-MSC (Zhang et al., 2021). Contrasting this information with the functional enrichment results, the set of MRs and their interacting genes are involved in numerous functions such as cell differentiation or cell matrix formation, which are critical functions for the MSCs.

Conversely, Master Regulators that are not expressed in MSCs have essential functions in the hematopoietic lineage. IKZF1, BCL11A, and their associated genes are involved in hematopoietic processes, activation of T cells or in the development of the immune system as is shown in the functional enrichment. Any mutation they suffer any of these MRs has serious consequences in the organism causing the development of hematopoietic disorders (Bielska et al., 2017; Gao et al., 2015). Notably, BCL11A is essential for normal lymphocyte development and a potential biomarker and therapeutic target in many hematopoietic diseases, including lymphoma or leukemia. This gene has functions primarily as a transcriptional repressor and is essential for multiple cell lineages, including B-cell development, plasmacytoid dendritic cells maturation, and maintenance of stemness in stem cell (Wang et al., 2020).

In addition to identifying transcription factors as Master Regulators, additional analysis was performed using the *Cytoscape* platform and the *iRegulon* algorithm to determine whether the identified the MR shared common sequences in the junction regions to the promoter. While only TEAD1 appeared as a MR after the application of *iRegulon*, new transcription factors were found. This set of new TF are: EP300, GADD45A, E2F1, SPL1 EP300, GADD45A, E2F1, SPL1. Subsequent analysis of these genes in different samples, indicated the possible functions of these genes are mostly expressed in the bone marrow (Fagerberg et al., 2014). However, further studies are needed to validate these findings due to the sparse bibliography on these additional TF candidates.

4. FINAL SUMMARY of CHAPTER 2: *Master regulators of MSCs*

The human bone marrow is a complex and dynamic niche composed of various non-hematopoietic cells, including stromal cells, endothelial cells, osteoblasts, adipocytes, and others. Several of these cells play critical roles in providing physical support to HSCs and regulating their homeostasis. MSCs are particularly important in this context because they are multipotent stem cells that influence the regulation and differentiation of the hematopoietic system. Therefore, research focused on gene regulators and transcription factors is of great importance to understand the molecular balance in the bone marrow microenvironment.

In this chapter, we present the results corresponding to the construction of a gene regulatory profile of human MSCs, generating a compendium of more than two hundred cell samples with genome-wide expression data. These data include five related primary cell types found in the bone marrow: BM-MSCs, HSCs, lymphocytes (LYM), fibroblasts (FIB), and osteoblasts (OSTB). These cell types show robust interrelationships, suggesting correlated expression patterns, possibly due to common lineage origins. Using the complete transcriptomic profiles of these cells, our study identifies master regulators for MSCs, which in order of significance were: SNAI2, STAB2, IRX3, EPAS1, HOXC6, TWIST1, TULP3, PRRX1, TEAD1, and NFE2L1. The SNAI2 and TWIST1 genes are associated with EMT processes and nervous system functions. In MSCs, SNAI2 exhibits a significant differential expression compared to hematopoietic cells, indicating the presence of markers associated with EMT. EMT confers MSCs with migratory and invasive properties, contributing to their multipotency. Additionally, TEAD1 emerges as an important gene regulator with potential implications in cell growth, while EPAS1 and other hypomethylated genes play a role in cell differentiation and potential stem cell properties. Finally, comparisons between MSCs and stMSCs revealed differential expression of genes such as VCAM1 and IL6, which are involved in the regulation of inflammation. These findings highlight the importance of stMSCs in maintaining homeostasis and their interactions with immune responses, including the modulation of macrophage polarization and T cell regulation.

In addition, our study found a number of master regulators that were not expressed in MSCs but over expressed in HSCs, which were also ranked by significance were: BCL11A, MYB, TFEC, HLF, GATA2, ERG, PLAGL2, DACH2, POU2F1, and GATA3. The IKZF1 and BCL11A genes are essential for hematopoiesis and the development of the immune system, and mutations in these genes cause hematopoietic disorders. BCL11A is particularly important in several cell lineages, including B-cell development and stem cell maintenance. Further analysis using the *iRegulon* algorithm identified potential transcription factors such as EP300, GADD45A, E2F1, and SPL1, suggesting a role in bone marrow function. However, these findings require additional validation due to the limited existing literature on these candidate TFs.

In conclusion, this study provides valuable insights into the gene expression profiles and regulatory mechanisms within the human bone marrow niche and sheds light on the role of MSCs and other stromal cells in hematopoiesis and tissue homeostasis.

CHAPTER 3:

Analysis of the proinflammatory or immunosuppressive effect of MSCs through stimulation

Short Title:

Stimulation of MSCs towards MSC1 and MSC2

1. MATERIAL AND METHODS

a. Characterization of MSCs

All Mesenchymal Stem Cells used in the current study were derived from the bone marrow of healthy donors. All the procedures were carried out in strict compliance with the institutional standards of ISCT and the University Hospital of Salamanca (HUS).

b. TLR4 and TLR3 stimulated RNA-seq datasets

In this Chapter 3, all data analyses were performed in R as described in the Materials and Methods section of Chapter 1. The RNA-seq data was a compendium of 21 samples from 3 data batches of MSCs. Sample identifiers were extracted from the NCBI under the accession numbers: GSE81478, GSE97723, GSE109181 and from platforms: Illumina HiSeq 2000 and Illumina HiSeq 2500, using paired-end sequencing. The samples were obtained from two different healthy donors; 6 samples (GSM) were stimulated with poly(I:C) 10µg/ml, 3 samples with LPS 10ng/ml and 3 samples with LPS 1 µg/ml, the other 9 samples were controls; as shown in **Table 11**.

Table 11. RNA-seq datasets with the samples associated, the labeled patients, the stimulation being with LPS, poly(I:C) or control (CTL), the platform where the data was analyzed and the type of sequencing: paired or single, and the corresponding articles.

GSE	GSM	PATIENT	STIMULATION	Patient	PLATFORM	TRIMMING	ARTICLE		
GSE81478	GSM2154690	Lonza, donor 7F3674	CTL LPS	1	Illumina HiSeq 2000	PAIRED	Kim SH, Das A, Chai JC, Binas B et al. Transcriptome sequencing wide functional analysis of human mesenchymal stem cells in response to TLR4 ligand. Sci Rep 2016 Jul 22;6:30311. PMID: 27444640		
GSE81478	GSM2154691								
GSE81478	GSM2154692								
GSE81478	GSM2154693	Lonza, donor 7F3674	LPS 10ng/ml	1	Illumina HiSeq 2000				
GSE81478	GSM2154694								
GSE81478	GSM2154695								
GSE97723	GSM2576182	Lonza, donor 7F3674	LPS 1ug/ml	1	Illumina HiSeq 2500	PAIRED	Kim SH, In Choi H, Choi MR, An GY et al. Epigenetic regulation of IFITM1 expression in lipopolysaccharide-stimulated human mesenchymal stromal cells. Stem Cell Res Ther 2020 Jan 7;11(1):16. PMID: 31910882		
GSE97723	GSM2576183								
GSE97723	GSM2576184								
GSE109181	GSM2934991	Lonza, donor 7F3674	CTL Poly(I:C)	1	Illumina HiSeq 2500			PAIRED	Kim SH, Das A, Choi HI, Kim KH et al. Forkhead box O1 (FOXO1) controls the migratory response of Toll-like receptor (TLR3)-stimulated human mesenchymal stromal cells. J Biol Chem 2019 May 24;294(21):8424-8437. PMID: 30944148
GSE109181	GSM2934992								
GSE109181	GSM2934993								
GSE109181	GSM2934994	Lonza, donor 7F3674	Poly(I:C) 10ug/ml	1	Illumina HiSeq 2500				
GSE109181	GSM2934995								
GSE109181	GSM2934996								
GSE109181	GSM2934997	Lonza, donor 0000127756	CTL Poly(I:C)	2	Illumina HiSeq 2500				
GSE109181	GSM2934998								
GSE109181	GSM2934999								
GSE109181	GSM2935000	Lonza, donor 0000127756	Poly(I:C) 10ug/ml	2	Illumina HiSeq 2500				
GSE109181	GSM2935001								
GSE109181	GSM2935002								

First, we performed a quality analysis using the *Trimmomatic* tool to remove the adapters and poor quality *fastqc* samples (Bolger et al., 2014). Then, we converted the *fastqc* files to count data using *STAR* and *RSEM* tools. *STAR* is a tool compiled in Linux open-source command line used to the alignment of samples to the reference genome Homo_sapiens.GRCh38.79.gtf, while *RSEM* was used to quantify and calculate the expression of the read alignments in counts. It's important to note that the choice to use *STAR* instead of *SALMON*, used in the previous chapter, is because *Salmon* is not a true mapper, but a pseudo-aligner, as we commented in the Introduction. *STAR* provides more accuracy in quantifying low expressed genes and is much better suited for isoform quantification (Dobin et al., 2013; Li & Dewey, 2011). Finally, we introduce the transcript data into R statistical programming using the *tximport* package from Bioconductor (Soneson et al., 2015).

c. *In vitro* Cell Cultures

All the human bone marrow samples were obtained from iliac crest aspirates with a volume of 10 to 15 ml according with the institutional standards (Villaron et al., 2004). The blood obtained from bone marrow was diluted in 1:3 of HBSS (Gibco, Thermo Fisher Scientific). The mononuclear cells were isolated by gradient centrifugation using Ficoll-Paque (GE Healthcare Biosciences), 1600 rpm, 25 minutes.

MNCs were obtained from the buffy coat layer and then the cells were seeded on plastic flasks at a concentration of 10^5 MNCs/cm² with DMEM 1g glycerin/l (Gibco, Thermo Fisher Scientific) supplemented with 10% FBS and 100 U/ml penicillin/streptomycin (Gibco, Thermo Fisher Scientific). MNCs were cultured at 37°C, 90% of humidity and 5% of CO₂. The medium was renewed every 3-4 days to eliminate no-adherent cells. The MSCs cells were cultured to 90% of confluence, where they were tripzinated with Tripsin (Gibco, Thermo Fisher Scientific) and counted with trypan blue (Gibco, Thermo Fisher Scientific) in a dilution (1:100) using a Neubauer chamber. The cells then were seeded at a concentration of 2500-5000 cells/cm² to allow their expansion and growth. All MSCs were grown *in vitro* from passage 3 to 5 depending on the experiment (Minguell et al., 2001). The procedure is shown in **Figure 24**.

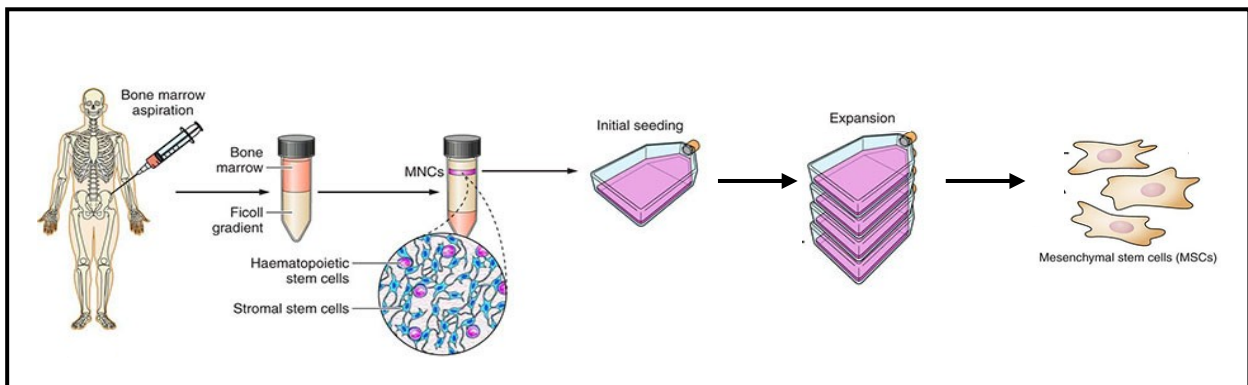


Figure 24. Schematic representation of the isolation and expansion of MSCs.

d. TLR3 stimulation of MSCs with poly(I:C)

MSCs were cultured to passage 4, in confluence of 70-85%, as indicated the section of “*In Vitro* Cell cultures” section of Materials and Methods. Poly(I:C) sodium salt (Sigma-aldrich) was diluted in 1 ml of DMEM 1g glycerin/l supplemented with 10% FBS and 100 U/ml penicillin/streptomycin from the stock of 1 mg, because poly(I:C) is soluble in 10 mg/ml. We then made a second dilution until 1mg/ml in DMEM + 10% FBS. We prepared dilution of 25µg/ml poly(I:C) was added in cultured cells and incubated at different times of 6 hours and 24 hours in the incubator with a temperature of 37°C, 90% of humidity and 5% of CO₂. After this incubation period, the cells were washed twice with PBS and DMEM + 10% FBS was added again.

e. RNA-sequencing Technique

The RNA-seq analysis was performed on a total of 9 samples from three healthy donors, with each condition consisting of 3 samples: Control, MSCs stimulated with poly(I:C) 25µ/ml at time 6 hours and MSCs stimulated with poly(I:C) 25µ/ml at time 24 hours. For the sample preparation, each sample was resuspended in 100µl of mixture of: 99µL of (Lysis/Binding Buffer for Dynabeads (Invitrogen), 1µL b-mercaptoethanol (Sigma-Aldrich) y 100 µL de isopropanol (Sigma-Aldrich), and added in labeled Eppendorf. Mixed samples were vortexed and stored at -80°C until they were ready for RNA-seq analysis. Sequencing was performed using mRNA-seq bulk in full length mRNA and paired-end reads, resulting in a sequencing depth of 30 million reads. The sequencer used was the Illumina NextSeq2000.

f. Differential Expression Analysis

As we explained in Materials and Methods section of the previous Chapter 1, we also used the *Limma* package for Differential expression Analysis. In the case of RNA-seq data, the variant of *Limma* package is *Limma-Voom*. *Limma-Voom* is specialized version of *Limma* package based on empirical Bayesian smoothing of gene-wise standard deviations, but especially for RNA-seq data (Law et al., 2014). The analysis of public NCBI data of stimulated MSCs with LPS and poly(I:C) was performed using different parameters of p-value (0.05) and logarithm Fold Change (no fold change, 1, 1.5 and 2). The analysis of different stimulation times with poly(I:C), we performed different analyses with p-value 0.05 and 0.01 combining each one with the different log(FC): no fold change, 1, 1.5 and 2. The markers obtained from the transcriptomic analysis, particularly those associated with differentiation, were obtained from different scientific articles, such as Robert et al., 2020; Sauer et al., 2022; and Sekiya et al., 2003. These sources provided relevant information for the identification and characterization of genes related to the differentiation lineages in MSCs (Robert et al., 2020; Sauer et al., 2022; Sekiya et al., 2003).

g. Functional Enrichment Analysis

This section was also explained in the Materials and Methods section of Chapter 1. Functional enrichment was performed using the DAVID bioinformatics platform (<https://david.ncifcrf.gov/>). This platform is commonly used to discover enriched functional annotations among differentially expressed genes, providing insights into their biological roles and potential pathways. Additionally, we used *clusterProfiler* package from Bioconductor to reveal the clusters associated with the differential expression analysis results. This package allows the identification and exploration of gene clusters with similar expression patterns. Moreover, the package performed network plots and dot plots to visualize and interpret the relationship of the enriched gene sets and their associated functions (Wu et al., 2021).

h. Differentiation assays of MSCs: Adipogenic differentiation

MSCs were plated and grown in plates of 9 cm² (Nunc, Roskilde) at a concentration of 2·10⁵ cells/cm² with DMEM supplemented with 10% FBS at 37°C, 90% of humidity and 5% of CO₂. When the cells were in 80% of confluence, the adipogenic differentiation medium Adipodiff (Lonza) was added and the cells were maintained with the medium changes twice a week for 21 days. The cells were then fixed with 4% paraformaldehyde and stained with Oil-Red-O during 30-45 minutes. Differentiation into adipocytes was determined by the bright reddish-orange stained vacuoles. The vacuoles were observed and photographed using an Olympus DP70 camera on the optical microscope Olympus31.

The adipogenic differentiation and counting were performed on 4 samples. Each sample contained the control MSC, stimulated MSC with 25µg/ml poly(I:C) at the time of 6 hours and stimulated MSC with 25µg/ml poly(I:C) at the time of 24 hours. For each sample, we selected ten random points around the slide of the plate. At these selected points, adipocytes were counted in each field using an optical microscope (Olympus31) at day 7, 14 and 21. These data were subjected to ANOVA (Analysis of Variance) statistical analysis, which allowed for the evaluation of significant differences between different conditions. Then, the pairwise post-Hoc comparison of the difference between groups was performed using the Tukey test, which provided valuable results in the pairwise comparisons of adipocyte counts between different time points and stimulation conditions.

i. Differentiation assays of MSCs: Osteogenic differentiation

MSCs were plated and grown in plates of 9 cm², in a concentration of 3·10⁴ cells/cm² in duplicate (control and osteogenic differentiation) with DMEM supplemented with 10% FBS at 37°C, 90% of humidity and 5% of CO₂ for the conditions: control MSC, stimulated MSC with 25µg/ml poly(I:C) at time 6 hours and stimulated MSC with 25µg/ml poly(I:C) at time 24 hours. When the cells were in 30-40% of confluence, the controls of each condition were not change the medium, while the medium of the samples for each condition to osteogenic differentiation was replaced with Osteodiff (Miltenyi Biotec). Both medium were changed twice a week for 10 days. The staining was evaluated with Osteoblastic alkaline phosphatase staining by NBT/BCIP colorimetric reactions (nitroblue tetrazolium chloride / 5-bromo-4-chloro-3-indolyl-phosphate) (Roche). Osteogenic differentiation was observed by morphology and the staining with a deep purple color, while the background was purple. These results were observed and photographed using an Olympus DP70 camera on the optical microscope. The samples were then quantified by to Alizarin Red method, using a solution of 20% methanol and 10% acetic acid in water. After 15 minutes, the liquid is transferred to a 96-well plate and the amount of Alizarin Red is read on the spectrophotometer at a wavelength of 450nm (Trivedi et al., 2020). Values were analyzed using the formula:

$$\text{Quantification value of the samples} = \frac{\text{Absorbance value} \cdot 1000}{1100}$$

Statistical analysis was then performed in the same manner as for the MSC Differentiation assays: Adipogenic Differentiation section of Materials and Methods, using ANOVA and post-Hoc tests.

2. RESULTS

a. Quality control of stimulated MSCs with LPS and poly(I:C)

To identify and study the relationship between changes in gene expression and the immunomodulatory stage of the MSCs, we performed a comprehensive analysis of the whole transcriptomic signal in the different samples collected in different datasets of RNA sequencing data, described in **Table 11**. Three conditions of differential expression analysis were performed: one with samples stimulated with LPS, the second one with a concentration of 10 μ g/ml of poly(I:C) and the last one with the respective Controls. We use 2 datasets with 6 samples of 10 μ g/ml of poly(I:C), 2 datasets with 3 samples of 1 μ g/ml LPS and 3 samples of 10 ng/ml LPS and 2 datasets shared with the stimulated samples with 9 samples of Control. That included a total of 4 datasets with a total of 21 samples, as was indicated in **Table 11** of the Materials and Methods.

We first performed a PCA and Kernel density plot of the samples to assess whether they had the same distribution and dimensionality in the different datasets along the 3 datasets and were comparable between them (**Figure 25**). The initial matrix obtained with raw counts contained 29190 genes and 21 samples.

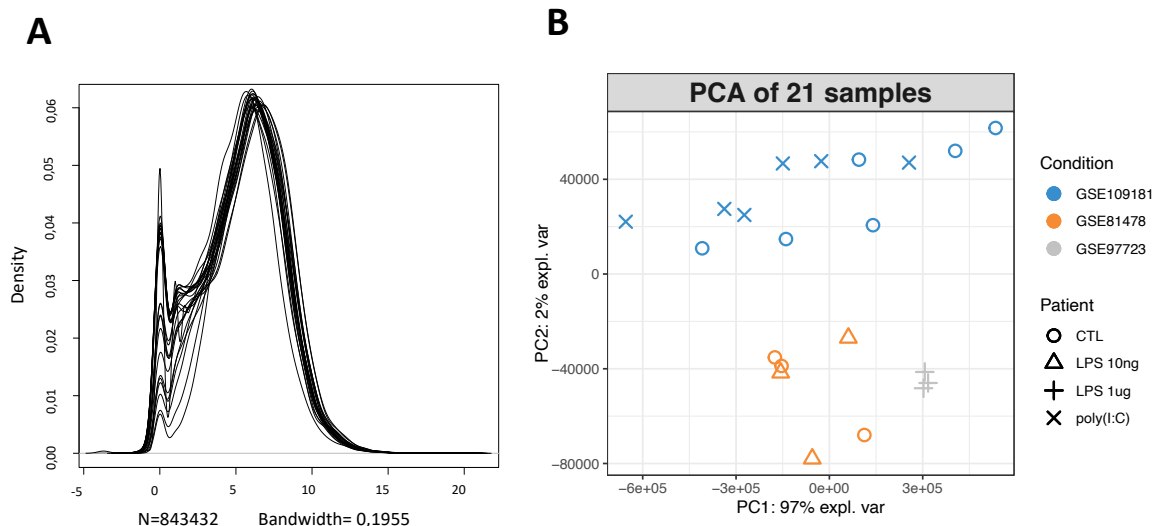


Figure 25. Quality Control analysis of 21 samples from 3 datasets: GSE1478, GSE7723, GSE109181.

A) Density Plot. **B)** PCA plot in two dimensions.

The Kernel plot (**Figure 25A**) was generated with the raw counts and revealed the presence of genes with 0 in counts values. The RNA-seq curve displayed a Gaussian fit on the right side of the main peak, while the left peak corresponds with $\text{Log}_2(\sim 1)$ counts. In fact, for each sample, the total number of reads tends to be in millions, while the counts per gene can vary considerably, tending to be in the tens to thousands. Consequently, the probability of a given read being mapped to a specific gene is small. This plot highlights the importance of pre-processing the samples in each dataset to filter out genes with low counts, as they can interrupt the analysis and lead to false positives in the contrasts.

Figure 25B shows the batch effect between the different datasets using the PCA plot. All samples were almost completely explained in the first dimension with $\text{PC1} = 97\%$. We performed a Meta-Analysis

approach combining the data as a function of stimulation of MSCs. The Control versus LPS analysis was performed with GSE1478 at a concentration of 10ng/ml LPS and GSE7723 at a concentration of 1µg/ml of LPS. These two datasets were analyzed independently. Similarly, poly(I:C) versus control analysis, only one dataset was analyzed: GSE109181 with a poly(I:C) concentration of 10 µg/ml and controls.

To perform a more in-depth analysis and to establish patterns of genes expressed and not expressed in the MSCs, we considered performing the differential expression analysis separately in each condition. These conditions were divided into two groups: one for the putative proinflammatory stage of LPS stimulation versus de respectively controls, and another for the putative immunosuppressor stage of poly(I:C) stimulation versus respectively controls. By performing separate analyses for each condition, we can more precisely characterize the specific gene expression changes associated with each type of stimulation. This approach will help us identify distinct gene signatures and gain a deeper understanding of the molecular mechanisms underlying the immunomodulatory properties of MSCs in response to LPS and poly(I:C) stimulation.

b. Differential and Functional Analysis of LPS stimulated MSCs

The GSES associated with the proinflammatory stage were led by the LPS stimulations and the respective controls in the datasets, corresponding to: GSE81478, GSE97723 and GSE109181.

According to the **Table 11**, LPS data were performed in two concentrations 1 µg/ml and 10ng/ml. To avoid missing the possible differentially expressed genes expressed in the different concentrations, the analysis was performed separately versus the controls.

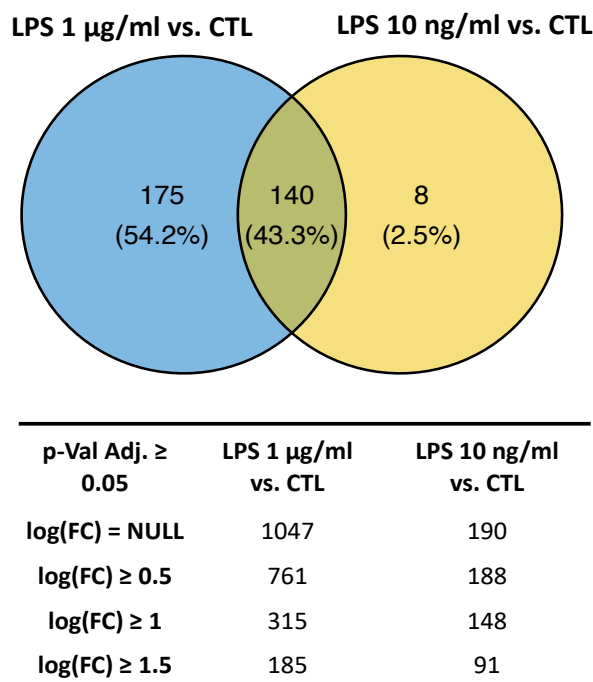


Figure 26. Venn Diagram of the interaction between MSCs stimulated with LPS 1 µg/ml vs. CTL and MSCs stimulated with LPS 10 ng/ml vs. CTL. The values associated correspond to the result with $\log(FC) \geq 1$ of the 4 comparisons of *Limma-Voom* Differential analysis.

The analysis with *Limma-Voom* package and the filtering by FDR p-value ≥ 0.05 showed a high cut-off of genes with 1047 genes in LPS 1 $\mu\text{g/ml}$ and 190 genes in 10 ng/ml . However, when the filtering was also done by the logarithm with the absolute number of fold change ($\log(\text{FC})$), we could observe that the filtering in the case of the LPS 1 $\mu\text{g/ml}$ was reduced to 315 genes in $\log(\text{FC}) \geq 1$ and to 148 genes in LPS 10 ng/ml whereas in $\log(\text{FC}) \geq 1.5$ the reductions were too strict with numbers of 185 and 91 genes, respectively (**Figure 26**).

These data revealed the large magnitude of genes in concentration 1 $\mu\text{g/ml}$ of LPS with fold change lower than 1, indicating the wide range of genes with a few variations in the differential gene analysis. On the other hand, the genes involved in the differential gene analysis of LPS 10 ng/ml versus control were stable along all the fold change which implied a robustness in the results. Aiming to obtain differential expression genes with sufficient significance and power, we worked with a filter of p-value adjusted ≥ 0.05 and $\log(\text{FC}) \geq 1$. In the intersection of both concentrations the number of genes up expressed was 131 and 9 down expressed.

The down expression cluster of 9 genes in LPS including ID1, MACIR, FGF1 and MLPH was also represented in a Heatmap, as shown in **Figure 27**. ID1 is a known marker of endothelial progenitor cells and plays an important role in EMT, in the stimulated LPS MSCs, the $\log(\text{FC})$ was down expressed with a value of -1.30 in both concentrations.

In contrast, the 131 up expressed genes in the stimulated MSCs were clustered into 7 groups, also represented in **Figure 27**. The cluster groups were composed of a main group of 118 genes, the second group reduced to 5 genes, two groups of 2 genes and four groups of 1 gene. According to the functional analysis, the main group was associated with defense response to virus, innate immunity. Moreover, this cluster was associated with cellular response to tumor necrosis factor, interleukin-1 and interferon-gamma. In this collection of genes, according to the previous data: BAMBI, NFX3-1, BMP2, CSF1, EREG, EMP2, IL1B, IL6, S1PR3 and TSLP are involved in the regulation of cell proliferation, in this cluster C3, ICAM1, IL6 and VCAM1 genes are relevant in the inflammatory response and BAMBI, BMP2, IL1N and IL6 are genes involved in EMT. The clustering of these genes once again supports the association of the proinflammatory MSC with the response functions of the immune system, causing an inflammatory alert system that leads to action on the entire cellular niche, as we can well observe in the case of macrophages.

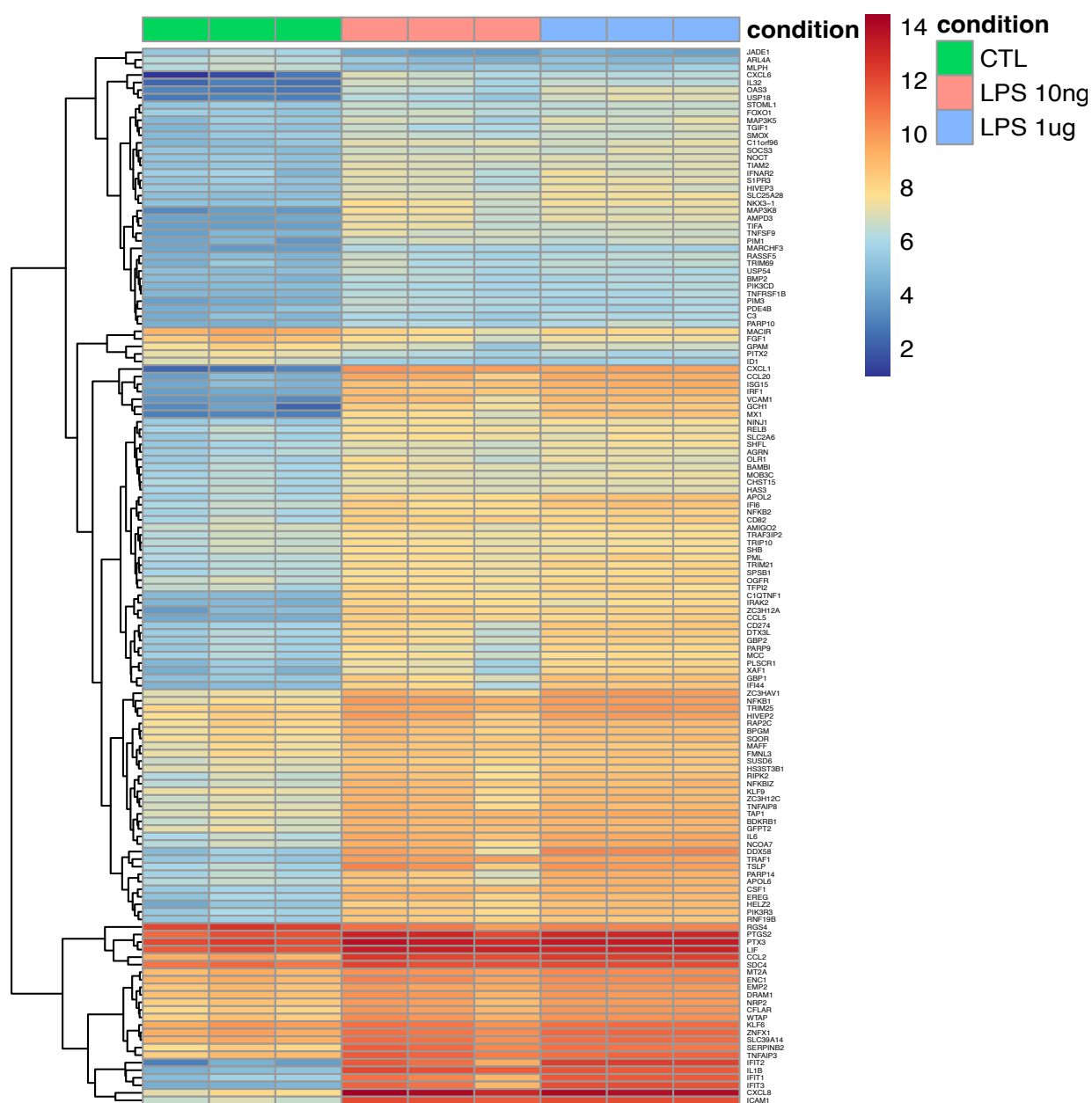


Figure 27. Heatmap of MSCs: CTLs of LPS, LPS 10ng/ml and 1 μ g/ml in 140 common genes in LPs stimulation. For each sample there were 3 technical replicates. The results were scaled in logarithm (cpm+1).

c. Differential and Functional Analysis of Poly(I:C) stimulated MSCs

The samples associated with an immunosuppressive stage were the poly(I:C) 10 µg/ml stimulated MSCs. The unique dataset associated with this stimulation, as shown in **Table 11**, was GSE109181. The analysis was performed as described in Materials and Methods with *Limma-Voom* and the filtering by p-value ≥ 0.05 . The values had low power, so the p-value was not filtered by the adjusted p-value (FDR). Considering the initial raw matrix of 29190 genes, the first filter of p-value reduces the number of genes to 419 genes. As in LPS stimulated MSCs analysis, the second filter was performed with the logarithm of absolute number of log(FC). As we could observe in the **Table 12**, the highest relative decrease in the percentage of the number of genes was from the $\log(\text{FC}) \geq 0.5$. Note that the analysis was performed with the filters of p-value ≥ 0.05 and $\log(\text{FC}) \geq 0.5$, obtaining a total of 207 genes within 120 genes were up and 87 were down expressed. Compared with the LPS stimulated MSCs analysis, the percentage between up and down expressed was more proportional; still, there is a slight increase in up regulated genes (**Table 12**).

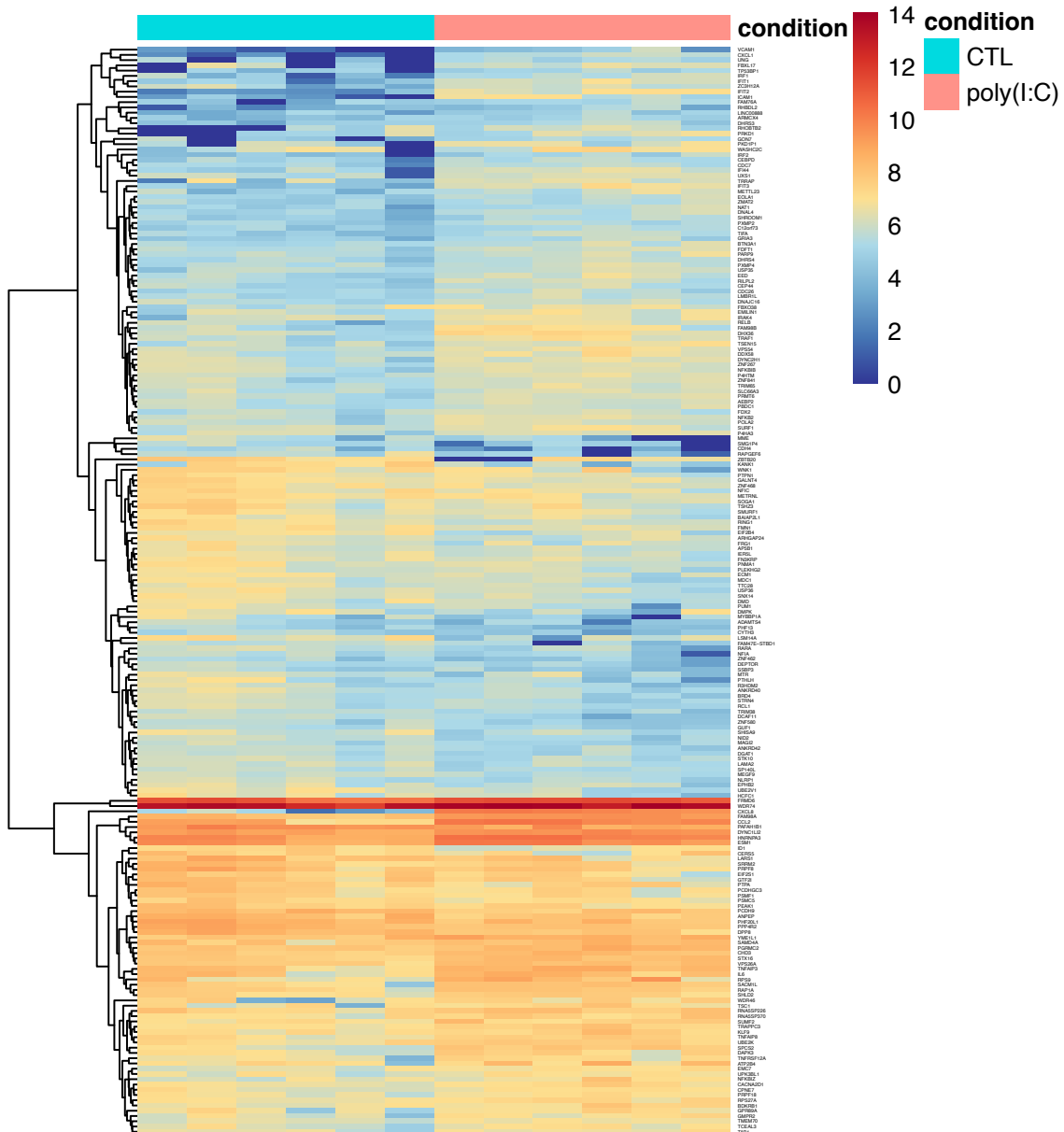
Table 12. Differential analysis of MSC stimulated with Poly(I:C) 10µg/ml vs. CLT of MSCs in two patients with 3 technical replicates each one. The table represent the differential genes in without log(FC), $\log(\text{FC}) \geq 0.5$, $\log(\text{FC}) \geq 1$ and $\log(\text{FC}) \geq 2$. Also are represented the up and down- expressed values.

p-Val ≥ 0.05	Poly(I:C) 10µg/ml vs. CTL		
		UP	DOWN
$\log(\text{FC}) = \text{NULL}$	419	244	175
$\log(\text{FC}) \geq 0.5$	207	120	87
$\log(\text{FC}) \geq 1$	58	37	21
$\log(\text{FC}) \geq 2$	20	15	5

The Heatmap and clustering was performed with the total of 207 genes, where 120 genes were up and 87 down expressed (**Figure 28**). The main group contained 151 genes. In this group there were functions related to cell regulation. It was interesting that specific genes ARHGAP24, TNFRSF12A, ANPEP, ECM1 and ZC3H12A were involved in the same functional cluster of angiogenesis and cell differentiation produced by DAVID Bioinformatics. ECM1 is an important prognostic marker for the invasion of cancer and the migration. In addition, ECM1 controls the expression of genes involved in EMT properties that present more mesenchymal invasive phenotype. In the analysis, the ECM1 gene was down regulated, with a $\log(\text{FC}) = -0.56$ and p-value = 0.034.

In this main group, the already known viral related functions become important, as determined by the functional analysis. Genes such as NFKBIB, IRAK4, PSMC5, TAP1, IL6 and CYTH3 presented functions associated with Epstein-Barr virus, COVID-19 or Influenza A. Moreover, FRG1, PRPF18, PRPF8, SRRM2 and ZMAT2 were involved in functions related to alternative splicing and the spliceosome, which is essential for cell function and defective pre-mRNA splicing causes disease. FRG1 and SRRM2 were also down expressed when the MSCs were stimulated with Poly(I:C), with a $\log(\text{FC}) = -0.623$ and $\log(\text{FC}) = -0.502$, respectively. Finally, ZMAT2 was up expressed in stimulated MSCs with a $\log(\text{FC}) = 0.688$. ZMAT2 is an interactor of the pre-spliceosome that is required to maintain cells in an undifferentiated, proliferative state.

The second largest cluster was composed of 32 genes. DDX58, TIFA, IFIT1, IFIT2 and PRKD1 were some of the genes related to the antiviral defense and innate immunity. All of them were up expressed but IFIT1, IFIT2 and PRKD1 genes had remarkably high values of fold change: $\log(\text{FC}) = 2.142, 4.096$ and 3.590 , respectively. The relationship between IFIT family (interferon-related genes) and poly(I:C) stimulation in MSCs is well known, as we commented in the Introduction section in the TRIF-mediated TLR3 signaling. The third cluster was much smaller than the previous ones with only 17 genes. Again, CCL2 and CXCL8 were up expressed together with TNFAIP3. Finally, the last third cluster consisted of 3 genes: PPP4R2, RING1, DPP8, which formed a down expressed cluster in stimulated poly(I:C) MSCs cells.



d. Stimulation of MSCs with poly(I:C): short treatment 6h versus longer treatment 24h

As we can see in **Figure 28**, the results of poly(I:C) 10 µg/ml were not so clear, and did not present confident results in the analysis due to the low values in the Fold Change, for this reason we decided to perform a new experiment with three independent BM-MSCs samples from human healthy donors from the University Hospital of Salamanca. All these samples were confirmed with the ISCT normative in the Clean and Sterile room of the Hospital.

- **Quality control of poly(I:C) stimulated MSCs at short (6h) and longer (24h) times**

We performed a quality control of three the BM-MSCs samples: s389-19, s592-21 and s98-19 in three different conditions: Controls (CTL), stimulated same samples at short time of 6 hours (6h) and in stimulated same samples at longer time of 24 hours (24h) with poly(I:C) 25µg/ml.

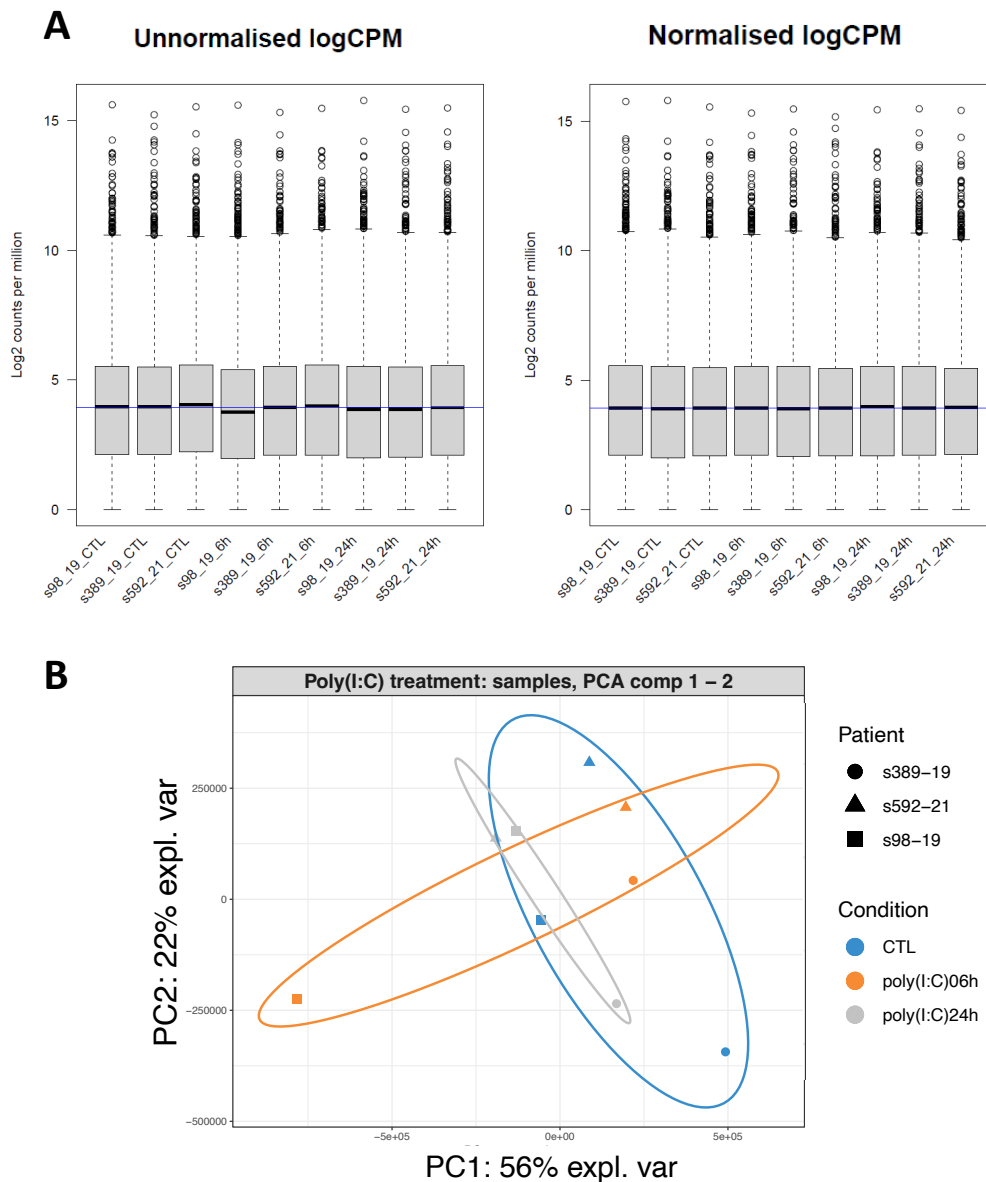


Figure 29. Quality control analysis of the samples: s389-19, s592-21 and s98-19 in different conditions: CTL, poly(I:C) at short time 6 hours and poly(I:C) at longer time 24 hours. **A)** Boxplots representing the unnormalized samples and normalized samples. **B)** PCA plot of the different samples in different conditions.

The raw expression matrix was consisted of 60604 genes and 9 samples. We filtered the raw matrix and obtained a result of 7838 genes and 9 samples. As we can see in the **Figure 29A**, the unnormalized samples were rather similar, which means that they were comparable in the analysis, in any case we performed a normalization with *cpm* method to standardize and avoid the small differences between them. As we can see, the logarithm of the counts was around 4, the maximum was around 11 and the minimum was 0 in all samples.

In the **Figure 29B**, the PCA plot was presented to visualize the distribution and relationships among the samples. As observed, there were no significant differences among the samples, which is not surprising since that they were derived from the same cell type (MSCs) and the same run. However, there was one exception: the sample s98-19 in poly(I:C) 6h condition appeared to be more distant from the other samples in the same condition. This outlier may indicate some unique gene expression patterns or biological variations specific to this particular sample, possibly due to experimental or biological factors. Samples from the controls and poly(I:C) at time 24 h conditions were the most similar among all cases.

- **Differential analysis of poly(I:C) stimulated MSCs at short (6h) and longer (24h) times**

Differential analysis was performed using *Limma-Voom* package as described in the Materials and Methods section. In **Table 13**, the summary of the contrast of poly(I:C) stimulated MSCs at short time of 6 hours against the control and the contrast of poly(I:C) stimulated MSCs at longer time of 24 hours against the control was shown. The summary statistics showed that in both contrasts the median was around 5 and the maximum was around 15, indicating that there were genes that showed significant up expression or down expression in response to the poly(I:C) stimulation.

Table 13. Summary table of numeric matrix of expression values on the Logarithm scale of differential analysis with *Limma-Voom* in the contrast of stimulated MSCs samples of poly(I:C) in 6 hours vs. MSCs CTLs and stimulated MSCs samples of poly(I:C) in 24 hours vs. MSCs CTLs.

MSCs samples poly(I:C) 6h vs CTL					
s98_19_6h	s98_19_CTL	s389_19_6h	s389_19_CTL	s592_21_6h	s592_21_CTL
Min. :-5.781	Min. :-5.799	Min. :-6.021	Min. :-5.977	Min. :-2.799	Min. :-5.561
1st Qu.: 4.329	1st Qu.: 4.560	1st Qu.: 4.549	1st Qu.: 4.590	1st Qu.: 4.634	1st Qu.: 4.718
Median : 5.418	Median : 5.568	Median : 5.566	Median : 5.538	Median : 5.611	Median : 5.619
Mean : 5.456	Mean : 5.537	Mean : 5.605	Mean : 5.529	Mean : 5.655	Mean : 5.623
3rd Qu.: 6.510	3rd Qu.: 6.590	3rd Qu.: 6.581	3rd Qu.: 6.584	3rd Qu.: 6.626	3rd Qu.: 6.626
Max. :15.670	Max. :15.696	Max. :15.397	Max. :15.310	Max. :15.555	Max. :15.618

MSCs samples poly(I:C) 24h vs CTL					
s98_19_24h	s98_19_CTL	s389_19_24h	s389_19_CTL	s592_21_24h	s592_21_CTL
Min. :-6.046	Min. :-5.799	Min. :-4.101	Min. :-5.977	Min. :-4.093	Min. :-5.561
1st Qu.: 4.504	1st Qu.: 4.560	1st Qu.: 4.492	1st Qu.: 4.590	1st Qu.: 4.552	1st Qu.: 4.718
Median :5.549	Median : 5.568	Median : 5.537	Median : 5.538	Median : 5.576	Median : 5.619
Mean : 5.558	Mean : 5.537	Mean : 5.551	Mean : 5.529	Mean : 5.603	Mean : 5.623
3rd Qu.: 6.589	3rd Qu.: 6.590	3rd Qu.: 6.576	3rd Qu.: 6.584	3rd Qu.: 6.595	3rd Qu.: 6.626
Max. :15.853	Max. :15.696	Max. :15.513	Max. :15.310	Max. :15.568	Max. :15.618

However, the minimum expression in both comparisons was more variable across the samples. For example, in the sample s389-19_6h, the stimulation 6h with poly(I:C) had a minimum of Min: -6.021, while in sample s592-21_6h the value of Min was: -2.799. However, the mean of the minimums was around 5 in both conditions. The summary statistics provide valuable information about the variability of the gene expression changes between poly(I:C) stimulated MSCs and the control MSCs, providing insight into the regulatory responses as a function of the different times of poly(I:C).

To determinate the best parameters for the differential analysis, we performed different comparisons as depicted the **Figure 30A**. We analyzed three different comparisons of: stimulated MSCs with poly(I:C) for 6 hours (poly(I:C) 6h) *versus* MSCs Control (CTL), stimulated MSCs with poly(I:C) for 24 hours (poly(I:C) 24h) *versus* MSCs Control (CTL), and stimulated MSCs with poly(I:C) for 24 hours *versus* stimulated MSCs with poly(I:C) for 6 hours. We tested different combinations of p-values and log(FC) of: 0.05 and 1.5, 0.01 and 1.5, 0.05 and 2; 0.01 and 2; and the last one 0.05 and 1; 0.01 and 1, respectively.

In the three comparisons the best results were obtained with p-value of 0.05 and log(FC) of 1.5. The number of genes determined as significant varied across the comparisons: 200, 219 and 130 genes in the order of the comparisons written in the previous paragraph. In **Figure 30B** we performed an analysis of the interaction of the comparisons from the **Figure 30A**, being the intersection of stimulated MSCs with poly(I:C) at 6 hours vs. MSCs Control and the stimulated MSCs with poly(I:C) at 24 versus MSCs Control, with name of "poly(I:C) 6h vs Poly(I:C) 24h", and the second intersection of the three comparisons in the **Figure 30A**. The best results were the intersection of poly(I:C) 6 hours vs. CTL and poly(I:C) 24 hours vs. CTL with 99 genes in p-value of 0.05 and log(FC)= 1.5. Additionally, the intersections between the three comparisons had a result of 40 genes in p-value of 0.05 and log(FC)= 1.5. The choice of p-value and log(FC) thresholds (0.05 and 1.5, respectively), was primarily aimed at ensuring a well-defined number of genes while maintaining homogeneity in the number of significant genes in terms of p-value and log(FC) across all comparisons.

Figure 30C displayed the number of up and down expressed genes in the differential expression analysis. In the intersection of two conditions, poly(I:C) 6 hours vs. CTL and poly(I:C) 24 hours vs CTL, all 99 genes were up expressed. Further analysis of these 99 genes in the comparison of poly(I:C) 24h vs poly(I:C) 6h revealed that 88 genes were up expressed and 11 were down expressed, indicating that there were genes that were down expressed in the condition of stimulated MSCs in 6 hours. However, in the intersection of three conditions of comparison, all 40 genes were up expressed in all comparisons. This observation was also represented in the **Figure 30D**, in the Venn diagram plots with p-value 0.05 and log(FC) of 1.5, where we could see that the number of differential genes increased in the condition of poly(I:C) 24h vs. control.

The choice of a p-value of 0.05 and a log fold change of 1.5 gave the best results for identifying differentially expressed genes in the comparisons, revealing consistent patterns of gene expression changes across the different conditions.

A

Limma Voom Analysis			
Comparison of MSCs	p-Value	log(FC)	N. of Genes
Poly(I:C) 6 h vs. CTL	0,05	1,5	200
	0,01	1,5	152
	0,05	2	132
	0,01	2	103
	0,05	1	357
	0,01	1	240
Poly(I:C) 24 h vs. CTL	0,05	1,5	219
	0,01	1,5	157
	0,05	2	151
	0,01	2	107
	0,05	1	425
	0,01	1	276
Poly(I:C) 24 h vs. Poly(I:C) 6 h	0,05	1,5	130
	0,01	1,5	91
	0,05	2	50
	0,01	2	38
	0,05	1	309
	0,01	1	176

B

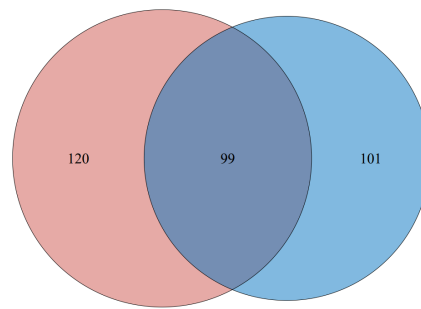
Limma Voom Analysis			
Intersection of MSCs	p-Value	log(FC)	N. of Genes
Poly(I:C) 6 h vs. Poly(I:C) 24 h	0,05	1,5	99
	0,01	1,5	68
	0,05	2	64
	0,01	2	43
	0,05	1	148
	0,01	1	97
Poly(I:C) 6 h vs. CTL	0,05	1,5	40
	0,01	1,5	25
Poly(I:C) 24 h vs. CTL	0,05	2	17
	0,01	2	10
Poly(I:C) 24 h vs. Poly(I:C) 6 h	0,05	1	68
	0,01	1	39

C

99 genes of intersection	UP	DOWN
poly(I:C) 6h vs CTL	99	0
poly(I:C) 24h vs CTL	99	0
poly(I:C) 24h vs poly(I:C) 6h	88	11
40 genes of intersection	UP	DOWN
poly(I:C) 6h vs CTL	40	0
poly(I:C) 24h vs CTL	40	0
poly(I:C) 24h vs poly(I:C) 6h	40	0

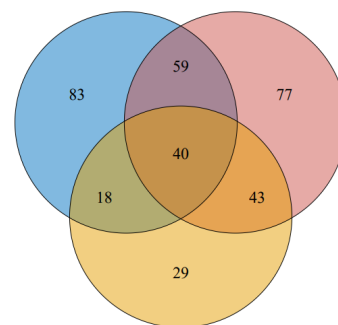
D

Poly(I:C) 24h vs. CTL Poly(I:C) 6 h vs. CTL



P-value= 0,05 logFC= 1,5

Poly(I:C) 6 h vs. CTL Poly(I:C) 24h vs. CTL



Poly(I:C) 6 h vs. Poly(I:C) 24 h

Figure 30. Limma-Voom differential analysis. **A)** Comparison between different conditions of MSCs: poly(I:C) at time 6 hours vs. CTL, poly(I:C) at time 24 hours vs. CTL and poly(I:C) at time 24 hours vs. poly(I:C) at time 6 hours; with p-values of 0.01 and 0.05 and log(FC) of 1, 1.5 and 2 for all comparisons. **B)** Intersection between different conditions of MSCs: poly(I:C) at time 6 hours vs. CTL, poly(I:C) at time 24 hours vs. CTL and poly(I:C) at time 24 hours vs. poly(I:C) at time 6 hours in the same p-values and log(FC) as **Figure 30A**. **C)** Up and down genes of the genes of intersection in each comparison. **D)** Venn Diagram plots of the intersection between: poly(I:C) at time 6 hours vs. CTL, poly(I:C) at time 24 hours vs. CTL and poly(I:C) at time 24 hours vs. poly(I:C) at time 6 hours with p-value 0.05 and log(FC)1.5.

Previously, in order to analyze each condition of each comparison in detail, we decided to design a volcano plot of the conditions of stimulation of poly(I:C) at 6 hours vs. CTL and stimulation of poly(I:C) at 24 hours vs. CTL, as it shown in the **Figure 31**. As we had seen in the **Figure 30**, most of the genes were up expressed when the MSCs were stimulated with poly (I:C), for this reason we increased the Fold Change of the up expressed genes to 4, which are the genes marked in red, to visualize the most significant ones.

In the volcano plot (**Figure 31**) of the contrast between MSCs stimulated with poly(I:C) during 6h vs. CTL MSCs, the up expressed genes were 42 with the most relevant genes in the major fold change: CXCL11, CXCL10, CXCL8, OASL, RSAD2, CCL8, IDO1, CX3CL1, IFIT2 and CCL2 the most significant gene with the largest p-value, these genes were related to the cytokine cascade in inflammatory processes as determined by an enrichment analysis. There were 15 down expressed genes, with the best FC and p-value in SNAI2, ID1 and SMAD6.

The contrast between MSCs stimulated with poly(I:C) during 24h vs. CTL MSCs causes a slight increase in the number of differentially expressed genes. The up expressed genes were 50 the genes with high fold change were more correlated with major p-value, except RSAD2, IFI27, CMOK2 and IFIT2. The most relevant genes were: MX1, IFIT1, IFIT3, OAS1, OAS2, OAS3, IFITM1, IFI6, ISG15, HERC6 and CXCL6, which were associated with regulation processes. In case of the down-regulated genes, there were 16, and the most relevant were: TNFRSF10D, GREM2, NT5DC2 and ZMAT3.

With the intention to describe the most variable genes in each condition, we evaluated in detail each of the contrasts with *Limma-Voom* package, which were printed in Heatmap plots and through the dendrogram of the heatmaps we established different pattern of genes classified in clusters as it is shown in **Figure 32**.

Figure 32A and **32B** represented the comparisons of stimulated MSCs with poly(I:C) at time 6h vs. CTL and stimulated MSCs with poly(I:C) at time 24h vs. CTL. The main difference that we can see in both Heatmaps was an increase of expression with more time of stimulation. The Heatmap of the **Figure 32A** included a total of 200 genes as indicated in previous **Figure 30** and the top 20 maximum of genes with the largest log(FC) were: CXCL10, RSAD2, CXCL11, CXCL8, OASL, IFIT2, GBP1P1, IDO1, MX1, CCL8, GBP5, CX3CL1, CCL3, ACTN2, BIRC3, CMPK2, TNFSF10, IFIT1 and MX2. The **Figure 32B** was the contrast of stimulated MSCs at time 24h vs. CTL with 219 genes and presented the top 20 maximum genes similar to the **Figure 32A**, being: RSAD2, MX1, BST2, CMPK2, IFIT2, GBP1P1, MX2, OAS1, CXCL8, IFIT1, IFIT3, IFIH1, HERC6, ISG15, IFI6, IFI27, CXCL6, OAS3, RTP4 and OAS2.

The maximum increment of the gene expression was represented in the pattern of genes of cluster 2 in both conditions, in the comparison of stimulated MSCs with poly(I:C) 6h vs. CTL the most relevant genes were: ICAM1, RIGI, CCL2, GBP1, TNFAIP3, HELZ2, IFI44L, and PARP14; with values of log(FC) of: 5.293, 4.999, 4.897, 3.531, 3.216, 4.373, 3.141 and 3.445, respectively and inflammation-related functions. Nevertheless, in the same cluster on the comparison of stimulated MSCs with poly(I:C) 24h vs. CTL, the genes were: IFI27, USP18, SLC15A3, PARP12, CLDN1, IRF7 and TLR3, which were associated with regulatory functions. With the log (FC) associated value of: 6.837, 5.685, 5.483, 5.393, 4.804, 4.664 and 4.436, respectively.

The **Figure 32A** presented the relevance in cluster 4. This presented a decrease in the set of genes in the stimulated MSCs at 6 hours compared to the control and stimulated MSCs at 24 hours. The cluster was formed by: SNAI2, CAMKK1, HOXA9, PPP1R3C, ID2, ID1, SMAD6, NR1D1, CHAC1, STARD9, CIART, TOX, KCTD16 and WEE1 with functional annotation in cell differentiation and circadian regulation of gene expression. On the other hand, other interesting cluster was presented in the cluster 5 of the **Figure 32B**, where the set of genes maintained their expression in Control and stimulated MSCs in 6h but decrease at the time of 24 hours, this cluster was formed by: ACO1, TNFRSF10D, ZMAT3, FAM172A, CERK, SESN3, PRKCA, GREM2, ANKH, NT5DC2, MMP16 and ASPM. DAVID bioinformatics analysis revealed the association of these genes with apoptosis and plasma membrane.

Finally, the **Figure 32C** showed the comparison of the 130 differential genes between stimulated MSCs with poly(I:C) for a longer time of 24 hours vs. stimulated MSCs with poly(I:C) at short time of 6 hours. As we can see in the Heatmap, the top 20 genes were: BST2, IFI27, MT1M, IFITM1, TRIM14, IFI6, CTSS, EGR3, ISG15, IFI35, SECTM1, TYMP, CXCL11, CFB, NR1D1, MX1, OAS3, SLC15A3, OAS1 and TLR3, the log(FC) of these genes presented a range of values from 2.5 to 5, was rather low than the comparisons of stimulated MSCs 24h vs CTL and stimulated MSCs 6h vs. CTL with ranges of 13 to 6.5. The most interesting clusters were presented in cluster 2 and 5. The 24 genes presented in cluster 2 with a decrease in stimulated MSCs at 24h were: STK38L, ACO1, FBN2, GPR68, PRKCA, ZMAT3, ANKH, FAM172A, NRXN3, CERK, TCF7, TNFRSF10D, NT5DC2 and GREM, related with plasma membrane functions; whereas in low peak of stimulation of MSCs at 6 hours presented the 24 genes: ADM, CD74, LINC00968, BDKRB1, OLFML1, BTN3A3, BTN3A1, SLC16A4, IFI30, NR1D1, EGR3, ID1, PDE4D, HES1, ID2, H4C14, NR1D2, TGFBR3, UNC93B1, BTN3A2, DDIT4, AVPI1, H1-10 and NT5C3A. The enrichment analysis of these genes presented an associated functions of: antigen processing and presentation of exogenous peptide antigens via MHC class II, immunity, regulation of lipid metabolism processed, cell differentiation and regulation of circadian rhythm.

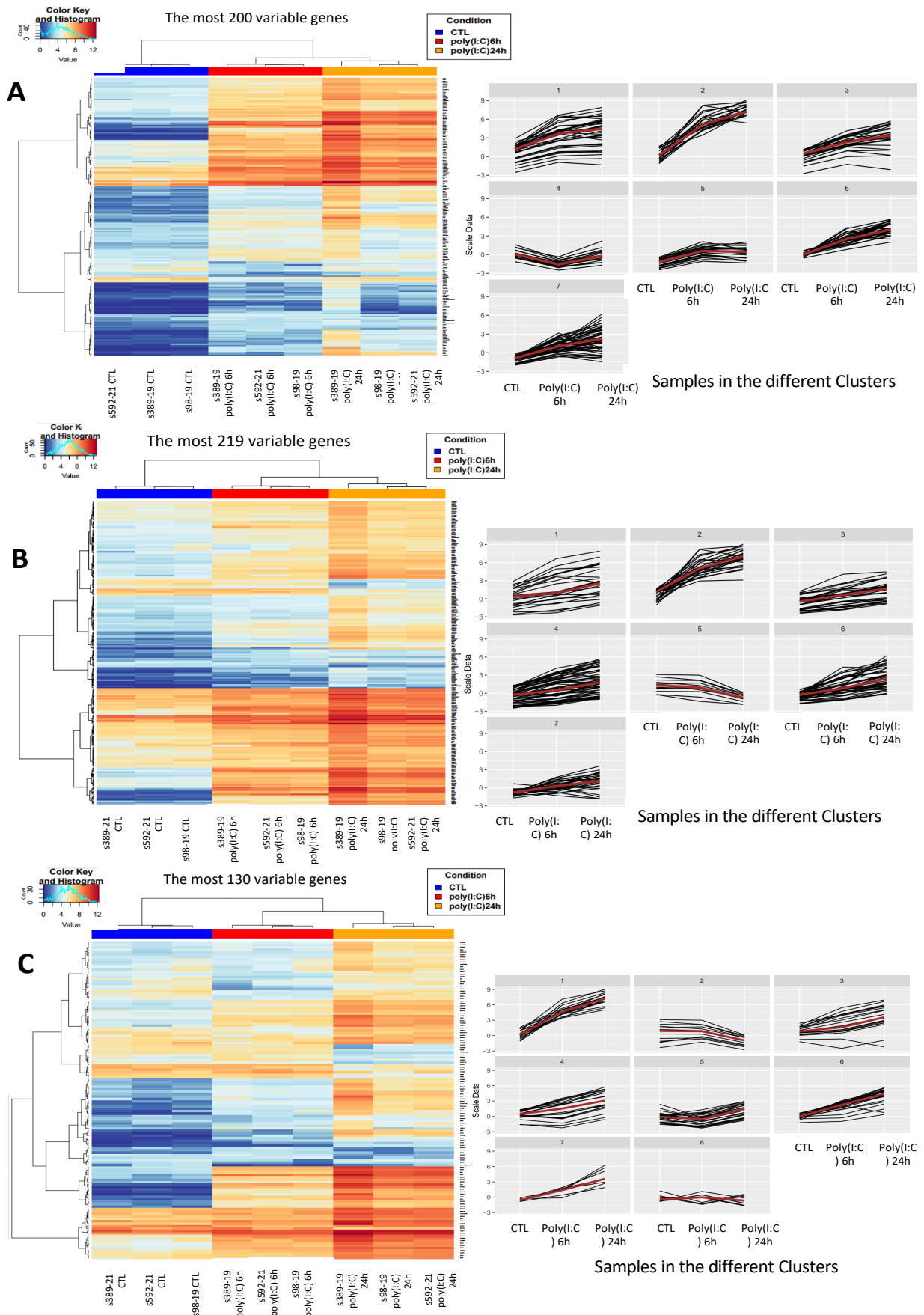


Figure 32. Heatmaps and expression pattern plots of the differential *Limma-Voom* analysis with p -value= 0.05 and $\log(\text{FC}) > 1.5$. **A)** Differential analysis of the contrast stimulated MSCs at time 6h vs CTL. **B)** Differential analysis of the contrast stimulated MSCs at time 24h vs CTL. **C)** Differential analysis of the contrast stimulated MSCs at time 24h vs stimulated MSCs at time 6h.

- **Functional Enrichment Analysis of poly(I:C) stimulated MSCs**

Both analyses summarize the relationship between poly(I:C) stimulation and the immune response against the virus. The analyses were divided based on the stimulation time of poly(I:C): one with a short stimulation time (6 hours) and the other with a longer stimulation time (24 hours). In the case of short stimulation time (6h) of poly(I:C) (**Figure 33A**), the main functions were associated with: the cytokine-mediated signaling pathway response, with the collection cytokines and interleukins such as: CXCL3, CXCL8, CXCL11, CXCL3CL1, CCL3, CCL2, CCL3, IL7R and IL15R. The response to virus with genes such as: TLR3, TRIM22, IFIT2M IFIT3M IFI6, NFKB1 or IRF2. The last subnetwork associated with the functions of the contrast stimulate MSCs in time (6h) of poly(I:C) vs. CTL was involved in processes of the regulation of viral function with genes as: CIITA, RSAD2, BST2, PARP10 or TNIP1. The network of the **Figure 33A** also presented genes that were related in the three main functions, such as: OAS1, OAS2, OAS3, OASL, MX1 or IFITM1. Other processes associated with this contrast (poly(I:C) stimulation for 6 hours vs. CTL) were: positive response to biotic stimulus or to external stimulus.

The **Figure 33B** shows the longer stimulation MSCs (24h) against CTL. In this case, the genes were associated with a regulatory response, such as: CCL2, ICAM1, PARP10, CD74, CXCL8, CLDN1, TNIP1, NECTIN2, TRIM21, IFI1, or MX1. Other relevant function was the regulation of response to biotic stimulus with genes: CXCL6, PARP14, OPTN, HLA-A, SERPING1, CFH, USP18, PARP14, ERAP1, PRKCA or HLA-B. Finally, there were functions such as in the short stimulation as a defense against virus with genes associated: IFIT2, IFIT3, MYD88, IFI44L, IFI6, IFI44, IL6, PARP4, TLR3, TREX1 or IRF7. There were few differences in functions associated with p-value compared to the short stimulation as: regulation of viral life cycle or response to interferon-gamma and beta. Taken together, these results suggest that the poly(I:C) stimulation in MSCs has interesting effects in the modulation of the immune and inflammatory systems. The differences observed between the 6 and 24 hours of stimulation may indicate time-dependent effects on MSC responses as a function of the poly(I:C) stimulation.

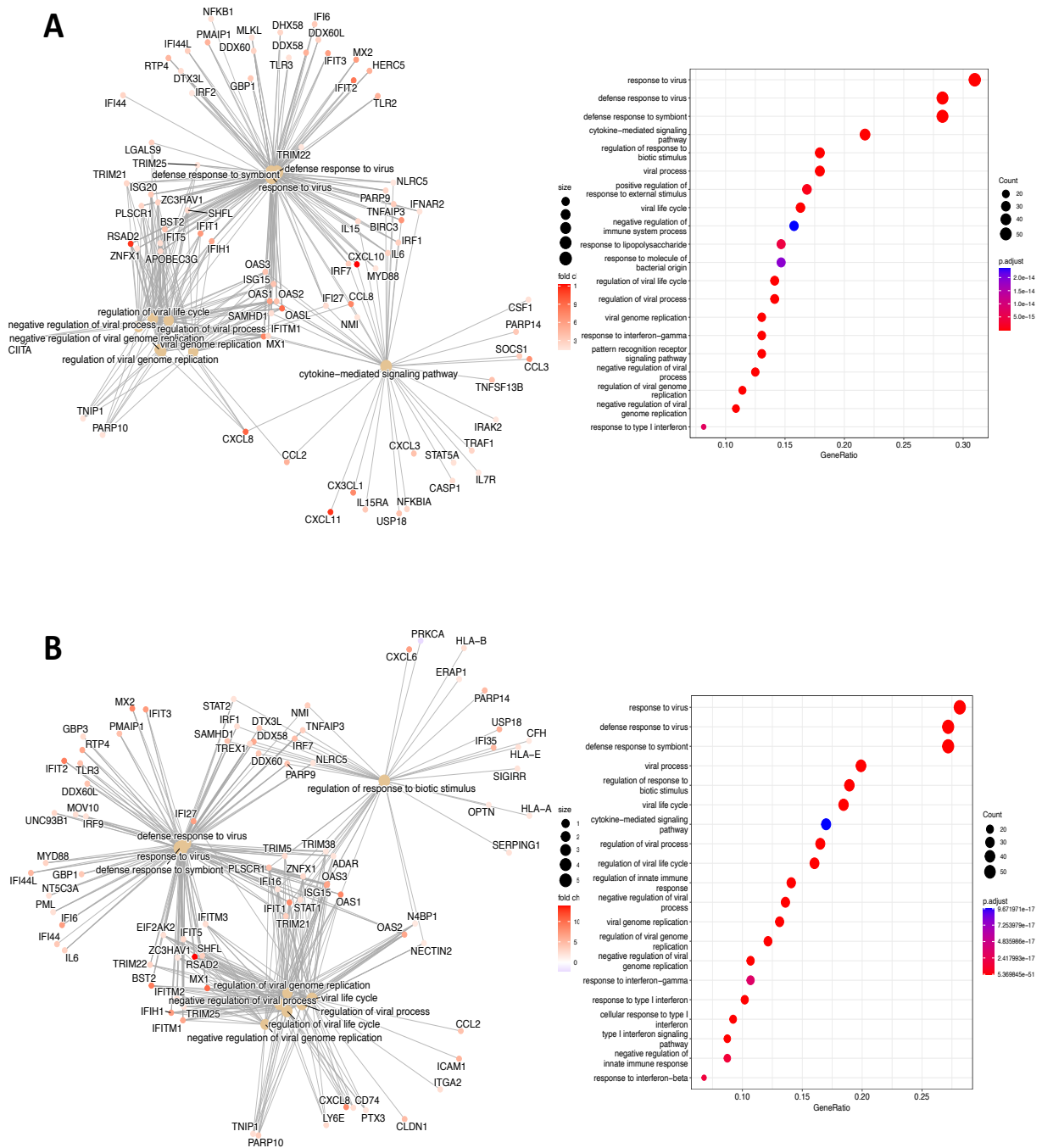


Figure 33. Network and dotplot of enrichment analysis. **A)** genes correlated with the functions associated with the contras of stimulated MSCs with poly(I:C) at time 6h vs. CTL. **B)** genes correlated with the functions associated with the contras of stimulated MSCs with poly(I:C) at time 24h vs. CTL.

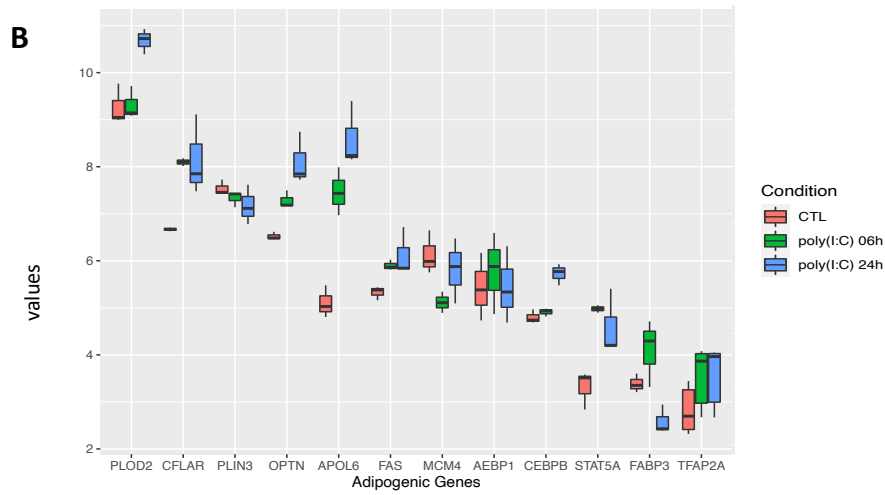
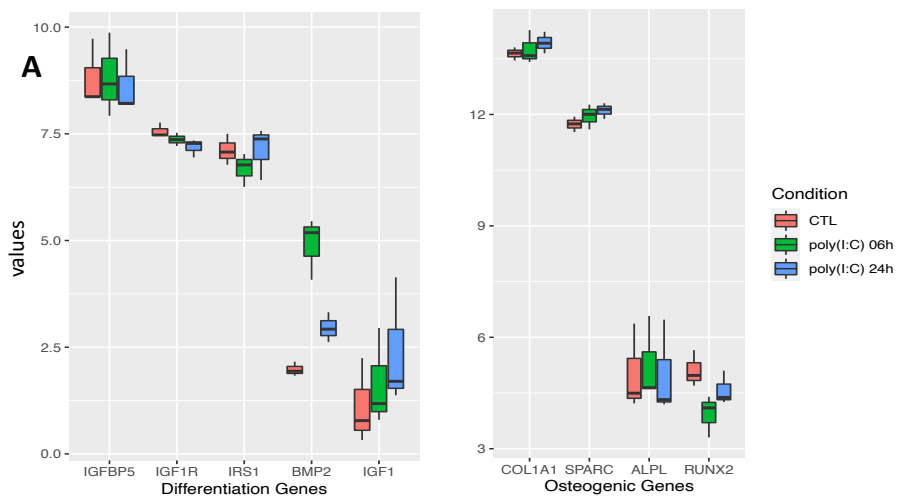
e. Differentiation in control and stimulated MSCs

In the final stage of the results, we performed a bioinformatic study of our samples of MSCs and stimulated MSCs in different points of poly(I:C) 6h and 24h, with the purpose of seeing the changes in the differential lineage.

The bioinformatic study of differentiation was performed using the differential markers collected in the literature and the analysis was performed as it is indicated in the Materials and Methods (Chen et al., 2014; Waterman et al., 2010). **Figure 34A** shows two different sets of boxplots of the general differentiation genes, included in all lineages, and osteogenic genes depending on the condition: control MSCs (CTL), stimulated MSCs poly(I:C) 6h or stimulated MSCs poly(I:C) 24h. It seems that BMP2 could have high differences, but it was not relevant in any differential analysis presented in *Limma-Voom* in previous section.

Figure 34B presents the boxplot of adipogenic differentiation genes, where the differences seem to be highest, overall, in PLOD2, CFLAR, OPTN, APOL6, FABP3 and CEBPB. However, in the **Figure 34C** shows a table with the significant differential genes, all of them which were adipogenic genes. In the contrast of the stimulated MSCs with poly(I:C) 6h vs. CTL, only three genes presented significance: CFLAR, APOL6 and STAT5A, with APOL6 being the most relevant gene with the highest $\log(\text{FC}) = 2.43$. Otherwise, in the case of the simulated MSCs with poly(I:C) 24h vs. CTL there were 4 differentially expression genes: APOL6, PLOD2, OPTM and CFLAR. As in the stimulation of MSCs in time of 6h, APOL6 was again the gene with the highest $\log(\text{FC}) = 3.65$.

In order to corroborate these results, which seems an increase in the adipogenic differentiation, we made an experimental analysis of differentiation to adipocytes and osteoblasts as it is indicated in Materials and Methods.



Limma-Voom Differential Analysis				
Stimulated MSCs Poly(I:C) 6h vs CTL				
Gene	Log(FC)	Average FC	p-Value	Adj. P-value
CFLAR	1.501	7.272	0	0.0007
APOL6	2.431	6.158	0.0002	0.0125
STAT5A	1.879	3.942	0.0003	0.0157
Stimulated MSCs Poly(I:C) 24h vs CTL				
Gene	Log(FC)	Average FC	p-Value	Adj. P-value
APOL6	3.653	6.756	0	0.014
PLOD2	1.541	9.902	0.001	0.047
OPTM	1.732	7.231	0.001	0.048
CFLAR	1.632	7.326	0.01	0.107

Figure 34. Analysis of differentiation genes in MSCs CTL, stimulated MSCs with poly(I:C) at 6h and stimulated MSCs with poly(I:C) at 24h. **A)** Boxplots of general gene markers of differentiation and osteogenic gene markers. **B)** Boxplots of adipogenic gene markers. **C)** Table of the significant genes of the **Figure 34A**.

Figure 35 represents the growth of the adipogenic differentiation after 21 days in MSC Control, stimulated MSCs with poly(I:C) a time 6h and stimulated MSCs with poly(I:C) a time 24h in one of the samples (s98-19), but the experiments were done for in four samples of MSC with the same results (not shown). In this visual comparison we could see several differences with an increase of adipocytes after the stimulation of poly(I:C). It seems that an increase of the stimulation time with poly(I:C) 25 μ g/ml caused an increase of the adipocyte differentiation in the sample. To confirm these results, we performed a statistical analysis with the counting of the adipocytes.

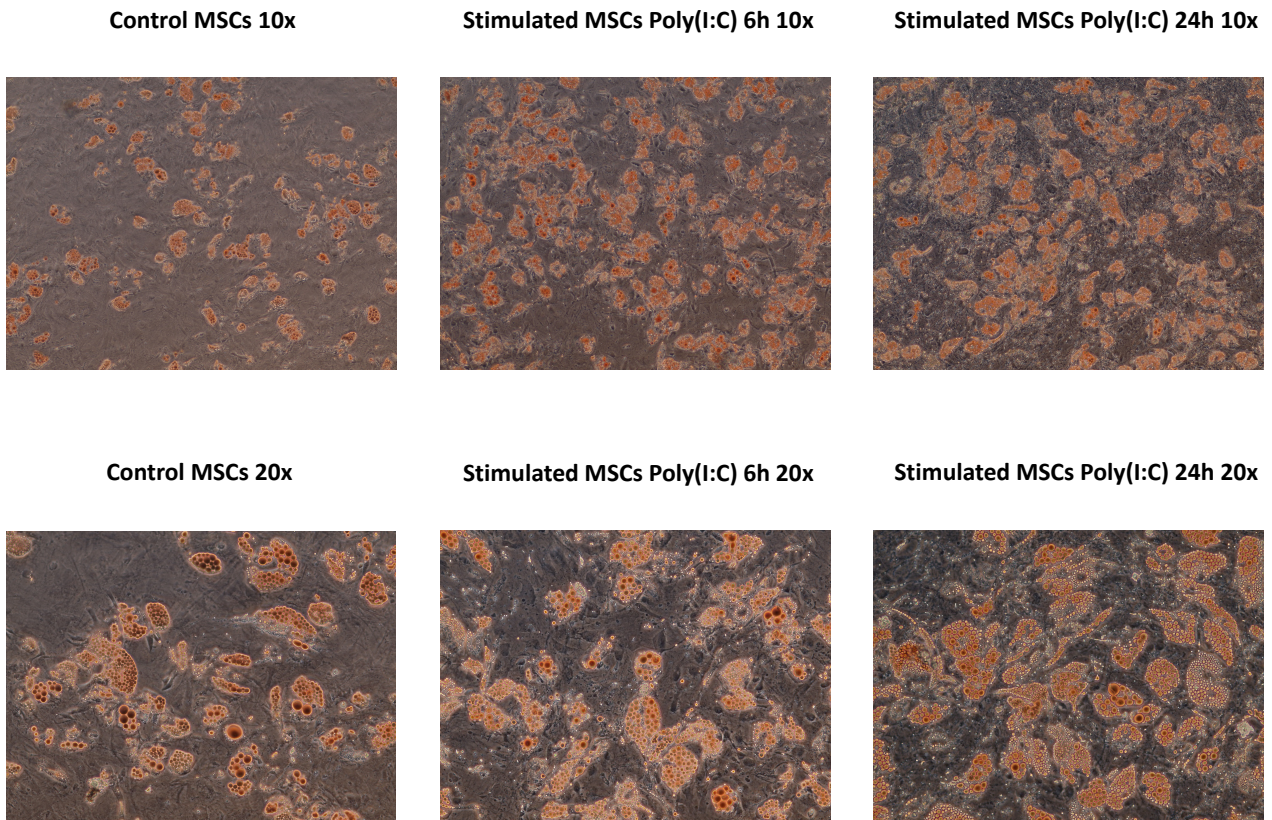


Figure 35. Control MSCs, stimulated MSCs with poly(I:C) in time 6h and stimulated MSCs with poly(I:C) in time 24h of sample s98-19 after 21 days of adipocyte differentiation. The pictures were taken in Optical Microscope at resolution 10x and 20x.

Adipocyte counting was performed in four samples at the time points of: 7, 14 and 21 days, as indicated in Materials and Methods. The statistical results presented relevant differences when the MSCs were stimulated. In **Figure 36A** showed that the growth of adipocytes increased with more time in culture. Moreover, the largest variance in growth appeared in the stimulated MSCs with poly(I:C) at 24 hours of stimulation, but at the same time, this condition was always above the conditions of stimulated MSCs with poly(I:C) at time of 6 hours and the control. To verify the boxplots, we performed an ANOVA and post-Hoc statistical analysis as shown in **Figure 36B**. The ANOVA analysis represented the differences between the three conditions, the effect of time and the correlation between the condition and time. Degrees of freedom (Df) is the total number of values minus 1. The Sum of squares (Sum sq) helps to express the total variation that can be attributed to different factors, in this case, condition, time or the relationship between condition over time. MSCs presented more variance along the time (with a value of 165713) than between the conditions and was reduced in the relationship between these two factors (with a value of 5000). ANOVA uses F-tests to statistically test the equality of means, the ratio between them, as the greater the result, the more significant is the factor, corresponding to the p-value of the F-statistic (Pr(>F)). In the results of these two variables, all factors: Condition and time: Time, presented significant values, that showed a significant increase of adipocytes in the stimulation of poly(I:C).

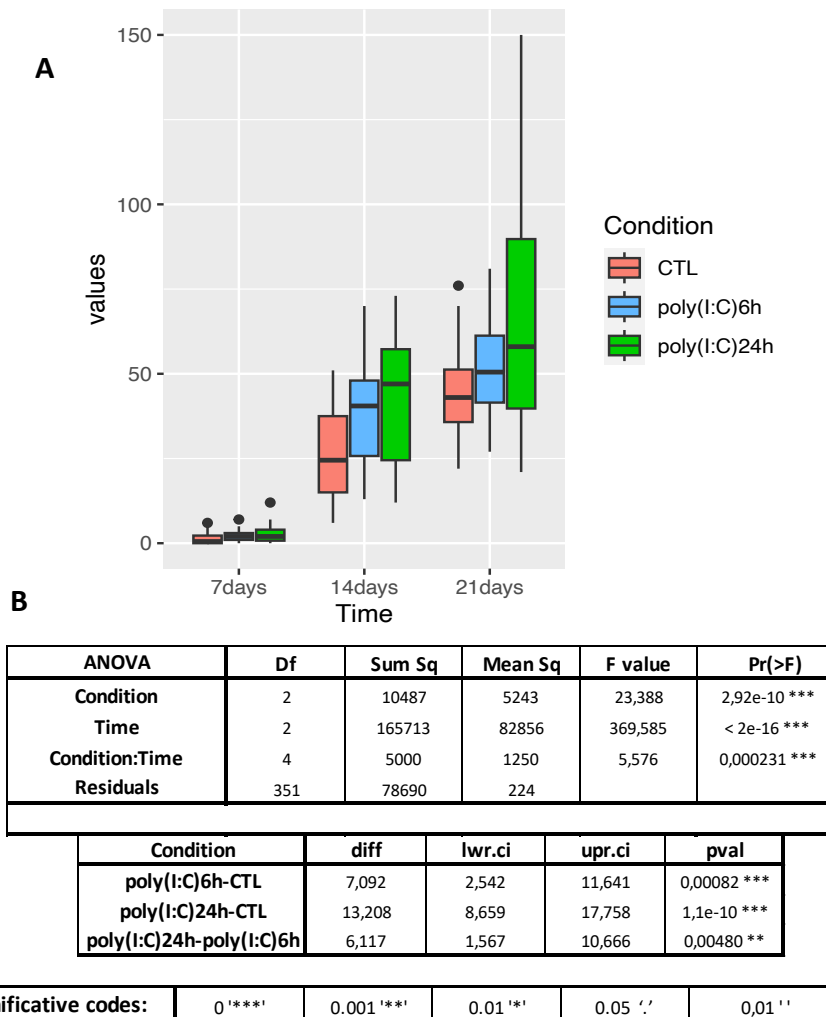


Figure 36. Statistical analysis of adipocyte differentiation of MSCs. **A)** Boxplot of the 4 samples in each condition in function of the 7, 14 or 21 days of adipocyte differentiation. **B)** ANOVA and post-Hoc statistical analysis of the samples.

To represent the statistical pair-wise means of all contrasts: stimulated MSCs at 6h vs. CTL, stimulated MSCs at 24h vs. CTL, and stimulated MSCs at 24h vs. stimulated MSCs at 6h, we performed a post-Hoc test. The output was the difference in means (diff), confidence levels (lower (lwr.ci) and upper (upr.ci)) and the adjusted p-values for all possible pairs (p-val). All the pair-wise comparisons were significant, the maximum significance being in the contrast of poly(I:C) stimulation of MSCs in the time 24h vs. CTL with p-value= 1.1e-10.

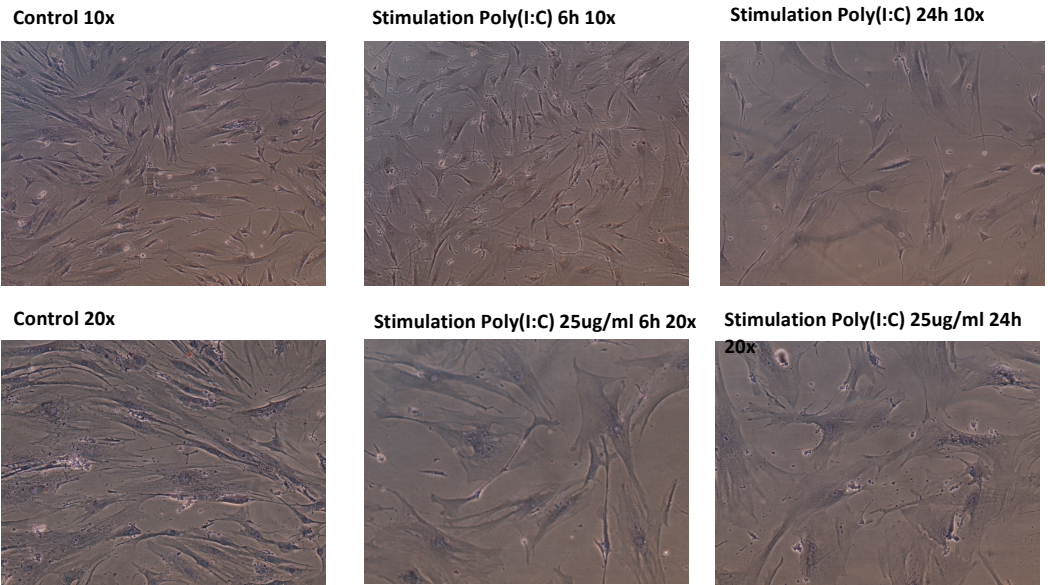
However, in contrast to adipocyte growth; in the **Figure 34A** we did not see significant changes in differential expression in osteogenic differentiation, as we can see in the experimental and statistical analysis performed.

Figure 37 shows the contrast between the MSCs and the MSCs treated with osteogenic medium in the different three conditions: Control of MSCs without stimulation: stimulated MSCs at 6h with poly(I:C) and stimulated MSCs at 24h with poly(I:C). The results observed by the Optical Microscope seem that there was a decrease in the growth of the stimulated MSCs with and without osteogenic differentiation medium. In the case of the stimulated MSCs at 6h with poly(I:C), the decrease seems to be more relevant in the osteoblasts than in the MSCs.

While the optical microscope appeared to have significant results, the absorbance of the mineralization analysis showed that the data did not have significant values. The **Figure 38A** showed the boxplot of 4 samples, each of them in the contrast of MSCs without differentiation media and Osteoblasts in cell differentiation in the three different conditions: CTL MSCs, stimulated MSCs at 6h with poly(I:C) and stimulated MSCs at 24h with poly(I:C), the plot did not present relevant differences between the conditions, but the absorbance between the cell type increase in the Osteoblasts. The **Figure 38B** shows the ANOVA and post-Hoc statistical analysis of 4 samples in the osteogenic differentiation. In the ANOVA analysis, only the cell type between MSCs and osteoblasts presented a significant F-value and p-Value, confirming a great osteogenic differentiation. However, the stimulation of MSCs with poly(I:C) in 6h and 24h did not affect to the MSCs with non-significant F-value of 0.28 and p-value of= 0.76. Neither was affected the relationship of Condition and Cell type, indicating that poly(I:C) did not affect in MSCs and Osteoblasts in the mineralization. The post-Hoc analysis revealed neither significant p-values in any of the contrasts: being stimulated MSCs at 6h vs. CTL with p-value 0.77, stimulated MSCs at 24h vs. CTL with p-value of 0.99 or in stimulated MSCs at 24h vs. stimulated MSCs at 6h with p-value of 0.82.

The results suggested that while seems to be a decrease in growth observed under optical microscope in the stimulation with poly(I:C) in MSCs, the absorbance of the mineralization did not show significant differences between the conditions. The Statistical analysis indicated that the stimulation of MSCs did not affect in the mineralization analysis in either in MSCs and osteoblasts. However, osteoblasts presented higher mineralization than the MSCs.

MSCs



Osteoblasts

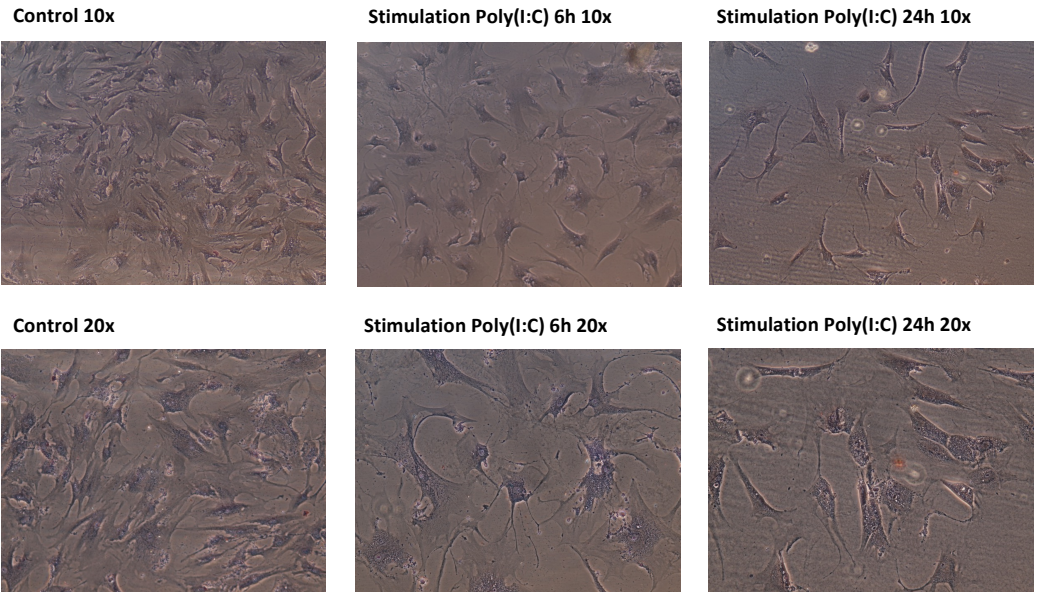
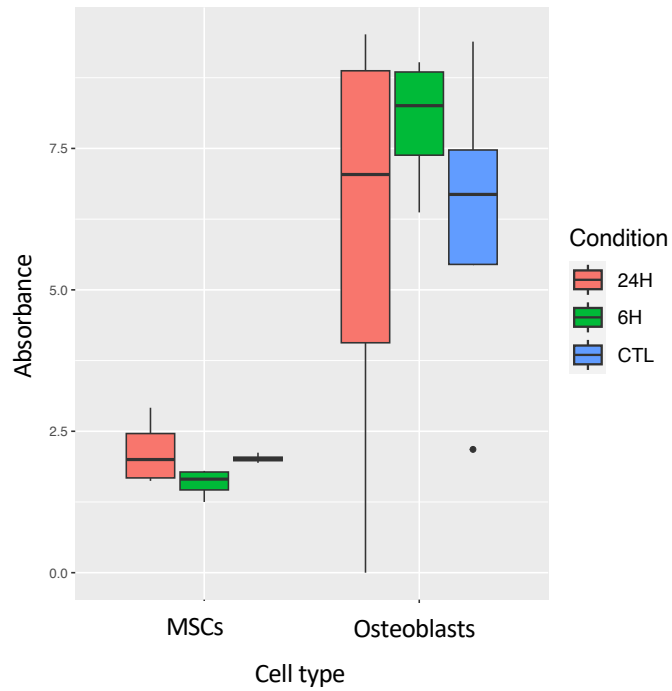


Figure 37. Control MSCs, stimulated MSCs with poly(I:C) in time 6h and stimulated MSCs with poly(I:C) in time 24h of sample 98-19 after 21 days with (osteoblasts) and without (MSCs) osteogenic differentiation medium. The pictures were taken in Optical Microscope at resolution 10x and 20x.



ANOVA	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Condition	2	2,74	1,37	0,280	0,759
Cell type	1	137,61	137,61	28,055	4,91e-5 ***
Condition:Cell type	2	7,86	3,93	0,801	0,464
Residuals	18	88,29	4,9		

Condition	diff	lwr.ci	upr.ci	pval
poly(I:C)6h-CTL	0,766	-2,060	3,592	0,771
poly(I:C)24h-CTL	0,110	-2,716	2,936	0,994
poly(I:C)24h-poly(I:C)6h	-0,656	-3,482	2,170	0,826

Significative codes:	0 '****'	0.001 '***'	0.01 '*'	0.05 '.'	0,01 ''	1
----------------------	----------	-------------	----------	----------	---------	---

Figure 38. Statistical analysis of osteogenic differentiation of MSCs. **A)** Boxplot of the 4 samples in MSCs without and with osteogenic differentiation medium. **B)** ANOVA and post-Hoc statistical analysis of the samples.

3. DISCUSSION

Mesenchymal Stem Cells have an important role in cell therapies due to their immunomodulatory properties, allowing them to regulate both adaptive and innate immunity. Toll-Like Receptor activation has been shown to modulate the functions and responses of MSCs (Najar et al., 2017). In this study, we aimed to understand and characterize the modulation of TLR3 and TLR4 in MSCs and characterize them.

In the first part of this chapter, we perform a transcriptomic analysis using samples obtained from public datasets. The data were derived from different stimulations of TLR4 and TLR3 in MSCs from healthy donors. Comparing LPS at two concentrations (1 μ g/ml and 10ng/ml) in stimulated MSCs versus control MSCs, we observed a significant up expression of most genes. In particular, genes such as CCL2, CXCL8, IL1B, and TNFAIP3, which are involved in the TNF signaling pathway and inflammatory response, showed significant differential expression.

For example, monocyte chemoattractant protein-1, MCP-1 (CCL2), is a potent cytokine involved in macrophage recruitment and induces accelerated wound healing in MSCs. CCL2 has demonstrated therapeutic potential in acute stroke (Lee et al., 2020; Whelan et al., 2020). C-X-C motif chemokine ligand 8 (CXCL8) is associated with autophagy and supports the survival and proliferation of acute myeloid leukemia cells through the PI3K/AKT pathway (Cheng et al., 2019). Both CCL2 and CXCL8 are associated with neutrophil chemotaxis (Han et al., 2022).

Interleukin-1beta (IL-1B) is a pro-inflammatory cytokine associated with tumorigenesis. Studies have shown that the combination of IL-1B and LPS can increase the production of pro-inflammatory cytokines such as IL6, IL8, and even IL-1B itself (Vézina Audette et al., 2013). Moreover, TNFAIP3 is also induced by tumor necrosis factor (TNF) and inhibits NF- κ B activation, leading to a pro-inflammatory response in human pericytes upon LPS induction (Guijarro-Muñoz et al., 2014).

The results of the published article on MSCs stimulated with LPS at 10ng/ml confirm our findings regarding the overexpression of these pro-inflammatory markers. However, our results, in contrast to the study by Hwa Kim et al., demonstrate an increase in these proinflammatory gene markers at a higher concentration of LPS (1 μ g/ml) (Hwa Kim et al., 2016). These results allow us to conclude that TLR4 priming with LPS in MSCs provides a well-defined understanding of the role of human bone marrow MSCs in inflammation and immunomodulation (Waterman et al., 2010).

The initial analysis of MSCs stimulated with poly(I:C) at 10 μ g/ml for 4 hours, using data obtained from public sources, yielded more diffuse results compared to LPS stimulation. The number of genes with significant log Fold Change in expression was reduced to 58 and low values of log(FC), even with the same stimulation parameters as in LPS stimulation. Notably, genes such as NFKB1B, which inhibits osteogenic differentiation of Mesenchymal Stem Cells by promoting β -catenin degradation (Chang et al., 2013), and IL6, CCL2, CXCL8, and TNFAIP3, which were also present in LPS stimulation, showed an increase in pro-inflammatory gene expression, but in this case with considerably lower log(FC) values. Additionally, Interferon Induced Protein With Tetratricopeptide Repeats 1/2: IFIT1 and IFIT2, presented significant fold changes (log(FC) of 2.142 and 4.096, respectively); and these genes are regulated by the NF- κ B family, which plays a key role in the induction of inflammatory cytokine genes and chemotaxis or cell migration (Kim et al., 2019). Furthermore, Protein Kinase D1 (PRDK1) was found to be overexpressed, and this gene is a target of the KRas-NF- κ B pathway in certain cancers (Döppler et al., 2016). Interestingly, another study revealed the relationship between the loss of PRDK1 expression in MSCs and a decrease in immunosuppression in CD4+ T cells (Wang et al., 2020). The results presented in the poly(I:C) stimulation source data define the transcription factors FOXO1, FOXO3, NF- κ B1, and the RELA proto-oncogene and

NF- κ B subunit as relevant genes, but these genes do not appear as significant genes in our differential analysis. They also note the up expression of CXCL8, IFIT2, IFIT3, IL1B and OASL indicating their involvement in response to viruses, inflammation, anti-apoptosis functions and immunosuppressive properties. However, the same article shows the results of genes such as NF- κ B that produce the induction of inflammatory cytokine genes (Kim et al., 2019). Nevertheless, due to the not relevant differential analysis in gene expression we decided to perform alternative experiments in order to clarify the characterization of stimulated MSCs with poly(I:C).

To address this, we increased the concentration of poly(I:C) to 25 μ g/ml, based on the relevant literature (Carpenter et al., 2011; Guinn et al., 2016; Weir et al., 2017), and those used by Balboa's group from Valladolid (Pindado et al., 2007). In addition, we found an article that discussed both short and longer time stimulation of poly(I:C) in MSCs, which may affect how these stimulated MSCs regulate Natural Killer cell immune and tissue-regenerative functions (Petri et al., 2017). Therefore, we decided to focus our experiments on differential, functional enrichment, and differentiation analysis of MSCs stimulated with poly(I:C) for short and longer times, 6 and 24 hours, respectively.

Our analysis revealed distinct gene expression patterns, which were further supported by volcano and Heatmap plots processed by the Differential Analysis of *Limma-Voom* package. In the 6-hour stimulated MSCs, we observed the up-regulation of cytokine-producing genes such as CXCL10, CXCL11, CXCL8, CCL8, CX3CL1, CCL3, and CCL2, indicating an inflammatory response. Other studies have also confirmed the inflammatory phenotype and the involvement of IL6, CXCL8, CXCL11, CCL2, and CXCL10 (Cassano et al., 2018; Souza-Moreira et al., 2022). In particular, C-X-C motif chemokine ligand 10 (CXCL10) has been associated with inflammation and plays a critical role in defense following coronavirus infection by recruiting and activating natural killer cells, resulting in reduced viral replication (Souza-Moreira et al., 2022). At the same time, genes related to apoptosis and inhibition of differentiation, such as ID1, ID2, and SMAD6, were down-regulated in the short time stimulation, suggesting an impact on immunosuppressive properties. Furthermore, ID1 is a transcriptional regulator that has been shown to be centrally involved in the induction of immunosuppressive properties in myeloid cells (Melief et al., 2020). In the functional enrichment analysis, in addition to the cytokine storm effect, we observed typical processes associated with poly(I:C) stimulation, including the regulation of viral functions, exemplified by members of the 2', 5'-oligoadenylate synthetase (OAS) gene family, CIITA, and IFITM1. All these data confirm that short stimulation of poly(I:C) for 6 hours in MSCs leads to a proinflammatory phenotype but also exhibit and regulatory properties, indicating that MSCs display duality in immunomodulation of MSCs in short time of stimulation.

In contrast, after 24 hours of stimulation, the MSCs exhibited functions related to the regulation of the immune system, with genes such as IFIT and TRIM family genes, OAS2, OAS3, BTN3A1, BTN3A2, and BTN3A3 being significantly expressed. IFIT genes encode a family of proteins that are induced after IFN treatment, viral infection, or PAMP recognition, as in our case with poly(I:C) stimulation. IFIT genes are induced in many virus-infected cells through IFN-dependent and -independent pathways, and they are also regulated by the NF- κ B family (Diamond & Farzan, 2013; Kim et al., 2019). Specifically, IFIT1 has been shown to promote EMT and modulate the proliferation, migration, and invasion of pancreatic cancer cells through the Wnt/ β -catenin pathway (Li et al., 2021). TRIM genes play an important role in the acquisition and maintenance of the stem cell phenotype and are also required for EMT (Jaworska et al., 2020). Moreover, TRIM21, which is significantly expressed in stimulated MSCs at 24 hours, inhibits the osteogenic differentiation of MSCs and bone formation (Xian et al., 2022). This gene also exhibits regulatory effects in the innate immune system by activating the NF- κ B, AP1, and IRF signaling pathways (Zhang et al., 2021). OAS family genes have antiviral functions and are involved in adaptive immune responses mediated by the RNase L-dependent pathway and are also affected by IFN stimulation and

participate in EMT (Choi et al., 2015; Doherty et al., 2017). Furthermore, Butyrophilins (BTN) are type 1 membrane proteins belonging to the immunoglobulin superfamily. Some BTN proteins have been associated with multiple autoimmune diseases by controlling T cell responses, and they also play a potential role in inhibiting anti-tumor immune responses. In particular, BTN3A1 and BTN3A2 are mainly expressed by CD4+ and CD8+ T cells, while BTN3A2 is the major form expressed in NK cells (Afrache et al., 2012). Additionally, the data revealed an increase in adipocyte differentiation, including PARP family genes, consistent with the duration of exposure to poly(I:C) in the MSCs (Cohen, 2020; Luo et al., 2017). All these genes corroborate the functional enrichment observed in the contrast between stimulated MSCs with poly(I:C) at 24 hours vs. the control, where the main functions are associated with the regulation of viral functions and IFN-dependent pathways, revealing a longer time of poly(I:C) stimulation in MSCs with regulatory phenotype.

In order to perform an in-depth analysis aimed at revealing differences in poly(I:C) stimulation over time (6 hours and 24 hours), we performed a differential expression analysis comparing stimulated MSCs at the longer time of 24 hours with stimulated MSCs at the short time of 6 hours. In the down-expressed genes in the 24-hour stimulation, such as ACO1, ZMAT3, ANKH, NRXN3, CERK, or TCF7, we observed their involvement in the increase of osteogenic differentiation, while TNFRSF10D is a gene involved in inflammatory responses (CUI et al., 2015; Granchi et al., 2010; Hsieh et al., 2010; Ullah et al., 2013; Wei et al., 2011; Zhang et al., 2019). This finding corroborates the inflammatory modulation of MSCs during the short time of 6-hour stimulation and their potential induction towards osteogenic differentiation. On the other hand, among the genes down expressed at 6 hours but up expressed at 24 hours of stimulation, we observed the BTN family genes, ID1, and ID2, which, as mentioned above, are down expressed at 6 hours of poly(I:C) stimulation. Also down expressed in time of 6 hours, were ADM, a potent apoptotic inhibitor, CD74 and HES1, potent regulators of hematopoietic stem cell maintenance, NRI1D1, which functions as a tumor suppressor, and DDIT4, which regulates adipogenic differentiation (Becker-Herman et al., 2021; Gharibi et al., 2016; S. Kim et al., 2023; Liu et al., 2015; Si et al., 2018).

This differential expression analysis revealed potential novel roles of poly(I:C) stimulation in MSCs. The data suggest that a short 6-hour stimulation induces an inflammatory stage in MSCs, similar to LPS stimulation, which is consistent with the findings of Petri et al., who found that short stimulation (4-8h) results in an early inflammatory phase (Petri et al., 2017) and contradicts the notion of only suppressive effects as indicated in the study by Waterman et al. (Waterman et al., 2010). Meanwhile, longer stimulation of poly(I:C) at 24 hours induces the known immunosuppressive and regulatory effects in MSCs.

Finally, we performed differentiation experiments in MSCs with poly(I:C). We examined the adipogenic and osteogenic differentiation of control MSCs, MSCs stimulated for a short time of 6 hours, and MSCs stimulated for a longer time of 24 hours. Adipogenic differentiation showed significant results with increased adipocyte growth with poly(I:C) stimulation, with more pronounced effects at 24 hours. This correlated with the differential gene expression, which revealed upregulation in PARP genes involved in adipogenic differentiation and APOL6 gene. Other studies have also supported our findings and showed a correlation between poly(I:C) stimulation of bone marrow MSCs and adipocyte growth (Chen et al., 2014; Waterman et al., 2010). However, we also found other studies suggesting that poly(I:C) stimulation does not affect adipogenic differentiation in adipogenic mesenchymal stem cells (DeLaRosa & Lombardo, 2010).

In contrast, osteogenic differentiation showed differences by the optical microscope, but our statistical analysis did not find significant differences. Differential expression analysis revealed genes associated with osteogenic differentiation in MSCs stimulated for 6 hours. Some studies have reported that osteogenic differentiation is only associated with LPS stimulation (Pevsner-Fischer et al., 2007;

[Waterman et al., 2010](#)), while others have suggested that poly(I:C) stimulation promotes both adipogenic and osteogenic differentiation ([Najar et al., 2017](#)). In our case, the experiments were conducted with only four samples, and further studies with increased sample size may help to determine significant results.

In conclusion, this chapter has provided promising results on the immunomodulation of MSCs. While our data confirm the well-established pro-inflammatory effect of LPS stimulation, the different time points of poly(I:C) stimulation in MSCs determine that short stimulation for 6 hours induces an inflammatory phase due to the cytokine cascade, whereas longer stimulation for 24 hours regulates the adaptive immune system and shows immunosuppressive properties in our human bone marrow MSCs. In terms of differentiation, our bone marrow MSCs clearly show increased adipogenic differentiation with longer stimulation time, while no significant differences were observed in osteogenic differentiation. This study is essential to identify new marker genes that, as we have shown, affect the modulation of MSCs in the microenvironment and can be utilized in new cell therapies. Further research is warranted to elucidate the intricate molecular mechanisms underlying these responses and to explore their implications for potential clinical applications.

4. FINAL SUMMARY of CHAPTER 3: *Immunomodulatory effect of MSCs*

In this chapter, we have presented the study of the immunomodulatory properties of the MSCs considering a dual effect, proinflammatory or immunosuppressive, as reported in different contexts and different experimental conditions. A thorough understanding of these immunomodulatory properties of MSCs is crucial for their use and application in cell therapy and in regenerative medicine.

In this context, our study focuses on investigating the activation of Toll-like receptors, specifically TLR3 and TLR4, to better understand their impact on MSCs. An initial transcriptomic analysis is performed on a collection of available datasets, comparing MSCs stimulated with different concentrations of the TLR4 agonist LPS (bacterial, potent innate immunostimulant targeting the toll-like receptor 4), and the TLR3 agonist poly(I:C) (a synthetic RNA analog of double-stranded that targets the toll-like receptor 3).

The results of TLR4 stimulation reveal the upregulation of genes intricately involved in the TNF signaling cascade and a broad inflammatory response. The upregulated genes include: CCL2, CXCL8, IL1B, and TNFAIP3; each of which plays an essential role in immune regulation, wound healing, and inflammatory processes. Stimulation with poly(I:C), initially performed at a concentration of 10 µg/ml, gives more diffuse results. However, when the stimulation concentration is increased to 25 µg/ml, clear patterns emerge. For a short period of 6 hours, stimulated MSCs exhibit a proinflammatory phenotype characterized by upregulation of cytokine-producing genes such as CXCL8, CXCL10, CXCL11, CX3CL1, CCL2, CCL3 and CCL8. In contrast, after longer stimulation (24 hours), MSCs shifted towards immune system regulation. Genes from families such as IFIT, TRIM, and OAS (such as OAS2, OAS3, BTN3A1, BTN3A2, and BTN3A3) are significantly expressed during this phase, highlighting the immunoregulatory properties of MSCs. Additionally, the 24-hour stimulation period promotes adipocyte differentiation, as exemplified by the upregulation of genes within the PARP family. In contrast, osteogenic differentiation, although showing differences in optical microscopy, does not show statistically significant changes. In conclusion, our differential expression analysis reveals that stimulation with poly(I:C) for a short time (6 h) induces an inflammatory phase in MSCs, while longer stimulation (24 h) exhibits immunosuppressive properties.

The results of our study challenge the conventional notion that poly(I:C) exerts only suppressive effects on MSCs. It highlights the complex, intricate and context-dependent immunomodulatory capabilities of MSCs that should be considered when using or applying them in cell-based therapeutic strategies. Indeed, the identification of novel marker genes that influence the modulation of MSCs within the microenvironment holds promise for advancing cellular therapies using these cells. However, further investigations are needed to unravel the complex molecular mechanisms underlying these dynamic responses and to delineate their clinical implications.

CHAPTER 4:

Analysis of BM-MSCs isolated in vitro and in vivo using single-cell transcriptomics

Short Title:

Single-Cell transcriptomics of MSCs

1. MATERIAL AND METHODS

a. Stimulation of MSCs at 6h with poly(I:C)

We prepared 2 samples of MSCs from healthy human donors, which were harvested and isolated *in vitro* from the PBMNCs. Each sample had two flasks: the control and its stimulated MSCs. All cells were cultured until passage four with DMEM supplemented with 10% FBS at 37°C, 90% of humidity and 5% of CO₂. Then, the MSCs were stimulated accordingly to the poly(I:C) 25µg/ml stimulation protocol cited in the Materials and Methods section of Chapter 3, for the conditions of stimulation poly(I:C) 6 hours.

After this stimulation time, the MSCs were washed twice with PBS and trypsinized with Trypsin (Gibco, Thermo Fisher Scientific). They were then counted them with trypan blue (Gibco, Thermo Fisher Scientific) at a dilution (1:100) using a Neubauer chamber. At the same time, we calculated the viability of each sample, with the percentage of viability = (number of viable cells/number of total cells) x 100. In addition, to ensure the best precision, the samples were also counted and measured their viability through Cellometer K2 Fluorescent Viability Cell Counter (Nexcelom).

b. Single-cell RNA sequencing Technique

The 10X Genomics Chromium workflow was based on microfluidic principles or the GemCode technology (10xGenomics). The first step of the experiment was cell labeling, which was performed using oligonucleotides known as Cell Multiplexing Oligos (CMOs), following the guidelines outlined in Protocol 1 Overview of 10x Genomics: Cell Multiplexing Oligo Labeling for Samples with >80% Viable Cells & for Nuclei Isolated from Fresh Cells/Tissues, document CG000391, Revision B. Once the cells were labeled with the appropriate oligos, we proceeded to the pool of the cells. Pool 1 contained the control MSCs, while Pool 2 contained the stimulated MSCs with poly(I:C). Pool 3 contained the control CMO and the last Pool 4 for the stimulated MSCs CMO.

The droplet-based encapsulation was achieved by gel beads, and then we proceeded to construct single-cell RNA-seq libraries under dual indexing. This process followed the Chromium Next Gel bead Emulsion (GEM) Single Cell 3' Reagent Kits v3.1 (Dual Index) protocol, document CG000388 Revision C. It should be noted that for the GEM Generation & Barcoding step, a targeted cell recovery of 25000 cells/µl each pool was used. DNA quantification was performed using Agilent 2100 Bioanalyzer System in High Sensitivity Chip (Agilent). The barcoded sequencing libraries were quantified by quantitative PCR (KAPA) and these libraries were subsequently loaded onto an Illumina NextSeq 1000/2000 sequencing platform.

c. scRNA-seq data preprocessing including demultiplexing

Sample demultiplexing, barcode processing and single-cell 3' gene counting were performed using 10X's Cell Ranger 7.1.0. Cell Ranger is a set of analysis pipelines that process Chromium single cell data to align reads, generate feature-barcode matrices, perform clustering and other secondary analyses. *fastqcs* were quantified using the *Cell Ranger* count package and the GRCh38 human genome reference (version refdata-gex-GRCh38-2020-A, 10X Genomics). The *Cellranger* count package performed the alignment, filtering, barcode counting, and UMI counting using the Unix platform. The output included the BAM files and the raw and filtered features, barcodes and count matrix of the samples (Zheng et al., 2017).

Freemuxlet is a demultiplexing algorithm designed to partition individual samples without the Single Nucleotide Polymorphism (SNP) genotype of each donor for the multiplexed capture. The algorithm takes a list of variant sites with known population allele frequencies. It examines single-cell RNA sequencing (scRNA-seq) reads that overlap these variant sites to cluster each uniquely barcoded droplet according to its respective sample origin (Figure 39). The application of *Freemuxlet* was performed by independently subjecting both control and stimulated MSCs samples to the *Demuxafy* Software, facilitated within the APTAINER container system for High Performance Computing (HPC) environments. The entire process was performed according to the protocol described in the Demuxafy 2.0.1 documentation (<https://demultiplexing-doublet-detecting-docs.readthedocs.io/en/latest/Freemuxlet.html>).

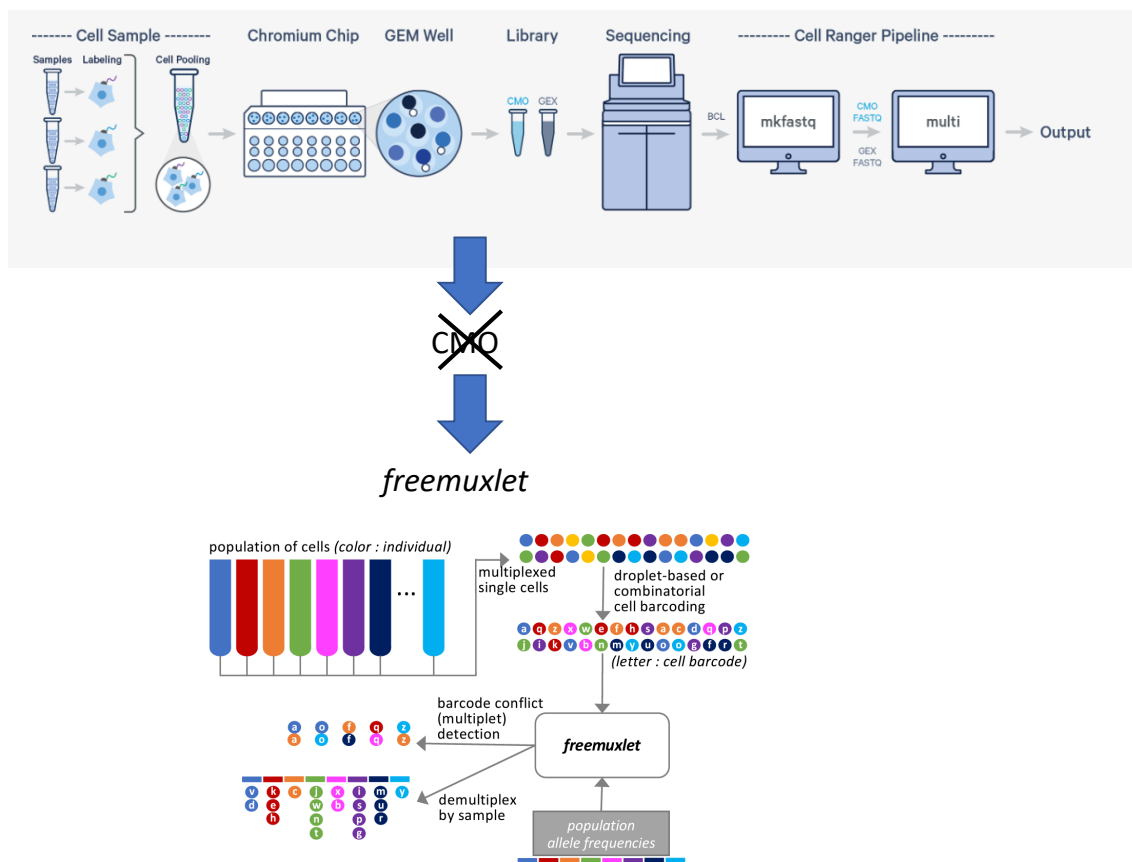


Figure 39. Workflow of single cell analysis downstream.

During this analysis, reference was made to the SNP genotype data obtained from the human genome assembly GRCh38, in our case: vcfGRCh38_1000G_MAF0.01_GeneFiltered_ChrEncoding.vcf. After applying the necessary sorting and filtering procedures as described in the Demuxafy Manual, the BAM files were processed. The Freemuxlet algorithm was executed using the *popsle* package.

d. Analysis of scRNA-seq data using algorithm *Seurat*

The analysis and exploration of single-cell RNA-seq data from our experimental samples was done in the R programming language using RStudio. The package to process all that was with *Seurat* version 4.3.0 (Hao et al., 2021).

Before the differential and exploratory analyses, the first step was to integrate the data of the samples with *Harmony* algorithm (Korsunsky et al., 2019). Subsequently, the data were processed using *Seurat* version 2.3.0 with *cells*. The criteria for filtering the data were to retain cells with less than 15% mitochondrial reads and at least 2500 number of genes.

Subsequently, all cells identified as doublets in the metadata were removed from the dataset. We then proceeded with log-normalization, which was accomplished through sequential steps including the creation of a *Seurat* object (*CreateSeuratObject*), data normalization (*NormalizeData*), identification of variable features (*FindVariableFeatures*), data scaling (*ScaleData*), principal component analysis (*RunPCA*), *Harmony* execution (*RunHarmony*), identification of neighborhood relationships (*FindNeighbors*), in our case across dimensions 1 to 30, and the cluster identification (*FindClusters*).

e. Functional Enrichment Analysis of the markers found for different cell populations

The enrichment analysis was performed in the same way as the Materials and Methods mentioned in Chapter 3. These methods include tools as DAVID Bioinformatics platform and *ClusterProfiler* package version 4 of Bioconductor able to use in R programming language.

2. RESULTS

2.1. Single Cell analysis of stimulated poly(I:C) and control MSCs

a. Quality control of Mesenchymal Stem Cells in scRNA-seq

To reveal the heterogeneity of the poly(I:C) 25µg/ml stimulation and the control MSCs, we performed a single-cell RNA sequencing experiment. We examined the dataset of control and stimulated MSCs individually, evaluating the number of clusters and the gene expression profiles for each condition. The second part of this chapter involved the integration of both datasets using bioinformatic techniques aimed at mitigating the batch effect. Our focus was on evaluating of the inference of the different clusters in the stimulated MSCs that differ and had in common from the control MSCs.

It is pertinent to highlight that the data was derived from a demultiplexing algorithm based on specific variant sites. This algorithm facilitated the extraction of the two healthy donor samples from the control run and the two stimulated MSCs samples from the second run. This approach was necessitated by the lack of expression data in the Cell Multiplexing Oligo (CMO) of the samples generated during the sequencing. The *Freemuxlet* algorithm allows us to associate the droplets with each patient and which are classified as doublets (ambiguous). **Table 14** shows the assignment of cells to each patient in the control and the stimulated MSCs. The filtered matrix contained the cells that met the standard quality controls established within the *Seurat* package. The criteria included unique feature counts over 2,500 or less than 200 and which have >5% mitochondrial counts, as we commented in Materials and Methods. This was done despite the loss of the amount quantity of cells to avoid low-quality cells or empty droplets, which often have very few genes, low-quality or dying cells, and exhibit extensive mitochondrial contamination.

Table 14. Distribution of samples (cells) in function of the two patients in raw matrix, and filtered matrix after quality control.

DATA	FEATURES (genes)	SAMPLES (cells)		
		Patient 1	Patient 2	Doublet
Matrix				
Control MSCs				
Non Filtered matrix	27477	4528	7965	4042
Filtered matrix	27477	162	1204	1076
Stimulated MSCs				
Non Filtered matrix	28017	7246	8150	4559
Filtered matrix	28017	485	556	618

After filtering both matrices, we performed an independent transcriptomic analysis for each condition and an integrated analysis of both conditions. The goal of these analyses was to characterize the subpopulations inherent to each condition and to discern the differences between healthy donors.

This comprehensive analytical effort provided us with a deep understanding of the effect of TLR3-priming in MSCs, the characterization of the stimulated and unstimulated MSCs, and the difference between the healthy donors in each condition. It is important to note that the central objective of this chapter was to identify and delineate subpopulations that differentiate and have in common the control and poly(I:C) Mesenchymal Stem Cells.

b. Global Analysis of stimulated MSCs versus Control MSCs in scRNA-seq

Within this analysis, we performed an integration of the stimulated poly(I:C) MSCs and control MSCs from the bone marrow of two healthy donors, previously filtered as indicated in the Materials and Methods section. As depicted in **Figure 40**, the integration of the samples was orchestrated with respect to their conditions and the resulting clusters. In the **Figure 40A**, the red dots correspond to the samples of the control and blue dots correspond to the stimulated MSCs. The samples were integrated together, although the poly(I:C) stimulated MSCs generated two clusters corresponding only to the poly(I:C) conditions, corresponding with cluster 1 and 3 (**Figure 40B**). It is noteworthy that **Figure 40B** demonstrated cluster 0 as the predominant cluster within the control MSCs, while the other clusters exhibited a mixture of characteristics from both conditions. The entire matrix of cells allowed a total of six clusters with a total number of 28852 genes and 2407 cells. The main clusters 0 and 1 contained 1198 and 403 cells, respectively. Cluster 2 consisted of 388 genes, while clusters 3, 4, and 5 contained 250, 97, and 71 cells, respectively. The filtered matrix contained a total of 27477 genes and 1366 cells.

Once the clusters were established, we performed a differential analysis using the *Seurat* package to obtain the most differentially expressed genes of each cluster with an adjusted p-value of <0.05. The significant gene clusters were very different among them, with 1198 genes for cluster 0, 403 genes for cluster 1 and 388, 250, 97 and 71 genes for clusters 2, 3, 4, and 5, respectively.

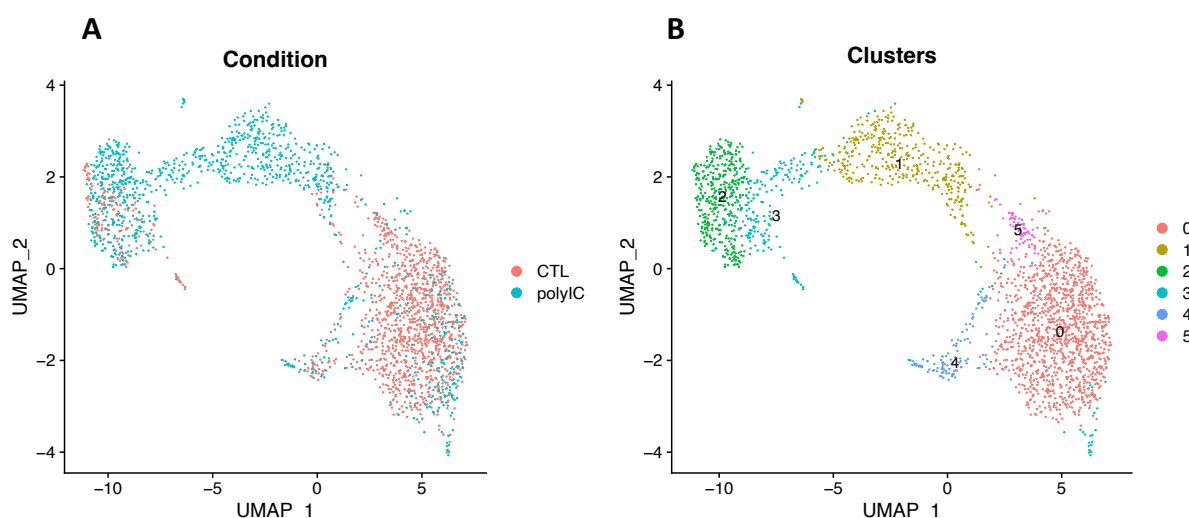


Figure 40. UMAP clustering plots. **A)** UMAP plot corresponding with the two conditions: Control MSCs and stimulated poly(I:C) MSCs. **B)** UMAP plot of clusters formed using Seurat package.

We generated a Heatmap of the top 50 gene markers for each cluster (ordered by adjusted p-value) in order to reveal the association of expression genes between clusters. **Figure 41** provides insight into the dynamics of the clusters. In the **Figure 41**, the largest cluster 0 showed relatively low expression, which was similar to cluster 5. Clusters 1 and 3 demonstrated similarity between them, with cluster 2 also showing a cohesive grouping of genes. These associations may be related to the effects of poly(I:C) stimulation. Interestingly, cluster 4 appeared to amalgamate elements from the two most distinct groups observed in the analysis. To reveal the information within each cluster, we performed an enrichment analysis in each cluster, indicating the most relevant genes in each cluster.

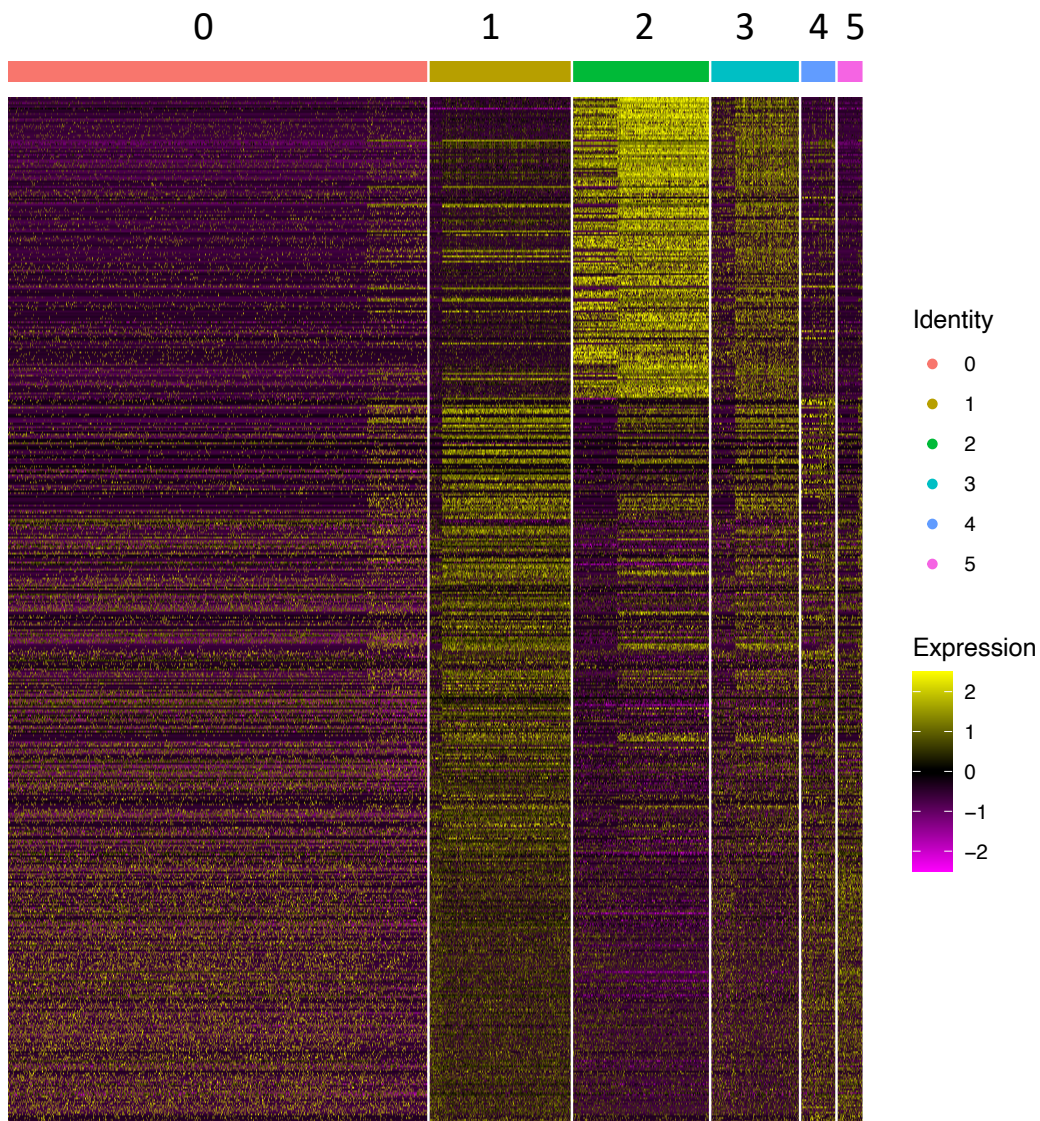


Figure 41. Heatmap of the top 50 differenced marker genes in each of 6 clusters of both MSC conditions.

The clusters were analyzed for functional enrichment using DAVID Bioinformatics and displayed in barplots (**Figure 42A**). Cluster 0 was associated with functions encompassing mitochondria, cytoplasmic translation, muscle protein and stress fiber; the genes contributing to these functions included two families, the myogenic genes MYL12A, MYL12B and MYL6 and the thrombopoietic genes TPM4, TPM1 and TPM2. Other genes involved in these processes were: ACTA2, CYBA, TAGLN, STMP1, MPC2, NDUFAB1, CHCHD1 and COX20.

Clusters 1 and 3 showed functions related to viral defense, innate immune response and negative regulation of viral function. Cluster 1 was characterized by genes from the IFIT and IFITM family genes: IFITM, IFITM2, IFI6, IFIT1, IFIT3, IFIT2, IFIH1, IFI16. OAS family genes: OAS1, OAS2 and OAS3. And other genes like MX2, MX1, IL6, IFI27, IRF1, IRF7, HLA-C, HLA-A, PARP14 and HLA-E. Moreover, cluster 1, presented inflammatory response with associated genes such as chemokine family: CXCL9, CXCL8, CCL20,

CXCL1, CXCL2, CXCL10, CXCL11, IL6, CCL8, CCL7, CCL5, CCL3 and CCL2. While cluster 3 represented the TNF signaling pathway: IL6, VCAM1, IRF1, VEGFC, PIK3R3, TNFAIP3, ICAM1 and positive regulation of interferon-beta production where genes IFIH1, OAS2, IRF1, OAS3 were involved.

Cluster 2 was mainly related to fundamental MSC functions such as respiratory chain, aerobic respiration, and electron transport, with genes such as MT-ND4, MT-CO1, MT-CO3, MT-ND3, and MT-ND1.

Cluster 4 was closely related to cell division, mitosis and cell cycle with genes such as CENPW, CKAP2, MKI67, SEPTIN7, CKS1B.

Cluster 5 was similar to cluster 0, with genes associated with muscle protein, muscle contraction, stress fiber or myosin genes such as: MYL12A, MYL12B, MYL6 and MYL9. Tropomyosin genes: TPM4, TPM3, TPM2, TPM1, and other genes such as, CRYAB, TAGLN and SEPTIN7.

In addition, **Figure 42B** presented a network of gene expression associated with different functions across all the clusters. Notably, the main functions were: aerobic respiration and oxidative phosphorylation with associated genes such as: COX family genes, MDH1, MDH2, ATP family genes, CYCS, NDUFA4. Nucleoside triphosphate biosynthetic process with genes such as: TYMS5, NME1, NME2. And the last cluster was formed in relation to response to virus, viral genome replication and regulation of viral cycle and genome replication, this cluster was formed with: CXCL10, CXCL8, CCL2, CCL8, IFIT family genes, OAS1, OAS2, OAS3, TRIM 22, TRIM25, NFKB1, MX1 or MX2.

Moreover, individual analyses were performed for each condition to identify clusters specific to these groups, further refining the subpopulation classification within the global analysis.

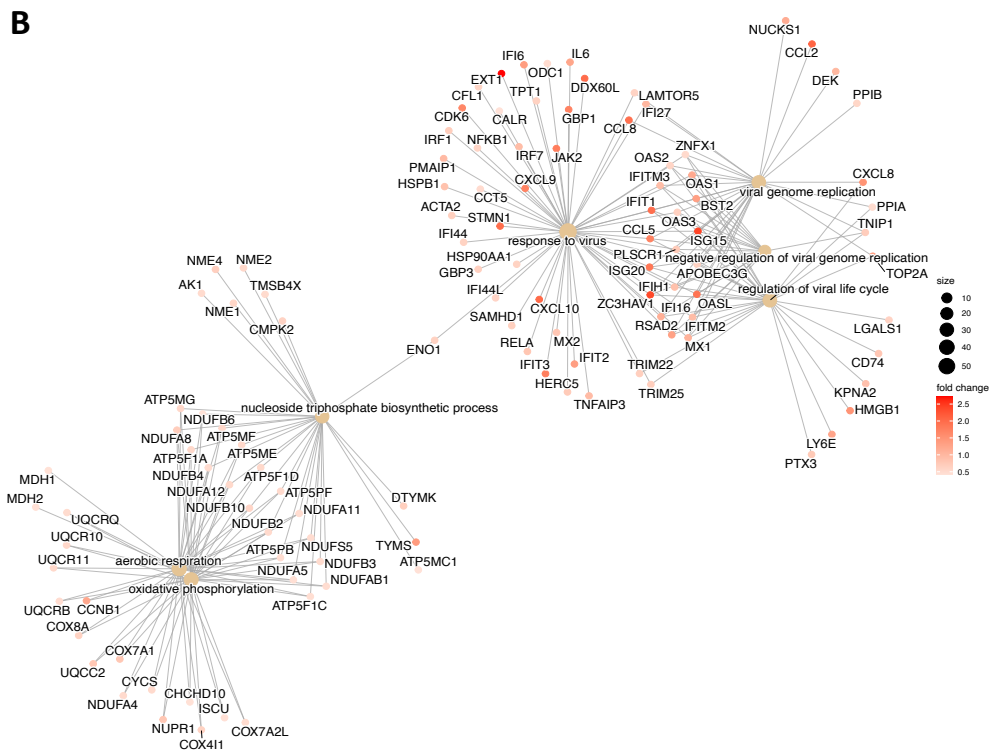
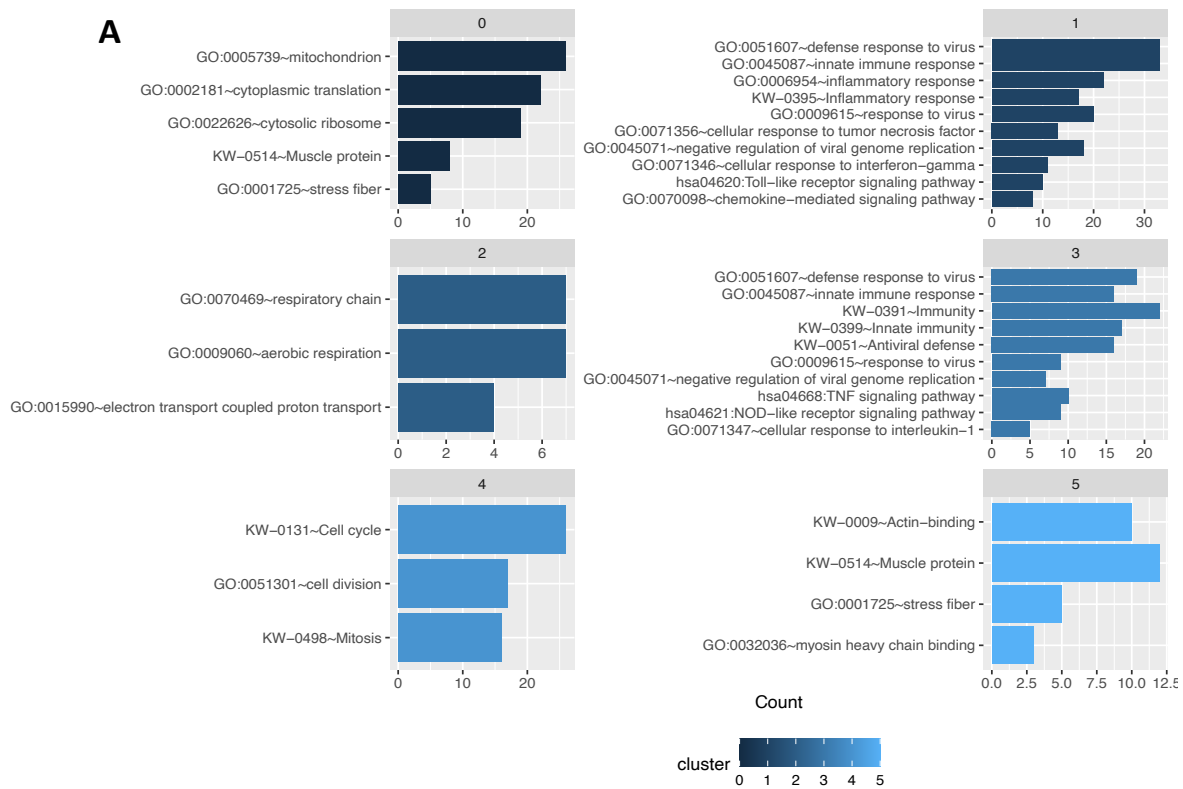


Figure 42. Enrichment analysis of unstimulated and stimulated MSCs. **A)** Barplot of the main functions of each cluster of **Figure 40B.** **B)** Network of the main functions of the most differentiated genes of the top 50 differential genes.

c. Identification of subpopulations in control MSCs

Table 14 shows the distribution of the two patients across the clusters, highlighting the robustness of the clusters across different cell quantities. The most significant result of this new analysis was the identification of three distinct clusters in the control MSCs. Among these clusters, cluster 0 was the dominant cluster, representing 77% (n= 1057) of the cells in two samples. Cluster 2, which showed the most pronounced differentiation from cluster 0 represented the 13% (n= 179) and the cluster 1 showed a mixed characteristic among the other samples and represented the 9,5% (n= 130) (**Figure 43**).

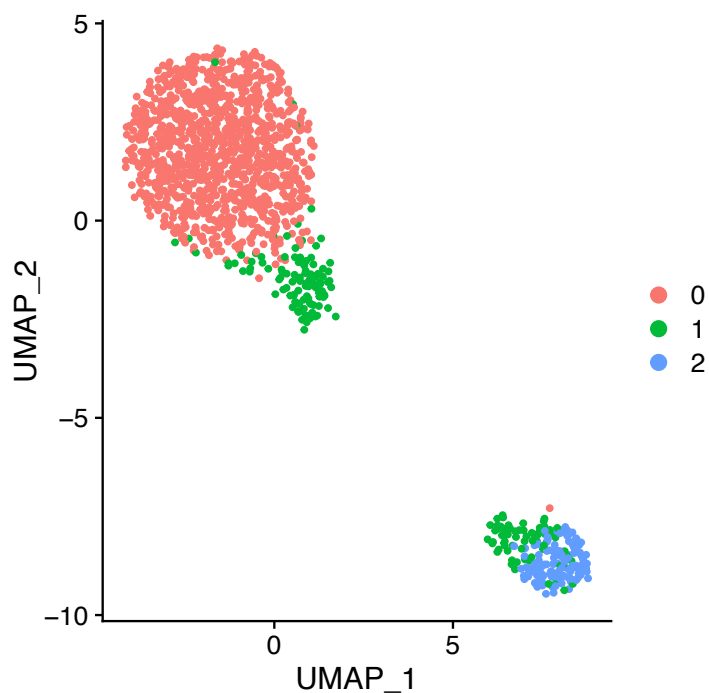


Figure 43. UMAP plot clustering of the control MSCs. The plot was analyzed in two samples grouped by clustering: 0, 1, 2.

We obtained 1680 genes from the differential analysis with a filter of $FDR < 0.05$. From this filter, 115 genes were from cluster 0, 112 genes from cluster 1 and 1453 genes from cluster 2.

Figure 44 shows Heatmaps of the top 50 genes with the highest adjusted p-value in the control samples. Consistent with **Figure 43**, the **Figure 44A** illustrates the heatmap of the control analysis, delineating the three clusters identified. Notably, similarities in gene expression levels were observed between clusters 0 and 1. However, cluster 2 contained an additional set of genes that were unique to this cluster. Comparing this specific Heatmap with the Global Heatmap, where both stimulated and unstimulated MSCs were considered, the clusters with higher similarities include clusters 0, 2, 4, and 5, a pattern consistent with **Figure 40**. Clusters 0, 4, and 5 reflect the characteristics of cluster 0 in the control samples, while cluster 2 aligns with cluster 2 in the control Heatmap (**Figure 44A**).

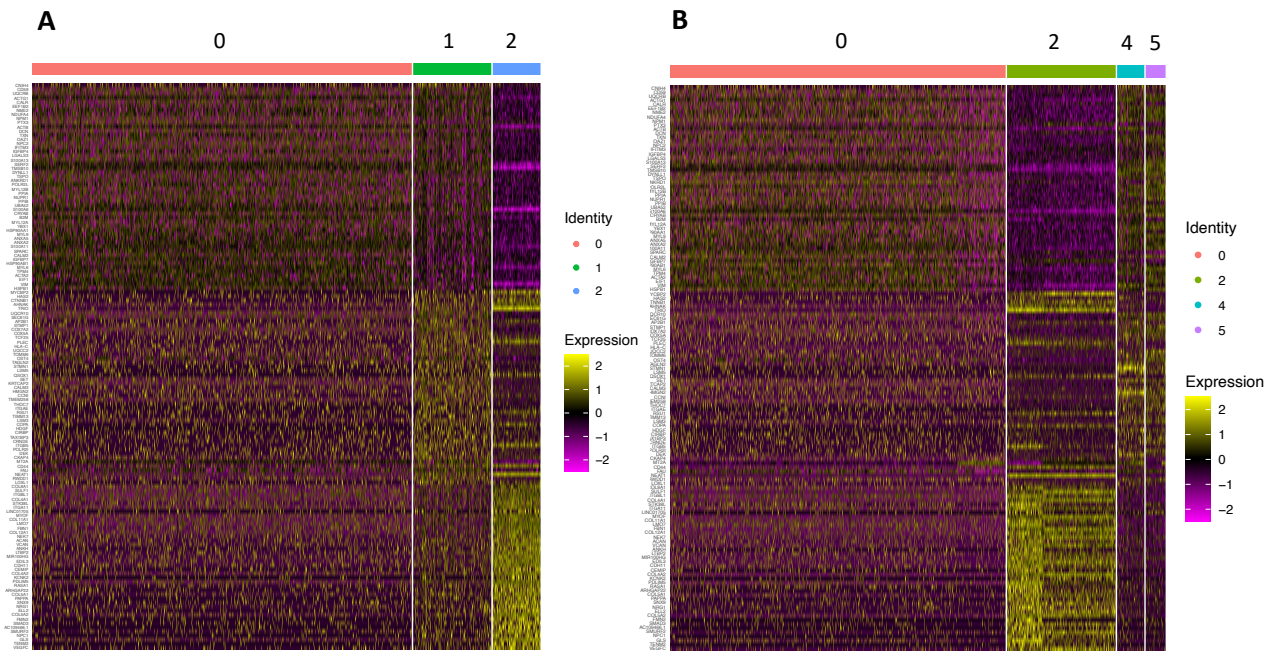


Figure 44. Heatmaps of the top 50 genes in the Control MSCs. **A)** Heatmap of top 50 genes in 3 clusters of control MSCs. **B)** Heatmap of top 50 genes represented in the stimulated and control MSCs.

In addition, the most differentially expressed genes in clusters 0, 1, and 2 were aggregated and displayed in violin plots compared to the global clusters, as shown in **Figure 45**. **Figure 45A** and **B** shows that the most prominent gene markers, including ACTA2, MYL12A, MYL12B, MYL6, TAGLN, TPM1, TPM2 and TPM4, were overexpressed in cluster 1 of our control analysis, whereas in the Global analysis of stimulated and unstimulated MSCs, they were mostly activated in clusters 1 and 5. Conversely, in **Figure 45C** and **D**, the most differentially expressed genes in cluster 2 of our control analysis, were Collagen family genes: COL8A1, COL4A1, COL4A2, COL11A1, COL5A2, COL12A1 and COL11A1. Other overexpressed genes included: FBN1, and VCAN. Interestingly, these genes were also overexpressed in cluster 2 in the Global analysis. This observation suggests that the similarities observed in **Figure 44** between cluster 2 of both Heatmaps were due to these specific genes. Notably, cluster 0 did not exhibit any distinct set of genes that were uniquely associated with this cluster.

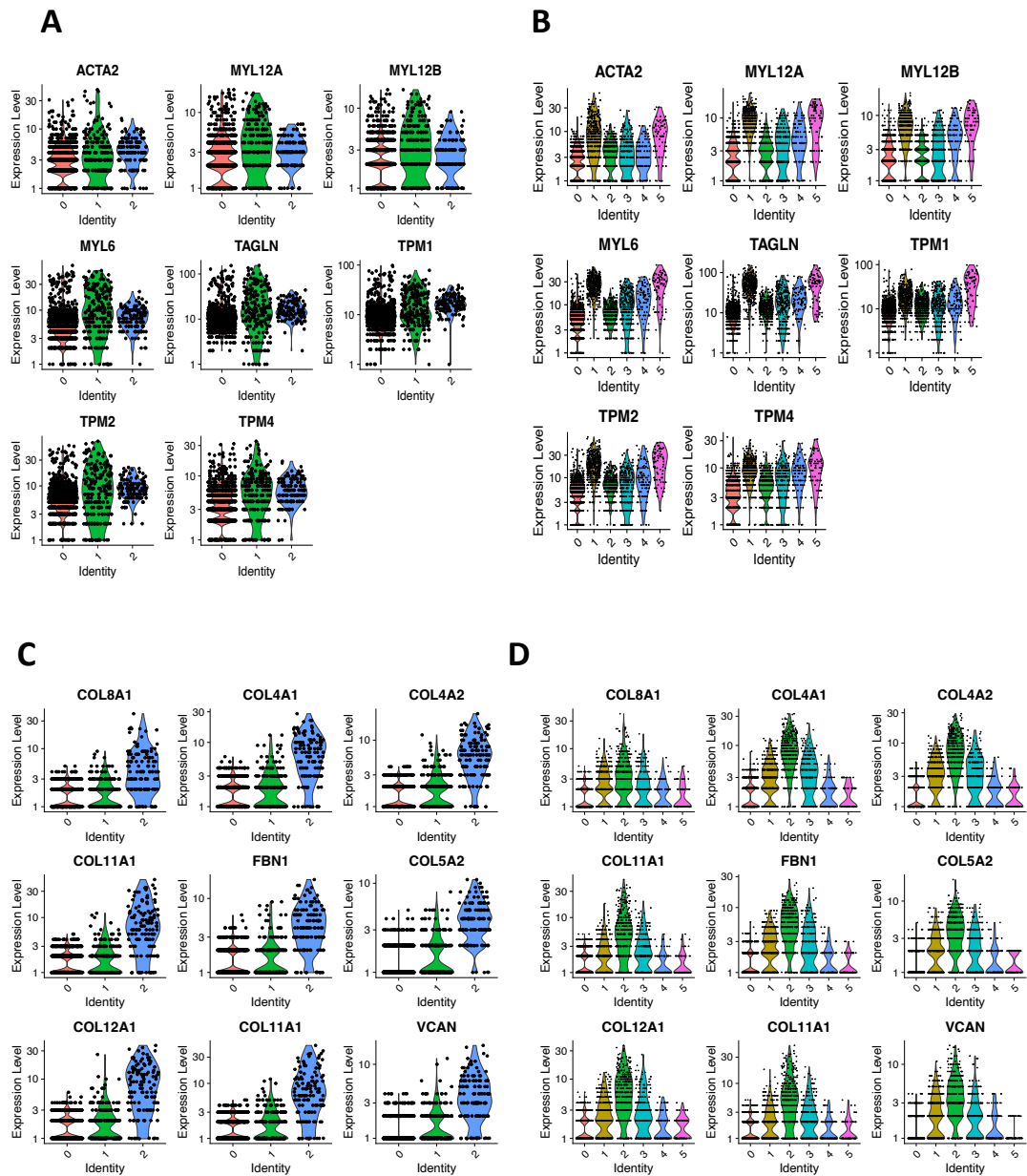


Figure 45. Violin plots associated with the most differential expressed genes in control MSCs. **A)** Most differentiated genes of Cluster 1 in 3 clusters of Control analysis. **B)** Most differentiated genes of Cluster 1 in 6 clusters of Global analysis. **C)** Most differentiated genes of Cluster 2 in 3 clusters of Control analysis. **D)** Most differentiated genes of Cluster 2 in 6 clusters of Global analysis.

We then performed an enrichment analysis for each cluster using DAVID Bioinformatics, as presented in the **Figure 46**.

The most differentiated genes (62) of cluster 0 revealed an association with muscle differentiation. Tropomyosin (TPM) family genes were widely distributed actin-binding proteins involved in the contractile system of striated and smooth muscle and the cytoskeleton. Notably, myosin genes were also prominent in these functions. Other relevant genes presented were TAGLN, ACTA2 and ACTB, which were of particular interest because they revealed new functions not previously associated with muscle functions in BM-MSCs.

Cluster 1 presented 68 differential genes associated with broader cell functions, such as mitochondrial membrane, cell adhesion or cellular respiration. A significant number of mitochondrial (MT) genes were evident, underscoring their importance in cluster 1. In cell adhesion, the most relevant genes were CD44, ITGB5 or CTNNB1.

Finally, the cluster 2, presented specific functions related to the epidermal growth factor domain with well-known genes present in MSCs such as ACAN or VCAN. Furthermore, there were specific functions associated with Collagen genes that included COL4A2, COL5A1, COL4A1, COL11A1, COL12A1, COL5A2, COL6A3 and COL8A1. Other genes associated with typical functions of the MSCs were extracellular matrix or cytoskeleton, including genes of the Collagen family, LTBP2, EDIL3, FBN1, ACAN, MACF1, LRP1, ANK2, FMN2, UTRN and SNTB1.

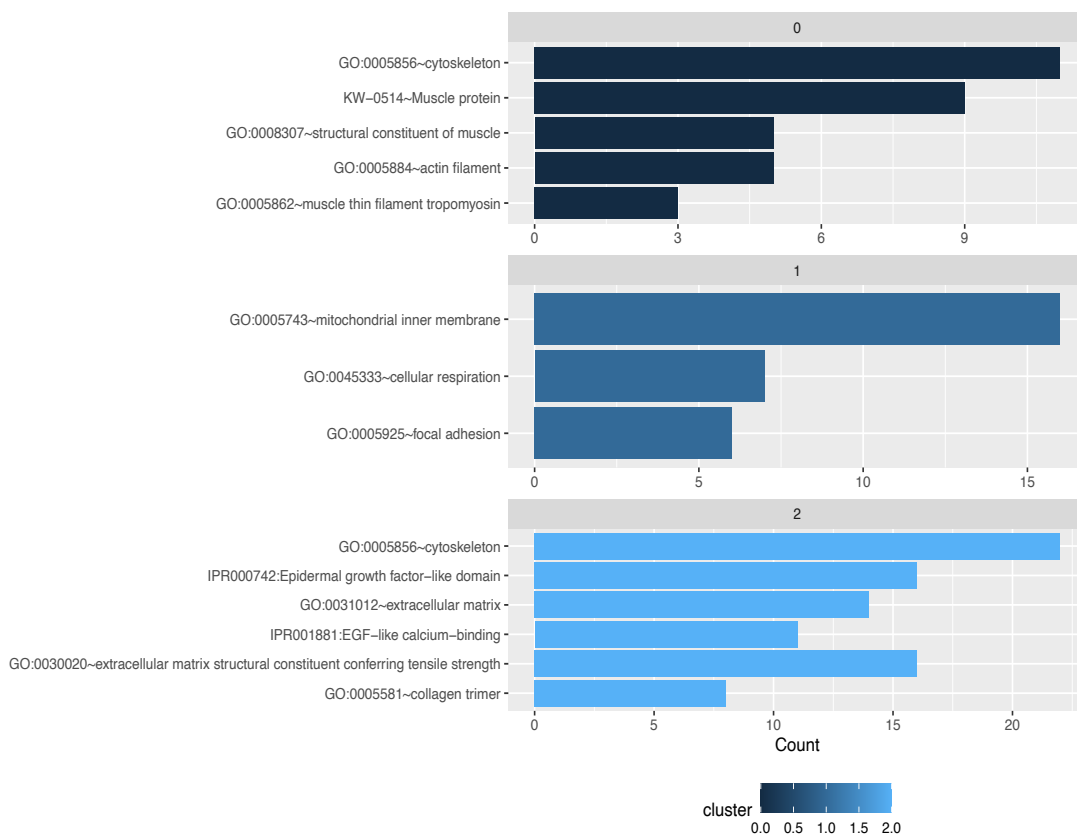


Figure 46. Barplots of enrichment analysis associated with the three cluster of control MSCs differential analysis.

d. Identification of subpopulations in stimulated poly(I:C) MSCs

In this section of results, we have been addressed to identify the different subpopulations allowed in the stimulated MSCs with poly(I:C) 25 μ g/ml in time of 6 hours. In the **Figure 47** was shown the formation of five different clusters in the stimulated MSCs. The most differentiated clusters were cluster 0, with 360 cells, cluster 1 with 264 cells and cluster 2 with 201 cells. Cluster 3 showed a more extensive distribution with 177 cells, while cluster 4 was added to cluster 2 with 39 cells. The total matrix contained 28017 genes in 1041 cells.

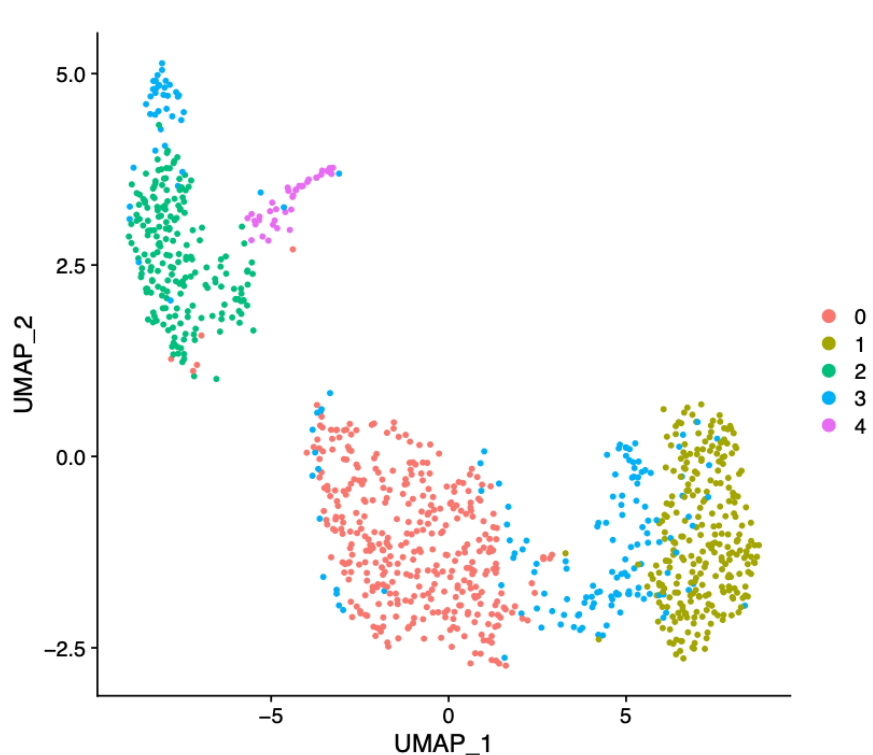


Figure 47. UMAP clustering of stimulated MSCs cells. The plot was analyzed in two samples grouped by clustering: 0, 1, 2, 3 and 4.

In parallel to our analysis of unstimulated MSCs, we performed a differential analysis of our stimulated MSCs. The results of this analysis were represented in **Figure 48A** by the Heatmap with the top 50 genes from each cluster. There were strong differences between cluster 0 and 1, which corresponded to cluster 1 and 2 in the Global Heatmap with unstimulated and stimulated MSCs (**Figure 48B**). Similarly, cluster 3 was overexpressed in all samples, similar to cluster 4 in the Global Heatmap (**Figure 48B**). Conversely, cluster 2 was down expressed, which was cluster 1, corresponded to cluster 0 in **Figure 48B**. Furthermore, cluster 4 of **Figure 48A** represented the last set of genes similar to cluster 0. Additionally, the genes that presented an up expressed pattern in cluster 4 of **Figure 48A** were also up expressed in cluster 1 of the Global Heatmap (**Figure 48B**). The Global Heatmap included all clusters due to the expanded distribution of stimulated MSCs, as observed in **Figure 41**. This expansion resulted in greater interplay between the clusters, creating a cohesive representation within the Global Heatmap.

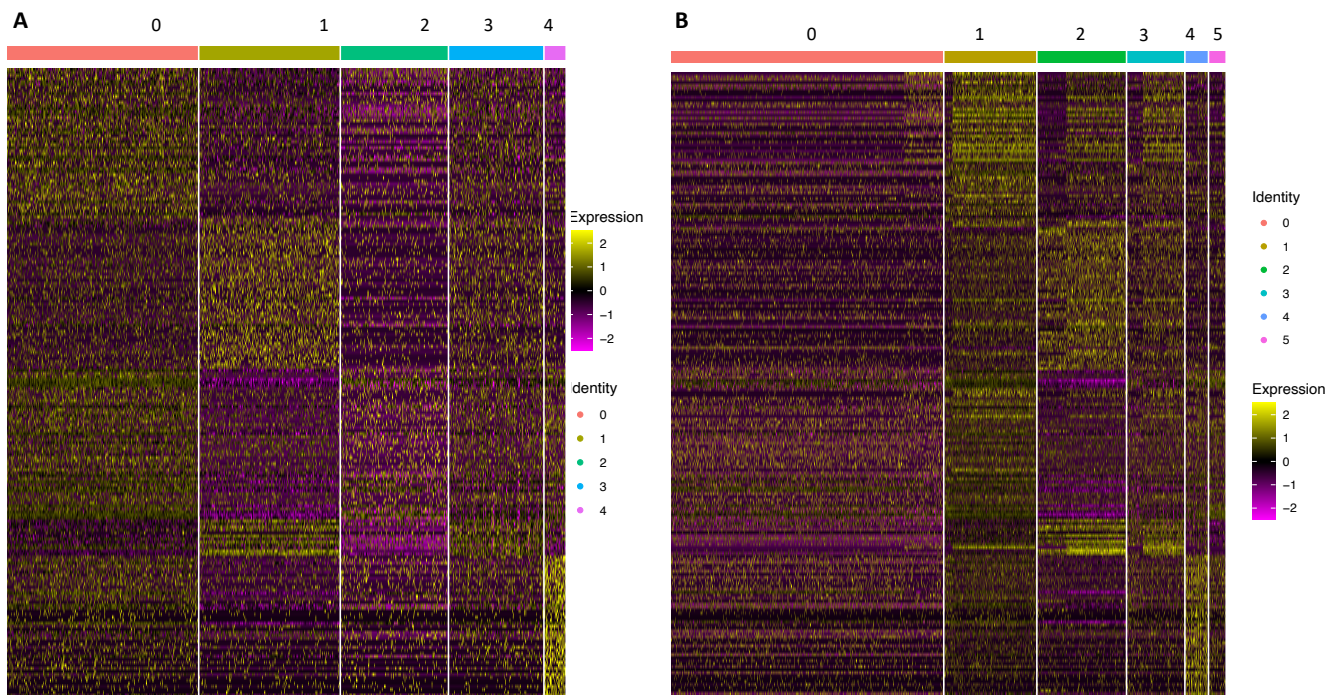


Figure 48. Heatmaps of the top 50 genes in the stimulated MSCs. **A)** Heatmap of top 50 genes in 35 clusters of stimulated MSCs. **B)** Heatmap of top 50 genes represented in the stimulated and control MSCs.

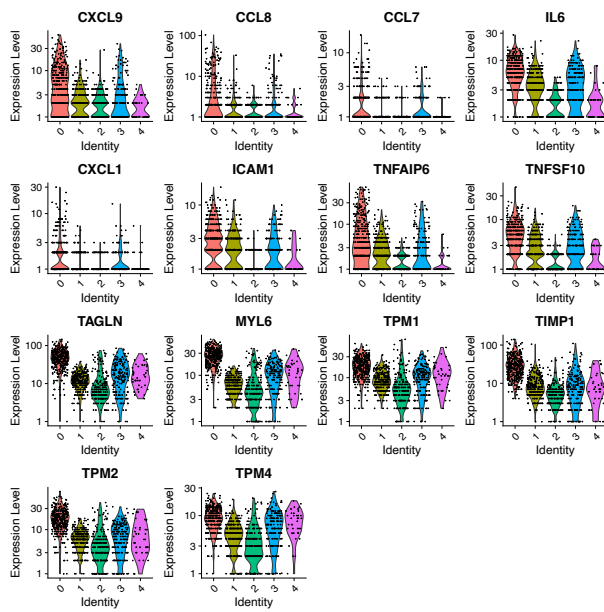
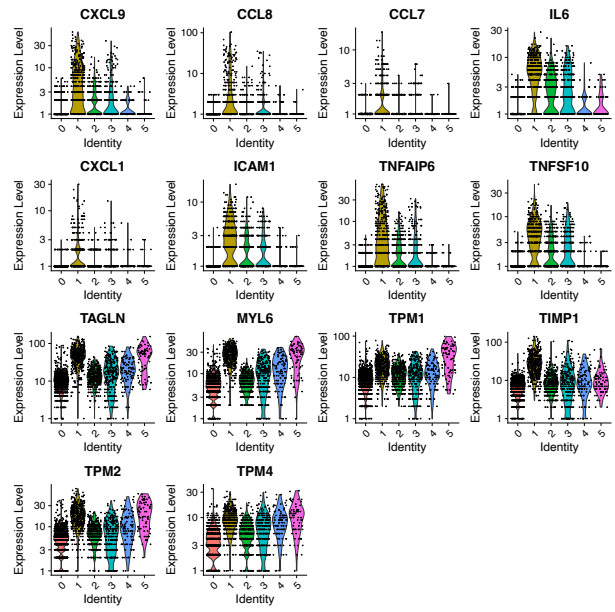
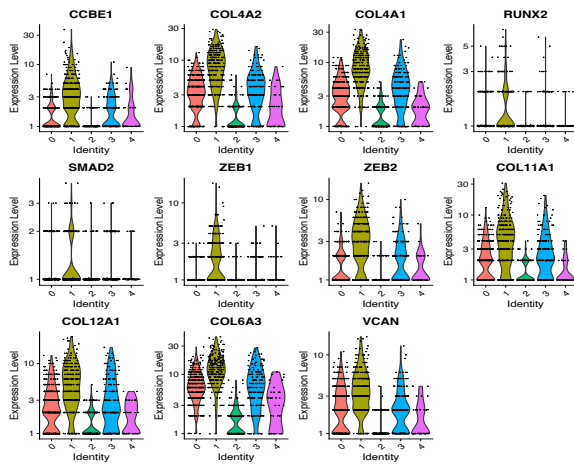
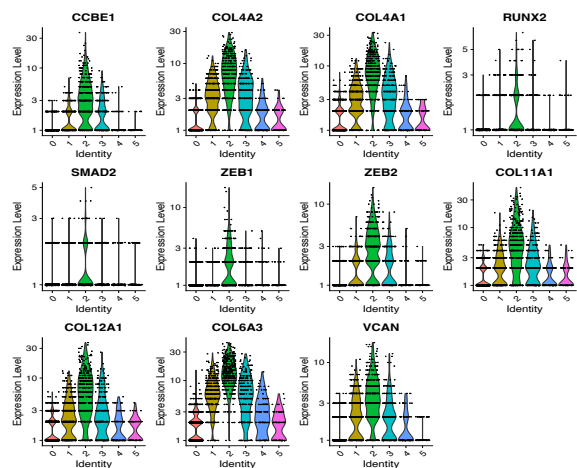
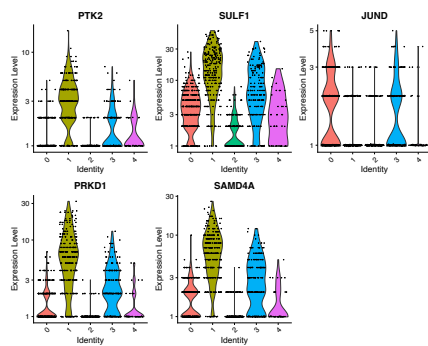
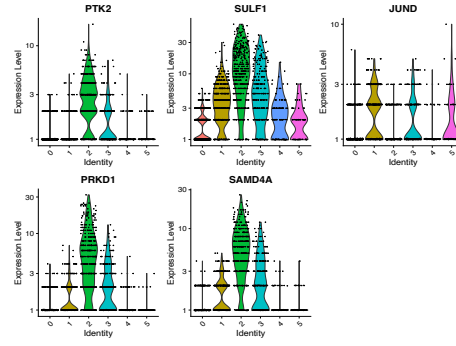
In addition to this study, we used a series of violin plots, as in control MSCs, to visually represent the most differentially expressed genes in each cluster, as shown in **Figure 49**.

The largest cluster, cluster 0, was enriched with relevant genes related to inflammatory processes, such as CXCL9, CCL8, CCL7, IL6, CXCL1, ICAM1, TNFAIP6 and TNFSF10, which correspond to cluster 1 of the Global analysis. We also incorporated genes from cluster 2 of Control MSCs analysis such as TAGLN, MYL6, TPM1, TIMP1, TPM2 and TPM4. These genes were also overexpressed in cluster 0, and corresponded to cluster 5 in Global analysis (**Figure 49A and B**).

Figure 49C shows that cluster 1 had overexpressed genes in all Collagen family genes: COL4A1, COL4A2M, COL11A1, COL12A1 and COL6A3. Notably, the gene RUNX2, previously associated with osteogenic differentiation, was also prominent as seen in Chapter 3. Moreover, we added ZEB1 and ZEB2, which are known genes related to EMT. In the corresponding Global analysis, osteogenic and EMT genes were up expressed only in cluster 2, while Collagen family genes were also up expressed in cluster 3 (**Figure 49D**).

Cluster 2 had similarities to both clusters 0 and 1, with relevant genes including JUND, PTK2, SULF1, PRKD1 and SAMD4A. Notably, cluster 2 exhibited substantial differentiation in the Global analysis, except for the gene JUND, which correlated with clusters 1 and 5 (**Figure 49E and F**). This cluster 2 was highly related to cluster 1 in all genes that appeared in the Global analysis of the **Figure 49**.

On the other hand, **Figure 49G** depicted cluster 3 which contained three genes related to adipogenic differentiation, as we previously discussed in Chapter 3: APOL6, PARP14 and FBXL7. Furthermore, SLIT2, FOXP1, and MAML2 genes were also overexpressed. However, it was evident that cluster 1 predominantly showed the highest differential expression levels. These genes were significantly up expressed in cluster 2 of the Global analysis (**Figure 49H**).

A**B****C****D****E****F**

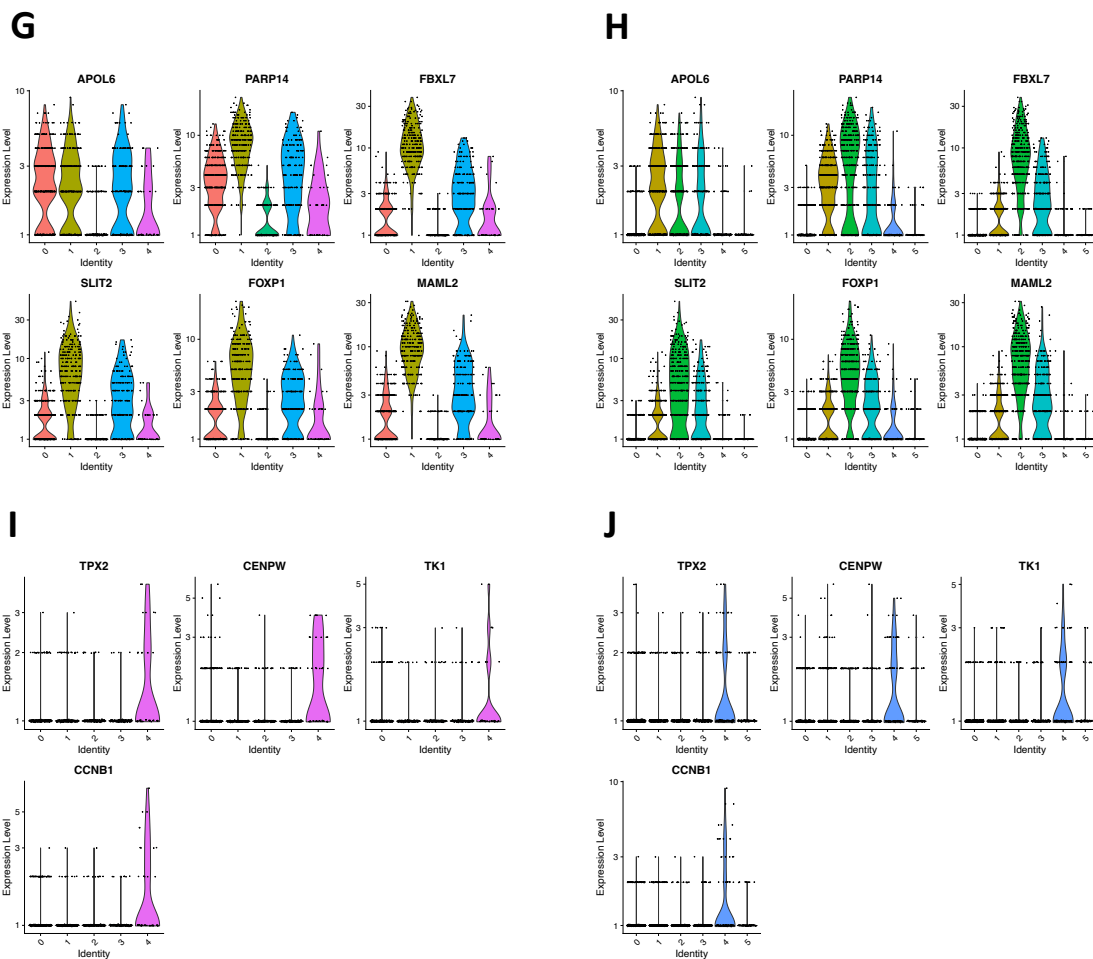


Figure 49. Violin plots associated with the most differential expressed genes in stimulated MSCs. **A and B)** Most differentiated genes of Cluster 0 in control and Global analysis. **C and D)** Most differentiated genes of Cluster 1 in control and Global analysis. **E and F)** Most differentiated genes of Cluster 2 in control and Global analysis. **G and H)** Most differentiated genes of Cluster 3 in control and Global analysis. **I and J)** Most differentiated genes of Cluster 4 in control and Global analysis.

To further understand the patterns and biological implications of each cluster, we performed to perform a functional enrichment analysis by grouping genes according to their associated biological functions. As shown in the **Figure 50**, this enrichment analysis was performed for clusters 0, 1, 2, and 4 using the DAVID Bioinformatics tool.

The genes inside cluster 3 could not be categorized in specific functions associated, there were only 12 significant genes that associated this cluster 3: SULF1, PRKD1, RSAD, KLF, ALCAM, CALU, MYOF, LRR75A, EMP1, CRIM1, ITGB and CLMP where DAVID Bioinformatics tool couldn't detect specific association. Nevertheless, the Violin plots of **Figure 49G and H** that had a clear association with adipogenic differentiation, although not limited to that cluster.

The main cluster 0 had the most specific functions associated with poly(I:C) stimulation, such as: response to virus, antiviral defense, response to interferon -alpha -beta and -gamma and innate immunity, reflected in the genes IFITM3, BST2, CCL8, IFITM2, OAS1, ISG15, OASL.

Cluster 1, on the other hand, had specific functions of MSCs, such as cytoskeleton or actin binding. Finally, cluster 4 with associated genes of TPX2, CENPW, TK1 and CCNB1 had a distinct and pronounced differential expression level in this cluster. Interestingly, these same genes were highly upregulated in cluster 4 in the Global analysis (**Figure 49I** and **J**).

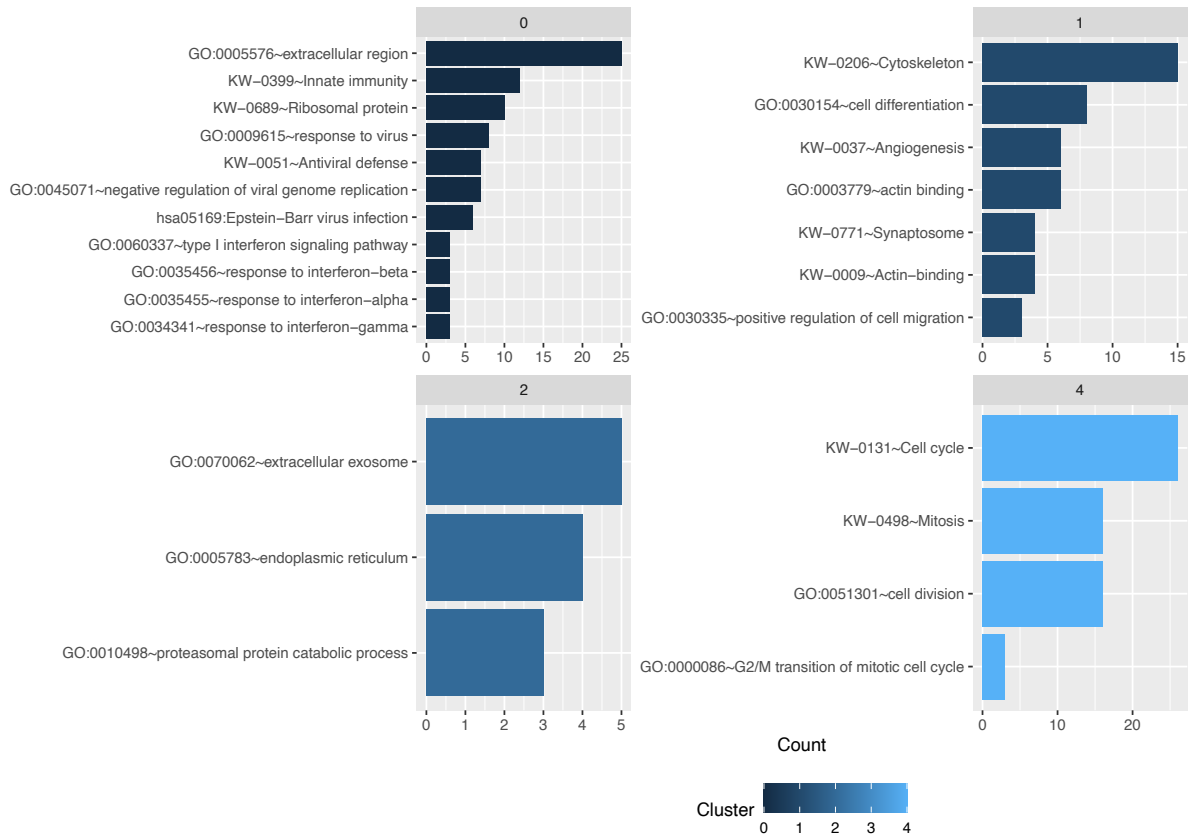


Figure 50. Barplots of enrichment analysis associated with clusters 0, 1, 2 and 4 of stimulated MSCs differential analysis.

In addition, this cluster was also involved in angiogenesis and synaptosome with genes such as CCBE1, RNF213, COL4A2, COL4A1, PRKD1, PTK2. As we saw in **Figure 49C** and **D**, this cluster was also involved in genes of cell migration or EMT such as ZEB1 and ZEB2 and cell differentiation such as RUNX2.

Cluster 2 displayed functions related to cell secretion, including extracellular exosome, endoplasmic reticulum or proteasomal catabolic process. Genes such as PSMA5, PSMB4, TPM4, DSTN, SDF4 were relevant within these functions.

Finally, cluster 4 had functions associated with cell cycle, mitosis and cell division, with specific genes such as MKI67, CCNB1, PTTG1, RAD21, NUSAP1, CENPW, KIF23 and SMC4, observed in the **Figure 49I** and **J**.

This independent analysis of stimulated MSCs revealed that not all clusters responded to stimulation in the same manner, including unstimulated cells that contained the same properties of the unstimulated MSCs. Performing an independent analysis of each condition allowed the identification of new subpopulations in the Global analysis of both stimulated and unstimulated MSCs.

e. Identification of new subpopulations in control and stimulated MSCs

The independent analysis of each condition and the association with the Global analysis, where include the two samples in control and stimulated with Poly(:C) 25µg/ml conditions, allows robustness in the the determination of these new MSC subpopulations. In **Figure 51**, we have presented the three primary UMAP plots as a function of the clusters determined by the *Seurat* package, the condition of the MSCs and the new UMAP plot reflecting the new subpopulations according to the previous sections of this chapter.

In **Figure 51A** and **B**, clusters 1 and 3 were transformed into the major groups of poly(I:C) stimulated MSCs. As we showed in **Figure 48**, these clusters corresponded to the clusters 0, 1 and 2 of the stimulated cells. These results demonstrate that the cluster 1 in the Global analysis corresponds to an inflammatory profile, characterized by genes such as CXCL9, CCL8, CCL7, IL6, CXCL1, ICAM1, TNFAIP5 and TNFSF10. Enrichment analysis further supports this observation, showing the typical functions of poly(I:C) stimulation at time 6 hours as we determined in Chapter 3, i.e. responses to virus. However, cluster 3 was also identified as poly(I:C) stimulated MSCs, exhibiting lower expression levels in contrast to cluster 0, while still sharing similar characteristics with cluster 3, such as adipogenic differentiation in up expression of PARP14 or FBXL7.

Moreover, cluster 4 displayed significant up expression in cluster 4 of stimulated MSCs, indicating the genes such as TPX2, TK1, HDGF and CEMPW, as a subpopulation related to cell division and mitosis of the MSCs.

Interestingly, clusters 0, 2 and 5 included a mixture of both conditions of MSCs. Upon separate analysis revealed specific genes MYL6, MYL12A, MYL12B, TAGLN, TPM1, TPM2 and TPM4 were highly up expressed in a distinct cluster, which in both cases corresponded to cluster 5. Enrichment analysis clarified the association of this set of genes with muscle functions, and TAGLN gene was also presented as a possible new gene marker in Chapter 1, corroborating the identification of MSCs. Cluster 2 had a high expression in Collagen family genes in both conditions: COL8A1, COL4A1, COL4A2, COL5A2, COL12A1 and COL11A1, along with VCAN and FBN1. Interestingly COL4A1 and COL4A2 genes were also presented in Chapter 1, and differing in the clustering with the TAGLN. Strikingly, stimulated cells in this cluster also showed differentiation properties with genes such as RUNX2, PARP14 or FBXL7, PRKD1 and APOL6. This cluster also showed genes related with EMT, such as ZEB1 and ZEB2, as we commented previously. This suggests that cluster 2 had a predisposition towards differentiation, particularly when were stimulated with poly(I:C).

Finally, cluster 0 was the largest cluster of the analysis, lacking specific genes or features that could be clearly associated with it. Enrichment analysis of this cluster associated it with mitochondria, cutoplasmatic trasnition and cytosolic ribosome (**Figure 42**).

This new analysis, derived from the previous independent analysis of stimulated and unstimulated MSCs, allowed a deeper understanding of the subpopulations established in both conditions. As depicted in **Figure 42**, the enrichment analysis alone failed to identify the newly established subpopulations derived from the separate analyses. Moreover, this approach allowed us to precisely classify our clusters as a function of the both comparisons, labeling cluster 0 as the most general and unscpecific MSCs, cluster 1 as typical poly(I:C) stimulated MSCs at time 6 hours, cluster 2 as MSCs prone to differentiation, cluster 3 as intermediate or low stimulation of MSCs with characteristics associated with cluster 2, cluster 4 as growth and division MSCs, and cluster 5 as MSCs with muscle-related properties.

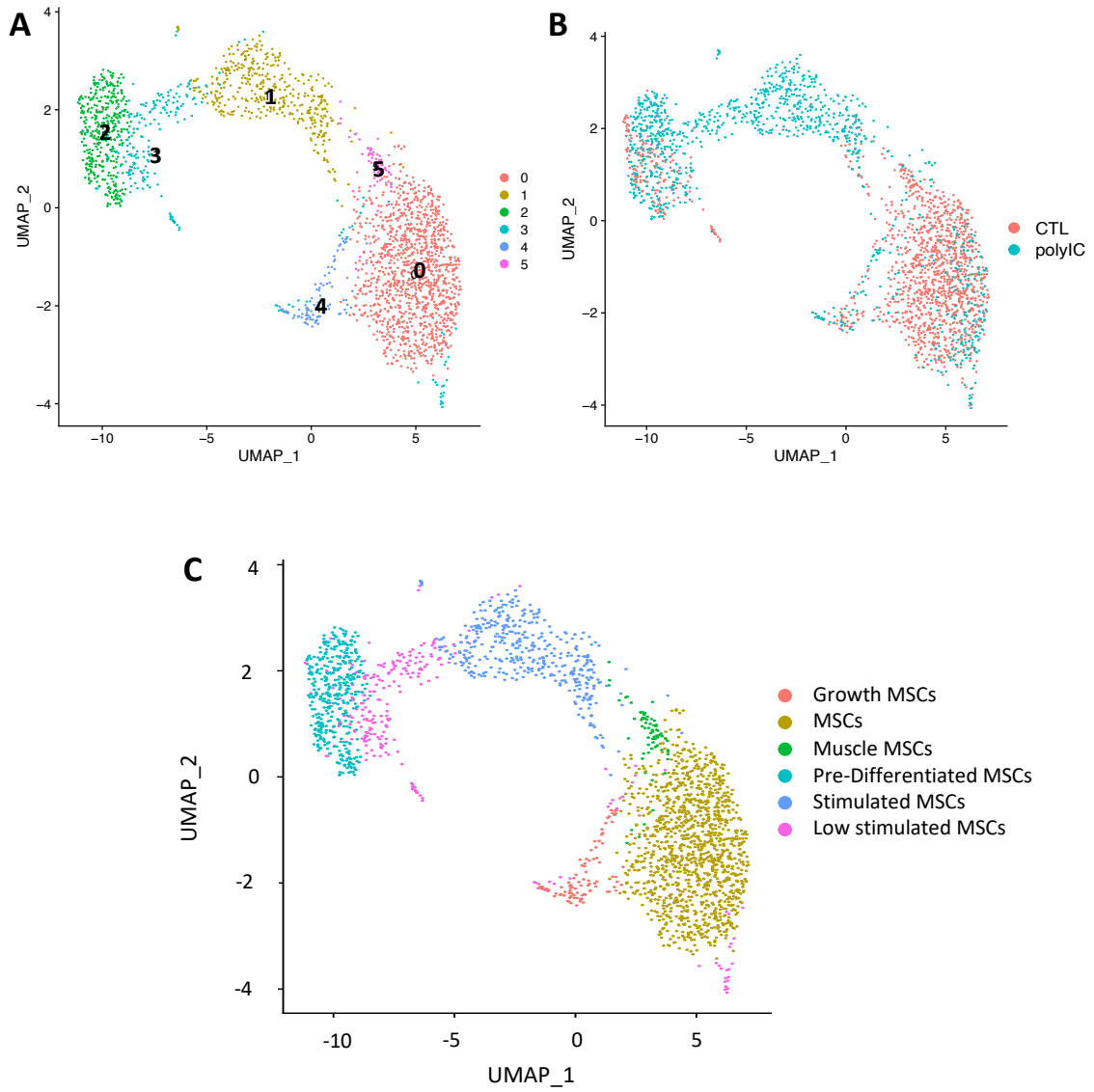


Figure 51. UMAP plots of unstimulated and stimulated MSCs. **A)** UMAP plot of Seurat Clusters. **B)** UMAP plot in function of control and poly(I:C) stimulation of the two patients. **C)** UMAP plot of new subpopulations established.

2.2 Comparison of *in vivo* and *in vitro* Mesenchymal stem Cells

This section of the chapter was focused on the fellowship in the Department of Oncohematology at the University and University Hospital of Zurich. All the data obtained during the stage were confidential, for this reason we have presented only the main results of the project carried out during the stage.

The fellowship consisted in the integration of the *in vitro* samples from our two healthy donors, (specifically the control MSCs mentioned in previous sections) with *in vivo* MSCs samples from 12 different healthy donors. As we presented in the **Figure 43** and **44**, our independent clustering yielded 3 clusters, where cluster 0 was the least unspecific, cluster 1 was associated with myogenic and thrombopoietic genes and cluster 2 with Collagen family genes (**Figure 45**).

The samples collected from the University Hospital of Zurich were categorized into 5 groups of MSCs: CAR cells (CAR), Osteogenic cells (osteo), myogenic cells (myo), and two more group cells. Among them, we used only CAR cells, osteogenic cells and myogenic cells. When we clustered the three groups (**Figure 52A**), the clear classification emerged with the three groups being completely separated from each other. However, the CAR cells showed two subgroups of cells, with only a few of them were interspersed with osteogenic cells. To better classify the *in vivo* cells with our *in vitro* cells, we briefly defined the two subpopulations of CAR cells. According to the similarities and proximity between the *Seurat* clusters (**Figure 52B**), we grouped the clusters: 1, 7, 8, 10 and 12 as CAR 1 and clusters: 0,2,3,9 and 11 designated as CAR 2.

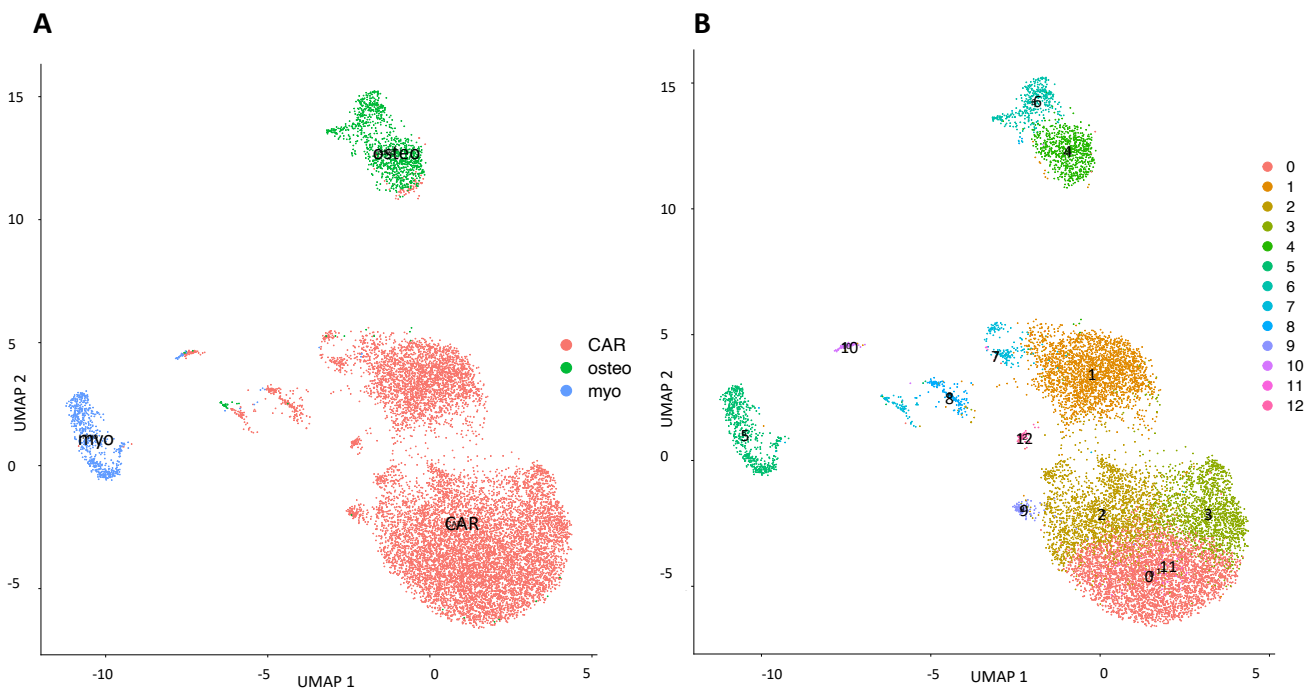


Figure 52. UMAP plot of *in vivo* MSCs from healthy donors. **A)** UMAP plot of CAR cells (CAR), osteogenic cells (osteo) and myogenic cells (myo) from *in vivo* MSCs from healthy donors. **B)** UMAP plot of Seurat clusters in the three groups.

Once the samples were aggregated, we integrated the *in vivo* samples with our *in vitro* samples using the *Harmony* algorithm as it was indicated in the Materials and Methods section. The **Figure 53A** shows the integration of the *in vitro* and *in vivo* MSCs. The most interesting results of this analysis was the approximation of the majority of *in vitro* MSCs with CAR 2 cells and the total integration of a few part of *in vitro* MSCs with CAR 1. Moreover, this few group of cells were clustered in the same cluster 7 as it was indicated in **Figure 53B**.

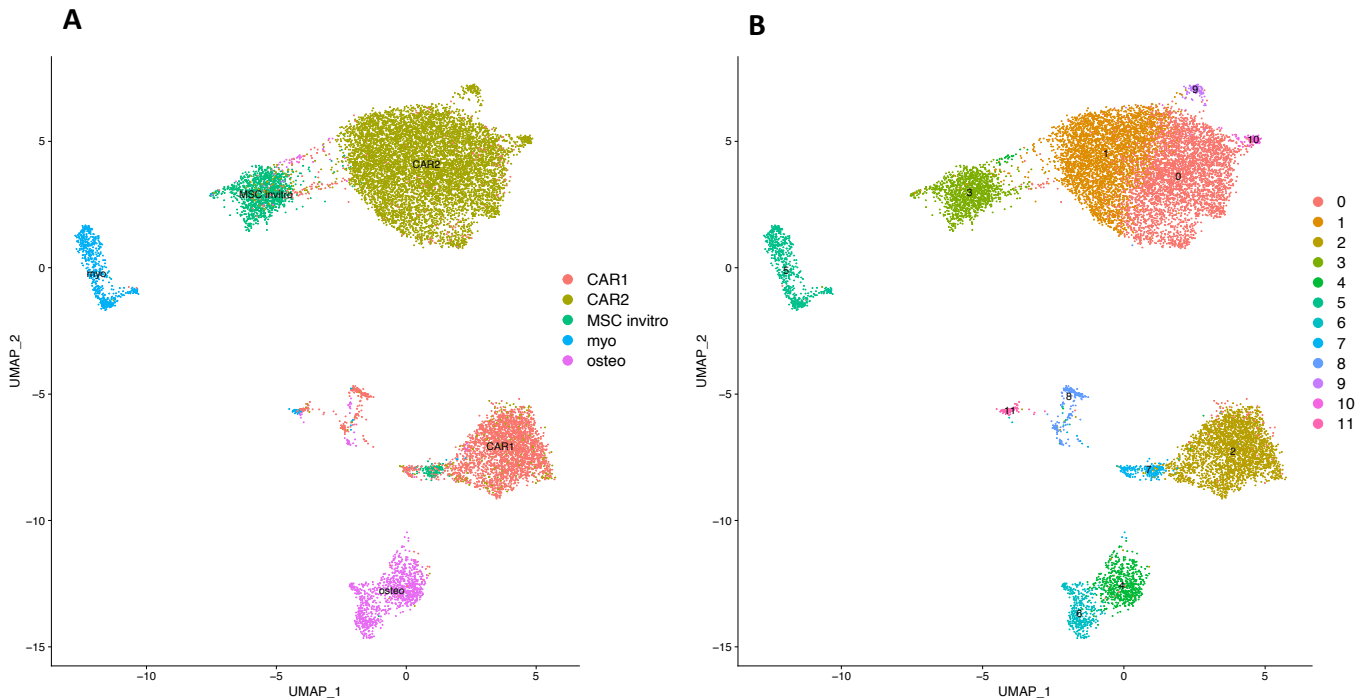


Figure 53. UMAP plots of *in vitro* and *in vivo* MSCs integration. **A)** UMAP plot of the different groups of MSCs: CAR1, CAR2, Myo, Osteo and MSC invitro. **B)** UMAP plot of Seurat clusters.

Due to the confidentiality of the data in the *in vivo* samples, the focus was directed to the similarities of the *in vitro* MSCs with the two subgroups of *in vivo* CAR cells. This led us to perform two enrichment analyses of the common genes in *in vitro* MSC clusters 3 and 1 with CAR 2; and other independent analysis of cluster 7 of the *in vitro* MSCs and CAR 1 cells; as depicted in **Figure 54** and **Figure 55**.

Figure 54A and **B** correlated the key functions and the significance in the differential analysis. The most correlated and the most significant functions were differentiation, regulation of hematopoiesis and regulation of myeloid cell differentiation. Conversely, functions such as ossification and wound healing were not correlated with any other function. Osteoblast differentiation was the most significant function in the enrichment analysis, followed by the ossification. Genes associated with the common functions were represented in **Figure 54C**.

Interestingly, known genes such as FN1, ACTA2, MYL9, TPM1, TIMP1 and TAGLN previously observed in our independent control *in vitro* MSCs analysis, were now related with wound healing. Moreover, TAGLN, was also observed in subgroup of Muscle MSCs in **Figure 52**, and was also a novel marker of MSCs in Chapter 1. The function of response to mechanical stimulus also included genes such as CXCL12, or FOS, which also appeared in our previous results. The functions of myeloid cell differentiation and regulation of hematopoiesis were highly associated with genes such as INHBA, TNFRSF11B, SFRP1, VCAM1 and LGALS1. Furthermore, the ossification function included typical

osteogenic differentiation-associated genes such as COL1A1. Additionally, this function contained other genes such as TPM4, ID4, IGFBP3 or TNGAIP6.

On the other hand, The **Figure 55** provided insight into the enrichment analysis of cluster 7, which contained the smaller group of *in vitro* MSCs and some CAR 1 cells. In this context, **Figure 55A and B** illustrate the correlations and significance of the specific functions. Notably, cell-substrate adhesion had the most genes associated, with additional correlations observed between endoderm development, endoderm formation, endodermal cell differentiation and formation of primary germ layer formation. The most significant functions were associated with cartilage development, and followed by the axon development. **Figure 55C** shows the relationship between the genes and functions. The top three clusters were related to extracellular matrix and collagen fibril organization; endoderm development and endodermal cell differentiation, and cell-substrate adhesion.

Extracellular matrix and collagen fibril organization had identity genes such as COL4A1, COL1A, POSTN, SULF1. Furthermore, there was a group of genes that shared functions related to endodermal cell differentiation, endoderm development, such as Collagen family genes: COL8A1, COL12A1, COL11A1, COL5A2, COL4A2, COL8A1 and EXT1. In this case, COL4A2 and COL4A1 genes were also observed in the subgroup of Pre-Differentiated MSCs in **Figure 52**, as well as novel markers of MSCs in Chapter 1. Finally, the function of cell-substrate adhesion was associated with genes such as SERPINE1, ITGA11, CDH13, S100A10, ITGA11 and FN1.

These two independent enrichment analyses determined the predominant functions in differentiation, ossification and response to stimuli within the larger cluster. In addition, the footprint of genes such as FN1, ACTA2, MYL9, TPM1, TIMP1 and TAGLN, which did not appear in either function analysis, could show a relationship with the muscle MSCs subgroup of **Figure 52**. While the smaller cluster, with the minority cells, was clearly associated with collagen organization, extracellular matrix and cell-substrate region.

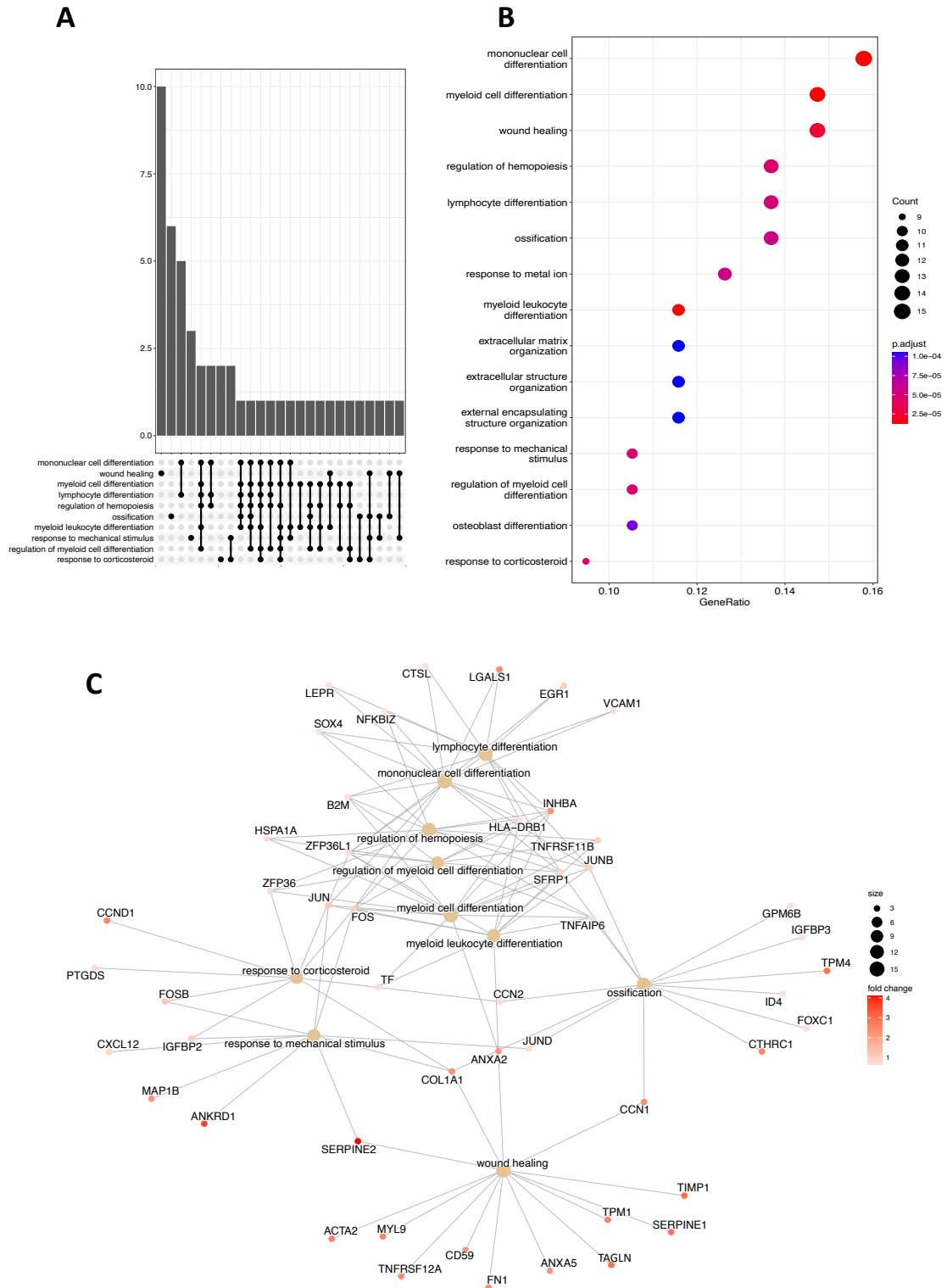


Figure 54. Enrichment analysis of Cluster 1 and 3. **A)** Barplot of the correlation of functions. **B)** Dot Plot of the significant functions. **C)** Network plot of the genes associated with the functions.

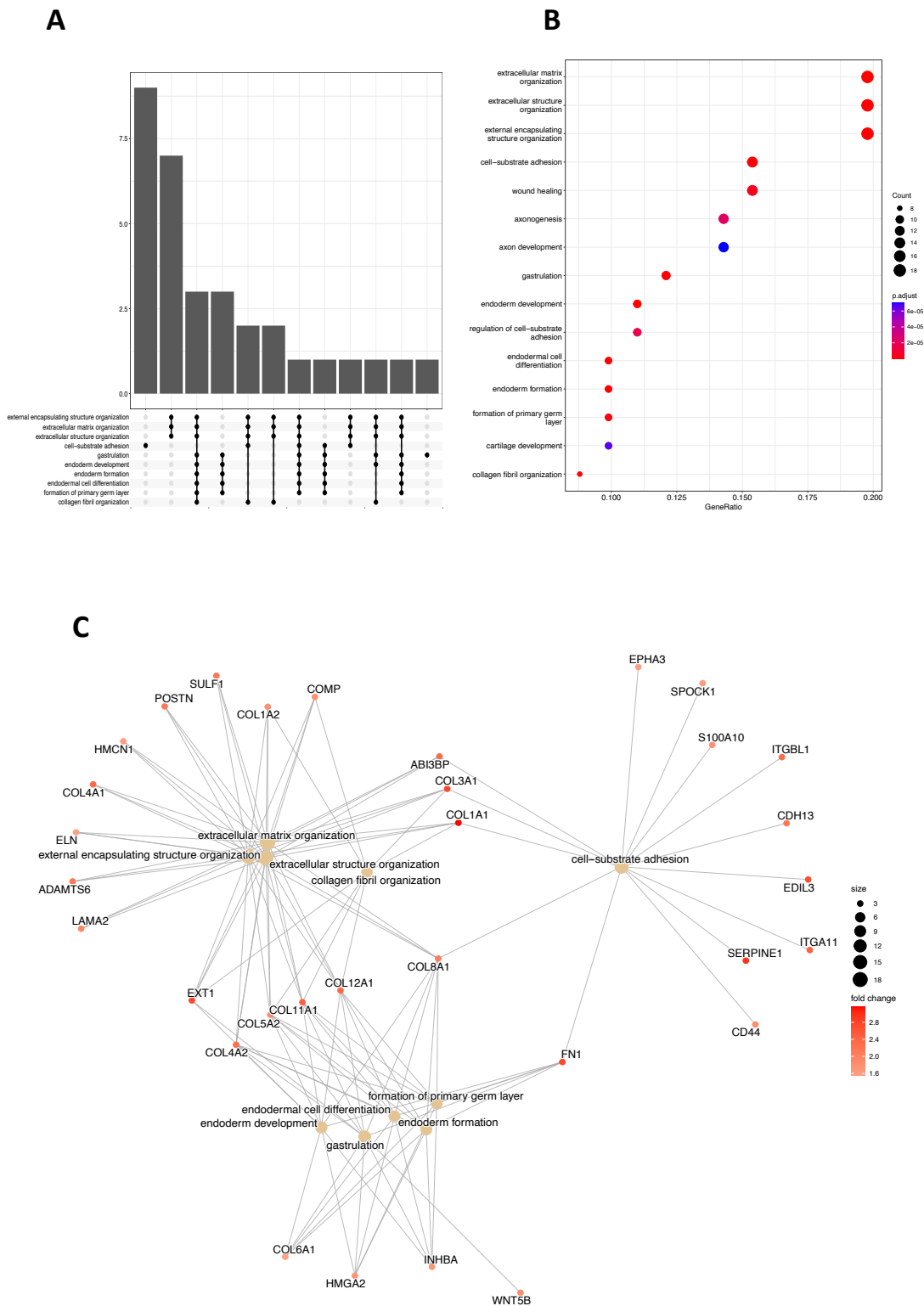


Figure 55. Enrichment analysis of Cluster 7. **A)** Barplot of the correlation of functions. **B)** Dot Plot of the significant functions. **C)** Network plot of the genes associated with the functions.

3. DISCUSSION

In light of the results from Chapter 3, we decided to perform a single-cell RNA sequencing to compare the control MSCs and poly(I:C) 25µg/ml stimulation at the time of 6 hours. These data were selected to determine potential subpopulations associated with each condition and their behavior when integrated. The decision to use the time of 6 hours of stimulation was driven by the substantial information available at this time in advance of the later time of 24 hours.

The initial analysis included both stimulations (Global analysis), but results due to the high expression of some genes could mask other relevant genes with lower expression that are important in our comparison. The major results within different clusters were associated with functions related with viral response and innate immunity, cytoplasmatic mitochondrion, aerobic respiration, cell division, muscle protein and stress fibers.

To increase the precision of our results, we performed two independent analyses in function of the control and stimulated poly(I:C) MSCs. The control MSCs were separated into three different clusters. The largest cluster, cluster 0, contains relatively unspecific genes related to muscle proteins, fibrin filaments and cytoskeleton. Conversely, the genes highlighted in the functional enrichment for cluster 0, such as Tropomyosin (TPM) family genes or myogenic genes, showed higher expression levels in the cluster 1. It's worth noting that when using the *Seurat* Clusters ratio, each cluster is compared to all other clusters to identify potential marker genes, and given the low expression levels of cluster 0 (**Figure 44**), this may cause discrepancies between the ratios and the absolute expression of the genes ([Hao et al., 2021](#)).

Cluster 1 appeared to overlap with the other clusters, this cluster exhibited potential expression in TPM family genes: TPM1, TPM2 and TPM4 and Myogenic family genes: MYL6, MYL12, MYL12B; and TAGLN gene. Tropomyosins, highly conserved genes, play indispensable roles in a variety of processes including cytokinesis, cell motility, cell transformation and more specialized functions such as myofibrillar contraction. They are present in both muscle and non-muscle cells. In striated muscle cells, tropomyosin contributes to the role of actin and myosin in contraction. In non-muscle cells, tropomyosin regulates actin filament organization ([Pasquet et al., 2006](#)). TPM1 and TPM2 in mesenchymal stem cells are secreted in response to hypoxia and have the potential to aid in the recovery of the myocardial damage. Studies suggest that these proteins may be located in extracellular vesicles or exosomes ([Song et al., 2016](#)). Meanwhile, TPM4 is implicated as an oncogenic gene in various malignancies, associated with EMT by altering the actin cytoskeleton ([J. Wang et al., 2022](#)). In addition, TAGLN, one of the most prominent MSC marker genes that we revealed in previous Chapters 1 and 3, was also involved in the muscle functions.

In contrast, cluster 2 exhibited strong associations with Collagen family genes such as: COL8A1, COL4A1, COL4A2, COL5A2, COL12A1 and COL11A1. COL4A1 and COL4A2, which we highlighted in Chapters 1 and 3, are relevant to improving outcomes in bone marrow transplantation using MSCs ([H.-J. Lee et al., 2021](#)). The COL12A1 gene, which encodes proteins, plays a structural role in the perichondrocytic matrix and at the bone-cartilage interface ([Perrier et al., 2011](#)), whereas COL5A2 is critical for osteogenic differentiation ([Yang et al., 2018](#)). These results allow us to classify of the control MSCs into two main groups related to muscle functions or osteogenic functions.

Moving to the analysis of poly(I:C) stimulation of MSCs, it contained 5 clusters. Cluster 0, the largest cluster, contained a typical inflammatory gene pattern, consistent with the observations in Chapter 4. Several chemokines, cytokines and interleukins were involved, such as: CXCL9, CCL8, IL6, TNFAIP6 and TNFPSDF10 and CXCL1. ICAM1 gene has also been reported in studies of stimulated poly(I:C) MSCs, which is associated with MSC migration ([S. H. Kim et al., 2019](#)).

Cluster 0 displayed overexpression in TAGLN, MYL6, TPM1, TPM2 and TPM4, corresponding to cluster 1 in control MSCs. In addition, this gene pattern was also overexpressed in cluster 5 of the Global analysis. Cluster 1 of stimulated MSCs presented typical genes associated with osteogenic differentiation, such as RUNX2. RUNX2 is also a master transcription factor of osteoblast differentiation, the up expression of RUNX2 occurs in undifferentiated MSCs, committed pre-osteoblasts in osteogenic fronts and developing membranous parietal bones (W.-J. Kim et al., 2020). Moreover, ZEB1 and ZEB2 genes were up expressed, which are known for their roles in EMT, as mentioned in Chapter 3. Furthermore, these genes are also associated with a wide variety of immune cells of both myeloid and lymphoid lineages, including dendritic cells, macrophages, monocytes, B, T, and NK cells. In these cells, ZEB1 and ZEB2 regulate important transcriptional networks associated with cell differentiation, maintenance, and function (Scott & Omilusik, 2019). Similarly, APOL6, PARP14 and FBXL7 were genes associated with adipogenic differentiation, analyzed in the adipogenic differentiation in Chapter 3, which were also up expressed in this cluster (Bengoechea-Alonso & Ericsson, 2010; X. Luo et al., 2017; Y. Tan et al., 2019).

Cluster 2, similar to Control cluster 2, displayed up expression of Collagen family genes like COL6A3, COL4A1, COL4A2, COL12A1 and COL11A1, indicating its robust potential in differentiation.

Cluster 3, appeared as a mixture of cluster 2 and cluster 0, without a clear gene pattern association. Nevertheless, there were genes closely related to cluster 2 in overall expression. SULF1 and SMAD4A are directly involved in the growth and differentiation of MSCs (Park et al., 2019; Zaman et al., 2016). These results reinforce the idea of the proximity between cluster 2 and 3, categorizing cluster 3 with low effect in the stimulation of poly(I:C) at time 6 hours.

Finally, cluster 4 was the most distinct cluster, with genes that exclusively present in this group. TPX2, CENPW, TK1, CCNB1 genes were involved in mitosis and cell division. Studies reflect the idea that CENP-W, identified as a centromeric component, plays a pivotal role in the cell cycle and is associated with the proliferation of cancer stem cell-like in hepatocellular carcinoma (Zhou et al., 2020). Notably, CCNB1, was also involved as a mitotic cyclin and there are studies that are using it as a prognostic factor for breast cancer death (P. Qiu et al., 2021). This cluster corresponded to the same cluster 4 of the Global analysis.

The independent analyses allowed for finer correlation between clusters from the independent and Global analyses, resulting in more accurate categorization. The newly defined subpopulations were represented in **Figure 51**. Notably, the subpopulation closely associated with the poly(I:C) stimulation was labeled as "Stimulated MSCs," known for their inflammatory profile. The pre-differentiated MSCs subpopulation presented all genes associated with adipogenic and osteogenic differentiation, as well as Collagen family genes from cluster 2 in Control MSCs analysis; this cluster presented a mixture of both cells. The low stimulated MSCs cluster, although fully defined as the stimulated MSC cluster, did not have the same characteristics. The cells of this cluster did not have the inflammatory profile, but were in concordance with the pre-differentiated MSCs. Muscle cells were clearly represented in both independent analyses as clusters associated with myogenic and tropomyogenic gene families whose functions were associated with muscle functions. Moreover, these genes were associated for their effects on skeletal myogenic differentiation potential, immune antigen expression, mixed lymphocytic responses and modulation of the immune response by regulatory T cells (Joo et al., 2014). The growth MSCs subpopulation displayed a gene pattern in stimulated MSCs associated with cell cycle and mitotic processes. This cluster was not represented in Control MSCs due to the low aggrupation of clusters in the control independent analysis. Finally, the subgroup of MSCs mainly composed of Control MSCs, lacked specific functions beyond their distribution in the Global analysis, being associated with cytoplasmic transition and stress fibers. In this point, future studies could focus on refining this last MSC cluster and increasing the robustness by including more samples and cells. This aspect, highlighted in this section of

Chapter 4, introduces a new focus in the study of poly(I:C) stimulation within emerging cell therapies and immunological research, as the projection of inflammatory properties within a more precisely defined subgroup of MSCs may increase the efficacy of such therapies. Additionally, the identification of new subgroups of MSCs within both stimulated and unstimulated MSCs enriches our understanding of MSC characterization. This process not only advances our understanding of the complex heterogeneity within MSCs subpopulations, but also facilitates the identification of the most promising MSC subgroups for further investigation. This broader perspective sheds light on a trajectory toward more precise research undertakings. Thus, it contributes significantly to the advancement of cell-based therapeutic strategies and immunological investigations.

The last part of this chapter dealt with the integration of *in vitro* and *in vivo* human bone marrow MSCs obtained from different healthy donors. As we previously mentioned, this collaborative project was carried out in partnership with the Department of Oncohematology at the University and University Hospital of Zurich, during my stage there. The main goal of this project was to merge and correlate the *in vivo* subgroups of MSCs: CAR cells, osteogenic cells and myogenic cells with our *in vitro* MSCs data.

From the *in vivo* MSCs data, we generated two new subgroups designated as CAR1 and CAR2. Remarkably, when integrated with our *in vitro* MSCs, the majority of cells were clustered closely with CAR2, while a smaller group of cells were integrated with the same cluster of CAR 1 cells.

Previous research from Cesar Nombela's laboratory identified CAR cells as the widely studied stromal subset comprising CXCL12 abundant reticular cells (CAR). These cells are ubiquitous in adult bone marrow and express high levels of the stem cell factor, interleukin-7 (IL-7). They have been implicated in the regulation of hematopoietic stem cells, lymphopoiesis and myelopoiesis. Studies in mice indicated the presence of both osteo- and adipo-primed progenitors in CAR cells, which are overrepresented in biological processes related to bone morphogenesis, mineralization and endochondral ossification (Helbling et al., 2019). Furthermore, other studies categorize the CAR cells into osteogenic CAR cells and adipogenic CAR cells in a transcriptional map of the bone marrow stromal niche, according to their transcriptional profile (Baccin et al., 2020).

Our results showed that the correlation of CAR 2 cells and the majority of *in vitro* cells was associated with the regulation of hematopoiesis, differentiation and an osteogenic profile. Interestingly, we were able to observe genes such as CXCL2 or TNF family genes, MYL9, TPM1, TIMP1, TPM4, COL1A1 and TAGLN that were identified in our subgroups as Muscle MSCs in the Control MSCs analysis and in pre-differentiated MSCs in our Global analysis that included control MSCs. Another study by J Mollentze and colleagues examined the osteogenic differentiation of mice MSCs in *in vivo* and *in vitro* and revealed similarities in their osteogenic differentiation due to the up expression of RUNX2 (Mollentze et al., 2021). Conversely, there is currently no existing literature on the relationship between tropomyogenic genes and *in vivo* and *in vitro* MSCs. In addition, the relationship of TAGLN between Chapters 1, 3 and this Chapter 4 is relevant as a possible new marker of a subpopulation in MSCs.

The second association was more robust due to the classification in the same cluster. The key functions were associated with extracellular matrix and cell-substrate adhesion. In particular, Collagen family genes such as COL12A1, COL11A1, COL5A2, COL4A2, COL4A1, COL8A1; gain relevant importance in this cluster, reflecting their presence in the second cluster of *in vitro* MSCs analysis. It is known that Collagen family genes promote higher adhesion, survival and proliferation of MSCs (Somaiah et al., 2015), although this specific subpopulation observed in *in vitro* and *in vivo* human MSCs has not been previously investigated. Interestingly, COL4A1 and COL4A2 were also present in Chapters 1 and 3 as possible novel markers in MSCs, which are present in a different subpopulation than TAGLN in MSCs clustering, revealing that these three genes could accurately identify different subpopulations in MSCs. However, further

experimental studies are needed to confirm this hypothesis.

In general, this study highlights the existence of two main groups within *in vitro* MSCs that share commonalities with human *in vivo* BM-MS-CAR cells. These two major groups are distinguished by a pronounced association of inflammatory and differentiation processes in the majority of *in vitro* cells, compared to the smaller group of *in vitro* MSCs, associated with extracellular matrix construction and cell adhesion of MSCs.

These novel findings introduce a new approach in the advancement of cell therapy research, potentially uncovering new gene markers, such as TAGLN, COL4A1 and COL4A2, that maintain their characteristics in both *in vitro* and *in vivo* contexts. Moreover, the categorization of the new human bone marrow MSCs subgroups based on specific capabilities allows us continue on further studies aimed at exploring the similarities and origins of human bone marrow Mesenchymal Stem Cell characterization.

4. FINAL SUMMARY of CHAPTER 4: *Single-Cell transcriptomics of MSCs*

Based on the results and observations obtained in Chapter 3, we embarked on a series of studies on MSCs using single-cell RNA sequencing technology. Thus, we performed a comparison of control MSCs with MSCs subjected to stimulation with poly(I:C) at 25 µg/ml for a duration of 6 hours (the choice of this time point was guided by the availability of other data and by the results of our previous study with bulk RNA-seq). Single cells were isolated using the 10xGenomics technology, followed by RNA sequencing using Illumina NextSeq platform.

Our initial analysis included both stimulations together, which is a global analysis including all samples. However, due to the smaller number of cells sequenced in some experiments, the global approach could hide or obscure low-expression genes that we expect to be significant in our samples. Nevertheless, the analysis of the main signal of the global change provided functions related to viral response and innate immunity, cytoplasmic mitochondria, aerobic respiration, cell division, muscle protein, and stress fibers.

In a second approach, we performed two independent analyses of the scRNA-seq data for control MSCs and stimulated MSCs to obtain a more precise analysis. Control MSCs clustered into three distinct cell clusters. The largest, cluster 0 of controls, contained genes related to muscle proteins, fibrin filaments and the cytoskeleton. Cluster 1 of controls, showed higher expression of genes associated with TPM family genes, myogenic genes, and TAGLN. Cluster 2 of controls was characterized by strong associations with collagen family genes, including COL4A1 and COL4A2. In the analysis of poly(I:C) stimulated MSCs, we identified five distinct cell clusters. Cluster 0 showed a typical inflammatory gene pattern with chemokines, cytokines, interleukins and ICAM1. It also showed overexpression of genes such as TAGLN, MYL6, TPM1, TPM2, and TPM4, similar to Cluster_c1 in control MSCs. Cluster_1 showed genes associated with osteogenic differentiation, including RUNX2. Cluster 2 showed upregulation of collagen family genes, including COL4A1 and COL4A2 again. Cluster 3 did not show a distinct gene pattern, but had similarities to cluster 2. Cluster 4 was the most distinct, with genes associated with mitosis and cell division, such as CENPW or CCNB1.

These independent analyses allowed a precise categorization of the different cell populations. Thus, the subpopulations found were classified as follows: (i) "stimulated MSCs", closely associated with poly(I: C) stimulation and characterized by an inflammatory profile; (ii) "pre-differentiated MSCs", over-represented in genes related to adipogenic and osteogenic differentiation, with a high presence of collagen family genes such as COL4A1 and COL4A2; (iii) "low-stimulated MSCs", which lack the inflammatory profile but show similarities to pre-differentiated MSCs; (iv) "muscle cells", associated with myogenic and TPM gene families, as well as muscle-related functions and the presence of TAGLN; (v) "growth MSCs", characterized by genes involved in cell cycle and mitotic processes.

In addition, we performed scRNA-seq data integration of bone marrow MSCs isolated *in vitro* (i.e., the standard primary MSCs isolated in the laboratory after growth in culture up to 4 passages) with MSCs obtained directly *in vivo* from human bone marrow biopsies (samples and data generated by the Zurich research group). The *in vivo* analysis of MSCs from different healthy donors allowed the identification of two cell subsets: CAR 1 and CAR 2 (defined in the Zurich phase). When we integrated our *in vitro* MSC data with the *in vivo* MSC data, the majority of the *in vitro* cells closely aligned with CAR 2, emphasizing an osteogenic character with a differentiation signal toward this lineage. A smaller set of *in vitro* cells were integrated with CAR 1 cells. These results indicated a significant association between CAR 2 cells and MSCs *in vitro*, with functions related to the osteogenic profile, wound healing and also regulation of homeostasis. In this context, an association with TPM genes and TAGLN was also identified.

The integration also revealed the similarity of CAR 1 cells and a small group of MSCs *in vitro*, presenting an enrichment in functions related to extracellular matrix and cell-substrate adhesion, with emphasis on genes of the collagen family, such as COL4A1 and COL4A2.

In summary, our studies in this chapter have revealed two primary subsets of *in vitro* MSCs that resemble *in vivo* human bone marrow-derived MSCs. The majority of *in vitro* MSCs are aligned with CAR 2, which is associated with inflammation and differentiation, while some *in vitro* MSCs are aligned with CAR 1, which is associated with extracellular matrix and adhesion. These results provide a significant advance in the exploration of specific gene markers across the *in vitro* and *in vivo* context of MSCs, improving their molecular characterization and their management and use for cell therapy.

CONCLUSIONS

In the light of the scientific and research objectives proposed in this Doctoral Thesis, and after discussing and integrating all our results, we can draw the following main CONCLUSIONS:

1. We collected, integrated and analyzed an extensive compendium of gene expression data from human mesenchymal stem cells, hematopoietic stem cells and fibroblasts generated by different transcriptomic platforms. Using all these data, we were able to identify four novel gene markers specific to hMSCs: COL4A1, COL4A2, TAGLN and SCUBE3; which outperform the CD markers currently used to isolate these stem cells. These new markers are particularly accurate in their ability to separate and avoid confusion with fibroblasts. These data were corroborated by experimental analyses of gene expression in different cell types using quantitative PCR.
2. For the second objective, we analyzed a collection of expression and methylation data of different cell types associated with bone marrow-derived hMSCs. Using this data, we generated an original gene coexpression network centered on the activity of hMSCs, followed by a gene regulatory network that allowed the identification of Master Regulators and Regulons. The most significant Master Regulators found to upregulate the expression of the generated network were: SNAI2, SATB2, IRX3, EPAS1, HOXC6.
3. For the third objective, we performed a study of MSCs subjected to different stimulations —with LPS and poly(I:C)— and analyzed their changes in the transcriptome using RNA-seq. The stimulation with poly(I:C) was done at various time points. In addition, we performed adipogenic and osteogenic differentiation experiments using BM-MSCs. The combination of these experimental studies followed by bioinformatic and statistical analyses allow us to identify gene signatures and biological processes associated with MSCs stimulation.
4. A set of MSCs samples isolated in vitro from healthy donors and stimulated with poly(I:C) were analyzed using single-cell RNA-seq. Different cell clusters were delineated in the samples, revealing distinct cell populations within both control and stimulated MSCs. These experiments demonstrated the feasibility and power of using single-cell data, both independently and synergistically with global bulk expression data, to better identify key gene markers and distinct cellular subpopulations.
5. For the last objective, we were able to harmoniously integrate MSCs isolated in vitro with MSCs obtained in vivo from fresh bone marrow biopsies (both from healthy donors and analyzed by single-cell RNA-seq). We distinguished two subgroups of cells in the in vitro MSCs, that were closely associated with two distinct subpopulations of the in vivo MSCs (CAR1 and CAR2). Functional enrichment analyses of these subpopulations were performed, and we were able to identify in the in vivo MSCs some of the novel gene markers that we found in our initial study with in vitro MSCs: TAGLN (Transgelin) associated with CAR2 cells; and COL4A1 and COL4A2 associated with CAR1 cells. The work highlighted the potential to identify common gene expression patterns between the in vivo and in vitro stem cell subpopulations.

BIBLIOGRAPHY

- Afrache, H., Gouret, P., Ainouche, S., Pontarotti, P., & Olive, D. (2012). **The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response.** *Immunogenetics*, 64(11), 781–794. <https://doi.org/10.1007/s00251-012-0619-z>
- Almalki, S. G., & Agrawal, D. K. (2016). **Key transcription factors in the differentiation of mesenchymal stem cells.** *Differentiation*, 92(1–2), 41–51. <https://doi.org/10.1016/j.diff.2016.02.005>
- Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., & Califano, A. (2016). **Functional characterization of somatic mutations in cancer using network-based inference of protein activity.** *Nature Genetics*, 48(8), 838–847. <https://doi.org/10.1038/ng.3593>
- Anasagasti, A., Irigoyen, C., Barandika, O., López de Munain, A., & Ruiz-Ederra, J. (2012). **Current mutation discovery approaches in Retinitis Pigmentosa.** *Vision Research*, 75, 117–129. <https://doi.org/10.1016/j.visres.2012.09.012>
- Aranda, P., Agirre, X., Ballestar, E., Andreu, E. J., Román-Gómez, J., Prieto, I., Martín-Subero, J. I., Cigudosa, J. C., Siebert, R., Esteller, M., & Prosper, F. (2009). **Epigenetic Signatures Associated with Different Levels of Differentiation Potential in Human Stem Cells.** *PLoS ONE*, 4(11), e7809. <https://doi.org/10.1371/journal.pone.0007809>
- Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.** *Bioinformatics*, 30(10), 1363–1369. <https://doi.org/10.1093/bioinformatics/btu049>
- Baccin, C., Al-Sabah, J., Velten, L., Helbling, P. M., Grünschläger, F., Hernández-Malmierca, P., Nombela-Arrieta, C., Steinmetz, L. M., Trumpp, A., & Haas, S. (2020). **Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization.** *Nature Cell Biology*, 22(1), 38–48. <https://doi.org/10.1038/s41556-019-0439-6>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2012). **NCBI GEO: archive for functional genomics data sets-update.** *Nucleic Acids Research*, 41(D1), D991–D995. <https://doi.org/10.1093/nar/gks1193>
- Becker-Herman, S., Rozenberg, M., Hillel-Karniel, C., Gil-Yarom, N., Kramer, M. P., Barak, A., Sever, L., David, K., Radomir, L., Lewinsky, H., Levi, M., Friedlander, G., Bucala, R., Peled, A., & Shachar, I. (2021). **CD74 is a regulator of hematopoietic stem cell maintenance.** *PLoS Biology*, 19(3), e3001121. <https://doi.org/10.1371/journal.pbio.3001121>
- Bengoechea-Alonso, M. T., & Ericsson, J. (2010). **The ubiquitin ligase Fbxw7 controls adipocyte differentiation by targeting C/EBPalpha for degradation.** *Proceedings of the National Academy of Sciences of the U.S.A.*, 107(26), 11817–11822. <https://doi.org/10.1073/pnas.0913367107>
- Berebichez-Fridman, R., & Montero-Olvera, P. R. (2018). **Sources and Clinical Applications of Mesenchymal Stem Cells: State-of-the-art review.** *Sultan Qaboos University Medical Journal [SQUMJ]*, 18(3), 264. <https://doi.org/10.18295/squmj.2018.18.03.002>
- Bianco, P., Robey, P. G., & Simmons, P. J. (2008). **Mesenchymal Stem Cells: Revisiting History, Concepts, and Assays.** *Cell Stem Cell*, 2(4), 313–319. <https://doi.org/10.1016/j.stem.2008.03.002>

- Bielska, M., Borowiec, M., Jesionek-Kupnicka, D., Braun, M., Bojo, M., Pastorczak, A., Kalinka-Warzocho, E., Prochorec-Sobieszek, M., Robak, T., Warzocho, K., Młynarski, W., & Lech-Marańda, E. (2017). **Polymorphism in IKZF1 gene affects clinical outcome in diffuse large B-cell lymphoma.** *International Journal of Hematology*, 106, 794–800. <https://doi.org/10.1007/s12185-017-2315-0>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Buccitelli, C., & Selbach, M. (2020). **mRNAs, proteins and the emerging principles of gene expression control.** *Nature Reviews Genetics*, 21(10), 630–644. <https://doi.org/10.1038/s41576-020-0258-4>
- Canalis, E., Parker, K., & Zanotti, S. (2012). **Gremlin1 is required for skeletal development and postnatal skeletal homeostasis.** *Journal of Cellular Physiology*, 227(1), 269–277. <https://doi.org/10.1002/jcp.22730>
- Caplan, A. (1991). **Mesenchymal stem cells.** *Journal of Orthopaedic Research*, 9(5), 641-650. <https://doi.org/10.1002/jor.1100090504>
- Carpenter, S., Wochal, P., Dunne, A., & O'Neill, L. A. J. (2011). **Toll-like Receptor 3 (TLR3) Signaling Requires TLR4 Interactor with Leucine-rich Repeats (TRIL).** *Journal of Biological Chemistry*, 286(44), 38795–38804. <https://doi.org/10.1074/jbc.M111.255893>
- Cassano, J. M., Schnabel, L. V., Goodale, M. B., & Fortier, L. A. (2018). **Inflammatory licensed equine MSCs are chondroprotective and exhibit enhanced immunomodulation in an inflammatory environment.** *Stem Cell Research & Therapy*, 9(1), 82. <https://doi.org/10.1186/s13287-018-0840-2>
- Chang, J., Liu, F., Lee, M., Wu, B., Ting, K., Zara, J. N., Soo, C., Al Hezaimi, K., Zou, W., Chen, X., Mooney, D. J., & Wang, C.-Y. (2013). **NF- κ B inhibits osteogenic differentiation of mesenchymal stem cells by promoting β -catenin degradation.** *Proceedings of the National Academy of Sciences of the U.S.A.*, 110(23), 9469–9474. <https://doi.org/10.1073/pnas.1300532110>
- Chen, X., Zhang, Z., Zhou, H., & Zhou, G. (2014). **Characterization of mesenchymal stem cells under the stimulation of Toll-like receptor agonists.** *Development, Growth & Differentiation*, 56(3), 233–244. <https://doi.org/10.1111/dgd.12124>
- Chen, Y., Ma, X., Zhang, M., Wang, X., Wang, C., Wang, H., Guo, P., Yuan, W., Rudolph, K. L., Zhan, Q., & Ju, Z. (2014). **Gadd45a regulates hematopoietic stem cell stress responses in mice.** *Blood*, 123(6), 851–862. <https://doi.org/10.1182/blood-2013-05-504084>
- Chen, Y., Wang, K., Qian, C.-N., & Leach, R. (2013). **DNA methylation is associated with transcription of Snail and Slug genes.** *Biochemical and Biophysical Research Communications*, 430(3), 1083–1090. <https://doi.org/10.1016/j.bbrc.2012.12.034>
- Cheng, J., Li, Y., Liu, S., Jiang, Y., Ma, J., Wan, L., Li, Q., & Pang, T. (2019). **CXCL8 derived from mesenchymal stromal cells supports survival and proliferation of acute myeloid leukemia cells through the PI3K/AKT pathway.** *The FASEB Journal*, 33(4), 4755–4764. <https://doi.org/10.1096/fj.201801931R>
- Choi, U. Y., Kang, J.-S., Hwang, Y. S., & Kim, Y.-J. (2015). **Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases.** *Experimental & Molecular Medicine*, 47(3), e144–e144. <https://doi.org/10.1038/emm.2014.110>
- Chomczynski, P., & Sacchi, N. (2006). **The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on.** *Nature Protocols*, 1(2), 581–585. <https://doi.org/10.1038/nprot.2006.83>

- Cohen, M. S. (2020). **Interplay between compartmentalized NAD⁺ synthesis and consumption: a focus on the PARP family.** *Genes & Development*, 34(5–6), 254–262. <https://doi.org/10.1101/gad.335109.119>
- Corchete, L. A., Rojas, E. A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N. C., & Burguillo, F. J. (2020). **Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis.** *Scientific Reports*, 10(1), 19737. <https://doi.org/10.1038/s41598-020-76881-x>
- Crippa, S., & Bernardo, M. E. (2018). **Mesenchymal Stromal Cells: Role in the BM Niche and in the Support of Hematopoietic Stem Cell Transplantation.** *HemaSphere*, 2(6), e151. <https://doi.org/10.1097/HS9.000000000000151>
- Cuesta-Gomez, N., Graham, G. J., & Campbell, J. D. M. (2021). **Chemokines and their receptors: predictors of the therapeutic potential of mesenchymal stromal cells.** *Journal of Translational Medicine*, 19(1), 156. <https://doi.org/10.1186/s12967-021-02822-5>
- CUI, X., CHEN, L., XUE, T., YU, J., LIU, J., JI, Y., & CHENG, L. (2015). **Human umbilical cord and dental pulp-derived mesenchymal stem cells: Biological characteristics and potential roles in vitro and in vivo.** *Molecular Medicine Reports*, 11(5), 3269–3278. <https://doi.org/10.3892/mmr.2015.3198>
- de Araújo Farias, V., Carrillo-Gálvez, A. B., Martín, F., & Anderson, P. (2018). **TGF- β and mesenchymal stromal cells in regenerative medicine, autoimmunity and cancer.** *Cytokine & Growth Factor Reviews*, 43, 25–37. <https://doi.org/10.1016/j.cytogfr.2018.06.002>
- de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N., & Dzierzak, E. (2013). **Gata2 is required for HSC generation and survival.** *Journal of Experimental Medicine*, 210(13), 2843–2850. <https://doi.org/10.1084/jem.20130751>
- DelaRosa, O., & Lombardo, E. (2010). **Modulation of Adult Mesenchymal Stem Cells Activity by Toll-Like Receptors: Implications on Therapeutic Potential.** *Mediators of Inflammation*, 2010, 1–9. <https://doi.org/10.1155/2010/865601>
- Diamond, M. S., & Farzan, M. (2013). **The broad-spectrum antiviral functions of IFIT and IFITM proteins.** *Nature Reviews Immunology*, 13(1), 46–57. <https://doi.org/10.1038/nri3344>
- Djouad, F., Ipseiz, N., Luz-Crawford, P., Scholtysek, C., Krönke, G., & Jorgensen, C. (2017). **PPAR β / δ : A master regulator of mesenchymal stem cell functions.** *Biochimie*, 136, 55–58. <https://doi.org/10.1016/j.biochi.2016.11.011>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Doherty, M. R., Cheon, H., Junk, D. J., Vinayak, S., Varadan, V., Telli, M. L., Ford, J. M., Stark, G. R., & Jackson, M. W. (2017). **Interferon-beta represses cancer stem cell properties in triple-negative breast cancer.** *Proceedings of the National Academy of Sciences of the U.S.A.*, 114(52), 13792–13797. <https://doi.org/10.1073/pnas.1713728114>
- Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F. C., Krause, D. S., Deans, R. J., Keating, A., Prockop, D. J., & Horwitz, E. M. (2006). **Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement.** *Cytotherapy*, 8(4), 315–317. <https://doi.org/10.1080/14653240600855905>

- Döppler, H., Panayiotou, R., Reid, E. M., Maimo, W., Bastea, L., & Storz, P. (2016). **The PRKD1 promoter is a target of the KRas-NF- κ B pathway in pancreatic cancer.** *Scientific Reports*, 6, 33758. <https://doi.org/10.1038/srep33758>
- D'Urso, M., & Kurniawan, N. A. (2020). **Mechanical and Physical Regulation of Fibroblast–Myofibroblast Transition: From Cellular Mechanoresponse to Tissue Pathology.** *Frontiers in Bioengineering and Biotechnology*, 8. <https://doi.org/10.3389/fbioe.2020.609653>
- Elsafadi, M., Manikandan, M., Dawud, R. A., Alajez, N. M., Hamam, R., Alfayez, M., Kassem, M., Aldahmash, A., & Mahmood, A. (2016). **Transgelin is a TGF β -inducible gene that regulates osteoblastic and adipogenic differentiation of human skeletal stem cells through actin cytoskeleton organization.** *Cell Death & Disease*, 7(8), e2321–e2321. <https://doi.org/10.1038/cddis.2016.196>
- Eskandari, F., Zolfaghari, S., Yazdanpanah, A., Shabestari, R. M., Fomeshi, M. R., Milan, P. B., Kiani, J., Zomorrod, M. S., & Safa, M. (2023). **TLR3 stimulation improves the migratory potency of adipose-derived mesenchymal stem cells through the stress response pathway in the melanoma mouse model.** *Molecular Biology Reports*, 50, 2293–2304. <https://doi.org/10.1007/s11033-022-08111-8>
- Espósito, A. C. C., Brianezi, G., Miot, L. D. B., & Miot, H. A. (2022). **Fibroblast morphology, growth rate and gene expression in facial melasma.** *Anais Brasileiros de Dermatologia*, 97(5), 575–582. <https://doi.org/10.1016/j.abd.2021.09.012>
- Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigartyo, C. A.-K., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., ... Uhlén, M. (2014). **Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics.** *Molecular & Cellular Proteomics*, 13(2), 397–406. <https://doi.org/10.1074/mcp.M113.035600>
- Feng, Y., Liao, Y., Huang, W., Lai, X., Luo, J., Du, C., Lin, J., Zhang, Z., Qiu, D., Liu, Q., Shen, H., Xiang, A. P., & Zhang, Q. (2018). **Mesenchymal stromal cells-derived matrix Gla protein contribute to the alleviation of experimental colitis.** *Cell Death & Disease*, 9(6), 691. <https://doi.org/10.1038/s41419-018-0734-3>
- Florian, M. C., Dörr, K., Niebel, A., Daria, D., Schrezenmeier, H., Rojewski, M., Filippi, M.-D., Hasenberg, A., Gunzer, M., Scharffetter-Kochanek, K., Zheng, Y., & Geiger, H. (2012). **Cdc42 Activity Regulates Hematopoietic Stem Cell Aging and Rejuvenation.** *Cell Stem Cell*, 10(5), 520–530. <https://doi.org/10.1016/j.stem.2012.04.007>
- Fontanillo, C., Nogales-Cadenas, R., Pascual-Montano, A., & De las Rivas, J. (2011). **Functional analysis beyond enrichment: non-redundant reciprocal linkage of genes and biological terms.** *PLoS One*, 6(9), e24289. <https://doi.org/10.1371/journal.pone.0024289>
- Fröbel, J., Landspersky, T., Percin, G., Schreck, C., Rahmig, S., Ori, A., Nowak, D., Essers, M., Waskow, C., & Oostendorp, R. A. J. (2021). **The Hematopoietic Bone Marrow Niche Ecosystem.** *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.705410>
- Galle, E., Thienpont, B., Cappuyens, S., Venken, T., Busschaert, P., Van Haele, M., Van Cutsem, E., Roskams, T., van Pelt, J., Verslype, C., Dekervel, J., & Lambrechts, D. (2020). **DNA methylation-driven EMT is a common mechanism of resistance to various therapeutic agents in cancer.** *Clinical Epigenetics*, 12(1), 27. <https://doi.org/10.1186/s13148-020-0821-z>

- Ganesan, R., Mallets, E., & Gomez-Cambronero, J. (2016). **The transcription factors Slug (SNAI2) and Snail (SNAI1) regulate phospholipase D (PLD) promoter in opposite ways towards cancer cell invasion.** *Molecular Oncology*, 10(5), 663–676. <https://doi.org/10.1016/j.molonc.2015.12.006>
- Gao, J., Chen, Y.-H., & Peterson, L. C. (2015). **GATA family transcriptional factors: emerging suspects in hematologic disorders.** *Experimental Hematology & Oncology*, 4(1), 28. <https://doi.org/10.1186/s40164-015-0024-z>
- Garcia-Gomez, A., Li, T., de la Calle-Fabregat, C., Rodríguez-Ubreva, J., Ciudad, L., Català-Moll, F., Godoy-Tena, G., Martín-Sánchez, M., San-Segundo, L., Muntión, S., Morales, X., Ortiz-de-Solórzano, C., Oyarzabal, J., San José-Enériz, E., Esteller, M., Agirre, X., Prosper, F., Garayoa, M., & Ballestar, E. (2021). **Targeting aberrant DNA methylation in mesenchymal stromal cells as a treatment for myeloma bone disease.** *Nature Communications*, 12(1), 421. <https://doi.org/10.1038/s41467-020-20715-x>
- Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). **affy—analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics*, 20(3), 307–315. <https://doi.org/10.1093/bioinformatics/btg405>
- Gentile, P., Calabrese, C., De Angelis, B., Pizzicannella, J., Kothari, A., & Garcovich, S. (2019). **Impact of the Different Preparation Methods to Obtain Human Adipose-Derived Stromal Vascular Fraction Cells (AD-SVFs) and Human Adipose-Derived Mesenchymal Stem Cells (AD-MSCs): Enzymatic Digestion Versus Mechanical Centrifugation.** *International Journal of Molecular Sciences*, 20(21), 5471. <https://doi.org/10.3390/ijms20215471>
- Gharibi, B., Ghuman, M., & Hughes, F. J. (2016). **DDIT4 regulates mesenchymal stem cell fate by mediating between HIF1 α and mTOR signalling.** *Scientific Reports*, 6(1), 36889. <https://doi.org/10.1038/srep36889>
- Goeman, J. J., van de Geer, S. A., de Kort, F., & van Houwelingen, H. C. (2004). **A global test for groups of genes: testing association with a clinical outcome.** *Bioinformatics*, 20(1), 93–99. <https://doi.org/10.1093/bioinformatics/btg382>
- Govindarajan, R., Duraiyan, J., Kaliyappan, K., & Palanisamy, M. (2012). **Microarray and its applications.** *Journal of Pharmacy And Bioallied Sciences*, 4(6), 310. <https://doi.org/10.4103/0975-7406.100283>
- Granchi, D., Ochoa, G., Leonardi, E., Devescovi, V., Baglio, S. R., Osaba, L., Baldini, N., & Ciapetti, G. (2010). **Gene Expression Patterns Related to Osteogenic Differentiation of Bone Marrow-Derived Mesenchymal Stem Cells During Ex Vivo Expansion.** *Tissue Engineering Part C: Methods*, 16(3), 511–524. <https://doi.org/10.1089/ten.tec.2009.0405>
- Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., & Koltay, P. (2015). **Technologies for Single-Cell Isolation.** *International Journal of Molecular Sciences*, 16(8), 16897–16919. <https://doi.org/10.3390/ijms160816897>
- Guijarro-Muñoz, I., Compte, M., Álvarez-Cienfuegos, A., Álvarez-Vallina, L., & Sanz, L. (2014). **Lipopolysaccharide Activates Toll-like Receptor 4 (TLR4)-mediated NF- κ B Signaling Pathway and Proinflammatory Response in Human Pericytes.** *Journal of Biological Chemistry*, 289(4), 2457–2468. <https://doi.org/10.1074/jbc.M113.521161>
- Guinn, Z., Lampe, A. T., Brown, D. M., & Petro, T. M. (2016). **Significant role for IRF3 in both T cell and APC effector functions during T cell responses.** *Cellular Immunology*, 310, 141–149. <https://doi.org/10.1016/j.cellimm.2016.08.015>

- Haas, B. J., & Zody, M. C. (2010). **Advancing RNA-Seq analysis.** *Nature Biotechnology*, 28(5), 421–423. <https://doi.org/10.1038/nbt0510-421>
- Han, Q., Yang, P., Wu, Y., Meng, S., Sui, L., Zhang, L., Yu, L., Tang, Y., Jiang, H., Xuan, D., Kaplan, D. L., Kim, S. H., Tu, Q., & Chen, J. (2015). **Epigenetically Modified Bone Marrow Stromal Cells in Silk Scaffolds Promote Craniofacial Bone Repair and Wound Healing.** *Tissue Engineering Part A*, 21(15–16), 2156–2165. <https://doi.org/10.1089/ten.tea.2014.0484>
- Han, Y., Yang, J., Fang, J., Zhou, Y., Candi, E., Wang, J., Hua, D., Shao, C., & Shi, Y. (2022). **The secretion profile of mesenchymal stem cells and potential applications in treating human diseases.** *Signal Transduction and Targeted Therapy*, 7(1), 92. <https://doi.org/10.1038/s41392-022-00932-0>
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). **Integrated analysis of multimodal single-cell data.** *Cell*, 184(13), 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
- Harkness, L., Zaher, W., Ditzel, N., Isa, A., & Kassem, M. (2016). **CD146/MCAM defines functionality of human bone marrow stromal stem cell populations.** *Stem Cell Research & Therapy*, 7(1), 4. <https://doi.org/10.1186/s13287-015-0266-z>
- Haydont, V., Neiveyans, V., Perez, P., Busson, É., Lataillade, J.-J., Asselineau, D., & Fortunel, N. O. (2020). **Fibroblasts from the Human Skin Dermo-Hypodermal Junction are Distinct from Dermal Papillary and Reticular Fibroblasts and from Mesenchymal Stem Cells and Exhibit a Specific Molecular Profile Related to Extracellular Matrix Organization and Modeling.** *Cells*, 9(2), 368. <https://doi.org/10.3390/cells9020368>
- He, X., Wang, H., Jin, T., Xu, Y., Mei, L., & Yang, J. (2016). **TLR4 Activation Promotes Bone Marrow MSC Proliferation and Osteogenic Differentiation via Wnt3a and Wnt5a Signaling.** *PLoS ONE*, 11(3):e0149876. <https://doi.org/10.1371/journal.pone.0149876>
- Helbling, P. M., Piñeiro-Yáñez, E., Gerosa, R., Boettcher, S., Al-Shahrour, F., Manz, M. G., & Nombela-Arrieta, C. (2019). **Global Transcriptomic Profiling of the Bone Marrow Stromal Microenvironment during Postnatal Development, Aging, and Inflammation.** *Cell Reports*, 29(10), 3313–3330.e4. <https://doi.org/10.1016/j.celrep.2019.11.004>
- Hemmings, B. A., & Restuccia, D. F. (2012). **PI3K-PKB/Akt Pathway.** *Cold Spring Harbor Perspectives in Biology*, 4(9), a011189–a011189. <https://doi.org/10.1101/cshperspect.a011189>
- Hepburn, A. C., Steele, R. E., Veeratterapillay, R., Wilson, L., Kounatidou, E. E., Barnard, A., Berry, P., Cassidy, J. R., Moad, M., El-Sherif, A., Gaughan, L., Mills, I. G., Robson, C. N., & Heer, R. (2019). **The induction of core pluripotency master regulators in cancers defines poor clinical outcomes and treatment resistance.** *Oncogene*, 38(22), 4412–4424. <https://doi.org/10.1038/s41388-019-0712-y>
- Ho-Pun-Cheung, A., Bascoul-Mollevis, C., Assenat, E., Boissière-Michot, F., Bibeau, F., Cellier, D., Ychou, M., & Lopez-Crapez, E. (2009). **Reverse transcription-quantitative polymerase chain reaction: description of a RIN-based algorithm for accurate data normalization.** *BMC Molecular Biology*, 10(1), 31. <https://doi.org/10.1186/1471-2199-10-31>
- Hou, W., Duan, L., Huang, C., Li, X., Xu, X., Qin, P., Hong, N., Wang, D., & Jin, W. (2021). **Cross-Tissue Characterization of Heterogeneities of Mesenchymal Stem Cells and Their Differentiation Potentials.** *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.781021>

- Hoy, M. A. (2019). **DNA Amplification by the Polymerase Chain Reaction: Providing a Revolutionary Method for All Biologists.** In *Insect Molecular Genetics* (pp. 263–314). Elsevier. <https://doi.org/10.1016/B978-0-12-815230-0.00007-8>
- Hsieh, J.-Y., Fu, Y.-S., Chang, S.-J., Tsuang, Y.-H., & Wang, H.-W. (2010). **Functional Module Analysis Reveals Differential Osteogenic and Stemness Potentials in Human Mesenchymal Stem Cells from Bone Marrow and Wharton's Jelly of Umbilical Cord.** *Stem Cells and Development*, 19(12), 1895–1910. <https://doi.org/10.1089/scd.2009.0485>
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., ... Morgan, M. (2015). **Orchestrating high-throughput genomic analysis with Bioconductor.** *Nature Methods*, 12(2), 115–121. <https://doi.org/10.1038/nmeth.3252>
- Hwa Kim, S., Das, A., Choul Chai, J., Binas, B., Ran Choi, M., Sun Park, K., Seek Lee, Y., Hwa Jung, K., & Gyu Chai, Y. (2016). **Transcriptome sequencing wide functional analysis of human mesenchymal stem cells in response to TLR4 ligand OPEN.** *Nature Publishing Group*. <https://doi.org/10.1038/srep30311>
- Janky, R., Verfaillie, A., Imrichová, H., Van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Naval Sanchez, M., Potier, D., Svetlichnyy, D., Kalender Atak, Z., Fiers, M., Marine, J.-C., & Aerts, S. (2014). **iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections.** *PLoS Computational Biology*, 10(7), e1003731. <https://doi.org/10.1371/journal.pcbi.1003731>
- Javed, A., Karki, S., Sami, Z., Khan, Z., Shree, A., Sah, B. K., Ghosh, S., & Saxena, S. (2022). **Association between Mesenchymal Stem Cells and COVID-19 Therapy: Systematic Review and Current Trends.** *BioMed Research International*, 2022, 1–17. <https://doi.org/10.1155/2022/9346939>
- Jaworska, A. M., Włodarczyk, N. A., Mackiewicz, A., & Czerwinska, P. (2020). **The role of TRIM family proteins in the regulation of cancer stem cell self-renewal.** *Stem Cells*, 38(2), 165–173. <https://doi.org/10.1002/stem.3109>
- Joo, S., Lim, H. J., Jackson, J. D., Atala, A., & Yoo, J. J. (2014). **Myogenic-induced mesenchymal stem cells are capable of modulating the immune response by regulatory T cells.** *Journal of Tissue Engineering*, 5, 2041731414524758. <https://doi.org/10.1177/2041731414524758>
- Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., & Luo, Y. (2022). **Single-cell RNA sequencing technologies and applications: A brief overview.** *Clinical and Translational Medicine*, 12(3). <https://doi.org/10.1002/ctm2.694>
- Karpenko, D., Kapranov, N., & Bigildeev, A. (2022). **Nestin-GFP transgene labels immunoprivileged bone marrow mesenchymal stem cells in the model of ectopic foci formation.** *Frontiers in Cell and Developmental Biology*, 10. <https://doi.org/10.3389/fcell.2022.993056>
- Kashima, Y., Sakamoto, Y., Kaneko, K., Seki, M., Suzuki, Y., & Suzuki, A. (2020). **Single-cell sequencing techniques from individual to multiomics analyses.** *Experimental & Molecular Medicine*, 52(9), 1419–1427. <https://doi.org/10.1038/s12276-020-00499-2>
- Khoshnoodi, J., Pedchenko, V., & Hudson, B. G. (2008). **Mammalian collagen IV.** *Microscopy Research and Technique*, 71(5), 357–370. <https://doi.org/10.1002/jemt.20564>

- Kim, S. H., Das, A., Choi, H. I., Kim, K. H., Chai, J. C., Choi, M. R., Binas, B., Park, K. S., Lee, Y. S., Jung, K. H., & Chai, Y. G. (2019a). **Forkhead box O1 (FOXO1) controls the migratory response of Toll-like receptor (TLR3)-stimulated human mesenchymal stromal cells.** *Journal of Biological Chemistry*, 294(21), 8424–8437. <https://doi.org/10.1074/jbc.RA119.008673>
- Kim, S. M., Jeon, Y., Jang, J. Y., & Lee, H. (2023). **NR1D1 deficiency in the tumor microenvironment promotes lung tumor development by activating the NLRP3 inflammasome.** *Cell Death Discovery*, 9(1), 278. <https://doi.org/10.1038/s41420-023-01554-3>
- Kim, W.-J., Shin, H.-L., Kim, B.-S., Kim, H.-J., & Ryoo, H.-M. (2020). **RUNX2-modifying enzymes: therapeutic targets for bone diseases.** *Experimental & Molecular Medicine*, 52(8), 1178–1184. <https://doi.org/10.1038/s12276-020-0471-4>
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., & Raychaudhuri, S. (2019). **Fast, sensitive and accurate integration of single-cell data with Harmony.** *Nature Methods*, 16(12), 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>
- Kyurkchiev, D. (2014). **Secretion of immunoregulatory cytokines by mesenchymal stem cells.** *World Journal of Stem Cells*, 6(5), 552. <https://doi.org/10.4252/wjsc.v6.i5.552>
- Lamouille, S., Xu, J., & Derynck, R. (2014). **Molecular mechanisms of epithelial–mesenchymal transition.** *Nature Reviews Molecular Cell Biology*, 15(3), 178–196. <https://doi.org/10.1038/nrm3758>
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). **voom: precision weights unlock linear model analysis tools for RNA-seq read counts.** *Genome Biology*, 15(2), R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
- Lee, H.-J., Kim, Y.-H., Choi, D.-W., Cho, K.-A., Park, J.-W., Shin, S.-J., Jo, I., Woo, S.-Y., & Ryu, K.-H. (2021). **Tonsil-derived mesenchymal stem cells enhance allogeneic bone marrow engraftment via collagen IV degradation.** *Stem Cell Research & Therapy*, 12(1), 329. <https://doi.org/10.1186/s13287-021-02414-6>
- Lee, J. Y., & Hong, S.-H. (2020). **Hematopoietic Stem Cells and Their Roles in Tissue Regeneration.** *International Journal of Stem Cells*, 13(1), 1–12. <https://doi.org/10.15283/ijsc19127>
- Lee, S. B., Shim, S., Kim, M.-J., Shin, H.-Y., Jang, W.-S., Lee, S.-J., Jin, Y.-W., Lee, S.-S., & Park, S. (2016). **Identification of a distinct subpopulation of fibroblasts from murine dermis: CD73 – CD105 + as potential marker of dermal fibroblasts subset with multipotency.** *Cell Biology International*, 40(9), 1008–1016. <https://doi.org/10.1002/cbin.10623>
- Lee, S., Kim, O. J., Lee, K. O., Jung, H., Oh, S.-H., & Kim, N. K. (2020). **Enhancing the Therapeutic Potential of CCL2-Overexpressing Mesenchymal Stem Cells in Acute Stroke.** *International Journal of Molecular Sciences*, 21(20), 7795. <https://doi.org/10.3390/ijms21207795>
- Li, B., & Dewey, C. N. (2011). **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics*, 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, F., Negi, V., Yang, P., Lee, J., Ma, K., Moulik, M., & Yeloor, V. K. (2022). **TEAD1 regulates cell proliferation through a pocket-independent transcription repression mechanism.** *Nucleic Acids Research*, 50(22), 12723–12738. <https://doi.org/10.1093/nar/gkac1063>
- Li, H., & Humphreys, B. D. (2021). **Single Cell Technologies: Beyond Microfluidics.** *Kidney360*, 2(7), 1196–1204. <https://doi.org/10.34067/KID.0001822021>

- Li, H.-C., Stoicov, C., Rogers, A. B., & Houghton, J. (2006). **Stem cells and cancer: Evidence for bone marrow stem cells in epithelial cancers.** *World Journal of Gastroenterology*, 12(3), 363. <https://doi.org/10.3748/wjg.v12.i3.363>
- Li, L., Miano, J. M., Cserjesi, P., & Olson, E. N. (1996). **SM22 α , a Marker of Adult Smooth Muscle, Is Expressed in Multiple Myogenic Lineages During Embryogenesis.** *Circulation Research*, 78(2), 188–195. <https://doi.org/10.1161/01.RES.78.2.188>
- Li, T.-H., Zhao, B.-B., Qin, C., Wang, Y.-Y., Li, Z.-R., Cao, H.-T., Yang, X.-Y., Zhou, X.-T., & Wang, W.-B. (2021). **IFIT1 modulates the proliferation, migration and invasion of pancreatic cancer cells via Wnt/ β -catenin signaling.** *Cellular Oncology*, 44(6), 1425–1437. <https://doi.org/10.1007/s13402-021-00651-8>
- Li, Y., Xu, X.-L., Zhao, D., Pan, L.-N., Huang, C.-W., Guo, L.-J., Lu, Q., & Wang, J. (2015). **TLR3 ligand Poly IC Attenuates Reactive Astrogliosis and Improves Recovery of Rats after Focal Cerebral Ischemia.** *CNS Neuroscience & Therapeutics*, 21(11), 905–913. <https://doi.org/10.1111/cns.12469>
- Li, Y., Zhong, H., Wu, M., Tan, B., Zhao, L., Yi, Q., Xu, X., Pan, H., Bi, Y., & Yang, K. (2019). **Decline of p300 contributes to cell senescence and growth inhibition of hUC-MSCs through p53/p21 signaling pathway.** *Biochemical and Biophysical Research Communications*, 515(1), 24–30. <https://doi.org/10.1016/j.bbrc.2019.05.061>
- Lin, T., Pajarinen, J., Nabeshima, A., Lu, L., Nathan, K., Yao, Z., & Goodman, S. B. (2017). **Establishment of NF- κ B sensing and interleukin-4 secreting mesenchymal stromal cells as an “on-demand” drug delivery system to modulate inflammation.** *Cytotherapy*, 19(9), 1025–1034. <https://doi.org/10.1016/j.jcyt.2017.06.008>
- Lin, Y.-C., Niceta, M., Muto, V., Vona, B., Pagnamenta, A. T., Maroofian, R., Beetz, C., van Duyvenvoorde, H., Dentici, M. L., Lauffer, P., Vallian, S., Ciolfi, A., Pizzi, S., Bauer, P., Grüning, N.-M., Bellacchio, E., Del Fattore, A., Petrini, S., Shaheen, R., ... Tartaglia, M. (2021). **SCUBE3 loss-of-function causes a recognizable recessive developmental disorder due to defective bone morphogenetic protein signaling.** *The American Journal of Human Genetics*, 108(1), 115–133. <https://doi.org/10.1016/j.ajhg.2020.11.015>
- Lin, Z., Wu, Y., Xu, Y., Li, G., Li, Z., & Liu, T. (2022). **Mesenchymal stem cell-derived exosomes in cancer therapy resistance: recent advances and therapeutic potential.** *Molecular Cancer*, 21(1), 179. <https://doi.org/10.1186/s12943-022-01650-5>
- Lindsay, S. L., & Barnett, S. C. (2017). **Are nestin-positive mesenchymal stromal cells a better source of cells for CNS repair?** *Neurochemistry International*, 106, 101–107. <https://doi.org/10.1016/j.neuint.2016.08.001>
- Liu, R., Lee, J., Kim, B. S., Wang, Q., Buxton, S. K., Balasubramanyam, N., Kim, J. J., Dong, J., Zhang, A., Li, S., Gupte, A. A., Hamilton, D. J., Martin, J. F., Rodney, G. G., Coarfa, C., Wehrens, X. H. T., Yechool, V. K., & Moulik, M. (2017). **Tead1 is required for maintaining adult cardiomyocyte function, and its loss results in lethal dilated cardiomyopathy.** *JCI Insight*, 2(17). <https://doi.org/10.1172/jci.insight.93343>
- Liu, Z.-H., Dai, X.-M., & Du, B. (2015). **Hes1: a key role in stemness, metastasis and multidrug resistance.** *Cancer Biology & Therapy*, 16(3), 353–359. <https://doi.org/10.1080/15384047.2015.1016662>

- Luo, L., Zhou, Y., Zhang, C., Huang, J., Du, J., Liao, J., Bergholt, N. L., Bünger, C., Xu, F., Lin, L., Tong, G., Zhou, G., & Luo, Y. (2020). **Feeder-free generation and transcriptome characterization of functional mesenchymal stromal cells from human pluripotent stem cells.** *Stem Cell Research*, 48, 101990. <https://doi.org/10.1016/j.scr.2020.101990>
- Luo, X., Ryu, K. W., Kim, D.-S., Nandu, T., Medina, C. J., Gupte, R., Gibson, B. A., Soccio, R. E., Yu, Y., Gupta, R. K., & Kraus, W. L. (2017). **PARP-1 Controls the Adipogenic Transcriptional Program by PARylating C/EBP β and Modulating Its Transcriptional Activity.** *Molecular Cell*, 65(2), 260–271. <https://doi.org/10.1016/j.molcel.2016.11.015>
- Mabuchi, Y., Okawara, C., Méndez-Ferrer, S., & Akazawa, C. (2021). **Cellular Heterogeneity of Mesenchymal Stem/Stromal Cells in the Bone Marrow.** *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.689366>
- Maltman, D. J., Hardy, S. A., & Przyborski, S. A. (2011). **Role of mesenchymal stem cells in neurogenesis and nervous system repair.** *Neurochemistry International*. <https://doi.org/10.1016/j.neuint.2011.06.008>
- Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. Ng., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., & Weinberg, R. A. (2008). **The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells.** *Cell*, 133(4), 704–715. <https://doi.org/10.1016/j.cell.2008.03.027>
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., & Califano, A. (2006). **ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context.** *BMC Bioinformatics*, 7(S1), S7. <https://doi.org/10.1186/1471-2105-7-S1-S7>
- Mary Piper; Meeta Mistry; Jihe Liu; William Gammerdinger; Radhika Khetani. (2022). **hbctraining/scRNA-seq_online: scRNA-seq Lessons from HCBC (first release).**
- McCall, M. N., Bolstad, B. M., & Irizarry, R. A. (2010). **Frozen robust multiarray analysis (fRMA).** *Biostatistics*, 11(2), 242–253. <https://doi.org/10.1093/biostatistics/kxp059>
- Melief, J., Pico de Coaña, Y., Maas, R., Fennemann, F.-L., Wolodarski, M., Hansson, J., & Kiessling, R. (2020). **High expression of ID1 in monocytes is strongly associated with phenotypic and functional MDSC markers in advanced melanoma.** *Cancer Immunology, Immunotherapy*, 69(4), 513–522. <https://doi.org/10.1007/s00262-019-02476-9>
- Meuwissen, M. E. C., Halley, D. J. J., Smit, L. S., Lequin, M. H., Cobben, J. M., de Coo, R., van Harsel, J., Sallevelt, S., Woldringh, G., van der Knaap, M. S., de Vries, L. S., & Mancini, G. M. S. (2015). **The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature.** *Genetics in Medicine*, 17(11), 843–853. <https://doi.org/10.1038/gim.2014.210>
- Minguell, J. J., Erices, A., & Conget, P. (2001). **Mesenchymal Stem Cells.** *Experimental Biology and Medicine*, 226(6), 507–520. <https://doi.org/10.1177/153537020122600603>
- Mollentze, J., Durandt, C., & Pepper, M. S. (2021). **An In Vitro and In Vivo Comparison of Osteogenic Differentiation of Human Mesenchymal Stromal/Stem Cells.** *Stem Cells International*, 2021, 9919361. <https://doi.org/10.1155/2021/9919361>
- Moran, A. P., Prendergast, M. M., & Appelmelk, B. J. (1996). **Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease.** *FEMS Immunology and Medical Microbiology*, 16, 105–120. <https://doi.org/10.1111/j.1574-695X.1996.tb00127.x>

- Morrison, S. J., & Scadden, D. T. (2014). **The bone marrow niche for haematopoietic stem cells.** *Nature*, 505(7483), 327–334. <https://doi.org/10.1038/nature12984>
- Muntión, S., Sánchez-Guijo, F. M., Carrancio, S., Villarón, E., López, O., Díez-Campelo, M., San Miguel, J. F., & del Cañizo, M. C. (2012). **Optimisation of mesenchymal stromal cells karyotyping analysis: implications for clinical use.** *Transfusion Medicine*, 22(2), 122–127. <https://doi.org/10.1111/j.1365-3148.2012.01134.x>
- Najar, M., Krayem, M., Meuleman, N., Bron, D., & Lagneaux, L. (2017). **Mesenchymal Stromal Cells and Toll-Like Receptor Priming: A Critical Review.** *Immune Network*, 17(2), 89. <https://doi.org/10.4110/IN.2017.17.2.89>
- Narwidina, A., Miyazaki, A., Iwata, K., Kurogoushi, R., Sugimoto, A., Kudo, Y., Kawarabayashi, K., Yamakawa, Y., Akazawa, Y., Kitamura, T., Nakagawa, H., Yamaguchi-Ueda, K., Hasegawa, T., Yoshizaki, K., Fukumoto, S., Yamamoto, A., Ishimaru, N., Iwasaki, T., & Iwamoto, T. (2023). **Iroquois homeobox 3 regulates odontoblast proliferation and differentiation mediated by Wnt5a expression.** *Biochemical and Biophysical Research Communications*, 650, 47–54. <https://doi.org/10.1016/j.bbrc.2023.02.004>
- Park, J. S., Kim, M., Song, N.-J., Kim, J.-H., Seo, D., Lee, J.-H., Jung, S. M., Lee, J. Y., Lee, J., Lee, Y. S., Park, K. W., & Park, S. H. (2019). **A Reciprocal Role of the Smad4-Taz Axis in Osteogenesis and Adipogenesis of Mesenchymal Stem Cells.** *Stem Cells*, 37(3), 368–381. <https://doi.org/10.1002/stem.2949>
- Pasquet, S., Naye, F., Faucheux, C., Bronchain, O., Chesneau, A., Thiébaud, P., & Thézé, N. (2006). **Transcription Enhancer Factor-1-dependent Expression of the α -Tropomyosin Gene in the Three Muscle Cell Types.** *Journal of Biological Chemistry*, 281(45), 34406–34420. <https://doi.org/10.1074/jbc.M602282200>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). **Salmon provides fast and bias-aware quantification of transcript expression.** *Nature Methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- Perrier, E., Ronzière, M.-C., Bareille, R., Pinzano, A., Mallein-Gerin, F., & Freyria, A.-M. (2011). **Analysis of collagen expression during chondrogenic induction of human bone marrow mesenchymal stem cells.** *Biotechnology Letters*, 33(10), 2091–2101. <https://doi.org/10.1007/s10529-011-0653-1>
- Petri, R. M., Hackel, A., Hahnel, K., Dumitru, C. A., Bruderek, K., Flohe, S. B., Paschen, A., Lang, S., & Brandau, S. (2017). **Activated Tissue-Resident Mesenchymal Stromal Cells Regulate Natural Killer Cell Immune and Tissue-Regenerative Function.** *Stem Cell Reports*, 9(3), 985–998. <https://doi.org/10.1016/j.stemcr.2017.06.020>
- Pevsner-Fischer, M., Morad, V., Cohen-Sfady, M., Rousso-Noori, L., Zanin-Zhorov, A., Cohen, S., Cohen, I. R., & Zipori, D. (2007). **Toll-like receptors and their ligands control mesenchymal stem cell functions.** *Blood*, 109(4), 1422–1432. <https://doi.org/10.1182/blood-2006-06-028704>
- Pindado, J., Balsinde, J., & Balboa, M. A. (2007). **TLR3-Dependent Induction of Nitric Oxide Synthase in RAW 264.7 Macrophage-Like Cells via a Cytosolic Phospholipase A2/Cyclooxygenase-2 Pathway.** *The Journal of Immunology*, 179(7), 4821–4828. <https://doi.org/10.4049/jimmunol.179.7.4821>
- Pittenger, M. F. (2008). **Mesenchymal Stem Cells from Adult Bone Marrow.** In *Mesenchymal Stem Cells* (pp. 27–44). Humana Press. https://doi.org/10.1007/978-1-60327-169-1_2

- Qiu, G., Zheng, G., Ge, M., Wang, J., Huang, R., Shu, Q., & Xu, J. (2019). **Functional proteins of mesenchymal stem cell-derived extracellular vesicles.** *Stem Cell Research & Therapy*, 10(1), 359. <https://doi.org/10.1186/s13287-019-1484-6>
- Qiu, P., Guo, Q., Yao, Q., Chen, J., & Lin, J. (2021). **Hsa-mir-3163 and CCNB1 may be potential biomarkers and therapeutic targets for androgen receptor positive triple-negative breast cancer.** *PLoS ONE*, 16(11), e0254283. <https://doi.org/10.1371/journal.pone.0254283>
- Querques, F., D'Agostino, A., Cozzolino, C., Cozzuto, L., Lombardo, B., Leggiero, E., Ruosi, C., & Pastore, L. (2019). **Identification of a Novel Transcription Factor Required for Osteogenic Differentiation of Mesenchymal Stem Cells.** *Stem Cells and Development*, 28(6), 370–383. <https://doi.org/10.1089/scd.2018.0152>
- Raicevic, G., Najar, M., Pieters, K., de Bruyn, C., Meuleman, N., Bron, D., Toungouz, M., & Lagneaux, L. (2012). **Inflammation and Toll-like receptor ligation differentially affect the osteogenic potential of human mesenchymal stromal cells depending on their tissue origin.** *Tissue Engineering. Part A*, 18(13–14), 1410–1418. <https://doi.org/10.1089/ten.TEA.2011.0434>
- Reagan, M. R., & Rosen, C. J. (2016). **Navigating the bone marrow niche: translational insights and cancer-driven dysfunction.** *Nature Reviews Rheumatology*, 12(3), 154–168. <https://doi.org/10.1038/nrrheum.2015.160>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Research*, 43(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Robert, A. W., Marcon, B. H., Dallagiovanna, B., & Shigunov, P. (2020). **Adipogenesis, Osteogenesis, and Chondrogenesis of Human Mesenchymal Stem/Stromal Cells: A Comparative Transcriptome Approach.** *Frontiers in Cell and Developmental Biology*, 8. <https://doi.org/10.3389/fcell.2020.00561>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). **edgeR : a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rogers, J. H., Gupta, R., Reyes, J. M., Gundry, M. C., Medrano, G., Guzman, A., Aguilar, R., Conneely, S. E., Song, T., Johnson, C., Barnes, S., Cristobal, C. D. D., Kurtz, K., Brunetti, L., Goodell, M. A., & Rau, R. E. (2021). **Modeling IKZF1 lesions in B-ALL reveals distinct chemosensitivity patterns and potential therapeutic vulnerabilities.** *Blood Advances*, 5(19), 3876–3890. <https://doi.org/10.1182/bloodadvances.2020002408>
- Roson-Burgo, B., Sanchez-Guijo, F., Del Cañizo, C., & De Las Rivas, J. (2014). **Transcriptomic portrait of human Mesenchymal Stromal/Stem cells isolated from bone marrow and placenta.** *BMC Genomics*, 15(1), 910. <https://doi.org/10.1186/1471-2164-15-910>
- Roson-Burgo, B., Sanchez-Guijo, F., Del Cañizo, C., & De Las Rivas, J. (2016). **Insights into the human mesenchymal stromal/stem cell identity through integrative transcriptomic profiling.** *BMC Genomics*, 17(1), 944. <https://doi.org/10.1186/s12864-016-3230-0>
- Sagaradze, G. D., Basalova, N. A., Efimenko, A. Yu., & Tkachuk, V. A. (2020). **Mesenchymal Stromal Cells as Critical Contributors to Tissue Regeneration.** *Frontiers in Cell and Developmental Biology*, 8, 576176. <https://doi.org/10.3389/fcell.2020.576176>
- San Segundo-Val, I., & Sanz-Lozano, C. S. (2016). **Introduction to the Gene Expression Analysis.** *Molecular Genetics of Asthma* (pp. 29–43). https://doi.org/10.1007/978-1-4939-3652-6_3

- Sauer, T., Facchinetti, G., Kohl, M., Kowal, J. M., Rozanova, S., Horn, J., Schmal, H., Kwee, I., Schulz, A.-P., Hartwig, S., Kassem, M., Habermann, J. K., & Gemoll, T. (2022). **Protein Expression of AEBP1, MCM4, and FABP4 Differentiate Osteogenic, Adipogenic, and Mesenchymal Stromal Stem Cells.** *International Journal of Molecular Sciences*, 23(5), 2568. <https://doi.org/10.3390/ijms23052568>
- Scadden, D. T. (2006). **The stem-cell niche as an entity of action.** *Nature*, 441(7097), 1075–1079. <https://doi.org/10.1038/nature04957>
- Schinke, C., Qu, P., Mehdi, S. J., Hoering, A., Epstein, J., Johnson, S. K., van Rhee, F., Zangari, M., Thanendrarajan, S., Barlogie, B., Davies, F. E., Yaccoby, S., & Morgan, G. J. (2018). **The Pattern of Mesenchymal Stem Cell Expression Is an Independent Marker of Outcome in Multiple Myeloma.** *Clinical Cancer Research*, 24(12), 2913–2919. <https://doi.org/10.1158/1078-0432.CCR-17-2627>
- Schofield, R. (1978). **The relationship between the spleen colony-forming cell and the haemopoietic stem cell.** *Blood Cells*, 4(1–2), 7–25.
- Scott, C. L., & Omilusik, K. D. (2019). ZEBs: **Novel Players in Immune Cell Development and Function.** *Trends in Immunology*, 40(5), 431–446. <https://doi.org/10.1016/j.it.2019.03.001>
- Sekiya, I., Larson, B. L., Vuoristo, J. T., Cui, J.-G., & Prockop, D. J. (2003). **Adipogenic Differentiation of Human Adult Stem Cells From Bone Marrow Stroma (MSCs).** *Journal of Bone and Mineral Research*, 19(2), 256–264. <https://doi.org/10.1359/JBMR.0301220>
- Semenova, E., Chroscinska-Krawczyk, M., Grudniak, M., Oldak, T., & Machaj, E. (2018). **Clinical application of AD-MSCs - A review.** *Journal of PreClinical and Clinical Research*, 12(3), 100–105. <https://doi.org/10.26444/jpccr/94910>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). **DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update).** *Nucleic Acids Research*, 50(W1), W216–W221. <https://doi.org/10.1093/nar/gkac194>
- Si, H., Zhang, Y., Song, Y., & Li, L. (2018). **Overexpression of adrenomedullin protects mesenchymal stem cells against hypoxia and serum deprivation induced apoptosis via the Akt/GSK3 β and Bcl 2 signaling pathways.** *International Journal of Molecular Medicine*. <https://doi.org/10.3892/ijmm.2018.3533>
- Somaiah, C., Kumar, A., Mawrie, D., Sharma, A., Patil, S. D., Bhattacharyya, J., Swaminathan, R., & Jaganathan, B. G. (2015). **Collagen Promotes Higher Adhesion, Survival and Proliferation of Mesenchymal Stem Cells.** *PLoS ONE*, 10(12), e0145068. <https://doi.org/10.1371/journal.pone.0145068>
- Soneson, C., Love, M. I., & Robinson, M. D. (2015). **Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.** *F1000Research*, 4, 1521. <https://doi.org/10.12688/f1000research.7563.1>
- Song, S.-W., Kim, K.-E., Choi, J.-W., Lee, C. Y., Lee, J., Seo, H.-H., Lim, K. H., Lim, S., Lee, S., Kim, S. W., & Hwang, K.-C. (2016). **Proteomic Analysis and Identification of Paracrine Factors in Mesenchymal Stem Cell-Conditioned Media under Hypoxia.** *Cellular Physiology and Biochemistry*, 40(1–2), 400–410. <https://doi.org/10.1159/000452555>

- Souza-Moreira, L., Tan, Y., Wang, Y., Wang, J.-P., Salkhordeh, M., Virgo, J., Florian, M., Murray, A. B. P., Watpool, I., McIntyre, L., English, S., Stewart, D. J., & Mei, S. H. J. (2022). **Poly(I:C) enhances mesenchymal stem cell control of myeloid cells from COVID-19 patients.** *iScience*, 25(5), 104188. <https://doi.org/10.1016/j.isci.2022.104188>
- Stefańska, K., Bryl, R., Moncrieff, L., Pinto, N., Shibli, J. A., & Dyszkiewicz-Konwińska, M. (2020). **Mesenchymal stem cells – a historical overview.** *Medical Journal of Cell Biology*, 8(2), 83–87. <https://doi.org/10.2478/acb-2020-0010>
- Stehman, S. V. (1997). **Selecting and interpreting measures of thematic classification accuracy.** *Remote Sensing of Environment*, 62(1), 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- Tan, P. Y., Chang, C. W., Duan, K., Poidinger, M., Ng, K. L., Chong, Y. S., Gluckman, P. D., & Stünkel, W. (2016). **E2F1 Orchestrates Transcriptomics and Oxidative Metabolism in Wharton’s Jelly-Derived Mesenchymal Stem Cells from Growth-Restricted Infants.** *PLoS ONE*, 11(9), e0163035. <https://doi.org/10.1371/journal.pone.0163035>
- Tan, Y., Gan, M., Fan, Y., Li, L., Zhong, Z., Li, X., Bai, L., Zhao, Y., Niu, L., Shang, Y., Zhang, S., & Zhu, L. (2019). **miR-10b-5p regulates 3T3-L1 cells differentiation by targeting Apol6.** *Gene*, 687, 39–46. <https://doi.org/10.1016/j.gene.2018.11.028>
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B., Siddiqui, A., Lao, K., Surani, M. (2009). **mRNA-Seq whole-transcriptome analysis of a single cell.** *Nature Methods*, 6(5), 377-382. <https://doi.org/10.1038/nmeth.1315>
- Tracy, L. E., Minasian, R. A., & Catterson, E. J. (2016). **Extracellular Matrix and Dermal Fibroblast Function in the Healing Wound.** *Advances in Wound Care*, 5(3), 119–136. <https://doi.org/10.1089/wound.2014.0561>
- Trivedi, S., Srivastava, K., Gupta, A., Saluja, T. S., Kumar, S., Mehrotra, D., & Singh, S. K. (2020). **A quantitative method to determine osteogenic differentiation aptness of scaffold.** *Journal of Oral Biology and Craniofacial Research*, 10(2), 158–160. <https://doi.org/10.1016/j.jobcr.2020.04.006>
- Tsui, Lin, Chang, Hou, Chen, Feng, & Juang. (2019). **Transgelin, a p53 and PTEN-Upregulated Gene, Inhibits the Cell Proliferation and Invasion of Human Bladder Carcinoma Cells in Vitro and in Vivo.** *International Journal of Molecular Sciences*, 20(19), 4946. <https://doi.org/10.3390/ijms20194946>
- Uccelli, A., Moretta, L., & Pistoia, V. (2008). **Mesenchymal stem cells in health and disease.** *Nature Reviews Immunology*, 8(9), 726–736. <https://doi.org/10.1038/nri2395>
- Ugurlu, B., & Karaoz, E. (2020). **Comparison of similar cells: Mesenchymal stromal cells and fibroblasts.** *Acta Histochemica*, 122(8), 151634. <https://doi.org/10.1016/j.acthis.2020.151634>
- Ullah, M., Stich, S., Notter, M., Eucker, J., Sittinger, M., & Ringe, J. (2013). **Transdifferentiation of mesenchymal stem cells-derived adipogenic-differentiated cells into osteogenic- or chondrogenic-differentiated cells proceeds via dedifferentiation and have a correlation with cell cycle arresting and driving genes.** *Differentiation*, 85(3), 78–90. <https://doi.org/10.1016/j.diff.2013.02.001>
- Vaculik, C., Schuster, C., Bauer, W., Iram, N., Pfisterer, K., Kramer, G., Reinisch, A., Strunk, D., & Elbe-Bürger, A. (2012a). **Human Dermis Harbors Distinct Mesenchymal Stromal Cell Subsets.** *Journal of Investigative Dermatology*, 132(3), 563–574. <https://doi.org/10.1038/jid.2011.355>

- Vaculik, C., Schuster, C., Bauer, W., Iram, N., Pfisterer, K., Kramer, G., Reinisch, A., Strunk, D., & Elbe-Bürger, A. (2012b). **Human Dermis Harbors Distinct Mesenchymal Stromal Cell Subsets.** *Journal of Investigative Dermatology*, 132(3), 563–574. <https://doi.org/10.1038/jid.2011.355>
- van Dongen, J., Macintyre, E., Gabert, J., Delabesse, E., Rossi, V., Saglio, G., Gottardi, E., Rambaldi, A., Dotti, G., Griesinger, F., Parreira, A., Gameiro, P., Díaz, M. G., Malec, M., Langerak, A., San Miguel, J., & Biondi, A. (1999). **Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease.** *Leukemia*, 13(12), 1901–1928. <https://doi.org/10.1038/sj.leu.2401592>
- Vézina Audette, R., Lavoie-Lamoureux, A., Lavoie, J.-P., & Laverty, S. (2013). **Inflammatory stimuli differentially modulate the transcription of paracrine signaling molecules of equine bone marrow multipotent mesenchymal stromal cells.** *Osteoarthritis and Cartilage*, 21(8), 1116–1124. <https://doi.org/10.1016/j.joca.2013.05.004>
- Villaron, E. M., Almeida, J., López-Holgado, N., Alcoceba, M., Sánchez-Abarca, L. I., Sanchez-Guijo, F. M., Alberca, M., Pérez-Simon, J. A., San Miguel, J. F., & del Cañizo, M. C. (2004). **Mesenchymal stem cells are present in peripheral blood and can engraft after allogeneic hematopoietic stem cell transplantation.** *Haematologica*, 89(12), 1421–1427.
- Wang, J., Yang, Y., & Du, B. (2022). **Clinical Characterization and Prognostic Value of TPM4 and Its Correlation with Epithelial–Mesenchymal Transition in Glioma.** *Brain Sciences*, 12(9), 1120. <https://doi.org/10.3390/brainsci12091120>
- Wang, S., Su, H., Feng, P., Deng, W., Su, C., Wu, Y., & Shen, H. (2020). **Loss of death-associated protein kinase 1 in human bone marrow mesenchymal stem cells decreases immunosuppression of CD4+ T cells.** *Journal of International Medical Research*, 48(6), 030006052093345. <https://doi.org/10.1177/0300060520933453>
- Wang, X.-X., Yin, G.-Q., Zhang, Z.-H., Rong, Z.-H., Wang, Z.-Y., Du, D.-D., Wang, Y.-D., Gao, R.-X., & Xian, G.-Z. (2020). **TWIST1 transcriptionally regulates glycolytic genes to promote the Warburg metabolism in pancreatic cancer.** *Experimental Cell Research*, 386(1), 111713. <https://doi.org/10.1016/j.yexcr.2019.111713>
- Wang, Z., Chai, C., Wang, R., Feng, Y., Huang, L., Zhang, Y., Xiao, X., Yang, S., Zhang, Y., & Zhang, X. (2021). **Single-cell transcriptome atlas of human mesenchymal stem cells exploring cellular heterogeneity.** *Clinical and Translational Medicine*, 11(12). <https://doi.org/10.1002/ctm2.650>
- Wangler, S., Menzel, U., Li, Z., Ma, J., Hoppe, S., Benneker, L. M., Alini, M., Grad, S., & Peroglio, M. (2019). **CD146/MCAM distinguishes stem cell subpopulations with distinct migration and regenerative potential in degenerative intervertebral discs.** *Osteoarthritis and Cartilage*, 27(7), 1094–1105. <https://doi.org/10.1016/j.joca.2019.04.002>
- Ward, A., & Hudson, J. W. (2014). **p53-Dependent and Cell Specific Epigenetic Regulation of the Polo-like kinases under Oxidative Stress.** *PLoS ONE*, 9(1), e87918. <https://doi.org/10.1371/journal.pone.0087918>
- Waterman, R. S., Tomchuck, S. L., Henkle, S. L., & Betancourt, A. M. (2010). **A New Mesenchymal Stem Cell (MSC) Paradigm: Polarization into a Pro-Inflammatory MSC1 or an Immunosuppressive MSC2 Phenotype.** *PLoS ONE*, 5(4), e10088. <https://doi.org/10.1371/journal.pone.0010088>
- Wei, K., Nguyen, H. N., & Brenner, M. B. (2021). **Fibroblast pathology in inflammatory diseases.** *Journal of Clinical Investigation*, 131(20). <https://doi.org/10.1172/JCI149538>

- Wei, Y., Chen, Y.-H., Li, L.-Y., Lang, J., Yeh, S.-P., Shi, B., Yang, C.-C., Yang, J.-Y., Lin, C.-Y., Lai, C.-C., & Hung, M.-C. (2011). **CDK1-dependent phosphorylation of EZH2 suppresses methylation of H3K27 and promotes osteogenic differentiation of human mesenchymal stem cells.** *Nature Cell Biology*, 13(1), 87–94. <https://doi.org/10.1038/ncb2139>
- Weir, G. M., Karkada, M., Hoskin, D., Stanford, M. M., MacDonald, L., Mansour, M., & Liwski, R. S. (2017). **Combination of poly I:C and Pam3CSK4 enhances activation of B cells in vitro and boosts antibody responses to protein vaccines in vivo.** *PLoS ONE*, 12(6), e0180073. <https://doi.org/10.1371/journal.pone.0180073>
- Whelan, D. S., Caplice, N. M., & Clover, A. J. P. (2020). **Mesenchymal stromal cell derived CCL2 is required for accelerated wound healing.** *Scientific Reports*, 10(1), 2642. <https://doi.org/10.1038/s41598-020-59174-1>
- Williams, A. R., & Hare, J. M. (2011). **Mesenchymal Stem Cells.** *Circulation Research*, 109(8), 923–940. <https://doi.org/10.1161/CIRCRESAHA.111.243147>
- Wilson, A., Hodgson-Garms, M., Frith, J. E., & Genever, P. (2019). **Multiplicity of Mesenchymal Stromal Cells: Finding the Right Route to Therapy.** *Frontiers in Immunology*, 10. <https://doi.org/10.3389/fimmu.2019.01112>
- Wolfien, M., David, R., & Galow, A.-M. (2021). **Single-Cell RNA Sequencing Procedures and Data Analysis.** *In Bioinformatics* (pp. 19–35). Exon Publications, Brisbane, Australia <https://doi.org/10.36255/exonpublications.bioinformatics.2021.ch2>
- Wu, L., Mickey Williams, P., & Koch, W. H. (2005). **Clinical applications of microarray-based diagnostic tests.** *BioTechniques*, 39(4S), S557–S582. <https://doi.org/10.2144/000112046>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). **clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.** *The Innovation*, 2(3), 100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Xavier, G. M., Panousopoulos, L., & Cobourne, M. T. (2013). **Correction: Scube3 Is Expressed in Multiple Tissues during Development but Is Dispensable for Embryonic Survival in the Mouse.** *PLoS ONE*, 8(9). <https://doi.org/10.1371/annotation/b65bdf5d-4917-4936-9d9d-f217ab2602cd>
- Xian, J., Liang, D., Zhao, C., Chen, Y., & Zhu, Q. (2022). **TRIM21 inhibits the osteogenic differentiation of mesenchymal stem cells by facilitating K48 ubiquitination-mediated degradation of Akt.** *Experimental Cell Research*, 412(2), 113034. <https://doi.org/10.1016/j.yexcr.2022.113034>
- Xie, L., Zeng, X., Hu, J., & Chen, Q. (2015). **Characterization of Nestin, a Selective Marker for Bone Marrow Derived Mesenchymal Stem Cells.** *Stem Cells International*, 2015, 1–9. <https://doi.org/10.1155/2015/762098>
- Xie, Z., Yu, W., Ye, G., Li, J., Zheng, G., Liu, W., Lin, J., Su, Z., Che, Y., Ye, F., Zhang, Z., Wang, P., Wu, Y., & Shen, H. (2022). **Single-cell RNA sequencing analysis of human bone-marrow-derived mesenchymal stem cells and functional subpopulation identification.** *Experimental & Molecular Medicine*, 54(4), 483–492. <https://doi.org/10.1038/s12276-022-00749-5>
- Yan, C., Chang, J., Song, X., Qi, Y., Ji, Z., Liu, T., Yu, W., Wei, F., Yang, L., & Ren, X. (2021). **Lung cancer-associated mesenchymal stem cells promote tumor metastasis and tumorigenesis by induction of epithelial–mesenchymal transition and stem-like reprogram.** *Aging*, 13(7), 9780–9800. <https://doi.org/10.18632/aging.202732>

- Yang, F., Luo, P., Ding, H., Zhang, C., & Zhu, Z. (2018). **Collagen type V $\alpha 2$ (COL5A2) is decreased in steroid-induced necrosis of the femoral head.** *American Journal of Translational Research*, 10(8), 2469–2479.
- Yu, M., Guo, G., Huang, L., Deng, L., Chang, C.-S., Achyut, B. R., Canning, M., Xu, N., Arbab, A. S., Bollag, R. J., Rodriguez, P. C., Mellor, A. L., Shi, H., Munn, D. H., & Cui, Y. (2020). **CD73 on cancer-associated fibroblasts enhanced by the A2B-mediated feedforward circuit enforces an immune checkpoint.** *Nature Communications*, 11(1), 515. <https://doi.org/10.1038/s41467-019-14060-x>
- Zaman, G., Staines, K. A., Farquharson, C., Newton, P. T., Dudhia, J., Chenu, C., Pitsillides, A. A., & Dhoot, G. K. (2016). **Expression of Sulf1 and Sulf2 in cartilage, bone and endochondral fracture healing.** *Histochemistry and Cell Biology*, 145(1), 67–79. <https://doi.org/10.1007/s00418-015-1365-8>
- Zaman, G., Sunters, A., Galea, G. L., Javaheri, B., Saxon, L. K., Moustafa, A., Armstrong, V. J., Price, J. S., & Lanyon, L. E. (2012). **Loading-related Regulation of Transcription Factor EGR2/Krox-20 in Bone Cells Is ERK1/2 Protein-mediated and Prostaglandin, Wnt Signaling Pathway-, and Insulin-like Growth Factor-I Axis-dependent.** *Journal of Biological Chemistry*, 287(6), 3946–3962. <https://doi.org/10.1074/jbc.M111.252742>
- Zeng, F., Gao, M., Liao, S., Zhou, Z., Luo, G., & Zhou, Y. (2023). **Role and mechanism of CD90+ fibroblasts in inflammatory diseases and malignant tumors.** *Molecular Medicine*, 29(1), 20. <https://doi.org/10.1186/s10020-023-00616-7>
- Zhang, C., Han, X., Liu, J., Chen, L., Lei, Y., Chen, K., Si, J., Wang, T., Zhou, H., Zhao, X., Zhang, X., An, Y., Li, Y., & Wang, Q.-F. (2022). **Single-cell Transcriptomic Analysis Reveals the Cellular Heterogeneity of Mesenchymal Stem Cells.** *Genomics, Proteomics & Bioinformatics*, 20(1), 70–86. <https://doi.org/10.1016/j.gpb.2022.01.005>
- Zhang, H.-M., Liu, T., Liu, C.-J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y., & Guo, A.-Y. (2015). **AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors.** *Nucleic Acids Research*, 43(D1), D76–D81. <https://doi.org/10.1093/nar/gku887>
- Zhang, L., Luo, Q., Shu, Y., Zeng, Z., Huang, B., Feng, Y., Zhang, B., Wang, X., Lei, Y., Ye, Z., Zhao, L., Cao, D., Yang, L., Chen, X., Liu, B., Wagstaff, W., Reid, R. R., Luu, H. H., Haydon, R. C., ... Kang, Q. (2019). **Transcriptomic landscape regulated by the 14 types of bone morphogenetic proteins (BMPs) in lineage commitment and differentiation of mesenchymal stem cells (MSCs).** *Genes & Diseases*, 6(3), 258–275. <https://doi.org/10.1016/j.gendis.2019.03.008>
- Zhang, Q., Dong, J., Zhang, P., Zhou, D., & Liu, F. (2021). **Dynamics of Transcription Factors in Three Early Phases of Osteogenic, Adipogenic, and Chondrogenic Differentiation Determining the Fate of Bone Marrow Mesenchymal Stem Cells in Rats.** *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.768316>
- Zhang, X.-Z., Li, F.-H., & Wang, X.-J. (2021). **Regulation of Tripartite Motif-Containing Proteins on Immune Response and Viral Evasion.** *Frontiers in Microbiology*, 12, 794882. <https://doi.org/10.3389/fmicb.2021.794882>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). **Massively parallel digital transcriptional profiling of single cells.** *Nature Communications*, 8(1), 14049. <https://doi.org/10.1038/ncomms14049>

- Zhou, Z., Zhou, Z., Huang, Z., He, S., & Chen, S. (2020). **Histone-fold centromere protein W (CENP-W) is associated with the biological behavior of hepatocellular carcinoma cells.** *Bioengineered*, 11(1), 729–742. <https://doi.org/10.1080/21655979.2020.1787776>
- Zhuang, H., Zhang, X., Zhu, C., Tang, X., Yu, F., wei Shang, G., & Cai, X. (2016). **Molecular Mechanisms of PPAR- γ ; Governing MSC Osteogenic and Adipogenic Differentiation.** *Current Stem Cell Research & Therapy*, 11(3), 255–264. <https://doi.org/10.2174/1574888X10666150531173309>
- Zuk, P. A., Zhu, M., Mizuno, H., Huang, J., Futrell, J. W., Katz, A. J., Benhaim, P., Lorenz, H. P., & Hedrick, M. H. (2001). **Multilineage Cells from Human Adipose Tissue: Implications for Cell-Based Therapies.** *Tissue Engineering*, 7(2), 211–228. <https://doi.org/10.1089/107632701300062859>

LIST OF PUBLICATIONS

List of publications and scientific output associated to this PhD thesis work.

Main publications associated to this PhD:

Sánchez-Luis E, Joaquín-García A, Campos-Laborie FJ, Sánchez-Guijo F, De Las Rivas J (2020). **Deciphering Master Gene Regulators and Associated Networks of Human Mesenchymal Stromal Cells.** *Biomolecules*, 10(4):557. doi: 10.3390/biom10040557. PMID: **32260546**; PMCID: PMC7226324.

Patents:

TITLE: "In vitro method for the identification of human Mesenchymal Stem/Stromal Cells"

INVENTORS (p.o.): Javier DE LAS RIVAS (40%), Sandra MUNTION (20%), Elena SÁNCHEZ-LUIS (20%), and Fermín SÁNCHEZ-GUIJO (20%)

APPLICATION and REFERENCE NUMBER: EP21383041

GRANT NUMBER: Granted, but under review regarding the number of specific marker genes that are included in the protection. / Granted, but in the process of reviewing the number of specific marker genes that will be included in the protection.

APPLICATION DATE: 17.November.2021

PRIORITY COUNTRY: Europe (EU)

OWNER ENTITY: Salamanca Biomedical Research Institute (IBSAL-IIS), Salamanca (also included as 2ary partial partners CSIC and USAL)

ENTITY THAT EXPLOITS IT: IBSAL-IIS

Additional publications:

Muntión S, Preciado S, Sánchez-Luis E, Corchete L, Díez-Campelo M, Osugui L, Martí-Chillón GJ, Vidriales MB, Navarro-Bailón A, De Las Rivas J, Sánchez-Guijo F (2022). **Eltrombopag increases the hematopoietic supporting ability of mesenchymal stem/stromal cells.** *Therapeutic Advances in Hematology*, 13:20406207221142137. doi: 10.1177/20406207221142137. PMID: **36601635**; PMCID: PMC9806379.

ACKNOWLEDGMENTS

En primer lugar, quiero agradecer especialmente a mi director de Tesis, el Dr. Javier De Las Rivas, por la oportunidad de trabajar y poder realizar la Tesis en su laboratorio, así como el gran aprendizaje en el mundo de la bioinformática. También darle las gracias por guiarme en la carrera investigadora, el apoyo y la confianza.

También agradecer a mis co-directores de Tesis, el Dr. Fermín Sanchez-Guijo Martín y la Dra. Sandra Muntión Olave, por su apoyo, sus conocimientos y total disponibilidad durante todo el proceso de la tesis, sobre todo en el terreno experimental. Y también a todo el equipo de Terapia Celular.

Agradecer también a todo el personal del Centro de Investigación del Cáncer y a la Universidad de Salamanca, por los trámites y los procesos que nos permiten seguir con la tesis día a día.

Agradecer al laboratorio del Dr. Cesar Nombela y a todo su equipo, en especial a Anjali, Serena, Ana y Candice por la increíble acogida y el cariño durante los meses que estuve en Zúrich, que para mí se ha convertido en un sitio muy especial.

No podría faltar mi propio laboratorio 19, por el cual ha pasado gente que se han convertido en personas muy queridas. Natalia, Alberto y Enrique (¡sigue duro baby!), y a los que estuvieron, Curro, Mónica, Santi, y un especial agradecimiento a ti Óscar, gracias por haberme enseñado tanto, por la paciencia, la amabilidad y el cariño. Y también agradecer el cariño y apoyo de Chus y José.

A toda la gente del laboratorio 17; Alba, Antonio, Cristina, Óscar, Luis, Chema y Carmen y en especial a Sara, ex -Lab 17. También a las chicas del 4: Helena, Aurora, Eva, Helena y a Patri y nuestro maravilloso viaje con curso de buceo incluido. También agradecer a Ignacio Jesús por su alegría e intensas muestras de cariño y a Víctor quien tampoco está ya en el CIC. Y en general a mucha más gente con la que he pasado muy buenos momentos y me han acompañado en esta aventura dentro del CIC.

A mis amigos y amigas de Girona y Barcelona, Alicia, Berta, Anna, Guti y todos los que habéis estado siempre apoyándome en el camino. Así como a mis amigas y amigos de Salamanca.

También quería dar un especial agradecimiento a mi equipo, Team Cadena Warrior. Sobre todo, a mi entrenador Guilherme, gracias por confiar en mi e introducirme en las artes marciales, por hacer del jiu-jitsu mi disciplina y desfogue en los momentos más tensos de la tesis, así como disfrutar inmensamente de este deporte y de la gran familia que ha construido. Agradecer también a Daniella, Gregory y Ana, quienes además de compañeros, se han convertido en grandes acompañantes en este viaje.

Por último, quería agradecer eternamente por la familia que tengo. Gracias a mi padre, mi madre y a mi hermano por el apoyo incondicional que me habéis brindado durante la tesis y el que me dais siempre, habéis sido y sois mi pilar fundamental.

ANNEX I: SUPPLEMENTARY TABLES

Supplementary Table 1.

Gene name and description of 151 gene candidates for the characterization of MSCs

Gene	Description
ACAN	aggrecan [Source:HGNC Symbol;Acc:HGNC:319]
ACTA2	actin alpha 2, smooth muscle [Source:HGNC Symbol;Acc:HGNC:130]
ACTB	actin beta [Source:HGNC Symbol;Acc:HGNC:132]
ADIPOQ	adiponectin, C1Q and collagen domain containing [Source:HGNC Symbol;Acc:HGNC:13633]
ALCAM	activated leukocyte cell adhesion molecule [Source:HGNC Symbol;Acc:HGNC:400]
ANPEP	alanyl aminopeptidase, membrane [Source:HGNC Symbol;Acc:HGNC:500]
ANXA2	annexin A2 [Source:HGNC Symbol;Acc:HGNC:537]
ANXA5	annexin A5 [Source:HGNC Symbol;Acc:HGNC:543]
APCDD1	APC down-regulated 1 [Source:HGNC Symbol;Acc:HGNC:15718]
ATL1	atlastin GTPase 1 [Source:HGNC Symbol;Acc:HGNC:11231]
B2M	beta-2-microglobulin [Source:HGNC Symbol;Acc:HGNC:914]
BAMBI	BMP and activin membrane bound inhibitor [Source:HGNC Symbol;Acc:HGNC:30251]
BDNF	brain derived neurotrophic factor [Source:HGNC Symbol;Acc:HGNC:1033]
BHLHE41	basic helix-loop-helix family member e41 [Source:HGNC Symbol;Acc:HGNC:16617]
BMP2	bone morphogenetic protein 2 [Source:HGNC Symbol;Acc:HGNC:1069]
BMP4	bone morphogenetic protein 4 [Source:HGNC Symbol;Acc:HGNC:1071]
BMP7	bone morphogenetic protein 7 [Source:HGNC Symbol;Acc:HGNC:1074]
BMPR1A	bone morphogenetic protein receptor type 1A [Source:HGNC Symbol;Acc:HGNC:1076]
BMPR1B	bone morphogenetic protein receptor type 1B [Source:HGNC Symbol;Acc:HGNC:1077]
BMPR2	bone morphogenetic protein receptor type 2 [Source:HGNC Symbol;Acc:HGNC:1078]
CASP3	caspase 3 [Source:HGNC Symbol;Acc:HGNC:1504]
CD14	CD14 molecule [Source:HGNC Symbol;Acc:HGNC:1628]
CD151	CD151 molecule (Raph blood group) [Source:HGNC Symbol;Acc:HGNC:1630]
CD19	CD19 molecule [Source:HGNC Symbol;Acc:HGNC:1633]
CD200	CD200 molecule [Source:HGNC Symbol;Acc:HGNC:7203]
CD226	CD226 molecule [Source:HGNC Symbol;Acc:HGNC:16961]
CD34	CD34 molecule [Source:HGNC Symbol;Acc:HGNC:1662]
CD44	CD44 molecule (Indian blood group) [Source:HGNC Symbol;Acc:HGNC:1681]
CD58	CD58 molecule [Source:HGNC Symbol;Acc:HGNC:1688]
CD59	CD59 molecule (CD59 blood group) [Source:HGNC Symbol;Acc:HGNC:1689]
CD79B	CD79b molecule [Source:HGNC Symbol;Acc:HGNC:1699]
CDH1	cadherin 1 [Source:HGNC Symbol;Acc:HGNC:1748]
CDH2	cadherin 2 [Source:HGNC Symbol;Acc:HGNC:1759]
CDX2	caudal type homeobox 2 [Source:HGNC Symbol;Acc:HGNC:1806]
CEBPA	CCAAT enhancer binding protein alpha [Source:HGNC Symbol;Acc:HGNC:1833]
CEMIP	cell migration inducing hyaluronidase 1 [Source:HGNC Symbol;Acc:HGNC:29213]
COL10A1	collagen type X alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2185]
COL1A1	collagen type I alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2197]
COL1A2	collagen type I alpha 2 chain [Source:HGNC Symbol;Acc:HGNC:2198]

COL2A1	collagen type II alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2200]
COL3A1	collagen type III alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2201]
COL4A1	collagen type IV alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2202]
COL4A2	collagen type IV alpha 2 chain [Source:HGNC Symbol;Acc:HGNC:2203]
CRISPLD1	cysteine rich secretory protein LCCL domain containing 1 [Source:HGNC Symbol;Acc:HGNC:18206]
CSRP1	cysteine and glycine rich protein 1 [Source:HGNC Symbol;Acc:HGNC:2469]
DLX2	distal-less homeobox 2 [Source:HGNC Symbol;Acc:HGNC:2915]
DLX5	distal-less homeobox 5 [Source:HGNC Symbol;Acc:HGNC:2918]
ENG	endoglin [Source:HGNC Symbol;Acc:HGNC:3349]
ERBB2	erb-b2 receptor tyrosine kinase 2 [Source:HGNC Symbol;Acc:HGNC:3430]
ERRFI1	ERBB receptor feedback inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:18185]
EYA2	EYA transcriptional coactivator and phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:3520]
FABP4	fatty acid binding protein 4 [Source:HGNC Symbol;Acc:HGNC:3559]
FGF18	fibroblast growth factor 18 [Source:HGNC Symbol;Acc:HGNC:3674]
FGF2	fibroblast growth factor 2 [Source:HGNC Symbol;Acc:HGNC:3676]
FGFR2	fibroblast growth factor receptor 2 [Source:HGNC Symbol;Acc:HGNC:3689]
FHL2	four and a half LIM domains 2 [Source:HGNC Symbol;Acc:HGNC:3703]
FZD9	frizzled class receptor 9 [Source:HGNC Symbol;Acc:HGNC:4047]
GAPDH	glyceraldehyde-3-phosphate dehydrogenase [Source:HGNC Symbol;Acc:HGNC:4141]
GATA4	GATA binding protein 4 [Source:HGNC Symbol;Acc:HGNC:4173]
GDF5	growth differentiation factor 5 [Source:HGNC Symbol;Acc:HGNC:4220]
GNL3	G protein nucleolar 3 [Source:HGNC Symbol;Acc:HGNC:29931]
GUSB	glucuronidase beta [Source:HGNC Symbol;Acc:HGNC:4696]
HPRT1	hypoxanthine phosphoribosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:5157]
HSPB3	heat shock protein family B (small) member 3 [Source:HGNC Symbol;Acc:HGNC:5248]
IHH	Indian hedgehog signaling molecule [Source:HGNC Symbol;Acc:HGNC:5956]
IL10	interleukin 10 [Source:HGNC Symbol;Acc:HGNC:5962]
IL1B	interleukin 1 beta [Source:HGNC Symbol;Acc:HGNC:5992]
INHBA	inhibin subunit beta A [Source:HGNC Symbol;Acc:HGNC:6066]
ITGA11	integrin subunit alpha 11 [Source:HGNC Symbol;Acc:HGNC:6136]
ITGA5	integrin subunit alpha 5 [Source:HGNC Symbol;Acc:HGNC:6141]
ITGAM	integrin subunit alpha M [Source:HGNC Symbol;Acc:HGNC:6149]
ITGAX	integrin subunit alpha X [Source:HGNC Symbol;Acc:HGNC:6152]
ITGBL1	integrin subunit beta like 1 [Source:HGNC Symbol;Acc:HGNC:6164]
JAG1	jagged canonical Notch ligand 1 [Source:HGNC Symbol;Acc:HGNC:6188]
KDR	kinase insert domain receptor [Source:HGNC Symbol;Acc:HGNC:6307]
KIT	KIT proto-oncogene, receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6342]
KITLG	KIT ligand [Source:HGNC Symbol;Acc:HGNC:6343]
KLF4	KLF transcription factor 4 [Source:HGNC Symbol;Acc:HGNC:6348]
KRT18	keratin 18 [Source:HGNC Symbol;Acc:HGNC:6430]
LEPR	leptin receptor [Source:HGNC Symbol;Acc:HGNC:6554]
LGALS3BP	galectin 3 binding protein [Source:HGNC Symbol;Acc:HGNC:6564]
LIF	LIF interleukin 6 family cytokine [Source:HGNC Symbol;Acc:HGNC:6596]
LIPE	lipase E, hormone sensitive type [Source:HGNC Symbol;Acc:HGNC:6621]
MCAM	melanoma cell adhesion molecule [Source:HGNC Symbol;Acc:HGNC:6934]
MGP	matrix Gla protein [Source:HGNC Symbol;Acc:HGNC:7060]
MLLT11	MLLT11 transcription factor 7 cofactor [Source:HGNC Symbol;Acc:HGNC:16997]
MME	membrane metalloendopeptidase [Source:HGNC Symbol;Acc:HGNC:7154]
MYC	MYC proto-oncogene, bHLH transcription factor [Source:HGNC Symbol;Acc:HGNC:7553]
NANOG	Nanog homeobox [Source:HGNC Symbol;Acc:HGNC:20857]

NCAM1	neural cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:7656]
NES	nestin [Source:HGNC Symbol;Acc:HGNC:7756]
NGFR	nerve growth factor receptor [Source:HGNC Symbol;Acc:HGNC:7809]
NOTCH3	notch receptor 3 [Source:HGNC Symbol;Acc:HGNC:7883]
NPR3	natriuretic peptide receptor 3 [Source:HGNC Symbol;Acc:HGNC:7945]
NPY	neuropeptide Y [Source:HGNC Symbol;Acc:HGNC:7955]
NT5E	5'-nucleotidase ecto [Source:HGNC Symbol;Acc:HGNC:8021]
NUDT6	nudix hydrolase 6 [Source:HGNC Symbol;Acc:HGNC:8053]
OMD	osteomodulin [Source:HGNC Symbol;Acc:HGNC:8134]
P3H2	prolyl 3-hydroxylase 2 [Source:HGNC Symbol;Acc:HGNC:19317]
PAX2	paired box 2 [Source:HGNC Symbol;Acc:HGNC:8616]
PAX3	paired box 3 [Source:HGNC Symbol;Acc:HGNC:8617]
PDGFRA	platelet derived growth factor receptor alpha [Source:HGNC Symbol;Acc:HGNC:8803]
PDGFRB	platelet derived growth factor receptor beta [Source:HGNC Symbol;Acc:HGNC:8804]
PDLIM5	PDZ and LIM domain 5 [Source:HGNC Symbol;Acc:HGNC:17468]
PDX1	pancreatic and duodenal homeobox 1 [Source:HGNC Symbol;Acc:HGNC:6107]
PECAM1	platelet and endothelial cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:8823]
PLIN4	perilipin 4 [Source:HGNC Symbol;Acc:HGNC:29393]
PLPP4	phospholipid phosphatase 4 [Source:HGNC Symbol;Acc:HGNC:23531]
POU5F1	POU class 5 homeobox 1 [Source:HGNC Symbol;Acc:HGNC:9221]
PPARG	peroxisome proliferator activated receptor gamma [Source:HGNC Symbol;Acc:HGNC:9236]
PROM1	prominin 1 [Source:HGNC Symbol;Acc:HGNC:9454]
PSMB4	proteasome 20S subunit beta 4 [Source:HGNC Symbol;Acc:HGNC:9541]
PTPRC	protein tyrosine phosphatase receptor type C [Source:HGNC Symbol;Acc:HGNC:9666]
RPS10	ribosomal protein S10 [Source:HGNC Symbol;Acc:HGNC:10383]
RPS18	ribosomal protein S18 [Source:HGNC Symbol;Acc:HGNC:10401]
RUNX2	RUNX family transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:10472]
S100A4	S100 calcium binding protein A4 [Source:HGNC Symbol;Acc:HGNC:10494]
SCD	stearoyl-CoA desaturase [Source:HGNC Symbol;Acc:HGNC:10571]
SCUBE3	signal peptide, CUB domain and EGF like domain containing 3 [Source:HGNC Symbol;Acc:HGNC:13655]
SEL1L3	SEL1L family member 3 [Source:HGNC Symbol;Acc:HGNC:29108]
SFRP4	secreted frizzled related protein 4 [Source:HGNC Symbol;Acc:HGNC:10778]
SLC25A4	solute carrier family 25 member 4 [Source:HGNC Symbol;Acc:HGNC:10990]
SNAI2	snail family transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:11094]
SORBS2	sorbin and SH3 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:24098]
SORT1	sortilin 1 [Source:HGNC Symbol;Acc:HGNC:11186]
SOX9	SRY-box transcription factor 9 [Source:HGNC Symbol;Acc:HGNC:11204]
SPP1	secreted phosphoprotein 1 [Source:HGNC Symbol;Acc:HGNC:11255]
SUSD2	sushi domain containing 2 [Source:HGNC Symbol;Acc:HGNC:30667]
TAGLN	transgelin [Source:HGNC Symbol;Acc:HGNC:11553]

TBC1D2	TBC1 domain family member 2 [Source:HGNC Symbol;Acc:HGNC:18026]
TBP	TATA-box binding protein [Source:HGNC Symbol;Acc:HGNC:11588]
TEK	TEK receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:11724]
TERT	telomerase reverse transcriptase [Source:HGNC Symbol;Acc:HGNC:11730]
TFAP2A	transcription factor AP-2 alpha [Source:HGNC Symbol;Acc:HGNC:11742]
TFRC	transferrin receptor [Source:HGNC Symbol;Acc:HGNC:11763]
THY1	Thy-1 cell surface antigen [Source:HGNC Symbol;Acc:HGNC:11801]
TIMP4	TIMP metalloproteinase inhibitor 4 [Source:HGNC Symbol;Acc:HGNC:11823]
TNFRSF11B	TNF receptor superfamily member 11b [Source:HGNC Symbol;Acc:HGNC:11909]
TNFRSF12A	TNF receptor superfamily member 12A [Source:HGNC Symbol;Acc:HGNC:18152]
TPD52L1	TPD52 like 1 [Source:HGNC Symbol;Acc:HGNC:12006]
TSLP	thymic stromal lymphopoietin [Source:HGNC Symbol;Acc:HGNC:30743]
TWIST1	twist family bHLH transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:12428]
TWIST2	twist family bHLH transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:20670]
UBC	ubiquitin C [Source:HGNC Symbol;Acc:HGNC:12468]
UBE2D3	ubiquitin conjugating enzyme E2 D3 [Source:HGNC Symbol;Acc:HGNC:12476]
VCAM1	vascular cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:12663]
VIM	vimentin [Source:HGNC Symbol;Acc:HGNC:12692]
VWF	von Willebrand factor [Source:HGNC Symbol;Acc:HGNC:12726]
WNT3A	Wnt family member 3A [Source:HGNC Symbol;Acc:HGNC:15983]
ZFP42	ZFP42 zinc finger protein [Source:HGNC Symbol;Acc:HGNC:30949]
HLA-DR	Major Histocompatibility Complex, Class II, DR Alpha [Source:HGNC Symbol;Acc:HGNC:4947]

ANNEX II: RESUMEN EN CASTELLANO

INTRODUCCIÓN

Las células madre mesenquimales (MSCs) han sido estudiadas durante los últimos 30 años en el campo de la biomedicina debido sus múltiples propiedades en terapia celular. Actualmente, muchos estudios están centrados especialmente en el campo de la medicina regenerativa gracias a sus capacidades inmunomodulatorias.

Desde el descubrimiento de las MSCs en 1976 por el investigador Friedenstein y sus colaboradores, numerosos estudios han sido realizados aportando relevantes datos a este descubrimiento. En 2006, la Sociedad Internacional de Terapia Celular (ISCT), establecía los criterios mínimos para la consideración de las células madre mesenquimales, en los que definía: la adherencia al plástico de las MSCs mantenida en condiciones de cultivo estándar, las MSCs debían ser positivas en la expresión de los marcadores de superficie: CD105, CD73 y CD90, así como negativas en CD45, CD34, CD14 o CD11b, CD79alpha o CD19 y HLA-DR. Por último, las MSCs debían tener capacidad para diferenciarse *in vitro* en osteoblastos, adipocitos y condrocitos. Definiendo las MSCs como células madre adultas con capacidad de auto-regeneración y diferenciación en gran cantidad de tejidos. Entre ellos lo que tienen más relevancia en el ámbito clínico son las MSCs procedentes de médula ósea y de tejido adiposo. Existen también numerosos tejidos más como: pulpa dental, tejidos derivados de nacimiento, líquido amniótico y placenta, sangre periférica sinovial y líquido sinovial, endometrio, gelatina de Wharton, tejido cervical, pulmón, piel y músculo. También recalcar el importante rol de las MSCs en la homeostasis de el nicho hematopoyético, entre otras múltiples funciones.

El nicho hematopoyético de médula ósea está comprendido por múltiples tipos celulares, incluidas las MSCs, éste constituye un microambiente que combina los estados quiescentes y la activación proliferativa de las células madre. Entre los tipos celulares, cabe resaltar la importancia del linaje hematopoyético y el no hematopoyético. Dentro del linaje hematopoyético, las células madre hematopoyéticas (HSCs) son fundamentales por su capacidad de diferenciación y regulación del nicho. Por otro lado, los fibroblastos (FIB), constituyen una parte fundamental en la síntesis de la matriz extracelular, con un fenotipo altamente similar a las MSCs, incluyendo las características típicas marcadas por la ISCT para las MSCs y sus propiedades inmunológicas. Por último, remarcar la importancia de las capacidades regulatorias de las MSCs en el nicho hematopoyético mediante sus propias propiedades inmunomodulatorias e interacción con otras células. En este marco, cabe resaltar la importancia de los Master reguladores (MR), factores de transcripción capaces de regular grupos de genes que a su vez activan o inhiben múltiples funciones que pueden afectar al nicho hematopoyético.

Las capacidades inmunológicas de las MSCs son muchas, en esta tesis doctoral nos centraremos en la estimulación por-inflamatoria e inmunosupresora mediante los receptores de tipo Toll (TLRs), en especial TLR4 y TLR3. Numerosos estudios reportan la estimulación con LPS en el receptor TLR4 de las MSCs, como promotor de sus capacidades inflamatorias mediante la segregación de diferentes citoquinas. Mientras la estimulación de TLR3 en MSCs, mediante poly(I:C), es reportada por sus capacidades regulatorias. Recientemente, se han descrito varios estudios que discuten varias discordancias acerca del tiempo de estimulación con poly(I:C) en las células madre mesenquimales, llegando incluso modificar sus propiedades moduladoras en el nicho hematopoyético.

En última instancia, remarcar que el notable avance en el estudio de las MSCs ha sido posible gracias a las nuevas técnicas de secuenciación que se han ido desarrollando en los últimos años. Técnicas como los Microarrays, RNA-seq o RNAseq (secuenciación de RNA) de célula única (scRNAseq) han permitido profundizar en la identificación de posibles nuevos genes marcadores, así como su

caracterización, funciones asociadas y el estudio de las subpoblaciones.

OBJETIVOS

Esta Tesis Doctoral titulada "Caracterización experimental y bioinformática del perfil transcriptómico de las Células Madre Mesenquimales (MSC) humanas", se centra en el estudio y caracterización de dos aspectos fundamentales de las Células Madre Mesenquimales aisladas de tejidos humanos adultos: (i) Caracterización transcriptómica: el primer aspecto corresponde a la expresión génica, el perfil transcriptómico y los mecanismos reguladores de genes de las MSC, centrándose en la identificación de distintos marcadores específicos que caracterizan a estas células; y (ii) Análisis inmunomodulador: el segundo aspecto corresponde al análisis de las propiedades inmunomoduladoras de las MSC bajo diferentes condiciones de estimulación (centrándose en la estimulación de TLR3 y TLR4), que proporcionan diferentes fenotipos y promueven distintos efectos dentro de su nicho celular.

Las Células Madre Mesenquimales son células madre adultas multipotentes que se encuentran en diversos tejidos humanos, incluida la médula ósea, el tejido adiposo y la pulpa dental, entre otros. Las MSC se caracterizan por su capacidad para diferenciarse en una variedad de tipos de células, incluidos osteoblastos, condrocitos y adipocitos, lo que las hace muy valiosas para la medicina regenerativa, aplicaciones de reparación de tejidos y otras terapias basadas en células. Si bien los efectos inmunomoduladores de las MSC han llamado la atención recientemente, muchas facetas de sus propiedades moleculares, funciones y heterogeneidad celular siguen sin resolverse. Para abordar estos problemas, se emplean técnicas pioneras de alto rendimiento, como la secuenciación de ARN del transcriptoma completo y la secuenciación de ARN unicelular, para lograr una caracterización más precisa, facilitando la identificación de nuevos biomarcadores y sus funciones asociadas. Siguiendo estas ideas, este estudio tiene como objetivo proporcionar un análisis transcriptómico profundo y detallado de las MSC humanas que arroje luz sobre sus características moleculares y celulares.

Objetivos principales:

Dentro de este marco temático, esta Tesis Doctoral utiliza varios métodos experimentales y bioinformáticos para cumplir los objetivos principales. La tesis ha sido escrita y se presenta en **cuatro capítulos principales** organizados para abordar cuatro objetivos específicos de nuestro trabajo científico:

OBJETIVO 1: Determinación de la firma transcriptómica de MSC humanas para identificar nuevos marcadores genéticos precisos que las distingan de otros tipos de células relacionadas.

a. Integración de un gran compendio de datos de expresión de todo el genoma (perfiles transcriptómicos) de muestras de MSC de diferentes orígenes de tejidos, junto con fibroblastos primarios y células madre hematopoyéticas utilizando diferentes técnicas transcriptómicas: Microarrays, Exon Arrays y RNA-seq.

b. Análisis de los datos transcriptómicos recopilados, empleando métodos estadísticos bioinformáticos y computacionales avanzados para identificar los genes marcadores óptimos que exhiben las diferencias más significativas entre los tipos de células y seleccionando nuevos marcadores genéticos específicos de MSC. Se utilizarán métodos de aprendizaje automático (machine learning) para evaluar la exactitud y precisión de los nuevos marcadores en comparación con los marcadores comúnmente utilizados para identificar MSC.

c. Validación experimental mediante PCR cuantitativa en tiempo real (RT-qPCR) de los nuevos marcadores genéticos encontrados en los diferentes tipos celulares testados.

Este Objetivo, incluyendo Material y Métodos, así como Resultados y Discusión, se presenta en

el Capítulo 1 de esta disertación.

OBJETIVO 2: Integración de datos transcriptómicos de diferentes tipos de células hematopoyéticas y no hematopoyéticas y construcción de redes reguladoras de genes para identificar los reguladores maestros (“Master Regulators”) de MSC.

a. Recopilación e integración de un compendio de datos transcriptómicos de diferentes tipos de células humanas asociadas a la médula ósea, incluidas células madre/progenitoras hematopoyéticas, linfocitos (LYM), Células Madre Mesenquimales, fibroblastos, osteoblastos (OSTB) y MSC estimuladas (stMSC).

b. Generación de las firmas genéticas derivadas del análisis de expresión diferencial de MSC frente a los diferentes tipos celulares descritos, seguido de un análisis de enriquecimiento funcional de dichas firmas.

c. Uso de algoritmos bioinformáticos para construir redes reguladoras de genes e identificar los reguladores maestros (“Master Regulators”) de las MSC (es decir, los factores de transcripción clave, TF, que regulan el perfil de expresión de las MSC). Análisis combinado de datos de expresión transcriptómica y datos de metilación para corroborar la actividad de los TF.

Este Objetivo, incluyendo Material y Métodos, Resultados y Discusión, se presenta en el Capítulo 2 de esta disertación. Además, este trabajo fue publicado como artículo de investigación en la revista *Biomolecules* en 2020 (doi: 10.3390/biom10040557). El artículo se adjunta al final de esta Tesis.

OBJETIVO 3: Análisis transcriptómico de las propiedades inmunomoduladoras de las MSC estimuladas para TLR3 y TLR4, junto con la exploración de los efectos dependientes del tiempo de la estimulación de TLR3.

a. Recopilación de datos de RNA-seq de la estimulación de MSC con lipopolisacárido, LPS, y con ácido poliinosínico-policitídílico, poli(I:C), seguido de análisis de expresión diferencial y enriquecimiento funcional utilizando diferentes métodos bioinformáticos.

b. Estimulación de MSC de médula ósea humana de donantes sanos con poli(I:C) en diferentes momentos (es decir, con tratamiento a corto y largo tiempo) para perfilar patrones y funciones de expresión génica utilizando tecnología de secuenciación de ARN transcriptómica completa.

c. Estudio de las diferencias en la diferenciación osteogénica y adipogénica de las MSC mediante análisis experimental y estadístico.

Este Objetivo, incluyendo Material y Métodos, Resultados y Discusión, se presenta en el Capítulo 3 de esta disertación.

OBJETIVO 4: Identificación y análisis de subpoblaciones de MSC estimuladas y MSC control *in vitro* y comparación entre MSC control *in vitro* e *in vivo* utilizando tecnología de expresión unicelular.

a. Estimulación de MSC de médula ósea humana de donantes sanos tratados con poli(I:C) durante un tiempo breve (6 horas). La tecnología de RNA-seq unicelular (scRNA-seq) se aplica al estudio comparativo de las 2 condiciones: MSC de control y MSC estimuladas con poli(I:C).

b. Implementación de un proceso bioinformático para el análisis de datos de MSC unicelulares, incluido el análisis de agrupamiento de diferentes subconjuntos de células en las muestras de MSC bajo cada condición. Identificación de patrones genéticos únicos y funciones biológicas en los grupos celulares encontrados.

c. Comparación de poblaciones celulares y firmas genéticas obtenidas a partir de scRNA-seq de MSC de médula ósea derivadas de diferentes donantes sanos, obtenidas *in vivo* e *in vitro*. Identificación de clusters (agrupaciones) compartidos entre ambas condiciones y establecimiento de las diferentes asociaciones gen-función dentro de cada subgrupo.

Este Objetivo, incluyendo Material y Métodos, Resultados y Discusión, se presenta en el Capítulo 4 de esta disertación. El tercer punto de este capítulo representa una investigación colaborativa con el Departamento de Oncohematología del Hospital Universitario de Zúrich, actualmente en curso y no divulgada en el momento de escribir esta Tesis, por lo que requiere medidas de confidencialidad para ciertos resultados.

CAPÍTULO 1:

En este primer capítulo de la tesis doctoral se ha realizado la caracterización transcriptómica de las MSCs mediante datos de tres metodologías distintas: Microarrays, ExonArrays y RNA-seq, con el objetivo de identificar nuevos marcadores más específicos y precisos que permitan la identificación de las MSCs. Este estudio consta de la comparación de tres tipos celulares: células mesenquimales madre, células madre hematopoyéticas y Fibroblastos.

En primera instancia, se determina un análisis diferencial mediante diferentes herramientas bioinformáticas sobre 151 genes para cada una de las metodologías en los tres tipos celulares y posterior análisis funcional de estos. Los resultados determinan los genes TAGLN, COL4A1, COL4A2 y SCUBE3 como posibles nuevos marcadores de MSCs.

Posteriormente, analizamos nuestros genes de forma independiente mediante análisis bioinformáticos, estadísticos y experimentales presentando los mejores resultados en la expresión de TAGLN. Adicionalmente, la combinación de nuestros cuatro marcadores propuestos: TAGLN, COL4A1, COL4A2 y SCUBE3 se compararon con los marcadores positivos estándar de MSCs determinados por la ISCT: THY1, ENG y NT5E, presentando mejores niveles de especificidad y sensibilidad.

Nuestros resultados también revelan la capacidad de esta combinación de genes para distinguir las MSC de las HSCs y los fibroblastos. Teniendo en cuenta la heterogeneidad de las MSC, podría ser interesante la combinación de nuestros cuatro candidatos para considerar las subpoblaciones desconocidas de MSCs. Para corroborar esta hipótesis, procedimos a realizar diseños experimentales mediante reacción en cadena de la polimerasa con transcripción inversa cuantitativa (RT-qPCR) para identificar de nuevo nuestros genes TAGLN, COL4A2, COL4A1 y SCUBE3 como los genes más expresados en MSCs en comparación a cinco tipos de células humanas diferentes: fibroblastos, células mononucleares, líneas de células estromales (HTERT and HY5) y células mesenquimales procedentes de tejido adiposo. Todos estos resultados podrían llevarnos a proponer estos genes noveles como una nueva combinación de marcadores génicos de MSCs.

CAPÍTULO 2:

Este capítulo se basa en la caracterización de la regulación de las células mesenquimales madre. Para conseguir este objetivo, realizamos un estudio bioinformático de MSCs y diferentes tipos celulares basado en obtener un compendio de datos de Microarrays procedentes de la plataforma Gene Expression Omnibus (GEO). Sobre estos datos se realiza un análisis de expresión diferencial, análisis de enriquecimiento funcional, análisis de metilación y analizar con diferentes herramientas bioinformáticas los posibles factores de transcripción (TF) y los llamados "Master Reguladores". Además, sobre todos los resultados, realizamos redes de coregulación para obtener una mejor identificación de los TFs. Todo el contenido que aparece en este capítulo ha sido publicado en nuestro artículo "Deciphering Master Gene Regulators and Associated Networks of Human Mesenchymal Stromal Cells" (DOI:

10.3390/biom10040557).

En la primera parte de este capítulo se realiza un análisis de expresión diferencial sobre los siguientes tipos celulares: Células madre/progenitoras hematopoyéticas y células hematopoyéticas diferenciadas (linfocitos (LYM)), células madre/estromales mesenquimales (MSC) y células estimuladas (stMSC), MSC derivadas de adipocitos; MSC derivadas de condrocitos; osteoblastos (OSTB), y fibroblastos de piel (FIB).

Los resultados indicaron una mayor relación entre los perfiles de expresión de las MSCs, osteoblastos, adipocitos, condroblastos y fibroblastos. Una relación comprensible, dado que los osteoblastos, adipocitos y condroblastos proceden de la diferenciación de MSCs. Por otro lado, muchos estudios científicos revelan varias similitudes entre fibroblastos y las MSC ya que comparten numerosas funciones, sobre todo como constituyentes de la matriz extracelular. Las diferencias entre las MSCs y las MSCs estimuladas radican sobre todo en procesos de inflamación, incluyendo genes como IL6 o VCAM1. Por último, la comparación que reveló más diferencias entre si fue el de las MSCs y el linaje hematopoyético, con genes altamente expresados en las MSCs como SNAI2, presente en la transición epitelio-mesenquimal (EMT) o genes de la familia del Colágeno, como son COL1A2, COL3A1, COL5A2 y COL6A3, con funciones relacionadas con el metabolismo celular, el crecimiento o supervivencia de las células.

La segunda parte de este capítulo engloba la búsqueda de Master Reguladores, en un contraste entre las MSCs y las HSCs. Los resultados basados en estudios bioinformáticos presentan los top 10 MR sobre expresados en MSCs, incluyendo: SNAI2, STAB2, IRX3, EPAS1, HOXC6, TWIST1, TULP3, PRRX1, TEAD1, y NFE2L1, y los top 10 MR sobre expresados en HSC: BCL11A, MYB, TFEC, HLF, GATA2, ERG, PLAGL2, DACH2, POU2F1, y GATA3. Además, estos datos fueron corroborados mediante estudios de metilación, resultando los top 10 MR en MSCs como hipometilados, lo que supone una sobreexpresión del gen, y los genes top 10 MR HSCs como hipermetilados, en consecuencia, silenciados en las MSCs.

Por último, se realizan estudios sobre los 20 MR en redes de regulación para determinar si los MR encontrados presentaban secuencias comunes en las áreas de unión al promotor. En este caso TEAD1 presenta nuevos relevantes TF asociados. Mediante este estudio también encontramos nuevos TF importantes en médula ósea: EP300, GADD45A, E2F1, SPL1 EP300, GADD45A, E2F1, SPL1. Aunque estos necesitan de mayor exploración científica debido a la falta de bibliografía presente.

CAPÍTULO 3:

Este capítulo se fundamenta en las características inmunomodulatorias de las MSCs. En él, se realiza un análisis transcriptómico mediante datos procedentes de GEO de RNAseq de las MSCS estimuladas con LPS, lo que causa la estimulación de TLR4, y MSCs estimuladas con poly(I:C), causando la estimulación de TLR3. Además, se añade un estudio experimental y bioinformático de la caracterización de las células BM-MSCs obtenidas de donantes sanos estimuladas con poly(I:C) a diferentes tiempos, siendo a tiempos tempranos de 6 horas y tiempos tardíos de 24 horas.

El estudio transcriptómico de las MSCs estimuladas con LPS, cabe considerar que se adquirieron datos de dos concentraciones distintas de LPS: 1 µg/ml and 10 ng/ml. En ambos casos se observa una clara sobre expresión de genes causantes de respuestas inflamatorias, tales como CCL2, CXCL8, IL1B, y TNFAIP3 en comparación las células MSCs control. Nuestros resultados, demuestran una mayor señal de expresión en estos genes inflamatorios a mayor concentración de LPS. Contrariamente, el primer estudio transcriptómico que se hizo sobre las células MSCs estimuladas con poly(I:C) 10 µg/ml, a tiempo 4 horas en base a datos obtenidos de GEO presentaron una expresión más difusa en el análisis diferencial respecto a los datos de la estimulación de MSCs con LPS.

Con el objetivo de mejorar estos resultados, realizamos estudios experimentales de cultivos celulares y transcriptómicos, mediante RNAseq de las células MSCs estimuladas con poly(I:C) a una concentración mayor de 25 µg/ml, a tiempos cortos de 6 horas y tiempos largos de 24 horas, y sus respectivos controles de MSCs. Además, también se realizaron estudios experimentales de diferenciación adipogénica y osteogénica sobre dichas estimulaciones.

Nuestro análisis reveló distintos patrones de expresión genética procesados por un análisis diferencial. En las MSC estimuladas durante 6 horas, observamos la regulación positiva de genes productores de citocinas como CXCL10, CXCL11, CXCL8, CCL8, CX3CL1, CCL3 y CCL2, lo que indica una respuesta inflamatoria. Al mismo tiempo, genes relacionados con la apoptosis y la inhibición de la diferenciación celular, como ID1, ID2 y SMAD6, mostraron una regulación negativa en el corto tiempo de estimulación, lo que sugiere un impacto en las propiedades inmunosupresoras. Por último, el análisis de enriquecimiento funcional, presenta procesos descritos por varios autores de la estimulación con poly(I:C), tales como la regulación de funciones virales. Todos estos datos confirman que la estimulación breve de poly (I:C) durante 6 horas en MSC conduce a un fenotipo proinflamatorio con características generales asociadas a la regulación viral, lo que indica que las MSC muestran dualidad en la estimulación de las MSC con poly(I:C). Por el contrario, después de 24 horas de estimulación, las MSC exhibieron funciones relacionadas con la regulación del sistema inmunológico, expresándose significativamente en genes de la familia IFIT y TRIM, así como también en los genes OAS2, OAS3, BTN3A1, BTN3A2 y BTN3A3.

Acorde a los resultados anteriores, realizamos un nuevo análisis diferencial para revelar diferencias en la estimulación poly(I:C) a lo largo del tiempo (6 horas y 24 horas). En los genes bajo expresados en la estimulación de 24 horas, como ACO1, ZMAT3, ANKH, NRXN3, CERK o TCF7, observamos su implicación en el incremento de la diferenciación osteogénica, y TNFRSF10D es un gen implicado en respuestas inflamatorias. Este hallazgo corrobora la modulación inflamatoria de las MSC durante la breve estimulación de 6 horas y su potencial inducción hacia la diferenciación osteogénica. Por otro lado, en los genes bajo expresados a las 6 horas, pero sobre regulados a las 24 horas, denotan genes de la familia BTN, ID1 e ID2, que, como se mencionó anteriormente. Este análisis de expresión diferencial reveló posibles nuevas funciones de la estimulación poly(I:C) en las MSCs. Los datos sugieren que una estimulación breve de 6 horas induce una etapa inflamatoria en las MSC, similar a la estimulación con LPS. Mientras tanto, la estimulación tardía de poly (I:C) a las 24 horas introduce los conocidos efectos inmunosupresores en las MSCs.

Finalmente, para corroborar estos resultados correspondientes con la diferenciación, realizamos varios experimentos de diferenciación osteogénica y adipogénica, así como su correspondiente análisis bioinformático y estadístico. La diferenciación adipogénica mostró resultados significativos con un mayor crecimiento de adipocitos en la estimulación poly(I:C), con efectos más pronunciados a las 24 horas. Esto se correlacionó con la expresión genética diferencial, que reveló una regulación positiva en los genes PARP implicados en la diferenciación adipogénica. Por el contrario, la diferenciación osteogénica mostró diferencias en el microscopio óptico, pero nuestro análisis estadístico no encontró diferencias significativas. Así como el análisis de expresión diferencial reveló genes asociados con diferencias osteogénicas en las MSC estimuladas durante 6 horas.

CAPÍTULO 4:

En este último capítulo de la tesis doctoral, se realizaron estudios transcriptómicos de célula única mediante la tecnología de scRNAseq en las MSCs y MSCs estimuladas con poly(I:C) 25 µg/ml a tiempo de 6 horas. Estas células madre mesenquimales, procedían de la médula ósea de dos donantes sanos. Además, se realiza un estudio comparativo de los datos control de nuestras células MSCs cultivadas *in vitro* respecto a los datos de scRNA de células MSCs *in vivo* procedentes de donantes sanos. Estos

últimos datos fueron proporcionados durante mi estancia doctoral en el Departamento Oncohematológico de la Universidad y Hospital Universitario de Zúrich, Suiza.

El análisis integral inicial incluyó ambas estimulaciones (análisis global), estos resultados pudieron enmascarar genes relevantes con menor expresión que tienen importancia en nuestra comparación. Los resultados clave dentro de distintos grupos se vincularon con funciones relacionadas con la respuesta al virus y la inmunidad innata, las mitocondrias citoplasmáticas, la respiración aeróbica, la división celular, las proteínas musculares y las fibras de estrés.

Para mejorar la precisión de nuestros resultados, realizamos dos análisis independientes en función de las MSC estimuladas y no estimuladas. Las MSC de control se segregaron en tres grupos diferentes. El grupo más grande, el grupo 0, incluye genes relativamente inespecíficos relacionados con proteínas musculares, filamentos de fibrina y citoesqueleto. El grupo 1 exhibió expresión potencial en genes de la familia TMP: TPM1, TPM2 y TPM4 y genes de la familia miogénica: MYL6, MYL12, MYL12B; y gen TAGLN. Genes relacionados con funciones musculares. Por el contrario, el Grupo 2 exhibió fuertes asociaciones con genes de la familia del colágeno como: COL8A1, COL4A1, COL4A2, COL5A2, COL12A1 y COL11A1, genes relacionados con la diferenciación osteogénica. Estos resultados nos permitieron clasificar las MSC control en dos grupos principales relacionados con funciones musculares o funciones osteogénicas.

Por otro lado, el análisis de estimulación poly(I:C) de MSC contenía 5 grupos. El grupo 0, el más grande, contenía un patrón genético de inflamación típico, en consonancia con las observaciones del Capítulo 4. En este están presentes varias quimiocinas, citocinas e interleucinas, como CXCL9, CCL8, IL6, TNFAIP6 y TNFSPDF10 y CXCL1. El grupo 0 también mostró sobreexpresión en TAGLN, MYL6, TPM1, TPM2 y TPM4, correspondiente al grupo 1 en las MSC de control. El grupo 1 de MSC estimuladas presentó un gen típico asociado con la diferenciación osteogénica, como el factor de transcripción relacionado con Runt 2 (RUNX2), así como genes asociados a la diferenciación adipogénica como APOL6, PARP14 y FBXL7, y genes asociados a la EMT como ZEB1 y ZEB2. El grupo 2, de manera similar al grupo de control 2, mostró expresión de genes de la familia del colágeno como COL6A3, COL4A1, COL4A2, COL12A1 y COL11A1, lo que indica su sólido potencial en la diferenciación. Mientras que el grupo 3 apareció como una mezcla del grupo 2 y el grupo 0, sin una asociación de patrón genético distinta. Finalmente, el grupo 4 fue el grupo más distinto, con genes que se presentan exclusivamente en este grupo. Los genes presentes en este grupo fueron: TPX2, CENPW, TK1, CCNB1, involucrados en la mitosis y la división celular.

Estos análisis independientes permitieron una correlación más específica entre los grupos de análisis independientes y globales, llevando a una categorización más precisa de los subgrupos identificados. Dentro de estos grupos en la categorización global encontramos: la subpoblación estrechamente asociada con la estimulación poly(I:C) se denominó "MSC estimuladas", conocidas por su perfil inflamatorio. La subpoblación de MSC denominada "MSCs prediferenciadas" presentó todos los genes asociados con la diferenciación adipogénica y osteogénica en el caso de las células MSCs estimuladas, así como genes de la familia del colágeno del grupo 2 en el análisis de MSC de control. La subpoblación "MSCs poco estimuladas", aunque se presentaron totalmente definidos como grupos de MSC estimuladas, no tenían las mismas propiedades. Las células de este grupo no tenían este perfil inflamatorio, pero estaban en concordancia con las MSC prediferenciadas. También se identifica la subpoblación de "MSCs musculares" claramente representadas tanto en el análisis control como el estimulado, presentan asociaciones con familias de genes miogénicos y tropomiogénicos, cuyas funciones estaban asociadas con las funciones musculares. Además, estos genes estaban relacionados por sus efectos potenciales de diferenciación miogénica esquelética, expresión de antígenos inmunes, reacciones linfocíticas mixtas y modulación de la respuesta inmune por células T reguladoras. La subpoblación de "MSCs en crecimiento", mostró un claro patrón genético en MSC estimuladas relacionadas con el ciclo

celular y los procesos mitóticos, dentro de este grupo no se encontraron asociaciones con las MSCs control, debido a la baja agrupación de grupos en el análisis independiente de estas. Finalmente, el subgrupo “MSCs”, compuesto principalmente por MSC de Control, carecía de funciones específicas más allá de su distribución en el análisis global, estando asociado con fibras de transición citoplasmática y estrés. La identificación de nuevos subgrupos en MSC dentro de MSC estimuladas y no estimuladas enriquece nuestra comprensión de la caracterización de las MSCs. Este proceso no solo mejora nuestra comprensión de la compleja heterogeneidad dentro de las subpoblaciones de MSC, sino que también facilita la identificación de los subgrupos de MSC más prometedores para próximas investigaciones.

La última parte de este capítulo involucró la integración de MSC de médula ósea humana *in vitro* e *in vivo*. El objetivo principal de este proyecto fue fusionar y correlacionar los subgrupos *in vivo* de MSC: células CAR, células osteogénicas y células miogénicas con nuestros datos de MSC *in vitro*. Dentro del grupo *in vivo* de las células CAR, generamos dos nuevos subgrupos en función de su caracterización determinada por técnicas bioinformáticas, denominados como CARv1 y CAR 2. Después de la integración de ambos datos, la mayoría de las células MSCs *in vitro* se agruparon estrechamente con CAR 2, mientras que un grupo más pequeño de células se integró con el mismo grupo de células CAR 1.

Nuestros resultados demostraron que la correlación de las células CAR 2 y la mayoría de las células *in vitro* estaba asociada con la regulación de la hematopoyesis, la diferenciación y un perfil osteogénico. Curiosamente, pudimos observar genes como CXCL2 o genes de la familia TNF, MYL9, TPM1, TIMP1, TPM4, COL1A1 que se identificaron en nuestros subgrupos como “MSC musculares” en el análisis de MSC de control y en “MSC prediferenciadas” en nuestro análisis global que contenía MSC de control.

La segunda asociación fue más sólida debido a que ambos grupos se encontraban dentro del mismo grupo de clusterización. Las funciones clave estaban asociadas con la matriz extracelular y la adhesión al sustrato celular. En particular, genes de la familia del colágeno tales como COL12A1, COL11A1, COL5A2, COL4A2, COL8A1; adquieren importancia relevante en este grupo, reflejando su presencia en el segundo grupo de análisis de MSC *in vitro*. Estos genes de la familia del colágeno promueven una mayor adhesión, supervivencia y proliferación de las MSC.

Este estudio destaca la existencia de dos grupos principales dentro de las MSC *in vitro* que presentan puntos en común con las células CAR BM-MSCs *in vivo* humanas. Estos dos grupos principales se distinguen por la pronunciada asociación de los procesos de inflamación y diferenciación en la mayoría de las células *in vitro* y células CAR 2, en comparación con el grupo más pequeño de MSC *in vitro* con células CAR 1, asociado con la construcción de la matriz extracelular y la adhesión celular de las MSC.

CONCLUSIONES

A la luz de los objetivos científicos y de investigación propuestos en esta Tesis Doctoral, y tras discutir e integrar todos nuestros resultados, podemos extraer las siguientes CONCLUSIONES principales:

1. Recopilamos, integramos y analizamos un extenso compendio de datos de expresión génica de células madre mesenquimales humanas, células madre hematopoyéticas y fibroblastos generados por diferentes plataformas transcriptómicas. Utilizando todos estos datos, pudimos identificar cuatro nuevos marcadores genéticos específicos de hMSC: COL4A1, COL4A2, TAGLN y SCUBE3; que superan a los marcadores CD utilizados actualmente para aislar estas células madre. Estos nuevos marcadores son particularmente precisos en su capacidad para separar y evitar confusión con los fibroblastos. Estos datos fueron corroborados mediante análisis experimentales de expresión génica en diferentes tipos de células mediante PCR cuantitativa.

2. Para el segundo objetivo, analizamos una colección de datos de expresión y metilación de diferentes tipos de células asociadas con hMSC derivadas de la médula ósea. Utilizando estos datos, generamos una red de coexpresión genética original centrada en la actividad de las hMSC, seguida de una red reguladora de genes que permitió la identificación de reguladores maestros (Master Regulators) y regulones. Los reguladores maestros más importantes que regulan positivamente la expresión de la red generada fueron: SNAI2, SATB2, IRX3, EPAS1, HOXC6.

3. Para el tercer objetivo, realizamos un estudio de MSC sometidas a diferentes estimulaciones —con LPS y poli(I:C)— y analizamos sus cambios en el transcriptoma mediante RNA-seq. La estimulación con poly(I:C) se realizó en varios momentos. Además, realizamos experimentos de diferenciación adipogénica y osteogénica utilizando BM-MS. La combinación de estos estudios experimentales seguidos de análisis bioinformáticos y estadísticos nos permite identificar firmas genéticas y procesos biológicos asociados con la estimulación de las MSC.



4. Se analizó un conjunto de muestras de MSC aisladas *in vitro* de donantes sanos y estimuladas con poly(I:C) utilizando RNA-seq. Se delinearon diferentes grupos de células en las muestras, revelando poblaciones de células distintas dentro de las MSC tanto de control como estimuladas. Estos experimentos demostraron la viabilidad y el poder de utilizar datos unicelulares, tanto de forma independiente como sinérgica con datos de expresión masiva global, para identificar mejor marcadores genéticos clave y subpoblaciones celulares distintas.

5. Para el último objetivo, pudimos integrar armoniosamente MSC aisladas *in vitro* con MSC obtenidas *in vivo* a partir de biopsias frescas de médula ósea (tanto de donantes sanos como analizadas mediante RNA-seq unicelular). Distinguimos dos subgrupos de células en las MSC *in vitro*, que estaban estrechamente asociadas con dos subpoblaciones distintas de las MSC *in vivo* (CAR1 y CAR2). Se realizaron análisis de enriquecimiento funcional de estas subpoblaciones y pudimos identificar en las MSC *in vivo* algunos de los nuevos marcadores genéticos que encontramos en nuestro estudio inicial con MSC *in vitro*: TAGLN (Transgelin) asociado con células CAR2; y COL4A1 y COL4A2 asociados con células CAR1. El trabajo destacó el potencial para identificar patrones de expresión genética comunes entre las subpoblaciones de células madre *in vivo* e *in vitro*.

ANNEX III: ARTICLE CHAPTER 2

Article

Deciphering Master Gene Regulators and Associated Networks of Human Mesenchymal Stromal Cells

Elena Sánchez-Luis ¹, Andrea Joaquín-García ¹, Francisco J. Campos-Laborie ^{1,2} ,
Fermín Sánchez-Guijo ³ and Javier De las Rivas ^{1,*} 

¹ Bioinformatics and Functional Genomics Group, Cancer Research Center (CiC-IMBCC, CSIC/USAL/IBSAL), Consejo Superior de Investigaciones Científicas (CSIC) and University of Salamanca (USAL), 37007 Salamanca, Spain; elenasl@usal.es (E.S.-L.); andreajoainguarcia@gmail.com (A.J.-G.); fjcamlab@gmail.com (F.J.C.-L.)

² Bioinformatics and Cancer genomics, Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, CB2 1QN Cambridge, UK

³ Cell Therapy Area and Department of Hematology, Institute of Biomedical Research of Salamanca -Hospital Universitario de Salamanca (IBSAL-HUS) and Department of Medicine, University of Salamanca (USAL), 37007 Salamanca, Spain; fermings@usal.es

* Correspondence: jrvivas@usal.es

Received: 30 December 2019; Accepted: 2 April 2020; Published: 5 April 2020



Abstract: Mesenchymal Stromal Cells (MSC) are multipotent cells characterized by self-renewal, multilineage differentiation, and immunomodulatory properties. To obtain a gene regulatory profile of human MSCs, we generated a compendium of more than two hundred cell samples with genome-wide expression data, including a homogeneous set of 93 samples of five related primary cell types: bone marrow mesenchymal stem cells (BM-MSC), hematopoietic stem cells (HSC), lymphocytes (LYM), fibroblasts (FIB), and osteoblasts (OSTB). All these samples were integrated to generate a regulatory gene network using the algorithm ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks; based on mutual information), that finds *regulons* (groups of target genes regulated by transcription factors) and *regulators* (i.e., transcription factors, TFs). Furtherly, the algorithm VIPER (Algorithm for Virtual Inference of Protein-activity by Enriched Regulon analysis) was used to inference protein activity and to identify the most significant TF regulators, which control the expression profile of the studied cells. Applying these algorithms, a footprint of candidate master regulators of BM-MSCs was defined, including the genes EPAS1, NFE2L1, SNAI2, STAB2, TEAD1, and TULP3, that presented consistent upregulation and hypomethylation in BM-MSCs. These TFs regulate the activation of the genes in the bone marrow MSC lineage and are involved in development, morphogenesis, cell differentiation, regulation of cell adhesion, and cell structure.

Keywords: mesenchymal stromal cells; transcription factor; regulons; master regulators; gene networks; transcriptomics; bioinformatic; meta-analysis

1. Introduction

Mesenchymal Stromal Cells (MSCs) are multipotent cells located in the stroma of multiple human tissues. In particular, they are present in the bone marrow (BM) hematopoietic niche, coexisting and regulating the maintenance of hematopoietic stem cells (HSCs). This BM niche also includes osteolineage cells (i.e., osteoblasts and osteoclasts), perivascular cells, endothelial cells, adipocytes, and macrophages. In this complex scenario, MSCs coordinate several critical activities, including self-renewal, mobilization, engraftment, and lineage differentiation [1].

When cultured *in vitro*, MSCs are characterized by their adherence to plastic; by their ability to differentiate *in vitro* to osteoblasts, adipocytes, and chondroblasts under specific culture conditions;

and by the expression of a characteristic immunophenotypic profile, being positive for CD73, CD90, and CD105 marker genes and negative for CD34, CD45, CD14 or CD11b, CD19 or CD79 α , and HLA-DR [2]. Close to MSCs, fibroblasts (FIBs) are non-stem cells that present a fairly similar phenotype to that of MSCs but differ in that they express other marker genes: CD10, CD26, CD106, and collagen VII (COL7A1). In addition, FIBs also share immunomodulatory properties with MSCs, such as the modulation of macrophages and the suppression of T cell proliferation [3]. Despite these similarities in phenotype, their transcriptomic signatures show significant differences between them, associating FIBs with a clear enrichment in genes related to the organization and function of the extracellular matrix and BM-MSCs in bone development tasks [4].

Due to their immunomodulatory ability, MSCs present great importance in the field of cellular therapy. Recent studies show the importance of the stimulation of MSCs to enhance their role in tissue repair and regeneration [5,6]. In addition, as it has been already mentioned, MSCs are being involved in the regulation of several relevant functions of the bone marrow. Nevertheless, it is still unclear how they perform such actions and what genes drive their activity at a regulatory level.

Given the relevance of MSCs in the BM niche, the present work pursues the identification of genes which regulate the most important functions of MSCs and compromise the MSC lineage, acting as master regulators of this cell type. The master regulators (MR) are defined as transcription factors (TF) that differentially regulate groups of target genes (called *regulons*). MRs regulate specific gene sets, activating or repressing their expression and, in consequence, activating or inhibiting the function of the corresponding proteins. To disclose the MRs of MSCs, we first selected and analysed a compendium of samples corresponding to primary human MSCs isolated from healthy donors and expanded in culture and the multiple cell types related to them. Over this compendium, after proper normalization and integration of the datasets, we applied ARACNe (*Algorithm for the Reconstruction of Accurate Cellular Networks*) and VIPER (*Algorithm for Virtual Inference of Protein-activity by Enriched Regulon analysis*) to infer dependency from similar expression patterns between target genes and TFs based on mutual information [7,8], to generate gene bipartite regulatory networks, and to identify master regulators. We also performed differential expression analysis and expression profiling to complement the data obtained with ARACNe and VIPER. DNA methylation profiles of MSCs were also analysed to validate the signal derived from the transcriptomic profiles. All these integrative analyses allow us to establish potential *master regulators* of MSCs in the hematopoietic niche. These candidate TF regulators and their associated gene sets, as *regulons*, define regulatory units that drive the mesenchymal lineage, providing the cellular characteristics of the MSCs and determining the specific functions that they play in the bone marrow.

2. Materials and Methods

2.1. Multiple Cell Sample Series with Transcriptomic Data Collected and Unified in a Compendium Set

For the transcriptomic analysis of this work, 18 datasets were used to create a large uniform compendium of 264 human samples with genome-wide expression data. All datasets were downloaded from public database GEO (Gene Expression Omnibus, www.ncbi.nlm.nih.gov/geo/). The IDs of the 18 datasets integrated are GSE2666, GSE3823, GSE6029, GSE6460, GSE7637, GSE7888, GSE9451, GSE9520, GSE9593, GSE9764, GSE9894, GSE10311, GSE10315, GSE10438, GSE11418, GSE12264, GSE18043, and GSE46053. The specific samples selected from each one of these datasets are described in Supplementary Table S1, which also indicates the authors and year of each set. In all cases, the transcriptomic profiles were obtained merging data of *Affymetrix* platforms HG-U133 A and B and from platform HG-U133 Plus 2.0, all corresponding to Human Genome high density oligo microarrays. The expression signals from the probes of these microarrays were mapped to genes (Ensemble genes (ENSG) done as described in Reference [9]), using as costume CDFs the R annotation packages from *BrainArray* version 23 (<http://brainarray.mbni.med.umich.edu>). As indicated in Supplementary Table S1, the biological samples were originally obtained from 10 different cell types: 47 samples of hematopoietic

stem cells (HSC), 10 of them isolated from bone marrow of healthy donors (BM-HSC); 9 samples of lymphocytes (LYM) as hematopoietic differentiated cells; 116 samples of mesenchymal stromal/stem cells (MSC) isolated from different tissues (50 isolated from bone marrow of healthy donors, BM-MSC); 27 MSCs stimulated with cytokines (stMSC), 6 of them stimulated with TGF β and selected for the comparison with MSCs; 11 samples of skin-derived primary fibroblasts (FIB); 13 primary osteoblasts (OSTB); 23 stimulated osteoblasts (stOST); 12 osteoblasts derived by differentiation from MSCs (dOSTB); 3 adipoblasts derived by differentiation from MSCs (dADIP); and 3 chondroblasts derived by differentiation from MSCs (dCHON). The transcriptomic signal from all these samples was normalized, and the batch effect was corrected as described in detail in an earlier publication of our laboratory [4]. In Supplementary Table S2, we also provide the given acronyms and the names of the cells included in the compendium, indicating the number of samples of each cell type, those which are primary cells and those which are derived from bone marrow. In particular, with respect to the 50 samples of BM-MSCs selected for our study, we checked that, in each corresponding GEO dataset, the samples were isolated using the standard protocol called “Minimal criteria for defining multipotent mesenchymal stromal cells” (from The international Society for Therapy position statement) (as indicated in Reference [10]). This means in practical terms that all samples selected correspond to primary MSCs from bone marrow of healthy donors isolated in culture (in pass 2–5) and characterized by the presence of specific CD surface markers: at least positive for CD73, CD90, and CD105 and negative for CD34 and CD45.

2.2. Regulatory Networks Based on Mutual Information

Target-TF regulatory networks were generated from the transcriptomic expression matrix obtained for the 264 samples and for the about 16,000 human genes measured. This expression matrix was analysed using the algorithm ARACNe (*Algorithm for the Reconstruction of Accurate Cellular Networks*), which applies information theory (Mutual Information (MI)) to calculate dependency between TFs and gene targets, avoiding many indirect interactions that are generally found through co-expression methods that are less accurate [7]. ARACNe was implemented using the R packages: *minet* and *parmigene* [11,12]. The MI values were filtered to select only the ones corresponding to the regulatory events that occur between Transcription Factors (TFs, considered *regulators*), and the linked genes (i.e., the targeted genes, considered as *regulons*). To map the TFs in the gene matrix, we used a comprehensive list that included 1544 *Homo Sapiens* transcription factors obtained from the database AnimalTFDB version 2.0. [13].

2.3. Differential Expression Between Six Types of Human Cells Related to Bone-Marrow MSCs

The normalized gene expression matrix was also analysed to obtain the differential expression (DE) between the MSCs and other 5 related cell types. Four of them were primary cells isolated from healthy individuals: HSC, LYM, FIB, and OSTB. The others were MSCs stimulated with TGF β (stMSC). Therefore, we created a subset of 93 samples, corresponding with 50 samples of BM-MSCs, 10 samples of HSCs, 9 samples of LYMs, 11 samples of FIB, 13 samples of OSTBs, and 6 samples of stMSCs. DE analyses were done using *limma* R package [14]. The comparisons were binary generating 6 groups: MSC-HSC, MSC-LYM, MSC-FIB, MSC-OSTB, MSC-stMSC, and stMSC-HSC. The selection of significant differentially expressed genes was done using a 5 % false discovery rate (FDR, that corresponded to adjusted *p*-value ≤ 0.05), plus the top 30 genes with most significant fold change (FC) in log₂ for downregulated).

2.4. Detection of Master Regulators in the Regulatory Networks

Master Regulators were detected using the algorithm called VIPER (*Virtual Inference of Protein-activity by Enriched Regulon analysis*), implemented in R and Bioconductor [8]. This computational algorithm allows an accurate assessment of protein activity from gene expression data. The method uses the expression matrix and the regulatory network provided by ARACNe to perform an enrichment statistical analysis on every *regulon* [8] and to identify the most significant TFs associated with the

regulatory models derived from the comparison of specific sample sets. Taking y as the value of MI returned by ARACNe, significant association TF-targets were filtered using as threshold = $mean(y)$. Furthermore, using VIPER, we compared the same groups of samples as we did with *limma*, that were MSC-HSC, stMSC-HSC and MSC-LYM (as cell types related to the hematopoietic niche); MSC-FIB (as cell types related to the stroma); and MSC-stMSC and MSC-OSTB (as cell types related to the mesenchymal lineage that can be originated from MSCs). Moreover, VIPER algorithm included bootstrapping, testing 100 times subsets of the samples to find and identify the most stable regulators. The algorithm also calculates the pleiotropy of the TFs [8]. From these analyses, we selected the most significant TF regulators found with a p -value < 0.05. The comparisons that provided the best signal corresponded to cell types related to the hematopoietic niche (MSC-HSC, stMSC-HSC, and MSC-LYM). In contrast, the comparison with FIB, OSTB, and stMSC did not find many significant regulators (only 4 upregulated TFs were found with a p -value < 0.05). The top 10 most overexpressed TFs and the top 10 most repressed TFs found significant, considering the 6 pairwise comparisons, were selected. The regulatory networks generated with the selected TFs were visualized using *Cytoscape* [15]. A complementary analysis based in the enrichment of TFs in the selected list of genes was done using *iRegulon* tool [16]. We applied this tool to the list of all the genes included the networks produced with VIPER. The tool *iRegulon* allows the identification of enrichment in specific *transcription factors binding sites* (TFBS) within the list of genes explored by comparative analysis of the sequences of the promoters of a query list of genes against curated datasets [16]. The tool also finds the corresponding TFs associated to the TFBSs.

2.5. Gene Set Functional Enrichment Analysis

The regulatory gene sets generated (that include each TF regulator and the corresponding regulated genes as regulons) were selected to perform functional enrichment analyses. The methods applied to do the enrichments were (i) DAVID bioinformatics tool (<https://david.ncifcrf.gov/>), that includes a functional annotation enrichment and clustering analysis [17], and (ii) GeneTerm-Linker bioinformatics tool (<http://gtlinker.cnbc.csic.es/>), that allows concurrent annotation and enrichment in several biological spaces in a unified way: GO Biological Process, GO Molecular Function, GO Cellular Component, KEGG Pathways, and InterPro Motifs and Domains [18].

2.6. DNA Methylation Analysis

To complement the gene regulatory analyses based on expression and transcriptomic profiling, we performed a global analysis of DNA methylation profiles of the CpG islands of the genes found as TF regulators in MSCs versus HSCs. This was done selecting and analysing three independent DNA methylation datasets of human bone marrow MSCs from healthy donors—GSE79695 (with 12 samples), GSE129266 (with 7 samples), and GSE87797 (with 6 samples)—compared with one DNA methylation dataset of human HSCs from healthy donors—GSE63409 (with 5 samples). The DNA methylation of all these samples was measured using Illumina Infinium HumanMethylation450 BeadChips (that corresponds to platform GPL13534 in the GEO database). This technology allows quantification of the global DNA methylation of the CpG islands across the genome based on the measurement of about 450,000 methylation sites per sample at single-nucleotide resolution. The raw data derived from these datasets was downloaded from GEO (<https://www.ncbi.nlm.nih.gov/geo/>) and analysed using algorithm *minfi* in R [19,20]. All these samples were preprocessed and integrated in a unified collection, after batch effect correction and normalization using the *combineArray* function from *minfi* [20], followed by the *preprocessIllumina* function also from *minfi* [19]. The technology of the Illumina HumanMethylation450 BeadChip consists of a two-color array that interrogates the methylation status of 485,512 methylation loci (mostly CpG sites but also a small number of cytosines outside of the CpG context), using bisulfite-converted DNA. For each methylation locus, two signals of interest are recorded: One signal measuring the amount of methylated DNA (Meth) and the other signal measuring the amount of unmethylated DNA (Unmeth). In principle, the proportion $Meth/(Unmeth +$

Meth) is the methylation ratio (referred to as *beta* value) in the population of cells from which the DNA was extracted [19]. Being Meth the methylated gene loci and Unmeth the unmethylated gene loci, the *beta* value is established within a range of 0 to 1 (where 0 corresponds to complete no-methylated status and 1 corresponds to complete methylated status). After the global normalization of the samples described and the calculation of the *Beta* values, we looked for the methylation signal of the CpG islands of the top 20 gene TF regulators that were found overexpressed or repressed in the BM-MSCs, and we compared such signals with the methylation profiles of HSCs.

3. Results and Discussion

3.1. Differential Gene Expression Profiling of MSCs versus Related Cell Types

To gain insight into the transcriptomic characterization of MSCs, genome-wide expression profiles were generated for these mesenchymal cells isolated from bone marrow and several related human primary cells: hematopoietic stem cells (HSCs) and lymphocytes (LYMs) as cells of the hematopoietic lineage; fibroblasts (FIBs) as cells of the stromal lineage; primary osteoblasts isolated from bone (OSTBs); several cell types derived by differentiation from MSCs (adipoblasts dADIPs, chondroblasts dCHONs, and osteoblasts dOSTBs); and finally MSCs stimulated with TGF β (stMSCs). After generating the global gene expression for the whole collection of 264 samples, as indicated in the Materials and Methods section, differential expression analyses were done for a subset of 99 samples to find the significant genes between 6 cell types of interest: MSC, HSC, LYM, FIB, OSTB, and stMSC. With these cells, the contrasts done were MSC-HSC, MSC-LYM, MSC-FIB, MSC-OSTB, MSC-stMSC, and stMSC-HSC. A differential expression signature of 188 genes corresponding to the top 30 most significant genes found in these six comparisons was produced, and these genes were used in the clustering expression analysis presented in Figure 1. The list of the 188 genes selected is also provided in Supplementary Table S3, that includes the statistical parameters corresponding to the differential expression analysis of MSCs versus HSCs. We also performed functional enrichment analyses of the selected set of 188 genes using GeneTerm-Linker and DAVID bioinformatic tools (as described in the Materials and Methods section). Gene expression clustering and gene functional association allow grouping the genes finding relevant links.

Upon examining specific genes in more detail, we found COL1A2, COL3A1, COL5A2, COL6A3, FN1, and CTGF among top 30 differentially expressed genes between MSCs and hematopoietic lineage (which includes HSCs and LYMs). These genes are common in extracellular matrix functions as well as in focal adhesion and skeletal system development and collagen binding. In fact, several of these genes take part in the PI3K-AKT pathway and interact with receptors and with the extracellular matrix. These interactions lead to direct or indirect control of cellular activities such as adhesion, migration, and differentiation. The differential signature of MSCs versus hematopoietic also contains numerous epithelial-mesenchymal transition (EMT) markers, such as the EMT-inducing gene SNAI2, and others: ACTA2, MMP2, and POSTN [21]. The genes COL5A2 and FN1 were also upregulated in stMSC and OSTB, indicating that upregulation of these genes is characteristic of the mesenchymal lineage. In the comparison between MSCs and stMSCs, it is important to consider that the stimulation of the cells (stMSCs) was done using TGF β (the transforming growth factor beta-3), that is a cytokine involved in cell differentiation, embryogenesis, and development. Moreover, this gene is involved in cellular adhesion and ECM formation during the process of human embryonic palate development and regulates the movements of epidermal and dermal cells in injured skin [21]. In this regard, A2M, APOD, COMP, CPE, DPT, PRELP, SERPINA3, SPARCL1, and SPP1 are genes found upregulated in stMSCs in the contrast with MSCs, and all of them are annotated to the functions of differentiation and epithelial-mesenchymal transition (EMT), showing a high expression due to the addition of TGF β [22,23]. Other genes such as FLG, IL6, MEST, PLK2, and VCAM1 were found upregulated in MSCs in contrast with stMSCs. These genes are involved in cell-cycle control. In particular, IL6 and PLK2 regulate JAK/STAT and p53 signalling pathways, respectively, to maintain the homeostasis and

quiescent cellular status [24,25]; this could be the reason for the upregulation of the MSC in contrast to stimulated or differentiated MSCs. The cluster linked to OSTBs, includes genes with functions related to immunoregulation, such as CSF2RB, IGDC4, NDNF, and VCAM1, and genes related with ECM, such as COL15A1, COL21A1, and ITIH5. Moreover, other genes in this cluster are EGFR2 and GREM1, which are genes necessary for the skeletal development and bone homeostasis and are also activated during bone cells response to mechanical strain [26,27].

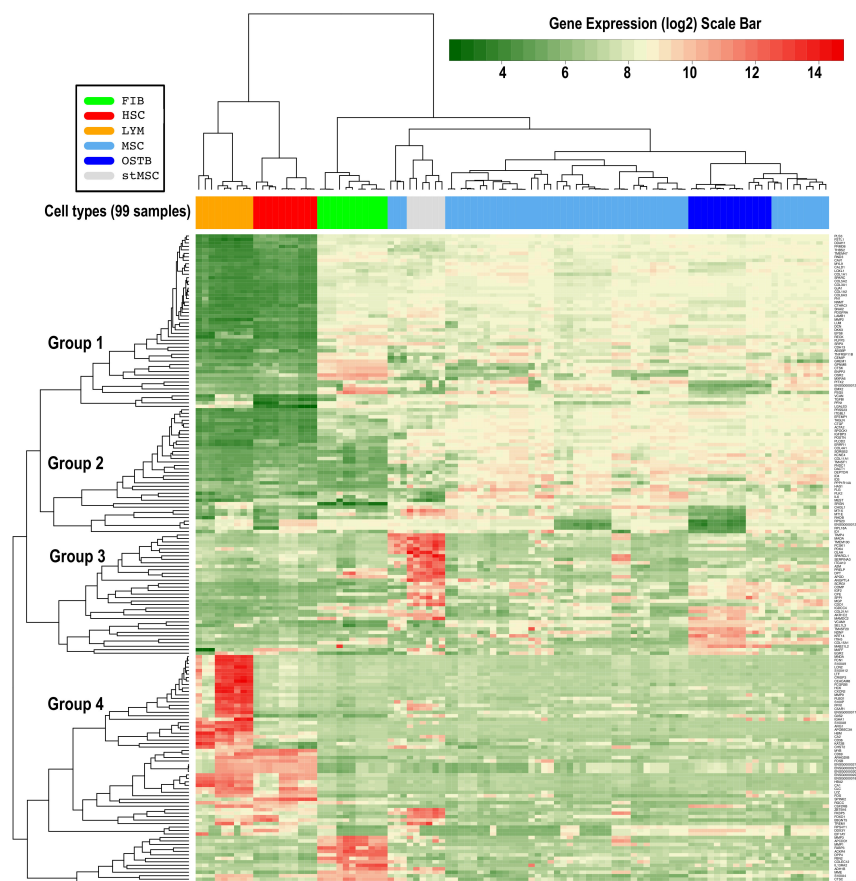


Figure 1. Heatmap presenting the expression profile of 188 genes found differentially expressed in the pair-wise comparisons of the six main cell types studied: Mesenchymal Stromal Cells (MSC), hematopoietic stem cells (HSC), fibroblasts (FIB), lymphocytes (LYM), osteoblasts (OSTB), and MSCs stimulated with cytokines (stMSC). The top 30 genes with most significant changes from each comparison were taken, and the union of these gave the list of 188 genes.

All the genes mentioned in this section are included in the heatmap presented in Figure 1. To facilitate the location of the genes, the dendrogram has been divided in 4 groups and the genes

in each group can be found in Supplementary Figure S4, which includes an enlarged version of the dendrogram with readable gene names.

In the comparison of MSCs with FIBs, even though significant differences were found for genes like CHI3L1, FNDC1, POSTN, SORBS2, SRGN, TM4SF1, and VCAM1; these genes are usually identified as expressed in fibroblasts, since they contribute to the structural integrity of the extracellular matrix, showing high similar functions in the stromal lineage (where FIBs and MSCs are included). Other genes, such as COL1A2, COL3A1, COL6A3, FN1, GJA1, and LUM were also present in both cell types, presenting a high expression in comparison with hematopoietic cell lineages. These genes are functionally related to focal adhesion, involving the PI3K-AKT signalling pathway and the extracellular matrix (ECM)-receptor interaction.

The heatmap (Figure 1) also includes other genes, such as ARHGBID, CD69, FOSB, HLA-DRB1 (ENSG00000206241), MYB, and SPINK2, which are a group of the most differentiated genes overexpressed in HSCs, forming an expression cluster associated to the HSCs. Moreover, many of them are consistently associated with gain and activation of hematopoietic-specific enhancers [28]. CD69 and HLA-DRB1 are genes that play a central role in the immune system and are constitutively expressed in T cells. This result is important since CD69 has been described as heterogeneous in HSCs, and our observation marks this gene as important for the mobilization and activation of stem cells in hematopoietic differentiation [29,30].

3.2. Identification of Master Regulators of MSCs and of Other Related Cell Types

To identify the possible role of the master regulators in MSCs, we used the ARACNe and VIPER algorithms [7,8]. The comparative analysis, based in VIPER, was done for the same sample groups used in the *limma* differential expression, that were MSC-HSC, stMSC-HSC, MSC-LYM, MSC-FIB, MSC-stMSC, and MSC-OSTB. The number of master regulators found in these 6 comparisons (five contrasts against MSCs and another contrast between stMSC and HSC) is presented in Figure 2A.

The analyses indicate that some comparisons detect many more changed TFs than others, revealing the dissimilarity or distance between the cell types. In this way, the results show major changes: (i) for MSC versus HSC (with 14 upregulators and 22 downregulators found after the bootstrapping analysis), (ii) for stMSC versus HSC (13 upregulators and 22 downregulators), and (iii) for MSC versus LYM (15 upregulators and 13 downregulators). By contrast, due to the similarities between MSC, FIB, OSTB, and stMSC, in the case of MSC versus FIB, only 3 upregulators and 10 downregulators appear, and this decreases more in the comparisons MSC-OSTB and MSC-stMSC, with 2 and 1 upregulators plus 5 downregulators, respectively (Figure 2A). The table also shows the number of TFs found considering pleiotropic effects due to the confluence of several genetic traits. This analysis greatly reduces the number of regulators since it implies cooperativity.

As a whole, the results identified the most relevant functional differences between the lineages mesenchymal and hematopoietic. Consequently, the genes found in the contrasts of MSCs versus HSCs were analysed in more detail. Figure 2B shows the top 10 gene upregulators—SNAI2, STAB2, IRX3, EPAS1, HOXC6, TWIST1, TULP3, PRRX1, TEAD1, and NFE2L1—and top 10 gene downregulators—BCL11A, MYB, TFEC, HLF, GATA2, ERG, PLAGL2, DACH2, POU2F1, and GATA3—common in all the three contrasts between these lineages (1, 2, and 3 in Figure 2A). Interestingly, the size of the regulon (N genes) indicates the number of target genes that were found associated with each master regulator. In the master upregulators, TEAD1 presents the largest number with 111 genes, followed by PRRX1 (57) and IRX3 (56). In the case of the master downregulators, POU2F1 (76) and GATA2 (70) have the largest number of associated genes. The table in Figure 2B also presents in which genes there is pleiotropy: SNAI2, EPAS1, and HOXC6 in the case of upregulators and BCL11A, MYB, HLF, PLAGL2, and GATA3 in the case of downregulators. With respect to the biological functions, the top 10 TFs upregulated in MSCs were involved mainly in morphogenesis and development functions and the top 10 TFs upregulated in HSCs presented mainly functions related to hematopoiesis, lymphocytes formation, and immune system regulation (see Table 1).

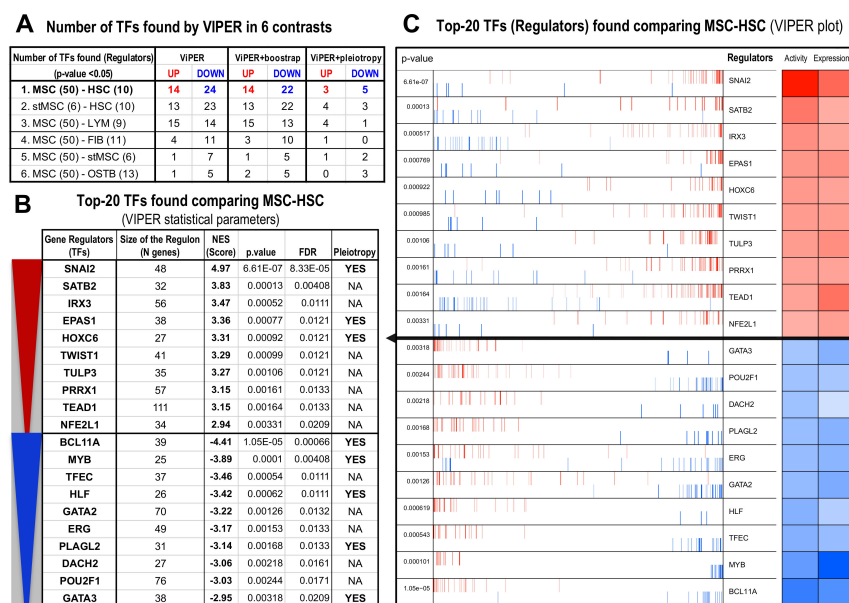


Figure 2. Gene regulators (TFs) and associated gene regulons found using VIPER (Algorithm for Virtual Inference of Protein-activity by Enriched Regulon analysis) in the comparison of the expression profiles of MSCs versus the other 5 related cell types, most of them present in the human bone marrow: (A) Table showing the number of TFs up- or downregulated and found in each comparison; (B) table with the top 10 up- and downregulated TFs found, including the parameters provided by VIPER (normalized enrichment score (NES), *p*-value, false discovery rate (FDR), and pleiotropy); and (C) plot produced by VIPER presenting the top 20 upregulated (red) and downregulated (blue) TFs that illustrates the strength of the protein activity and the RNA expression (darker colors higher values).

Figure 2C corresponds to the enrichment plot with the top 20 TF master regulators found in our analysis. The colour panels on the right represent (i) *expression*, the expression value of each TF in the dataset, and (ii) *activity*, the estimated protein activity corresponding to this regulator in the system. These parameters provide a measure of the importance of these genes to shape the biological characteristics of the MSCs and the mesenchymal lineage. SNAI2 is the most significant upregulator gene and has been described as a master TF in organogenesis and wound healing and is directly involved in the epithelial to mesenchymal transition (EMT) [31]. The highest activity in downregulators of MSCs was observed for BCL11A and MYB. The MYB gene is an important regulator of hematopoietic cell development and plays a central role in cell-cycle progression in B- and T-lymphoid progenitor cells [32].

3.3. Differential Expression Analysis of the MSCs Master Regulators

To support the characterization of the top 20 master regulators found with VIPER in the comparison of BM-MSCs with other cell types, we elaborated a parallel differential expression analysis (using *limma*) with the cells described before (performing the same 6 comparisons of cells, Figure 2A). In the results of these DE analyses, we searched for the changes of the 20 TFs found with VIPER. The numbers corresponding to the log₂ of the fold change (log₂FC) observed for each one of these 20 TFs are presented in Supplementary Table S5. The fold change represents a good measure of the intensity

of the changes. Corroborating the results, the gene *SNAI2* is found the most upregulated with a high differential expression in MSCs with respect to the hematopoietic lineage (compared to HSCs $\log_2FC = 5.93$ and compared to LYMs $\log_2FC = 6.49$), while *MYB* is found the most in the opposite case ($\log_2FC = -6.04$ compared to HSCs and $\log_2FC = -3.62$ compared to LYMs). In the case of the comparisons between cell types that belong to the mesenchymal or stromal lineage (which are MSCs, stMSCs, FIBs, and OSTBs), the top 20 TFs do not show large changes and the values of \log_2FC are always between 1 and -1 (i.e., close to 0). Such small changes reflect that these cell types are closer in biological genetic terms. For example, *PRRX1* is a paired-related homeobox (*PRRX*) transcription factor that regulates mesenchymal cell fate and stands at the centre of a network coordinating fibroblast differentiation; therefore, it is described as a regulator of FIBs that also regulates MSCs during the development [33]. By contrast, *IRX3* is not so similar for all the mesenchymal lineage cells, showing a clearer overexpression in MSCs with respect to FIBs ($\log_2FC = 2.13$).

Table 1. Functional enrichment analysis done with the top 10 master regulators (TFs) found upregulated (UP) and their corresponding gene regulons and with the top 10 master regulators (TFs) downregulated (DOWN) and their corresponding gene regulons. The genes assigned to each function are included in Supplementary Table S6.

Enriched Functional Term	N Genes (in the Function)	N Genes (in the Query)	N in Function/N in Query (%)	Regulation (UP/DOWN)	p-value (adj. Benjamini)
generation of neurons	26	283	7.34	UP	0.015168
neurogenesis	30	283	8.47	UP	0.001735
nervous system development	42	283	11.86	UP	0.007296
cell-substrate adhesion	10	283	2.82	UP	0.014732
extracellular matrix	22	323	6.21	UP	0.001305
cytoskeleton	51	323	14.41	UP	0.003200
face morphogenesis	4	283	1.13	UP	0.047319
embryonic development	29	283	8.19	UP	0.001836
organ morphogenesis	31	283	8.76	UP	0.000405
organ development	57	283	16.10	UP	0.016957
cell differentiation	61	283	17.23	UP	0.000488
T cell activation	17	293	5.00	DOWN	0.000004
lymphocyte activation	21	293	6.18	DOWN	0.000005
hemopoiesis	22	293	6.47	DOWN	0.000010
hemopoietic or lymphoid organ development	22	293	6.47	DOWN	0.000043
immune system development	22	293	6.47	DOWN	0.000074
regulation of immune system process	23	293	6.76	DOWN	0.002043
calponin-like actin-binding	10	323	2.94	DOWN	0.002577
actin filament-based process	19	293	5.59	DOWN	0.000509
cytoskeleton organization	26	293	7.65	DOWN	0.000664
GTPase regulator activity	24	304	7.06	DOWN	0.001932

Related to the TFs downregulated in MSCs, it may be interesting to remark the case of *HLF*, which acts as a negative activator, because *HLF* is a relevant TF in some leukemia due to *TCF3-HLF* fusion protein, which suppresses *RUNX1* transcription and activates expression of *LMO2* and several *Groucho-related* genes as well as antiapoptotic genes like *SNAI2* [34].

3.4. Methylation Profiles of the MSCs Master Regulators

As an independent validation of the top 20 MSC master regulators found in the gene regulatory analysis performed with VIPER and corroborated by the differential expression analysis performed with *limma*, we analysed the methylation profiles of the promoters of these TF genes to verify whether they correlated with their relative expression signal. To do this, as outlined in the Materials and Methods section, we used three independent DNA methylation profiles of BM-MSCs and compared them with a DNA methylation profile obtained for HSCs. After adequate robust normalization of the datasets (as described in the Materials and Methods section), differential methylation of the CpG islands of these genes was determined and compared as presented in Figure 3. The results provided

relative hypomethylation for the genes EPAS1, NFE2L1, SATB2, SNAI2, TEAD1, and TULP3 from our list of upregulated TFs and relative hypermethylation for ERG, GATA2, GATA3, HLF, MYB and POU2F1 from our list of TF downregulated in BM-MSCs (Figure 3).

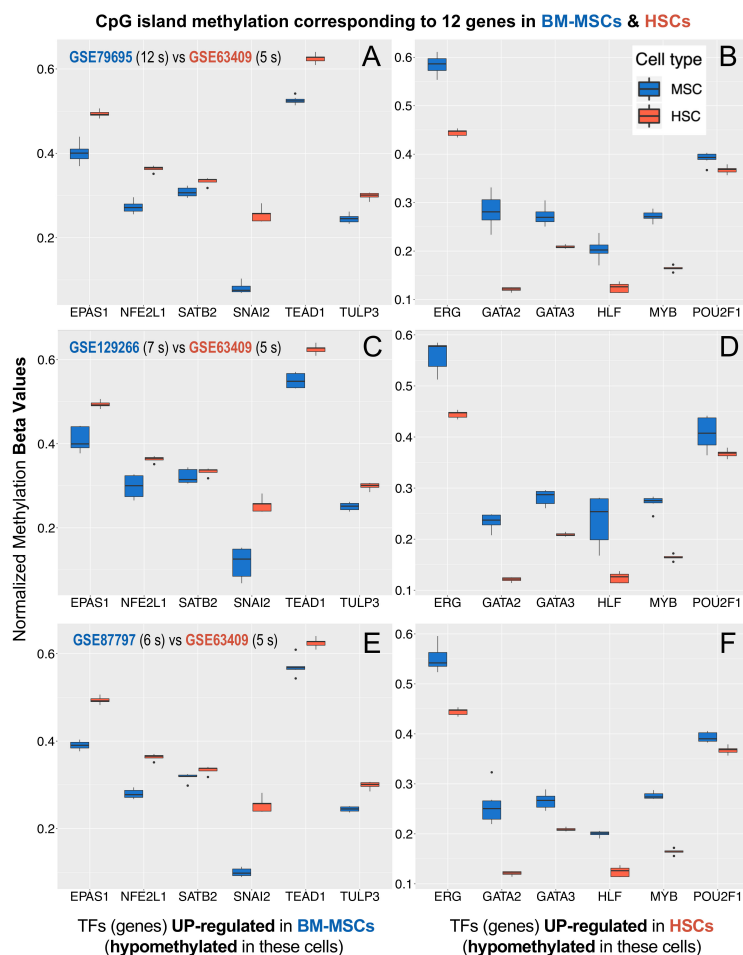


Figure 3. DNA methylation levels (measured as Beta values) of 12 TFs (blue boxplots) in 3 independent datasets of BM-MSCs (A,C,E) (GSE79695, GSE129266, GSE87797) in comparison with the methylation levels of the same TFs (red boxplots) in a dataset of HSCs (B,D,F) (GSE63409) (s indicates the number of samples). The TFs represented are: EPAS1, NFE2L1, SATB2, SNAI2, TEAD1, TULP3 hypomethylated in BM-MSCs; and ERG, GATA2, GATA3, HLF, MYB, POU2F1 hypermethylated.

The methylation results provide special support to the SNAI2 gene as a master regulator of BM-MSCs. A study with fibroblasts showed that the associated proximal promoters of SNAI1 and SNAI2 genes were hypomethylated due to EMT, being characteristic of stromal cells [35]. Furthermore, the association between DNA methylation and transcription levels for the SNAI2

gene has been demonstrated in iPSC (induced pluripotent stem cells) generated from fibroblasts [35]. Hypomethylation of TEAD1 gene, also found in our analysis, has been reported to be a requirement for the ability of MSCs to undergo proper differentiation [36]. These studies also suggested that aberrant DNA hypermethylation of the loci of genes of the TEAD family could compromise their role in the development of BM-MSCs and could promote malignant diseases originated in the bone marrow, such as multiple myeloma (MM) [36]. EPAS1 gene was also detected in our results hypomethylated in BM-MSCs, and this gene showed a decrease in repressed chromatin marks in mesenchymal stromal cells isolated from bone marrow and adipose tissue, suggesting that epigenetic mechanisms are probably involved in determining the stem potential of MSCs via this gene [37]. In addition, our data also showed a decreasing in methylation of the gene locus of SATB2 (special AT-rich sequence-binding protein 2). In this sense, regulatory studies have shown that the activity of SATB2 is modulated epigenetically and that, when this gene shows decrease in methylation, this triggers bone marrow stromal cells osteogenic differentiation, facilitating bone formation and regeneration [38]. All the reports referenced in this section and our results in the search for master regulators confirm that epigenetic regulation of gene expression is a central mechanism that governs cell stemness, determination, commitment, and differentiation. Taken together, these results reinforce our findings of candidate master regulators of BM-MSCs.

3.5. Construction of Bipartite Networks Including Regulators and Regulons

Using the links found with ARACNe between each gene regulator and its target genes, we can build regulatory networks including all the most stable and significant regulators and their genes, marking also the expression level detected. These regulatory networks are represented as a bipartite graph that includes two types of nodes (regulator TFs and regulon targets) plus directed links between them (TF to Target). Figures 4 and 5 present two views of these type of networks: first (Figure 4) including the top 20 regulators (10 up- and 10 downregulated) linked to their gene regulons (presenting the nodes in red when they are upregulated and in blue when they were downregulated); second (Figure 5) including only the top 10 upregulated TFs linked to their regulons (again presenting the upregulated nodes in red and the downregulated nodes in blue).

In both networks, the intensity of the colour of the nodes is proportional to the expression signal values of the corresponding genes. In the case of the second network (Figure 5), 6 other TFs were found using the *iRegulon* tool over the entire list of genes and TFs of VIPER (as indicated in the Materials and Methods section) [16].

The network in Figure 4 shows the central regulatory role of SNAI2. This gene regulates important genes (such as CEBPB, EBF1, ERG, MYB, TGIF1, and ZMAT1), which are TFs in charge of DNA-binding transcriptional activation and transcription regulation. Other meta-regulators are TEAD1, which regulates TFs related to phosphorylation such as COPS6, KUF1BP, MYO6, and RGL2; and TULP3, which shows the regulation of TFs as CCND1, LAMB2, and THBS1, and regulates the PI3K/AKT signalling pathway that performs a critical role in regulating diverse cellular functions including metabolism, growth, proliferation, survival, transcription, and protein synthesis as well as in extracellular matrix organization, ECM-receptor interactions, cell adhesion, and integrin binding.

With respect to the TFs found repressed, MYB shows the strongest signal both in the regulatory network derived from VIPER and in the differential expression analysis, and it was found hypomethylated in HSCs. MYB regulates TFs related to cellular transport (such as AP2M1, ATP8A1, KDELR1, LAPTM5, TMED10, and UCP2) and regulates other TFs responsible of the cell proliferation (such as EPS8, SYK, and YAP1). The role of TFs that are downregulated in MSCs versus HSCs may reveal a positive action or a more relevant function in the hematopoietic lineage.

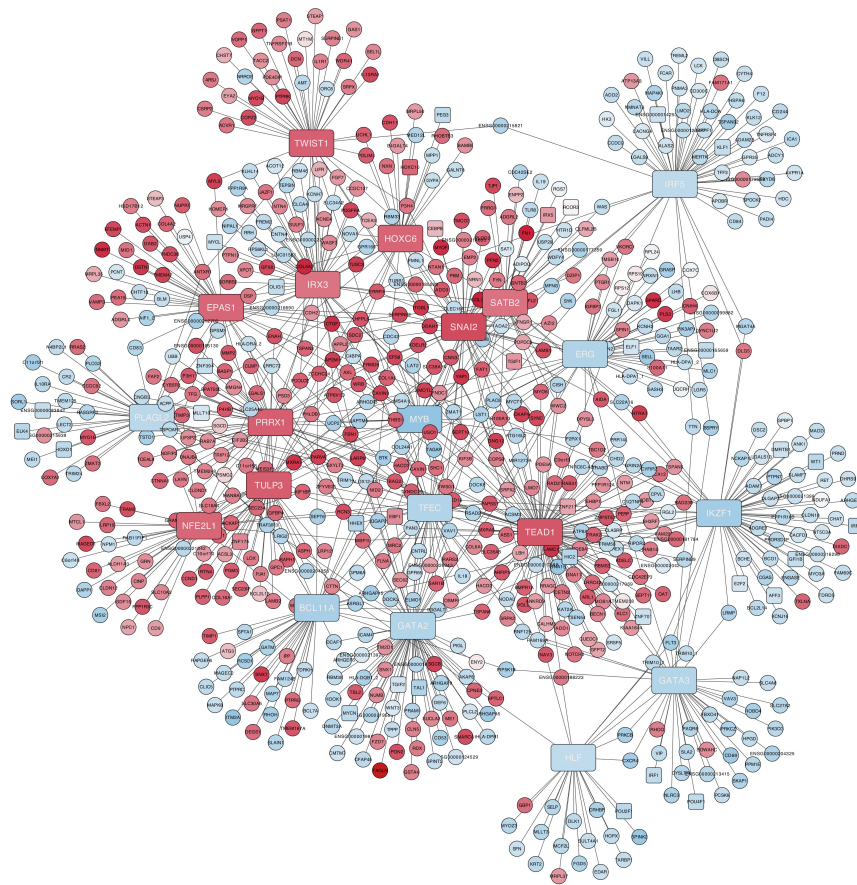


Figure 4. Gene co-regulation network presenting the top 10 up- and top 10 downregulated master regulators (in red and blue rectangles, respectively): The master regulators are connected with their regulons. Upregulated genes have a red background, while downregulated genes have a pale blue color.

Figure 5 presents the regulatory network produced by selecting only the top TF regulators that were overexpressed in MSCs. These TFs are presented in large red rectangles in the figure. Those marked with a purple frame were the ones that also showed a low methylation level in the analysis of the CpG islands of these genes in BM-MSCs (as shown in Figure 3). Furthermore, other 6 TFs (E2F1, EP300, GADD45A, MAFK, TCF12, and TEAD4) are included in Figure 5 in yellow ellipses and correspond to the TFs found enriched in the promoters of all the genes of this network. A search of public data revealed that E2F1, EP300, GADD45A, and SPL1 are highly expressed in bone marrow and that E2F1, EP300, and GADD45A are genes expressed in response to the detection of DNA damage and can cause a the reduction of the cell cycle rate, characteristic of the quiescent state of stem cells.

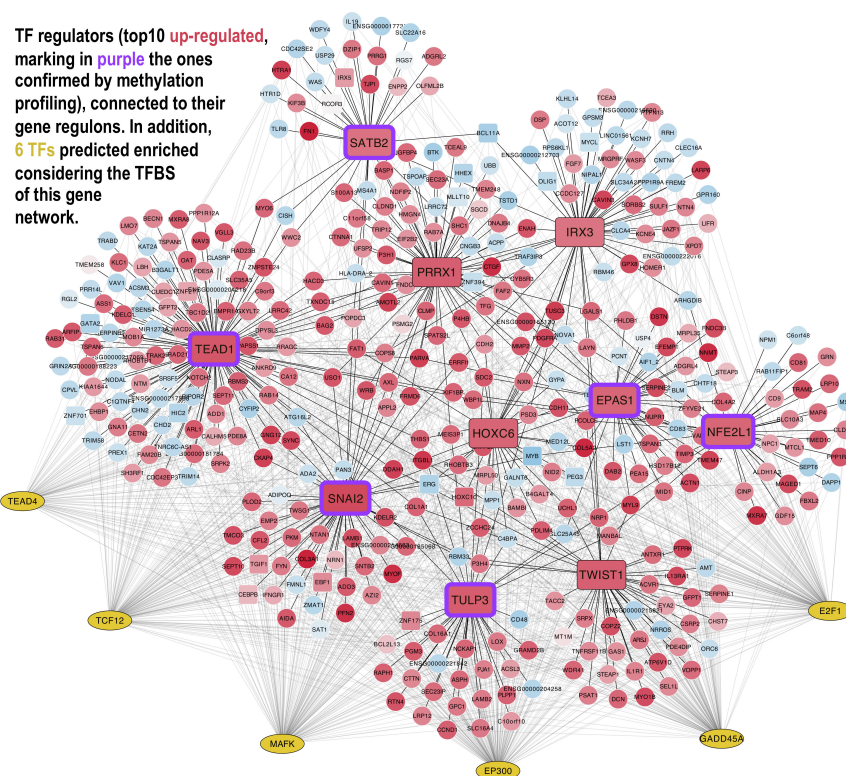


Figure 5. Gene co-regulation network presenting the top 10 upregulated master regulators (red rectangles): The master regulators are connected with their regulons that are the gene sets that each one regulates: Upregulated genes have a red background, while downregulated genes have a pale blue color. The 6 TFs (E2F1, EP300, GADD45A, MAFK, TCF12, and TEAD4) included in yellow ellipses correspond to the TFs found enriched with the *iRegulon* tool in the promoters of all the genes of the network.

3.6. Functional Enrichment Analysis of the Master Regulators and Their Regulons

The results of the functional enrichment analyses are presented in Table 1. The analysis was done first with the list of upregulated TFs and their gene regulons (marked UP in Table 1) and second with the list of downregulated TFs and their gene regulons (marked DOWN in Table 1). The upregulated TFs and their regulons show first a significant enrichment in functions associated to the nervous system: generation of neurons, neurogenesis, and nervous system development, being IRX3, SATB2, TULP3, and TWIST1 the master regulators included in these functions (the complete list of all the genes associated to each enriched function is included in Supplementary Table S6). The classical differentiation paths of MSCs are adipogenesis, chondrogenesis, and osteogenesis, but recent studies show that MSCs have also the plasticity to differentiate into cells of ectodermic origin like neurocytes [39]. Moreover, TULP3 regulates hedgehog signalling pathway and promotes the development of multipotent neural crest progenitors endowed with both mesenchymal and neural potentials [40]. Functional enrichment on organ morphogenesis and development is the second most significant biological trait found in the regulators and regulons overexpressed in MSCs. These biological features involve again as master

regulators SATB2 and TULP3 and bring together EPAS1, TEAD1, and SNAI2. Endothelial PAS domain Protein 1 (EPAS1) promotes adipose differentiation and is a TF specific in endothelial cells as an important regulator of vascularization. These functions are related to the role and action of MSCs inside the bone marrow, indicating that the regulators found are essential to the function of MSCs [41]. All these TFs were found to be overexpressed and hypomethylated in MSCs, being postulated as key master regulators of this cell lineage.

Regarding the top 10 master regulators that are downregulated in MSCs and upregulated in HSCs, many of them have hematopoietic functions: BCL11A and IKZF1 are strongly related to hematopoiesis and to the immune system. IKZF1 was described as a regulator of gene expression and chromatin remodelling, playing an important role in the correct development of the immune system and acting as an important tumour suppressor in lymphoblastic leukaemia (ALL) [42]. BCL11A is also important in hematopoiesis, with a particular role in B-cell development and in the maintenance of stemness in HSCs. Besides, BCL11A is highly expressed in the initial phases of myeloid and lymphoid malignancies, indicating that a high level of BCL11A can cause leukaemia cells' continuous replication, blocking differentiation [43]. All these data suggest that the master regulators found repressed are responsible for the maintenance of the hematopoietic cell lineage.

4. Conclusions

Transcription factors are known to maintain stemness and to drive differentiation of cell lineages. Our study presents a selection of relevant master regulators that define human primary MSCs, classifying different groups of related cell types and providing a transcriptomic footprint with the most relevant gene regulators and regulons. Likewise, we found links between the MSCs and gene-regulating extracellular matrix functions as well as other functions, such as adhesion, migration and differentiation, and maintenance of the BM niche. Many of these functions are also directly related to the epithelial–mesenchymal transition (EMT), a process by which epithelial cells lose their cell polarity and cell–cell adhesion and gain migratory and invasive properties to become mesenchymal stem cells or cells of the mesenchymal lineage. Our finding of SNAI2 as one of the master regulators of MSCs gives strong support to the functional link of these cells with the EMT.

This article also highlights the importance of the specific links and relationships between master regulators, forming networks, to clarify the complexity and cooperation between them beyond the individual regulation of each one in MSCs. As an example of these interactions, SNAI2, SATB2, and TULP3 have been identified as a group of relevant upregulated master regulators, being involved in the maintenance of MSCs through Hedgehog and PI3K/AKT signalling pathways, that are essential to induce stem cell traits, immunosuppression, senescence, drug resistance, and metastasis. With respect to previous studies, TULP3 has not been directly related to MSCs. By contrast, SNAI2 has previously been associated with MSC, involved in ECM organization, and functionally associated with EMT. Among the downregulated master regulators, MYB and BCL11A have been related to immune system regulation in hematologic malignancies, suggesting their participation in the maintenance of the hematopoiesis and in the regulation of other downregulated master regulators such as IKZF1. Finally, a complementary epigenetic study was carried out to obtain the DNA methylation profiles of the TFs found, which corroborated the gene expression profiles and gave support to EPAS1, NFE2L1, SATB2, SNAI2, TEAD1, and TULP3 as candidate positive master regulators of MSCs.

To summarise, the work presents a set of transcription factors associated to the mesenchymal lineage as well as their direct links with other regulators and other genes, deciphering the regulatory networks of the human BM-MSCs. For future work, we are interested in developing new trials and tests with the top 10 overexpressed master regulators found in this work to investigate their modulation in different contexts and to better understand the dynamic behaviour of MSCs within the hematopoietic niche, with a particular focus on their immunomodulatory properties.

Supplementary Materials: The following files are available online at <http://www.mdpi.com/2218-273X/10/4/557/s1>. Supplementary Table S1: List of human cells used in this study that correspond to 18 datasets including a compendium of 264 human samples with genome-wide expression data: All datasets were downloaded from the public database GEO (Gene Expression Omnibus, www.ncbi.nlm.nih.gov/geo/). The IDs of the 18 datasets integrated are GSE2666, GSE3823, GSE6029, GSE6460, GSE7637, GSE7888, GSE9451, GSE9520, GSE9593, GSE9764, GSE9894, GSE10311, GSE10315, GSE10438, GSE11418, GSE12264, GSE18043, and GSE46053. The specific number of samples selected from each one of these datasets are indicated within the table. Supplementary Table S2: List of 10 human cell types used in this study, indicating the ones that were primary cells, stimulated cells, or cells derived from MSCs. The set of 93 samples is the one included in the comparisons done with VIPER algorithm for regulation and with *limma* algorithm for differential expression. Supplementary Table S3: List of 188 genes included in the heatmap presented in Figure 1 of the article: The table presents for these genes the results of the differential expression analysis of BM-MSCs (50 samples) versus HSCs (10 samples) done with *limma*: log2 fold-change; mean expression of the gene across all the samples; F-statistic; raw *p*-value; and adjusted *p*-value. This list of genes was composed selecting the top 30 genes that provided the most significant differences in each one of the 6 comparisons of cell types performed in this work: MSC-HSC, stMSC-HSC, MSC-LYM, MSC-FIB, MSC-stMSC, and MSC-OSTB. Supplementary Figure S4: Dendrogram divided into 4 groups corresponding to the genes included in the heatmap of Figure 1, presented in an enlarged version with readable gene names to facilitate the location of each gene. Supplementary Table S5: Results of the differential expression (DE) analysis done with *limma* corresponding to the top 20 gene regulators (TFs) found in this work: The data correspond to the DE parameters (F-statistic, raw *p*-values, and adjusted *p*-values) of the comparison of MSCs versus HSCs. The log2FoldChanges of all the 6 comparisons (MSC-HSC, stMSC-HSC, MSC-LYM, MSC-FIB, MSC-stMSC, and MSC-OSTB) are also included. The mean expression of each of the 20 genes in all samples is also presented. Supplementary Table S6: Functional enrichment analysis (i) done with the top 10 master regulators (TFs) found upregulated (UP) and their corresponding gene regulons and (ii) done with the top 10 master regulators (TFs) downregulated (DOWN) and their corresponding gene regulons. The genes marked in bold in the column “Genes” correspond to TFs.

Author Contributions: Conceptualization, J.D.L.R. and F.S.-G.; methodology, E.S.-L., F.J.C.-L., and J.D.L.R.; software development, application, and validation, E.S.-L., A.J.-G., and F.J.C.-L.; formal analysis, investigation, resources and data curation, E.S.-L., A.J.-G., and F.J.C.-L.; writing—original draft preparation, E.S.-L. and J.D.L.R.; writing—review and editing, E.S.-L., F.J.C.-L., F.S.-G., and J.D.L.R.; supervision and project administration, F.S.-G. and J.D.L.R.; funding acquisition, J.D.L.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by *Fondo de Investigación Sanitaria—Instituto de Salud Carlos III* (FIS—ISCIII, Spanish Ministry of Health, project reference P18/00591) where J.D.L.R. was the PI. The research was also partially funded by *Fondo de Investigación Sanitaria* (FIS—ISCIII, project reference P116/01407). The work within these projects was also cofunded by the FEDER program of the European Union.

Acknowledgments: J.D.L.R. lab acknowledges also the support given by the European Project H2020 ArrestAD (Project ID: 737390, H2020-FETOPEN-1-2016-2017).

Conflicts of Interest: All the authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

BM, Bone Marrow; CDF, Chip Description File; CpG, CG sequence islands in the genome; GO, Gene Ontology; HSC, Hematopoietic Stem Cell; KEGG, Kyoto Encyclopedia of Genes and Genomes; MSC, Mesenchymal Stromal/Stem Cell; TF, Transcription Factor.

References

1. Yu, V.W.C.; Scadden, D.T. Hematopoietic Stem Cell and Its Bone Marrow Niche. *Current Top. Dev. Biol.* **2016**, *118*, 21–44. [[CrossRef](#)]
2. Horwitz, E.M.; Le Blanc, K.; Dominici, M.; Mueller, I.; Slaper-Cortenbach, I.; Marini, F.C.; Keating, A. Clarification of the nomenclature for MSC: The International Society for Cellular Therapy position statement. *Cytotherapy*. **2005**, *7*, 393–395. [[CrossRef](#)] [[PubMed](#)]
3. Denu, R.A.; Nemcek, S.; Bloom, D.D.; Goodrich, A.D.; Kim, J.; Mosher, D.F.; Hematti, P. Fibroblasts and Mesenchymal Stromal/Stem Cells Are Phenotypically Indistinguishable. *Acta Haematol.* **2016**, *136*, 85–97. [[CrossRef](#)] [[PubMed](#)]
4. Roson-Burgo, B.; Sanchez-Guijo, F.; Del Cañizo, C.; De Las Rivas, J. Insights into the human mesenchymal stromal/stem cell identity through integrative transcriptomic profiling. *BMC Genom.* **2016**, *17*, 944. [[CrossRef](#)]

5. Waterman, R.S.; Henkle, S.L.; Betancourt, A.M. Mesenchymal Stem Cell 1 (MSC1)-Based Therapy Attenuates Tumor Growth Whereas MSC2-Treatment Promotes Tumor Growth and Metastasis. *PLoS ONE* **2012**, *7*, e45590. [[CrossRef](#)]
6. Najjar, M.; Krayem, M.; Meuleman, N.; Bron, D.; Lagneaux, L. Mesenchymal stromal cells and toll-like receptor priming: A critical review. *Immune Netw.* **2017**, *17*, 89–102. [[CrossRef](#)]
7. Margolin, A.A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Favera, R.D.; Califano, A. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **2006**, *7*, S7. [[CrossRef](#)]
8. Alvarez, M.J.; Shen, Y.; Giorgi, F.M.; Lachmann, A.; Ding, B.B.; Ye, B.H.; Califano, A. Network-based inference of protein activity helps functionalize the genetic landscape of cancer. *Nat. Genet.* **2016**, *48*, 838–847. [[CrossRef](#)]
9. Risueño, A.; Fontanillo, C.; Dinger, M.E.; De Las Rivas, J. GATEExplorer: Genomic and transcriptomic explorer; mapping expression probes to gene loci, transcripts, exons and ncRNAs. *BMC Bioinform.* **2010**, *11*, 221. [[CrossRef](#)]
10. Dominici, M.; Le Blanc, K.; Mueller, I.; Slaper-Cortenbach, I.; Marini, F.; Krause, D.; Deans, R.; Keating, A.; Prockop, D.J.; Horwitz, E. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **2006**, *8*, 315–317. [[CrossRef](#)]
11. Meyer, P.E.; Lafitte, F.; Bontempi, G. Minet: A R/bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinform.* **2008**, *9*, 461. [[CrossRef](#)] [[PubMed](#)]
12. Sales, G.; Romualdi, C. Parmigene—a parallel R package for mutual information estimation and gene network reconstruction. *Bioinformatics* **2011**, *27*, 1876–1877. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, H.M.; Liu, T.; Liu, C.J.; Song, S.; Zhang, X.; Liu, W.; Guo, A.Y. AnimalTFDB 2.0: A resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **2015**, *11*, 321. [[CrossRef](#)] [[PubMed](#)]
14. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *7*, e47. [[CrossRef](#)] [[PubMed](#)]
15. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Nucleic Acids Res.* **2003**, *13*, 2498–2504.
16. Janky, R.; Verfaillie, A.; Imrichová, H.; van de Sande, B.; Standaert, L.; Christiaens, V.; Aerts, S. iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput. Biol.* **2014**, *10*, e1003731. [[CrossRef](#)] [[PubMed](#)]
17. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)]
18. Fontanillo, C.; Nogales-Cadenas, R.; Pascual-Montano, A.; De Las Rivas, J. Functional analysis beyond enrichment: Non-redundant reciprocal linkage of genes and biological terms. *PLoS ONE* **2011**, *6*, e24289. [[CrossRef](#)]
19. Aryee, M.J.; Jaffe, A.E.; Corrada-Bravo, H.; Ladd-Acosta, C.; Feinberg, A.P.; Hansen, K.D.; Irizarry, R.A. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **2014**, *30*, 1363–1369. [[CrossRef](#)]
20. Fortin, J.P.; Triche, T.J., Jr.; Hansen, K.D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **2017**, *33*, 558–560. [[CrossRef](#)]
21. Cheng, W.-Y.; Kandel, J.; Yamashiro, D.; Canoll, P.; Anastassiou, D. Slug-based epithelial-mesenchymal transition gene signature is associated with prolonged time to recurrence in glioblastoma. *Nat. Preced.* **2011**. [[CrossRef](#)]
22. Herpin, A.; Lelong, C.; Favrel, P. Transforming growth factor- β -related proteins: An ancestral and widespread superfamily of cytokines in metazoans. *Dev. Comp. Immunol.* **2004**, *28*, 461–485. [[CrossRef](#)] [[PubMed](#)]
23. Zhai, L.J.; Zhao, K.Q.; Wang, Z.Q.; Feng, Y.; Xing, S.C. Mesenchymal stem cells display different gene expression profiles compared to hyaline and elastic chondrocytes. *Int. J. Clin. Exp. Med.* **2011**, *4*, 81–90. [[PubMed](#)]
24. Xu, C.; Sun, L.; Jiang, C.; Zhou, H.; Gu, L.; Liu, Y.; Xu, Q. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. *Biomed. Pharmacother.* **2017**, *91*, 1167–1177. [[CrossRef](#)]

25. Ward, A.; Hudson, J.W. p53-dependent and cell specific epigenetic regulation of the polo-like kinases under oxidative stress. *PLoS ONE* **2014**, *9*, e87918. [[CrossRef](#)]
26. Heinrich, P.C.; Behrmann, I.; Haan, S.; Hermanns, H.M.; Müller-Newen, G.; Schaper, F. Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem. J.* **2003**, *374*, 1–20. [[CrossRef](#)]
27. Canalis, E.; Parker, K.; Zanotti, S. Gremlin1 is required for skeletal development and postnatal skeletal homeostasis. *J. Cel. Physiol.* **2012**, *227*, 269–277. [[CrossRef](#)]
28. Zaman, G.; Sunters, A.; Galea, G.L.; Javaheri, B.; Saxon, L.K.; Moustafa, A.; Lanyon, L.E. Loading-related regulation of transcription factor EGR2/Krox-20 in bone cells is ERK1/2 protein-mediated and prostaglandin, Wnt signaling pathway-, and insulin-like growth factor-i axis-dependent. *J. Biol. Chem.* **2012**, *289*, 25509. [[CrossRef](#)]
29. González, A.J.; Setty, M.; Leslie, C.S. Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. *Nat. Genet.* **2015**, *47*, 1249–1259. [[CrossRef](#)]
30. Bujanover, N.; Goldstein, O.; Greenspan, Y.; Turgeman, H.; Klainberger, A.; Scharff, Y.; Gazit, R. Identification of immune-activated hematopoietic stem cells. *Leukemia* **2018**, *32*, 2016–2020. [[CrossRef](#)]
31. Ganesan, R.; Mallets, E.; Gomez-Cambrotero, J. The transcription factors Slug (SNAI2) and Snail (SNAI1) regulate phospholipase D (PLD) promoter in opposite ways towards cancer cell invasion. *Mol. Oncol.* **2016**, *10*, 663–676. [[CrossRef](#)] [[PubMed](#)]
32. Lorenzo, P.I.; Brendeford, E.M.; Gilfillan, S.; Gavrilov, A.A.; Leedsak, M.; Razin, S.V.; Gabrielsen, O.S. Identification of c-Myb Target Genes in K562 Cells Reveals a Role for c-Myb as a Master Regulator. *Genes Cancer* **2011**, *2*, 805–817. [[CrossRef](#)] [[PubMed](#)]
33. Froidure, A.; Marchal-Duval, E.; Ghanem, M.; Gerish, L.; Jaillet, M.; Crestani, B.; Mailleux, A. Mesenchyme associated transcription factor PRRX1: A key regulator of IPF fibroblast. *Eur. Respir. J.* **2016**, *48*, OA506.
34. Tjichon, E.; Havinga, J.; Van Leeuwen, F.N.; Scheijen, B. B-lineage transcription factors and cooperating gene lesions required for leukemia development. *Leukemia* **2013**, *27*, 541–552. [[CrossRef](#)] [[PubMed](#)]
35. Chen, Y.; Wang, K.; Qian, C.N.; Leach, R. DNA methylation is associated with transcription of Snail and Slug genes. *Biochem. Biophys. Res. Commun.* **2013**, *430*, 1083–1090. [[CrossRef](#)]
36. Garcia-Gomez, A.; Li, T.; Rodríguez-Ubreva, J.; Ciudad, L.; Català-Moll, F.; Martín-Sánchez, M.; San-Segundo, L.; Morales, X.; Ortiz de Solórzano, C.; Oyarzabal, J.; et al. Targeting aberrant DNA methylation in mesenchymal stromal cells as a treatment for myeloma bone disease. *bioRxiv* **2019**, 767897. [[CrossRef](#)]
37. Aranda, P.; Agirre, X.; Ballestar, E.; Andreu, E.J.; Román-Gómez, J.; Prieto, I.; Martín-Subero, J.I.; Cigudosa, J.C.; Siebert, R.; Esteller, M.; et al. Epigenetic signatures associated with different levels of differentiation potential in human stem cells. *PLoS ONE* **2009**, *4*, e7809. [[CrossRef](#)]
38. Han, Q.; Yang, P.; Wu, Y.; Meng, S.; Sui, L.; Zhang, L.; Yu, L.; Tang, Y.; Jiang, H.; Xuan, D.; et al. Epigenetically modified bone marrow stromal cells in silk scaffolds promote craniofacial bone repair and wound healing. *Tissue Eng. Part A* **2015**, *21*, 2156–2165. [[CrossRef](#)]
39. Krampera, M.; Marconi, S.; Pasini, A.; Galiè, M.; Rigotti, G.; Mosna, F.; Bonetti, B. Induction of neural-like differentiation in human mesenchymal stem cells derived from bone marrow, fat, spleen and thymus. *Bone* **2007**, *40*, 382–390. [[CrossRef](#)]
40. Calloni, G.W.; Glavieux-Pardanaud, C.; Le Douarin, N.M.; Dupin, E. Sonic Hedgehog promotes the development of multipotent neural crest progenitors endowed with both mesenchymal and neural potentials. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19879–19884. [[CrossRef](#)]
41. Tian, H.; McKnight, S.L.; Russell, D.W. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev.* **1997**, *11*, 72–82. [[CrossRef](#)] [[PubMed](#)]
42. Payne, K.J.; Dovat, S. Ikaros and tumor suppression in acute Lymphoblastic leukemia. *Crit. Rev. Oncog.* **2011**, *16*, 3–12. [[CrossRef](#)] [[PubMed](#)]
43. Yin, J.; Zhang, F.; Tao, H.; Ma, X.; Su, G.; Xie, X.; Yin, B. BCL11A expression in acute phase chronic myeloid leukemia. *Leuk. Res.* **2016**, *47*, 88–92. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).