Investigating the functional role of lncRNA *CASC20* in heterotopic bone formation

Favour Felix-Ilemhenbhio

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Division of Clinical Medicine

School of Medicine and Population Health

September 2023

Acknowledgment

I would like to express my heartfelt gratitude to my supervisors, Professors J Mark Wilkinson and Endre Kiss-Toth, for their continuous and exceptional guidance, support, and enthusiasm. Their supervision was invaluable throughout both my MRes and PhD studies. I couldn't have asked for better mentors for my doctoral journey. I also extend my thanks to my supervisor from Newcastle University, Professor David Young, for their unwavering availability and mentorship. Special appreciation goes to Dr Matt Barter (Newcastle University) and the Young team for providing me with invaluable lentiviral training during my time in Newcastle. In addition, I am grateful to Professor John Loughlin of Newcastle University for generously donating the ASC52Telo and ASC52TeloSOX9 cells, which were used in chapters 4 and 6 of my thesis. I acknowledge the funding provided by CIMA and their support in fostering collaborations among research groups in the captivating field of Musculoskeletal Ageing.

I wish to extend my sincere appreciation to the EKT/HLW joint group, especially Dr Heather L Wilson, Charlotte Moss and Saff Foster for their support. I would like to thank the Sudbery group, with special recognition for Sumeet Deshmukh and Charlotte Vandermeulen, who provided me with essential bioinformatics training. I would like to extend my gratitude to Drs Klaudia Kocsy, Chiara Niespolo, Laura Campesino, and Taewoo Kim, who were esteemed members of the IICD during my academic journey. Although they left the department up to two years ago, our connection has remained strong, and their support has been unwavering. My gratitude also goes to Professor Christine Le Maitre for all the chats we had about *in vitro* chondrogenic and osteogenic conditions. I extend my thanks to the IICD department, the IICD Student Society, the OM department, and the AUBM group for their contributions and support.

I am profoundly thankful to my family, my parents and my sisters, Onoseta and Onomhen, for their unwavering love, positivity, and support throughout my academic journey. Many thanks to my beautiful wife, Lorenza. We have been best friends since my first year of university in 2014, and she has been a great support throughout my academic career. Marrying her on the 25th March 2023 was one of the best days yet. Thanks to my mother-in-law, Mercy, whose radiant smile and tenacity continue to illuminate my life. I am also grateful to my friends, Hassan and Kerry Hassanpour, and my god-daughter, Jasmine Hassanpour, for their friendship and encouragement. Many thanks to my golf buddy and fellow EKT/HLW group member, Josh Norwood. Lastly, I offer a special thanks to God for blessing me with these wonderful people.

I dedicate this thesis to my parents, for their encouragement, sacrifices, and support they have shown me over the years. I wouldn’t have made it without them.

Thank you all!

“A person who never made a mistake never tried anything new.” Albert Einstein

COVID Impact Statement

The work contained within this thesis was conducted in part during the COVID-19 pandemic. Much of the second six months of my first year of studies coincided with the first COVID-19 lockdown and its following severe lab restrictions. As a result, I lost a large part of my first-year lab activity, which resulted in a change of focus during that period to in-silico work that is presented in Chapter 2. The first COVID-19 lockdown also gave me the opportunity to write the now-published literature review that is presented in Chapter 1. At the end of my studies, I also returned to review the in-silico predictions by comparing the predicted micro-RNA targets against those found in in-vitro experiments. This comparison is presented at the end of chapters 5 and 7.

I would also like to make the examiners aware that the human P512MSC primary cells described in Chapters 5 to 7 are actually murine in origin. This identification came to light only after all of the experimental work was complete and the results chapters were written. The P512 MSC cells originally came from a patient undergoing joint replacement under the care of my first supervisor, Mark Wilkinson. The cells were shared with Sheffield Hallam University (SHU) as part of a project studying human bone-derived MSCs. When we encountered problems with the high passage number of the hMADs we were using for our experiments, a flask of the P512MSC cells was shared back with us from SHU. These cells behaved in the same way as we expected, differentiating into both chondrocytes and osteoblasts and thus the thought of possible contamination did not occur. Because of some cell number and RNA quality issues with our initial sequencing run of the final experiments described in Chapters 5 and 7, we made a choice to run miR Seq only in order to complete the miRome story for my thesis. The sequencing data generated matched to the human genome (hg38), as expected, and the studies were completed. However, after this work was done my lab sent off a set of duplicate samples to Novogene for mRNA sequencing. The returned data underwent FastQC on 31.08.2023 and returned a match to the mouse genome (GRCm39) and not the human genome. A subsequent investigation conducted in September 2023 identified that work was going on in the same SHU group at the time on a mouse pre-osteoblast cell line. We can only conclude that although the flask shared was labelled P512MSCs, a cross-contamination episode took place, and the mouse cell line outgrew the human primary cells. Further sequencing of the original flask shared back with us for karyotyping is currently being conducted to further clarify.

Once this contamination was identified, my supervisors discussed the situation with our PGR team. It was decided that the best course of action within the University of Sheffield regulations at this late stage was to leave the written chapters “as is”, but to document the findings, their consequences, and future work in the context of the actual cell type studied. A full description of this is given in Chapter 8.5 P512MSCs cell population and implications. The practical consequence of this late finding has been that our P512 studies represent a murine late MSC cell line CASC20 knock-in model (given CASC20 is a human-only gene), rather than a

human primary cell CASC20 over-expression model. Given that the miRome is highly conserved between mice and humans, it is unlikely that the CASC20 miR targets are substantially different between the species. The subsequent prediction work on the downstream mRNA target genes is also likely to be similar. Whilst repeating this work is out of the scope of this PhD, a first step for my following PhD colleague will be to replicate my findings in confirmed human-derived MSCs in an over-expression model.

Print name: Favour Felix-Ilemhenbhio

Date: 30.09.2023

Abstract

Heterotopic Ossification (HO) is a lamellar bone formation in the soft tissues and usually follows injury, trauma or joint replacement. A genome wide association study of HO patients after total hip arthroplasty identified lncRNA Cancer Susceptibility 20 (*CASC20*) as a gene that is strongly associated with HO severity. Previous findings from my MRes demonstrated an unreported upregulation of *CASC20* during BMP2-induced osteodifferentiation of hMSCs. Based on these studies, I hypothesised that *CASC20* is a novel regulator of bone formation.

This doctoral thesis investigated the role of *CASC20* in regulating osteogenic and chondrogenic differentiation processes. Using a diverse range of experimental models, including hMADs, ASC52teloSOX9, and P512MSCs, this research explored *CASC20's* impact on key osteogenic and chondrogenic markers. Techniques such as *in silico* miR prediction, *CASC20* lentiviral expression, RT-qPCR, and miR-Seq were used to assess the effect of *CASC20* on microRNAs (miRs), genes and biological pathways.

Here I found that *CASC20* consistently played a pivotal role in the early stages of osteogenesis and chondrogenesis. *CASC20*'s modulation of miRs emerged as a consistent regulatory mechanism throughout the experiments. The study identified 13 genes commonly targeted by the putative *CASC20*-interacting miRs, including well-known osteogenic and chondrogenic players such as MAPK1.

To confirm the findings, future steps include genotyping human mesenchymal stem cells and conducting comprehensive experiments to validate the observed effects. The creation of human CRISPR *CASC20* KO models will further elucidate *CASC20*'s role. In conclusion, this doctoral thesis provides valuable insights into the intricate mechanisms governed by *CASC20* in osteogenesis and chondrogenesis. Further research and exploration are needed to comprehensively elucidate the extent of *CASC20*'s influence in these critical biological processes.

Declaration

I, Favour Felix-Ilemhenbhio, confirm that the thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

Chapter 1 describes a literature review that I authored as part of the work of this thesis. I contributed 90% of this manuscript by carrying out data curation, formal analysis, original draft preparation, and design of the figures and tables. The co-authors listed at the beginning of that paper contributed as outlined there. Elsewhere in my thesis I had support from Sumeet Deshmukh, who provided me with training in the bioinformatics necessary for me to conduct the *in-silico* analyses presented in Chapter 2. Charlotte Vandermeulen gave me training in matching the MSC miR expression data to the human and mouse genomes and helped me to develop the scripts to perform the differential expression analyses described in chapters 5 and 7. All other data presented in this thesis are exclusively my own work.

Associated Publications

The publications below are associated with the work presented in this thesis.

- Pickering, G.A.E, **Felix-Ilemhenbhio F**, et al. The Kinesin Gene KIF26B Modulates the Severity of Post-Traumatic Heterotopic Ossification. Int. J Mol Sci. 2022 Aug 16;23(16):9203. doi:10.3390/ijms23169203.
- **Felix-Ilemhenbhio F**, Pickering GAE, Kiss-Toth E, Wilkinson JM. Pathophysiology and Emerging Molecular Therapeutic Targets in Heterotopic Ossification. Int J Mol Sci. 2022 Jun 23;23(13):6983. doi: 10.3390/ijms23136983.
- **Felix-Ilemhenbhio F**, Deshmukh SR, Sudbery I, Kiss-Toth E, Wilkinson JM. Functional Interactions of the Long Non-Coding (LNC) RNA *CASC20* in Heterotopic Ossification. Orthop Procs. 2021;103-B(SUPP_16):66-66. doi:10.1302/1358-992X.2021.16.066
- Hatzikotoulas K, Pickering GA, Clark MJ, **Felix-Ilemhenbhio F,** et al. Genome-wide association and functional analyses identify *CASC20* and KIF26B as target loci in heterotopic ossification. bioRxiv; 2019. doi:10.1101/845958.

Associated Presentations

I presented results arising from the thesis in the meetings below.

- IICD Drip Meeting: Functional interactions of *CASC20* in Heterotopic Ossification – Royal Hallamshire Hospital, Sheffield. 3rd March 2023. (Oral)
- CIMA-CMAR Joint Meeting: Functional interactions of *CASC20* in Endochondral Ossification - Birmingham, UK. 22nd-24th May 2023. (Poster)
- Orthopaedic Research Society (ORS) Annual Meeting 2023: Functional interactions of *CASC20* in Endochondral Ossification - Dallas, Texas. 10th-14th Feb 2023. (Poster)
- Mellanby Centre Research Day 2022: Functional interactions of *CASC20* in Heterotopic Ossification - The University of Sheffield, Sheffield. 25th March 2022. (Poster)
- Orthopaedic Research Society (ORS) Annual Meeting 2022: Functional interactions of *CASC20* in Heterotopic Ossification - Tampa, Florida. 4th-8th Feb 2022. (Poster)
- AUBM Group Meeting: Functional interactions of *CASC20* in Heterotopic Ossification – Northern General Hospital, Sheffield. 11th October 2021. (Oral)
- IICD Drip Meeting: Functional interactions of *CASC20* in Heterotopic Ossification – Royal Hallamshire Hospital, Sheffield. 1st October 2021. (Oral)
- The British Orthopaedic Research Society (BORS) Annual Meeting 2021: Functional interactions of *CASC20* in Heterotopic Ossification – Held Online. 13th-14th September 2021. (Oral)
- Orthopaedic Research Society (ORS) Annual Meeting 2021: Functional interactions of *CASC20* in Heterotopic Ossification – Held Online. 12th-16th Feb 2021 (Poster)
- CIMA Annual Meeting 2019: Investigating the functional role of lncRNA *CASC20* in Heterotopic ossification – The Principal Hotel, York. 14th-15th October 2019. (Oral)
- Mellanby Centre Internal Seminar: Is the long non-coding RNA *CASC20* driving Heterotopic ossification? – The University of Sheffield, Sheffield. 7th June 2019 (Oral)
- Blast Injury Conference 2019: Does the lncRNA *CASC20* modulate Heterotopic ossification? – Imperial College, London. 11th-12th July 2019 (Oral)
- IICD Drip Meeting: Exploring *CASC20* and KIF26B expression and functional role in Heterotopic ossification. – Royal Hallamshire Hospital, Sheffield. 26th July 2019 (Oral).

Table of Contents

List of Figures

List of Tables

Abbreviations

ACVR1/ALK2	Activin A receptor type I/activin receptor-like kinase 2
ALP	Alkaline phosphatase
AON	Antisense oligonucleotide
AS	Ankylosing spondylitis
ASC52telo	hTERT immortalised human adipose-derived mesenchymal stem cells
ASC52teloSOX9	SOX9-overexpressing ASC52telo
ASCs	Adipose-derived stem cells
ASP-RNAi	Allele-specific siRNA
BMP	Bone morphogenetic protein
BP	Biological processes
caACVR1	Constitutively activated ACVR1/ALK2
CAS9	CRISPR associated protein 9
CASC20	Cancer susceptibility 20
CC	Cellular compartments
ceRNAs	Competitive endogenous RNAs
CHD	Chromodomain helicase DNA-binding
CLIP studies	UV cross-linking and immunoprecipitation
COL1A1	Collagen type I alpha 1 chain
COL2A1	Collagen, type II, alpha 1
COMP	Cartilage oligomeric matrix protein
CRIPSR	Clustered regularly interspaced short palindromic repeats
DMEM	Dulbecco's modified eagle medium
DMMB	1,9-dimethyl methylene blue
DNMT1	DNA methyltransferase 1
DSB	Double-stranded break
EA	Effect allele
ECM	Extracellular matrix
ENCORI	Encyclopedia of RNA Interactomes
EndMT	Endothelial-mesenchymal transition
FAPs	Fibro-adipogenic precursors
FBS	Foetal bovine serum
FGF	Fibroblast growth factor
FOP	Fibrodysplasia ossificans progressiva
FVB/NJ mice	Friend leukemia virus B/ NIH Jackson mice
GADD45	Growth arrest and DNA-damage-inducible 45 alpha
GAGs	Glycosaminoglycans
GO	Gene ontology
GWAS	Genome-wide association study
HATs	Histone acetyltransferases
hBMSCs	Human bone-marrow stem cells
HDACs	Histone deacetylases
HDMs	Histone demethylases

HIFs	Hypoxia-inducible factors
hMADS	Human multipotent adipose-derived stem cell
hMSCs	Human mesenchymal stem cells
HMTs	Histone methyltransferases
HO	Heterotopic ossification
INO80	Inositol-requiring protein 80
IRES	Internal ribosome entry sites
ISW1	Imitation SWI/SNF
KEGG	Kyoto encyclopaedia of genes and genomes
KIF26B	Kinesin family member 26 B
KO	Knockout
LD	Linkage disequilibrium
lncRNA	Long non-coding RNA
LSD1	Histone lysine demethylase 1
MAPK	Mitogen-activated protein kinase
MEG3	Maternally expressed gene 3
MRSA	Methicillin-resistant Staphylococcus aureus
MF	Molecular functions
miR	Micro-RNA
mTOR	Mammalian target of rapamycin
NATs	Natural antisense transcripts
NSAID	Non-steroidal anti-inflammatory drug
NSE-BMP4	Neuron-specific enolase-BMP4 mouse
OCN/BGLAP	Osteocalcin
OE	Overexpression
OM	Osteogenic induction media
OSX	Osterix
P/S	Penicillin-streptomycin
P512MSCs	Primary human explant MSCs from patient 512
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGFR	Platelet-derived growth factor receptor
PDLSCs	Periodontal ligament stem cells
PMOP	Postmenopausal osteoporosis
POH	Progressive osseous heteroplasia
PRC2	Polycomb repressive complex 2
PTMs	Post-translational modifications
RA	Retinoic acid
RAR	Retinoic acid receptor
RARE	RA-response elements
RAR-RXR	Retinoid X receptor
rhBMP2	Recombinant BMP2
RNAi	RNA interference
RT-qPCR	Real-time quantitative polymerase chain reaction

RUNX2	Runt related transcription factor 2/Core-binding factor-α
SAM	S-adenosylmethionine
s-IBM	Sporadic inclusion-body myositis
SMAD	SMA/MAD homology proteins
SNP	Single nucleotide polymorphism
SNVs	Single nucleotide variants
SOX	SRY-related HMG box
SP cells	Hematopoietic side-population cells
SUMO	Small ubiquitin-like modifier
SWI/SNF	Switch/sucrose non-fermentable
TALE	Transcription activator-like effector
TDMD	Target-directed mir degradation
TET	Ten-eleven translocation
TFs	Transcription factors
TGF	Transforming growth factor
THA	Total hip arthroplasty
TSS	Transcription start site
uORFs	Upstream open reading frames
UTRs	Untranslated regions
WT	Wild type

Chapter 1 – Introduction

1.1 Outline of the PhD Thesis

The term Heterotopic ossification (HO) describes abnormal bone formation in soft tissues where bone is normally not found. Here, using the publication format, this thesis presents my research focused on the molecular basis of common, complex HO. Chapter 1 presents a literature review that I conducted as part of my PhD to appraise the depth and breadth of current research into the cellular and molecular basis of HO, and to discuss investigational molecular therapeutic targets. This review was published in 2022 and is presented in the introduction as it discusses the background to the PhD project. Next, I introduce the gene of interest of this PhD, the long non-coding RNA (lncRNA) *CASC20*. In this section, I present the research that preceded the PhD: 1) An investigation into the heritable biology of HO after total hip arthroplasty (THA) using Genome-wide Association Study (GWAS), which identified *CASC20* as a genetic susceptibility locus. 2) My CIMA MRes research project in which I characterised the expression of *CASC20* in human waste bone samples and studied its expression during the osteodifferentiation of human mesenchymal stem cells. *CASC20* is expressed in normal bone and is upregulated during the osteodifferentiation of mesenchymal stem cells stimulated with BMP2. I then introduce the functional biology of lncRNA and micro-RNA (miR) in HO, and the hypothesis and aims of the PhD.

The subsequent chapters describe experimental work conducted in the course of my PhD. In Chapter 2, I discuss in-vitro studies conducted during the initial COVID-19 lockdown showing that *CASC20* may interact with miRs that are enriched for osteogenic and chondrogenic function during bone formation. In Chapter 3, I present preliminary studies aimed to provide insight into *CASC20* function in osteoblast differentiation of human multipotent adipose-derived stem cells (hMADs). In Chapter 4, the functional role of *CASC20* lentiviral overexpression in osteoblast differentiation is explored using hMADs, immortalised adipose-derived stem cells (ASCs) and P512-hBMSCs (human bone-marrow stem cells isolated from a patient undergoing joint replacement for osteoarthritis), together with upscaling experiments to collect sufficient material for total RNA sequencing. In Chapter 5, miR-sequencing of datasets generated by chapter 4 experiments is used to examine the effect of *CASC20* overexpression on the osteodifferentiation miRome and to predict downstream target genes. In Chapter 6, the functional role of *CASC20* lentiviral overexpression in chondrocyte differentiation is explored using hMAD, ASCs and P512-hBMSCs. In Chapters 7, miR-sequencing of datasets generated by chapter 6 experiments is used to examine the effect of *CASC20* overexpression on chondrogenesis miRome and to predict downstream target genes. Chapter 8 follows as general discussion, future work and conclusions.

The GWAS data together with the expression data gathered in the CIMA MRes was previously posted in bioRxiv. The functional analysis of *KIF26B,* a genetic locus that was found to be

associated with HO severity was published in Int J Mol Sci in 2022. Where relevant, these manuscripts are cited in the thesis and are included in its appendices.

1.2 Review: Pathophysiology and Emerging Molecular Therapeutic Targets in Heterotopic Ossification

Favour Felix-Ilemhenbhio [1,2], George A. E. Pickering [1], Endre Kiss-Toth [2] and J. Mark Wilkinson [1,*]

[1] Department of Oncology and Metabolism, The University of Sheffield, The Medical School, Beech Hill Road, Sheffield S10 2RX, UK;
ffelixilemhenbhio1@sheffield.ac.uk (F.F.-I.);
georgeaepickering@gmail.com (G.A.E.P.)
[2] Department of Infection, Immunity and Cardiovascular Disease, The University of Sheffield, Sheffield, UK; e.kiss-toth@sheffield.ac.uk
* Correspondence: j.m.wilkinson@sheffield.ac.uk

Author Contributions:

Conceptualisation, J.M.W. and E.K.-T.; Methodology, F.F.-I. and G.A.E.P.; Formal Analysis, F.F.-I.; Data Curation, F.F.-I.; Writing—Original Draft Preparation, F.F.-I. and G.A.E.P.; Writing—Review & Editing, J.M.W. and E.K.-T.; Visualisation, F.F.-I.; Supervision, J.M.W. and E.K.-T.; Funding Acquisition, J.M.W. All authors have read and agreed to the published version of the manuscript.

Abstract: The term heterotopic ossification (HO) describes bone formation in tissues where bone is normally not present. Musculoskeletal trauma induces signalling events that in turn trigger cells, probably of mesenchymal origin, to differentiate into bone. The aetiology of HO includes extremely rare but severe, generalised and fatal monogenic forms of the disease; and as a common complex disorder in response to musculoskeletal, neurological or burn trauma. The resulting bone forms through a combination of endochondral and intramembranous ossification, depending on the aetiology, initiating stimulus and affected tissue. Given the heterogeneity of the disease, many cell types and biological pathways have been studied in efforts to find effective therapeutic strategies for the disorder. Cells of mesenchymal, haematopoietic and neuroectodermal lineages have all been implicated in the pathogenesis of HO, and the emerging dominant signalling pathways are thought to occur through the bone morphogenetic proteins (BMP), mammalian target of rapamycin (mTOR), and retinoic acid receptor pathways. Increased understanding of these disease mechanisms has resulted in the emergence of several novel investigational therapeutic avenues, including palovarotene and other retinoic acid receptor agonists and activin A inhibitors that target both canonical and non-canonical signalling downstream of the BMP type 1 receptor. In this article I illustrate the key cellular and molecular mechanisms involved in the pathogenesis of HO and outline recent advances in emerging molecular therapies to treat and prevent HO that have had early success in the monogenic disease and are currently being explored in the common complex forms of HO.

Keywords heterotopic ossification; genetics; bone morphogenetic protein; activin A/ALK2; retinoic acid receptor; Hoxa11+ mesenchymal stromal cells

Citation: Felix-Ilemhenbhio, F.; Pickering, G.A.E.; Kiss-Toth, E.; Wilkinson, J.M. Pathophysiology and Emerging Molecular Therapeutic Targets in Heterotopic Ossification.
Int. J. Mol. Sci. 2022, 23, 6983. https://doi.org/10.3390/ijms23136983

Academic Editor: José Manuel López

Received: 26 May 2022

Accepted: 22 June 2022

Published: 23 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1.2.1 Introduction

Heterotopic ossification (HO) is a disorder characterised by bone development within tissues where bone does not normally exist. Several presentations of HO have been described since its early documentation in 1883 by Riedel and its first association with musculoskeletal trauma in World War One combatants in 1918 (1). There are two forms of HO traditionally described: the rare 'genetic disease', and the more common acquired, or 'post-traumatic', HO. The monogenic HO diseases, which follow a Mendelian pattern of inheritance, include fibrodysplasia ossificans progressiva (FOP) and progressive ossific heteroplasia (POH). FOP is a rare debilitating disease with a prevalence of 1-2 cases per million persons in which muscle and connective tissues are gradually substituted by bone that is commonly triggered by minor trauma events (2, 3). POH is an extremely rare disease affecting less than 60 people

worldwide, (4, 5) in which ossification develops initially in the deeper layers of the dermis and subcutaneous fat and spreads to include muscle and tendons as the disease progresses. Both diseases are associated with progressive disability and early death (6). The term acquired, or "post-traumatic" HO describes extra-skeletal bone formation that occurs following musculoskeletal or neurological trauma and burns (7). Acquired HO occurs in 20-30% of patients with spinal cord injury (8), 10-20% of patients with closed head injury (8), up to 50% of patients after total hip replacement (9), and up to 70% of patients following high-energy combat trauma (10). The present review provides an overview of our current understanding of the molecular biology of HO initiation and development, including the cellular and genetic origins of HO. Based on these molecular advances in our understanding of the disease, I also review the current status of evolving molecular therapies for HO prevention and treatment. Throughout the article, I use the term HO to describe acquired HO and the terms FOP and POH to describe the specific monogenic disorders.

1.2.2 Overview of Normal Bone Formation

In order to understand the mechanisms of bone formation in HO, a brief review of normal bone formation is given against which HO development will be compared. Normal mature bone is formed through one of two mechanisms, termed intramembranous and endochondral ossification. The progenitor cell for both processes is the mesenchymal precursor, but the mechanism and site at which ossification occurs differs (reviewed in (11, 12)). In intramembranous ossification, a sheet of mesenchymal connective tissue, termed the fibrous membrane, forms the template of the future bone. Bones forming through this mechanism are typically flat, including the cranium, sternum, ribs, and scapula. The mesenchymal precursor cells differentiate into osteoblasts or into supporting blood vessels. The osteoblasts secrete osteoid, an extracellular matrix comprising collagen and other organic proteins that entraps the osteoblasts as the osteoid mineralises. Once entrapped, the osteoblasts trans-differentiate into osteocytes that remain as mechanosensing cells within the bone matrix. Osteoblasts on the surface of the bone transdifferentiate to form a cellular layer termed the periosteum. The periosteum is responsible for cortical bone synthesis, and envelopes the cancellous bone that is continuous with the haematopoietic red bone marrow. In endochondral ossification, bone formation occurs through an intermediate, cartilaginous stage that serves as a template for the final bone. The long bones, including the clavicle, humerus, radius, ulna, metacarpals, phalanges, femur, fibula, tibia, metatarsals, and phalanges form through endochondral ossification. The process commences as mesenchymal stem cells condense and differentiate into chondrocytes to form the cartilage template. This is followed by hypertrophy and subsequent apoptosis of the central cells, whilst mesenchymal progenitors at the template surface differentiate into osteoblasts and osteoclasts. The hypertrophic and apoptotic cartilage core is innervated, vascularised, and replaced by bone and bone marrow in the primary ossification centre. At the developing bone metaphysis, a hypertrophic component of the growing cartilage is constantly substituted by trabecular bone to mediate longitudinal bone growth. The non-vascularised cartilage at the ends of the bone

is invaded by epiphyseal vessels to initiate the secondary ossification centre. Between the epiphyseal and metaphyseal bone centres reside layers of chondrocytes that form growth plates to further support longitudinal growth. Longitudinal growth stops as the growth plate is fully resorbed to leave a single marrow cavity within the long bone.

1.2.3 Cellular Origins of HO

In HO development, following the initiating stimulus bone may form within a range of extraosseous tissues of mesenchymal origin and may involve either of the above mechanisms of bone formation. The architectural features of HO resemble normal bone and include a zonal mineralisation pattern, mature cortical bone at the periphery, and a central marrow component (13). Chalmers et al. (1975) first proposed the basic common requirements for HO formation: osteogenic precursor cells, a permissive environment and an inducing agent (14). This model is consistent with HO formation through either the endochondral or intramembranous routes. A summary of current concepts of the cellular origins of HO is given below and is reviewed further elsewhere (15, 16).

1.2.3.1 Hematopoietic Cells

In the 1970s, hematopoietic stem cells or other precursors recruited to the lesions from bone marrow were suggested to contribute to the induction and formation of ectopic bone in patients with FOP (17). Lymphocytes taken from FOP patients were subsequently shown to overexpress Bone Morphogenetic Protein 4 (BMP4), a potent bone-inducing protein (18). Olmsted-Davis and colleagues investigated hematopoietic side-population (SP) cells as possible precursors for HO (19). These cells were known to possess multi-lineage potential, with the ability to differentiate into skeletal myocytes (20) and vascular endothelial cells (21). SP cells were isolated from the bone marrow of C57BL/6 CD45.2 Rosa26 mice and their osteogenic potential was tested by transplantation into C57BL/6 CD45.1 mice. Both osteoblasts and osteocytes from the subsequent newly formed bone stained positively for markers of donors SP cells indicating osteogenic potential (19). Dominici and colleagues demonstrated in Friend leukemia virus B/ NIH Jackson (FVB/NJ) mice that transplantable fluorescently-labelled marrow cells from the non-adherent population can produce functional osteoblasts, osteocytes and hematopoietic cells (22). Kaplan and colleagues observed in a patient with FOP that bone marrow transplantation for treating anaemia was not sufficient to inhibit FOP, but that pharmacological suppression of the donor's immune system following transplantation inhibited FOP (23). In mice, hematopoietic cells contributed to the inflammatory and bone marrow-repopulating stages of BMP4-induced HO by recruiting and activating osteogenic precursors, but they did not act directly as a cellular precursor of HO (23). These findings contrast with those of Otsuru and colleagues who showed a contribution of hematopoietic cells to bone formation in BMP2-induced intramuscular HO, although only a minority of bone-marrow derived cells were embedded in the definitive heterotopic bone (24, 25). More recently, analysis of clinical tissue following musculoskeletal injury in humans has demonstrated that circulating osteogenic progenitor cells of bone

marrow origin, characterised by both type 1 collagen and CD45 immunopositivity, are found in early fibroproliferative and neovascular HO lesions, supporting the concept that circulating mononuclear progenitors can seed inflammatory sites to initiate HO formation (26). Taken together these studies suggest that haematopoietic cells of bone marrow origin contribute to both FOP and acquired HO development most likely through their creation of the necessary pro-osteogenic environment, but are unlikely to be significant direct osteogenic progenitors.

1.2.3.2 Endothelial Cells

Vascular endothelial cells have been suggested as a primary candidate for HO formation due to their multilineage potential via endothelial-mesenchymal transition (EndMT) and the expression of endothelial markers in FOP lesions (27, 28). During EndMT, endothelial cells loose cell-cell adhesion and change polarity, reconfiguring into a spindle-shape, and reducing the expression of endothelial markers whilst increasing mesenchymal marker expression. Following transition, these cells are highly motile and invasive and play an important role in both tissue development and disease (29-31). Medici and colleagues showed *in vitro* that endothelial cells over-expressing Activin Receptor-like Kinase 2 (ALK2, also called ACRV1), or treated with the ALK2 ligands TGF-β2 or BMP4, can dedifferentiate into stem cells with the capacity to re-differentiate into cartilage or bone cells (28). *In vivo* data in the neuron-specific enolase-BMP4 (NSE-BMP4) mouse also show that ectopic cartilage and bone cells express endothelial biomarkers such as vWF, VE-cadherin, Tie1, and Tie2 after injection of purified BMP (27, 28), after transgenic over-expression of ALK2 (28), or after muscle injury (27). Tie2 and vWF are also expressed in chondrogenic and osteogenic lesions from FOP patients, whereas osteoblasts and chondrocytes from normal cartilage or bone do not express these biomarkers (28). Lineage tracing in Tie2-Cre transgenic mice found that 50% of the cartilage and bone cells in HO lesions were of endothelial origin (27, 28). However, CD31+ endothelial cells were shown to not contribute to heterotopic cartilage or bone formation directly in the mouse following intramuscular BMP2 injection, but they did participate in lesion angiogenesis (32) and to HO development following burn/tenotomy injury (33). The different outcomes in the last two studies may be attributed to differences in the Cre drivers or in the HO models used (16). A further limitation of the lineage trace studies is that the markers expressed by endothelial cells can also be expressed by other cell types. Tie2, which is expressed in endothelial cells to regulate development and maintenance of vasculature (34), is also expressed in hematopoietic cells (35, 36), and by a population of Tie2+PDGFRα+Sca1+ multipotent mesenchymal progenitors that was shown to contribute to HO initiation (32). Furthermore, musculoskeletal injury induces expression of endothelial markers Tie2, CD31 and VE-cadherin in mesenchymal, non-endothelial cells (33). Taking together, the studies outlined above suggest that endothelial cells can undergo EndMT to initiate HO but they are unlikely to be pivotal, as Tie2+, CD31+ or VeCadherin+ progenitors also arise from other cell types and vary with the HO induction model used. These inconsistencies underscore the idea that the cellular populations contributing to HO development are highly tissue and context-specific.

1.2.3.3 Fibro-Adipogenic Cells

Fibro-adipogenic precursors (FAPs) are a population of PDGFRα+SCA1+ multipotent cells located within, but not exclusive to, skeletal muscle (32, 37, 38). FAPs are found near vascular elements, but are unlike pericytes in that they do not share a basal lamina with the endothelium and are NG2- (32, 38). Muscle resident FAPs support muscle regeneration but lack myogenic potential (29, 37, 38). FAPs were first discovered due to their fibrogenic and adipogenic capacities (37, 38). They were later shown to possess osteogenic potential when stimulated with BMP in culture and *in vivo* (32). Wosczyna and colleagues observed that Tie2-Cre lineage labelled FAPs made up ~50% of heterotopic bone and cartilage in the mouse (32). These cells have been proposed to play a major role in human FOP (39, 40). Several studies in the mouse show that progenitors of intramuscular and intratendinous HO are frequently PDGFRα+ and positive for cartilage and bone formation markers (38, 39, 41-44). Using a mouse FOP model in which *ACVR1* (that encodes ALK2) was genetically manipulated, Dey and colleagues showed that FAP-like cells can be divided into two lineages, Scx+ tendon-derived progenitors and a muscle-resident interstitial Mx1+ population (39). The Scx+ progenitors mediated endochondral HO without exogenous injury, whilst the Mx1+ population mediated injury-dependent HO. PDGFRα+ cells made up a minor subgroup of Mx1+ and Scx+ lineages; however, constitutive activation of *ACVR1* signalling demonstrated that PDGFRα+ subsets had an enhanced osteogenic and chondrogenic potential compared to unfractionated Scx+ or Mx1+ cells. Eisner and colleagues demonstrated that tissue resident FAPs in skeletal muscle are the primary source of osteogenic cells in the murine BMP2-Matrigel model of post-traumatic HO (44). In the same study using Notexin to induce muscle damage, they demonstrated that FAPs contribute to the formation of mature bone without the addition of exogenous BMP2. Moreover, when FAPs were cleared by macrophages at day 3-4 after injury, osteogenic genes were downregulated. Taken together, these findings suggest that FAPs can contribute to most HO presentations due to their broad distribution across tissue types and their documented presence in HO and that cells of hematopoietic origin play a role in stimulating their osteogenic potential.

1.2.3.4 Myosatellite Cells

Myosatellite cells are myogenic muscle-resident stem cells that are pivotal in skeletal muscle regeneration (45). They are located between the myofibre sarcolemma and basal lamina, and give rise to myodifferentiated cells following muscle injury (46). They were initially considered a primary precursor for muscle HO due to their osteogenic potential in culture in response to BMPs in the C2C12 murine myoblast model (47) and in human myogenic progenitor cells (48). However, lineage and transplantation studies indicate that they contribute minimally to BMP-induced HO *in vivo* (27, 32, 49, 50). Further, targeted expression of constitutively-activated ACVR1/ALK2 (caACVR1) (39, 51) and ACVR1 (R206H) (40) in myosatellite cells is insufficient to induce HO. Although Lees-Shephard and Goldhamer (16) have proposed that myosatellite

cells do not contribute to HO initiation, several lines of *in vivo* data do support their role in its pathogenesis. BMP signalling is a primary mechanism leading to the formation of acquired and genetic HO and has also been associated with the physiological regulation of skeletal muscle mass (52). When transplanted into the quadriceps of nude mice, skeletal muscle myoblasts have been shown to promote osteogenic differentiation (53). Muscle-derived stem cells express BMP4 and differentiate into bone (54). BMPs at the location of muscle injury inhibit myogenesis and promote osteogenesis of myoblasts, both *in vitro* (47) and *in vivo* (32). Further, serum taken from animals following a burn injury increases the osteogenic capacity of myosatellite cells, suggesting a role in burn-induced HO (55). Taken together, these findings indicate that the muscle tissue provides a permissive environment for HO and that following musculoskeletal trauma BMPs can modulate endogenous muscle progenitors to form heterotopic bone.

1.2.3.5 Other Cell Types

Through *in vivo* models, several other progenitor cell types have also been identified, including pericytes, tendon and ligament progenitors, and transient brown adipocyte-like cells (Table 1.1). Although these cell types are associated with HO initiation, their precise contributions remain unclear. More recently, using a burn/tenotomy injury in *Hoxa11-CreER^{T2}*; *ROSA-LSL-TdTomato* mice, Pagani and colleagues have traced the cell fate of MSCs in HO development using single-cell sequencing (56). They found that MSCs of the *Hoxa11* lineage differentiate through both the endochondral and osteogenic route into HO bone in the mouse forelimb following burn/tenotomy injury. During HO progression, the *Hoxa11*-lineage cells expressed transcriptional profiles characteristic of both osteogenesis and chondrogenesis. Previous studies have shown that *Hoxa11+* multipotent stromal cells are self-renewing and persistent throughout the life of mice, and that Hoxa11 contributes to bone formation, maintenance and repair (57-59).

Table 1.1. Overview of cell types investigated for their contribution to heterotopic ossification.

Cell Type	**Location**	**Description**	**Key Papers**
Hematopoietic cells	Bone marrow	Contribute to inflammation and marrow-repopulating stages. Contribution to HO is unclear.	(19, 23, 25, 60)
Endothelial cells	Blood and lymphatic vessels	Contribute to HO through EndMT route, but may be overestimated due to lack of surface marker endothelial cell-specificity.	(28, 35)

FAPs	Muscle and related soft tissues; widely spread in other tissues	Support muscle regeneration. Contribute to a high percentage of HO.	(32, 43, 61)
Myosatellite cells	Muscle	BMP2-induced HO. Contribution low based on most lineage studies.	(32, 48)
Pericytes	Vascular basement membrane	BMP-induced HO but assessment of contribution unclear due to high degree of heterogeneity.	(50, 62-64)
Hoxa11+ Mesenchymal stromal cells	Tendon, muscle and skeletal tissue	Contribute to skeletal repair, express chondrogenic and osteogenic transcription profile following injury.	(56-59)
Tendon and ligament progenitor cells	Tendon Ligament	Account for 25 and 40% of heterotopic bone and cartilage, respectively, after bone/tendonectomy based on Scx-Cre labelling. Molecularly heterogeneous.	(39, 43, 65)
Sensory neurons	Dermis, epidermis, and muscle spindle.	Mediate HO formation via substance P and calcitonin gene-related peptide. BMP2 may induce neurogenic inflammation to remodel nerve and release HO precursor cells. May explain how HO occurs following traumatic brain injury. Mice lacking sensory neurons cells do not develop HO. Tie2+ endoneurial progenitors the major HO cell contributors in a mice model, however, Tie2 marker is also expressed in endothelial and mesenchymal cells.	(66-69)
Transient brown adipocyte-like cells	Adipose	Specialised pool of brown adipocytes that contribute to HO. Associated to deposition of cartilage. Detected in human traumatic injury-induced HO.	(70, 71)

Due to the heterogenic nature of HO aetiology, several cell types contribute depending on the site and initiating factors. This raises the issue of which cell and experimental model is most appropriate for investigating the function of HO susceptibility genes in culture and/or *in vivo*. A conclusive answer to this question remains elusive, nevertheless, the role of specific genes may be best examined by investigating how they affect the signalling response of precursor cells to promote bone formation and/or maintenance using an experimental model most appropriate to the type of HO investigated.

1.2.4 Signalling Pathways in HO

1.2.4.1 BMP Signalling

BMPs are a family of signalling molecules that belong to the Transforming Growth Factor-β (TGF-β) superfamily of proteins. Discovered by Urist in 1965 (72), they play a crucial role in bone formation and repair, and in HO development (73). During normal bone development and physiological homeostasis, BMP ligands bind to a heterotetrameric complex of two BMPRI and two BMPRII transmembrane serine/threonine kinase receptors to initiate chondrogenesis and osteogenesis. The BMPs that initiate signalling through this mechanism and the osteogenic processes that they initiate are summarised in Table 1.2.

Table 1.2. Overview of BMPs and their role in major cellular process and heterotopic ossification.

Signalling Protein	Function	Key Papers
BMP1	Bone formation and homeostasis.	(74)
BMP2	Induces bone and cartilage development. Induces EndMT transition. Also involved in hedgehog pathway, cardiac cell differentiation, embryonic development.	(75-78)
BMP3	Bone and cartilage development; antagonises other BMPs in osteo-differentiation.	(79)
BMP4	Potently induces chondro- and osteogenic differentiation; induces EndMT transition. Also involved in embryonic development, adipogenesis, neurogenesis.	(80-83)
BMP5	Bone and cartilage development; may play a role in some cancer types; expressed in the visual apparatus.	(84-86)

BMP6	Osteogenic differentiation; closely related to BMP5 and BMP7; regulates iron metabolism	(87-89)
BMP7	Bone homeostasis; induces osteoblast differentiation through SMAD canonical pathway; involved in embryonic development, adipogenesis.	(90-92)
BMP8	Expressed in developing skeleton; osteogenesis and germ cell generation.	(93-96)
BMP9/GDF2	Induces chondro- and osteogenesis; cannot be blocked by BMP3 unlike most BMPs; involved in lymphatic development.	(97-99)
BMP10	Involved in the trabeculation of the heart and regulates monocyte recruitment to the vascular endothelium.	(100-102)
BMP11/GDF11	Augments bone formation; induces embryonic development.	(103, 104)
BMP12/GDF7	Inhibits endochondral bone growth; induces tenogenic differentiation; regulates bone structure	(105)
BMP13/GDF6 /CDMP2	Establishes the boundaries between skeletal elements during development; induces tenogenic differentiation	(105, 106)
BMP14/GDF5 /CDMP1	Regulates skeletal development and joint formation; promotes fracture healing.	(106-108)
BMP15	Involved in fertilisation and ovulation	(109, 110)

Four type I BMP receptors (ALK1, ALK2 (also termed ACVR1), ALK3 and ALK6) bind BMP ligands. Three receptors (BMPR2, ALK4 and ALK7) serve as type II BMP receptors. ALK4 and ALK7 (also termed ActR-IIA and ActR-IIB), also act as receptors for activins, whilst BMPR2 only binds BMPs (Figure 1.1).

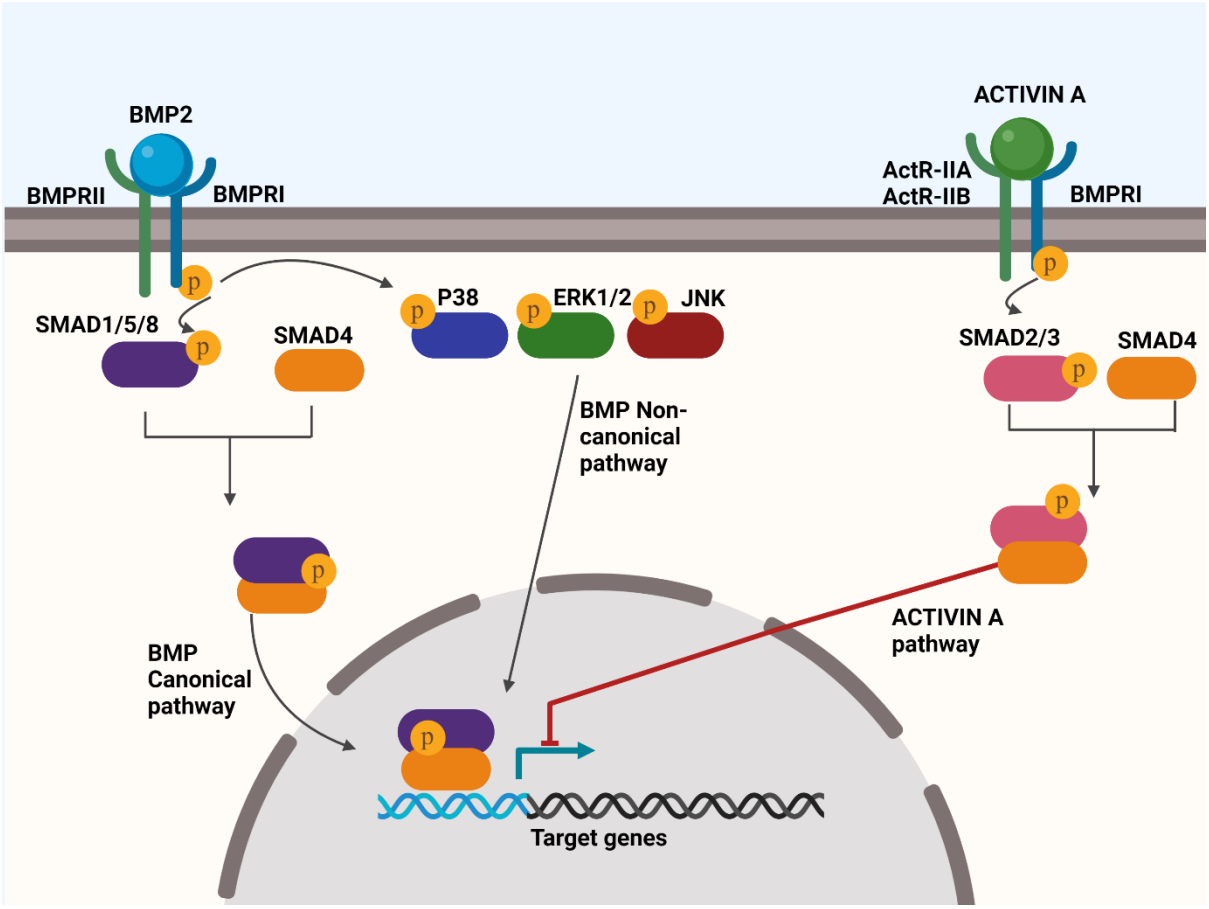


Figure 1.1. BMP receptor activation and downstream signalling and its antagonism through the Activin A pathway. In the canonical pathway, SMAD1/5/8 is activated and interacts with SMAD4 to promote expression of target genes that induce bone formation. In the non-canonical SMAD pathway, p38 MAPK, ERK1/2 and/or JNK are activated to promote the expression of osteogenic target genes. BMP signalling is antagonised by the binding of Activin A to its receptor complex to initiate SMAD2/3 signalling that acts to suppress BMP target gene transcriptional activation.

Downstream signalling following BMP receptor activation occurs through 2 distinct pathways: 1. SMAD canonical pathways, in which SMAD 1/5/8 proteins are phosphorylated to promote expression of chondro- or osteogenic genes (111); 2. Non-canonical SMAD pathways where p38 MAPK, ERK or JNK are activated (112-115). Under normal physiological conditions, these chondro- and osteogenic signalling pathways are antagonised by Activin A (another TGF-β superfamily member) binding to a heterotetrameric receptor complex comprising two ActR BMPRII receptors and two BMPRI receptors to initiate SMAD2/3 phosphorylation and downstream signalling as a negative feedback mechanism for gene transcriptional activation that is initiated by BMP signalling (116). These pathways should not be viewed as independent, as crosstalk between them occurs (117-120). BMP2 is overexpressed in clinically evolving HO tissue after trauma (121, 122). Augmented BMP signalling also occurs following trauma-induced HO development in animal models whilst BMP antagonism reduces HO severity (123, 124). Experimental models of HO therefore commonly use exogenous BMP2

(27, 68) or overexpression of BMP4 (125), or recombinant BMP2 (rhBMP2) (27, 126) as the HO initiator. BMP signalling is also a key feature of the heritable forms of the disease (127). In FOP, a mutation in *ACVR1* that encodes the BMP type 1 receptor ALK2, causes its constitutive activation, initiating downstream BMP signalling regardless of BMP ligand binding (124).

1.3.4.2 mTOR Signalling

The mammalian target of rapamycin (mTOR) signalling pathway is involved in several cellular processes, including chondrogenesis, osteogenesis and skeletal development (128, 129). The FOP activating mutation in *ACVR1* has been shown to increase mTOR signalling (130). Conversely, rapamycin suppresses bone formation in experimental models for FOP (41, 130), trauma-induced HO (41, 131), and in leptin-induced osteogenesis in both *in vitro* and *in vivo* models (132) through inhibition of mTOR complexes mTORC1 and mTORC2 (133) (Figure 1.2). Rapamycin is currently being studied in a phase 2 clinical trial (UMIN000028429) of the disease. BMP2 also promotes osteogenesis through an mTORC1-dependent mechanism (134), whilst mTORC2 modulates osteogenesis in response to a range of mechanical or chemical cues (128, 135, 136).

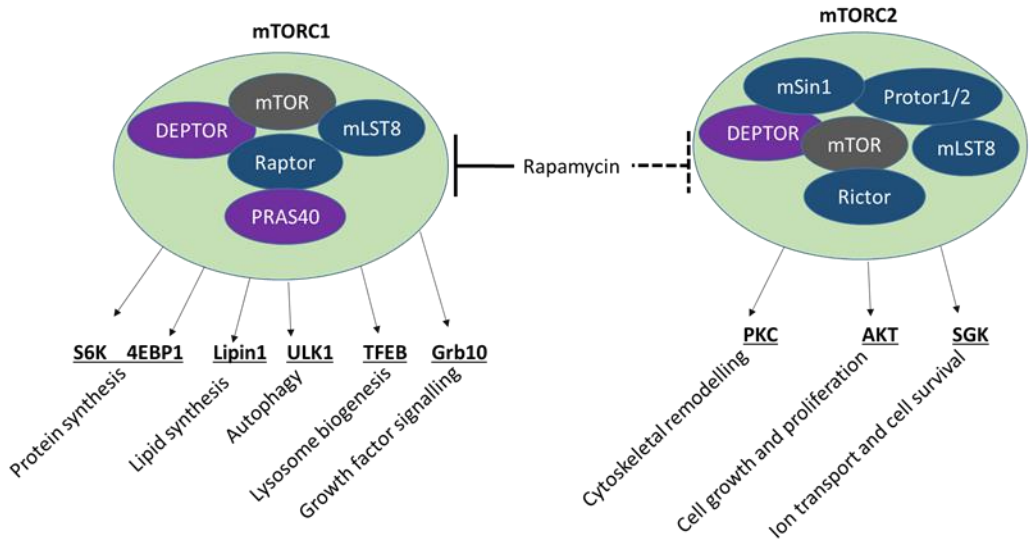


Figure 1.2. mTOR signalling pathway. Rapamycin inhibits mTORC1 and mTORC2, which in turn modulate several downstream osteogenic pathways. Acute rapamycin treatment inhibits mTORC1 whilst repeated dosing of rapamycin also inhibits mTORC2. Both mTORC1 and mTORC2 are activated by Wnt and IGF. mTORC1 is also activated by BMP2 and mTORC2 is also activated by mechanical and chemical signals to promote osteogenesis.

1.2.4.3 Other Signalling Pathways

Hypoxia-inducible factors (HIFs) activate genes that mediate adaptive responses to reduced oxygen tension (137, 138). HIFs augment HO formation (41) and couple bone and vascular growth during development (138). Retinoic acid receptor (RAR) signalling is mediated by

retinoids (metabolic derivatives of vitamin A), which are potent morphogens that promote both chondro- and osteogenesis to shape skeletal development (139). In retinoic acid (RA) mediated gene activation, RA binds to a heterodimer complex comprising RAR and the retinoid X receptor (RAR-RXR). RAR-RXR then activates gene transcription by binding to DNA motifs termed RA-response elements (RARE) located within enhancer regions of RA target genes (140). In the absence of RA, unliganded RAR-RXR recruits histone deacetylases and nuclear corepressors to inhibit transcriptional activation at the RARE (140, 141). Chondrogenesis requires the absence of RA signalling, in which the repressor function of unliganded RAR-RXR on RAREs dominates (141, 142), whilst active RA signalling prevents the chondrogenic differentiation of precursor cells (143). Crosstalk between the HIF and RAR signalling systems is well documented, but how they co-operate to modulate bone formation is still incompletely understood (144-148). Due to the pleiotropic function of these pathways, it is anticipated that any therapeutic application to inhibit HO may have off-target effects, as these pathways also dynamically regulate several other critical cellular processes (149).

1.2.5 Therapeutic Strategies for HO

Treatment strategies for acquired HO to date have included the use of anti-inflammatory agents, bisphosphonates, local radiation therapy, and surgical resection. Systematic reviews have shown that patients treated with either selective or non-selective non-steroidal anti-inflammatory drugs (NSAIDs) showed a significant decrease in post-traumatic HO formation when compared with placebo (150-152), but were associated with a higher rate of drug discontinuation due to gastrointestinal side effects. Low-dose local radiation therapy also decreases the incidence of HO after surgery (153, 154), but carries the risk of irradiation-induced malignancy (155) and side-effects such as delayed wound-healing, progressive soft-tissue contracture, non-union, and inhibited ingrowth of cementless hip implants (156, 157). The treatment of mature HO after trauma involves surgical resection, although complete excision may not be feasible and recurrence is common (158-160)). Simple bisphosphonates, such as etidronate, have also been studied as a prophylactic intervention in HO, as they delay matrix mineralisation. However, bisphosphonates do not inhibit bone matrix synthesis, and mineralisation recommences after drug discontinuation (161, 162). None of these strategies specifically target molecular pathways involved in HO pathogenesis. However, as our understanding of these cells and pathways evolves, molecular mechanism-specific investigative therapeutic approaches are beginning to emerge, as outlined below).

1.2.5.1 Palovarotene and Other RAR Agonists

The observation that RA signalling suppresses chondrogenesis has stimulated its investigation as a therapeutic target for HO. Synthetic retinoid agonists selective for nuclear RARα or RARγ have been tested in mouse models of injury-induced intramuscular HO, implantation of rhBMP-2 and constitutive activation of mutant Acvr1(Q207D) (51, 163). Whilst RA agonists targeting both RARα and RARγ inhibited endochondral HO, those targeting RARγ were most effective as RARγ is more strongly and selectively expressed in chondrogenic cells than other

RAR members (164, 165). Chakkalakal and colleagues showed that palovarotene prevented HO, restored long bone growth, and preserved growth plate function in transgenic mice carrying the human ACVR1 (R206H) mutation for classic FOP (166). In juvenile FOP mice, palovarotene reduced HO both in vitro and in vivo, but resulted in aggressive synovial joint overgrowth and long bone growth plate ablation (167). In a rat model of post-traumatic HO (in which rats were subjected to blast overpressure via a shock tube resulting in femur fracture, soft tissue crush injury, and amputation through the zone of injury (168)), Palovarotene treatment suppressed the systemic and local inflammatory response, decreased osteogenic progenitor colonies by 98% in both *in vitro* and *in vivo*, and decreased the expression of osteo-and chondrogenic genes, including BMP4 (168). In another trauma-induced model, rats were subjected to blast-related limb injury, femoral fracture, quadriceps crush injury, amputation, and infection with methicillin-resistant Staphylococcus aureus (MRSA) (169). Palovarotene treatment decreased HO by 50-60%, however 63% of rats treated with palovarotene and inoculated with MRSA experienced delayed healing or dehiscence compared to 25% of MRSA rats in the placebo arm of the study. Palovarotene is currently the subject of several clinical trials of efficacy and safety for the prevention of new HO lesions in both children and adults with FOP (www.clinicaltrials.gov; accessed on 5 May 2022 NCT02190747, NCT03312634, NCT02979769, NCT02521792, NCT05027802). However, whether Palovarotene or other RAR agonists represent a viable approach for treating acquired HO in humans remains unstudied.

1.2.5.2 Targeting ACVR1/ALK2 and Other Related Signalling Pathways

Under physiological conditions in normal tissues, the ligand Activin A interacts with ALK2 to mediate SMAD2/3 phosphorylation to regulate cell proliferation, apoptosis, and differentiation (Figure 1.1) (170-174). In ALK2R206H+ FOP cells (that carry the common *ACVR1* mutation) ALK2 is activated constitutively in the absence of BMPs, enhancing both canonical and non-canonical BMP signalling pathways (175-179) to augment chondrogenesis (39, 179-183). Although the ACVR1mutation is not implicated in other forms of HO, ALK2 signalling has been explored as an investigational target due to its BMP agonism (184). Table 3 provides a summary of molecular targets and investigational therapeutic strategies explored to date in HO prevention and treatment.

Table 1.3. Summary of investigational therapeutic strategies for the inhibition of heterotopic ossification, based on ALK2 signalling and other pathways. FOP = fibrodysplasia Ossificans Progressiva, tHO = acquired post-traumatic Heterotopic Ossification.

Type of HO Pathways	Type of molecule	Molecule	Description and Function	Key Papers
FOP	Antibody	REGN2477 (Garetosmab)	Anti-activin-A human monoclonal antibody in phase 2 clinical trial for FOP (LUMINA‑1 study, NCT03188666). Blocks signalling of activin A, AB, and AC. Inhibits HO in animal model of FOP.	(179, 185-187)
FOP	Antibody	Perhexiline maleate (Pex)	Identified in screening of 1040 FDA-approved drugs for suppression of the Id1 promoter activated by mutant ACVR1/ALK2 in mouse C2C12 myoblasts. Pex reduced HO volume in BMP-induced mouse model, but failed to inhibit HO in an open-label clinical trial in FOP.	(188, 189)
tHO	Antibody	Metformin	Regulates osteogenic differentiation via AMPK, and RUNX2/CBFA1 *in vitro* and *in vivo*. Prevents traumatic HO in mouse by decreasing ALK2 and AMPK regulation of Smad2.	(190-192)
FOP	Alpha-2 blocker	Fendiline hydrochloride	Identified in screen of 1040 FDA-approved drugs for suppression of the Id1 promoter activated by mutant ACVR1/ALK2. Mice administered with fendiline showed a slight reduction in HO.	(188)
FOP	Small molecule inhibitor	Dorsomorphin	Identified by chemical library screen for small molecules that dorsalise zebrafish embryos. Selectively inhibited ALK2 to block BMP-mediated SMAD1/5/8 phosphorylation. Preclinical use precluded by the inhibition of other ALKs (ALK3 and ALK6) and other kinases.	(176, 193)
FOP, tHO	Small molecule inhibitor	LDN-193189	An optimised version of dorsomorphin with greater potency and selectivity. Inhibits transcriptional activity of ALK2, ALK3, and constitutively active ALK2 mutant proteins.	(124)
FOP, tHO	Small molecule inhibitor	LDN-212854	Derivative of dorsomorphin with increased selectivity for ALK2. LDN-212854 and LDN-193189 reduce osteogenic differentiation of tissue-resident MPCs from injured tissue	(194, 195)

			following burn or tenotomy insult in animal model. In a blast-induced rat tHO model, LDN193189 and LDN212854 effective at limiting tHO.	
FOP, tHO	Small molecule inhibitor	Other dorsomorphin derivatives	Currently undergoing investigation, including K02288, DMH-1, ML347, LDN 214117 and VU465350.	(196-198)
FOP	Small-molecule inhibitor	Saracatinib (AZD-0530)	Identified by screening compounds in an ALK2-mutated chondrogenic ATDC5 cell line. Inhibited both BMP and TGF-β signalling in vivo. Currently undergoing phase 2 clinical trial for FOP (NCT04307953). Well tolerated and potently inhibits the development of HO in inducible ALKQ207D transgenic and ACVR1R206H knock-in mouse.	(199-202)
FOP	Small-molecule inhibitor	PD 161570	Identified by screening compounds in an ALK2-mutated chondrogenic ATDC5 cell line. Inhibits both BMP and TGF-β signalling in vivo.	(199)
FOP	Small-molecule inhibitor	TAK 165	Identified by screening compounds in an ALK2-mutated chondrogenic ATDC5 cell line. Indirectly modulates mTOR signalling in vivo.	(199)
FOP	Ligand traps	sActR-IIA-Fc and sActR-IIB-Fc	ACVR1-Fc fusion proteins comprising the extracellular domain of human WT ACVR1 and the Fc portion of human immunoglobulin γ1. Inhibits dysregulated BMP signalling caused by FOP mutant ACVR1 and abrogates chondro-osseous differentiation in vitro.	(203-205)
FOP	Platelet inhibitor	Dipyridamole	Identified in screening of 1280 FDA-approved compounds for suppression of ACVR1 gene expression. Showed the highest inhibitory effect on SMAD signalling, chondrogenic and osteogenic differentiation in vitro. Reduced HO in BMP-induced model in mice.	(206, 207)
FOP, tHO	Nucleotides	microRNAs	Altered expression of miRNA detected in HO. mir148b and mir365 down-regulate ACVR1/Alk-2 expression, whereas mir26a showed a positive effect on its mRNA. Inhibition of miRNAs, miR-146b-5p and -424 suppresses osteocyte maturation. Manipulating miR-574-3p levels both in vitro and in vivo inhibits chondrogenesis.	(208-214)

			miR-630 downregulated in early HO and used to distinguish HO from other processes in tHO. miR-17-5p upregulated in ankylosing spondylitis (AS) patients versus non-AS individuals. Knockdown and overexpression of miR-17-5p in fibroblasts derived from AS patients modulates osteogenesis.	
FOP, tHO	Nucleotides	Antisense oligonucleotide (AON)	AON binds to specific exons in the primary mRNA transcript to prevent splicing and enable the skipping of specific exons. AONs designed to knockdown ALK2 expression in mice impair ALK2 signalling in both C2C12 end endothelial cells. However, AON affects both wild-type and mutated allele.	(215-217)
FOP, tHO	Nucleotides	RNA interference (RNAi)	Allele-specific siRNA (ASP-RNAi) duplexes tested for specific inhibition of mutant c.617A allele in mesenchymal progenitor cells from FOP patients. ASP-RNAi decreased BMP signalling to control cell levels.	(218, 219)
tHO	Nucleotides	LncRNAs	Several lncRNAs regulate bone formation. Downregulation of *MANCR* inhibits osteoinduction *in vitro*. In a mouse *in vivo* tHO model, *Brd4-Mancr* signalling attenuated HO.	(220-222)

1.2.6. Conclusions

In summary, heterotopic ossification may arise from both rare, heritable and common complex diseases. The downstream molecular pathways that underpin these heterogeneous aetiologies are broadly similar in both patterns of disease, although the diseases differ in extent and severity. Whether genetic or acquired, initiation of a new HO lesion involves tissue injury that results in a signal to initiate endochondral or intramembranous ossification. The dominant cell types in HO include are FAPs, endothelial cells, hematopoietic cells, tendon and ligament progenitor cells, pericytes and Hoxa11+ mesenchymal stromal cells. The dominant pathways in HO include BMP, mTOR and RAR signalling. Several therapeutic strategies have been developed to target these signalling pathways. RAR agonists have been shown to be effective in preventing HO in pre-clinical models. Although the RAR agonist Palovarotene is undergoing clinical trials for FOP, further pre-clinical animal studies will be required to investigate its efficacy and safety for the post-traumatic HO indication. Several strategies have been developed to target ACVR1/ALK2 with REGN2477, metformin and dorsomorphin derivatives being a few prospects for clinical therapeutic applications. These future studies would benefit from translational experimental approaches that incorporate clinically relevant

animal models in parallel with clinical investigations, population epidemiology studies and relevant molecular medicine techniques.

References for this article are included in the general references section at the end of the thesis.

1.3 Introduction to the human-only lncRNA *CASC20* as the topic of this thesis

The genetic susceptibility of post-traumatic HO was investigated by conducting a case-control GWAS in THA-treated patients not less than 1 year after surgery (223). Single Nucleotide Variants (SNVs) linked to the radiographic phenotype of HO were identified. The GWAS discovery cohort consisted of 411 cases and 480 controls, and the replication GWAS was conducted with a cohort made up of 198 cases and 205 controls. The GWAS discovery analysis revealed that the strongest signal linked to HO susceptibility was located downstream *ARHGAP18* and the second strongest localised within the intronic region (rs11699612) of the human-only lncRNA Cancer Susceptibility 20 (*CASC20)*. The replication GWAS demonstrated that the SNV within *CASC20* is the strongest signal linked to HO, reaching genome wide significance (Effect Allele (EA) T, EA Frequency 0.24, OR 1.94 [1.59-2.35], p=2.71x10^{-11}) (Figure 1.3). The SNV downstream ARHGAP18 was the third strongest signal, but did not reach genome-wide significance. Further statistical analysis was conducted to confirm the locus of the SNV observed within *CASC20*. Following up on the SNV's discovery, preliminary studies on *CASC20* expression were undertaken. Bone-marrow derived human mesenchymal stem cells (hMSCs) and human multipotent adipose-derived stem cells were differentiated into osteoblasts and analysed to evaluate the expression of *CASC20*. It was found that *CASC20* is differentially expressed and upregulated across the osteogenic differentiation suggesting that it has an osteogenic-related function. Further gene expression analysis confirmed *CASC20* expression in several musculoskeletal cell types (223).

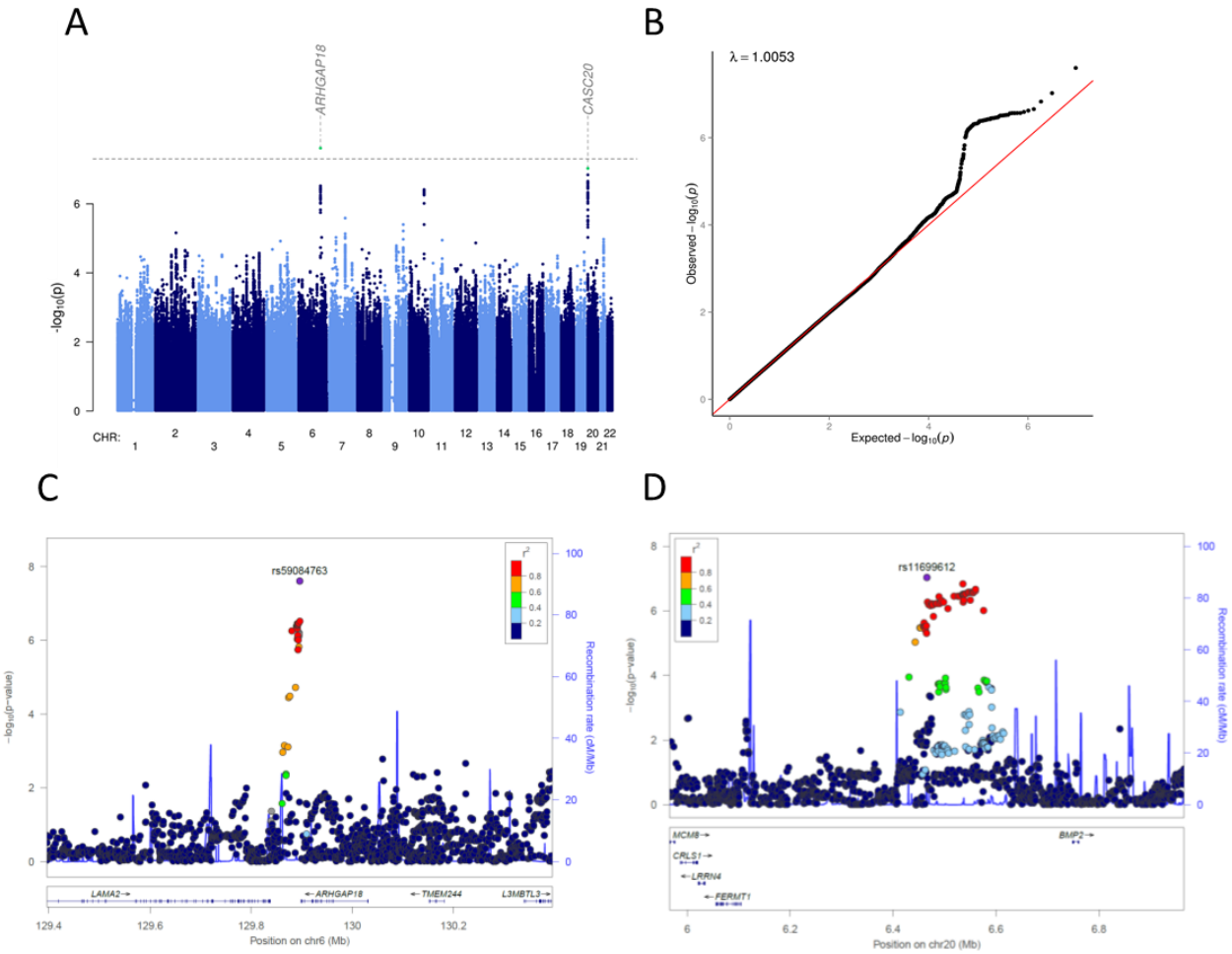


Figure 1.3. Discovery genome-wide association analysis. A) Manhattan plot showing the -log10 p-values for each variant (y axis) plotted against their respective chromosomal position (x axis). The horizontal dashed line denotes the genome-wide significance threshold p=5x10-8. Lead signals are indicated in green. B) Quantile-quantile plot of the data used in the GWAS. The x-axis indicates the expected −log10 p-values and the y-axis the observed ones. The red line represents the null hypothesis of no association at any locus and λ is the genomic inflation factor. Regional association plots for **C)** rs59084763, and **D)** rs11699612 with HO susceptibility. Each filled circle represents the *p*-value of analysed variants in the discovery stage plotted against their physical position (NCBI Build 37). The purple circle denotes the variant with the lowest *p*-value in the region. The colours of variants in each plot indicate their r^2 with the lead variant according to a scale from $r^2 = 0$ (blue) to $r^2 = 1$ (red).

Despite recent investigation of the possible functional role of *KIF26B* in the severity of HO (224), how *CASC20* may modulate HO remains uninvestigated. Little was known about *CASC20* prior to this PhD project, *CASC20* was previously found to be expressed in the testis, placenta, and prostate cancer. The intronic SNV at rs11699612 within *CASC20* is associated at genome-wide significance with other musculoskeletal traits, including height (GIANT UK Biobank GWAS, p=1.80 x 10-95) (225), and estimated bone mineral density (UK Biobank eBMD and fracture GWAS, p=2.40 x10-22) (226). However, the mechanism of these associations

remained unknown. Preceding work during my CIMA MRes aimed to characterize the expression of *CASC20* during the osteodifferentiation of hMADs and human bone marrow-derived mesenchymal stem cells (hBMSC) and in waste bone samples retrieved at joint replacement (n= 3 subjects). I found that *CASC20* is expressed at a low-endogenous level and that it is upregulated during the osteodifferentiation of hMADs and hBMSC alongside *RUNX* and *OSX*, which are master regulators of osteoblast differentiation (223). To investigate the potential mechanisms through which *CASC20* may exert its function in HO, it would be appropriate to evaluate how lncRNAs operate. The next section of this thesis introduces the concept of epigenetic modification of gene function in molecular biology and is followed by a section dealing with the epigenetic modification of osteogenesis in HO.

1.4 Epigenetics and its role in the modification of gene expression

Epigenetics is a rapidly growing field that includes heritable or long-term variations in gene activity and expression that are not caused by changes in DNA sequence (227-230). These changes can occur at various levels, including chromatin remodelling, histone modification, DNA methylation, and translational regulation (231, 232), which will be described in the following subsections. Genetic and environmental factors, such as diet, exercise, and stress, can influence these epigenetic modifications, with significant consequences for health (233-235). For example, alterations in epigenetic markers have been linked to a range of pathological and physiological processes, including cancer, autoimmune diseases (236, 237), development, growth, and ageing (238, 239).

While some epigenetic modifications can be inherited across generations, not all are heritable (240). Somatic cells, which make up an organism's body but are not involved in reproduction, can undergo epigenetic changes that are not passed on to offspring (241). Epigenetic alterations that occur in reproductive (germline) cells are more likely to be heritable (240). However, not all epigenetic modifications in germline cells are inherited due to a process called epigenetic reprogramming that occurs in early embryonic development (240). During this process, epigenetic modifications in germline cells are erased to establish a totipotent state in the zygote (240). Despite this, some epigenetic changes in germline cells can persist across multiple generations (242). For example, a study by Cunningham and colleagues found that offspring of mice exposed to a stressful environment before breeding exhibited increased susceptibility to stress phenotypes and modifications in lncRNA expression and behaviour (243).

The extent and mechanisms of epigenetic inheritance are still being investigated (244), but this example highlights the potential for epigenetic changes to have intergenerational effects on health and behaviour (243). As the field of epigenetics continues to advance, researchers are exploring novel methods for modifying epigenetic markers to promote health and prevent disease (245, 246). These methods include using transcription activator-like effector (TALE) or clustered regularly interspaced short palindromic repeats (CRISPR)-based tools to promote DNA methylation or demethylation (247, 248).

1.4.1 Chromatin remodelling

Eukaryotic genetic information is packaged into chromatin, which consists of DNA wrapped around a core of histone proteins to form a structural unit called the nucleosome (249, 250). Chromatin remodelling is a dynamic process in which the structure of the chromatin is modified by condensation or relaxation of DNA-histone interactions to regulate gene expression. This process enables genes to become transcriptionally accessible when the chromatin is in a relaxed, open state (249, 250) (See Figure 1.4). Chromatin remodelling complexes are divided into two main classes: ATP-independent and ATP-dependent (251). The ATP-independent class includes enzymes that modify chromatin through methylation, phosphorylation, acetylation, sumoylation, adenosine diphosphate–ribosylation, glycosylation, or ubiquitination (252). The ATP-dependent class encompasses enzymes that modify chromatin structure using energy from ATP hydrolysis. There are four families of ATP-dependent remodelling complexes: Switch/sucrose non-fermentable (SWI/SNF), Imitation SWI/SNF (ISW1), chromodomain helicase DNA-binding (CHD), and inositol-requiring protein 80 (INO80) (251). It is difficult to estimate the exact proportion of chromatin remodellers that interact with lncRNA to function correctly, as this varies with the experimental system and specific cellular context examined (253). Several studies have shown that many chromatin remodellers, both ATP-independent and ATP-dependent, require the involvement of lncRNA for their activity, Examples of such chromatin remodellers include SWI/SNF (254), histone lysine demethylase 1 (LSD1) (255), polycomb repressive complex 2 (PRC2) (256), DNA methyltransferase 1 (DNMT1) (257), and growth arrest and DNA-damage-inducible 45 alpha (GADD45) (258).

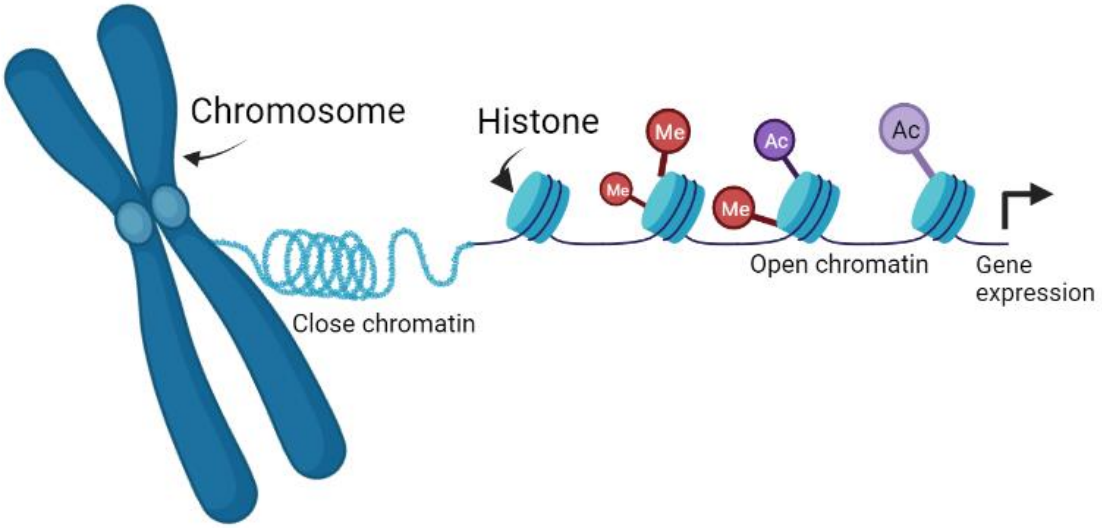


Figure 1.4. Schematic representation of chromatin remodelling. On the left is the tightly packed or closed chromatin structure, also termed heterochromatin. This conformation results in low or no gene expression (259). The chromatin structure can be remodelled by demethylating and acetylating histones, as shown on the right. This process results in an open

chromatin structure termed euchromatin, allowing transcription factors and RNA polymerase easier access to the DNA to promote RNA expression (260).

1.4.2 Histone modification

The histone proteins comprise eight subunits, including two copies each of H2A, H2B, H3, and H4, that form a histone octamer (261, 262) that is bound (wrapped) by about 145-147 base pairs of DNA to form a nucleosome (263-265). Histones are made of two regions: a globular region that coils DNA in a core, and flexible amino- and carboxyl terminal tails (266). For each nucleosome core, there are 10 histone tails: amino terminals of H2A, H2B, H3, and H4, and carboxyl terminal of H2AX, which is a variant of H2A (267). The term "histone modification" describes the post-translational modification (PTM) of the histone proteins. PTM affects how accessible DNA is to the transcriptional machinery, which is important for controlling gene expression (231). The extremely high positive charge of histones, concentrated in the histone tails, is thought to contribute to nucleosome stability by neutralising the negative charge of the DNA (268, 269). A variety of covalent modifications such as acetylation, methylation, phosphorylation, ubiquitination, and SUMOylation, are applied to the N-terminal tails of histones (231). Specific residues on histones undergo modifications that can influence chromatin structure or the binding of effector molecules, ultimately leading to the activation or repression of gene expression (270).

1.4.2.1 Histone acetylation and methylation

Histone acetylation is one of the most well-studied modifications and has been linked to gene activation (231, 270). Histone acetyltransferases (HATs) catalyse the addition of acetyl-CoA to the ε-amine of lysine residues to neutralise the positive charge of the histone and reduce DNA-histone affinity (271-274). This process promotes an open chromatin structure, allowing transcription factors and RNA polymerase to access the DNA (271). Histone deacetylation, on the other hand, is catalysed by histone deacetylases (HDACs) (275). They reverse the reaction mediated by HATs to restore the positive charge of the histone, promoting chromatin condensation and gene repression (231). Histone methylation can either promote or repress gene expression, depending on the lysine, arginine, or histidine residue methylated and the amount of methyl groups added (276-279). Histone methyltransferases (HMTs) catalyse methylation, resulting in mono-, di- or trimethylation of the amino acid residue (231). Histone methylation promotes gene regulation by creating binding sites for transcriptional activators/repressors and/or modulating chromatin structure and accessibility by recruiting proteins involved in chromatin remodelling (280-283). For example, tri-methylation of histone H3 at lysine 4 (H3K4me3) is associated with transcriptional activation and is essential for many types of cell differentiation (284, 285), including osteoblast differentiation (286). On the other hand, H3K9me3 and H3K27 are associated with gene repression, and their levels have been shown to decrease during MSC differentiation into osteoblasts (287). Histone methylation was previously believed to be irreversible (288). In 2004, LSD1 (lysine specific demethylase 1, also known as KDM1A) was discovered to demethylate H3K4, revealing that

histone methylation is reversible (289). Since then, a plethora of HMTs and histone demethylases (HDMs) have been identified (279).

1.4.2.2 Histone phosphorylation

Histone phosphorylation is an important and reversible histone modification that affects many cellular processes, including cell division, DNA damage, and transcriptional regulation (for review, see (290). Phosphorylation of specific histone residues can act as a signal to recruit specific proteins involved in these processes, allowing the appropriate cellular response to take place (291, 292). A well-studied example of this modification is phosphorylation of histone H2AX at serine 139, termed γ H2AX (293, 294). This modification serves as double-stranded break (DSB) marker and recruits DNA repair machinery to the site of the break (291). Phosphorylation of serine 10 residue in histone H3 (H3S10ph) is an emerging epigenetic marker associated with chromatin condensation and gene regulation (295), particularly during mitosis and meiosis (292, 296). H3S10ph is also involved in DNA-RNA hybridisation, resulting in regulation of transcription and chromosomal stability (297). H3S10ph is associated with several cancer types (298-300) and several H3S10 kinases, such as AURORA and CDK8, are considered promising targets for cancer therapy (301, 302).

1.4.2.3 Histone ubiquitination and SUMOylation

Histone ubiquitination and SUMOylation are less studied PTMs, but increasing evidence shows that they play important roles in chromatin organisation and gene regulation (303, 304). Ubiquitination is a reversible process in which ubiquitin molecules are added to a protein (305), resulting in either mono- or polyubiquitination of histone tails (306). Ubiquitin is a highly conserved protein made up of 76 amino acids (307). Ubiquitination is associated with several biological processes such as DNA repair (308, 309), cell cycle progression (308, 310), and transcriptional regulation (311, 312). The two most copious ubiquitinated histones are H2A and H2B, as 5-15% of H2A and 1-2% of H2B are ubiquitinated in vertebrate cells (306, 313-315). Monoubiquitination of H2A (H2Aub) and H2B (H2Bub) is generally correlated with gene silencing (316-318) and activation (319, 320), respectively. Polyubiquitination of H2A and H2AX recruits DNA repair proteins at DNA damage loci (321). SUMOylation is a similar process in which small ubiquitin-like modifier (SUMO) proteins are added to a protein by SUMO ligases (322). Examples of SUMOylation include SUMOylation of H4, H2A, and H2AX, which are involved in transcriptional repression (323, 324).

1.4.2.4 Cross-talk between histone modifications

It is important to note that histone modifications are not independent events (231). They interact with each other and with DNA methylation to regulate gene expression (325). Lysine residues on histone proteins can undergo various types of modifications (281) (See Table 1.4), resulting in antagonism as different types of modifications on lysine are often mutually exclusive (281). Additionally, the binding of a protein can be disrupted by nearby post-

translational modifications (PTMs). For example, phosphorylation of histone H3 serine 10 (H3S10ph) can affect the binding of heterochromatin protein 1 (HP1) to methylated histone H3 lysine 9 (H3K9me), leading to changes in chromatin structure and gene expression (326). Moreover, the catalytic activity of an enzyme can be compromised by modifications at its substrate recognition site. For instance, isomerisation of histone H3 proline 38 (H3P38) can affect the methylation of histone H3 lysine 36 (H3K36) by the enzyme Set2, resulting in alterations in gene expression (327). These examples emphasize the complex and dynamic interplay between histone modifications, which is crucial for controlling chromatin structure and gene expression.

Table 1.4. Overview of histone-modifying enzymes investigated for their contribution to gene regulation.

Histone-Modifying enzymes	**Type of Modification**	**Effect on Chromatin Structure**	**Examples of Molecules**	**Key papers**
Histone acetyltransferases (HATs)	Acetylation of Lysine residues	Loosens chromatin structure, promotes gene expression	p300, CBP, GCN5	(328, 329)
Histone deacetylases (HDACs)	Deacetylation of Lysine residues	Condenses chromatin structure, represses gene expression	Class I (HDAC1, 2, 3, and 8), Class II (HDAC4, 5, 6, 7, 9, 10), Class III (Sir2-related enzymes), and Class IV (HDAC11).	(275)
Histone methyltransferases (HMTs)	Methylation of Lysine or Arginine residues	May condense or loosen chromatin structure, depending on the site and degree of methylation	EZH2, G9a, SETD2	(330)
Histone demethylases (HDMs)	Demethylation of Lysine or Arginine residues	May loosen chromatin structure, promoting gene expression	LSD1, JMJD3, UTX	(331, 332)

Histone kinases	Phosphorylation of Serine or Threonine residues	Can affect chromatin structure and gene expression, depending on the specific enzyme and the site	Aurora B Kinase, MSK1	(290)
Histone phosphatases	Dephosphorylation of Serine or Threonine residues		PP1, PP2A	(290)
Histone ubiquitin ligases	Addition of ubiquitin respectively	Can affect gene expression, transcriptional activation or repression, and DNA repair	RNF20, RNF40, MDM2	(333)
Histone deubiquitinating enzymes	Removal of ubiquitin respectively		UBE2D1, UBE2D2, and UBE2D3	(306, 334)
SUMO ligases	Addition of SUMO respectively	Can affect gene expression, transcriptional activation or repression	PIAS	(304)
SUMO-specific proteases	Removal of SUMO respectively		SENPs	(304)

1.4.3 DNA methylation

DNA methylation is a crucial epigenetic modification that primarily occurs at CpG dinucleotides and involves adding a methyl group to the cytosine base of DNA (335). The dynamic regulation of CpG methylation patterns has a profound impact on gene expression and cellular function (336), affecting embryogenesis (337), tissue differentiation (338), cell fate determination (339), tissue-specific expression (340), and ageing (341). Growing evidence points to the association between abnormal DNA methylation patterns and a variety of human diseases (342), including neoplastic (343), musculoskeletal (344), neurological (345), and metabolic disorders (346). Therefore, DNA methylation-based therapies are being investigated as prospective therapeutic interventions (345), and DNA methylation biomarkers are being studied for disease diagnosis, prognosis, and therapy response testing (347-349). DNMTs catalyse DNA methylation by transferring a methyl group from S-adenosylmethionine (SAM) to the carbon-5 position of cytosine (350). DNMT1, also termed the maintenance methyltransferase, ensures the accurate replication of DNA methylation patterns during cell division (351). DNMT3A and DNMT3B are the de novo methyltransferases that create DNA methylation patterns during embryogenesis and tissue differentiation (352).

Various and intricate mechanisms regulate DNA methylation, including environmental factors like nutrition and chemical exposure, as well as epigenetic and genetic factors (353, 354) (See Figure 1.5). Two main mechanisms are involved in the regulation of DNA methylation (355): active processes involving enzymes, also termed active methylation, and passive processes

involving DNA replication and cell division, also known as passive methylation (356). An example of active methylation is ten-eleven translocation (TET) proteins that oxidise 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (357), which can then be modified to other forms of cytosine, such as 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (358). These alterations are reversible and result in changes in gene expression (359). Passive methylation describes how cell division and DNA replication affect DNA methylation patterns (360). During DNA replication, new unmethylated DNA strands are synthesised based on the original template strands (361). After replication, DNMTs restore the methylation patterns (361). However, errors can occur during this process, leading to changes in the DNA methylation patterns of daughter cells following cell division (362).

DNA methylation can have diverse effects on chromatin structure depending on its locus and context (325). Methylation of CpG islands in gene promoter regions generally cause gene repression (363). In this scenario, the methyl group promotes a closed chromatin structure by recruiting methyl-binding domain proteins, which in turn recruit other proteins that can regulate histone modification or obstruct the binding of transcription factors (364). In the genomes of normal cells, there is a negative association between elevated 5mC levels in CpG islands and transcriptional repression of the neighbouring transcription start site (TSS) (365). DNA methylation in other regions of the genome, such as the gene body (366) or enhancer region (367), can correlate with gene activation and an open chromatin state (366, 368). This mechanism is considered non-canonical, dynamic, and tissue-specific (366). Here, the methyl group may stabilise the binding of transcription factors or attract proteins that support an open chromatin state (368, 369).

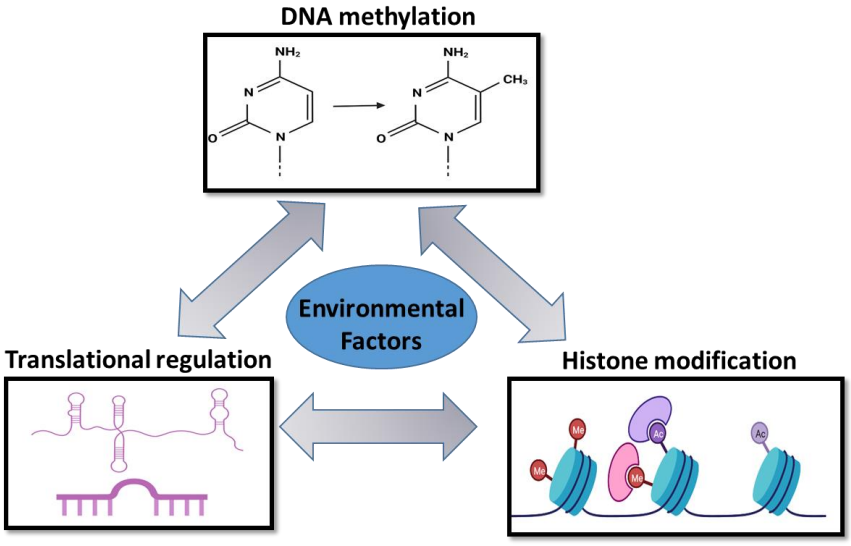


Figure 1.5. The interplay of environmental factors, cellular factors, and epigenetics in DNA methylation is a complex process. Environmental factors contribute to epigenetic modifications, which are sustained by two major processes: DNA methylation and histone modification. DNA methylation and histone modification, both epigenetic factors, interact with translational regulation, a cellular factor, to shape gene expression and cellular function.

1.4.4 Translational regulation

Translational regulation is a fundamental process that controls gene expression in cells by acting on the protein synthesis from mRNA molecules (370). Translational regulation is generally considered a cellular factor rather than an epigenetic factor according to the formal definition of epigenetics (229, 370, 371). However, translational regulation can interact with leading epigenetic factors to control gene expression (371, 372). Dysregulation of translational control is involved in several disorders, including neoplastic (373), musculoskeletal (374), and neurological disorders (375). The regulation of mRNA translation can occur at various stages, including initiation, elongation, and termination of protein synthesis (376), as well as the modulation of ribosome activity and mRNA stability (377). As described in Table 1.5, several factors can influence translational control, including non-coding RNAs, RNA-binding proteins, and ribosomal proteins (378).

Translational regulation can occur due to the presence of regulatory regions in the mRNA sequence (379) (See Table 1.4), as well as the binding of translational repressors or activators that control ribosome recruitment or activity during translation (378). For example, upstream open reading frames (uORFs) in the mRNA sequence regulate translation initiation by controlling ribosome accessibility to the main coding region (380). uORFs are minor ORFs that precede the main coding sequence and can be translated into short peptides (380). uORFs can play a positive or negative regulatory role in translation depending on the cellular context (381). Other mRNA sequences also mediate translational regulation, including internal ribosome entry sites (IRES), which allow ribosomes to bypass the canonical cap-dependent translation initiation and start translation at internal locations (382, 383). This mechanism is crucial when cap-dependent translation is impaired in cellular stress or viral infection (383). Apart from regulatory regions in the mRNA sequence, translational regulation involves mechanisms mediated by miRs and lncRNAs (384) (See Table 1.4). These regulatory factors interact with sequences or structures in mRNA to influence its stability, activity, and ribosome recruitment (385).

Table 1.5. Overview of molecules and factors that mediate translational regulation.

Factors	Description	Examples	Key papers
lncRNAs	Long non-coding RNAs that can interact with mRNA molecules and affect their translation.	MALAT1, HOTAIR, XIST, and NEAT1.	(386)
miRs	Small non-coding RNAs that bind to mRNA molecules and inhibit their translation	miR-21, miR-155, and let-7.	(387)
mRNA sequence	Specific sequence and structure of mRNA molecules that harbours regulatory elements.	RNA motifs, uORFs, IRES, 5' and 3' untranslated regions (UTRs).	(388)
RNA Modifications	Chemical modifications to RNA molecules, such as m6A methylation and pseudouridylation. They affect mRNA stability, localisation, and translation efficiency.	N6-(m6A) methylation, and N1- (m1A) methylation	(389)
RNA Binding Proteins	Proteins that bind to RNA molecules and regulate their stability, localisation, and translation.	HuR, PUM2, and IGF2BP1.	(390)
Translation Initiation Factors	Proteins that facilitate the initiation of translation by binding to mRNA and ribosomes.	eIF4E, eIF4G, and eIF4A.	(391)
Ribosomal Proteins	Proteins that are part of the ribosome and directly involved in translation.	RPL11, RPS3, RPS14, and RPS28.	(392)

1.4.4.1 Roles and classification of lncRNAs in gene regulation

LncRNAs are a class of transcripts longer than 200 nucleotides with no or low coding RNA content (393). Although the exact roles of this class of RNAs is still under investigation, they play critical regulatory roles in many physio-pathological processes including development, growth and oncogenesis (394-396). LncRNAs take part in these processes by regulating the pattern of gene expression at the transcriptional and post-transcriptional level (397, 398).

LncRNAs are a heterogenous group of RNAs that may be classified in several ways (399) (See Figure 1.6). Based on their function, lncRNAs can be classified as decoy, scaffold, and guide RNA (400) (See Figure 1.7). Decoy lncRNA are competitive endogenous RNAs (ceRNAs) that interact with miR in the cytoplasm to enable the expression of the mRNAs that would have been otherwise downregulated (401). Several lncRNAs acts as scaffold to recruit and assemble chromatin remodelling complexes, resulting in altered gene expression (402).

lncRNAs can act as guides, directing TFs to specific genomic regions to influence transcription of nearby genes (403).

Based on their genomic location, lncRNAs are divided into sense, antisense, bidirectional, intronic and intergenic (See Figure 1.6) (399). The sequence of sense or antisense lncRNAs overlaps respectively with the sense or antisense strand of a coding gene (404, 405). Most sense and antisense lncRNAs are mRNA-like, with the sense lncRNA being less investigated for their function (404, 406, 407). Some sense lncRNAs, such as *SRA*, have been found to function both as lncRNA by acting as a scaffold for regulating gene expression in the nucleus and as protein-coding transcripts in the cytoplasm (408). The sequences of intronic and bidirectional lncRNAs are respectively derived from an intron of a coding gene, and in opposite orientation to a nearby coding gene (<1kb) (409). Intergenic lncRNA are not located near protein genes (410). LncRNAs can also be generated from promoter or enhancer regions. These are termed respectively pRNA or PROMPTs and eRNAs, they contribute to gene control by modifying the chromatin architecture (411, 412). Based on their mechanism of action, lncRNAs can be classified as cis-, trans and ceRNA. Cis- and transRNA respectively regulate the expression of genes near and distant to their transcription site (413, 414). Based on their subcellular fate, lncRNAs are classified as cytoplasmic, nuclear, exosomal, and nuclear body-residing RNAs (415, 416).

The biosynthesis of lncRNAs is analogous to that of mRNA with some variations in the processing. Most lncRNAs are capped, polyadenylated and spliced by the canonical mRNA transcription mechanism (410, 417, 418). These lncRNAs are termed mRNA-like lncRNAs (419-421). Other lncRNAs can be generated through non-canonical mechanisms such as cleavage by ribonuclease P (RNase P) to process mature 3’ ends, capping by snoRNA to enhance their stability, and generation of non-polyadenylated circular structures to protect the lncRNA from degradation (417, 422, 423).

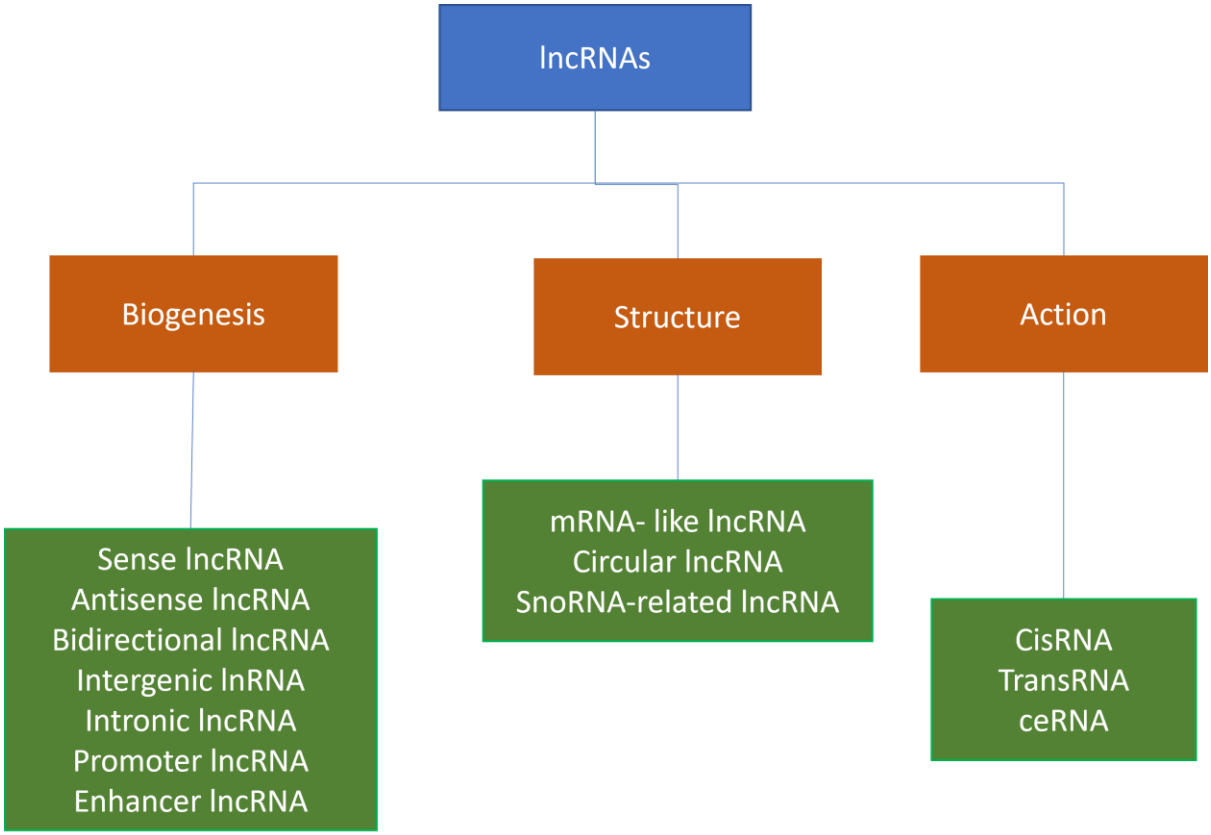


Figure 1.6. Classification of lncRNAs. The schematic shows how lncRNAs can be organised into classes and subclasses based on their action, biogenesis, and structure.

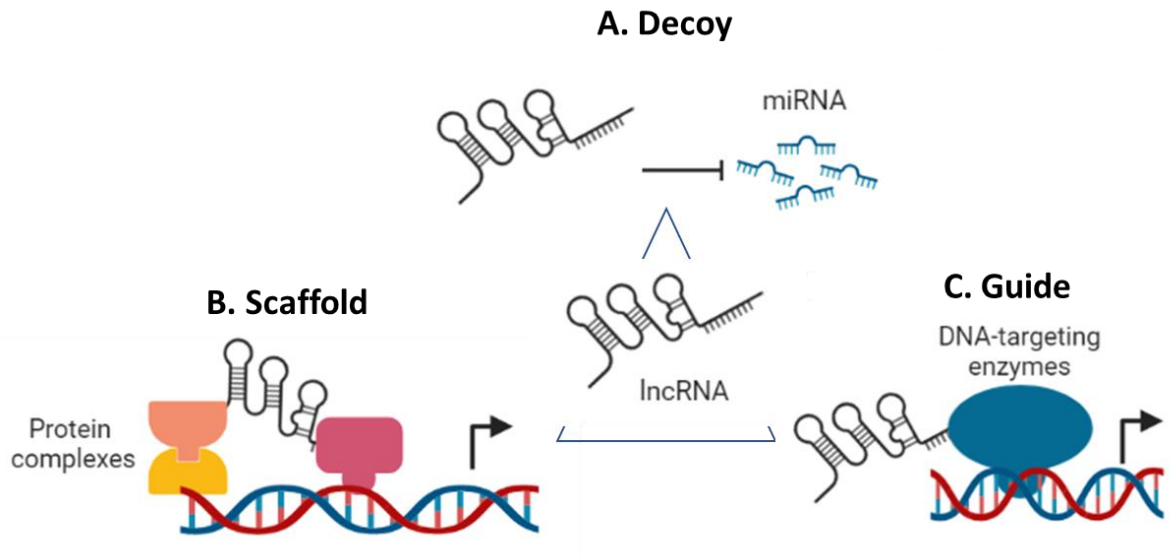


Figure 1.7. Functions of lncRNAs. A) Some lncRNAs can operate as sponges to sequestrate miRs away from their mRNA targets, thereby altering miR activity and indirectly influencing gene expression. B) Certain lncRNAs can act as scaffolds to aid in the recruitment and assembly of chromatin remodelling complexes, which can affect chromatin remodelling and alter gene expression. C) Other lncRNAs operate as guides to lead transcription factors (TFs)

to specific genomic areas, where they impact the transcriptional activity of neighbouring genes.

1.4.4.2 MicroRNAs and gene regulation

Small, non-coding RNA molecules called microRNAs (miRs) are a highly conserved class of RNA that range in size from 19 to 25 nucleotides. miRs were once believed to be evolutionary debris with no discernible function, however they are now known to regulate gene expression predominantly by silencing their target genes. miRs have been since documented in regulating physiological and pathological processes such as development, cell differentiation, carcinogenesis, and autoimmune disorders (424). Several biological processes are involved in the highly complex process of miR synthesis as shown in Figure 1.8 (425). The mature miR sequence is found inside a lengthy molecule known as the primary miR transcript (pri-miR), which has a distinctive local hairpin structure (426-428). The endonuclease DROSHA, which forms the microprocessor complex with the RNA binding protein DGCR8 (Di George syndrome critical-related gene 8), cleaves the pri-miR into pre-miR (429-431). Some pre-mRNA introns contain RNA sequences termed mirtrons that encode certain miRs (432). In this case, pre-mRNA splicing is required to generate the pre-miR (432). Approximately 50% of all presently known miRs are generated primarily from introns and a few exons of coding transcripts (intragenic), while the remainder are transcribed independently of a host gene and regulated by their own promoters (intergenic) (433-435). miRs can be transcribed as a single long transcript called a cluster, which may have similar seed regions and is considered a family (433, 436). After pre-miR formation, XPO5/RanGTP complex exports the RNA from the nucleus to the cytoplasm (437). Pre-miR is further processed in the cytoplasm by DICER, an endonuclease that creates a miR duplex (438, 439). Although both miR strands can be functional, one of them is destroyed during the miR processing and the other joins AGO and GW182 proteins to form the RNA-induced silencing complex (RISC) (440, 441). Pre-miRs can be processed independently of DICER, in this case AGO2 mediates the cleavage as shown in Figure 1.8 (442-444). The mature miR structure possesses a distinctive RNA sequence termed the seed domain, which binds to the target mRNA to promote its destabilisation or degradation (445, 446). Additionally, miRs can be exported into the extracellular environment to act as autocrine, paracrine, and/or endocrine regulators (433, 447). Once generated, miRs in most cases interact with the 3'UTR of target mRNAs to inhibit their expression by promoting the decapping of the mRNA and consequent mRNA degradation (445, 446). Although most research demonstrate the inhibitory functions of miRs, some research has shown that in peculiar cases miRs may upregulate mRNA expression by binding to the 5' UTR (448-451). Here, The AGO2-miR complex may promote interaction with ribosomes or acts as scaffold for binding translation factors (452). This molecular action is found in quiescent (GO) somatic cells and frog oocytes, and is mediated by AGO2 and FXR1, which form the RISC complex (453, 454).

.

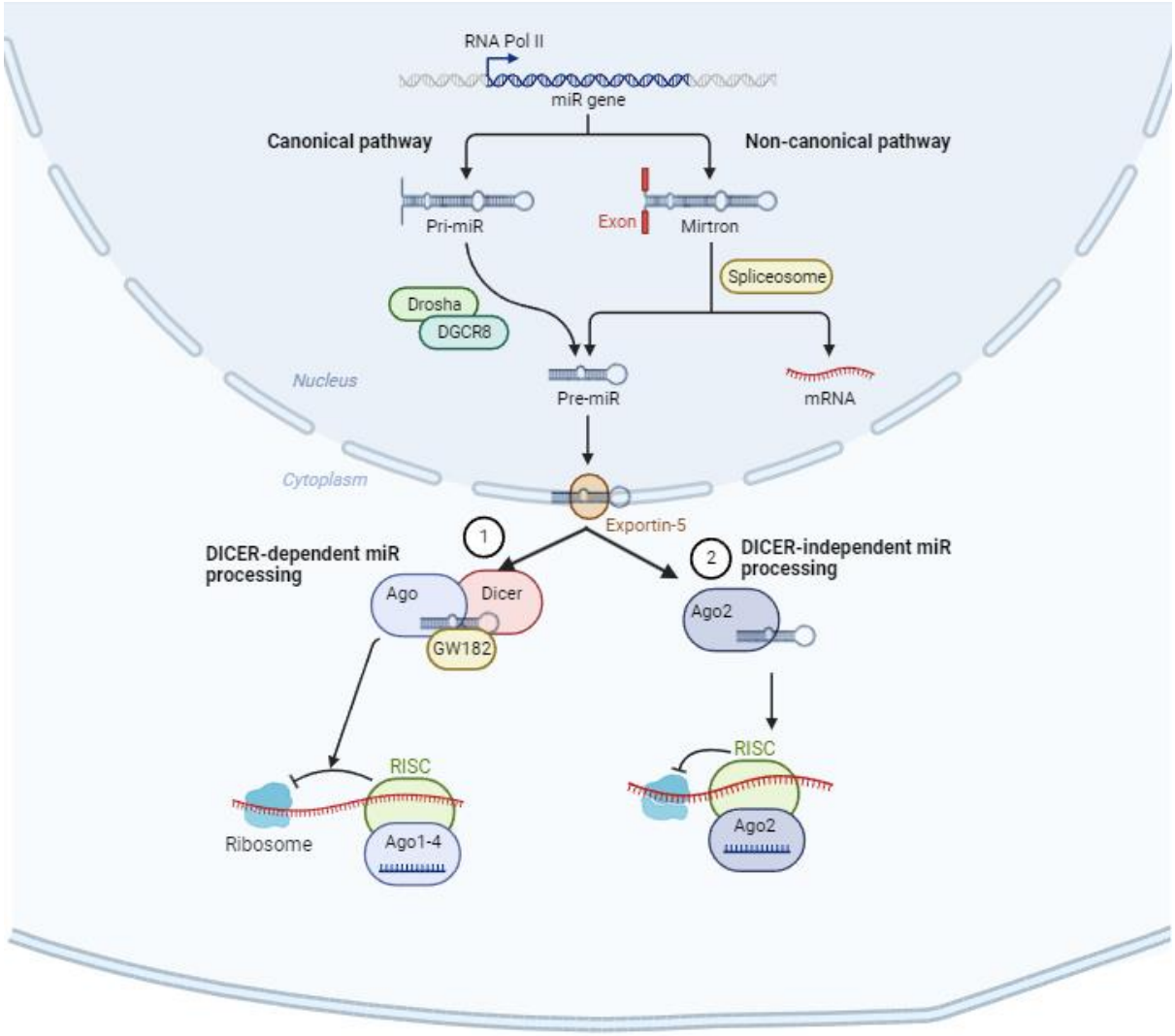


Figure 1.8. Schematic representation of the canonical and non-canonical miR biogenesis pathways. Pre-miRs are transcribed in the nucleus. In the canonical pathway, the transcript is independent from a host gene and processed via DROSHA/DGCR8, while in the non-canonical pathway the transcripts a processed from the intronic regions of coding transcript via the spliceosome. The miRs are exported into the nucleus, where they are further processed through two alternative pathways. 1) In the DICER-dependent pathway, the miRs are processed by DICER to form the RNA-induced silencing complex made up of the miR, AGO and GW182. 2) In the DICER-independent pathway, AGO2 catalyses the pre-miR into miR. An example of a miR that is produce through the DICER-independent pathway is miR-451 (442, 455).

1.4.4.3 Regulating miR abundance: mechanisms in gene regulation and role of ceRNAs

The differential expression of miRs regulates the expression of genes that modulate several disorders including HO. A question arises, what regulates miR differential expression? Multiple factors and molecular mechanisms have been shown to regulate miR differential expression such as DNA copy number (456), methylation of CpGs (457), transcription factors downstream signalling pathways (458). Regulatory RNA termed ceRNAs or miR sponges can also regulate miR abundance. This class of RNA include lncRNAs, circular RNAs and pseudogene transcripts (459). Moreover, abundantly expressed mRNAs have been shown to act as miR sponges by sequestrating miR from other mRNA targets (460). Generally, miRs are stable and promote the degradation of their mRNA target. However, recent studies have shown that ceRNAs such as lncRNAs may promote the degradation of miRs (460-463). This process is termed target-directed miR degradation (TDMD), and it is believed to be regulated by the complementarity between miR and RNA target (461, 464). The stability of the miR is promoted when partial pairing is formed with the RNA target as shown in Figure 1.9 (A-B). However, extensive pairing with the RNA target induces the degradation of miR as shown in Figure 1.9 (C-D) (461, 464). Despite current advances in investigating how miR may interact with lncRNAs to regulate HO (451, 465), further work is required to describe the regulatory functions of lncRNA-miR axis in HO and other conditions.

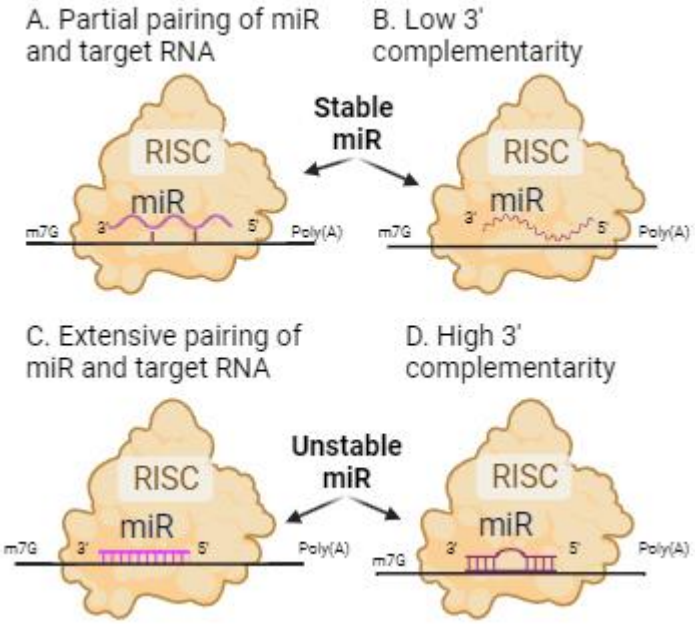


Figure 1.9. Stable miR versus TDMD. miRs are stable if the pairing with target RNA is partial (A) or have a low 3' complementarity (B). This process leads to the degradation of the target RNA. MiRs are degraded when they are extensively paired with the target RNA (C) or a high 3' complementarity (D).

1.5 LncRNA biology in HO

Growing evidence has shown that lncRNAs play a major role in HO. An increasing number of lncRNAs, comprising *H19* (466), *DANCR* (467), *MEG3* (468), *MALAT1* (469), *HOTAIR* (470), are differentially expressed and regulate markers and pathways in bone/cartilage formation (471). The following lncRNAs have been shown subsequently to regulate normal and/or heterotopic bone formation *in vivo*: *H19* (472), *MIAT* (473), *lncRNA-OG* (474), *HOXA-AS3* (475). Below I discuss the biogenesis of lncRNAs and their functional role in bone formation and HO by considering how they interact with DNA, proteins, and miRs.

1.5.1 LncRNAs modify chromatin structure to regulate bone formation

LncRNAs interact with nucleosome-remodelling factors and chromatin-modifying enzymes to mediate chromatin remodelling, (253). For example, *HOTAIR* recruits and binds to lysine-specific histone demethylase 1A (LSD1) and polycomb repressive complex 2 (PRC2) to target and promote demethylation of histone H3K4 and trimethylation of H3K27, which ultimately condenses the chromosomes and silences the expression of targeted genes (255) (Figure 1.10). Knockdown data using osteoblastic SaOS-2 cells has shown that *HOTAIR* decreases H3K4 methylation levels in the alkaline phosphatase (ALP) promoter region to repress mineralisation (476). The lncRNA *XIST* accelerates methylation of the *TIMP-3* promoter to induce collagen degradation in osteoarthritic chondrocytes after tibial plateau fracture (477). Hypoxia-inducible factor 1α (*HIF1α-AS1*) has been shown to promote TGF-β-induced osteoblastic differentiation of hBMSCs by promoting acetylation, which then leads to expression of homeobox D10 that plays a key role in osteoblast differentiation (478). Anti-differentiation non-coding RNA (*ANCR*) is downregulated during osteogenic differentiation of MSCs (479). *ANCR* has been shown to recruit EZH2 to promote the trimethylation of H3 lysine-27 in the *RUNX2* gene promoter to inhibit transcription of *RUNX2*, and therefore, the osteogenic differentiation of periodontal ligament stem cells (PDLSCs) (480). In an *in vivo* model for HO, in which stem cells were seeded onto TCP/HA and implanted in 10 weeks old BALB/c nude mice to induce the HO, lncRNA *HOXA-AS3* has been shown to interact with EZH2 for H3K27 trimethylation to upregulate RUNX2 (475).

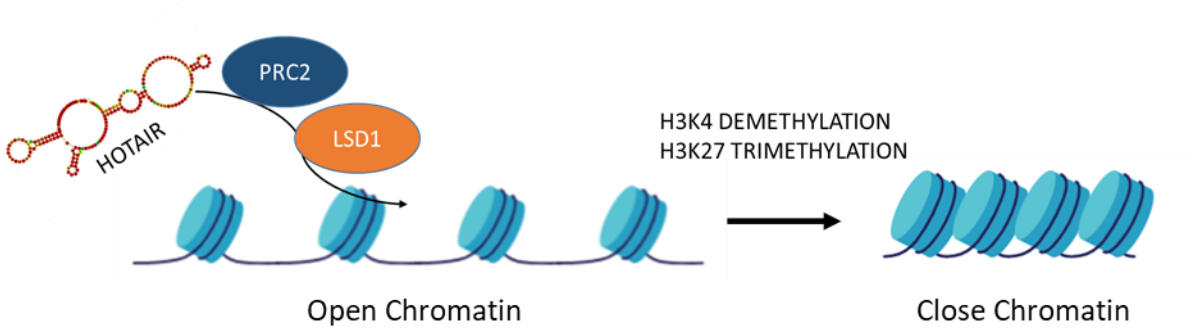


Figure 1.10. *HOTAIR* scaffolding. *HOTAIR* interacts with chromatin and recruits PRC2 and LSD1 to demethylate histone *H3K4* and trimethylate *H3K27*. *HOTAIR* directly interacts with the chromatin and act as scaffold for these complexes. This event leads to the condensation of chromatin, which ultimately silences gene expression of target genes (255).

1.5.2 LncRNAs modulate transcriptional process to regulate bone formation

LncRNAs can recruit transcription factors (TFs), bind with RNA Pol II, or interfere with the polymerase-promotor bind to promote or inhibit gene transcription (481). Maternally expressed gene 3 (*MEG3*) has been shown to increase during osteogenic differentiation of MSCs. *MEG3* knockdown inhibited the expression of RUNX2, OSX and OCN and the overexpression enhanced the expression of these markers. MEG3 was mechanistically shown to activate BMP4 transcription by disassociating the transcription factor SOX2 from the BMP4 promoter (482). *MEG3* was shown to be antiosteogenic in MSCs from patients with postmenopausal osteoporosis (PMOP) by upregulating the expression of miR-133-a3p (468). *MEG3* has been shown to positively regulate HO *in vivo* in mice by enhancing BMP4 expression (483). In this study, stem cells were seeded on TCP/HA and then implanted in 10-week-old BALB/c mice to induce HO. These data indicate that the direction of *MEG3* effect on osteoblast differentiation depends on the disease state (Figure 1.11). In some cases, the mature transcripts of the lncRNA are not necessary as the transcription process itself recruits TFs to regulate nearby genes (481). This feature is typical of low abundance lncRNAs and was demonstrated by Engreitz and colleagues, who showed that manipulation of genomic loci that produce lncRNAs affected the expression of a neighbouring cis-acting gene (484).

Aside from promoting the differential expression of genes, a new lncRNA class termed sno-lncRNAs can promote expression of specific splice variants by targeting splicing factors (485). These lncRNAs are flanked by small nucleolar RNA (snoRNAs) and derive from introns. The alternative splicing factor FOX2 is inhibited by a sno-lncRNA from the 15q11-13 chromosomal region (485). LncRNAs termed natural antisense transcripts (*NATs*) can also modulate mRNA processing during transcription owing to their anti-sense orientation to protein-coding genes, hence their name (486, 487).

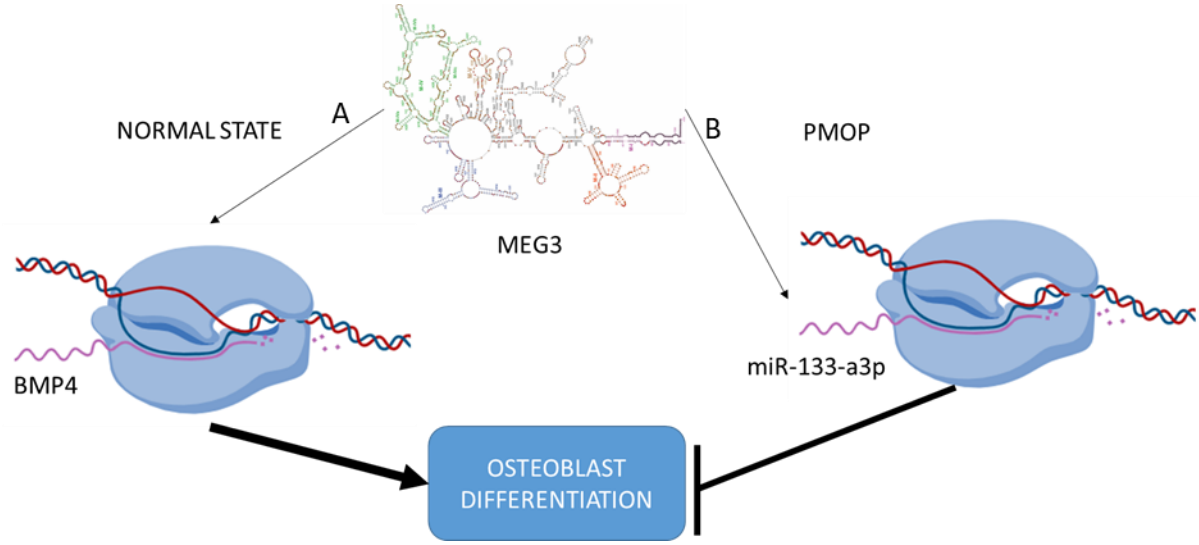


Figure 1.11. *MEG3* binds with TFs to regulate osteoblast differentiation. A) In normal MSCs *MEG3* promotes the transcription of BMP4, which induces osteodifferentiation (482). B) In MSCs from PMOP patients, MEG3 promotes the expression of miR-133-a3p to inhibit the differentiation (468). *In vivo* mice data showed that MEG3 promotes BMP4 expression to regulate HO, this is consistent with the model A (483).

1.5.3 LncRNAs modulate translational processes to regulate bone formation

LncRNAs have been associated with both negative and positive regulation of mRNA stability. For example, *BACE1AS*, an antisense lncRNA that derives from *BACE1* locus, increases the stability of *BACE1* mRNA and competes with mir-485-5p for the same region (488, 489). *BACE1* and *BACE1AS* are suggested to be involved in Sporadic inclusion-body myositis (s-IBM) that is an inflammatory myopathy (490). Alu-repeat containing lncRNAs target mRNA transcripts for Staufen-mediated decay, which is a cellular mechanism involving the degradation of mRNA (491). In the last decade, the ceRNA hypothesis has been developed. The theory proposes that there is a prevalent network of crosstalk between non-coding and coding RNAs that reveals itself through competition for miR binding (401). The LncRNA *H19* and *IGF1*, the gene encoding a fundamental growth factor in skeletal health (IGF1), were highly co-expressed and miR-185-5p was poorly expressed in mineralized cells (492). By knocking down *H19*, matrix mineralisation and *IGF1* expression were inhibited and miR-185-5p was highly expressed. Wu and colleagues showed that H19 directly interacts with miR-185-5p to promote its degradation and to sequestrate it from suppressing *IGF1* expression (492) (Figure 1.12a). This study demonstrated and characterized the function of a lncRNA-miR-mRNA axis in modulating bone formation. Jia and colleagues have characterized the function of the *LINC00707*-mir370-3p-*WNT2B* axis in regulating osteogenesis of hBMSCs, where lncRNA *LINC00707* promotes osteogenic differentiation by sequestering mir370-3p from interacting with *WNT2B* (493). Several lncRNAs have been shown to target miRs to drive bone formation (468-470, 479, 494-496). As well as interacting with miRs to regulate the translation of genes, lncRNAs can act as a reservoir of miRs. In this mechanism, the lncRNA is transcribed and then

further processed to generate mature miRs that can affect biological processes (472, 497, 498). For instance, *H19* has been shown to act as a developmental source of miR-675, which suppresses growth through its action on IGFR1 (497). The controlled release of mir-675 was suggested to inhibit cell proliferation in response to oncogenic signals or cellular stress (497). It was later demonstrated that *H19*/miR-675 inhibits the expression of *TGF-β1* and *HDAC* to inhibit SMAD3 phosphorylation and increase the expression of osteogenic genes in an *in vivo* model for HO (472) (Fig 1.12b). In this study, hMSCs were induced under osteogenic medium for 1 week, mixed with collagen scaffolds and implanted in BALB/c nude mice to induce the HO. Interestingly mir675-5p has been shown to target *H19* in self-regulatory feedback to regulate osteoblast differentiation (499).

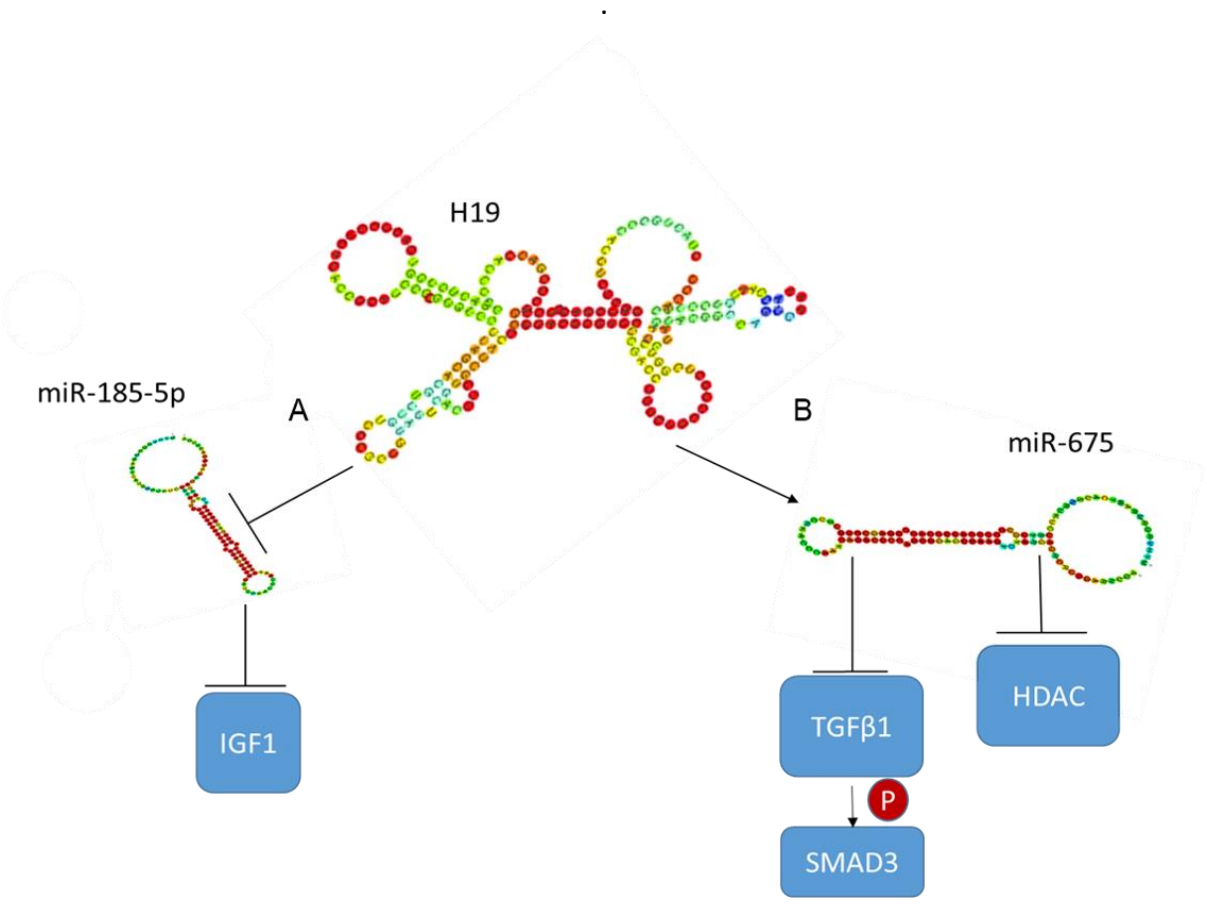


Figure 1.12. A *H19* act as a ceRNA and as a source of miR. a *H19* inhibit miR-185-5p to promote *IGF1* expression, which in turns drive osteogenesis (492). B *H19* act as a source of miR-675, which in turn targets *TGF-β1* to inhibit the phosphorylation of SMAD3 and inhibits the expression of *HDAC* to promote the expression of osteogenic genes (472).

1.5.4 miRs regulate HO

Recent studies have been focused on investigating the functional role of miRs to comprehend the ethology of musculoskeletal ageing diseases (500, 501). Dicer was deleted in osteo-chondroprogenitor cells in Prx-1-Cre mice to test whether miR synthesis and function affects the musculoskeletal system (502). This resulted in formation of smaller limbs due to increased cell death (502). Other research studies have shown how miR synthesis is necessary for physiological musculoskeletal development (503, 504). Several miRs have been described to specifically regulate HO (451). Ji and colleagues found that miR-205 and -215 were upregulated in HO patients versus control. In the same study, muscle-specific miRs (miR1, miR26a, miR133a, miR133b, miR146b, and miR206) also termed myomiRs were found to be differentially regulated (451). MyomiRs have been shown to regulate skeletal muscle and promote the proliferation and differentiation of muscle stem cells (505). Therefore, differential expression of myomiRs in HO may support the notion that defective differentiation of myogenic cells contributes to HO (451, 505). By comparing the miR profile in the serum of patients with immature post-traumatic HO (1 month after HO became visible) versus mature HO samples (6 months after HO appearance), Sun and colleagues found that miR-630 is downregulated during HO and that can serve as an early HO marker (212). The miR was found to target Slug, which is involved in the endMT of endothelial cells. Downregulation of miR630 was found to increase the osteogenic differentiation of endothelial cells (212). By comparing miR expression in normal human bone with traumatic heterotopic bone, Tu and colleagues have described how miR-203 may regulate HO. The miR was found to be decreased in HO and to directly target *Runx2*. The downregulation of miR-203 increased *Runx2* expression and promoted osteoblast differentiation *in vitro* and HO *in vivo*. Other miRs have been shown to regulate HO, bone/cartilage development and musculoskeletal health such as miR-320e (506), miR-433 (507), miR-17-5p (213), miR-140-5p (508), miR-342-3p (509), and miR-485 (510, 511). Taking together, these data demonstrate a role of miRs in regulating HO and that they represent a potential strategy in the diagnosis and treatment of traumatic HO.

1.6 Hypothesis and Aims

The PhD project hypothesises that *CASC20* is a novel regulator of bone formation. *CASC20* may regulate biological processes by modifying chromatin structure, modulating transcriptional processes or by modulating translational processes. The specific aims of this study were as follows:

1) Explore whether *CASC20* has a 3' poly-A tail to enable potential action as a ceRNA to miRs active in osteo and chondrogenesis using *in vitro* and *in silico* approaches:
2) Examine the effect of *CASC20* overexpression on osteogenesis and key marker genes *in vitro*.
3) Examine the effect of *CASC20* overexpression on chondrogenesis and key marker genes *in vitro*.
4) Conduct RNA-Sequencing to characterise the miRome changes in *CASC20* overexpressing cells during osteo- and chondrogenesis *in vitro* and predict the effect of these differentially regulated miRs on target genes.

Chapter 2 – *In silico* modelling of *CASC20* interactions with micro-RNAs (miRs) and *in vitro* analysis of selected candidate miRs

2.1 INTRODUCTION

During my MRes, I determined whether *CASC20* is expressed in human bone tissue by extracting total RNA from fresh frozen, surgically excised bone from patients undergoing joint replacement. *CASC20* expression was confirmed by real time quantitative polymerase chain reaction (RT-qPCR, Figure 2.1A). Next, I explored whether *CASC20* is differentially expressed in human multipotent adipose-derived stem cells (hMAD, Figure 2.1B)(512) and in primary human bone marrow-derived mesenchymal stem cells (hMSCs, Figure 2.1C-E) in response to stimulation with BMP2 and osteogenic supplements. I found that *CASC20* expression was significantly upregulated by day 8 in hMADs, and by day 16 or 24 in hMSCs. This was associated with robust upregulation of markers of osteogenic differentiation runt-related transcription factor 2 (*RUNX2*) and the transcription factor Sp7/Osterix (*OSX*) in both the hMADs and the hMSCs at these timepoints, and a significant increase in percentage mineralisation per well at day 24, measured by Alizarin Red S stain.

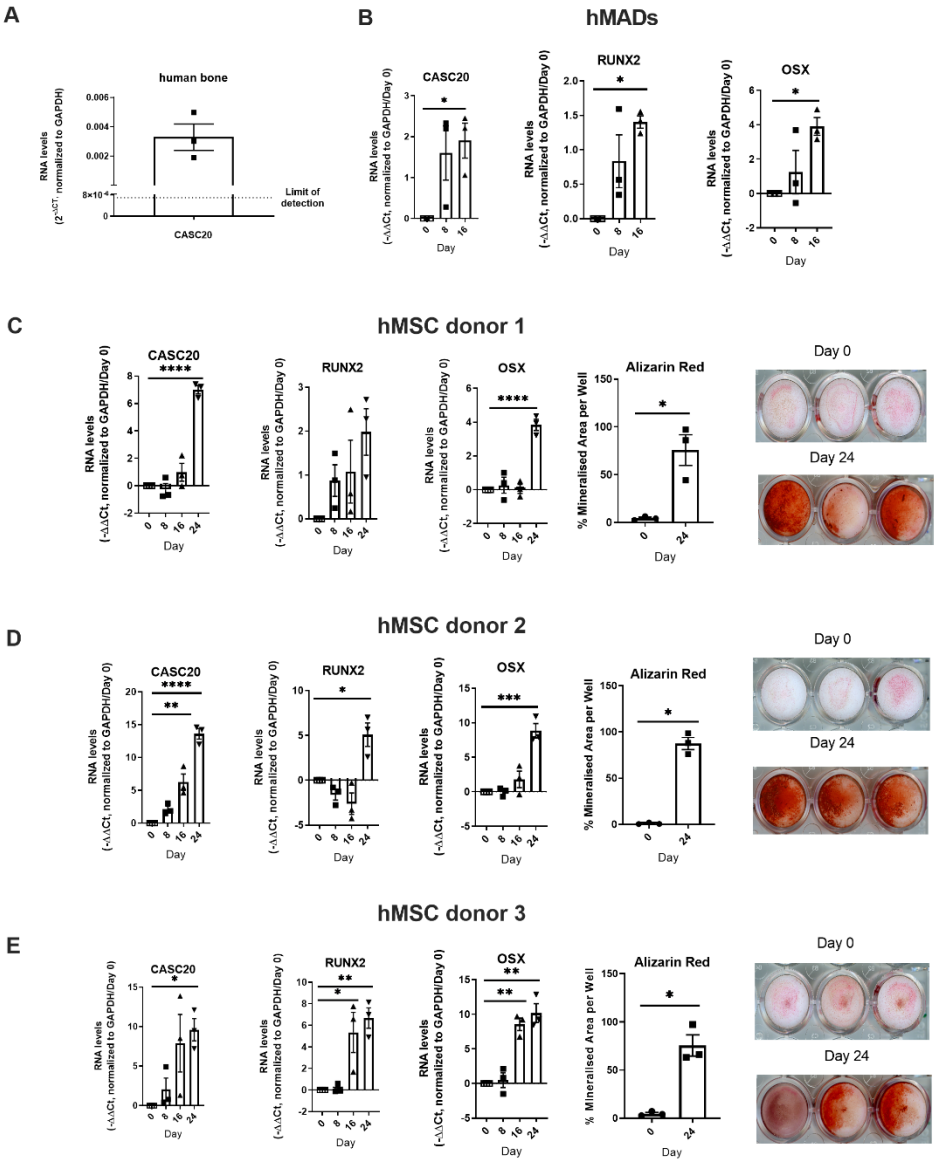


Figure 2.1. Reproduced from MSc thesis. *CASC20* is expressed in human bone and is induced in mesenchymal stem cells by BMP2 *in vitro*. A) RT-qPCR was used to measure the expression of *CASC20* RNA in waste bone samples retrieved at joint replacement (N=3 subjects). B) *CASC20* is induced in hMADs. RT-qPCR was used to measure the expression of *CASC20*, *RUNX2* and *OSX* at days 0, 8 and 16 of hMAD differentiation. Data were analysed using 2^(-ΔCt) by normalising to GAPDH (N=3, technical replicates). C-E) RT-qPCR was used to measure the expression of *CASC20*, *RUNX2* and *OSX* at days 0, 8, 16, and 24 of hMSCs differentiation. Data were analysed using (-ΔΔCt) by normalising to GAPDH/Day0 (N=3, technical replicates).

Several molecular mechanisms for lncRNA action have been described. Transcripts without 5'capping and a 3' poly-A tail may act as antisense transcripts or interfere with enhancer regions (513). In contrast, lncRNAs that have canonical 5' cap and poly-A tail are exported to the cytoplasm where they can serve as a competing endogenous RNA (ceRNA) for specific miRs (514), thus indirectly regulating the expression of the miR targets (515).

As a first step in my PhD studies, I used DU145 prostate cancer cells that are known to express *CASC20* (516) to demonstrate that *CASC20* has a poly-A tail. Based on this proposition, it was further hypothesised that *CASC20* may serve as a competing endogenous RNA (ceRNA). However, due to the unforeseen circumstances of the COVID-19 pandemic coinciding with my doctoral research, I was compelled to adapt my methodology and focus on an extended *in silico* investigation during the period of initial long lockdown. In this work I explored *in silico* modelling of *CASC20* interactions with miRs. I predicted *CASC20*-interacting miRs using miRanda (http://www.microrna.org), and then characterised the expression and functional role of putative *CASC20*-interacting miRs using Gene Ontology (GO) analysis and by analysis of published datasets.

Upon the resumption of laboratory access, I examined selected candidate miRs *in vitro* as an initial exploration of how *CASC20* may interact with miRs in osteogenesis and chondrogenesis. To examine the impact of *CASC20* on potential miR targets, I transduced hMADs with *CASC20*-overexpressing lentivirus, differentiated the cells for 10 days into osteoblasts and collected the cells for RNA extraction and qPCR. I found that *CASC20*-overexpression robustly inhibited the expression of hsa-miR-485-3p during osteoblast differentiation.

2.2 MATERIALS and METHODS

2.2.1 Poly A-tail specific RT-qPCR analysis in DU145 cells

DU145 cells were cultured in DMEM (Thermo Fisher Scientific, 61965-059 – Waltham, Massachusetts) supplemented with 10% FBS (Thermo Fisher Scientific, 16000044), and 1%Penicillin/Streptomycin (P/S) (Sigma-Aldrich, P4333-100ML – Dorset, UK). The cells were seeded in 6-well plates and allowed to grow for 24 hours. RNA extraction was performed using the miRNeasy mini kit (Qiagen, 217004 – Manchester, UK) according to the manufacturer's instructions. For cDNA synthesis, the Precision nanoscript2 Reverse Transcriptase kit (Primer design, RT-NanoScript2-150 – Southampton, UK) was used. Reverse transcription of miR into cDNA was completed using the Veriti 96-well thermal cycler (Applied Biosystems, Waltham, MA) following the manufacturer's instructions. The cDNA synthesis was achieved by targeting the poly A tail using OligoDT or by targeting RNA sequences using random nonamers. For RT-qPCR, 2ng of cDNA was loaded per well. The qPCR samples were run on the C1000 Touch™ Thermal Cycler (Bio-Rad, 1851138 – Hercules, CA) using 384-well plates. Triplicate technical repeats were performed for each assay, and CT values were plotted and analysed to detect differences between OligoDT and random nonamers using GraphPad Prism software. The primers for SYBR Green qPCR (Sigma-Aldrich) were designed using Primer-BLAST (NCBI: www.ncbi.nlm.nih.gov/tools/primer-blast). To ensure primer quality, all primers were screened for self-complementarity using Oligonucleotide Properties Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html). PrecisionPLUS SYBR-Green master mix (Primer design, PPLUS-machine type-1ML) was used with SYBR primers.

Gene	Human qPCR Primer Sequence (5' -> 3')
GAPDH	FW ATTGCCCTCAACGACCACTTT
	REV CCCTGTTGCTGTAGCCAAATTC
β-Actin	FW GGATGACAGAAGGAGATCACT
	REV CGATCCACACGGAGTACTTG
CASC20	FW TCATATGGATTTCAAGCTGGGT
	REV TCCCAGTCTTCTGCATCACTTC

2.2.2 lncRNA-miR target prediction

lncRNA-miR target prediction analysis was made in MiRanda (miRanda-aug2010, http://www.microrna.org/) (517) using the default parameters with the following modification: Energy Threshold ≤-15.000000 kCal/mol (518). Query and Reference files were used in FASTA format: human mature miRs sequences were downloaded from miRbase 22.1 (http://www.mirbase.org/ftp.shtml) (519), the *CASC20* sequence was obtained from NCBI database (https://www.ncbi.nlm.nih.gov/)(520). The Encyclopedia of RNA Interactomes (ENCORI) database was downloaded (https://rnasysu.com/encori/) and used to filter the list of *CASC20*-interacting miRs to select only miRs that are experimentally validated to be functional by UV cross-linking and immunoprecipitation with high-throughput sequencing

(CLIP-Seq) (521). The RISmed R package (version 2.1.7, https://rdrr.io/cran/RISmed/) was used to perform an electronic search using the keywords osteoblast, chondrocyte, chondrogenesis, osteogenesis, BMP2, RUNX2, SOX9, bone, cartilage, skeletal, ossification, mineralisation, for all ENCORI-validated miRs with a role in osteo- or chondrogenesis that are reported in PubMed between January 01 2012 and May 15 2020 (search date May 15 2020).

The ENCORI database was also used to identify genes that are targeted by at least 1 *CASC20*-interacting miR. This list of genes was then ranked in order of the number of miRs with which they interact to create a *CASC20*-miR interaction hierarchy. The RISmed PubMed search strategy was then repeated to identify a published role for these genes in chondro or osteogenesis, comparing their frequency in the top versus the bottom quartile of the ranked gene list.

2.2.3 Pathway enrichment analysis

Pathway enrichment analysis was performed using the R package GOSeq (522) selecting Gene Ontology terms for molecular function and biological process (GOTERM_MF and GOTERM_BP) and using REViGO (http://revigo.irb.hr/) (523) to summarise the GO terms based on semantic similarity. Bonferroni multiple testing-corrected p values using log10(FDR) equal or less than -1.3 were considered a significant difference. Heatmaps of the GOTERM_MF and GOTERM_BP associated miRs were created using the R package Heatmap.2 (https://www.rdocumentation.org/packages/gplots/versions/3.0.3/topics/heatmap.2).

2.2.4 Expression of *CASC20*-interacting miRs in osteo- and chondrogenesis datasets

To characterise the expression pattern of the miRs predicted to be targeted *CASC20* during osteoblast differentiation, we analysed publicly available miR-Seq dataset GSE107279 (524). In this dataset, hMSCs (n=3) were differentiated into osteoblast using osteogenic medium comprising: 10nM dexamethasone, 0.2mM l-ascorbic acid, 10mM β-glycerophosphate and 10mM 1,25-dihydroxyvitamin D3 (Calcitriol). The cells were differentiated for 13 days and collected at the following time-points: 0h, 6hour, 12hour, 24hour, day 3, day 7, day 10, day 13. The dataset was downloaded from GEO database (https://www.ncbi.nlm.nih.gov/geo/) and miR with less than 10 raw counts across all the samples were filtered. Differential expression analysis was performed with limma (v 3.38.3) in R (525, 526). Log count per million+1 values were calculated using the EdgeR (v 3.24.3) in R (527). logFC < 0 was used identify *CASC20*-interacting miRs that were downregulated between day 3 and day 13, the timeframe over which *CASC20* expression was found to increase in our MSC stimulation studies. Normalised reads of *CASC20*-interacting miRs were used for constructing the heatmap using gplots (v 3.0.1.1) in R.

To characterise the expression pattern of the miRs predicted to be targeted by *CASC20* during chondrogenic differentiation, we analysed publicly available miR-Seq dataset GSE109503, from Barter et al (528). In this dataset, hMSCs were differentiated into chondrocyte using chondrogenic media: DMEM containing 100µg/ml sodium pyruvate (Lonza), 10ng/ml TGF-β3,

100nM dexamethasone, 1×ITS-1 premix, 40µg/ml proline, and 25µg/ml ascorbate-2-phosphate. The cells were collected at day 0, day 1, day 3, day 6, day 10, day 14. The Illumina whole-genome expression array Human HT-12 V4 (Illumina, Saffron Walden, U.K.) was used to profile gene expression. Normalised values of *CASC20*-interacting miRs were extracted and visualised in a heatmap constructed using the gplots package in R (v 3.0.1.1). logFC < 0 was used identify *CASC20*-interacting miRs that were downregulated between day 3 and day 13.

2.2.5 Tissue culture and osteogenic differentiation

Human multipotent adipose-derived stem cells (hMADs) were cultured in growth medium which consisted of: DMEM (Lonza, BE12-707F – Slough, UK), 10% FBS, 1%glutamine (Thermo Fisher Scientific, 25030-024), 0.2%P/S, 1%HEPES (Thermo Fisher Scientific, 15630-056) and 0.01%hFGF2 (Sigma-Aldrich, F0291). Confluency was avoided to prevent differentiation, with hMADs cells split every 2-3 days. For lentiviral transduction and osteogenic differentiation, hMADs were seeded for 24 hours in growth media, the cells were then transduced with lentivirus for 24 hours, as detailed in the following section. After that, the cells were washed with PBS and incubated with normal hMAD growth media for 24 hours. Then 300ng/ml human recombinant BMP2 (GenScript, Z02913-1 – Wanchai, Hong Kong) was added for 48 hours. The media was replaced with the osteogenic media on day 0 of the differentiation. Osteogenic media was made as to the growth media, except for the addition of 300ng/ml BMP2, 10mM β-glycerophosphate (Sigma-Aldrich, G9422), 10nM dexamethasone (Sigma-Aldrich, D8893) and 50µg/ml L-ascorbic acid (Sigma-Aldrich, A4544).

2.2.5 Lentiviral packaging and transduction

CASC20 gene was cloned from pUCIDT-AMP plasmid (IDT - Berkshire, UK) into pcDNA4/TO/Myc-His A plasmid (Thermo Fisher Scientific, V863-20 –Paisley PA49RF, UK) using BAMHI and EcoRV enzymes following manufacturer's instructions. After that, *CASC20* was cloned into the lentiviral plasmid pTwist+Lenti+SFFV+Puro+WPRE-Amp (Twist Bioscience, South San Francisco, CA) with enzymes BamHI and Xho following manufacturers' instructions (Figure 2.2). For lentiviral packaging, HEK293T cells were plated (3.8 x 10^6 cells) in 10cm dish (Cat. No. 10062-880, VWR) for 24 hours. The cells were cultured using 6mL of DMEM, 10% FBS, 1% glutamine, 0.2%P/S. After that, the cells were transfected with the following 3 plasmids using FuGene HD (Promega, E2311 – Southampton, UK):

- 3.15µg pTWIST-*CASC20* plasmid for *CASC20* OE or pGFP-C-shLenti plasmid for negative control (Origene, TR30023 – Herford, Germany).
- 2.5µg psPAX2 (a gift from David Young, Addgene plasmid #12260;http://n2t.net/addgene:12260; RRID: Addgene_12260)
- 0.63µg pCMV-VSV-G (a gift from David Young, Addgene plasmid #8454;http://n2t.net/addgene:8454; RRID: Addgene_8454)

The three plasmids were mixed in a 1.5 mL tube, and 19 µL of FuGENE HD was added. The volume was then adjusted to 220 µL using DMEM (no additional substances added). The mixture was incubated at room temperature for 15 minutes and subsequently added dropwise to the cells, which were then incubated for 24 hours. Following incubation, the cells were rinsed with PBS, and the media was changed. After another 24 hours, the media containing the virus was harvested and stored at 4°C. This process was repeated, and 24 hours later, the collected media was combined and centrifuged at 500g for 10 minutes. The resulting supernatant was transferred to a new tube. To concentrate the virus, Lenti-X™ Concentrator (Takara, 631231 – London, UK) was added to the media at a 1:3 ratio. The mixture was refrigerated for 30 minutes, then centrifuged at 4°C and 1500g for 45 minutes. The media was discarded, and the virus pellet was resuspended in PBS at a 10X concentration. This virus-containing PBS was divided into individual aliquots and stored at -80°C.

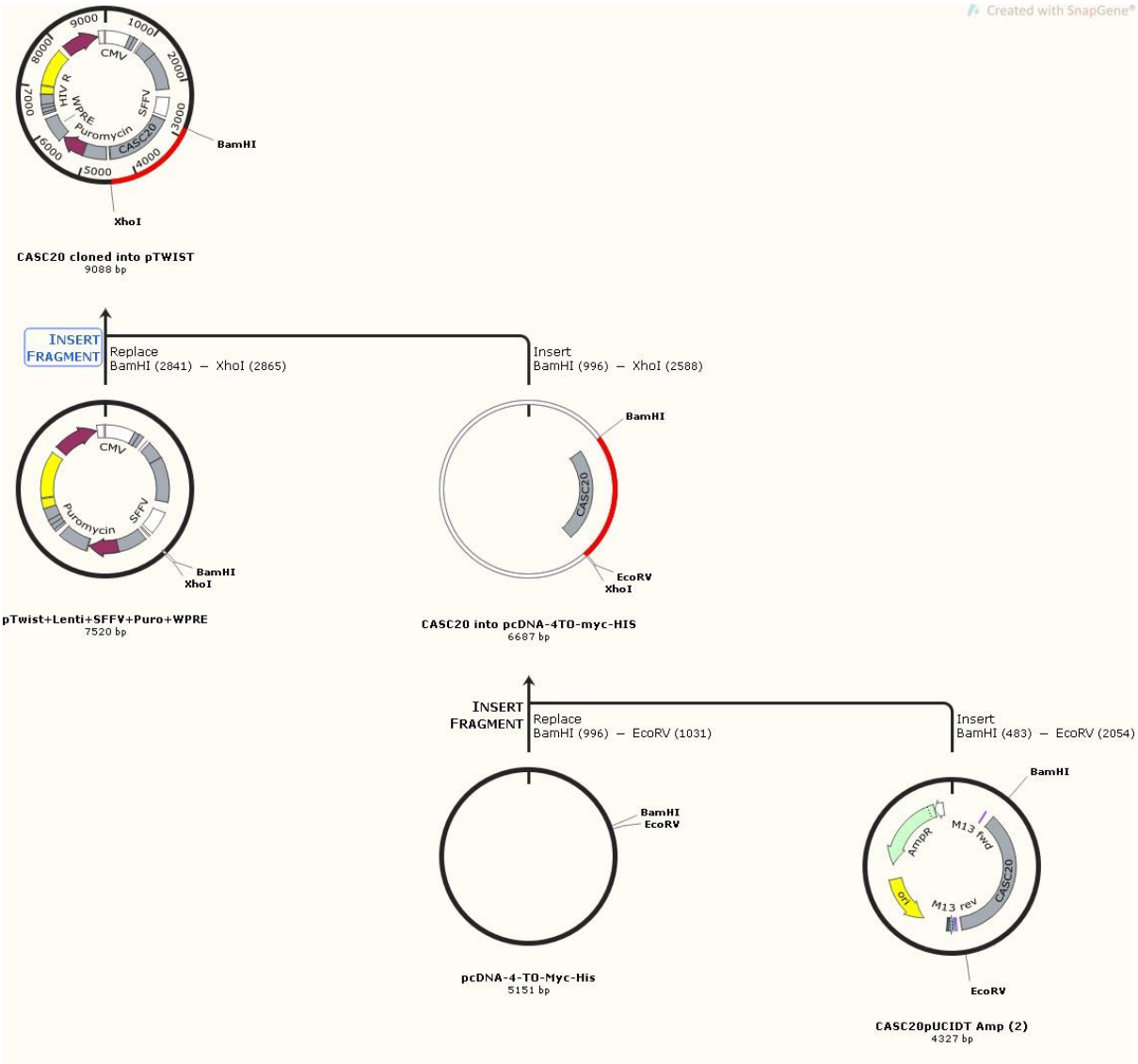


Figure 2.2. Cloning history of *CASC20* gene into lentiviral plasmid.

For lentiviral transduction, 20,000 stem cells were transduced with 10µL of lentivirus in the presence of polybrene (Sigma-Aldrich, TR-1003-G) at 8µg/mL, with a multiplicity of infection (MOI)= 0.9-1.0. Virus:cell ratio was determined following optimisations carried out using DU145 prostate cancer cell lines and then validated using hMADs (Figure 2.3 and 2.4). For subsequent downstream studies, tissue culture was conducted post-transduction without antibiotic selection to minimise any negative effects on physiological cell biology.

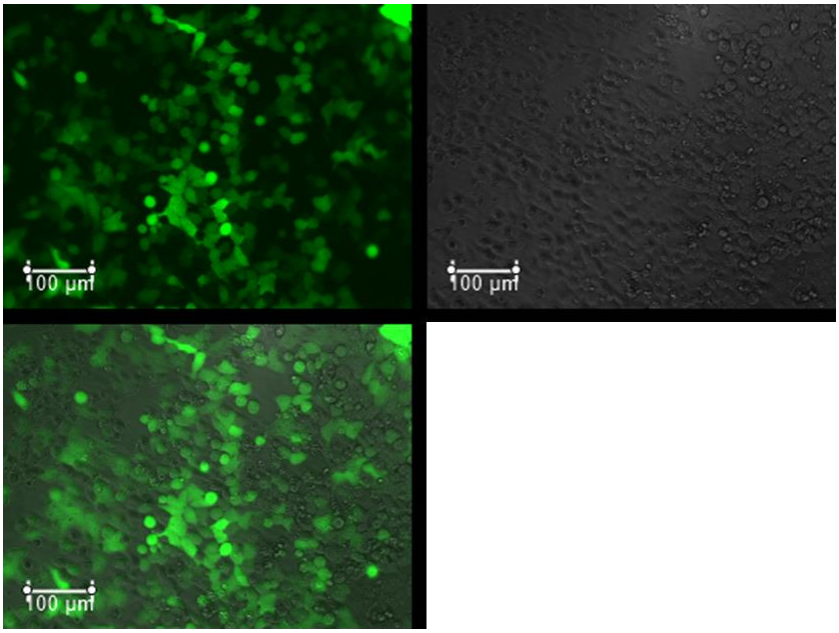


Figure 2.3. DU145 cells transduced with lentivirus expressing GFP for transduction control (representative image).

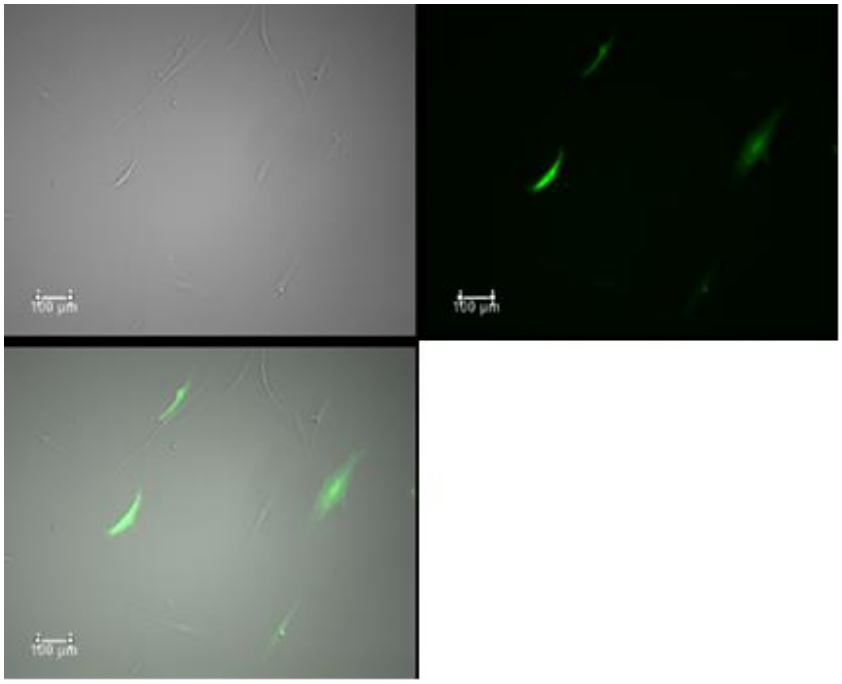


Figure 2.4. hMADs cells transduced with lentivirus expressing GFP for transduction control (representative image).

2.2.6 miRs isolation and RT-qPCR

For running qPCR for miRs, reverse transcription of miR into cDNA was completed using miRCURY LNA RT Kit (Qiagen, 339340) and the Veriti 96-well thermal cycler following the manufacturer's instructions. For RT-qPCR, miRCURY LNA SYBR Green PCR Kit (Qiagen, 339345) was used following the manufacturer's instructions and qPCR samples were run on the C1000 Touch™ Thermal Cycler in 384-well plates. Triplicate technical repeats were conducted for each assay and normalised to a *U6* housekeeping gene in human. Primers for miRCURY LNA SYBR green were ordered from Qiagen. Ct values were presented as normalisation to the housekeeping control ($2^{-\Delta CT}$) or as fold change in expression compared to day 0 and after normalisation to the housekeeping control ($2^{-\Delta\Delta CT}$).

2.2.7 Statistics

Data are presented as mean ± SEM. Technical replicates were used for analysis (N = 3-5). Student's t-test or, one-way or two-way ANOVA was used, with various post-hoc tests as indicated in the figure legends. Categorical data are analysed by chi-squared test with Yates' correction, where applicable. GraphPad Prism 9 (GraphPad software) was used to present and analyse quantitative data.

2.3 RESULTS

2.3.1 The *CASC20* transcript is 3' poly-adenylated

To determine whether *CASC20* has a poly-A tail, I purified total RNA from DU145 (prostate cancer) cells that express abundant *CASC20* (516) and synthesised cDNA using either random oligonucleotide or oligo-dT priming followed by qPCR. The housekeeping genes *β-ACTIN* and *GAPDH* that both harbour a 3' poly-A tail were used as positive controls. I found no difference in abundance of *CASC20* cDNA between the oligod/T and random primed templates (Figure 2.5), confirming the presence of a poly-A tail and the ability for cytoplasmic export.

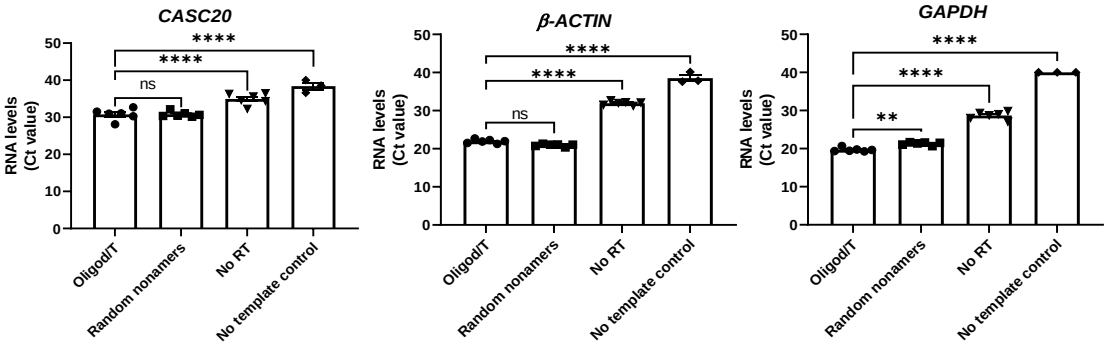


Figure 2.5. *CASC20* has a 3' poly-A tail. DU145 were collected for RNA extraction and qPCR was run for *CASC20*, *B-ACTIN*, and *GAPDH*. I used 2 methods for cDNA synthesis (Oligod/T and Random nonamers) and 2 controls for qPCR (no reverse transcriptase and no template control). Data were analysed using Ct and one-way ANOVA with Dunnett's multiple comparisons test; data is plotted as mean ±SEM; **P<0.01, ****P<0.0001; N = 6, technical replicates.

2.3.2 *CASC20* is predicted to interact with miRs involved in osteo- and chondrogenic differentiation

I used miRanda (517) to test the potential of *CASC20* to interact with miRs, enabling it to act as a ceRNA. After filtering over 2,000 miRs to include only those predicted to interact with *CASC20*, I obtained 327 putative miRs (Table 2.1). Next, I removed miRs absent from the Encyclopedia of RNA Interactomes (ENCORI) database of experimentally validated human miRs (search date May 15, 2020) (521). This left a list of 64 confirmed human miRs with predicted *CASC20* interaction (Table 2.2). I then performed an electronic search of PubMed to determine whether a published function in osteo- or chondrogenesis was enriched in these *CASC20*-interacting miRs versus presence of the same function across the non-*CASC20* interacting miRs within the ENCORI database. This search, performed using RISMed in R, identified 19 of the 64 *CASC20*-interacting miRs versus 88 of 554 non-*CASC20*-interacting miRs with such a role (Table 2.3, chi-squared p=0.0096), confirming an enrichment of osteo- and chondrogenic function associated with *CASC20* interaction from published data.

Table 2.1. Predicted Interactions of *CASC20* with miRs using miRanda. The table presents the results of the analysis conducted to predict the interactions between *CASC20* and miRs using the miRanda algorithm. Score = grading assigned to each interaction; Energy-Kcal/Mol = energy values; Query-Aln(start) and Query-Aln(end) = alignment positions within the miR Subject-Al(start) and Subject-Al(end) = the alignment positions within *CASC20;* Al-Len= alignment length between the miR and *CASC20*; Subject-identity = extent to which the *CASC20* region aligns with the miR sequence; Query-identity = extent to which the miR region aligns with *CASC20*.

miR	Score	Energy-Kcal/Mol	Query-Aln (start)	Query-Aln (end)	Subject-Al (start)	Subject-Al (end)	Al-Len	Subject-Identity	Query-Identity
hsa-miR-1303	175	-27.12	2	21	1205	1227	20	80.00%	90.00%
hsa-miR-143-5p	172	-31.61	2	21	383	404	19	73.68%	89.47%
hsa-miR-6089	172	-45.85	2	23	346	368	21	76.19%	85.71%
hsa-miR-3613-3p	168	-15.82	3	21	869	892	18	88.89%	94.44%
hsa-miR-619-5p	168	-33.44	2	21	1120	1141	19	94.74%	94.74%
hsa-miR-767-5p	168	-22.9	2	22	102	127	23	73.91%	78.26%
hsa-miR-519d-5p	167	-19.56	2	20	936	960	18	72.22%	88.89%
hsa-miR-6134	167	-28.71	2	17	1240	1259	16	87.50%	93.75%
hsa-miR-103b	165	-17.55	2	22	1136	1158	20	70.00%	80.00%
hsa-miR-5095	165	-25.51	2	18	1114	1134	16	81.25%	87.50%
hsa-miR-6771-3p	165	-29.17	2	18	288	308	16	81.25%	87.50%
hsa-miR-6776-5p	165	-26.25	2	14	390	408	12	100.00%	100.00%
hsa-miR-873-5p	165	-25.1	2	20	1171	1190	18	77.78%	88.89%
hsa-miR-32-3p	164	-18.51	2	21	670	689	19	78.95%	89.47%
hsa-miR-363-5p	164	-23.6	2	17	336	357	15	86.67%	86.67%
hsa-miR-4774-5p	164	-21.8	2	21	453	474	19	73.68%	94.74%
hsa-miR-515-5p	164	-20.7	3	21	938	961	18	88.89%	88.89%
hsa-miR-224-5p	163	-17.96	2	16	828	848	14	85.71%	92.86%
hsa-miR-3915	163	-22.93	2	21	731	753	20	80.00%	90.00%
hsa-miR-4717-3p	163	-24.18	2	16	715	735	14	78.57%	100.00%
hsa-miR-8070	163	-17.48	2	21	1118	1140	20	75.00%	80.00%
hsa-miR-29b-2-5p	162	-25.11	2	20	719	741	19	78.95%	94.74%
hsa-miR-4793-5p	162	-21.5	2	19	304	327	17	76.47%	82.35%
hsa-miR-762	162	-29.12	2	19	390	411	17	76.47%	82.35%
hsa-miR-205-3p	161	-15.13	2	18	93	113	16	75.00%	87.50%
hsa-miR-4728-3p	161	-21.51	2	22	71	95	20	70.00%	75.00%
hsa-miR-487a-5p	161	-26.84	2	18	141	162	16	75.00%	87.50%
hsa-miR-3682-3p	160	-20.39	2	20	418	436	18	72.22%	88.89%
hsa-miR-4659a-3p	160	-16.83	2	21	232	251	19	78.95%	84.21%
hsa-miR-487b-5p	160	-23.29	2	18	140	162	17	82.35%	82.35%
hsa-miR-6856-5p	160	-25.07	2	22	86	105	20	75.00%	80.00%
hsa-miR-1273h-5p	159	-22.91	2	18	352	371	16	81.25%	87.50%
hsa-miR-150-5p	159	-22.84	2	21	27	49	20	70.00%	80.00%

hsa-miR-3619-5p	159	-28.36	2	20	88	109	18	77.78%	88.89%
hsa-miR-6769b-5p	159	-29.75	2	22	332	360	26	69.23%	73.08%
hsa-miR-6858-5p	159	-29.51	2	16	155	176	14	85.71%	85.71%
hsa-miR-3679-3p	158	-23.33	2	20	35	58	20	75.00%	75.00%
hsa-miR-6883-5p	158	-21.13	2	17	149	169	15	86.67%	86.67%
hsa-miR-450a-1-3p	157	-15.32	2	21	1004	1028	22	72.73%	81.82%
hsa-miR-6511b-5p	157	-29.68	2	23	14	38	22	63.64%	77.27%
hsa-miR-7160-5p	157	-21.39	2	19	1240	1262	19	73.68%	78.95%
hsa-miR-1587	156	-23.27	2	19	394	412	17	76.47%	82.35%
hsa-miR-3620-5p	156	-28.53	2	19	392	412	17	76.47%	82.35%
hsa-miR-4659b-3p	156	-17.41	2	21	232	251	19	73.68%	84.21%
hsa-miR-487a-3p	156	-17.29	2	21	952	973	19	63.16%	78.95%
hsa-miR-519e-5p	156	-20.12	3	21	940	961	18	77.78%	88.89%
hsa-miR-6772-3p	156	-22.42	2	21	635	657	19	68.42%	73.68%
hsa-miR-6871-5p	156	-31.52	3	21	351	372	18	77.78%	88.89%
hsa-miR-7112-3p	156	-23.8	2	22	259	283	22	68.18%	72.73%
hsa-miR-922	156	-18.98	2	17	87	109	15	80.00%	80.00%
hsa-miR-1285-5p	155	-28.64	3	20	1195	1215	17	82.35%	88.24%
hsa-miR-29b-3p	155	-16.99	2	22	102	128	24	66.67%	75.00%
hsa-miR-30b-3p	155	-18.05	2	16	350	371	14	78.57%	85.71%
hsa-miR-3190-5p	155	-22.49	2	18	1177	1195	16	81.25%	81.25%
hsa-miR-4496	155	-23.92	2	21	510	533	21	66.67%	90.48%
hsa-miR-4498	155	-27.48	2	21	393	411	19	78.95%	78.95%
hsa-miR-5586-5p	155	-27.04	2	21	199	221	20	75.00%	85.00%
hsa-miR-6514-3p	155	-17.54	2	16	1268	1288	14	78.57%	85.71%
hsa-miR-3679-5p	154	-18.56	2	19	1301	1323	17	70.59%	94.12%
hsa-miR-4286	154	-22.79	2	15	38	54	13	69.23%	100.00%
hsa-miR-4437	154	-21.75	2	16	1282	1303	15	80.00%	86.67%
hsa-miR-4473	154	-15.98	2	21	132	152	19	68.42%	78.95%
hsa-miR-6134	154	-20.62	2	15	344	362	13	84.62%	84.62%
hsa-miR-6502-3p	154	-16.77	2	20	310	331	19	63.16%	84.21%
hsa-miR-6731-3p	154	-23.14	3	19	41	61	16	87.50%	87.50%
hsa-miR-7162-3p	154	-20.35	2	18	1245	1261	16	81.25%	81.25%
hsa-miR-1304-3p	153	-18.35	2	18	1192	1213	16	75.00%	75.00%
hsa-miR-1304-3p	153	-19.81	2	20	1	20	18	72.22%	77.78%
hsa-miR-4452	153	-28.75	3	22	1166	1188	19	73.68%	100.00%
hsa-miR-4492	153	-24.02	2	14	395	411	12	83.33%	91.67%
hsa-miR-4643	153	-20.34	3	19	124	146	17	88.24%	88.24%
hsa-miR-4692	153	-17.27	2	20	13	36	20	70.00%	80.00%
hsa-miR-5589-5p	153	-30.87	2	19	385	409	21	61.90%	80.95%
hsa-miR-6726-5p	153	-30.79	2	20	1247	1266	18	77.78%	88.89%
hsa-miR-6756-5p	153	-28.55	2	22	335	357	20	70.00%	80.00%
hsa-miR-7158-5p	153	-20.13	2	23	1277	1301	22	63.64%	72.73%
hsa-miR-892a	153	-18.92	2	18	1190	1210	16	68.75%	81.25%

hsa-miR-3165	152	-15.4	2	18	334	356	17	76.47%	76.47%
hsa-miR-3619-3p	152	-31.32	2	21	310	331	19	63.16%	89.47%
hsa-miR-381-5p	152	-16.7	2	18	326	348	17	76.47%	76.47%
hsa-miR-3929	152	-20.75	2	21	345	367	19	73.68%	78.95%
hsa-miR-4435	152	-20.05	2	14	1172	1194	13	84.62%	92.31%
hsa-miR-4477b	152	-15.62	3	21	553	574	18	72.22%	88.89%
hsa-miR-4689	152	-25.21	2	18	154	176	17	76.47%	76.47%
hsa-miR-5011-3p	152	-15.93	2	17	724	745	15	66.67%	86.67%
hsa-miR-545-5p	152	-17.67	2	21	1433	1454	19	73.68%	78.95%
hsa-miR-629-3p	152	-24.33	2	21	633	654	19	73.68%	78.95%
hsa-miR-7847-3p	152	-20.45	2	13	1309	1329	11	90.91%	90.91%
hsa-miR-9-3p	152	-16.41	3	21	511	532	18	83.33%	94.44%
hsa-miR-1178-3p	151	-24.44	2	18	3	22	16	81.25%	93.75%
hsa-miR-1258	151	-15.22	2	18	320	343	19	68.42%	84.21%
hsa-miR-1261	151	-15	2	16	739	757	14	78.57%	78.57%
hsa-miR-128-3p	151	-16.83	2	20	463	483	18	72.22%	83.33%
hsa-miR-3922-3p	151	-21.91	2	21	1159	1181	20	60.00%	80.00%
hsa-miR-4651	151	-36.01	2	18	344	362	16	81.25%	93.75%
hsa-miR-520f-5p	151	-16.85	2	20	939	960	18	66.67%	88.89%
hsa-miR-6069	151	-26.14	2	20	1173	1193	18	72.22%	83.33%
hsa-miR-6748-3p	151	-22.68	3	20	279	299	17	82.35%	82.35%
hsa-miR-6758-5p	151	-22.33	2	21	85	105	20	75.00%	75.00%
hsa-miR-6779-5p	151	-24.22	2	17	348	371	18	72.22%	77.78%
hsa-miR-6825-5p	151	-24.9	2	14	150	170	12	91.67%	91.67%
hsa-miR-7843-3p	151	-21.8	2	21	288	310	20	65.00%	90.00%
hsa-miR-8058	151	-22.13	2	20	70	91	18	72.22%	83.33%
hsa-miR-93-3p	151	-21.71	3	21	73	95	19	78.95%	84.21%
hsa-miR-1322	150	-19.18	2	11	80	98	9	100.00%	100.00%
hsa-miR-20b-3p	150	-18.46	2	19	1264	1285	17	70.59%	88.24%
hsa-miR-20b-3p	150	-18.53	2	19	367	388	17	70.59%	88.24%
hsa-miR-4639-5p	150	-18.61	2	20	891	913	19	73.68%	84.21%
hsa-miR-4797-5p	150	-19.51	2	16	981	1002	15	73.33%	86.67%
hsa-miR-490-3p	150	-15.3	2	20	1290	1308	18	77.78%	77.78%
hsa-miR-5001-5p	150	-20.66	2	19	388	411	17	70.59%	70.59%
hsa-miR-5093	150	-19.83	2	22	725	750	23	69.57%	73.91%
hsa-miR-519e-3p	150	-16.93	2	20	378	400	19	63.16%	78.95%
hsa-miR-6083	150	-16.06	3	19	659	678	16	75.00%	93.75%
hsa-miR-6754-5p	150	-21.46	2	20	84	106	19	68.42%	89.47%
hsa-miR-6865-5p	150	-20.36	2	11	1236	1258	9	100.00%	100.00%
hsa-miR-6881-3p	150	-17.44	2	21	1144	1167	21	66.67%	76.19%
hsa-miR-920	150	-24.68	2	19	1247	1266	17	76.47%	82.35%
hsa-miR-485-3p	149	-16.91	2	18	951	972	16	68.75%	75.00%
hsa-miR-515-3p	149	-15.4	2	19	378	400	18	61.11%	83.33%
hsa-miR-542-3p	149	-16.08	2	19	983	1005	18	72.22%	88.89%

hsa-miR-6767-3p	149	-22.75	2	19	4	26	18	72.22%	88.89%
hsa-miR-6780b-5p	149	-21.48	2	22	1304	1326	20	60.00%	85.00%
hsa-miR-6863	149	-19.68	2	18	1311	1333	16	75.00%	87.50%
hsa-miR-1285-3p	148	-30.09	2	21	302	323	19	78.95%	84.21%
hsa-miR-181c-5p	148	-15.6	2	21	810	828	19	63.16%	78.95%
hsa-miR-302b-3p	148	-16.7	2	22	637	660	21	57.14%	90.48%
hsa-miR-3162-5p	148	-24.1	2	22	1245	1268	21	61.90%	85.71%
hsa-miR-33b-3p	148	-25.29	2	20	379	400	19	68.42%	78.95%
hsa-miR-4311	148	-16.42	2	17	518	535	15	66.67%	100.00%
hsa-miR-500b-3p	148	-16.91	2	19	204	222	17	70.59%	76.47%
hsa-miR-532-3p	148	-30.73	2	21	34	55	19	68.42%	78.95%
hsa-miR-633	148	-18.4	2	21	44	66	19	68.42%	78.95%
hsa-miR-6512-5p	148	-26.64	2	21	776	797	19	68.42%	94.74%
hsa-miR-6751-5p	148	-25.93	2	21	1450	1472	19	63.16%	100.00%
hsa-miR-6760-3p	148	-17.96	2	13	1	18	11	81.82%	90.91%
hsa-miR-6780a-5p	148	-20.23	2	22	348	371	21	66.67%	66.67%
hsa-miR-6811-5p	148	-17.72	2	17	17	38	15	66.67%	80.00%
hsa-miR-769-5p	148	-19.04	2	17	247	268	15	66.67%	80.00%
hsa-miR-218-2-3p	147	-15.94	2	13	1274	1296	12	83.33%	91.67%
hsa-miR-222-3p	147	-19.6	2	16	1100	1120	14	78.57%	92.86%
hsa-miR-3153	147	-18.28	2	22	795	816	20	70.00%	80.00%
hsa-miR-3190-5p	147	-18.34	2	16	304	323	14	71.43%	78.57%
hsa-miR-323b-5p	147	-24.6	2	16	139	161	14	78.57%	92.86%
hsa-miR-378g	147	-18.64	2	17	1284	1305	17	64.71%	82.35%
hsa-miR-4268	147	-19.14	2	13	635	656	12	83.33%	91.67%
hsa-miR-4514	147	-16.06	2	16	19	36	14	64.29%	85.71%
hsa-miR-4802-3p	147	-15.17	2	12	712	734	10	80.00%	100.00%
hsa-miR-487b-3p	147	-16.73	2	20	952	973	18	66.67%	83.33%
hsa-miR-5695	147	-19.86	2	21	936	960	22	68.18%	77.27%
hsa-miR-6511a-5p	147	-22.08	2	21	12	35	20	60.00%	75.00%
hsa-miR-6741-5p	147	-21.07	2	14	387	407	12	83.33%	91.67%
hsa-miR-6749-3p	147	-28.95	2	20	635	654	19	68.42%	78.95%
hsa-miR-6811-5p	147	-16.2	2	21	1112	1134	20	70.00%	80.00%
hsa-miR-6843-3p	147	-16.75	2	20	721	741	18	61.11%	72.22%
hsa-miR-6867-5p	147	-17.12	2	20	470	492	18	66.67%	83.33%
hsa-miR-876-3p	147	-15.15	2	17	142	164	16	75.00%	75.00%
hsa-miR-1294	146	-20.85	3	21	1241	1261	18	77.78%	83.33%
hsa-miR-148b-5p	146	-17.17	2	20	1123	1145	19	73.68%	78.95%
hsa-miR-181a-5p	146	-15.21	2	15	806	828	13	76.92%	76.92%
hsa-miR-3156-3p	146	-28.12	2	19	34	54	17	64.71%	88.24%
hsa-miR-34a-5p	146	-23.84	3	20	381	403	18	72.22%	88.89%
hsa-miR-3653-5p	146	-22.4	3	19	257	276	16	75.00%	87.50%
hsa-miR-3925-5p	146	-15.41	2	20	511	533	19	63.16%	89.47%
hsa-miR-4468	146	-19.25	3	16	91	109	14	92.86%	92.86%

hsa-miR-4698	146	-19.12	3	19	1313	1335	16	87.50%	93.75%
hsa-miR-4742-5p	146	-15.44	2	21	216	241	22	63.64%	86.36%
hsa-miR-520a-5p	146	-25.47	2	19	939	959	17	76.47%	94.12%
hsa-miR-532-3p	146	-27.24	2	21	626	652	24	62.50%	70.83%
hsa-miR-5588-5p	146	-17.9	2	19	176	196	17	70.59%	82.35%
hsa-miR-5787	146	-25.17	2	17	392	410	15	73.33%	80.00%
hsa-miR-583	146	-18.82	2	15	1487	1507	13	76.92%	100.00%
hsa-miR-6748-5p	146	-21.16	2	11	338	359	9	88.89%	100.00%
hsa-miR-7151-3p	146	-16.08	2	15	1114	1134	13	76.92%	76.92%
hsa-miR-935	146	-21.45	3	20	786	809	18	72.22%	88.89%
hsa-miR-1294	145	-19.13	2	20	328	348	18	72.22%	83.33%
hsa-miR-154-5p	145	-16.01	3	18	141	162	15	80.00%	86.67%
hsa-miR-3156-3p	145	-19.38	2	10	631	651	8	100.00%	100.00%
hsa-miR-3176	145	-17.46	2	18	1163	1181	16	62.50%	75.00%
hsa-miR-3663-3p	145	-18.58	2	10	107	129	8	100.00%	100.00%
hsa-miR-3680-3p	145	-22.15	2	20	261	286	21	71.43%	80.95%
hsa-miR-3689a-3p	145	-21.61	2	21	350	371	20	65.00%	75.00%
hsa-miR-3689b-3p	145	-21.66	2	21	350	371	20	65.00%	75.00%
hsa-miR-3689c	145	-21.66	2	21	350	371	20	65.00%	75.00%
hsa-miR-374b-5p	145	-15.03	2	18	953	974	16	87.50%	87.50%
hsa-miR-3944-3p	145	-34.31	2	22	1178	1200	20	70.00%	85.00%
hsa-miR-4461	145	-21.99	3	22	313	335	19	63.16%	84.21%
hsa-miR-4483	145	-16.99	2	16	150	165	14	71.43%	85.71%
hsa-miR-450a-2-3p	145	-15.97	2	20	1005	1028	20	70.00%	85.00%
hsa-miR-451b	145	-15.67	2	19	217	239	18	72.22%	83.33%
hsa-miR-4539	145	-19.9	2	14	71	92	12	75.00%	83.33%
hsa-miR-4666a-3p	145	-15.41	2	19	461	484	20	70.00%	80.00%
hsa-miR-4739	145	-21.8	2	19	80	105	18	72.22%	83.33%
hsa-miR-4781-3p	145	-16.87	2	21	101	122	20	65.00%	75.00%
hsa-miR-4786-5p	145	-24.66	2	16	312	332	14	78.57%	78.57%
hsa-miR-4787-3p	145	-34.76	2	23	632	658	24	70.83%	79.17%
hsa-miR-5699-3p	145	-20.56	2	18	282	303	16	68.75%	87.50%
hsa-miR-612	145	-32.69	2	23	1170	1195	22	72.73%	81.82%
hsa-miR-656-3p	145	-15.01	2	20	956	975	18	72.22%	83.33%
hsa-miR-6745	145	-24.47	2	20	335	357	20	70.00%	70.00%
hsa-miR-6785-5p	145	-19.2	2	17	150	169	15	73.33%	80.00%
hsa-miR-6849-5p	145	-30.21	4	22	737	759	18	72.22%	100.00%
hsa-miR-6880-5p	145	-23.4	3	14	1309	1330	11	100.00%	100.00%
hsa-miR-6886-3p	145	-20.59	2	19	61	83	19	63.16%	73.68%
hsa-miR-7851-3p	145	-18.82	2	20	1287	1307	18	66.67%	72.22%
hsa-miR-8069	145	-18.57	2	22	1427	1449	20	60.00%	80.00%
hsa-miR-1236-3p	144	-24.94	2	21	634	655	19	63.16%	78.95%
hsa-miR-1273f	144	-18.44	2	17	1206	1224	15	73.33%	86.67%
hsa-miR-26a-5p	144	-20.02	2	13	1154	1175	11	90.91%	100.00%

hsa-miR-302a-3p	144	-16.09	2	22	637	660	21	52.38%	90.48%
hsa-miR-4433b-3p	144	-29.24	2	17	154	174	15	86.67%	93.33%
hsa-miR-4655-3p	144	-29.61	2	19	286	305	17	76.47%	82.35%
hsa-miR-4680-5p	144	-19.06	2	20	1167	1187	18	66.67%	88.89%
hsa-miR-4752	144	-16	3	21	335	356	18	72.22%	77.78%
hsa-miR-4766-5p	144	-18.11	2	19	497	519	19	68.42%	89.47%
hsa-miR-509-3p	144	-20.91	2	13	1325	1346	11	90.91%	100.00%
hsa-miR-519d-3p	144	-16.11	2	19	636	661	21	71.43%	80.95%
hsa-miR-657	144	-20.73	2	21	11	33	19	68.42%	73.68%
hsa-miR-6769a-5p	144	-23.44	2	18	338	360	18	66.67%	72.22%
hsa-miR-6826-3p	144	-20.73	2	21	36	57	19	63.16%	78.95%
hsa-miR-939-3p	144	-16.16	2	18	1285	1306	17	70.59%	70.59%
hsa-let-7b-5p	143	-23.5	2	21	1237	1259	20	75.00%	85.00%
hsa-let-7c-5p	143	-23.5	2	21	1237	1259	20	75.00%	85.00%
hsa-miR-100-3p	143	-15.26	3	21	196	218	19	73.68%	78.95%
hsa-miR-1205	143	-15.7	2	17	17	39	18	61.11%	77.78%
hsa-miR-1231	143	-18.7	2	16	609	628	14	64.29%	78.57%
hsa-miR-1233-5p	143	-27.96	3	16	351	372	13	84.62%	92.31%
hsa-miR-217	143	-16.13	2	21	1095	1120	22	68.18%	72.73%
hsa-miR-217	143	-20.58	3	20	264	286	17	76.47%	76.47%
hsa-miR-302c-3p	143	-18.58	3	22	131	152	19	68.42%	84.21%
hsa-miR-320a	143	-18.87	3	20	1244	1265	17	70.59%	82.35%
hsa-miR-320b	143	-18.87	3	20	1244	1265	17	70.59%	82.35%
hsa-miR-3929	143	-21.29	2	18	1243	1264	16	75.00%	87.50%
hsa-miR-3940-5p	143	-29.91	2	18	391	412	18	72.22%	88.89%
hsa-miR-4321	143	-15.89	2	12	665	685	10	80.00%	90.00%
hsa-miR-4459	143	-27.66	3	21	157	175	18	83.33%	83.33%
hsa-miR-4521	143	-16.43	3	21	732	754	19	73.68%	78.95%
hsa-miR-485-5p	143	-26.01	3	16	1178	1199	13	84.62%	92.31%
hsa-miR-526b-3p	143	-16.85	2	18	641	661	16	68.75%	93.75%
hsa-miR-649	143	-17.31	2	21	836	858	20	60.00%	70.00%
hsa-miR-6815-5p	143	-17.46	2	15	1238	1258	13	76.92%	84.62%
hsa-miR-6877-5p	143	-28.01	2	17	344	367	17	76.47%	82.35%
hsa-miR-6888-3p	143	-16.09	2	20	279	299	18	61.11%	83.33%
hsa-miR-8078	143	-25.43	2	22	1181	1202	20	65.00%	80.00%
hsa-let-7e-5p	142	-19.81	2	21	1204	1224	19	63.16%	84.21%
hsa-let-7i-5p	142	-16.34	2	21	1204	1224	19	63.16%	84.21%
hsa-miR-103a-2-5p	142	-15.33	3	17	229	250	14	85.71%	92.86%
hsa-miR-149-3p	142	-25.12	2	20	153	174	19	68.42%	78.95%
hsa-miR-181b-2-3p	142	-22.32	2	19	126	146	18	83.33%	88.89%
hsa-miR-320c	142	-18.91	3	19	1246	1265	16	75.00%	81.25%
hsa-miR-3913-5p	142	-18.79	2	17	314	334	15	73.33%	93.33%
hsa-miR-4284	142	-17.06	3	15	132	149	12	91.67%	91.67%
hsa-miR-4429	142	-17.57	3	19	1246	1265	16	75.00%	81.25%

hsa-miR-4498	142	-24.37	2	21	1178	1198	19	68.42%	78.95%
hsa-miR-4639-3p	142	-18.66	2	17	33	51	15	80.00%	86.67%
hsa-miR-4640-5p	142	-25.7	3	21	1173	1194	18	72.22%	83.33%
hsa-miR-4674	142	-23.92	2	11	1284	1304	9	77.78%	100.00%
hsa-miR-492	142	-19.16	2	22	1287	1309	21	71.43%	76.19%
hsa-miR-5189-5p	142	-17.64	2	23	1090	1113	21	57.14%	76.19%
hsa-miR-520a-3p	142	-17.38	2	21	636	660	22	63.64%	81.82%
hsa-miR-520d-3p	142	-18.21	2	15	639	660	13	69.23%	100.00%
hsa-miR-5699-5p	142	-18.52	2	21	941	961	19	78.95%	84.21%
hsa-miR-6807-3p	142	-16.33	3	19	264	286	16	68.75%	87.50%
hsa-miR-6823-5p	142	-18.36	2	15	1329	1349	13	76.92%	92.31%
hsa-miR-6824-5p	142	-27.24	3	15	151	172	12	91.67%	91.67%
hsa-miR-6864-3p	142	-18.41	2	15	249	269	13	76.92%	92.31%
hsa-miR-6878-3p	142	-22.31	2	20	1158	1180	20	60.00%	70.00%
hsa-miR-6884-5p	142	-21.08	2	21	1241	1265	22	68.18%	77.27%
hsa-miR-7855-5p	142	-22.6	2	15	292	313	13	76.92%	92.31%
hsa-miR-936	142	-19.16	2	16	87	109	15	80.00%	86.67%
hsa-miR-1288-5p	141	-16.43	3	22	1478	1500	19	68.42%	89.47%
hsa-miR-181b-5p	141	-15.38	2	22	806	828	20	55.00%	65.00%
hsa-miR-205-3p	141	-17.86	3	20	1354	1376	19	73.68%	84.21%
hsa-miR-26b-5p	141	-17.19	2	19	1153	1175	19	68.42%	78.95%
hsa-miR-302b-3p	141	-16.31	3	22	130	152	19	63.16%	78.95%
hsa-miR-3182	141	-19.61	2	16	27	42	14	78.57%	92.86%
hsa-miR-3668	141	-16.66	2	19	1398	1420	19	73.68%	89.47%
hsa-miR-3919	141	-20.7	2	20	1294	1313	18	72.22%	77.78%
hsa-miR-4505	141	-22.16	2	12	394	410	10	90.00%	90.00%
hsa-miR-4715-3p	141	-15.97	2	22	776	798	20	65.00%	70.00%
hsa-miR-4728-5p	141	-26.92	2	14	152	174	12	83.33%	91.67%
hsa-miR-494-5p	141	-15.46	2	20	140	161	18	66.67%	83.33%
hsa-miR-614	141	-18.62	3	19	369	393	18	72.22%	83.33%
hsa-miR-6502-3p	141	-16.52	2	20	438	457	18	77.78%	88.89%
hsa-miR-6806-5p	141	-16.28	2	16	884	906	14	78.57%	92.86%
hsa-miR-6848-3p	141	-19.32	2	15	721	741	15	73.33%	73.33%
hsa-miR-6894-5p	141	-23.5	2	21	82	103	19	63.16%	84.21%
hsa-miR-8054	141	-19.75	2	20	402	425	21	71.43%	76.19%
hsa-miR-1293	140	-19.8	2	9	144	165	7	100.00%	100.00%
hsa-miR-140-5p	140	-17.85	2	21	1276	1297	19	57.89%	78.95%
hsa-miR-181d-5p	140	-15.38	2	9	806	828	7	100.00%	100.00%
hsa-miR-2467-3p	140	-17.98	2	21	86	107	19	63.16%	73.68%
hsa-miR-30c-1-3p	140	-15.72	2	20	1008	1027	18	66.67%	83.33%
hsa-miR-3192-5p	140	-24.46	3	22	353	372	19	73.68%	84.21%
hsa-miR-372-3p	140	-17.24	2	22	637	660	21	61.90%	76.19%
hsa-miR-379-3p	140	-19.38	2	17	949	970	15	80.00%	93.33%
hsa-miR-383-3p	140	-15.37	2	9	1258	1276	7	100.00%	100.00%

hsa-miR-3916	140	-16.05	3	17	1302	1327	14	78.57%	85.71%
hsa-miR-4441	140	-18.64	2	14	89	106	13	84.62%	92.31%
hsa-miR-4459	140	-25.84	3	21	1247	1268	18	72.22%	88.89%
hsa-miR-4472	140	-17.43	2	17	1312	1329	15	66.67%	86.67%
hsa-miR-4476	140	-15.71	2	21	1304	1325	19	57.89%	78.95%
hsa-miR-4507	140	-21.28	2	19	394	412	17	70.59%	82.35%
hsa-miR-4748	140	-16.65	3	17	1427	1447	14	78.57%	85.71%
hsa-miR-4763-3p	140	-27.74	2	22	83	108	22	63.64%	86.36%
hsa-miR-4772-3p	140	-19.13	2	21	264	285	19	73.68%	78.95%
hsa-miR-4804-3p	140	-19.36	2	17	73	93	15	86.67%	86.67%
hsa-miR-5580-3p	140	-17.13	3	21	189	207	18	77.78%	77.78%
hsa-miR-5698	140	-17.34	2	21	150	171	19	63.16%	73.68%
hsa-miR-584-5p	140	-17.11	3	18	1274	1297	17	76.47%	82.35%
hsa-miR-654-3p	140	-18.64	2	20	1267	1289	20	70.00%	80.00%
hsa-miR-655-3p	140	-16.12	2	21	985	1007	20	75.00%	85.00%
hsa-miR-664a-5p	140	-25.71	2	13	1172	1195	11	81.82%	100.00%
hsa-miR-6734-5p	140	-24.66	2	21	156	176	19	73.68%	78.95%
hsa-miR-6778-5p	140	-24.92	3	18	350	372	16	81.25%	81.25%
hsa-miR-6802-5p	140	-18.84	3	17	1240	1259	14	78.57%	85.71%
hsa-miR-6802-5p	140	-21.32	3	17	343	362	14	78.57%	85.71%
hsa-miR-6821-5p	140	-24.89	2	22	722	745	21	61.90%	76.19%
hsa-miR-6826-5p	140	-16.72	2	21	1458	1480	19	68.42%	84.21%
hsa-miR-6832-5p	140	-20.67	2	18	84	107	17	70.59%	82.35%
hsa-miR-6859-5p	140	-20.46	2	22	811	834	21	66.67%	85.71%
hsa-miR-6862-3p	140	-20.56	3	13	201	224	10	100.00%	100.00%
hsa-miR-6884-5p	140	-25.27	2	19	345	368	19	73.68%	78.95%
hsa-miR-7107-3p	140	-16.79	3	20	275	301	18	77.78%	83.33%
hsa-miR-765	140	-21.47	3	19	157	176	16	81.25%	81.25%
hsa-miR-7855-5p	140	-22.72	2	21	251	272	19	63.16%	73.68%
hsa-miR-937-5p	140	-17.59	2	9	1107	1126	7	100.00%	100.00%

Table 2.2. List of the 64 putative *CASC20* interacting miRs.

miR name
hsa-let-7b-5p
hsa-let-7c-5p
hsa-let-7e-5p
hsa-let-7i-5p
hsa-miR-128-3p
hsa-miR-1294
hsa-miR-140-5p
hsa-miR-150-5p
hsa-miR-154-5p
hsa-miR-181a-5p

hsa-miR-181b-5p
hsa-miR-181c-5p
hsa-miR-181d-5p
hsa-miR-217
hsa-miR-222-3p
hsa-miR-224-5p
hsa-miR-2467-3p
hsa-miR-26a-5p
hsa-miR-26b-5p
hsa-miR-29b-3p
hsa-miR-302a-3p
hsa-miR-302b-3p
hsa-miR-302c-3p
hsa-miR-320a
hsa-miR-320b
hsa-miR-320c
hsa-miR-34a-5p
hsa-miR-3619-5p
hsa-miR-3679-5p
hsa-miR-372-3p
hsa-miR-374b-5p
hsa-miR-378g
hsa-miR-379-3p
hsa-miR-3944-3p
hsa-miR-4429
hsa-miR-4640-5p
hsa-miR-4739
hsa-miR-4766-5p
hsa-miR-485-3p
hsa-miR-485-5p
hsa-miR-487a-3p
hsa-miR-487b-3p
hsa-miR-490-3p
hsa-miR-509-3p
hsa-miR-515-5p
hsa-miR-519d-3p
hsa-miR-519e-5p
hsa-miR-520a-3p
hsa-miR-520a-5p
hsa-miR-520d-3p
hsa-miR-526b-3p
hsa-miR-532-3p
hsa-miR-542-3p
hsa-miR-545-5p

hsa-miR-5586-5p
hsa-miR-584-5p
hsa-miR-654-3p
hsa-miR-655-3p
hsa-miR-656-3p
hsa-miR-6807-3p
hsa-miR-6884-5p
hsa-miR-769-5p
hsa-miR-873-5p
hsa-miR-9-3p

Table 2.3. 19 confirmed human miRs with predicted interactions with *CASC20* and published roles in osteo- or chondrogenesis using ENCORI.

miR	**Published role in osteo- or chondrogenesis (PMID:) (search date May 15, 2020)**
hsa-miR-217	30551361, 30367466, 30098434, 26054690, 25289936
hsa-miR-140-5p	30666640, 25941324
hsa-miR-150-5p	29113181, 26025627
hsa-miR-181a-5p	30286747, 29032608
hsa-miR-26a-5p	29063236, 26854724
hsa-let-7b-5p	24617339
hsa-let-7c-5p	24617339
hsa-let-7e-5p	24617339
hsa-miR-181c-5p	29032608
hsa-miR-222-3p	29102598
hsa-miR-26b-5p	29534118
hsa-miR-320a	31276478
hsa-miR-320b	25724494
hsa-miR-320c	30509082
hsa-miR-34a-5p	30564968
hsa-miR-4739	28340487
hsa-miR-487b-3p	29102319
hsa-miR-584-5p	25475098
hsa-miR-769-5p	31465725

2.3.3 Osteo and chondrogenic gene and pathway enrichment

Next, I used the ENCORI database to identify genes that are targeted by *CASC20* interacting miRs. This search identified 13463 genes that are targeted by at least 1 miR. Next, I ranked the genes in order of the number of putative *CASC20* miRs with which they interact to create a *CASC20*-miR interactome hierarchy on the assumption that the genes in the top quartile of the ranking are more likely to be regulated via *CACS20*-miR interactions than those in the bottom quartile. I then repeated the RISMed PubMed search strategy here to identify a

published role of the top vs. bottom quartile of the ranked target genes in chondro- or osteogenesis. This analysis yielded 2283 genes with a published role versus 1083 with no published role in the top quartile, and 1960 versus 1406 in the bottom quartile (chi-squared $p<0.0001$). Within the top quartile, 19 genes are predicted to be targets for ≥50 putative *CASC20* interacting miRs (Figure 2.6A), suggesting that these targets may be the most profoundly regulated by altered *CASC20* expression. Ten of these 19 genes also have published experimental data demonstrating their role in osteo or chondrogenesis (Table 2.4). Next, I used Gene Ontology (GO) to identify molecular functions and biological processes that are targeted by *CASC20*-interacting miRs (Figure 2.6B). The most frequently-targeted molecular functions included SMADs, MAPKK and other kinase activities that are known to play a central role in osteo and chondrogenesis (529). The most frequently targeted biological processes included endothelial cell proliferation and migration, cell-matrix adhesion, cell junction assembly and MAPKK activation, consistent with the role of endothelial cells and matrix assembly in HO initiation (27, 28, 530).

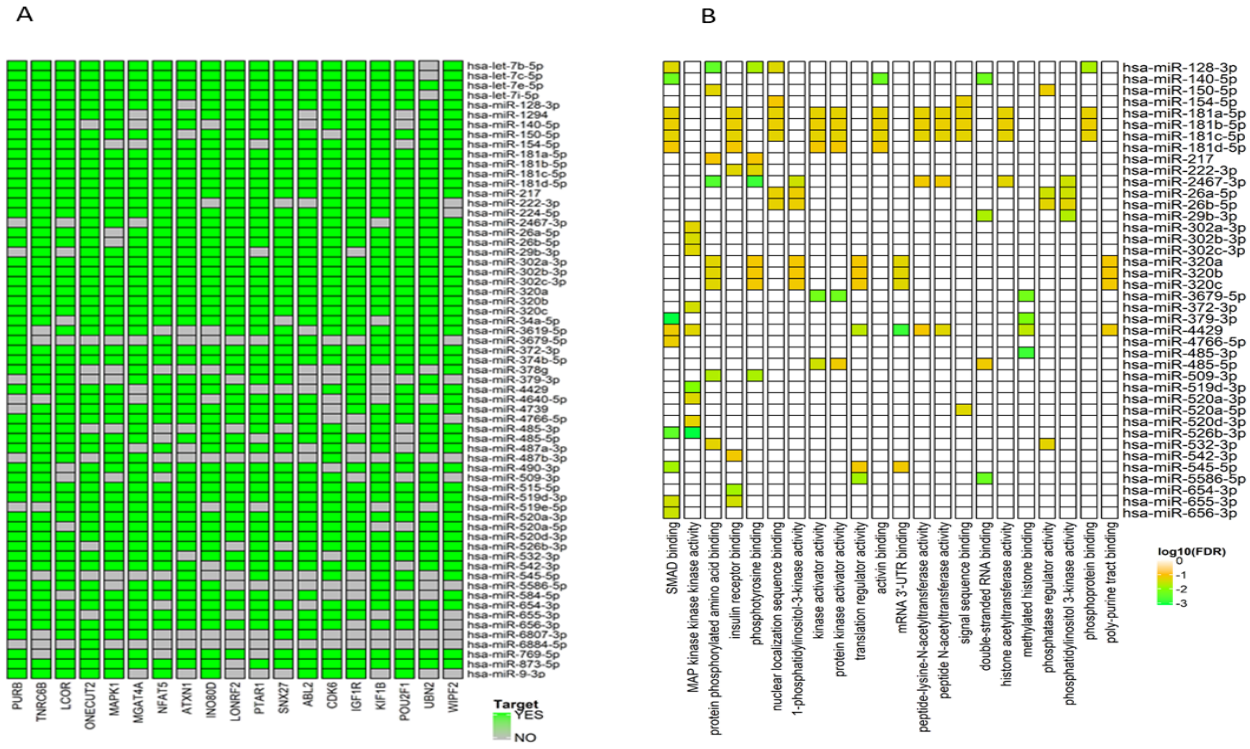


Figure 2.6. *CASC20* interacts with multiple miRs that have enriched representation in gene targets that modulate osteo and chondrogenesis, and in association with HO-relevant Gene Ontology (GO) terms. A) Heatmap shows mRNAs targeted by at least 50 *CASC20*-interacting miRs (subset of all the targeting miRs is shown). B) Molecular function (MF) GO terms. GOSeq was used to perform the MF analysis. The GO terms are ranked so that those associated with the highest number of miRs are furthest to the left of the map. Only GO terms associated with ≥ 4 miRs are shown.

Table 2.4. Putative gene targets for *CASC20*-interacting miRs and most recent PMIDS for their published role in osteo or chondrogenesis.

Target gene	Number of *CASC20*-interacting miRs	Most recent PMID number corresponding with published role in osteo or chondrogenesis (search date May 15, 2020)
LCOR	54	29663375
NFAT5	52	26418500
MAPK1	52	31013682
SNX27	51	26912788
ATXN1	51	30210606
PTAR1	51	31323019
KIF1B	50	29113313
POU2F1	50	30783479
IGF1R	50	22729283
CDK6	50	30466085

2.3.4 Expression of putative *CASC20*-interacting miRs in osteo- and chondrogenesis experimental datasets

Finally, to identify whether *CASC20*-interacting miRs are differentially expressed by primary human MSCs during osteo and chondrogenesis over the time-period when *CASC20* is induced in response to a differentiating stimulus, I explored published datasets. To explore expression during osteoblast differentiation I used miR-Seq dataset GSE107279 (524). 42 of the 64 ENCORI *CASC20*-interacting miRs were found in this dataset at raw expression values >10. Of these 42, 24 became downregulated between day 3 and 13 (Figure 2.7A), consistent with the upregulation of *CASC20* expression after day 3 following an osteogenic stimulus and a putative role of *CASC20* as a ceRNA to these miR targets. I found that 169 out of 352 non-interacting miRs were downregulated after day 3. This finding suggested that there is no greater likelihood of downregulation among *CASC20*-interacting miRs compared to non-interacting miRs (chi-squared = 1.252, df = 1, p-value = 0.263).

Having established that *CASC20* is also expressed in differentiating chondrocytes within mRNA dataset GSE109503 (531), I used the previously published chondrocyte differentiation miR microarray dataset of Barter and colleagues (528) to explore miR expression during chondrocyte differentiation. 21 of the 64 ENCORI *CASC20*-interacting miRs were found in this dataset. Of these, 16 became downregulated between day 3 and 14 (chi-squared test) (Figure 2.7B). I found that 17 out of 187 non-interacting miRs were downregulated after day 3. This finding suggested that *CASC20*-interacting miRs are more likely to be downregulated compared to non-interacting miRs (chi-squared = 20.884, df = 1, p-value <0.00001). 10 miRs, hsa-miR-485-5p, hsa-miR-485-3p, hsa-miR-3944-3p, hsa-miR-520a-5p, hsa-miR-3619-5p, hsa-

miR-515-5p, hsa-miR-542-3p, hsa-miR-520d-3p, hsa-miR-769-5p, and hsa-miR-654-3p demonstrated down-regulation across both datasets suggesting a possible common *CASC20*-mediated regulatory mechanism during osteo and chondrogenesis.

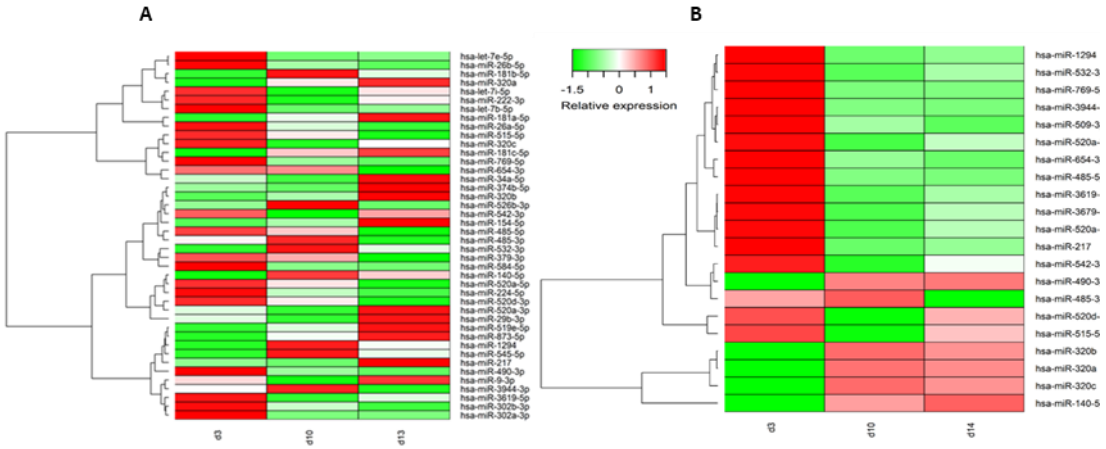


Figure 2.7. *CASC20* interacting miRs are also downregulated during osteo and chondrogenic differentiation**.** A) Characterisation of miRs targeted by *CASC20* during osteoblast differentiation in dataset GSE107279, analysed using Limma and EdgeR with reads normalised the using log(CPM+1) method. A heatmap for the normalised reads of *CASC20*-interacting miRs expressed in the dataset for day 0, 7 and 13 was constructed using Heatmap.2. B) Characterisation of miRs targeted by *CASC20* during chondrocyte differentiation. The normalised chondrocyte differentiation microarray data was obtained from (528). A heatmap for the normalised data of *CASC20*-interacting miRs expressed in the dataset for day 3, 10 and 14 was constructed using Heatmap.2.

2.3.5 Testing *CASC20* overexpression on candidate miRs

Finally, I tested the effect of *CASC20* on 6 out of the 10 miRs identified by the bioinformatics analysis, using *in vitro* approaches. The miRs hsa-miR-3944-3p and hsa-miR-3619-5p were excluded from the analysis as they were lowly expressed in the osteoblast dataset at baseline (day 3). The miRs hsa-miR-520a-5p and hsa-miR-542-3p were excluded from the *in vitro* analysis because they showed the least downregulation in the datasets analysed. This left us with the following 6 candidate miRs: hsa-miR-485-3p, hsa-miR-485-5p, hsa-miR-520d-3p, hsa-miR-654-3p, hsa-miR-769-5p, and hsa-miR-515-5p. I transduced *CASC20*-overexpressing lentivirus into hMADs and confirmed that the cells overexpress *CASC20* using RT-qPCR (Figure 2.8A). *CASC20*-overexpressing hMADs were differentiated for 10 days into osteoblasts and qPCR was run for the six candidate miRs. Two miRs were not detected (hsa-miR-520d-3p and hsa-miR-515-5p), three miRs showed no statistical differential expression (hsa-miR-485-5p, 654-3p, and 769-5p), and hsa-miR-485-3p was shown to be significantly downregulated at day 10 in *CASC20*-overexpressing samples compared to wild type (FC= 0.24, p= 0.002696) as shown in Figure 2.8B.

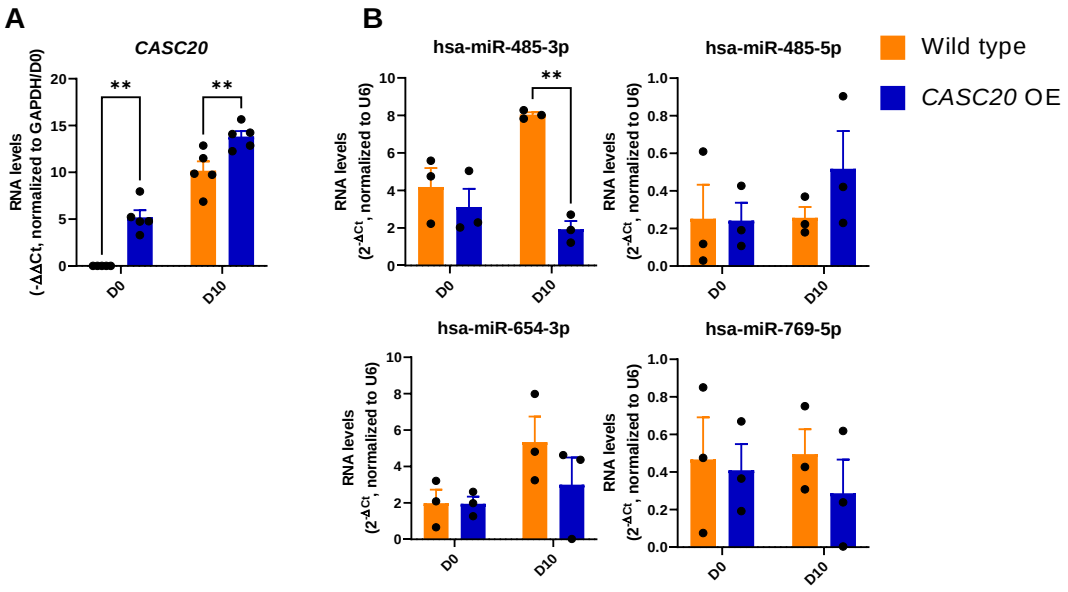


Figure 2.8. Pilot data indicating that *CASC20* may negatively regulate miR-485-3p in osteoblast differentiation. A) RT-qPCR was used to confirm *CASC20* lentiviral overexpression in hMAD at days 0 and 10. Day 0 samples were collected 24 hours after transduction. Data were analysed using - Δ Δ Ct by normalising to *GAPDH* and day 0 (N= 5, technical replicates). Analyses are multiple unpaired t test; data is plotted as mean ±SEM. B) hMAD were transduced to overexpress *CASC20* and RT-qPCR was used to measure the expression of 6 candidate miRs at day 0 and 10 of osteoblast differentiation. Data were analysed using 2^(- Δ Ct) by normalising to U6 (N=3, technical replicates). Multiple unpaired t-test; data is plotted as mean ±SEM. **P<0.01

2.4 DISCUSSION

Our finding that *CASC20* has a 3' poly-A tail indicates that this lncRNA may function as a cytoplasmic ceRNA to miRs that modulate chondro- and osteogenesis. Several other lncRNAs are already established as functional ceRNAs in bone and cartilage formation. For example, lncRNA *H19* acts as a ceRNA to miR-185-5p, a miR that modulates *IGF1* activity in osteoblast mineralisation (532). LncRNA *ADAMTS9-AS2* acts as a ceRNA to miR-942-5p, promoting the expression of *Scrg1*, a transcription factor for chondrogenesis (533). In the ENCORI database I found several putative *CASC20* interacting miRs, and strong enrichment of their target genes and molecular function pathways with a known role in osteo and chondrogenesis.

Upon examining the direction of change, I observed that several miRs displayed a downregulation pattern in osteo- and chondrogenesis. This finding provides *in-silico* support for the ceRNA function hypothesis in the context of osteo- and chondrogenesis, as the downregulation observed in the expression datasets coincided with the timeframe of *CASC20* upregulation resulting from our MSC stimulation experiments. Notably, all 10 of the identified miRs that were semi-quantitatively downregulated in both datasets are known to possess inhibitory roles in osteo- or chondrogenesis. The identification of *CASC20* as a common interacting lncRNA suggests the existence of a potential shared regulatory link across multiple signalling pathways.

By integrating bioinformatics analysis with *in vitro* experiments, I identified and prioritised potential *CASC20*-interacting miRs for further investigation. The successful overexpression of *CASC20* in hMAD cells and subsequent differentiation into osteoblasts provided a relevant cellular context to evaluate the effect of *CASC20* on miR expression. Similar lentiviral models have been widely utilised in the literature to investigate the role of genes in osteo- and chondrogenesis (534-537). The significant downregulation of miR-485-3p in *CASC20*-overexpressing samples versus wild type further supports the hypothesis that *CASC20* may modulate osteo- and chondrogenesis through its interaction as a ceRNA with this specific miR.

There are several limitations that should be acknowledged in this study. Firstly, the focus was placed on a specific subset of miRs identified through bioinformatics analysis of published datasets. It is important to recognise that the availability of miR expression data is limited, and the selection of miRs may not encompass the entire landscape of potential *CASC20*-interacting miRs. Moreover, the use of published datasets generated using different platforms or technologies introduces inherent variations in data consistency and quality (537, 538). In this bioinformatics analysis, the osteogenesis dataset was generated using an RNA-Seq platform, while the chondrogenesis dataset was generated using a microarray platform. By comparing data from disparate sources, there is a risk of introducing inconsistencies that may impact the reliability of the analysis.

Additionally, the methods used for differentiating cells into osteoblasts or chondrocytes in this study do not align precisely with the standard protocols I routinely employ. While I utilise

BMP2 and OM (β-glycerophosphate, ascorbic acid, dexamethasone) media for osteogenic differentiation, the published osteogenic dataset studied used calcitriol and OM media. These variations in differentiation methods could potentially introduce differences in cellular behaviour and gene expression profiles. Therefore, it is essential to exercise caution when interpreting the results, and wet-lab validation of *CASC20*-miR interactions using transcriptome-wide sequencing of samples generated within this PhD project are warranted to thoroughly investigate and validate the findings.

It is also important to acknowledge that the ceRNA function of *CASC20* represents only one possible mechanism of action among several. Alternative mechanisms may involve the binding of transcription factors or enhancers to drive HO genes. To gain a more comprehensive understanding of *CASC20* function, further investigations employing lentiviral overexpression and RNA-Sequencing are planned. These experiments aim to elucidate how *CASC20* influences the expression profile of miRs. By addressing this inquiry, valuable insights into the molecular mechanisms underlying the effects of *CASC20* on osteo- and chondrogenesis can be obtained. These forthcoming experiments will provide a clearer picture of the specific molecular pathways and processes through which *CASC20* exerts its function.

Chapter 3 – Initial experiments on the effect of *CASC20* overexpression on osteoblast differentiation

3.1 INTRODUCTION

In this chapter, I present experiments exploring the impact of *CASC20* overexpression on osteoblast differentiation. To investigate the effect of *CASC20* on osteodifferentiation, I transduced *CASC20* lentiviral virus into hMADs and investigated their differentiation into osteoblasts. The presence of robust calcium deposits demonstrated the successful, functional maturation of osteoblasts (539). The aim was to shed light on the effects of *CASC20* hMAD osteo-differentiation, as defined by *RUNX* and *OSX* osteogenic gene expression, and on mineralisation by Alizarin Red S staining as the biological read-out. This staining method allowed for the visualisation and quantification of calcium deposits indicative of osteoblast differentiation (539). Here I confirmed that *CASC20* is overexpressed during the differentiation of hMADs. *CASC20* positively regulated the expression of osteogenic genes and calcium deposit in the early stages of the differentiation. By day 20, minimal disparities were discernible in calcium deposition between *CASC20* OE and wild type, indicating that by this time-point the wild-type cells may have achieved a comparable level of calcium deposition as the *CASC20* OE cells. This chapter represents a pilot study in which a minimal sample size is employed to determine which specific areas warrant further investigation.

3.2 MATERIALS and METHODS

3.2.1 Tissue culture

hMADs were cultured and differentiated into osteoblasts, as previously described in Chapter 2.

3.2.2 RNA isolation and RT-qPCR

Total RNA was isolated using the miRNeasy mini Kit, as previously described in Chapter 2. Reverse transcription of 400ng of RNA into cDNA was completed using the iScript™ cDNA Synthesis Kit (Bio-Rad, 1708891BUN – Hercules, CA) and the Veriti 96-well thermal cycler according to the manufacturer's instructions (Applied Biosystems, Waltham, MA). For RT-qPCR, 2ng of cDNA was loaded per well. qPCR samples were run on the C1000 Touch™ Thermal Cycler in 384-well plates. Triplicate technical repeats were conducted for each assay and normalised to a *GAPDH* or *GUSB* housekeeping gene in human. Primers for SYBR green qPCR (Sigma-Aldrich) were designed using Primer-BLAST (NCBI: (www.ncbi.nlm.nih.gov/tools/primer-blast). All primers were screened to avoid self-complementarity with Oligo Calc: Oligonucleotide Properties Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html). PrecisionPLUS SYBR-Green master mix was used with SYBR primers. Ct values were presented as normalisation to the housekeeping control ($2^{-\Delta CT}$).

Gene	Human qPCR Primer Sequence (5' -> 3')
GAPDH	FW ATTGCCCTCAACGACCACTTT
	REV CCCTGTTGCTGTAGCCAAATTC
CASC20	FW TCATATGGATTTCAAGCTGGGT
	REV TCCCAGTCTTCTGCATCACTTC
RUNX2	FW GGTTAATCTCCGCAGGTCACT
	REV CACTGTGCTGAAGAGGCT
OSX	FW CCACCTACCCATCTGACTTTG
	REV CCACTATTTCCCACTGCCTT
GUSB	FW CTGTCACCAAGAGCCAGTTCCT
	REV GGTTGAAGTCCTTCACCAGCAG
COL1A1	FW ACCGCCCTCCTGACGCAC
	REV GCAGACGCAGATCCGGCAG
ALP	FW GCAGACGCAGATCCGGCAG
	REV CCTGGCTTTCTCGTCACTCTCA

3.2.3 Alizarin Red S staining

Cells were washed twice with PBS and fixed overnight at 4°C in 100% ethanol. Cells were then washed twice with PBS before the addition of 40mM Alizarin Red S (Sigma-Aldrich, A5533), pH 4.2. The wells were washed extensively with 95% ethanol until all unbound stain was removed, the same number of washes was used for each well. Plates were air dried overnight

and scanned on high-resolution flat-bed scanner at 1200dpi. ImageJ Software (http://rsb.info.nih.gov/ij/) was used to quantify the percentage mineralised area in each well. The area fraction positive for the stain was recorded, representing percentage mineralisation. Identical settings were used for all wells. The wild type wells were used as reference for setting the threshold values.

3.2.4 Statistics

Statistical analyses were conducted as previously described in Chapter 2.

3.3 RESULTS

3.3.1 *CASC20* enhances mineralisation

I first transduced *CASC20* lentiviral virus into human multipotent adipose-derived stem cells (hMAD) in 24-well tissue culture plate and confirmed consistent *CASC20* over-expression in the *CASC20* OE cells versus wild type across all differentiation time points (Figure 3.1).

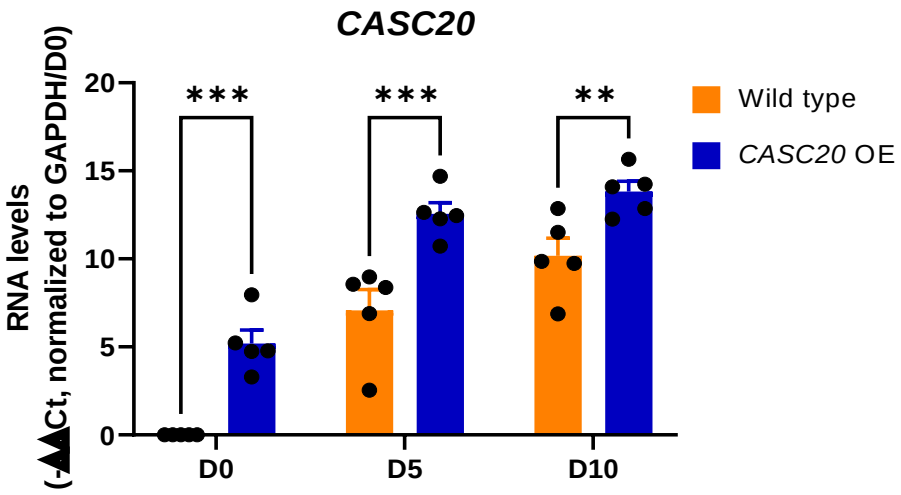


Figure 3.1. *CASC20* is overexpressed during osteodifferentiation of hMADs. RT-qPCR was used to confirm *CASC20* lentiviral overexpression in hMAD at days 0, 5, and 10. Data were analysed using two-way ANOVA with Sisak multiple comparison test; **P<0.01, ***P<0.001; N = 5, technical replicate.

Next, the wells were stained for Alizarin Red S at day 0 and 10. I found that *CASC20* overexpression significantly increased the mineralisation at day 10 compared to wild type (FC= 14, p-value= 0.002), consistent with increased osteogenesis (Figure 3.2).

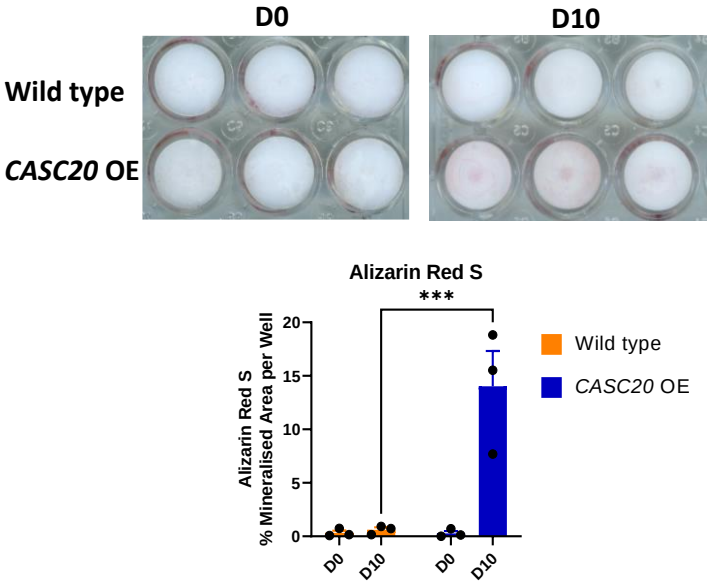


Figure 3.2. *CASC20* augments osteodifferentiation of hMADs. Alizarin Red S demonstrated increase calcium deposits in *CASC20* OE versus wild type. Data were analysed using two-way ANOVA with Sidak multiple comparison test; ***P<0.001; N = 3, technical replicate.

To test whether *CASC20* induces key osteogenic genes, *CASC20*-overxpressing cells were treated with BMP2 for 48hrs and expression of *RUNX2* and *COL1A1* was measured. I found that *CASC20* upregulated the expression of *COL1A1* (FC = 8.1, p = 0.0061) (Figure 3.3) There was a trend of upregulation in *RUNX2* expression in *CASC20* OE versus wild type, however it did not reach statistical significance (FC = 1.9, p = 0.1070).

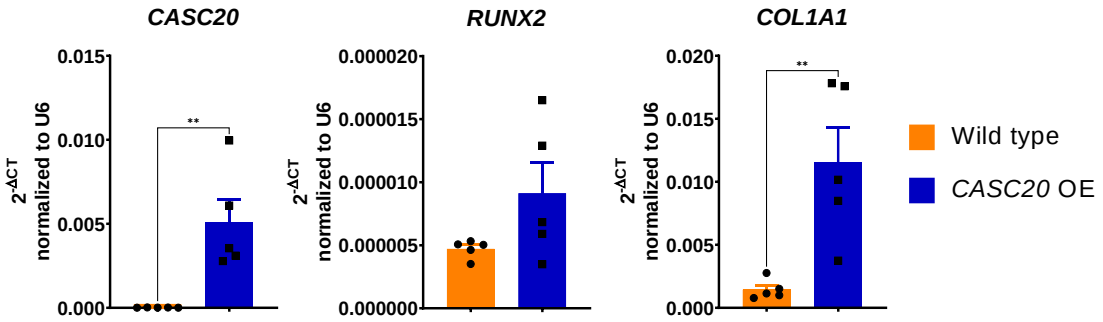


Figure 3.3. *CASC20* OE in hMADs augmented the expression of key osteogenic genes. Cells were collected 48H after BMP2 stimulation. RT-qPCR for key genes following lentiviral transduction showed increase in *RUNX2* and *COL1A1* expression in *CASC20*-overespressing cells versus wild type. Analyses are unpaired t test; **P<0.01; N = 5, technical replicate.

3.3.2 *CASC20* promotes the expression of other osteogenic genes during osteoblast differentiation

To test whether *CASC20* regulates other osteogenic genes during osteoblast differentiation, *CASC20*-overxpressing and wild type hMADs were differentiated for 5 days. After that, I measured the expression of *RUNX2*, *COL1A1 and ALP*. At day 5, I found a trend in upregulation in *RUNX2 and ALP,* but not in *COL1A1*.

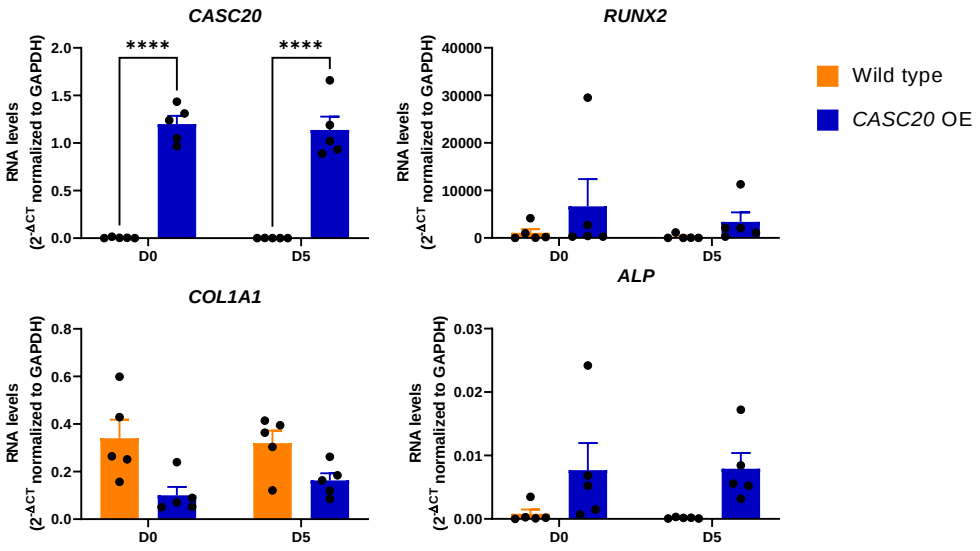


Figure 3.4. The effect of *CASC20* OE on *RUNX2*, *ALP* and *COL1A1* during the osteodifferentiation of hMADs. Analyses are two-way ANOVA with Sidak's multiple comparison test; ****P<0.0001; N = 5, technical replicate.

3.3.3 Evaluating the effect of *CASC20* in the later stages of differentiation

I then tested the effect of *CASC20* on the later stages of osteodifferentiation of hMADs by differentiating the cells for 20 days. At day 20, I found no difference in calcium deposit between *CASC20* OE and wild type (Figure 3.5). This lack of difference could be attributed to the cells already being committed to the osteoblast lineage, making it too late to observe any augmentation of the differentiation process.

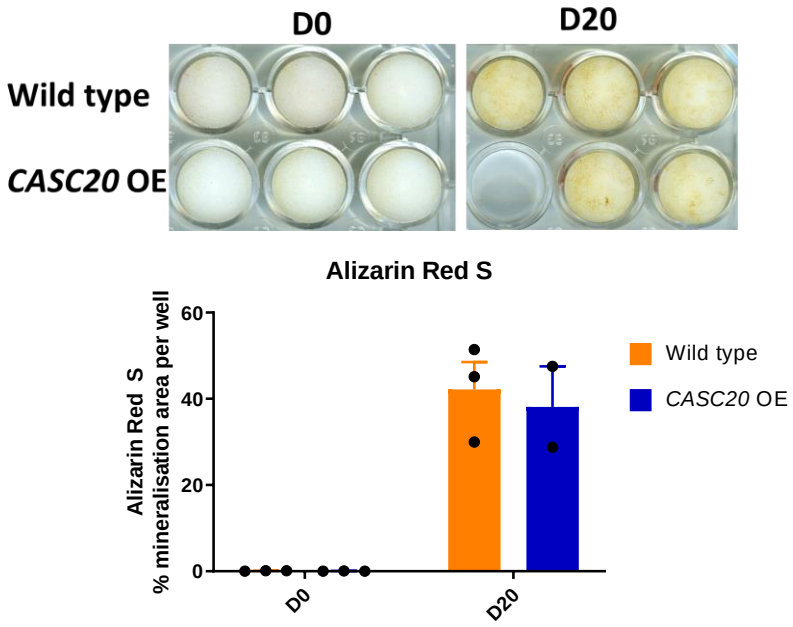


Figure 3.5. Comparison of calcium deposition in *CASC20* OE and wild type at day 20. Data were analysed using a two-way ANOVA with Sidak multiple comparison test; N=3, technical replicate.

3.4 DISCUSSION

The results demonstrated that *CASC20* overexpression enhances mineralisation compared to the wild type. This observation aligns with the notion that *CASC20* acts as a positive regulator of the mineralisation process during osteogenesis. The substantial fold change (FC= 14) and low p-value (p-value= 0.0020) underscore the robustness of this effect.

Another significant outcome of our study is the upregulation of *COL1A1* expression in *CASC20*-overexpressing cells following BMP2 treatment. The fold change (FC = 8.1) and low p-value (p = 0.0061) suggest a strong regulatory link between *CASC20* and *COL1A1* expression. COL1A1 is a key structural protein of bone and thus *COL1A1* an important marker of osteogenic differentiation (540).

While not reaching statistical significance, our findings indicate a trend in the upregulation of *RUNX2* and *ALP* expression in *CASC20*-overexpressing cells during early osteoblast differentiation. Although the mean level difference was substantial between treatment groups, the ANOVA was not significant, indicating that due to possible technical variability between biological replicates, further repeats would be required to confirm this trend. Further discussion of the role of these genes in osteogenesis is covered in the next chapter.

In summary, this pilot study gives support to the hypothesis that *CASC20* plays a role in enhancing mineralisation and promoting the expression of osteogenic genes during osteoblast differentiation. These results, in particular the biological readout of differential mineralisation in response to *CASC20* over-expression, provide a basis for follow up investigations into the molecular mechanisms by which *CASC20* influences bone formation.

Chapter 4 – Confirmation and scale up effects of *CASC20* overexpression on osteoblast differentiation

4.1 INTRODUCTION

In this chapter, the primary aim was to explore the potential regulatory role of *CASC20* in osteoblast differentiation by upscaling to allow sufficient total RNA output to allow transcriptome wide sequencing of the outputs. To achieve this, a lentiviral overexpression approach is employed in hMADs, P512MSCs (Primary human explant MSCs from patient 512), and ASC52telo hTERT immortalised human adipose-derived mesenchymal stem cells (ATCC® SCRC-4000™). By manipulating the expression levels of *CASC20*, I examined its impact on the process of osteoblast differentiation. To quantitate the biological endpoint of osteoblast differentiation Alizarin Red S staining was again implemented (536, 537, 539, 540).

My hypothesis was that *CASC20* enhances the process of osteoblast differentiation, leading to anticipate the presence of early calcium deposits at this time point. This is consistent with existing literature that indicates that the detection of Alizarin Red S staining typically occurs by the end of the second week of the differentiation process (541, 542). Therefore, examining the differences between the wild type and *CASC20* overexpression groups at day 10 allows me to assess the potential accelerating effect of *CASC20* on the formation of calcium deposits during osteoblast differentiation.

In addition to the biological endpoint assessment through Alizarin Red S staining, I employed RT-qPCR to measure the expression levels of key osteogenic markers. These markers, including *RUNX2*, *OSX*, *ALP*, and *COL1A1*, play pivotal roles in osteoblast differentiation and bone formation (543). *RUNX2*, also known as Runt-related transcription factor 2, is a master regulator of osteoblast differentiation and is responsible for initiating osteogenic lineage commitment (544). It controls the expression of various osteoblast-specific genes, including *OSX* (Osterix) (545). *OSX* is another transcription factor essential for osteoblast differentiation, functioning downstream of *RUNX2* (545). It promotes the maturation and mineralisation of osteoblasts by regulating the expression of osteoblast-specific genes such as *ALP* (Alkaline phosphatase) and *COL1A1* (Collagen type I alpha 1) (546). ALP is an enzyme involved in mineralisation, playing a crucial role in the deposition of calcium and phosphate during bone formation (547). *COL1A1*, a major component of the extracellular matrix, provides structural integrity to bone tissue and contributes to its mechanical strength (548). Together, these factors orchestrate the intricate processes of osteoblast differentiation and bone mineralisation, crucial for maintaining bone integrity and function (549).

To further assess the phenotypic characteristics of the differentiating cells, RT-qPCR was performed to analyse the expression of key chondrogenic markers, including *ACAN*, *COL2A1*, and *COMP* (550). *ACAN* encodes for aggrecan, a major proteoglycan component of the

extracellular matrix in cartilage, providing structural support and hydration (551). *COL2A1* is involved in the synthesis of type II collagen, the main collagen type found in cartilage, contributing to its structural integrity (552). *COMP* is a non-collagenous extracellular matrix protein that interacts with other cartilage matrix components, playing a crucial role in maintaining cartilage structure and function (553). The expression levels of these markers serve as molecular indicators for assessing the progress of chondrocyte maturation and cartilage formation (554). By monitoring the expression of chondrogenic markers alongside osteogenic markers, this ensures that the observed changes in gene expression are specific to the osteogenic lineage and not influenced by chondrogenic differentiation (555-557).

After completing the initial experiments, I upscaled the tissue culture to obtain a higher yield of RNA for subsequent RNA-Seq analysis. This expansion phase is necessary to generate sufficient RNA for reliable and comprehensive transcriptomic analysis that includes miR expression. The findings presented in this chapter provided evidence suggesting that *CASC20* exerts a positive regulatory effect on osteoblast differentiation in hMADs and P512MSCs. I found that ASCs were not suitable for osteoblast differentiation, as indicated by lack of Alizarin Red S staining. The observed increase in calcium deposits and the upregulation of osteogenic marker genes, as indicated by the qPCR data, supported the hypothesis that *CASC20* plays a role in promoting osteogenic differentiation.

4.2 MATERIALS and METHODS

4.2.1 Cell culture

hMADs were cultured as previously described in Chapter 2. P512MSCs were cultured in growth medium which consisted of: DMEM (Thermo Fisher Scientific, 61965-059) supplemented with 10%FBS, 1%P/S, and 25µg/mL ascorbic acid. During subculturing, when the cells reached 80-90% confluency, I passaged them and seeded into new flasks at a density of 5,000 viable cells/cm^2. Confluency was avoided to prevent differentiation, with P512MSCs cells split every 2-3 days. ASC52telo were cultured using a complete growth medium composed of Mesenchymal Stem Cell Basal Medium (ATCC, PCS-500-030 – Manassas, Virginia), Mesenchymal Stem Cell Growth Kit for Adipose and Umbilical derived MSCs Low Serum Components (ATCC, PCS-500-040), and G418 disulphate salt solution (Sigma Aldrich, G8168). The complete growth medium was prepared by adding the Mesenchymal Stem Cell Growth Kit and G418 to the basal medium as follows: 482 ml of basal medium, 10ml of MSC supplement (2% FBS, 5ng/ml rh FGF basic, 5ng/ml rh FGF acidic, 5ng/ml rh EGF), 6ml of L-Alanyl-L-Glutamine (2.4mM final concentration), and 2ml of 50mg/ml G418 solution (0.2mg/ml final concentration). For trypsin neutralisation, I used Mesenchymal Stem Cell Basal Medium (ATCC, PCS-500-030) with 10% FBS to create Trypsin Neutralisation Medium (TNM). The seeding density for cell culture was determined as follows: 5,000 cells/cm2 for T75 flask (375,000 cells). During subculturing, when the cells reached 80-90% confluency, I passaged the cells and seeded into new flasks at a density of 5,000 viable cells/cm^2.

4.2.2 Osteogenic differentiation

Cells were transduced and differentiated into osteoblasts as previously described in Chapter 2. The cells were collected at the designated time points for further analysis. For testing the differential effect of osteogenic media (OM, which includes β-glycerophosphate and dexamethasone) and BMP2, P512MSCs cells were differentiated into osteoblasts using +/-OM and +/-BMP2. Ascorbic acid was added regardless of the differential treatment, as it is necessary for collagen synthesis (558). The supplements were added 24 hours after transduction.

4.2.3 RNA isolation and RT-qPCR

Total RNA was isolated using the miRNeasy mini Kit, as previously described in Chapter 2. Reverse transcription of RNA into cDNA and RT-qPCR were conducted, as previously described in Chapter 3.

Gene	**Human qPCR Primer Sequence (5' -> 3')**
GAPDH	FW ATTGCCCTCAACGACCACTTT
	REV CCCTGTTGCTGTAGCCAAATTC
GUSB	FW CTGTCACCAAGAGCCAGTTCCT
	REV GGTTGAAGTCCTTCACCAGCAG
CASC20	FW TCATATGGATTTCAAGCTGGGT

	REV TCCCAGTCTTCTGCATCACTTC
RUNX2	FW GGTTAATCTCCGCAGGTCACT
	REV CACTGTGCTGAAGAGGCT
OSX	FW CCACCTACCCATCTGACTTTG
	REV CCACTATTTCCCACTGCCTT
COL1A1	FW ACCGCCCTCCTGACGCAC
	REV GCAGACGCAGATCCGGCAG
ALP	FW GCAGACGCAGATCCGGCAG
	REV CCTGGCTTTCTCGTCACTCTCA
ACAN	FW CAGGCTATGAGCAGTGTGACGC
	REV GCTGCTGTCCTTGTCACCCACG
COL2A1	FW GCCGAGGTGATAGTGTGGTT
	REV AACGGGGATGGCCTTGTATG
COMP	FW GGAGATGCTTGTGACAGCGATC
	REV TGAGTCCTCCTGGGCACTGTTA

4.2.4 Alizarin Red S staining

Alizarin Red S was conducted as previously described in Chapter 3.

4.2.5 Statistics

Statistical analyses were conducted as previously described in Chapter 2.

4.3 RESULTS

4.3.1 *CASC20* overexpression results in increased mineralisation during osteoblast differentiation of stem cells

I performed osteoblast differentiation experiments for over 20 days using *CASC20*-overexpressing hMADs, P512MSCs cells, and ASC52telo. I observed a significant increase in calcium deposits in hMADs undergoing osteoblast differentiation at day 20 compared to day 0 (FC > 50) in wild type cells, as indicated by Alizarin Red Staining. Additionally, *CASC20* overexpression led to increased mineralisation at both day 10 (FC = 23.2, p = 0.0093), but by day 20 mineralisation under both conditions was similar. (Figure 4.1).

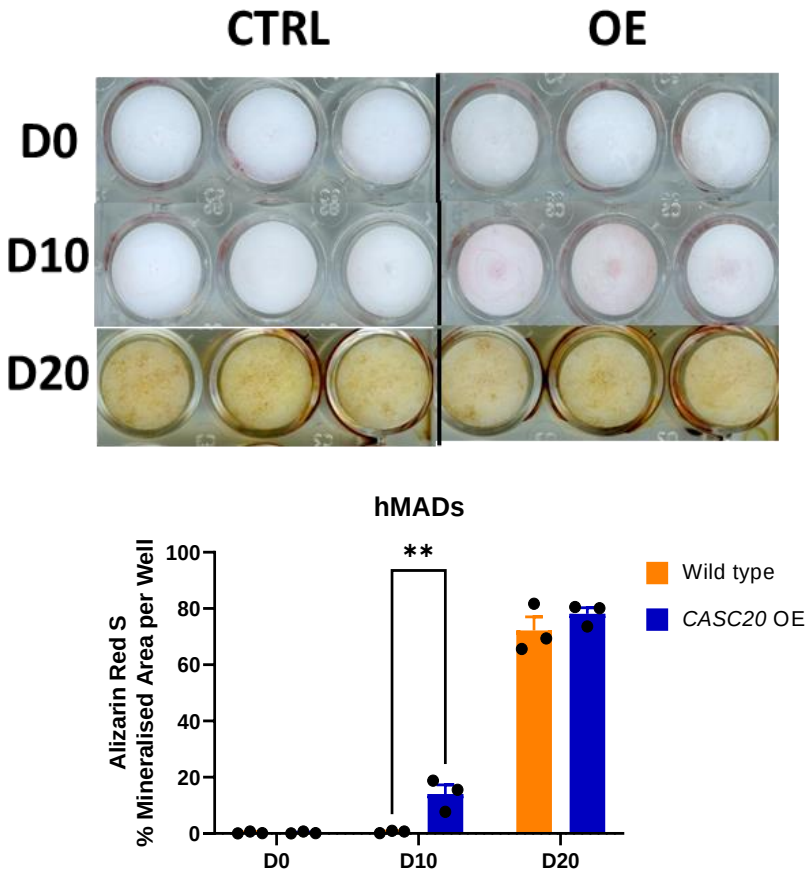


Figure 4.1. Alizarin Red Staining reveals enhanced calcium deposits in *CASC20* overexpressing hMADs during early osteoblast differentiation. Alizarin Red was used to stain calcium deposits at days 0, 10, and 20. Analyses are two-way ANOVA with Sidak's multiple comparisons test; data is plotted as mean ±SEM. **: p<0.01, N=3, technical replicates.

I next performed qPCR analysis for *RUNX2*, *OSX*, *ALP*, and *COL1A1* using hMAD cells to examine the gene expression patterns and determine if similar associations are observed. The results demonstrated a tendency of increased expression for *RUNX2*, *OSX*, and *ALP* at day 10 in *CASC20*-overexpressing cells compared to the wild type group, however, these differences

did not reach statistical significance. Also, *COL1A1* was already expressed at a high endogenous level in hMAD cells and did not follow the same trend observed for the other genes as shown in Figure 4.2.

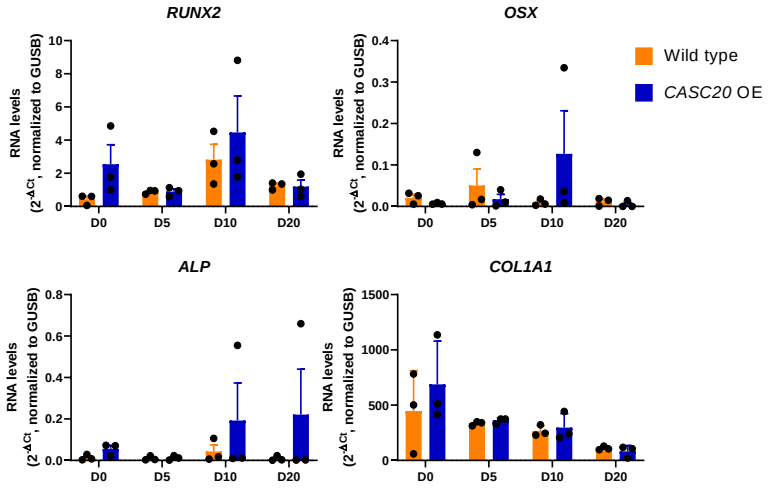


Figure 4.2. RT-qPCR for osteogenic genes in *CASC20* overexpressing hMADs at day 10 of the osteoblast differentiation. Alizarin Red was used to stain calcium deposits at days 0, 10, and 20. Analyses are two-way ANOVA with Sidak's multiple comparisons; data is plotted as mean ±SEM; N=3, technical replicates.

I then induced lentiviral *CASC20* overexpression in P512-MSCs. RT-qPCR demonstrated robust expression of *CASC20* in the OE group compared to no detectable expression in the wild type (Figure 4.3) at both day 0 and day 20 of osteodifferentiation (Figure 4.3).

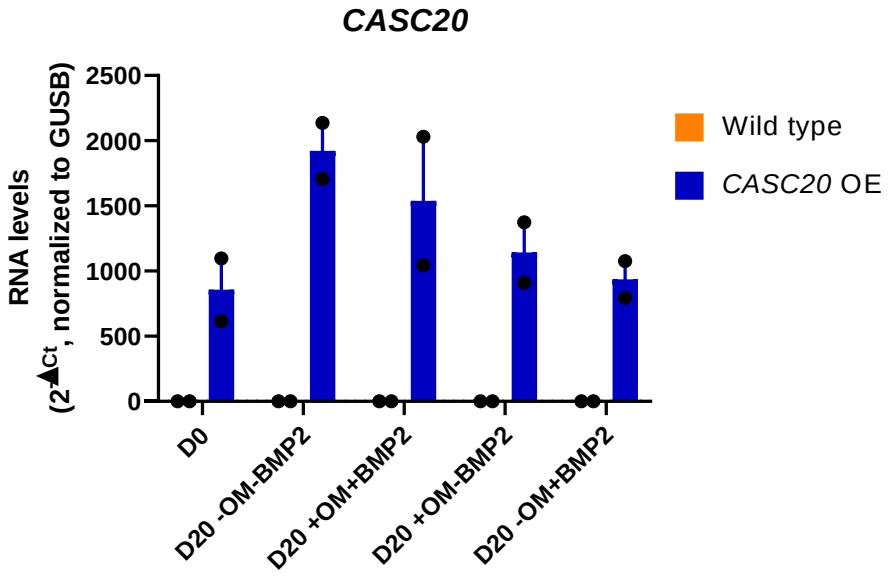


Figure 4.3. RT-qPCR demonstrated expression of *CASC20* in OE versus no expression in wild type in P512-MSCs. The cells were differentiated for 20 days using +/-OM+/-BMP2 media. Data is plotted as mean ±SEM; N=2, technical replicates.

I observed a significant increase in calcium deposits in P512MSCs cells undergoing osteoblast differentiation at day 20 compared to day 0 (FC > 50, p = 0.0137), as indicated by Alizarin Red S, as shown in Figure 4.4. The results did not show statistically significant differences between the conditions. However, there was a trend indicating an increase in expression of osteogenic genes with *CASC20* overexpression compared to the wild type group.

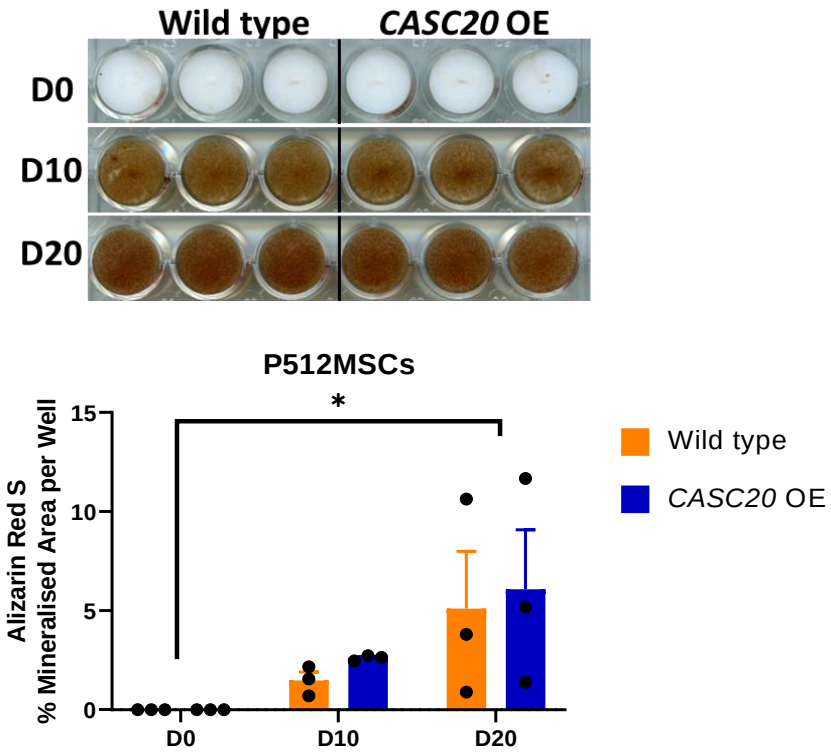


Figure 4.4. Alizarin Red S deposits in *CASC20* overexpressing P512MSCs Cells during osteoblast differentiation. Alizarin Red was used to stain calcium deposits at days 0, 10, and 20. Analyses are unpaired t test; data is plotted as mean ±SEM; *p<0.05; N=3, technical replicates.

The ASC52telo did not undergo osteoblast differentiation, as evidenced by the absence of calcium deposits visualised by Alizarin Red S. Microscopic examination revealed a fibroblastic phenotype, as shown in Figure 4.5.

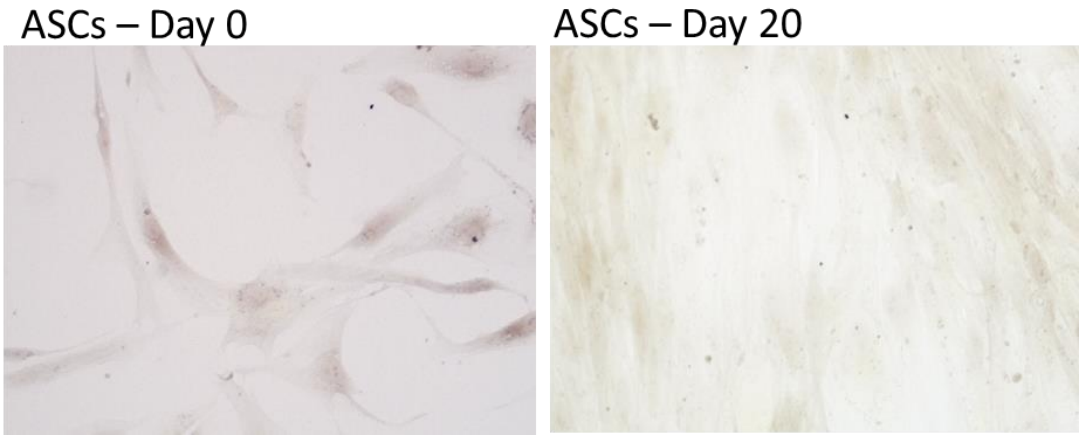


Figure 4.5. Representative microscopic images of ASC52telo at days and 20 of the differentiation. Cells were stained with Alizarin Red S and captured at a magnification of 20x.

The findings from these experiments demonstrate that *CASC20* overexpression exerts a positive effect on mineralisation in hMADs that is most marked at day 10, with a similar, but not significant, trend in P512MSCs. Correspondingly, the examination of osteogenic marker gene expression in hMADs aligns with these trends. Hence, I concluded that a more focused investigation on day 10 is warranted to unravel the effects of *CASC20* overexpression on osteoblast differentiation.

4.3.2 Upscaling the osteoblast differentiation setup for RNA-sequencing

To facilitate total RNA-sequencing analysis, I next upscaled the osteoblast differentiation setup to generate a higher yield of RNA for comprehensive transcriptomic profiling. As I previously used 24-well plates, the cells were cultured in larger tissue culture plates to increase the RNA output to a minimum of 1.5μg, which would be sufficient for RNA-Seq analysis encompassing both miRs and mRNAs. The differentiation process was carried out for 10 days, with samples collected at days 0 and 10 for subsequent analysis. Considering more promising pilot data indicating superior responsiveness at day 10 in hMAD cells compared to P512MSCs, the upscale experiments were conducted using hMAD cells.

I first differentiated hMADs into osteoblasts in the 12-well plate. Upon analysis, I observed an increase in mineralisation at day 10 in the *CASC20*-overexpressing cells compared to the baseline ($p = 0.0497$), as depicted in Figure 4.6. However, the observed trend between *CASC20* expression groups, although consistent with previous data, did not reach statistical significance ($p= 0.2127$), potentially due to the experiment being underpowered, given n=3 repeats.

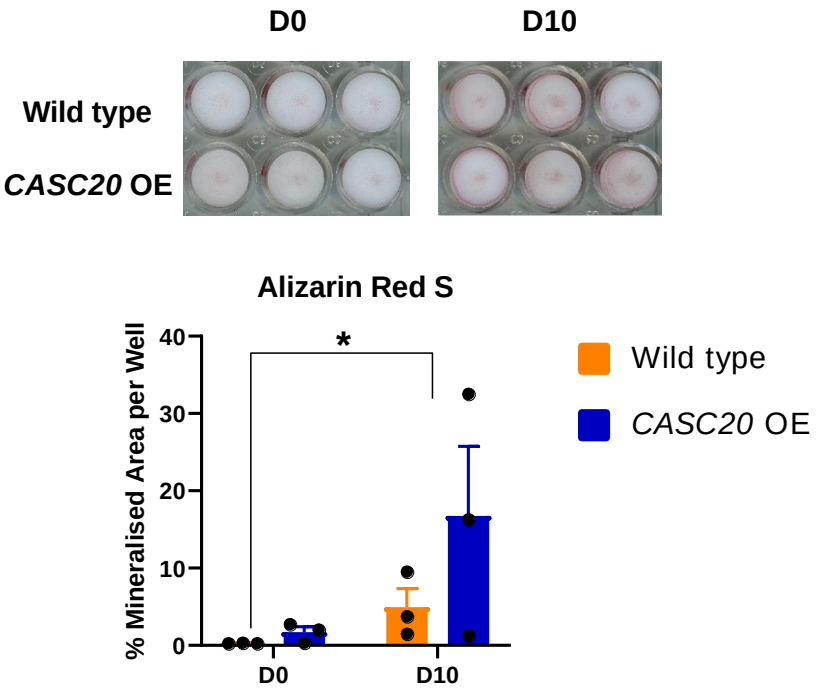


Figure 4.6. *CASC20* overexpression in hMAD using a 12-well plate. Alizarin Red S was used to stain calcium deposits at days 0 and 10. Analyses are unpaired t test; data is plotted as mean ±SEM.; *: p<0.05; N=3, technical replicates.

I then proceeded to upscale the experiment by using a 10-cm dish to increase the yield of cells for analysis. This decision was made because the 12-well wells did not yield enough RNA for the desired analysis. The cells were differentiated into osteoblasts for a period of 10 days, and samples were collected at days 0, 5 and 10 for subsequent qPCR analysis. The qPCR results demonstrated that *CASC20* overexpression led to a significant increase in *ALP* RNA expression at day 5 in *CASC20* overexpressing cells versus wild type (FC = 2.7, p = 0.0385). At day 10, a similar trend towards upregulation was observed, although it did not reach statistical significance (FC = 2.9, p = 0.0547). However, no significant difference was observed in the expression levels of *COL1A1*. For *RUNX2*, there was a trend towards downregulation in *CASC20*-overexpressing cells compared to the wild type group, but it was not statistically significant, as depicted in Figure 4.7.

Taken together, my data thus far indicates that *CASC20* may positively regulate *ALP* to control osteogenesis. However, as the hMAD cells used were at a higher passage number (>9), it is possible that they may have lost their differentiating capacity and giving rise to the disparity in results compared to my previous findings. It is also important to acknowledge the challenges posed by using Alizarin Red S on the larger surface area of the 10-cm dish, which could compromise accuracy and reliability. Based on these observations, I decided to continue using the P512MSCs cells as these had greater osteogenic capacity and were of earlier passage number, and the 12-well plate setup.

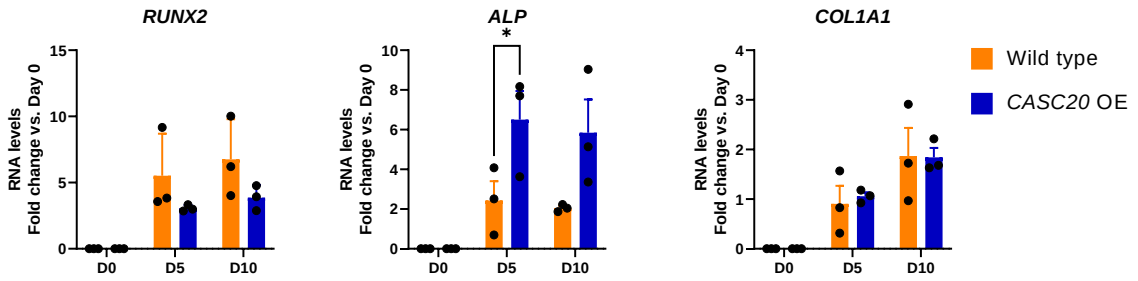


Figure 4.7. Effect of *CASC20* overexpression on the expression of key genes in a 10cm-dish using hMAD cells. The data were analysed by normalising to GAPDH and day 0. Analyses are two-way ANOVA Sidak's multiple comparisons test; the data are presented as the mean ± SEM; * P<0.05; N = 3, technical replicates.

4.3.3 Comparatively analysis of *CASC20*-overexpressing P512MSCs versus wild type cells at day 10

In view of the finding that the hMADs were no longer differentiating consistently and the lack of a resupply from the source lab, I changed to comparative analysis of the expression levels of key osteogenic and chondrogenic markers in P512MSCs at day 10 of the differentiation process. Specifically, I examined the expression of osteogenic markers, including *RUNX2*, *ALP*, and *COL1A1*. Additionally, I assessed the expression of chondrogenic markers, namely *ACAN*, *COL2A1*, and *COMP*, to test for chondrogenic differentiation during the osteogenic differentiation protocol.

Both *CASC20* over-expressing and wild type cells demonstrated an increase in expression of *RUNX2* ($p = 0.0133$), *COL1A1* ($p = 0.0002$ and $p = 0.0008$, respectively) between day 0 and day 10. *ALP* demonstrated a trend of upregulation between day 10 and day 0; however, this trend was not statistically significant. The p-values for wild type and overexpressing cells were 0.2670 and 0.0694, respectively.

However, again, although there was a general trend towards increased differential expression for *RUNX2* and *COL1A1*, this trend was inconsistent across the various transcription factors (Figure 4.8). Chondrogenic transcription factors tended to decrease over 10 days, but again there was inconsistent differential expression between genes (2-way ANOVA P>0.05).

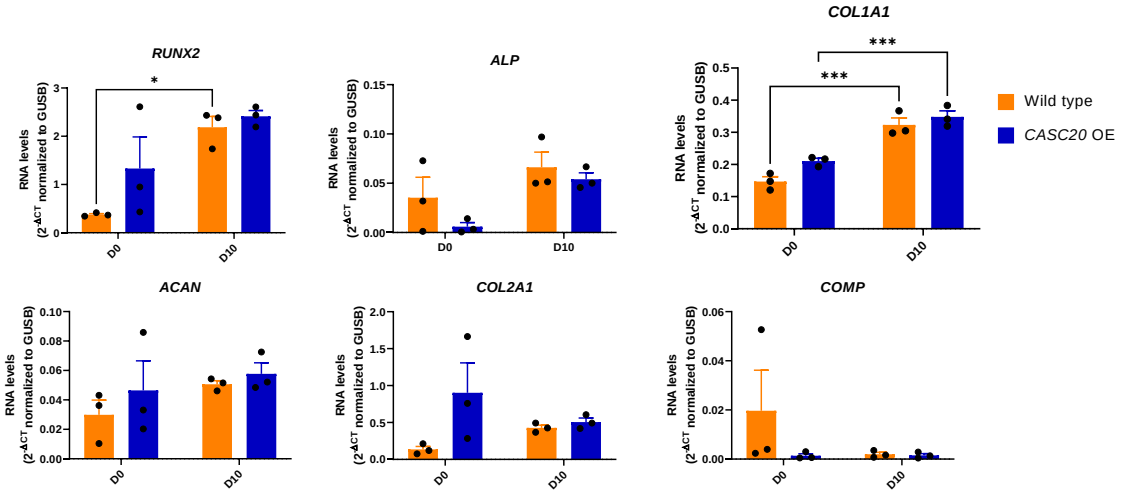


Figure 4.8. Expression analysis of osteogenic (*RUNX2*, *ALP*, *COL1A1*) and chondrogenic (*ACAN*, *COL2A1*, *COMP*) markers in wild type and *CASC20*-overexpressing P512MSCs cells at day 0 and day 10. The analysis included three replicates. Analyses are two-way ANOVA Sidak's multiple comparisons test; the data are presented as the mean ± SEM; *P<0.05, ***P<0.001; N= 3, technical replicates.

4.3.4 The effect of *CASC20* overexpression on osteoblast differentiation using P512MSCs and +/-*CASC20*, +/-OM, +/-BMP2

Given the issues with upscaling identified in the previous 2 experiments, I next elected to take a step back and review the osteogenic culture conditions and their influence on cell differentiation, as defined by the biological readout of Alizarin Red S, in the P512MSCs as these cells were of the lowest passage number and most dynamic in terms of differentiation capacity. In this experiment the culture conditions were thus +/- OM and +/- BMP2 an +/- *CASC20* over-expression. Samples were collected at days 0, 3, 10, and 20 and stained using Alizarin Red S.

The results demonstrated that *CASC20* overexpression again had a positive trend in its impact on osteoblast differentiation, although the differences are not statistically significant (Figure 4.9). In the absence of *CASC20*, the cells exhibited lower calcium deposition at day 20, except for the -OM-BMP2 condition. This suggested that the function of *CASC20* may be dependent on the presence of osteogenic stimuli. Interestingly, at day 10, I observed higher calcium deposits in the +BMP2 conditions, consistent with the importance of BMP2 in promoting early osteogenic differentiation.

Based on the mineralisation data, it was proposed to perform RNA-Sequencing analysis on the +OM-BMP2 condition as it exhibited the most consistent results. Although no significant difference was observed at day 10, the data revealed that by day 20, there was a positive trend in the *CASC20*-overexpressing cells, exhibiting higher calcium deposition compared to

the wild type group. This suggested that the presence of OM played a role in enhancing the osteoblast differentiation potential of *CASC20*-overexpressing cells. Therefore, conducting RNA-Sequencing on this specific condition would provide insight into the molecular mechanisms underlying the observed differences in calcium deposition.

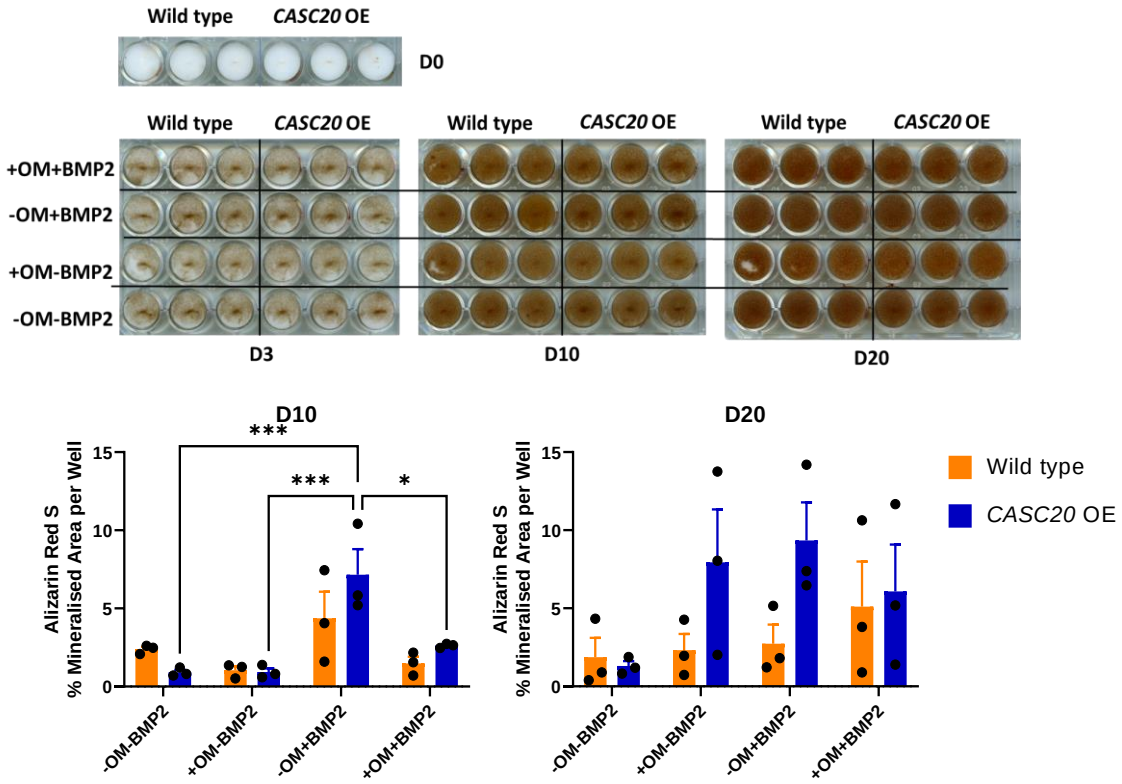


Figure 4.9. Effect of +/-*CASC20*, +/-OM and +/- BMP2 in osteoblast differentiation of P512MSCs. Alizarin Red S was used to stain calcium deposits at days 0, 3, 10, and 20. Analysis are two-way ANOVA with Sidak's multiple comparisons; data is plotted as mean ±SEM; *P<0.05, ***P<0.001; N= 3, technical replicates.

4.4 DISCUSSION

In this chapter, I investigated the effects of *CASC20* overexpression on osteoblast differentiation using hMADs, P512MSCs cells, and ASC52telo. The results demonstrated that *CASC20* overexpression enhanced mineralisation in both P512MSCs and hMADs, particularly at day 10 of differentiation. Furthermore, gene expression analysis revealed differential expression patterns for osteogenic markers. The findings suggest the need for further investigation focused on day 10. In the osteoblast differentiation experiments, I observed that ASC52telo did not undergo osteoblast differentiation, as evidenced by the absence of calcium deposits and the presence of a fibroblastic phenotype. This confirms the lack of differentiation potential in ASC52telo that led to our discontinuing experiments on this cell type.

In hMADs, *CASC20* overexpression led to increased mineralisation at day 10. These results suggest that *CASC20* plays a positive role in enhancing the mineralisation process during osteoblast differentiation. Furthermore, the qPCR analysis revealed increased expression *ALP*, in *CASC20*-overexpressing hMADs at day 10 compared to the wild type group. However, the observed differential expression of *RUNX2* and *COL1A1* did not reach statistical significance, which may suggest a differential regulation of these genes during osteoblast differentiation or could be attributed to the experiment being underpowered. These findings highlight the complexity of gene expression patterns and emphasise the need for further investigation to understand the specific mechanisms underlying *CASC20's* effects on osteogenic gene expression.

In the case of P512MSCs cells, *CASC20* overexpression led to increased mineralisation at both day 10 and day 20, although the differences were not statistically significant. This consistent trend indicates a positive impact of *CASC20* on osteoblast differentiation. The analysis of osteogenic and chondrogenic markers in P512MSCs cells further supports the notion that *CASC20* may play a regulatory role in directing cell fate towards the osteogenic lineage.

Additionally, I investigated the effects of *CASC20* overexpression on osteoblast differentiation using P512MSCs cells in the presence or absence of *CASC20*, OM, and BMP2. The results indicated that *CASC20* RNA alone is not sufficient to induce osteoblast differentiation, as the presence of OM and/or BMP2 is necessary to promote differentiation. This finding aligns with existing literature on HO, which emphasises the requirement of a conducive microenvironment for successful differentiation (559-562). The study also reveals that *CASC20* RNA is responsible for controlling osteoblast differentiation, supporting the exploration of miR expression to elucidate the molecular mechanisms of *CASC20* on bone formation.

Chapter 5 – Effects of *CASC20* overexpression on miR profile during osteogenesis

5.1 INTRODUCTION

The first step in osteoblastic differentiation is osteogenic stimulation of mesenchymal stem cells (MSCs) (563). MSCs undergo differentiation during this phase to become proliferative pre-osteoblasts to create the extracellular matrix, a critical scaffold for the development of mineralised bone tissue (548). Pre-osteoblasts eventually develop into fully functional osteoblasts as the extracellular matrix stiffens and mineralisation occurs through the deposition of hydroxyapatite (548). Precise genetic regulation is necessary to coordinate this dynamic cellular change (548).

Although the biological characterisation of osteoblast development has been extensively studied, a thorough transcriptional analysis that concentrates on MSCs that overexpress *CASC20* has not yet been published. The main goal of the studies described in this chapter was to examine miR expression profile associated with osteoblast differentiation with a focus on *CASC20* modulation of their expression. To do this, I used osteogenic induction medium (OM) to differentiate wild type and *CASC20* overexpressing P512MSCs into osteoblasts and collected samples on days 0, 10, and 20. The dynamic changes in gene expression that occurred during the osteogenic differentiation process were then captured by miR sequencing.

Using osteogenic induction media (OM), I was able to stimulate osteoblastic development in both wild-type and *CASC20*-overexpressing P512MSCs. The selection of the timepoints (days 0, 10, and 20) was informed by previous data collected in Chapter 4, which demonstrated the significant impact of *CASC20* on osteoblast differentiation by day 10.

To quantitate miR expression, I used miR sequencing, a high-throughput technology that permits the analysis of the complete miR transcriptome. Using this method, I was able to identify miRs that exhibited differential expression during the stimulation of osteogenesis in the wild type versus *CASC20*-overexpressing P512MSCs at various time points.

I identified distinctive miR expression patterns using unsupervised clustering analysis and identified the mRNA targets of differentially expressed miRs. Additionally, I functionally annotated the differentially expressed miRs to identify linked pathways using Kyoto encyclopaedia of genes and genomes (KEGG) analyses (564).

5.2 MATERIAL and METHODS

5.2.1 Tissue culture and osteoblast differentiation

P512MSCs were cultured and differentiated into osteoblasts using +OM-BMP2 media and 24-well plates, as previously described in Chapter 4.

5.2.2 RNA quality and sequencing

RNA samples were sent to Novogene (Beijing, China) for sequencing. The following QC steps were then conducted: RNA was run on an agarose gel to determine RNA integrity, contamination and degradation; Next, RNA was run on an Agilent 2100 (Beijing, China) bioanalyser to quantify RNA concentration, as well as RNA integrity number (RIN) scores based on the rRNA peaks. Then nanodrop was used to look for key wavelength ratios that are indicative of contamination by gDNA or protein.

RNA was sequenced according to the following method. 3' and 5' adaptors were ligated to 3' and 5' end of small RNA, respectively. Then the first strand cDNA was synthesised after hybridisation with reverse transcription primer. The double-stranded cDNA library was generated through PCR enrichment. After purification and size selection, libraries with insertions between 18~40 bp were checked with Qubit and real-time PCR for quantification and Bioanalyser for size distribution detection. Quantified libraries were pooled and sequenced on Illumina platforms (Beijing, China), according to effective library concentration and data amount required. Sequencing was run on an Illumina NovaSeq 6000 on an SP flow cell using the SE50 method. The sequencing data was transferred back using Novogene's Customer Service System (CSS).

5.2.3 RNA-Sequencing analysis

Upon receiving the sequencing data, I used FastQC to assess quality, and removed poor quality and adapter contamination using Trimmomatic (version 0.39). Reads were aligned to the human reference genome (hg38, Ensembl release 93) using STAR (version 2.7.10b). A count matrix containing the number of reads mapped to each gene in each sample was generated using FeatureCounts (version 2.0.3).

To compare aligned reads across samples, I used DESeq2) (565) (version 1.12.3) (script in Appendix 9.2.1), which normalises raw read counts considering the sample size factor. The dataset underwent statistical analysis using the DESeq method to estimate size factors and dispersion values for each gene. A variance stabilising transformation (VST) was then applied to ensure the data's suitability for analysis. The resulting transformed data was used to generate PCA plots. For hierarchical clustering, the PCA plot was graphed using plotPCA (565) for visualising the overall effect of experimental covariates and batch effects. Heatmaps of the count matrix and sample-to-sample distance matrix were plotted using pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html) (version 1.0.12) function

(script in Appendix 9.2.1). Log2FC were calculated using lfcshrink in DESeq2 (script in Appendix 9.2.2) (566).

5.2.4 Pathway enrichment analysis

The Encyclopedia of RNA Interactomes (ENCORI) database was downloaded (StarBase v3, https://rnasysu.com/encori/) and utilised to predict miR-mRNA interactions. Specifically, mRNAs that have been experimentally validated to interact with miRs and are predicted to bind with the same miR using at least one prediction algorithm within ENCORI were selected.

The RISmed R package (version 2.1.7, https://rdrr.io/cran/RISmed/) was used to conduct an electronic PubMed search to test whether targeted mRNAs are enriched for osteo- and chondrogenesis-related publications (Script in Appendix 9.2.3). The following relevant keywords were used: osteoblast, chondrocyte, chondrogenesis, osteogenesis, *BMP2*, *RUNX2*, *SOX9*, bone, cartilage, skeletal, ossification, and mineralisation. This search aimed to identify genes associated with osteo- or chondrogenesis, reported in PubMed between January 01, 2000, and August 04, 2023 (search date: August 04, 2023).

For GO/KEGG enrichment analyses, the GOSeq package (version 1.52.0) was utilised. Terms with an adjusted p-value of less than 0.05 were considered statistically significant. The GO/KEGG data were plotted using pheatmap in R (script in Appendix 9.2.4).

R (version 4.3.1) was used to run all packages except for the RISmed package, which was executed using version 4.1.3 due to its incompatibility with recent R versions.

5.3 RESULTS

5.3.1 *CASC20* differential expression affects global miR expression profile

Exploration of genome-wide miR expression profile was conducted by PCA (Figure 5.1A). The first principal component accounted for 57% of the variability, while the second component explained 17% of the variance. The PCA analysis demonstrated distinct segregation among the different time points (days 0, 10, 20) with no overlapping, confirming temporal miR transcriptional variation between day 0, 10 and 20. This Initial examination of the data also indicated that the maximum variation between treatment conditions occurred at day 10. This is also shown in the sample-set heatmap (Fig 5.1B). The gene level heatmap analysis (Figure 5.1C) confirmed clustering of individual replicates of each treatment condition and miR level variation between conditions, seen most clearly at day 10. The data displayed in figure 5.1 confirmed success of the experiment in terms of the progression of miR expression variation during induced human osteogenesis and modulation of miR expression over this time course by *CASC20*. Of the timepoints studied, day 10 demonstrated the greatest difference between treatments.

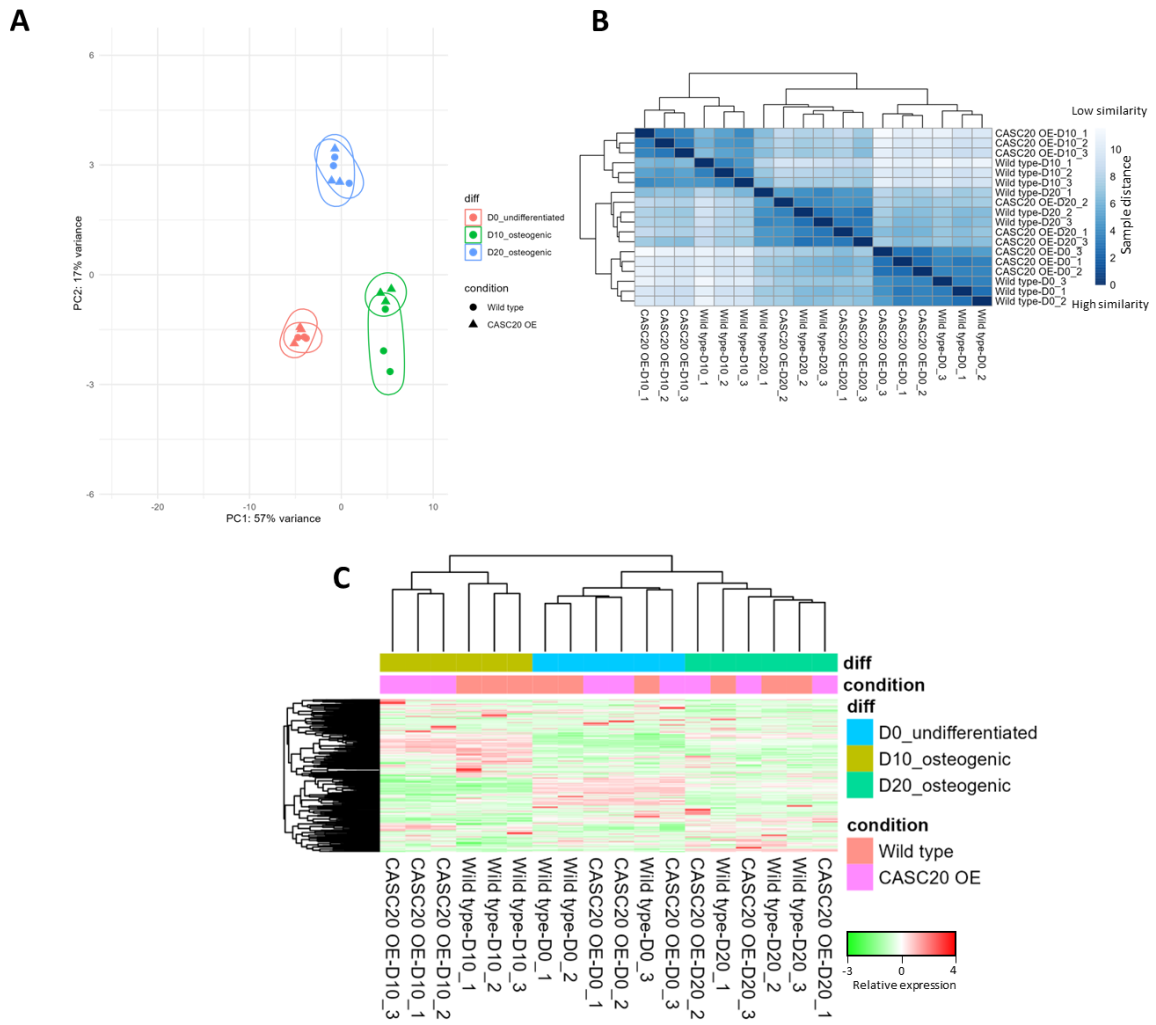


Figure 5.1. Global miR expression profiles during osteoblast differentiation in wild type versus *CASC20* overexpressing P512 MSCs. **A**. Principal component analysis was conducted on a matrix of gene expression data, which was transformed for analysis. Each point on the plot represents an experimental sample. The colour of the point indicates the time point (D0, D10, and D20), and the shape represents the experimental condition (OE and CTRL). **B**. A plot showing the distance between experimental groups based on a Euclidean distance metric. **C**. The expression heatmap was plotted to display the miR expression levels across different samples. Each row represents to a specific gene and each column represents an individual sample. The colour intensity signals the expression level, with dark red indicating higher expression and dark green lower expression.

5.3.2 Pairwise analysis of miR expression in *CASC20* OE versus WT P512 MSCs at day 0

To investigate the differences between *CASC20* OE and wild type, pairwise analysis was conducted at day 0 of the differentiation (see Figure 5.2). The results showed differential expression of 34 miRs at day 0 in *CASC20* OE versus wild type ($p < 0.05$), with 6 being upregulated and 28 downregulated. This is consistent with a negative regulation of miRs by *CASC20*. 8 miRs had log2FC greater than 0.5. 7 of these were downregulated in *CASC20* OE

versus wild type (hsa-let-7a-5p, hsa-mir-7-5p, hsa-mir-221-5p, hsa-let-7f-5p, hsa-mir-125b-5p, hsa-mir-26a-5p, hsa-mir-30c-5p, hsa-mir-3529-3p). 1 of the miRs was upregulated (hsa-mir-1298-5p). These miRs would therefore make the most likely targets for further investigation of the potential effect of *CASC20* on bone and cartilage formation.

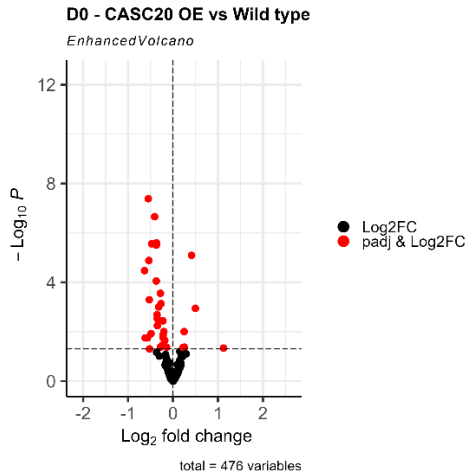


Figure 5.2. Volcano plot comparing *CASC20* OE vs WT at Day 0. The volcano plot depicts the differential expression analysis of miRs in *CASC20* OE compared to wild type at day 0. Each dot on the plots represents an individual miR, with the x-axis representing the log2 fold change (log2FC) and the y-axis representing the negative logarithm of the adjusted p-value (−log10p).

Table 5.1. Differential expression analysis of miRs in *CASC20* OE vs Wild type at Day 0 during osteoblast differentiation.

miR	log2FoldChange	padj
hsa-let-7a-5p	-0.6342	3.36E-05
hsa-mir-7-5p	-0.6186	1.78E-02
hsa-mir-221-5p	-0.5716	1.78E-02
hsa-let-7f-5p	-0.5506	4.10E-08
hsa-mir-125b-5p	-0.5377	1.31E-05
hsa-mir-26a-5p	-0.5284	5.08E-04
hsa-mir-30c-5p	-0.5229	4.94E-02
hsa-mir-3529-3p	-0.4955	1.22E-02
hsa-mir-181b-5p	-0.4893	1.21E-02

hsa-mir-362-5p	-0.4754	2.74E-06
hsa-mir-145-5p	-0.4084	2.19E-07
hsa-mir-10b-5p	-0.3903	3.09E-06
hsa-mir-708-5p	-0.3722	8.83E-05
hsa-mir-30a-5p	-0.3684	2.53E-06
hsa-mir-361-3p	-0.3547	2.05E-03
hsa-mir-125a-5p	-0.3506	2.96E-03
hsa-mir-1249-3p	-0.3472	5.68E-03
hsa-mir-92a-3p	-0.3202	9.90E-04
hsa-mir-16-5p	-0.2828	4.23E-02
hsa-mir-30e-5p	-0.2782	2.79E-04
hsa-mir-30d-5p	-0.2728	7.27E-04
hsa-mir-708-3p	-0.2380	3.56E-02
hsa-mir-182-5p	-0.2270	3.63E-03
hsa-mir-146b-5p	-0.2236	1.33E-02
hsa-mir-20a-5p	-0.2189	1.78E-02
hsa-mir-30a-3p	-0.2040	9.87E-03
hsa-mir-10a-5p	-0.1887	2.25E-02
hsa-mir-19a-3p	-0.1438	4.23E-02
hsa-mir-125a-3p	0.2219	4.36E-02
hsa-mir-99b-3p	0.2479	9.87E-03
hsa-mir-222-5p	0.2519	4.22E-02
hsa-mir-149-5p	0.4131	8.10E-06
hsa-mir-10395-3p	0.4987	1.13E-03
hsa-mir-1298-5p	1.1259	4.60E-02

5.3.3 *In silico* analysis of genes that may interact with miRs downregulated in *CASC20* OE versus wild type P512MSCs at day 0

Next, to explore the potential regulatory role of *CASC20* on undifferentiated cells through miRs, I investigated the genes targeted by the downregulated miRs, utilising the StarBase v3 database. Subsequently, I established a ranking of these targeted genes based on the number of miRs with which they exhibit interactions. I chose to concentrate on genes exhibiting the most extensive miR interactions, as depicted in Figures 5.3 and 5.4.

The compiled data showed that 63 genes are subject to targeting by ~74% of miRs (20 out of the 28 downregulated miRs), including osteogenic and chondrogenic genes ACVR2B, INO80D and LONRF2 (Figure 5.4). Next, I used RISMed to ascertain the extent to which all of the target genes are documented participants in osteo- and chondrogenesis. This *in silico* analysis showed that of the 63 genes, 55 genes had experimental evidence for a role in osteo- or chondrogenesis (Table 5.2).

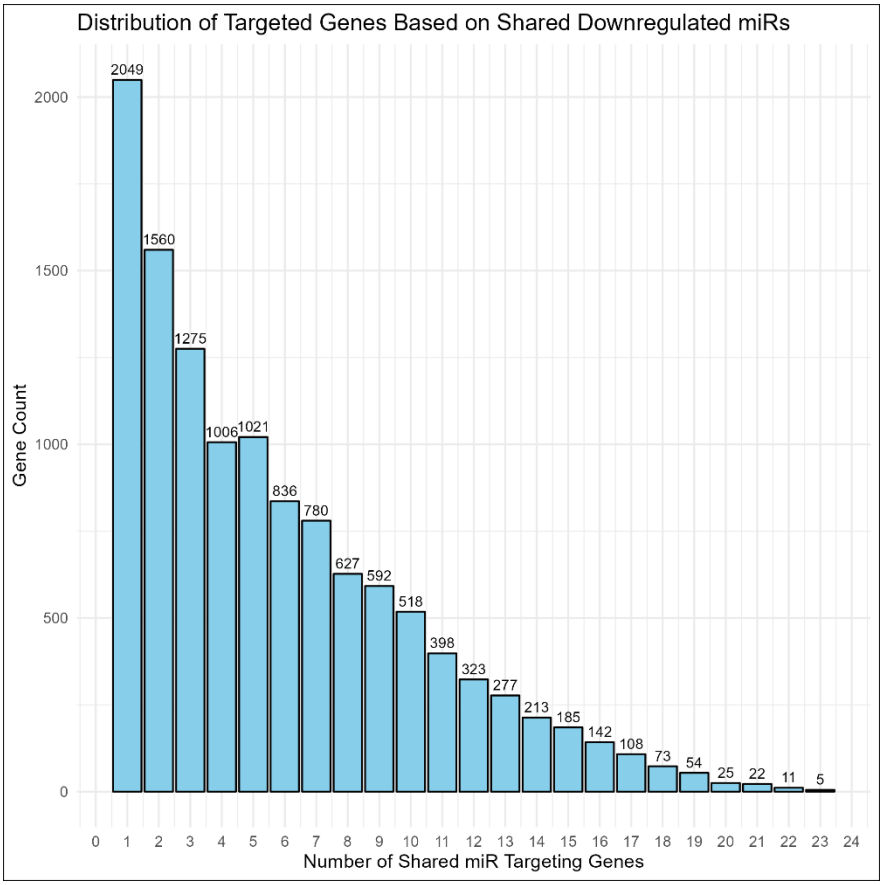


Figure 5.3. Distribution of targeted genes based on shared downregulated miRs in *CASC20* OE vs wild type at day 0. The figure illustrates the number of genes targeted by miRs, ranked by the number of shared miR-targeting genes on the x-axis and the gene count on the y-axis.

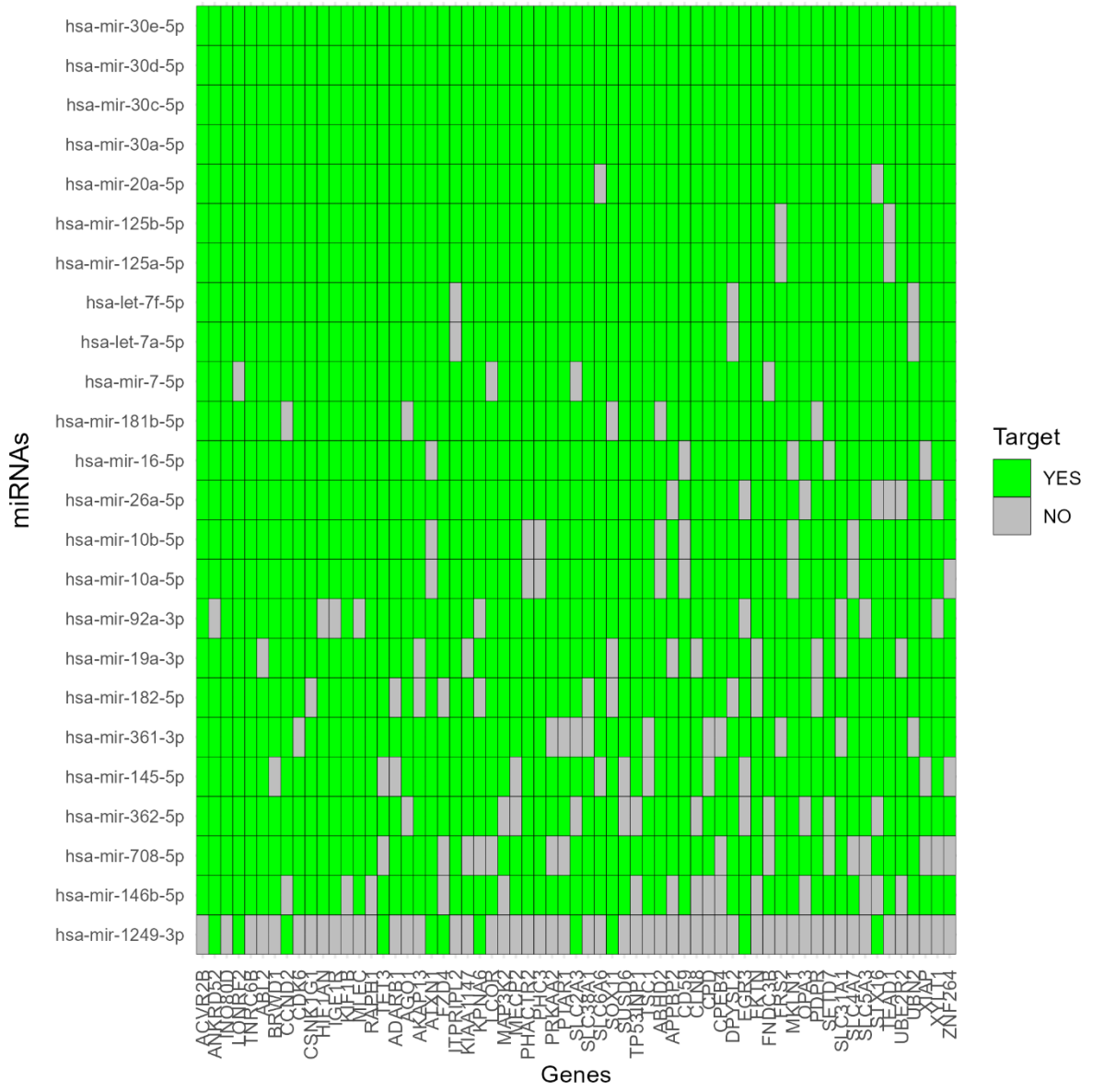


Figure 5.4. Top 63 genes targeted by downregulated miRs in *CASC20* OE vs wild type at day 0. The heatmap showcases the top 53 genes targeted by downregulated miRs. The genes are ranked so that those interacting with the highest number of miRs are positioned towards the left, and the miRs targeting the greatest number of mRNA are placed towards the top of the heatmap.

Table 5.2. Putative gene targets for downregulated miRs and most recent PMIDs for their published role in osteo or chondrogenesis.

Target gene	**Most recent PMID number corresponding with published role in osteo or chondrogenesis (search date August 04, 2023)**
ACVR2B	37486509
ANKRD52	34054925
INO80D	30253751
LONRF2	36888978
TNRC6B	33889172
ABL2	37587494
BRWD1	34082824
CCND2	37510277
CDK6	37602543
CSNK1G1	33416120
HIF1AN	33619902
IGF1R	37605180
KIF1B	35234284
MLEC	16440295
RAPH1	26695371
TET3	37095518
ADARB1	37287534
AGO1	36890226
AKAP13	33380835
ATXN1	36810326
FZD4	37340863

LCOR	29663375
MAP3K2	35707845
MECP2	37340152
PHC3	37468993
PRKAA2	37005694
PTAR1	32128853
SLC2A3	34031595
SLC38A1	37055385
SLC6A6	37202645
SOX11	37543028
TP53INP1	34718338
TSC1	37634327
ABHD2	37298205
APPBP2	30210606
CD59	37603218
CLN8	32518749
CPD	37518893
CPEB4	37310402
DPYSL2	37479784
EGR3	37612521
FKTN	36760122
FNDC3B	27541078
FRS2	37264620
MKLN1	35138470
OPA3	31119193

PDPR	30024048
SETD7	36209579
SLC31A1	31658633
SLC4A7	20079835
SLC5A3	35899258
STX16	35119251
TEAD1	37197086
XIAP	37527738
XYLT1	37296099

5.3.4 KEGG pathway enrichment analysis of inferred target genes of downregulated miRs at day 0

Next, I applied KEGG to identify key molecular pathways affected by the combined mRNA target list to understand the network of interactions that the downregulated miRs at day 0 have with their target mRNAs (Figure 5.5). One of the prominent pathways identified was "Pathways in Cancer". This is expected, as *CASC20* was first identified in Cancer-associated pathway and underscores the complexity of *CASC20's* miR-mediated regulatory networks that can impact cell proliferation, apoptosis, and metastasis. The next 4 top pathways affected by the miRs were the MAPK signalling pathway, endocytosis, focal adhesion, and protein processing in endoplasmic reticulum. Osteo- and chondro related pathways targeted included the Wnt signalling pathway and mTOR signalling pathway.

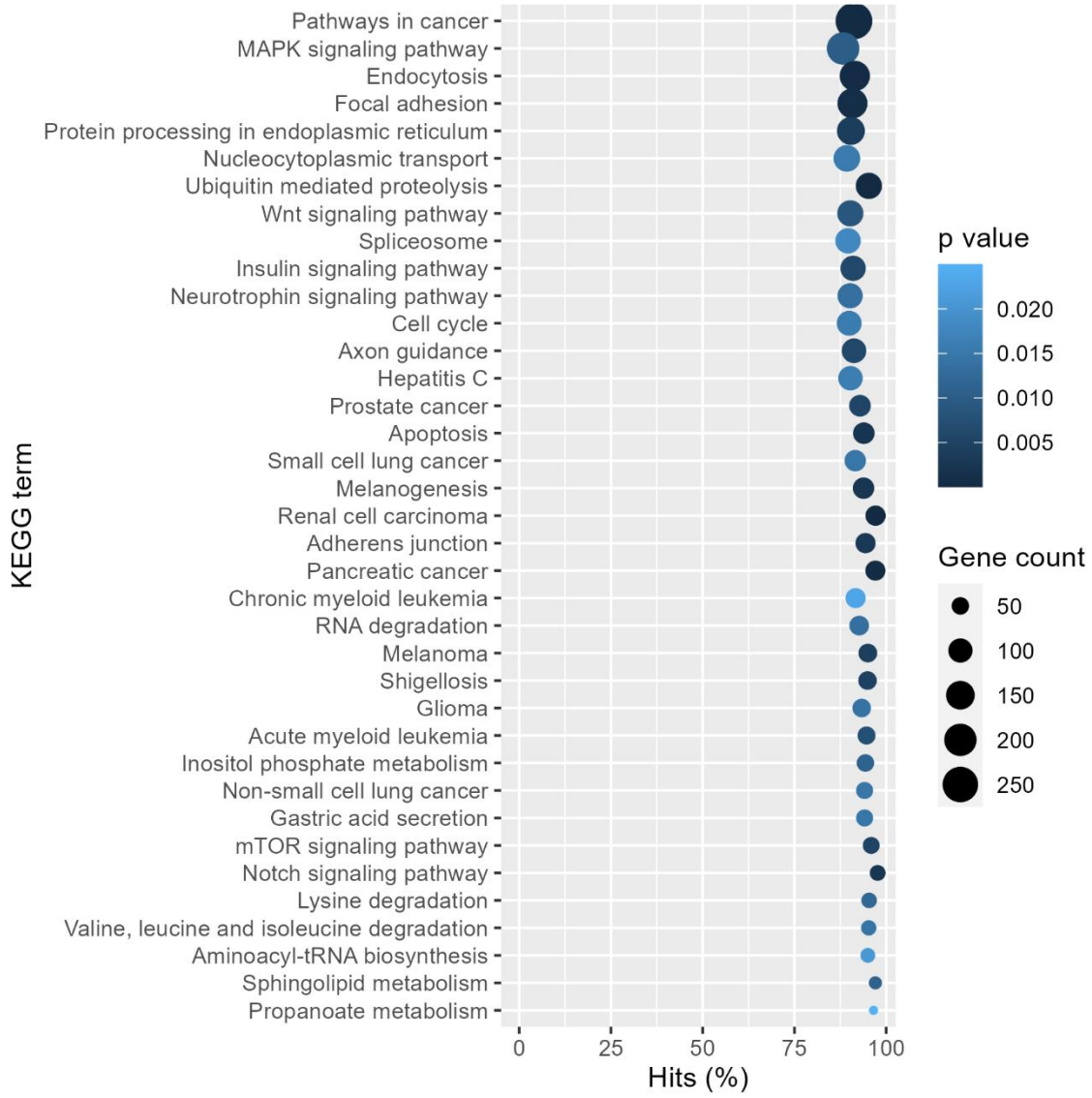


Figure 5.5. Enrichment plot showing KEGG pathways targeted by downregulated miRs at day 0. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the KEGG pathway terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

5.3.5 *In silico* analysis of genes that may interact with miRs upregulated in *CASC20* OE versus wild type P512MSCs at day 0

Subsequently, I conducted an analysis of the genes targeted by the upregulated miRs using the same methodological approach, prioritising the list of targeted genes based on the extent of their interactions with miRs (Figure 5.6).

The outcome of this approach showed 950 genes as targets of 2 out of the 6 upregulated miRs (hsa-mir-1298-5p and hsa-mir-149-5p), as shown in Figures 5.6. To identify the most prominent target genes, genes that were demonstrated to be degraded through interaction with these miRs were selected. The analysis showed that 60 genes out of the 950 have been demonstrated to be degraded by interaction with these miRs, with 40 being degraded by miR-149-5p and 20 by miR-1298-5p (Figure 5.7).

Next, I applied the RISMed tool to the identified targets to determine If they had a confirmed published role in these processes. This analysis confirmed that 52 of the 60 target genes had published functional evidence for such a role (Table 5.3).

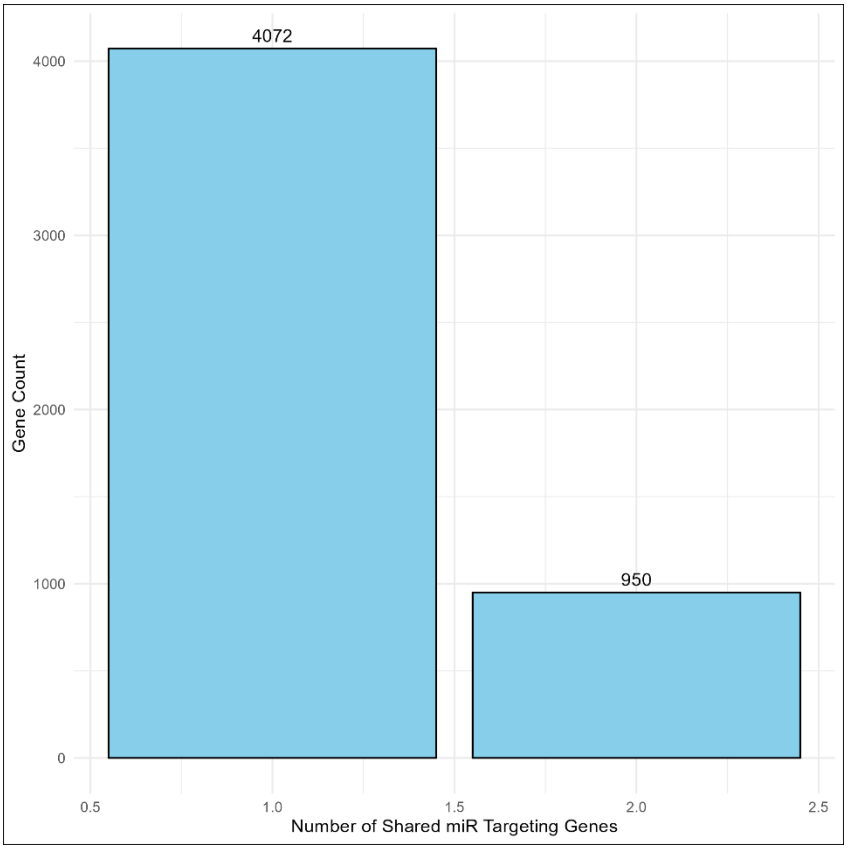


Figure 5.6. Distribution of targeted genes based on shared upregulated miRs in *CASC20* OE vs wild type at day 0. The figure illustrates the number of genes targeted by miRs, ranked by the number of shared miR-targeting genes on the x-axis and the gene count on the y-axis.

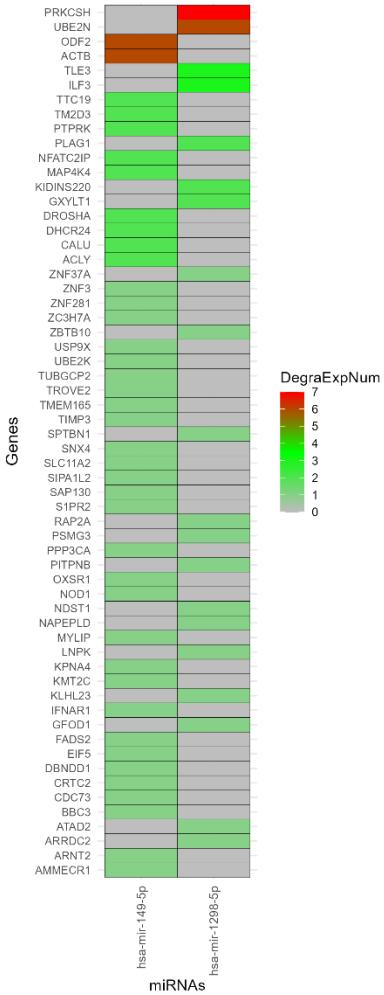


Figure 5.7. 60 Genes degraded by upregulated miRs in *CASC20* OE vs wild type at day 0. The heatmap shows the 60 genes targeted and degraded by the upregulated miRs at day 0. DegraExpNum = number of experimental studies validating miR mediated degradation. The genes are ranked so that those with the highest DegraExpNum are positioned towards the top of the heatmap.

Table 5.3. Putative gene targets for upregulated miRs and most recent PMIDs for their published role in osteo or chondrogenesis.

Target gene	Most recent PMID number corresponding with published role in osteo or chondrogenesis (search date August 04, 2023)
ACLY	36787367
ACTB	37571423
AMMECR1	29787394
ARNT2	32021278
ARRDC2	35736796
ATAD2	34922489
BBC3	37335333
CALU	34740309
CDC73	36928741
CRTC2	37441497
DBNDD1	35474152
DHCR24	36877347
DROSHA	37549955
EIF5	29151592
FADS2	37614423
GXYLT1	37614303
IFNAR1	37559153
ILF3	36040165
KIDINS220	36982724
KMT2C	37387515
KPNA4	33537823

LNPK	35179257
MAP4K4	37312710
MYLIP	32072135
NAPEPLD	37531659
NDST1	35354833
NFATC2IP	35883195
NOD1	37551879
ODF2	36933475
OXSR1	31085334
PITPNB	30042096
PLAG1	37486535
PPP3CA	37098184
PRKCSH	30149291
PTPRK	32217638
RAP2A	36366779
S1PR2	37443710
SAP130	27628766
SIPA1L2	36802481
SLC11A2	34142171
SPTBN1	37391169
TIMP3	37647255
TLE3	37392376
TMEM165	34930890
TROVE2	29448111
UBE2K	33875618

UBE2N	30834083
USP9X	36653407
ZBTB10	11282075
ZC3H7A	18682727
ZNF281	32788627
ZNF3	15257610

5.3.6 KEGG pathway enrichment analysis of inferred target genes of upregulated miRs at day 0

Next, I applied KEGG to identify key molecular pathways affected by the combined mRNA target list to understand the network of interactions that the downregulated miRs at day 0 have with their target mRNAs (Figure 5.8). As previously, the most prominent pathway identified was "Pathways in Cancer". The next 4 top pathways affected by the miRs were MAPK signalling pathway, endocytosis, focal adhesion, and regulation of actin cytoskeleton. Osteo- and chondro related pathways targeted included osteoclast differentiation, VEGF signalling pathway, and mTOR signalling pathway.

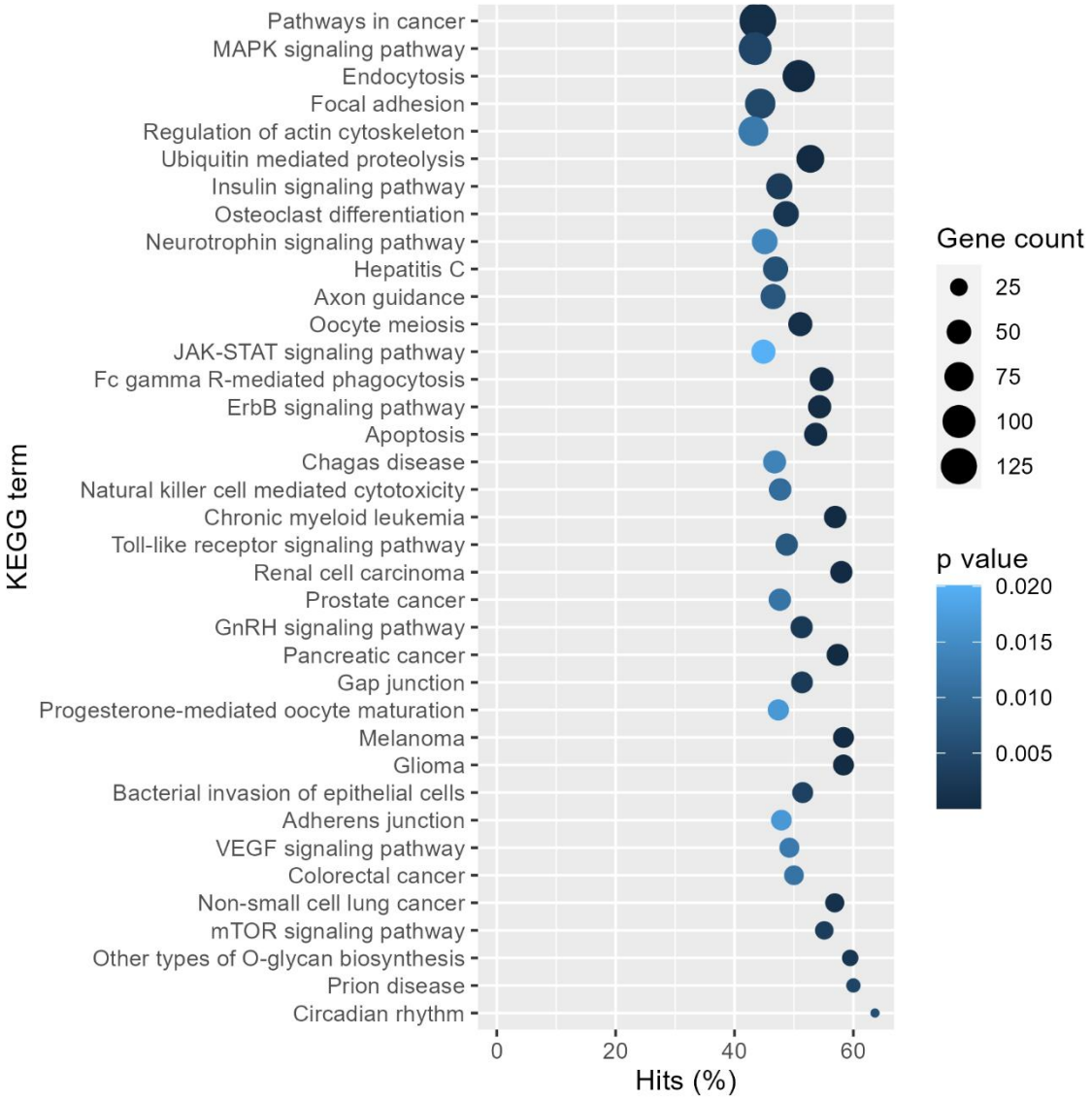


Figure 5.8. Enrichment plot showing KEGG pathways targeted by upregulated miRs at day 0. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the KEGG pathway terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

5.3.7 Pairwise analysis of miR expression in *CASC20* OE versus WT P512MSCs at day 10 and 20

To further investigate the differences between *CASC20* OE and wild type, pairwise analysis was conducted at each time point during osteodifferentiation (see Figure 5.9). In comparison to the 34 miRs differentially expressed at day 0 (28 down and 6 upregulated in the *CASC20* OE cells), at day 10, 43 miRs exhibited differential expression in *CASC20* OE versus wild type (p <

0.05), with 28 upregulated and 15 downregulated. However, at day 20, only one miR was differentially downregulated in *CASC20* OE versus wild type and none were upregulated.

At day 10, the miRs with the greatest fold change, hsa-mir-5100 (FC = -1.31988, p = 0.000111) and hsa-mir-1249-3p (FC = -1.08343, p = 4.19E-13), are downregulated in *CASC20* OE versus wild type. In contrast, the upregulated miR with the highest fold change is hsa-mir-155-5p (FC = 0.89998, p = 0.010408) as shown in Table 5.4. These findings indicate that miR-5100 exhibits a fold difference 146% greater than that of miR-155-5p, which would be consistent with a negative regulation of miRs by *CASC20* in osteoblast differentiation, at least in terms of simple magnitude of differential miR effects.

At day 20, hsa-mir-1249-3p (FC = -0.589704735, p = 0.013305985) was the only differentially regulated miR, suggesting that the examined timepoint is relatively late for discerning *CASC20* effects, consistent with the hypothesis that *CASC20* modulates early differentiation events during osteogenesis.

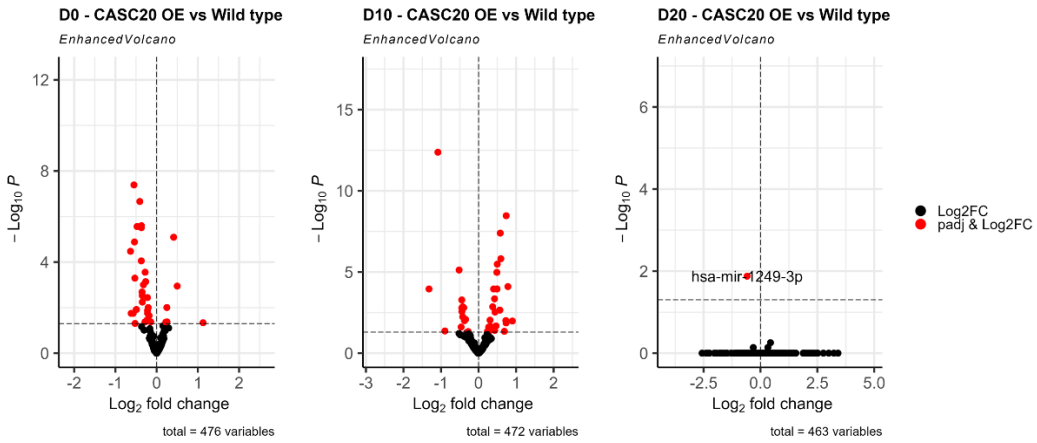


Figure 5.9. Volcano plots comparing *CASC20* OE vs WT P512MSCs at days 0, 10, and 20 during osteoblast differentiation. The volcano plots depict the differential expression analysis of miRs in *CASC20* OE compared to wild type at different time points (days 0, 10, and 20) during osteoblast differentiation. Each dot on the plots represents an individual miR, with the x-axis representing the log2 fold change (log2FC) and the y-axis representing the negative logarithm of the adjusted p-value (−log10p).

Table 5.4. Differential expression analysis of miRs in *CASC20* OE vs wild type at day 10 during osteoblast differentiation.

	log2FoldChange	padj
hsa-mir-5100	-1.31988	0.000111
hsa-mir-1249-3p	-1.08343	4.19E-13
hsa-mir-370-3p	-0.89686	0.042644
hsa-mir-149-5p	-0.5186	7.54E-06
hsa-mir-451a	-0.46407	0.024585
hsa-let-7d-3p	-0.447	0.000524
hsa-mir-22-3p	-0.44626	0.001633
hsa-mir-24-2-5p	-0.44446	0.002697
hsa-let-7d-5p	-0.42161	0.001387
hsa-mir-22-5p	-0.42149	0.005718
hsa-mir-23a-3p	-0.39516	0.001576
hsa-mir-142-5p	-0.37612	0.009891
hsa-mir-27a-3p	-0.34694	0.008314
hsa-mir-125a-3p	-0.28522	0.049806
hsa-mir-615-3p	-0.27672	0.046569
hsa-mir-21-5p	0.233547	0.046569
hsa-mir-140-3p	0.259896	0.046569
hsa-mir-140-5p	0.269718	0.024585
hsa-mir-34a-5p	0.282477	0.043202
hsa-mir-200b-3p	0.288502	0.028436
hsa-mir-148b-3p	0.305557	0.009238
hsa-let-7f-5p	0.329209	0.023473

hsa-mir-146b-5p	0.376974	0.001387
hsa-mir-30d-5p	0.401323	0.000111
hsa-mir-10a-5p	0.426092	0.000446
hsa-mir-181b-5p	0.430844	0.039337
hsa-mir-425-5p	0.43791	0.002942
hsa-mir-362-5p	0.465965	0.020986
hsa-mir-148a-3p	0.484313	0.000111
hsa-mir-30e-5p	0.487176	1.05E-05
hsa-mir-361-3p	0.492182	0.000111
hsa-mir-30a-5p	0.497908	3.29E-06
hsa-mir-29c-5p	0.56236	0.002225
hsa-mir-210-3p	0.56825	0.002225
hsa-mir-182-5p	0.578747	4.00E-08
hsa-mir-10b-5p	0.593101	1.52E-06
hsa-mir-190a-5p	0.687058	0.044951
hsa-mir-103a-3p	0.727464	0.009891
hsa-mir-103b	0.727464	0.009891
hsa-mir-221-5p	0.73346	0.013628
hsa-mir-145-5p	0.736331	3.42E-09
hsa-mir-16-5p	0.781399	7.91E-05
hsa-mir-155-5p	0.89998	0.010408

5.3.8 *In silico* analysis of genes that may interact with miRs downregulated in *CASC20* OE versus wild type P512MSCs at day 10

Next, to explore the potential regulatory role of *CASC20* in osteoblast differentiation through miRs, I investigated the genes targeted by the downregulated miRs, utilising the StarBase v3 database. Subsequently, I established a ranking of these targeted genes based on the number of miRs with which they exhibit interactions. I chose to concentrate on genes exhibiting the most extensive miR interactions, as depicted in Figures 5.10 and 5.11.

The compiled data showed that 53 genes are subject to targeting by ~50% of miRs (8 of the 15 downregulated miRs in in Figure 5.10), including osteogenic and chondrogenic genes LONRF2, ONECUT2, and MAPKs (Figure 5.11). Next, I used RISMed to ascertain the extent to which of the target genes are documented participants in osteo- and chondrogenesis. This *in silico* analysis showed that of the 53 genes, 49 genes had published evidence for a role in osteo- or chondrogenesis (Table 5.5).

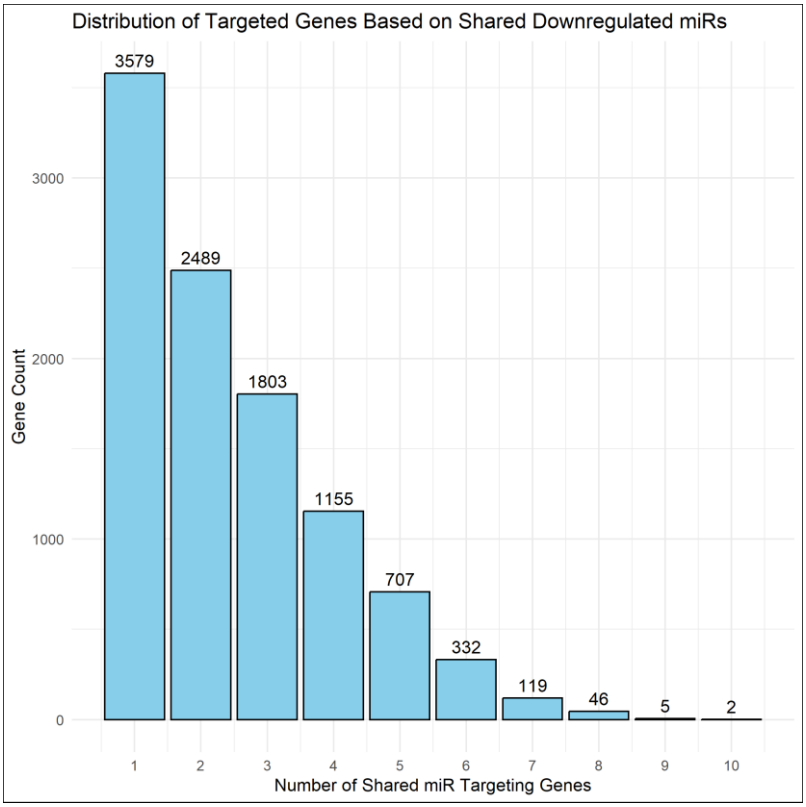


Figure 5.10. Distribution of targeted genes based on shared downregulated miRs. The figure illustrates the number of genes targeted by miRs, ranked by the number of shared miR-targeting genes on the x-axis and the gene count on the y-axis.

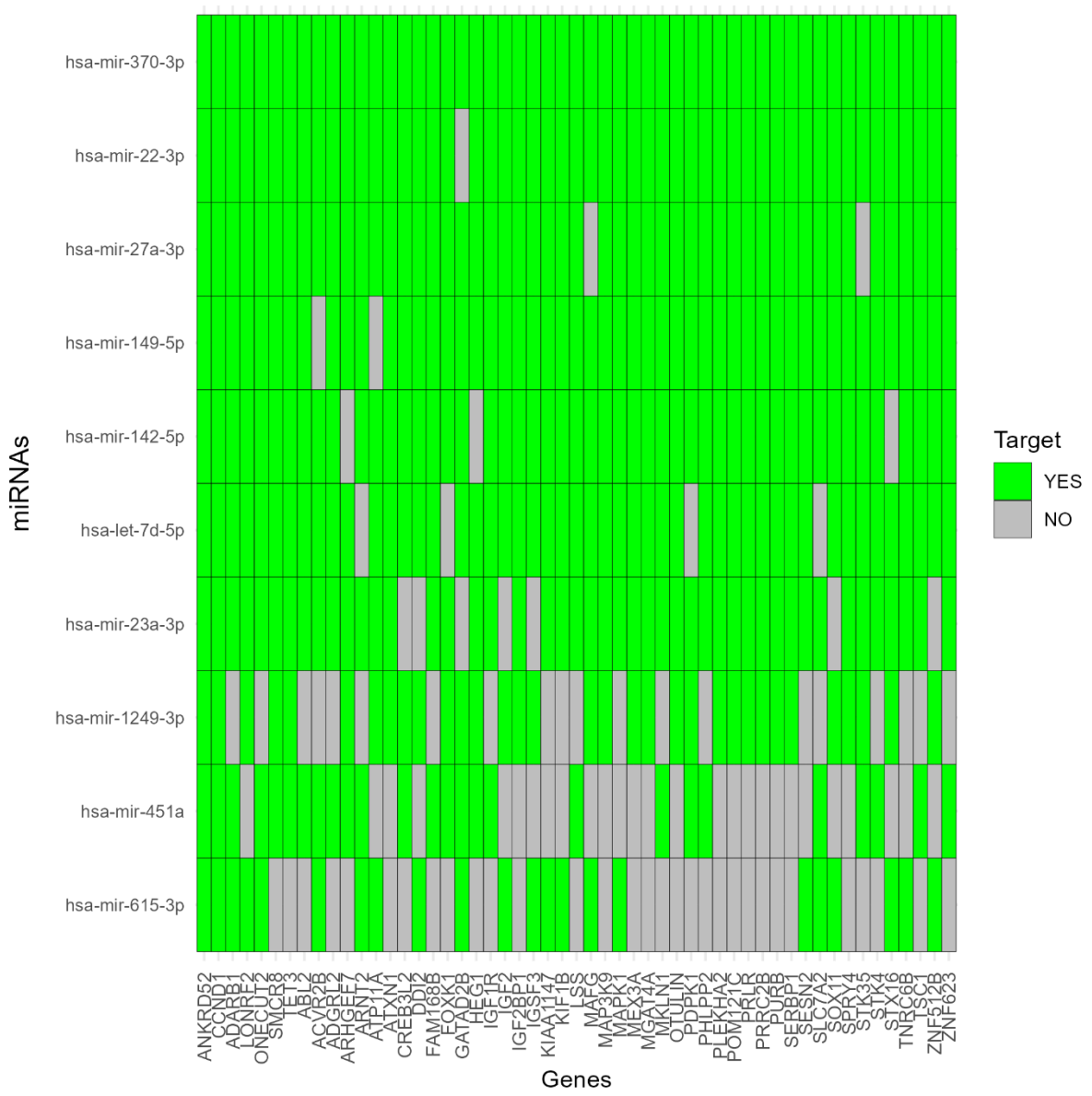


Figure 5.11. Top 53 genes targeted by downregulated miRs. The heatmap showcases the top 53 genes targeted by downregulated miRs. The genes are ranked so that those interacting with the highest number of miRs are positioned towards the left, and the miRs targeting the greatest number of mRNA are placed towards the top of the heatmap.

Table 5.5. Putative gene targets for downregulated miRs and most recent PMIDs for their published role in osteo or chondrogenesis.

Target gene	**Most recent PMID number corresponding with published role in osteo- or chondrogenesis (search date August 04, 2023)**
ANKRD52	34054925
CCND1	37535926
ADARB1	37287534
LONRF2	36888978
ONECUT2	34860830
SMCR8	31847700
TET3	37095518
ABL2	37376546
ACVR2B	37486509
ADGRL2	30340542
ARHGEF7	37290677
ARNT2	32021278
ATP11A	34472226
ATXN1	36810326
CREB3L2	32769431
DDI2	32344880
FAM168B	22543972
FOXK1	36695573
GATAD2B	36442307
HEG1	29382913

IGF1R	37522970
IGF2	37454090
IGF2BP1	36982941
IGSF3	31560140
KIF1B	35234284
LSS	37416814
MAFG	36454022
MAP3K9	35978511
MAPK1	37350000
MEX3A	35433464
MGAT4A	34632546
MKLN1	35138470
OTULIN	37395936
PDPK1	36982289
PHLPP2	36343848
PLEKHA2	32496000
PRLR	37232379
PRRC2B	29109093
PURB	37373006
SERBP1	35892886
SESN2	37336368
SLC7A2	35202090
SOX11	37391758
SPRY4	37424727
STK35	29414823

STK4	37027967
STX16	35119251
TNRC6B	33889172
TSC1	37108467

5.3.9 *In silico* analysis of genes that may interact with miRs upregulated in *CASC20* OE versus wild type P512MSCs

Subsequently, I conducted an analysis of the genes targeted by the upregulated miRs using the same methodological approach, prioritising the list of targeted genes based on the extent of their interactions with miRs (Figure 5.12).

The outcome of this approach showed 53 genes as targets of ~70% (20 of 28) of the upregulated miRs (as shown in Figure 5.12).

The gene targets included LONRF2, ONECUT2, and MAPKs (Figure 5.13) that play a role in osteo and/or chondrogenesis. Next, I applied the RISMed tool to all the Identified target to determine If they had a confirmed published role in these processes. This analysis confirmed that 48 of the 53 target genes had published functional evidence for such a role (Table 5.6).

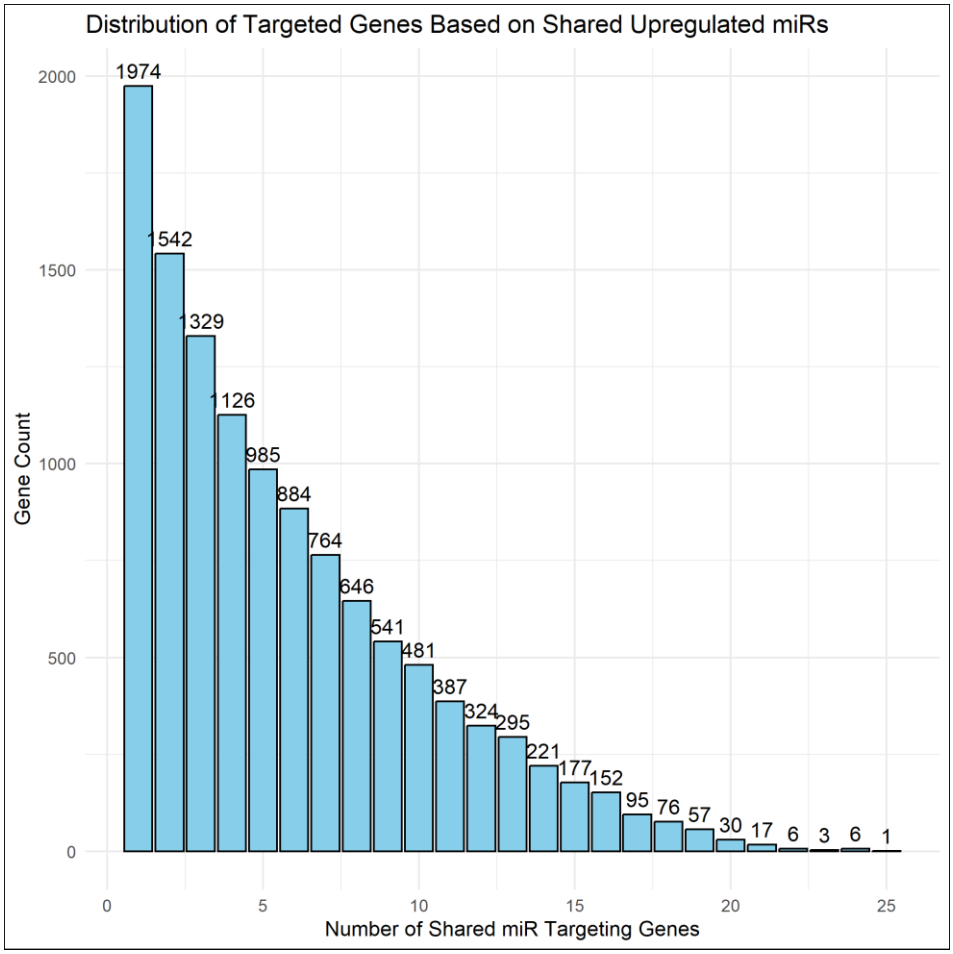


Figure 5.12. Distribution of targeted genes based on shared upregulated miRs. The figure illustrates the number of genes targeted by miRs, ranked by the number of shared miR-targeting genes on the x-axis and the gene count on the y-axis.

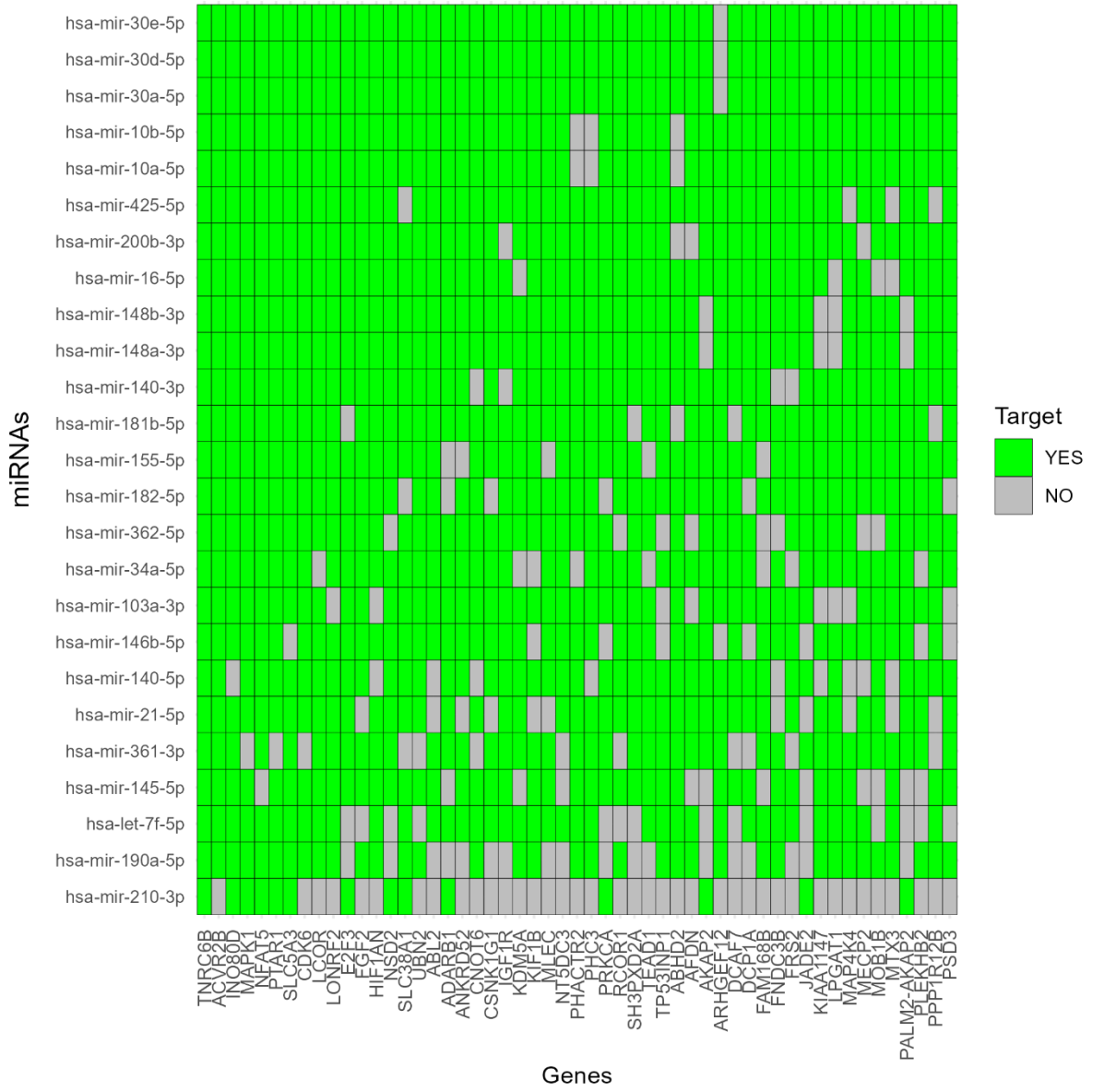


Figure 5.13. Top 53 genes targeted by upregulated miRs. The heatmap showcases the top 53 genes targeted by upregulated miRs. The genes are ranked so that those interacting with the highest number of miRs are positioned towards the left, and the miRs targeting the greatest number of mRNA are placed towards the top.

Table 5.6. Putative gene targets for upregulated miRs and most recent PMIDS for their published role in osteo or chondrogenesis.

Target gene	**Most recent PMID number corresponding with published role in osteo or chondrogenesis (search date August 04, 2023)**
TNRC6B	33889172
ACVR2B	37486509
INO80D	30253751
NSD2	28338204
FNDC3B	34382874
MAPK1	37350000
NFAT5	36836762
PTAR1	32128853
SLC5A3	35899258
CDK6	37534476
LCOR	29663375
LONRF2	36888978
E2F3	37395281
FGF2	37534610
HIF1AN	33619902
SLC38A1	37055385
ABL2	37376546
ADARB1	37287534
ANKRD52	34054925
CNOT6	31797865
CSNK1G1	33416120
IGF1R	37522970
KDM5A	36733232
KIF1B	35234284
MLEC	16440295
PHC3	37468993
PRKCA	37287061
RCOR1	37442513
SH3PXD2A	33143131
TEAD1	37197086
TP53INP1	34718338
ABHD2	37298205
AFDN	35484498
AKAP2	36927779
ARHGEF12	34547282
DCAF7	27880803
DCP1A	36640348
FAM168B	22543972

FRS2	37264620
JADE2	36008159
LPGAT1	37217003
MAP4K4	37312710
MECP2	37340152
MOB1B	35685465
MTX3	23173263
PLEKHB2	22543972
PPP1R12B	28681629
PSD3	31989994

5.3.10 KEGG pathway enrichment analysis of inferred target genes of downregulated miRs at day 10

Next, I applied KEGG to identify key molecular pathways affected by the combined mRNA target list to understand the network of interactions that the downregulated miRs have with their target mRNAs (Figure 5.14). Again, the top hit was “Pathways in Cancer”. The next 4 top pathways affected by the miRs were included endocytosis, focal adhesion, protein processing in endoplasmic reticulum, and Wnt signalling pathway. Other osteo- and chondro related pathways targeted included TGF-beta signalling pathway and mTOR signalling pathway.

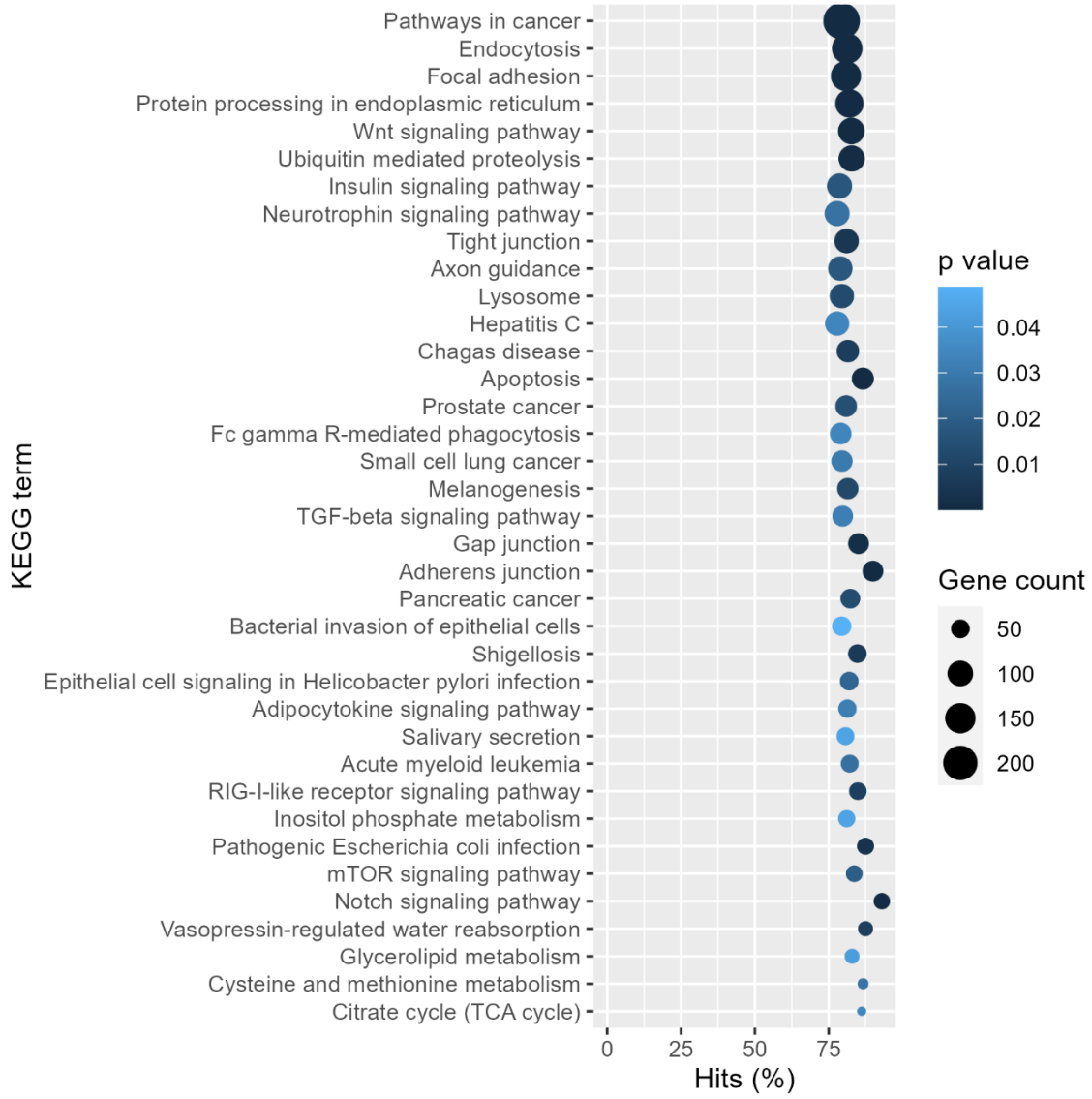


Figure 5.14. Enrichment plot showing KEGG pathways targeted by downregulated miRs. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the KEGG pathway terms. Pathways are

color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

5.3.11 KEGG pathway enrichment analysis of inferred target genes of upregulated miRs at day 10

Following the investigation into combined mRNA targets for the upregulated miRs, I identified the network of the interactions that these miRs have with their target mRNAs. The most prominent pathway identified was once again "Pathways in Cancer (Figure 5.15). The remaining top 5 pathways affected by the miRs included MAPK signalling pathway, endocytosis and focal adhesion, and regulation of actin cytoskeleton. Osteo- and chondro related pathways targeted included the mTOR signalling pathway.

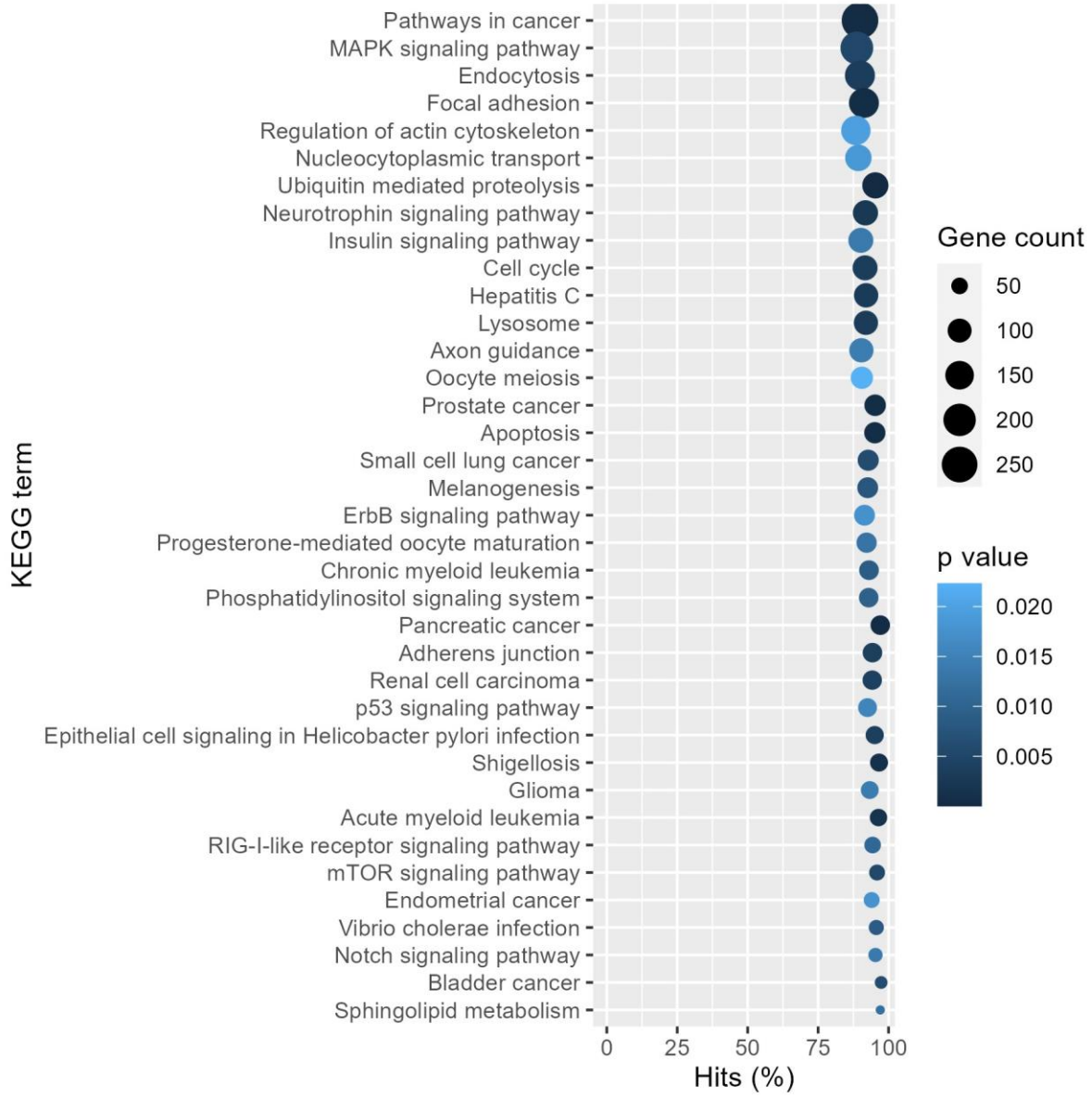


Figure 5.15. Enrichment plot showing KEGG pathways targeted by upregulated miRs. The data has been processed and sorted to focus on the top pathways based on their significance

scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the KEGG pathway terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

5.3.12 KEGG pathway enrichment analysis of individual downregulated miRs

To identify how specific downregulated miRs affect molecular pathways, I plotted their targeted pathways in Figure 5.16. The analysis showed that the most targeted pathways included adherens junction, pathways in cancer, endocytosis, focal adhesion, and Wnt signalling (log10FDR < -1.30103). Other pathways involved in bone and cartilage development also emerged, including MAPK signalling, osteoclast differentiation, mTOR, and TGF-beta signalling pathways. The data showed that miR-22-3p displayed the lowest -log10(FDR) values for MAPK, Wnt, and mTOR signalling pathways. MiR-142-5p displayed the lowest -log10(FDR) value for TGF-beta signalling. Furthermore, miR-149-5p the lowest -log10(FDR) value for osteoclast differentiation. GO analysis suggested that *CASC20* could be involved in metabolic processes, catalytic activity, and kinase activity (Appendix 5.4).

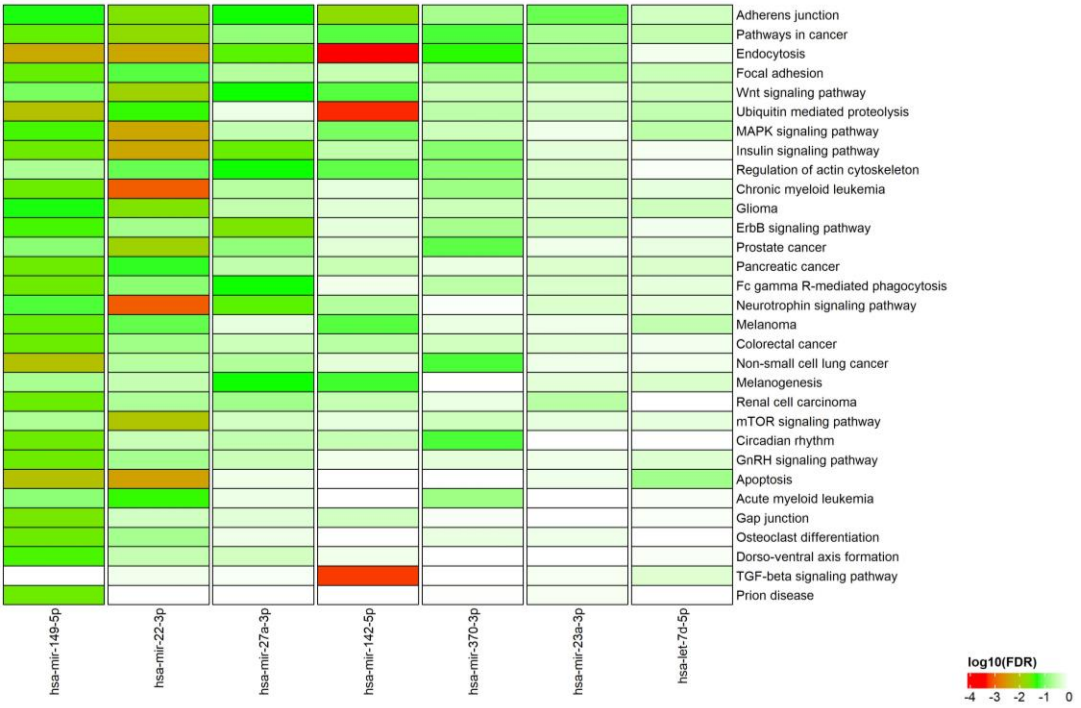


Figure 5.16. Heatmap displaying the GO/KEGG terms targeted by the individual miRs downregulated on day 10 (*CASC20* OE versus Wild type). The terms in the heatmap are arranged in a ranking order, with the GO terms targeted by the highest number of miRs positioned on the top, and the miRs targeting the greatest number of GO terms positioned at the left of the heatmap.

5.3.13 Comparison between our experimental dataset and published osteodifferentiation dataset from Chapter 2

For the published osteogenesis dataset, of the 231 miRs that were differentially expressed miRs at D10 versus D0, 11 miRs were present in our experimental dataset. Pearson correlation analysis demonstrated no meaningful correlation between the miRs between the datasets (r=-0.12, P=0.7205). For the 19 *CASC20*-interacting miRs that are differentially expressed at D10 versus D0 in the in-silico dataset, 3 miRs are present in our experimental dataset. The Pearson correlation between the shared miRs across the 2 datasets was 0.935, however the number of shared miRs was insufficient for statistical significance (P=0.229).

5.4 DISCUSSION

The miR sequencing demonstrated distinct separation among the different time points (days 0, 10, and 20) as revealed by the plot PCA and the expression heatmaps. Pairwise analysis to compare *CASC20* OE with wild type demonstrated that *CASC20* overexpression affected the expression of 34, 43, and 1 miR, respectively at days 0, 10 and 20. Further analysis of day 10 showed that the top genes targeted by the downregulated and upregulated miRs played a role in osteo- and chondrogenesis. Ten of these mRNA targets were found to be shared between the downregulated and upregulated miRs. These targets were *ANKRD52, ADARB1, LONRF2, ABL2, ACVR2B, FAM168B, IGF1R, KIF1B, MAPK1,* and *TNRC6B*. This highlights the importance of quantifying mRNA expression levels through mRNA-Sequencing.

The data shows that top genes KEGG pathways targeted by the downregulated miRs included pathways in cancer, Wnt signalling pathway, TGF-beta signalling pathway and mTOR signalling pathway. KEGG pathways targeted by the upregulated miRs included pathways in cancer, MAPK signalling and mTOR signalling pathway. Analysis of the pathways targeted by the individual miRs showed that miR-22-3p has the lowest FDR value for MAPK, Wnt, and mTOR signalling pathways. MiR-142-5p was shown to have the lowest FDR value for TGF-beta signalling.

Comparison between the experimental dataset and the published osteoblast differentiation dataset showed limited to no overlapping, evidenced by the correlation analysis. It is not unusual for *in silico* miR-target predictions to yield a larger set of candidates than what is experimentally confirmed (567, 568). The *in-silico* prediction and analysis of published dataset provided a starting point for further investigation (569). The miR-Seq analysis and results underscored the importance of empirical detection and validation of *CASC20*-miR interactions.

The data analysis showed various internal and external consistencies. Internal consistencies includes PCA separation, which indicated that miR transcriptional variations over time are captured in the data. Moreover the consistent targeting of specific genes by multiple miRs, both downregulated and upregulated, indicates internal consistency in the potential regulatory mechanisms mediated by these miRs. External consistencies included the identification of pathways in cancer as prominently affected by *CASC20*-regulated miRs, as this aligns with existing biological knowledge (570). The identification of pathways related to osteogenesis and chondrogenesis, such as Wnt, TGF-beta, and MAPK signalling pathways is consistent with the expected impact of the miRs on bone and cartilage development (571). As these pathways are targeted by up- and/or down-regulated miRs, it highlights the importance of quantifying the cumulative impact of *CASC20* differential expression on mRNA and KEGG pathways. Moreover, the overlap of targeted genes with documented functions in osteogenesis and chondrogenesis, as evidenced by their associated PubMed IDs, supports external consistency with established literature.

The differential expression of specific miRs at days 0 and 10 suggests that *CASC20* might play a role in early events during osteogenesis, potentially influencing cell differentiation. The finding that miRs with higher fold changes, such as miR-5100, are downregulated in *CASC20* OE suggest a potential suppressive effect on these miRs. The identified genes targeted by both downregulated and upregulated miRs play roles in osteo- and chondrogenesis. This suggest that *CASC20* may influence key biological processes through multifaceted miR-mediated interactions. The findings suggest that *CASC20* may affect the expression of miRs such as miR-22-3p and miR-142-5p to regulate MAPK and TGF-beta signalling pathways, respectively.

It is important to acknowledge the limitations of the study. The experimental setup involves *in vitro* tissue culture conditions, which may not fully replicate *in vivo* data, as 3D architecture and systemic influences are absent (572, 573). The study compared *CASC20* OE with wild type cells and may not accurately replicate the effects of endogenous *CASC20*. The analysis presumes that differential miR expression directly translates into modification of target gene expression. However, *CASC20* may affect miR function without affecting miR expression or detection (574, 575).

In summary, it is important to approach these results with caution, as they represent preliminary findings. To confirm these insights, we are conducting mRNA sequencing to gain a better understanding of the interactome. Even in cases where sponging is the proposed mechanism, over-expression of *CASC20* may not necessarily lead to a decrease in measured miR levels, as sequencing could detect inactive miRs (576). Therefore, additional biological testing methods may be necessary to validate this hypothesis.

Chapter 6 –Effects of *CASC20* overexpression on chondrogenic differentiation

6.1 INTRODUCTION

In this chapter, the primary aim is to explore the potential regulatory role of *CASC20* in chondrogenic differentiation. To achieve this, a lentiviral overexpression approach was employed in hMADs, P512MSCs, and ASC52teloSOX9, *SOX9*-overexpressing hTERT immortalised adipose-derived mesenchymal stem cells.

SOX9 is a transcription factor that plays a critical role in the development and differentiation of various tissues, particularly in skeletal development (577). It is a member of the *SOX* (*SRY*-related HMG box) family of transcription factor family and is known for its involvement in chondrogenesis, the process of cartilage formation (578). *SOX9* regulates the expression of genes that are essential for the differentiation and maintenance of chondrocytes, the cells responsible for producing and maintaining cartilage tissue (579). *SOX9* is considered a master regulator of chondrogenesis, as it controls key steps in the differentiation of mesenchymal cells into chondrocytes (580). Dysregulation of *SOX9* has been linked to skeletal disorders, such as campomelic dysplasia, highlighting its importance in skeletal development and function (581).

Glycosaminoglycans (GAGs) are a class of long, unbranched polysaccharides that are an essential component of the extracellular matrix (ECM) (582). Examples of GAGs include chondroitin sulphate and hyaluronic acid (583, 584). GAGs help cartilage maintain its unique characteristics, such as its ability to distribute and absorb mechanical forces (585). GAGs provide hydration and flexibility to the cartilage matrix, enhancing its resistance to compression (586). Moreover, GAGs act as signalling molecules, regulating cell adhesion, migration, and proliferation during chondrogenic differentiation (587). Understanding the fluctuations in GAG content and metabolism is crucial for maintaining healthy cartilage function and comprehending the mechanisms underlying chondrogenesis (587).

By manipulating the expression levels of *CASC20*, I examined its impact on the process of chondrogenic differentiation. To quantitate the biological endpoint of chondrogenic differentiation, a widely-used technique called the GAG assay was implemented (588). This assay uses a molecule termed 1,9-dimethyl methylene blue (DMMB) that can selectively bind with GAGs, allowing for their precise quantitation (588).

In addition to the biological endpoint assessment through GAG assay, I employed RT-qPCR to quantitate the expression of key chondrogenic markers, including *ACAN*, *COL2A1*, and *COMP* (550). To further assess the phenotypic characteristics of the differentiating cells, RT-qPCR was performed to quantitate the expression of key osteogenic markers, including *RUNX2*, *OSX*, *ALP*, and *COL1A1* (589) to determine whether *CASC20* over-expression affected the

differential expression of osteogenic versus chondrogenic genes that might suggest an action as a molecular switch between these processes (589).

To assess the influence of *CASC20* overexpression on chondrocyte differentiation, 3 different sources of MSCs (hMAD, P512MSC and ASC52teloSox9) were differentiated into chondrocytes using +/- TGFβ3, and samples were collected at day 0, 7, and 14. TGFβ3, or transforming growth factor beta 3, is a key signalling molecule involved in chondrogenic differentiation (590). TGFβ3 belongs to the TGF-beta superfamily and plays a crucial role in promoting and maintaining the chondrocyte phenotype (591). It stimulates the production of ECM components, such as collagen and proteoglycans, which are essential for cartilage formation (592).

The data showed that *CASC20* overexpression led to a decrease in chondrogenic differentiation, as evidenced by GAG assay and RT-qPCR at day 7 and 14. As a working hypothesis, I proposed that *CASC20* upregulation may divert chondrogenesis towards osteogenesis and that variation within the *CASC20* may differentially regulate miR levels in osteogenesis and chondrogenesis. The results from sequencing data will help confirm or refute this mechanistic model.

6.2 MATERIALS and METHODS

6.2.1 Cell culture

hMADs, P512MSCs and ASC52teloSOX9 were cultured as previously described in Chapters 2 and 4.

6.2.2 Chondrogenic differentiation

Cells were seeded in T75 flasks for 24 hours in growth media and then transduced with lentivirus for 24 hours. After that, the cells were washed with PBS and incubated with fresh growth media for 24 hours. The cells were trypsinised, and growth media was added to the cell-trypsin mixture. Subsequently, the cells were transferred to a falcon tube and centrifuged. Following this, the cells were resuspended in chondrogenic media at a concentration of 50,000 cells per 150µL. The differentiation media was prepared as DMEM (4.5g/L glucose), 1%glutamine, 1% P/S, 10ng/mL TGFβ3 (Proteintech, HZ-1090 - Rosemont, IL), 100nM dexamethasone, 50µg/ml ascorbic acid, 40µg/ml proline (Sigma-Aldrich, 81709), and 1X ITS+L premix (insulin, transferrin, selenium, linoleic acid) (Thermo Fisher Scientific, 51500056). After this, 150µl of cells were transferred into wells of UV-irradiated V-bottomed 96-well plates (Greiner Bio-One, 651101 – Kremsmünster, Austria). The plates were centrifuged, and the media was changed every three days until the designated time points for further analysis. To collect the cells, the differentiation media was removed, and the pellets were washed with PBS. The pellets were transferred to tubes, snap-frozen in dry ice and stored at -80°C. To test the differential effect of TGFβ3, the cells were differentiated into chondrocytes using +/-TGFβ3.

6.2.3 RNA isolation and RT-qPCR

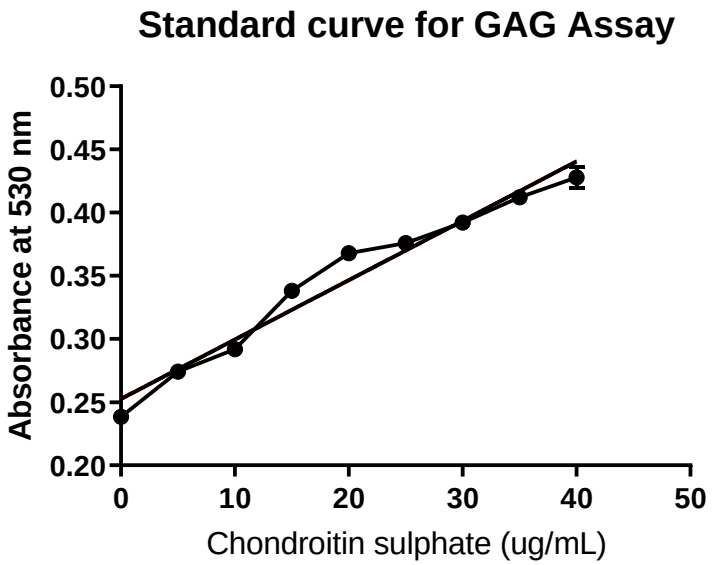
Total RNA was isolated using the miRNeasy Micro Kit, as previously described in Chapter 2. Reverse transcription of RNA into cDNA and RT-qPCR were conducted, as previously described in Chapter 3.

Gene	**Human qPCR Primer Sequence (5' -> 3')**
GUSB	FW CTGTCACCAAGAGCCAGTTCCT
	REV GGTTGAAGTCCTTCACCAGCAG
CASC20	FW TCATATGGATTTCAAGCTGGGT
	REV TCCCAGTCTTCTGCATCACTTC
RUNX2	FW GGTTAATCTCCGCAGGTCACT
	REV CACTGTGCTGAAGAGGCT
COL1A1	FW ACCGCCCTCCTGACGCAC
	REV GCAGACGCAGATCCGGCAG
ALP	FW GCAGACGCAGATCCGGCAG
	REV CCTGGCTTTCTCGTCACTCTCA
ACAN	FW CAGGCTATGAGCAGTGTGACGC
	REV GCTGCTGTCCTTGTCACCCACG
COL2A1	FW GCCGAGGTGATAGTGTGGTT

	REV AACGGGGATGGCCTTGTATG
COMP	FW GGAGATGCTTGTGACAGCGATC
	REV TGAGTCCTCCTGGGCACTGTTA

6.2.4 Cartilage digestion and GAG Assay

To perform cartilage digestion, the phosphate buffer with was prepared by combining solution A (0.1M NaH_2PO_4) (Sigma-Aldrich, S3139) with solution B (0.1M Na_2HPO_4) (Sigma-Aldrich, 567547) in the ratio of 137/63 respectively. Next, papain solution was prepared as 25µg/mL papain (Sigma-Aldrich, 1071440025), 7.8µg/mL cysteine-HCl (Sigma-Aldrich, C7880), and 19µg/mL EDTA (disodium salt) (Sigma-Aldrich, E5134) in phosphate buffer. After that, 70µL of phosphate buffer and 40µL of papain solution were added to each tube containing cartilage pellets. The tubes were incubated at 65°C for 4 hours. To aid the digestion process, the tubes were vortexed and centrifuged every hour. For the GAG assay, the DMMB solution was prepared as 2.6g/mL DMMB (Sigma-Aldrich, 341088), 3.4g/mL NaCl, and 9.5mM HCl in dH_2O. GAG content was measured by combining 40µL of sample with 250µL of DMMB solution. The resulting mixture was read using a plate reader at a wavelength of 530nm. To determine GAG concentrations, chondroitin sulphate (Sigma-Aldrich, C4384) served as the standard reference.



6.2.5 Statistics

Statistical analyses were conducted as previously described in Chapter 2.

6.3 RESULTS

6.2.1 *CASC20* negatively regulates chondrogenic differentiation in ASC52teloSOX9 and P512MSCs

Initially, I confirmed the stable overexpression of *CASC20* in the *CASC20*-overexpressing cells compared to wild type (FC = 7989.505, p = 0.0002) as shown in Figure 6.1, as indicated by the enhanced *CASC20* upregulation observed. This stable overexpression of *CASC20* allowed further investigation of *CASC20* differential expression on chondrocyte differentiation.

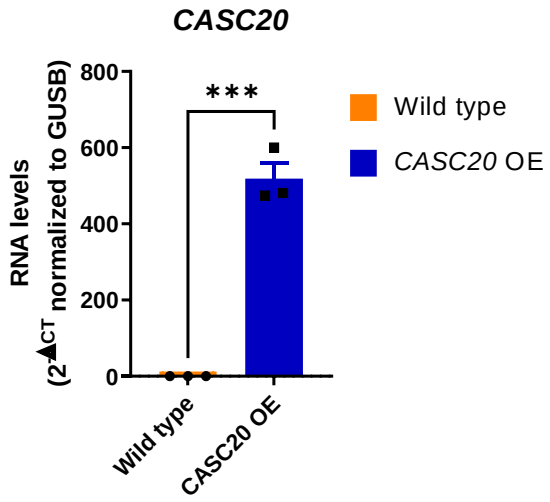


Figure 6.1. RT-qPCR reveals enhanced *CASC20* upregulation in *CASC20* overexpressing ASC52teloSOX9s compared to control. Analyses are unpaired t test; data is plotted as mean ±SEM; ***P<0.001; N = 3, technical replicates.

I observed an increase in GAG synthesis in wild type ASC52teloSOX9 cells undergoing chondrocyte differentiation at day 14 compared to day 0, both in the absence and presence of TGFβ3 ($p < 0.0001$), as indicated by GAG assay. Furthermore, in presence of TGFβ3, *CASC20* overexpression resulted in a decrease in GAG deposition at day 14 compared to wild type (FC = 0.58, p = 0.0003), as shown in Figure 6.2.

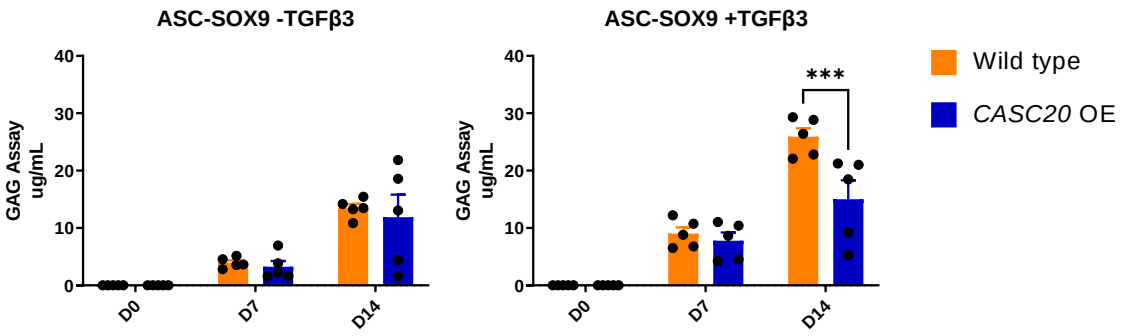


Figure 6.2. Effect of *CASC20* overexpression on chondrocyte differentiation of ASC52telosox9s. GAG assay was used to stain quantify GAG deposition at days 0, 7, and 14. Analyses are two-way ANOVA with Sidak's multiple comparisons test; data is plotted as mean ±SEM; ***P<0.001; N = 5, technical replicates.

I observed a significant increase in GAG deposition in wild type P512MSCs undergoing chondrocyte differentiation at day 14 compared to day 0, both in the absence and presence of TGFβ3 ($p < 0.0001$), as indicated by GAG assay. In the absence of TGFβ3, *CASC20* overexpression resulted in decreased GAG deposition at day 14 compared to wild type (FC = 0.51, $p = 0.0007$). Furthermore, in the presence of TGFβ3, *CASC20* overexpression led to decreased GAG deposition at day 7 compared to wild type (FC = 0.079, $p < 0.0001$), as shown in Figure 6.3. However, by day 14 this difference was lost.

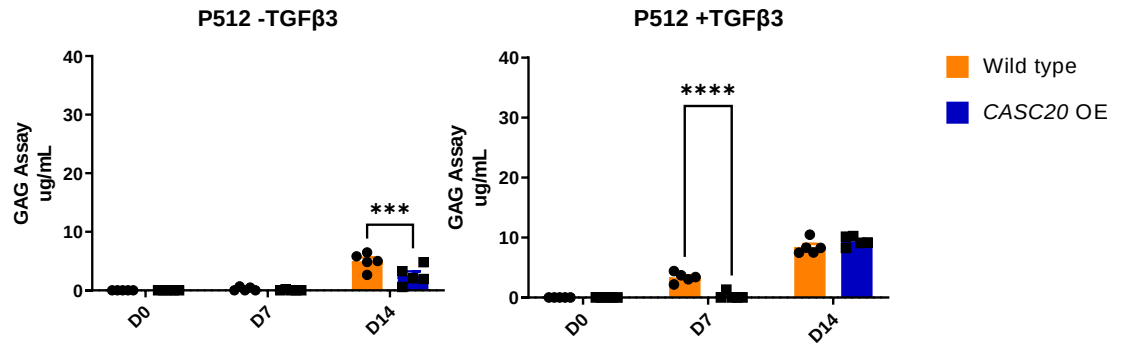


Figure 6.3. Effect of *CASC20* overexpression on chondrocyte differentiation of P512MSCs. GAG assay was used to stain quantify GAG deposition at days 0, 7, and 14. Analyses are two-way ANOVA with Sidak's multiple comparisons test; data is plotted as mean ±SEM; ***P<0.001, ****P<0.0001; N = 5, technical replicates.

I performed chondrocyte differentiation experiments for over 14 days using *CASC20*-overexpressing hMADs, P512MSCs cells, and ASC52teloSOX9. The hMADs did not exhibit chondrocyte differentiation as evidenced by the absence of GAG synthesis on day 14 (Figure

6.4). This lack of differentiation could be attributed to intrinsic properties of the hMADs or the high passage number of the cells during the experiment.

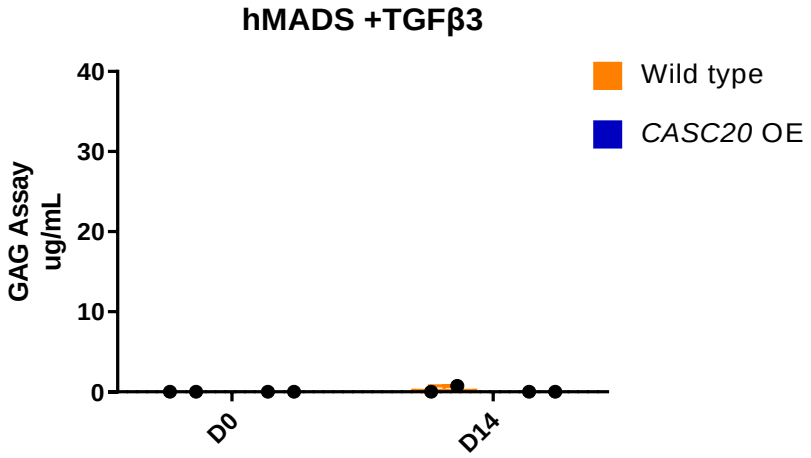


Figure 6.4. Effect of *CASC20* overexpression on chondrocyte differentiation of hMADs. GAG assay was used to quantify GAG at days 0 and 14. Data is plotted as mean ±SEM. N = 2, technical replicates.

Taken together, these findings indicate that *CASC20* exerts a negative regulatory effect on chondrogenic differentiation in both ASC52teloSOX9 and P512MSCs. ASC52teloSOX9 cells exhibited higher levels of GAG production compared to P512MSCs. This difference in GAG levels is likely attributed to the overexpression of *SOX9* in ASC52teloSOX9 cells, which is known to play a crucial role in chondrogenesis. These observations suggest that *CASC20* may modulate chondrogenic differentiation in a cell-type-specific manner, and the presence of *SOX9* could potentially influence the response to *CASC20* overexpression.

The pro-chondrogenic effect of TGFβ3 is evident through the observed increase in GAG synthesis over time in both wild type ASC52teloSOX9 and P512MSCs during chondrogenic differentiation.

In both ASC52teloSOX and in P512MSCs, the presence of TGFβ3 appears to be modulated by *CASC20*, resulting in a decrease in GAG deposition in *CASC20* OE versus wild-type cells at day 14 and 7, respectively. In the absence of TGFβ3, the temporal effect of *CASC20* becomes evident only at day 14, as at day 7, the cells have produced a negligible amount of GAG. These results collectively emphasise the significance of both the differentiation media and the specific time points examined in understanding the interplay between *CASC20* and TGFβ3 in chondrogenesis.

6.2.2 *CASC20* modulates the expression of chondrogenic genes during chondrogenesis

I examined the expression changes of chondrogenic genes (*ACAN*, *COL2A1* and *COMP*) during chondrogenic differentiation using ASC52teloSOX9 cells. I investigated ASCs as they showed promising results in the GAG assay. The cells were differentiated into chondrocytes and collected at days 0, 7, and 14 for RT-qPCR analysis (see Figure 6.5).

In wild type cells, the expression of *ACAN* and *COL2A1* was upregulated at day 7 and 14 compared to baseline, both in presence and absence of TGFβ3. However, *CASC20* overexpression in cells resulted in reduced expression of *ACAN* (FC = 0.53, p = 0.0033) and *COL2A1* (FC = 0.56, p < 0.0001) compared to wild type cells in the absence of TGFβ3 at day 14, but not in the presence of TGFβ3. A trend towards downregulation of *COL2A1* was observed in the presence of TGFβ3 at both day 7 (FC = 0.13) and day 14 (FC = 0.82), but not for *ACAN*.

These findings provide evidence to suggest that *CASC20* may act as a negative regulator of chondrogenesis, modulating the expression of *ACAN* and *COL2A1* in the absence of TGFβ3. In the presence of TGFβ3, the influence of *CASC20* on *ACAN* and COL2A1 expression appears to be mitigated by the chondrogenic effect of TGFβ3.

In contrast, *COMP* exhibited different behaviour compared to the other chondrogenic genes examined. In wild type cells differentiated without TGFβ3, *COMP* showed no significant upregulation during chondrogenesis. However, in wild type cells differentiated with TGFβ3, *COMP* was significantly upregulated at both day 7 (p = 0.0005) and day 14 (p = 0.0394) compared to baseline. *CASC20* overexpression resulted in relatively reduced *COMP* expression at day 7 (FC = 0.18, p = 0.0026), and at day 14 (FC = 0.67), although the latter was not statistically significant. These findings suggest that the expression of *COMP* during chondrogenesis may be dependent on the presence of TGFβ3. Additionally, the results indicate that *CASC20* may act as a negative regulator of *COMP*, exerting wild type over the chondrogenic process in the presence of TGFβ3.

Collectively, these findings indicate that *CASC20* acts as a negative regulator of chondrogenesis that influences the expression of key chondrogenic genes.

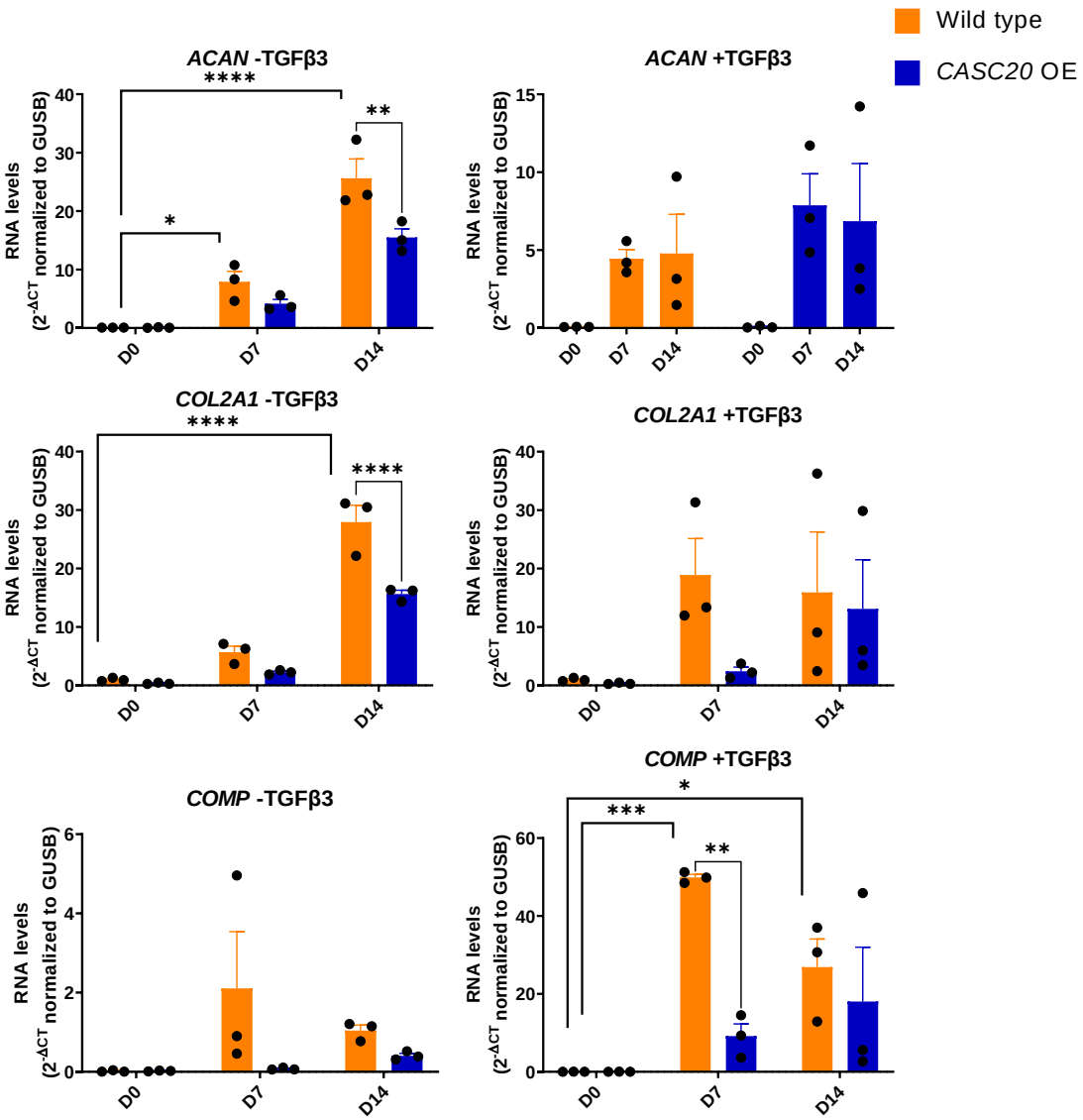


Figure 6.5. Expression analysis of chondrogenic (*ACAN*, *COL2A1*, and *COMP*) markers in wild type and *CASC20*-overexpressing ASC52teloSOX9 cells at days 0, 7 and 14. The analysis included three replicates. The data are presented as the mean ± SEM. Analyses are two-way ANOVA with Sidak's multiple comparisons test; data is plotted as mean ±SEM.*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001; N = 3, technical replicates.

6.2.3 *CASC20* does not substantially modulate the expression of osteogenic genes during chondrogenesis in ASC52telSOX9 cells

To investigate the potential role of *CASC20* in diverting chondrogenesis to osteogenesis, I examined the expression changes of osteogenic genes (*RUNX2*, *ALP*, and *COL1A1*) during chondrogenic differentiation using ASC52teloSOX9 cells. The cells were differentiated into chondrocytes and collected at days 0, 7, and 14 for RT-qPCR analysis. Although statistical significance was limited, trends were observed in the RT-qPCR data (see Figure 6.6).

I identified no significant changes in RUNX2 throughout the chondrogenesis time course in either the absence or presence of TGFβ3 and between the *CASC20* OE versus wild type cells. For ALP, I saw no significant change in expression throughout the time course in the presence or absence of TGFβ3. In contrast, in the *CASC20* overexpressing cells I found a significant Increase in ALP expression at day 14 in the absence of TGFβ3 , (FC = 7.2, p = 0.0217). In the presence of TGFβ, a similar trend towards increased ALP expression was observed at both day 7 (FC=10) and 14 (FC = 1.4) though this did not reach statistical significance versus wild type cells. In contrast, for *COL1A1*, although this marker was not significantly differentially expressed in wild type cells between days 0, 7, and 14 during chondrogenesis, for the *CASC20* over-expressing cells I found a relative downregulation at day 7 versus the wild type cells in the presence of TGFβ3.

Taken together, these data suggest that *CASC20* OE in SOX9 primed cells being driven down a chondrogenic lineage does not substantially change lineage direction towards osteogenesis. It is unclear whether a more convincing pro-osteogenic effect might occur in non-SOX9 primed MSCs.

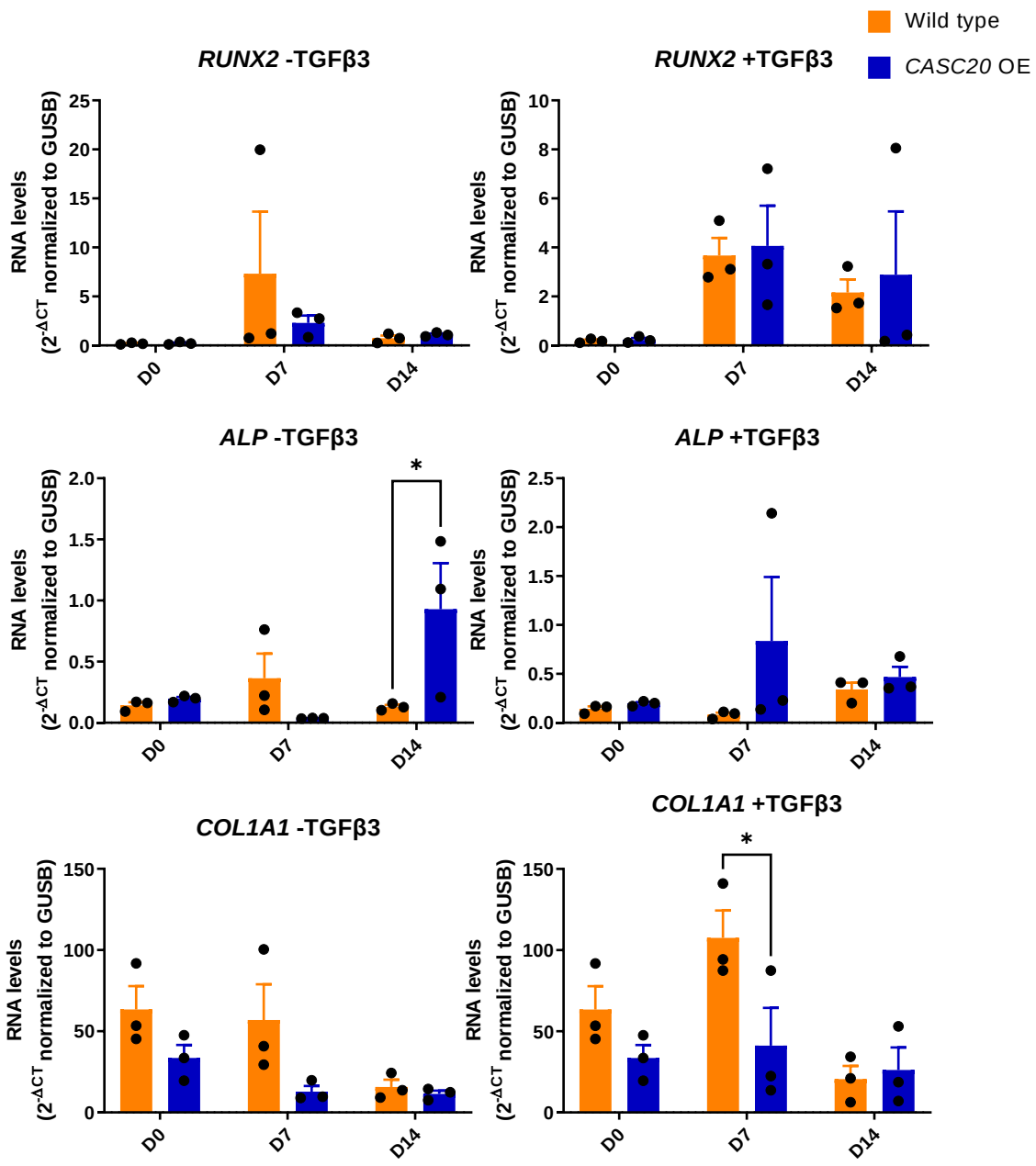


Figure 6.6. Expression analysis of osteogenic (*RUNX2*, *ALP*, and *COL1A1*) markers in control and *CASC20*-overexpressing ASC52teloSOX9 cells at days 0, 7 and 10. The analysis included three replicates. The data are presented as the mean ± SEM. Analyses are two-way ANOVA with Sidak's multiple comparisons test; data is plotted as mean ±SEM; *P<0.05; N = 3, technical replicates.

6.4 DISCUSSION

In this study, I investigated the effects of *CASC20* overexpression on chondrocyte differentiation using hMADs, P512MSCS, and ASC52teloSOX9. Interestingly, hMADs did not undergo chondrocyte differentiation, as evidenced by the absence of GAG synthesis at day 14. This lack of differentiation could be attributed to intrinsic properties of the hMADs or the high passage number of the cells during the experiment. My focus therefore shifted to ASC52teloSOX9 and P512MSCs cells, which did exhibit chondrocyte differentiation.

In both ASC52teloSOX9 and P512MSCs cells, I observed a significant increase in GAG synthesis in wild type cells undergoing chondrocyte differentiation at day 14 compared to day 0, both in the absence and presence of TGFβ3. This finding confirms the successful chondrogenic differentiation in the experimental setup.

CASC20 inhibited GAG deposition in both ASC52teloSOX9 and P512MSCs cells at specific time points and culture conditions. These observations suggest that *CASC20* may negatively regulates chondrogenic differentiation in both ASC52teloSOX9 and P512MSCs cells.

I explored the expression of chondrogenic genes during chondrogenic differentiation. The data suggests that *CASC20* may act as a negative regulator of chondrogenesis by modulating the expression of *ACAN* and *COL2A1* in the absence of TGFβ3. In the presence of TGFβ3, the influence of *CASC20* on *ACAN* and *COL2A1* expression appears to be overridden, as indicated by the lack of significant changes in their expression levels.

Moreover, the upregulation of *COMP* in the presence of TGFβ3 and its negative modulation by *CASC20* overexpression indicate a potential role for *CASC20* in controlling *COMP* expression. *COMP* is upregulated in later stages of chondrogenic differentiation *in vitro* (at days 14 and 21) using the full chondrogenic media (593). This means that in the absence of TGFβ3, more time may be needed to induce the upregulation of *COMP* expression.

I then explored the expression of osteogenic genes (*RUNX2*, *ALP*, *COL1A1*) during chondrogenesis. This data did not show any consistent effects of *CASC20* over-expression on changing the expression phenotype in the ASC52teloSOX9 cells. However, this experiment was not conducted in non-SOX9 primed cells and thus a potential role in diverting chondrogenesis to osteogenesis cannot be excluded.

Taken together, the findings provided evidence that *CASC20* acts as a negative regulator of chondrogenic differentiation in ASC52teloSOX9 and P512MSCs cells.

It is important to note that this study has some limitations. The sample sizes were relatively small, which may have affected the statistical power of the analyses. Further studies with larger sample sizes are needed to validate and strengthen the findings. Additionally, the mechanisms by which *CASC20* exerts its regulatory effects on chondrogenic differentiation remain to be elucidated. Future studies should investigate the specific molecular pathways

and signalling networks involved in the *CASC20*-mediated regulation of chondrogenesis. Consequently, the next chapter of this thesis will focus on miR-Seq analysis of P512MSCS during chondrocyte differentiation to further investigate the regulatory mechanisms of *CASC20* in chondrogenesis in non-SOX9 primed MSCs.

Chapter 7 – Effects of *CASC20* overexpression on miR profile during chondrogenesis

7.1 INTRODUCTION

Chondrogenesis is the first step of cartilage formation, and also for bone formation during endochondral ossification (594). Chondrogenic signals stimulate mesenchymal stem cells (MSCs) to transition into proliferative pre-chondrocytes (595). These pre-chondrocytes play a role in constructing the ECM, a foundational scaffold essential for subsequent mineralised cartilage tissue formation (595). As the matrix matures, hydroxyapatite deposition occurs, coinciding with the maturation of pre-chondrocytes into fully functional chondrocytes (596). Precise genetic orchestration drives this dynamic cellular differentiation (594).

Although the biological characterisation of chondrogenesis has been extensively studied (597), transcriptional analysis that concentrates on MSCs that overexpress *CASC20* has not yet been published. The main goal of the studies described in this chapter was to examine miR expression profile associated with chondrocyte differentiation with a focus on *CASC20* modulation of their expression. To do this, I used chondrogenic media and TGFB3 to chondro-differentiate wild type and *CASC20* overexpressing P512MSCs. The cells were collected on days 0 and 7 for miR-Seq. The selection of the timepoints (days 0 and 7) was informed by previous data collected in Chapter 6, which demonstrated the significant impact of *CASC20* on chondrocyte differentiation by day 7.

I identified distinctive miR expression patterns using unsupervised clustering analysis and identified the mRNA targets of differentially expressed miRs. Additionally, I functionally annotated the differentially expressed miRs to identify linked pathways using KEGG and GO analyses.

7.2 MATERIAL and METHODS

7.2.1 Chondrogenic differentiation and RNA isolation

P512MSCs were cultured and differentiated into chondrocytes using +TGFβ3 chondrogenic media, as previously described in Chapter 6.

7.2.2 RNA quality and sequencing

RNA samples were sent to Novogene for QC and sequencing, as previously described in Chapter 5.

7.2.3 RNA-Sequencing analysis

RNA-Sequencing analysis was conducted as in Chapter 5.

7.2.4 Pathway enrichment analysis

Pathway enrichment analysis was conducted as in Chapter 5.

Correlation analysis was conducted using corrplot (version 0.92) to illustrate the relationships between the parameters within Starbase v3 and the degradation of targeted mRNA.

7.3 RESULTS

7.3.1 *CASC20* differential expression affects global miR expression profile

Exploration of genome-wide miR expression profile was conducted by PCA (Figure 7.1A). The first principal component accounted for 90% of the variability, while the second component explained 4% of the variance. The PCA analysis demonstrated distinct segregation among the different time points (days 0 and 7) with no overlapping, confirming temporal miR transcriptional variation. This initial examination of the data also indicated only limited variation between treatment conditions that was greater at day 0 versus day 7. This is also shown in the sample-sample distance heatmap (Fig 7.1B). However, the gene level heatmap analysis (Figure 7.1C) did show more separation between time points and treatment conditions. The data displayed in Figure 7.1 confirmed the success of the experiment in terms of the progression of miR expression variation during induced human chondrogenesis and the modulation of miR expression over this time course by *CASC20*.

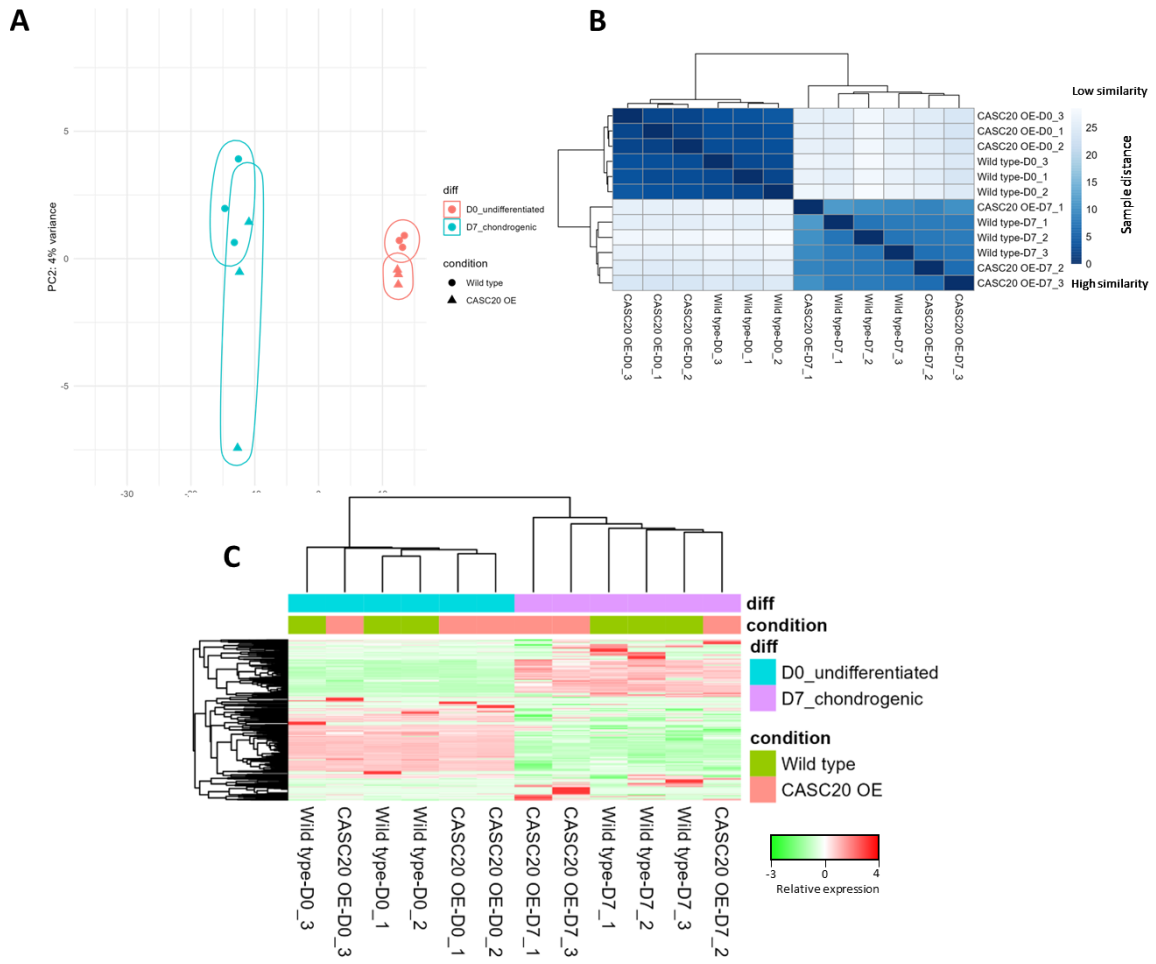


Figure 7.1. Global miR expression profiles during chondrocyte differentiation in wild type versus *CASC20* overexpressing P512 MSCs. A. Principal component analysis was conducted on

a matrix of gene expression data, which was transformed for analysis. Each point on the plot represents an experimental sample. The colour of the point indicates the time point (D0, and D7) and the shape represents the experimental condition (wild type and *CASC20* OE). B. A plot showing the distance between experimental groups based on a Euclidean distance metric. C. The expression heatmap was plotted to display the miR expression levels across different samples. Each row represents a specific miR and each column represents an individual sample. The colour intensity signals the expression level, with dark red indicating higher expression and dark green lower expression.

7.3.2 Pairwise analysis of miR expression in *CASC20* OE versus wild type P512 MSCs during chondrogenesis

To further investigate the differences between *CASC20* OE and wild type, pairwise analysis was conducted at each time point during chondrogenesis (see Figure 7.2). The results showed differential expression of 34 miRs at day 0 in *CASC20* OE versus wild type ($p < 0.05$), with 6 being upregulated and 28 downregulated, as previously described in chapter 5 as the day 0 cells are the same cells for both the chondrogenesis and ostoegenesis experiments. At day 7, only one miR (hsa-mir-1249-3p, log2FC= -0.87442, p = 0.000321) remained differentially expressed in *CASC20* OE versus wild type. This suggests that the effects of *CASC20* overexpression on miR profile are for the most part lost during chondrogenesis.

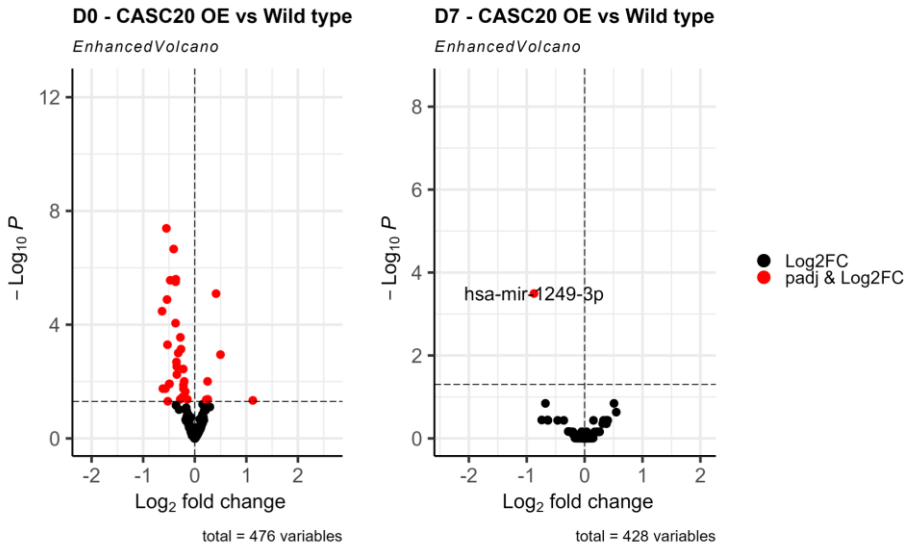


Figure 7.2. Volcano plots comparing *CASC20* overexpression vs wild type at days 0 and 7 during chondrocyte differentiation. The volcano plots depict the differential expression analysis of miRs in *CASC20* OE compared to wild type at different time points (days 0 and 7) during chondrocyte differentiation. Each dot on the plots represents an individual miR, with the x-axis representing the log2 fold change (log2FC) and the y-axis representing the negative logarithm of the adjusted p-value (−log10p).

7.3.3 In-silico analysis of genes that may interact with the miR differentially regulated in *CASC20* OE versus WT P512MSCs at day 7

Next, to explore the potential regulatory role of *CASC20* in chondrocyte differentiation through miRs, I investigated the genes targeted by the only differentially expressed miR (miR-1249-3p), using the StarBase v3 database. A total of 507 mRNAs interact with miR-1249-3p. All 507 mRNAs have been predicted to interact with the miR-1249-3p *in silico*, with 70 mRNAs predicted to interact with the miR through at least two miR-mRNA prediction algorithms within StarBase. Additionally, 295 mRNAs have demonstrated a direct interaction with miR-1249-3p through at least two UV cross-linking and immunoprecipitation (CLIP) studies. Of the 295 mRNAs, 50 are targeted by the miR in at least 8 CLIP studies (Figure 7.3B). Furthermore, 42 mRNAs have shown interaction with the miR via at least 2 binding sites (Figure 7.3C). Analysis of the mRNA target list showed that 12 mRNAs have been demonstrated to undergo degradation by miR-1249-3p (Figure 7.3A).

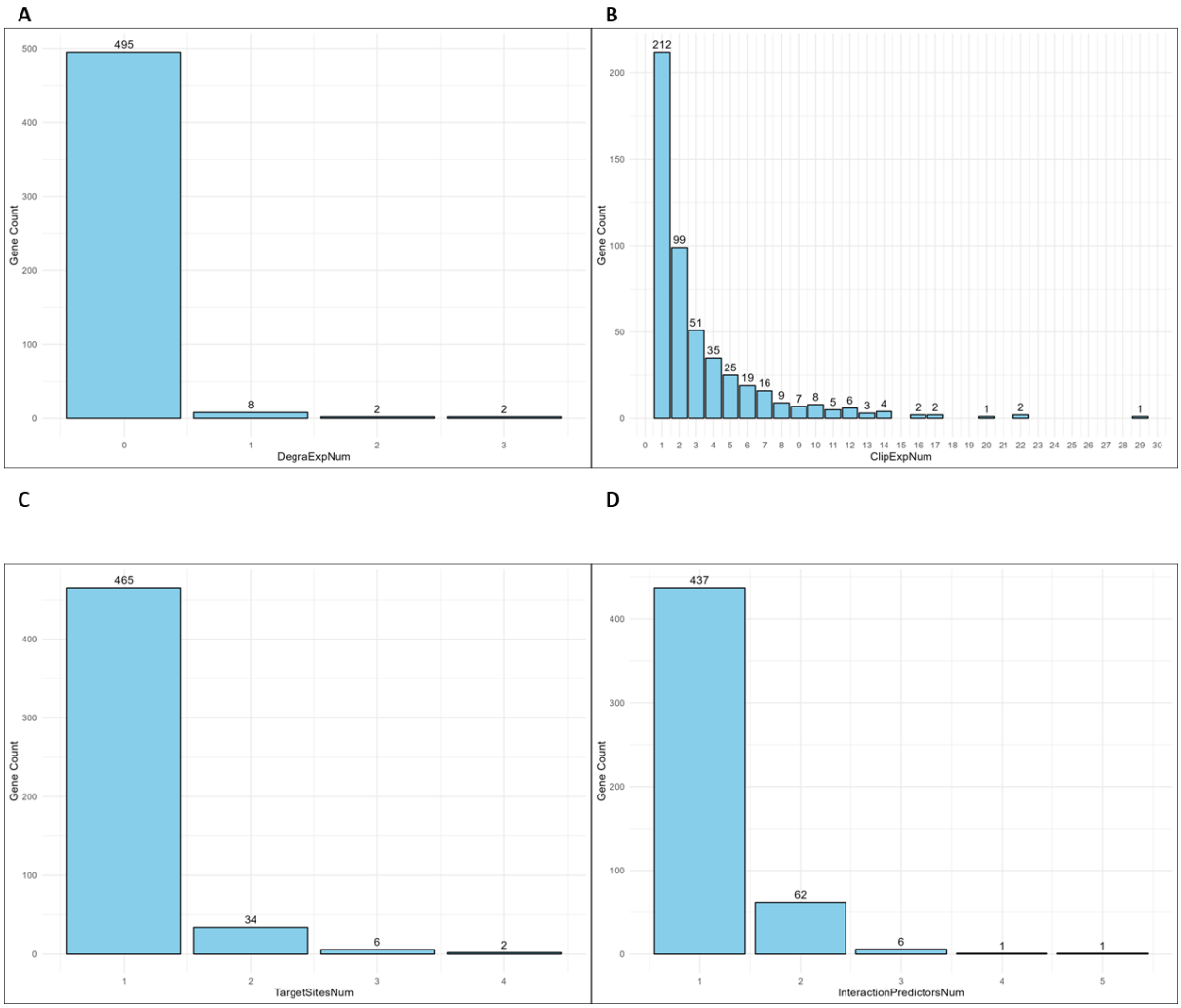


Figure 7.3. Distribution of targeted genes based on differing criteria. The figure illustrates the number of genes targeted by miRs (gene count on the y-axis), ranked by the following criteria: A) DegraExpNum = number of experiments, in which the degradation of mRNA has been observed. B) ClipExpNum = number of CLIP experiments, in which the miR-mRNA interaction has been observed. C) TargetSitesNum = number of distinct binding sites on the mRNA sequence where the miR can potentially interact. D) InteractionPredictorsNum = number of miR-mRNA prediction algorithms that have identified potential miR interactions for the gene.

Subsequently, correlation analysis was employed to test the extent to which each parameter exhibits a correlation with the degradation of targeted mRNA. The findings demonstrated that the parameter with the highest correlation to degradation is the number of interaction predictors. This is followed by the number of Clip experiments validating the interaction, and the number of binding sites (Figure 7.4). RISmed analysis showed that 8 out of the 12 mRNA degraded by miR-1249-3p had experimental evidence for a role in osteo- or chondrogenesis (Table 7.1).

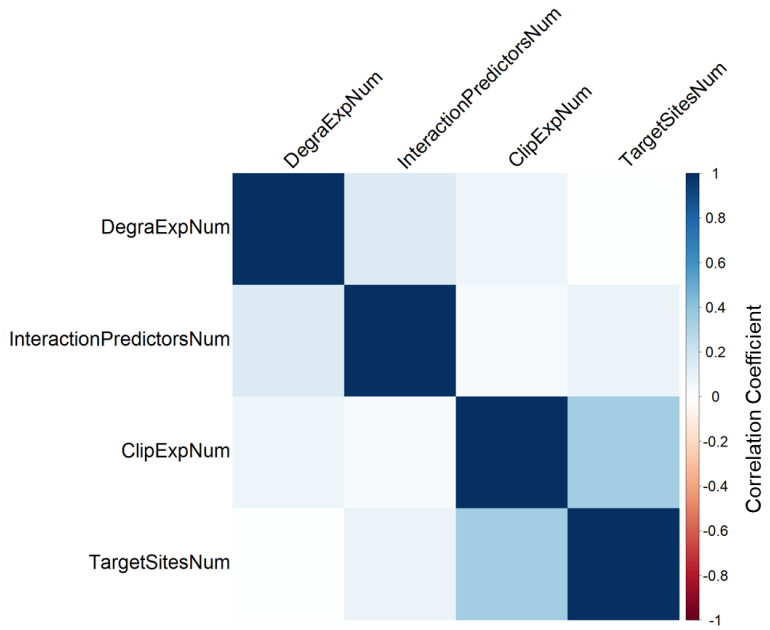


Figure 7.4. Correlation analysis plot. The image illustrates the relationships between the parameters within Starbase v3 and the degradation of targeted mRNA. The matrix represents the Pearson correlation coefficients between pairs of parameters. Each cell in the matrix is color-coded to indicate the strength and direction of the correlation. Positive correlations are shown in blue and negative correlations are shown in red. The diagonal cells are in dark blue to indicate self-correlation.

Table 7.1. Genes degraded by their interaction with miR-1249-3p and most recent PMIDs for their published role in osteo or chondrogenesis.

Target gene	**Most recent PMID number corresponding with published role in osteo or chondrogenesis (search date August 04, 2023)**
HNRNPK	37592256
CSDE1	36042979
ING3	36943599
STC2	37568262
NSD2	30683853
NXF1	28296067
GNB1L	22326833
ID1	37603563

7.3.4 KEGG pathway enrichment analysis of inferred target genes of the differentially expressed miR

Next, I applied KEGG analysis to identify key molecular pathways affected by the mRNA target list (Figure 7.5). The pathways significantly targeted by miR-1249-3p included nucleocytoplasmic transport, p53 signalling pathway, galactose metabolism, endocytosis. MiR-1249-3p was not shown to significantly target pathways that are relevant to osteo- and chondrogenesis. GO analysis suggested that miR-1249-3p could be involved in metabolic processes, cell adhesion, enzyme binding and calcium ion binding (Figure 7.6-8).

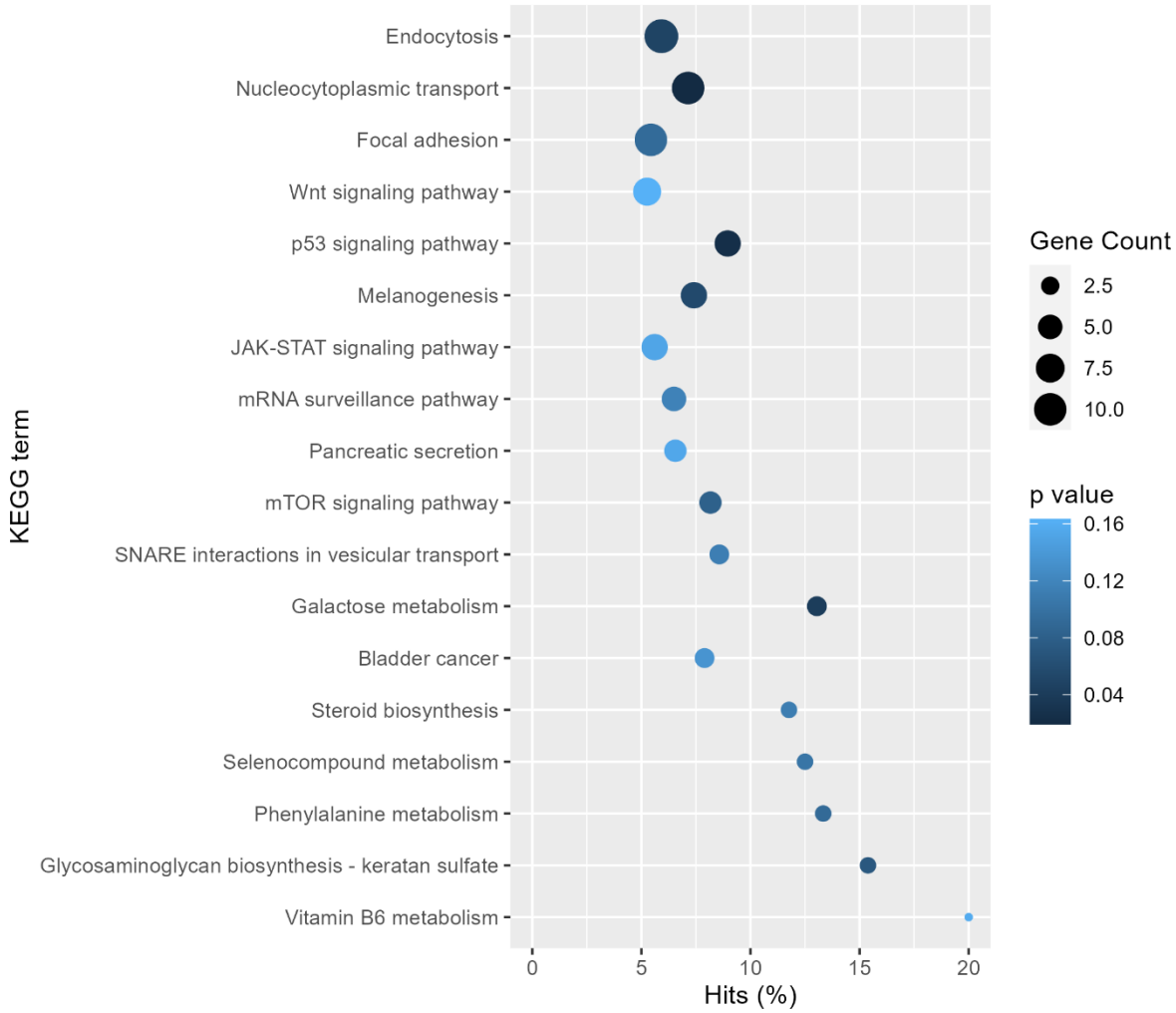


Figure 7.5. Enrichment plot showing KEGG pathways targeted by miR-1249-3p. The scatter plot illustrates the results of KEGG pathway analysis conducted on the mRNA target genes associated with miR-1249-3p. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the KEGG pathway terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

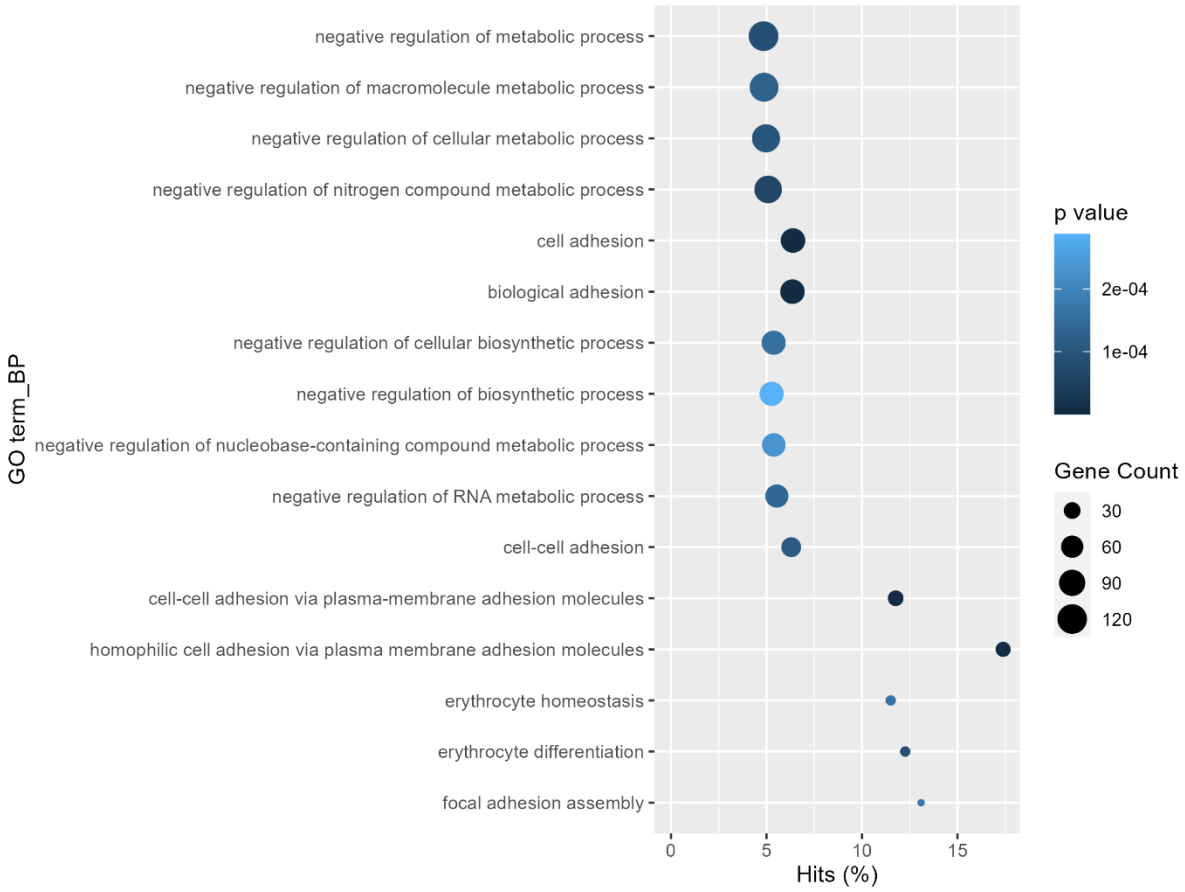


Figure 7.6. Enrichment plot showing GO terms for biological processes (BP) targeted by miR-1249-3p. The scatter plot illustrates the results of GO terms (BP) analysis conducted on the mRNA target genes associated with miR-1249-3p. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the GO terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

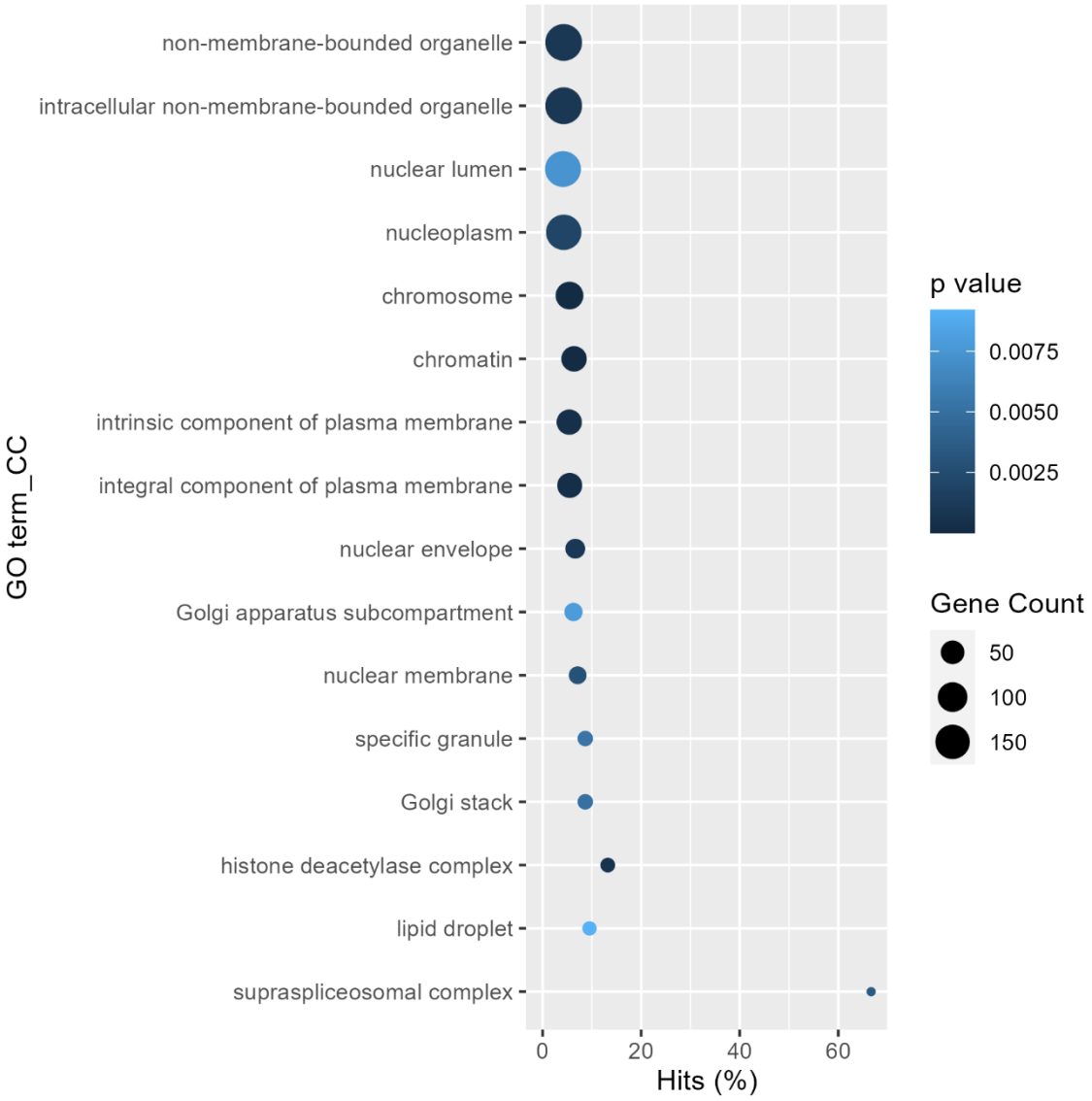


Figure 7.7. Enrichment plot showing GO terms for cellular compartments (CC) targeted by miR-1249-3p. The scatter plot illustrates the results of GO terms (CC) analysis conducted on the mRNA target genes associated with miR-1249-3p. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the GO terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

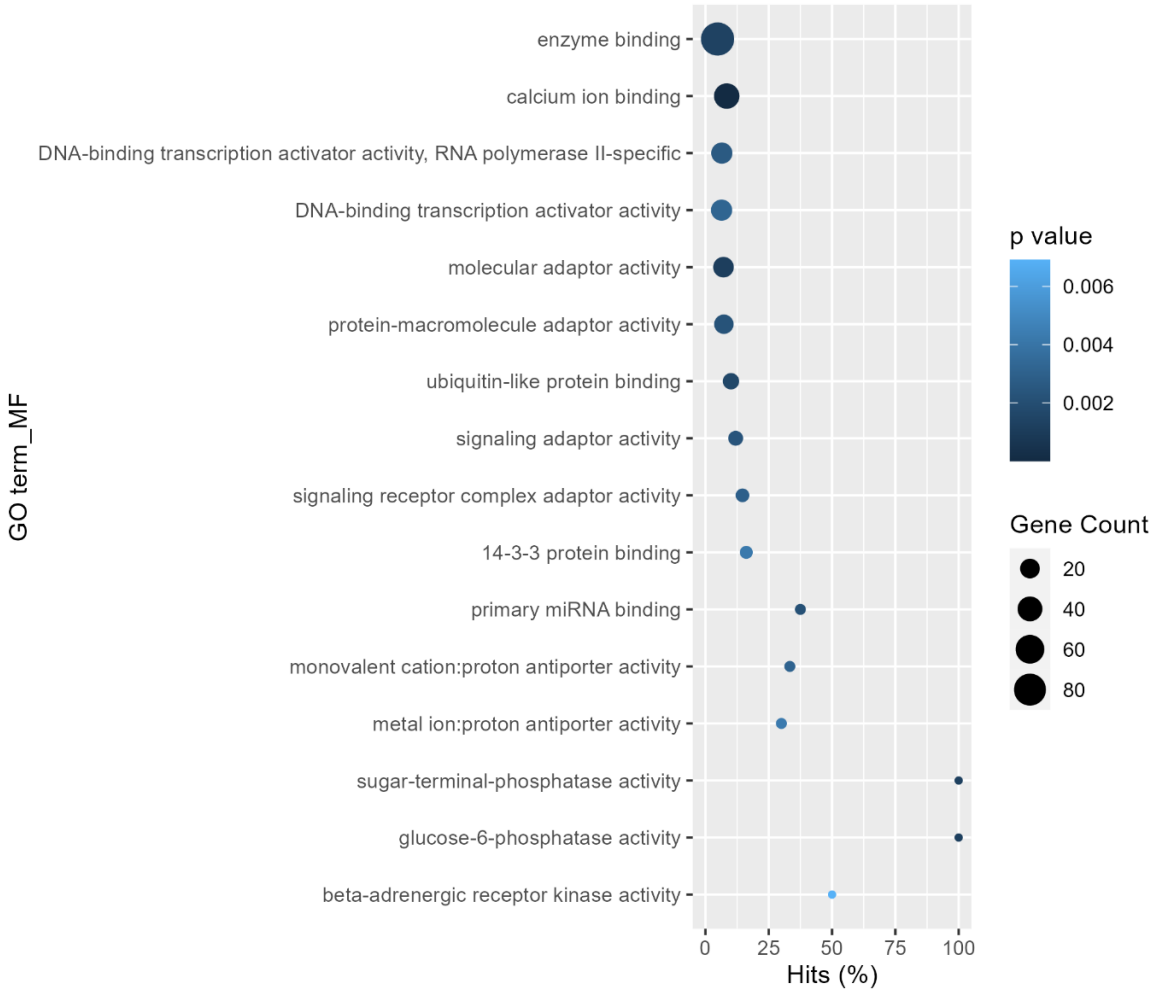


Figure 7.8. Enrichment plot showing GO terms for molecular functions (MF) targeted by miR-1249-3p. The scatter plot illustrates the results of GO terms (MF) analysis conducted on the mRNA target genes associated with miR-1249-3p. The data has been processed and sorted to focus on the top pathways based on their significance scores and gene count. The x-axis represents the percentage of genes in the pathway that are targeted by the miR (Hits %), and the y-axis shows the GO terms. Pathways are color-coded based on their significance represented by the p value, and the size of each point corresponds to the gene count within that pathway.

7.3.5 Comparison between our experimental dataset and published chondrogenesis dataset from Chapter 2

For the published chondrogenic dataset, of the 57 miRs that were differentially expressed miRs at D6 versus D0, none were identified in the experimental dataset and thus the experimental set ups are likely to be sufficiently different to not be able to draw any conclusions from the *in silico* chondrogenesis work for miR profile.

7.4 DISCUSSION

In this study, PCA of miR expression showed distinct temporal variations during chondrocyte differentiation and more limited variation between treatment conditions. Pairwise analysis between *CASC20* OE and wild type at different time points showed that only one miR was differentially expressed at day 7. MiR-1249-3p was shown to be differentially expressed at day 7 and to interact with 507 mRNAs. Whilst 8 targeted genes were demonstrated to play a role in osteo- and chondrogenesis, I found no evidence of significant interactions in respect of pathway analysis to suggest a significant impact on osteochondral pathways.

The data analysis showed various internal and external consistencies. Internal consistencies included PCA analysis, which demonstrates time-dependent miR transcriptional variation. The internal consistency is evident in the pairwise analysis, which demonstrates the early effects of *CASC20* on miR expression. External consistency included interaction predictions, where top target genes have experimental evidence for their roles in osteo- or chondrogenesis. The external consistency is demonstrated by the enrichment analysis, where identified pathways are biologically relevant to the context of osteo- and chondrogenesis is only present at day 0 and not by day 7.

The observed temporal miR transcriptional variation aligns with the well-established concept of temporal regulation during chondrocyte differentiation (598). This suggests that miRs play a role in orchestrating different stages of differentiation, contributing to the formation of chondrocytes. The limited differential expression of miRs at day 7 in *CASC20* OE versus wild type suggests that the examined timepoint might be relatively late to discern the effects of *CASC20* overexpression on miR expression. This could imply that *CASC20's* impact on miRs might be more pronounced during earlier stages of chondrocyte differentiation. Another hypothesis is that *CASC20*'s effects and sensitivity may be specific to MSCs, suggesting that committed osteoblasts or chondrocytes may not exhibit the same response to *CASC20*.

The differential expression of miR-1249-3p and its interaction with numerous mRNAs raise the possibility that this miR might be involved in regulating early pathways related to chondrogenesis. The expression pattern analysis of putative *CASC20*-interacting miRs highlights the dynamic nature of miR expression during chondrocyte differentiation. The absence of differential expression at day 7 suggests that *CASC20* might not significantly impact these miRs at this particular timepoint.

Similar to Chapter 5, the study presented several strengths including the use of whole miRNome analysis, controlled experimental conditions. Limitations of the study included focus on a single cell type, limited time points and a single overexpression model. The absence of experimental conditions without *CASC20* overexpression (e.g., gene knockout) might limit the understanding of *CASC20* impact. Additional experimental assays are needed to confirm the functional relevance of miR-1249-3p, critically including confirmatory mRNA-Seq. Addressing these limitations in future studies would provide a more comprehensive

understanding of the complex regulatory mechanisms at play during chondrocyte differentiation.

Chapter 8 – General Discussion and Conclusion

8.1 Key findings presented in this Thesis

In Chapter 2, I first confirmed in DU145 cells that *CASC20* has a 3' poly-A tail, thus enabling its export into the cytoplasm and theoretical activity as a ceRNA. Next, and because of the first COVID lockdown, I used *in silico* approaches to identify potential miRs that interact with *CASC20* and subsequently validated a subset of these *CASC20* interactions using hMADs, once the approximately 6 month lockdown was over. The results revealed that 64 miRs have the potential to interact with *CASC20*, and among them, 10 are downregulated in both published studies on osteogenesis and chondrogenesis. One of these miRs, miR-485-3p was downregulated during the osteodifferentiation of hMADs following *CASC20* overexpression, providing experimental evidence for the capacity of *CASC20* to regulate miR expression levels.

In Chapter 3, I conducted a pilot study to assess the impact of *CASC20* overexpression (OE) in hMADs differentiated towards osteoblasts, and the results demonstrated that *CASC20* OE may enhance calcium deposition, as indicated by positive Alizarin Red S staining as the biological endpoint of mineralisation.

In Chapter 4, I performed *CASC20* OE and osteogenic differentiation using three types of mesenchymal stem cells: ASC52telo, hMADs, and P512MSCs. The findings showed that *CASC20* OE may promote osteodifferentiation in hMADs and P512MSCs, as evidenced by enhanced Alizarin Red S staining and the upregulation of key osteogenic genes such as RUNX2 and ALP, as confirmed by RT-qPCR.

In Chapter 5, I observed that *CASC20* OE had a temporal regulatory effect on miR expression during osteogenic differentiation, as demonstrated by miR-Seq. Specifically, 34, 43, and 1 miRs were differentially regulated at days 0, 10, and 20 of differentiation, respectively. The top mRNA targets of these differentially regulated miRs were found to have experimental evidence supporting their roles in osteogenesis.

In Chapter 6, I conducted *CASC20* OE and chondrocyte differentiation using three types of stem cells: ASC52teloSOX9, hMADs, and P512MSCs. The results showed that *CASC20* OE may inhibit the chondrogenesis of ASC52teloSOX9 and P512MSCs, as demonstrated by reduced GAG deposition as a biological endpoint of chondrogenesis and the downregulation of RNA levels for key chondrogenic genes, including ACAN, COL2A1, and COMP.

In Chapter 7, I found that *CASC20* OE had a temporal regulatory effect on miR expression during chondrocyte differentiation, as demonstrated by miR-Seq. Specifically, 34 miRs and 1 miR were differentially regulated at days 0 and 7 of differentiation, respectively. The top mRNA targets of the differentially regulated miRs were shown to have experimental evidence supporting their roles in chondrogenesis.

8.2 *CASC20* in osteogenesis

CASC20 may play a significant role in the early stages of osteogenesis in the various *in vitro* model systems used in this thesis. The effects of *CASC20* on osteogenesis exhibit a temporal regulation pattern, with its influence being more pronounced during the earlier stages of differentiation. This is discussed in chapters 2, 3, 5, and 6.

CASC20's modulation of miRs appears to be consistent across the experiments conducted within this thesis. This is elaborated in both Chapter 2 and Chapter 5. This regulation involves both the upregulation and downregulation of specific miRs. I observed that 17 miRs exhibited regulation at both days 0 and 10. Of these, 14 miRs were downregulated at day 0 but upregulated at day 10. Two miRs were upregulated at day 0 but downregulated at day 10. One miR (has-mir-1249-3p) was consistently downregulated at day 20, as well as day 0 and 10 .

8.3 *CASC20* in chondrogenesis

Similar to its role in osteogenesis, *CASC20*'s effects on chondrogenesis also may exhibit temporal regulation. *CASC20*'s influence is shown to be more pronounced during the early stages of differentiation. This is described in both Chapters 2, 6, and 7. *CASC20*'s modulation of miRs appears to be a consistent mechanism throughout the experiments within this thesis. This is discussed in both Chapter 2 and Chapter 7. One miR is consistently downregulated at both days 0 and 7 of the differentiation (hsa-mir-1249-3p).

8.4 Commonly targeted miRs and genes

miR-1249-3p consistently exhibited downregulation across all time points during both osteo- and chondrogenic differentiation of P512MSCs. This consistent pattern is evident in both Chapters 5 and 7. However, miR-1249-3p was not among the 64 putative *CASC20*-miRs identified in the published datasets from Chapter 2. This suggests that miR-1249-3p may be a false negative in the *in silico CASC20*-miR interaction prediction or may be a secondary effect of *CASC20* upregulation.

Limited information is available regarding miR-1249-3p. Published research indicates that miR-1249-3p plays a role in regulating epithelial-mesenchymal transition (EMT) (599, 600). In addition, the lncRNA MIF-AS1 has been shown to act as a sponge for miR-1249-3p, influencing the regulation of HOXB8 in EMT (599). This process has been shown to promote breast cancer cell proliferation, migration and EMT process (599). Further studies are needed to explore whether miR-1249-3p may contribute to EMT in HO or osteo- and chondrogenesis.

In Chapter 2, I identified that the 64 putative *CASC20*-interacting miRs commonly target 19 genes. 13 out of these 19 genes are also targets of the miRs that exhibited differential regulation in the miR-Seq dataset. These mRNA targets have been previously implicated in osteo- or chondrogenesis in published studies. These mRNAs are PURB (601), TNRC6B (602), LCOR (603), ONECUT2 (604), MAPK1 (605), NFAT5 (606), INO80D (607), LONRF2 (608), ABL2 (609), IGF1R (610), KIF1B (611), CDK6 (612), and UBN2 (although no previously published research study tested for its role in osteo- or chondrogenesis). 6 of these mRNA targets were found to be shared between the downregulated and upregulated miRs. These targets were *LONRF2, ABL2, IGF1R, KIF1B, MAPK1,* and *TNRC6B*. This demonstrates the importance of using mRNA-Sequencing to determine *CASC20*'s overall effect on mRNAs.

In Chapter 2, I also found that miR-485-3p may interact with *CASC20* and observed its downregulation during the osteodifferentiation of hMADs. However, this correlation did not reproduce in the miR-Seq experiments presented in Chapters 5 and 7. This discrepancy could be attributed to differences in the cell types and experimental conditions used, and highlights the putative nature of *in silico* datasets and their requirement for experimental validation.

Across the experiments, two pathways consistently emerged as the targets of interest for *CASC20*-modulated miRs, aside from the cancer-associated pathway. The first is the MAPK signalling pathway, which is targeted by both 9 of the 64 putative miRs and 6 differentially expressed miRs in undifferentiated P512MSCs and on day 10 of osteoblast differentiation. The second is the endocytosis pathway, which is targeted by miRs differentially regulated on days 0 (undifferentiated), 10 (osteodifferentiated), and 7 (chondrodifferentiated).

The MAPK signalling pathway has been extensively documented for its contributions to HO, bone development, and cartilage formation (144, 613). A direct link between endocytosis and HO is not established (562, 614). However, endocytosis can modulate various signalling pathways and processes critical for HO, bone, and cartilage formation by regulating receptor-mediated signalling (615) and receptor levels (616). For example, hypoxia stimulates oxygen sensors, resulting in HIF-1 overexpression, endocytosis, and suppression of the degradation of cell surface protein kinase receptors. These promote the retention of mutant ACVR1 on the cell membrane, extending the activation of BMP (617, 618). Moreover, endocytosis has been shown to regulate TGF-β receptors levels and distribution (616).

8.5 P512MSCs cell population and implications

Although there was insufficient time to conduct formal mRNA-Seq analysis as part of this thesis, sequencing of this element of our experimental dataset was conducted over summer 2023. FAST QC analysis of the mRNA-Seq dataset performed on 31.08.2023 showed that the P512 cell RNA mapped to the mouse genome and not the human genome. This was after all experimental work contributing to this thesis was completed, and all but the last results chapter was written. The original cells came from a patient undergoing knee replacement by

my supervisor JMW at the Royal Hallamshire Hospital and were shared with researchers at Sheffield Hallam University (SHU) for human primary bone marrow derived MSC culture work. A flask of the cells was shared back with us once we found that the hMADs were no longer differentiating consistently. The P512 MSCs behaved as expected, showed MSC morphology and differentiated down both osteo and chondrogenic routes, as shown by Alizarin Red S and GAG staining. Around this time at SHU, work was also being conducted using MC3t3-E1 murine pre-osteoblasts, MLO-Y4 murine calvarial osteocytes, and MG63 murine osteosarcoma cells. I speculate that at some point the P512s were cross-contaminated by murine cells and the murine cell line outpopulated the human cells before being transferred back to us. The suspect line is most likely the MC3t3-E1 as they behave like MSCs and may differentiate into both osteoblasts and chondrocytes.

As *CASC20* is a human-only gene, this would mean that the work presented in Chapters 5 and 7 was actually a *CASC20* knock-in model, and thus we were comparing the effect of *CASC20* null versus *CASC20* + on the miRome in osteo and chondrogenesis. miRs are well conserved between human and mice, as demonstrated by equal mapping of the miRs to the mouse and the human genomes. The sequencing data mapped equally well to the human and mouse miRome (Figure 8.1). This means that results from this study are still able to provide meaningful insight into *CASC20*'s effect on miRs and the downstream predicted effects on the human transcriptome.

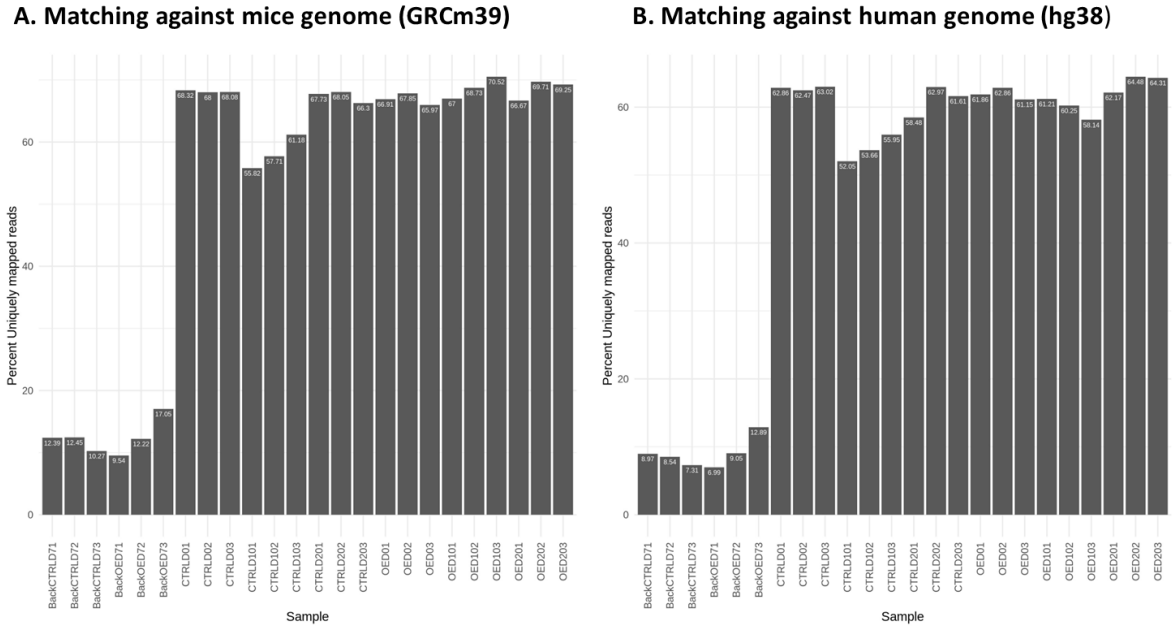


Figure 8.1. Matching miR-Seq generated in this thesis against a) mice genome (GRCm39), and b) human genome (hg38).

This unexpected set of experimental designs could be argued as fortuitous, as the analogous KO model we have been trying to develop using CRISPR in parallel has, as yet been unsuccessful. *CASC20*'s impact on osteo and chondrogenesis in mouse "P512MSCs" is

experimentally similar to our experiments with hMADs and ASCs, and consistent with human MSC behaviour. *CASC20* knock-in shows that *CASC20* expression exhibits a temporal effect over miR expression in osteo- and chondrogenesis similar to that we would expect in hMSCs based on previous analysis.

However, if I were to repeat this set of experiments 'as intended' , I would karyotype the cell by sequencing upon receipt and before conducting experiments.

8.6 Future steps for confirming data in human cells

My lab has a plan to karyotype by sequencing residual cells from original P512 flask sent to us and cells remaining at SHU and also last experimental set of P512s to confirm if contamination occurred prior to their being sent back to us. They plan to repeat the experiments in known human primary cells with over-expression versus wild type, and CRISPR *CASC20* KO versus wild type to examine if similar effects to the mouse knock-in are found, as follows:

1. Extract hMSCs from waste bone samples (in-house), genotype cells, induce *CASC20* OE and repeat +/-OM+/-BMP2 experiments, total RNA sequencing for miRome and expressed transcriptome to confirm the actual gene targets of *CASC20* over-expression
2. Create human CRISPR *CASC20* KO models as proper analogous validation of the mouse knock-in model and repeat the experiments outlined in 1 above to compare the relative effects of over-expression versus KO.

8.7 Further work following validation

After validating the findings, I propose using crosslinking, ligation, and sequencing of hybrids (CLASH) to investigate the precise binding sequences through which *CASC20* interacts with miRs and other RNAs (619). These insight would enrich the understanding of *CASC20*'s regulatory role and may lay the groundwork for potential therapeutic interventions (619).

After that, I would use functional assays to research the impact of *CASC20* on cell proliferation and migration. This could help elucidate the broader cellular processes influenced by *CASC20* (620). Then, I would use cross-linking immunoprecipitation followed by sequencing (CLIP-Seq) to investigate whether *CASC20* can directly interact with proteins (621). At this stage, chromosome conformation capture (3C) could be used to investigate whether *CASC20* regulates chromatin interactions and looping, resulting in gene regulation (596).

Mice implanted with human stem cells expressing *CASC20* could be used to investigate the impact of *CASC20* in HO *in vivo* (622, 623). This model is well-established and can be used to study human-only genes *in vivo* (16, 623, 624). After this, I would conduct functional rescue experiments to rescue the effects of *CASC20* overexpression or knockout by manipulating downstream targets (625). This would help establish causality in the observed phenotypes

(626). Finally, novel therapeutic agents could be developed by screening for small molecules or compounds that can modulate *CASC20* expression or activity in bone and cartilage-related diseases (627, 628).

8.8 Concluding remarks

In summary, the research presented in this thesis has shed some light on the functional role of *CASC20* in osteogenesis and chondrogenesis. Our findings reveal consistent patterns of *CASC20* involvement in the early stages of differentiation, both in terms of its temporal regulation and its modulation of miRs. While the unexpected contamination issue altered the nature of our model, it offered valuable insights into *CASC20*'s impact on miRs and its predicted downstream effects on the human transcriptome.

Further research is warranted to confirm these findings in pure human cell models, validate the role of *CASC20*, and explore the mechanisms involved. The unexpected challenges encountered during this study highlight the importance of rigorous quality control measures in experimental research.

As we continue to unravel the complexities of *CASC20*'s role in musculoskeletal development, these findings contribute to our understanding of the regulatory networks involved and may have implications for future therapies targeting osteogenesis and chondrogenesis.

Chapter 9 – Appendix

9.1 *CASC20* is a susceptibility locus for heterotopic ossification in the human

*Konstantinos Hatzikotoulas[1,5], ***Favour Felix-Ilemhenbhio[2,3]**, George AE Pickering[2], Matthew J Clark[2], Klaudia Kocsy[3], Chiara Niespolo[3], Scott J MacInnes[2], Mine Koprulu[4,5], Lorraine Southam[1,4,6], Ilaria Bellantuono[2], Kajus Baidžajevas[3], David A Young[7], *Eleftheria Zeggini[1], *Endre Kiss-Toth[2,3], *J Mark Wilkinson[2]

*Equal contribution

[1]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

[2]Healthy Lifespan Institute, Department of Oncology & Metabolism, University of Sheffield, Sheffield, UK

[3]Department of Infection, Immunity, and Cardiovascular Disease, University of Sheffield

[4]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

[5]University of Cambridge, Department of Medical Genetics, Cambridge Biomedical Campus, Cambridge, UK

[6]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

[7]Biosciences Institute, Newcastle University, Newcastle Upon Tyne, UK

9.1.1 INTRODUCTION

Heterotopic ossification (HO) is the pathological formation of bone within extra-skeletal tissues that do not normally ossify. HO is a common sequel of trauma, developing in approximately two thirds of casualties after blast injury (10, 629). It is also common after hip replacement surgery with a reported incidence of up to 5% for severe disease and up to 90% for milder forms (630, 631). HO also occurs after traumatic brain injury and burns (632, 633). The clinical impacts of HO include pain and restricted movement, and symptoms due to the compression of adjacent structures, such as nerves and blood vessels. Clinical risk factors for HO after hip replacement include male sex, hypertrophic osteoarthritis, ankylosing spondylitis, hip ankylosis, and African-American ethnicity that suggest a common, complex aetiology (634, 635).

Rare, monogenic diseases that share some phenotypic similarities to post-traumatic HO have also been described. Fibrodysplasia ossificans progressiva (FOP) is a devastating and invariably fatal disease that is also characterised by bone formation at extra-skeletal sites (3). In 97% of cases it is caused by a constitutively-activating mutation in the Activin receptor A

type I gene (*ACVR1*) that codes for a type I bone morphogenetic protein (BMP) receptor (3). Progressive osseous heteroplasia (POH) (4) arises as a result of an extremely rare inactivating mutation in the *GNAS* gene (636) that results in intramembranous ossification within subcutaneous tissue and progressively extends into deeper tissues to muscle, tendons, and ligaments (3). In contrast to these rare monogenic disorders, the molecular pathogenesis of post-traumatic HO is poorly understood although, in common with FOP, BMP-induced signalling is also thought to play a role (4, 636).

Here, we report the first exploration of the genetic architecture of post-traumatic HO through genome-wide association analysis for HO susceptibility in patients after hip replacement. Prioritised signals were followed up in an independent patient cohort. Expression of implicated genes was confirmed in human bone and mechanistically explored using *ex vivo* and *in vitro* models of gene expression and osseous differentiation.

9.1.2 MATERIALS and METHODS

Discovery cohort. 891 British Caucasian men and women (481 controls and 410 cases) who had previously undergone hip replacement for idiopathic osteoarthritis were studied. Controls comprised subjects who had no evidence of HO on plain AP radiographs of the pelvis taken not less than 1 year following primary hip replacement. Cases comprised subjects with radiographic evidence of post-operative HO and were graded (0-4) using the Brooker classification (9). Brooker described: no HO formation (class 0); small islands of bone (class 1); bone spurs from pelvis and proximal femur leaving at least 1cm between opposing surfaces (class 2); bone spurs from pelvis and proximal femur leaving gap less than 1cm (class 3); apparent ankylosis of the hip (class 4). The distribution of patient demographics and the presence and severity of HO by Brooker classification is shown in Supplemental Table 11.

Replication cohort. 419 British Caucasian subjects (207 cases and 212 controls) were recruited not less than 1 year following hip replacement for idiopathic osteoarthritis. HO was graded from plain pelvic radiographs as outlined above using the Brooker grading. Cohort demographics and distribution of Brooker grades for the samples proceeding to case control association analyses are shown in Supplemental Table 11.

Discovery GWAS. DNA from subjects in the discovery cohort was extracted from either whole blood or saliva and genotyped using the Illumina 610k beadchip (Illumina, San Diego, CA). Standard GWAS QC was conducted at the samples and variants level. The exclusion criteria have been previously described (642, 643). Briefly, individuals with 1) gender discrepancy; 2) call rate <95% (<97% in arcOGEN); 3) excess homozygosity or heterozygosity (more than ±3 SD of the mean); 4) duplicates and related samples (π^ >0.2); 5) non-UK European ancestry (ethnicity outliers) were excluded from further analyses. Variants with 1) minor allele

frequency (MAF) ≥5% and a call rate <95%, or a MAF <5% and call rate <99%; 2) monomorphic; 3) exact Hardy Weinberg Equilibrium (HWE) p<0.05 (p<0.0001 in arcOGEN) were also excluded from the merged dataset. Variant QC was carried out on autosomal variants. Genotype-calling intensity plots were examined and single nucleotide polymorphisms (SNPs) with poorly clustering plots were not taken forward. Following QC, 891 subjects (481 controls and 410 cases) and 448770 variants were imputed with IMPUTE2 (644) using the European reference panel from the 1000 Genomes Project (Dec 2010 phase I interim release) (645). Variants with an imputation information score <0.4 and MAF <0.05 were excluded from further analysis.

An HO susceptibility case-control analysis was undertaken on >10 million variants under the additive model using method score implemented in SNPTESTv2 (644). The analysis were adjusted for age and sex as they are known risk factors for HO (646). Data were pruned for linkage disequilibrium (LD) using the clumping function in PLINK (637). Parameters used: (a) significance threshold for index SNP: 1e-5, (b) LD threshold for clumping: 0.20, and (c) physical distance threshold for clumping: 500 kb. Statistical independence of the signals were also confirmed through conditional single-variant association analyses as implemented in SNPTESTv2. A variant was considered independent of the index SNP if the pre- and post-conditioning *p*-value difference was smaller than two orders of magnitude. The top twenty index SNPs were prioritised for replication.

Replication and meta-analysis. DNA from the subjects in the replication cohort was extracted from saliva and genotyped using the iPLEX® Assay of the MassARRAY® System (Agena Bioscience, Inc) to conduct *de novo* replication. Twenty independent and prioritised variants of the discovery stage (Supplemental Table 2) were genotyped. All variants had high Agena design metrics and thus no replacement with highly-correlated proxy SNPs was required. QC was conducted at the sample and variant levels. Sample exclusions were based on sex inconsistencies and a sample call rate <60%. Variants with a call rate <75%, minor allele frequency (MAF) <1% and exact HWE *P*<0.001 in controls were also excluded. Following QC, 205 controls and 198 cases and 13 variants proceeded to association analyses. An HO susceptibility case-control analysis was undertaken under the additive genetic model using the "method maximum likelihood" option as implemented in SNPTESTv2.5.2. Age and sex were used as covariates. The significance threshold for association in the replication study was 0.05/23=0.0022. Finally, we performed a fixed-effects inverse-variance-weighted meta-analysis in METAL (647) across the discovery and replication datasets, comprising a total of 608 cases and 686 controls. Genome-wide significance was defined as $p<5.0 \times 10^{-8}$.

Cell culture. Human mesenchymal stem cells (hMSCs) from 3 independent subjects were obtained from the bone marrow of children undergoing osteotomy and cultured in growth

medium which consisted of: DMEM containing L-glutamine (61965-059) and 4.5g/L Glucose (Gibco), with the addition of 10% hyclone (Scientific Laboratories Supplies, SH30070.03). Human multipotent adipose-derived stem cells (hMADs) were cultured in growth medium which consisted of: DMEM (Lonza, BE12-707F), 10% FBS, 1%glutamine (Gibco, 25030-024), 0.2% Penicillin-Streptomycin (P/S), 1% HEPES (Gibco, 15630-056) and 0.01% hFGF2 (Invitrogen, F0291). Confluency was avoided to prevent differentiation, with hMSCs and hMADs cells split every 3-4 days. For osteogenic differentiation studies, cells were cultured in growth medium for 24 hours, followed by 48 hours of exposure to BMP2. At this timepoint the cells were confluent, and the media was changed into differentiation media.

Osteogenic differentiation. MSCs and hMADs were seeded for 24 hours in growth media, then 300ng/ml human recombinant BMP2 (GenScript, Wanchai, Hong Kong) was added for 48 hours. The media was replaced with the osteogenic media on day 0 of the differentiation. Osteogenic media was made as to the growth media, except for the addition of 300ng/ml BMP2, 10mM β-glycerophosphate, 10nM dexamethasone and 50µg/ml ascorbic acid (Sigma-Aldrich).

RNA isolation and RT-qPCR. Total RNA was isolated using the Promega ReliaPrep™ RNA Cell, Tissue Miniprep System (Promega, Madison, WI) and RNeasy UCP Micro Kit (Qiagen) according to the manufacturer's instructions. Reverse transcription of 400ng of RNA into cDNA was completed using the iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA) and the Veriti 96-well thermal cycler according to the manufacturer's instructions (Applied Biosystems, Waltham, MA). For RT-qPCR, 2ng of cDNA was loaded per well. qPCR samples were run on the C1000 Touch™ Thermal Cycler (Bio-Rad) in 384-well plates. Triplicate technical repeats were conducted for each assay and normalised to a *β-Actin* housekeeping control in murine and *GAPDH* in human. Primers for SYBR green qPCR (Sigma-Aldrich) were designed using Primer-BLAST (NCBI: (www.ncbi.nlm.nih.gov/tools/primer-blast, Supplemental Table 12). All primers were screened to avoid self-complementarity with Oligo Calc: Oligonucleotide Properties Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html). PrecisionPLUS SYBR-Green master mix and TaqMan master mix (Primer design, Southampton, UK) were used with SYBR primers. Ct values were presented as fold change in expression compared to day 0 and after normalisation to the housekeeping control ($2^{-\Delta\Delta CT}$).

Alizarin Red S staining. Cells were washed twice with PBS and fixed overnight at 4°C in 100% ethanol. Cells were then washed twice with PBS before the addition of 40mM Alizarin Red S (Sigma-Aldrich), pH 4.2. The wells were washed extensively with 95% ethanol until all unbound stain was removed, the same number of washes was used for each well. Plates were air dried overnight and scanned on high-resolution flat-bed scanner at 1200dpi. ImageJ

Software (http://rsb.info.nih.gov/ij/) was used to quantify the percentage mineralised area in each well. The area fraction positive for the stain was recorded, representing percentage mineralisation. Identical settings were used for all wells. The wild type wells were used as reference for setting the threshold values.

Statistics. Data are presented as mean ± SEM. Student's t-test or, one-way or two-way ANOVA was used, with various post-hoc tests as indicated in the figure legends. Categorical data are analysed by chi-squared test with Yates' correction, where applicable. All analyses are 2-tailed and statistical significance is represented as $P<0.05$. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001. GraphPad Prism 7 (GraphPad software) was used to present and analyse quantitative data.

Study Approval. For the cohort analyses, all subjects were recruited as part of previous ethically approved studies (642, 648), and provided written, informed consent obtained prior to participation. For the *in vitro* analyses, samples were collected under ethics approval from Oxford NHS REC C (10/H0606/20 and 15/SC/0132), Yorkshire and Humber REC (13/YH/0419), and Human Tissue Authority license 12182, South Yorkshire and North Derbyshire Musculoskeletal Biobank, University of Sheffield, UK.

9.1.3 RESULTS

The genetic architecture of post-traumatic heterotopic ossification. To examine the genetic architecture of HO, we conducted a genome-wide association analysis for disease susceptibility (HO cases versus controls) within patients not less than 1 year after hip replacement surgery for osteoarthritis. Genome-wide analysis of the discovery cohort identified an excess of signals associated with HO susceptibility (Figure 9.1A & 9.1B, Supplemental Table 1), consistent with a heritable component to the condition. Following linkage disequilibrium (LD) pruning using the clumping function in PLINK (637), we selected the lead 20 independent signals for follow up (Supplemental Table 2). Thirteen of these passed our variant-level quality control (QC) pipeline for *de novo* replication (see Methods).

The strongest signal, which reached genome-wide significance, resides in an intergenic region (rs59084763, effect allele (EA) T, effect allele frequency (EAF) 0.19, OR [95% CI] 1.87 [1.47–2.37], p=2.48x10^{-8}; Figure 9.1C), just downstream of *ARHGAP18* that encodes a protein involved in the modulation of cell signalling, cell shape and motility (638). The second strongest signal resides within the long non-coding (lnc) RNA-encoding gene *CASC20* (rs11699612, EA T, EAF 0.25, OR 1.73 [1.40-2.16], p=9.39x10^{-8}; Figure 9.1D). *CASC20* is a human-only lncRNA that has no orthologues in species outside apes. Both its mechanism of action and functional importance in health and disease are currently unexplored.

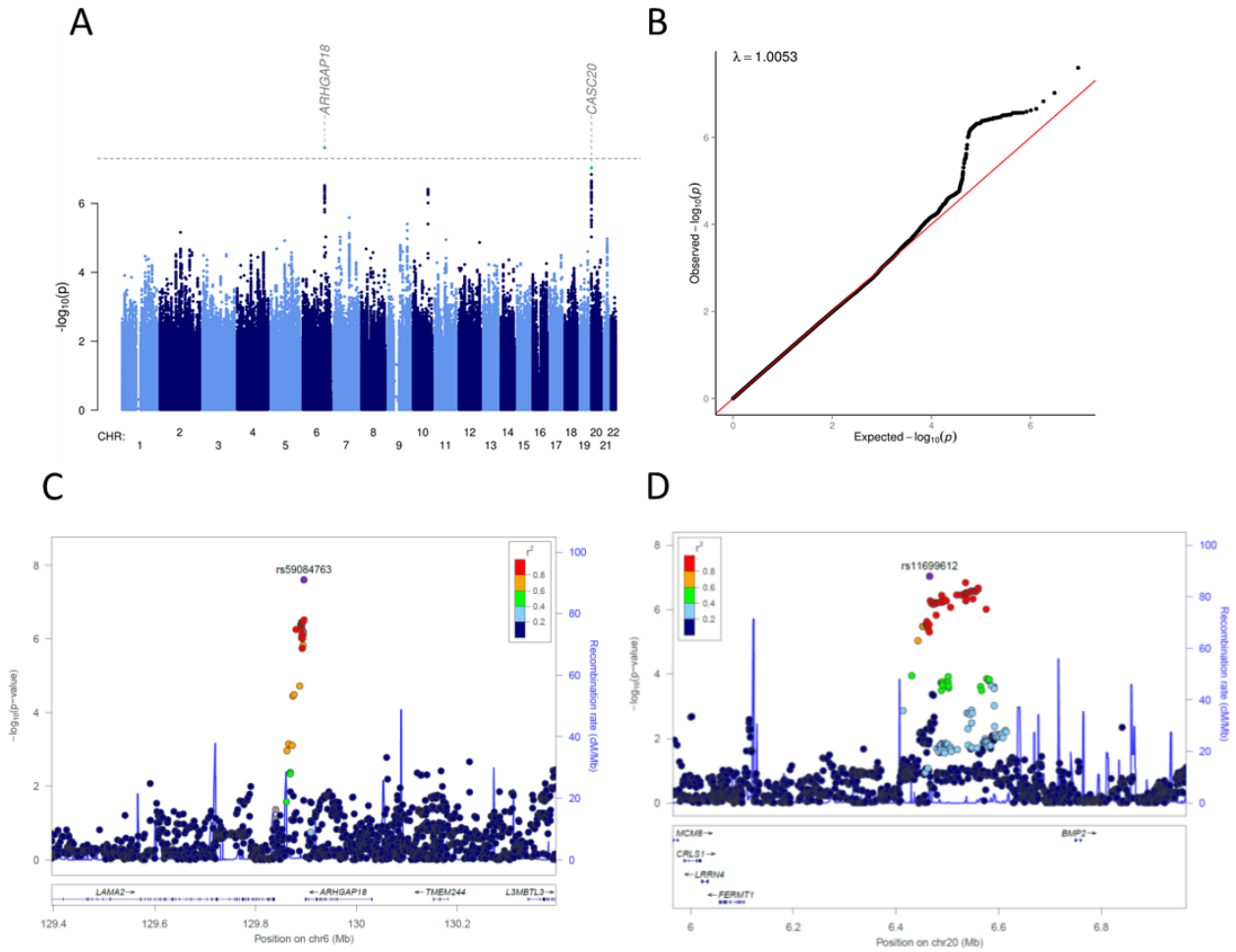


Figure 9.1 Discovery genome-wide association analysis. A) Manhattan plot showing the -log10 *p*-values for each variant (y axis) plotted against their respective chromosomal position (x axis). The horizontal dashed line denotes the genome-wide significance threshold $p=5x10^{-8}$. Lead signals are indicated in green. B) Quantile-quantile plot of the data used in the GWAS. The x-axis indicates the expected −log10 *p*-values and the y-axis the observed ones. The red line represents the null hypothesis of no association at any locus and λ is the genomic inflation factor. Regional association plots for C) rs59084763, and D) rs11699612 with HO susceptibility. Each filled circle represents the *p*-value of analysed variants in the discovery stage plotted against their physical position (NCBI Build 37). The purple circle denotes the variant with the lowest *p*-value in the region. The colours of variants in each plot indicate their r^2 with the lead variant according to a scale from $r^2 = 0$ (blue) to $r^2 = 1$ (red).

Variation within CASC20 is robustly associated with HO susceptibility and is independent of the adjacent BMP2. At replication in an independent hip replacement patient cohort, four of the thirteen independent HO susceptibility signals that passed variant-level QC showed a concordant direction of effect and one residing within the *CASC20* locus replicated at the Bonferroni-corrected significance threshold (rs11699612, EA T, EAF 0.23, OR 1.90 [1.34-2.68], p=2.70x10^{-5}; Supplemental Table 3). Following meta-analysis, this association reached genome-wide significance (EA T, EAF 0.24, OR 1.94 [1.59-2.35], p=2.71x10^{-11}; Figure 9.2A).

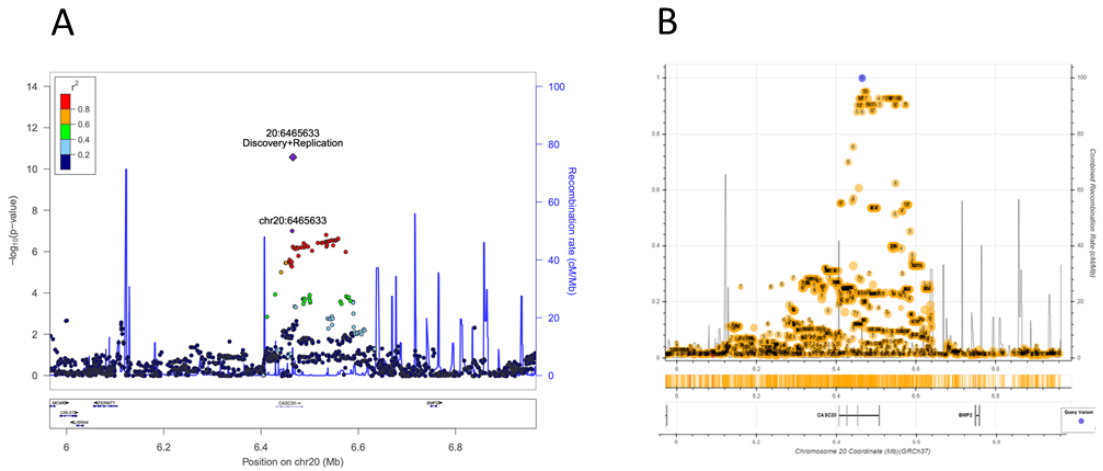


Figure 9.2. *CASC20* is a robust susceptibility locus for HO susceptibility independent of BMP2. A) Regional association plot showing meta-analysis of rs11699612 in discovery and replication cohorts. Each filled circle represents the *p*-value of analysed variants in the discovery stage plotted against their physical position (NCBI Build 37). The purple circle denotes rs11699612 (chr20:6465633), which is the variant with the lowest *p*-value in the region. The colours of variants in each plot indicate their r^2 with the lead variant according to a scale from $r^2 = 0$ (blue) to $r^2 = 1$ (red). B) Fine mapping of the *CASC20* locus using a clumping threshold of 2000 kb either side of rs11699612 to include the whole *BMP2* locus and most of the proximal coding genes identified no variants within *BMP2* in linkage disequilibrium >0.03 with *CASC20* variants. Variant rs11699612 is indicated in purple.

CASC20 resides in close proximity to *BMP2* on chr20. In order to confirm the origin of the signal as lying within *CASC20* rather than *BMP2*, we increased the physical distance threshold used for clumping to 2000 kb either side of rs11699612 to include the whole *BMP2* locus and most of the proximal coding genes. We found no HO-associated variants within *BMP2* nor any *BMP2* variants in higher LD than r^2=0.03 with the *CASC20* variant, confirming *CASC20* as the origin of the genetic signal (Figure 9.2B and Supplemental Table 4). Given the proximity of *CASC20* to *BMP2*, we further examined whether rs11699612 is a cis-acting expression quantitative trait locus (cis-eQTL) for *BMP2* expression using RNA sequencing data from unstimulated primary chondrocytes and synovium taken from an independent cohort of 100

patients undergoing joint replacement (See data availability section and biorxiv.org/content/10.1101/835850v1). Data were analysed using a GTEx-modified version of FastQTL (639), and confirmed no evidence of rs11699612 action as a cis-eQTL on *BMP2* or any other adjacent genes within this 1Mb window, with the exception of *MCM8* that encodes minichromosome maintenance protein 8 (Supplemental Table 5).

CASC20 is expressed in human bone and is induced in BMP2-stimulated human mesenchymal stem cells. To determine whether *CASC20* is expressed in human bone tissue, we extracted total RNA from fresh frozen, surgically excised bone from patients undergoing joint replacement. *CASC20* expression was confirmed by real time quantitative polymerase chain reaction (RT-qPCR, Figure 9.3A). Next, we explored whether *CASC20* is differentially expressed in human multipotent adipose-derived stem cells (hMAD, Figure 9.3B)(512) and in primary human bone marrow-derived mesenchymal stem cells (hMSCs, Figure 9.3C-E) in response to stimulation with BMP2 and osteogenic supplements. We found that *CASC20* expression was significantly upregulated by day 8 in hMADs, and by day 16 or 24 in hMSCs. This was associated with robust upregulation of markers of osteogenic differentiation runt-related transcription factor 2 (*RUNX2*) and the transcription factor Sp7/Osterix (*OSX*) in both the hMADs and the hMSCs at these timepoints, and a significant increase in percentage mineralisation per well at day 24, measured by Alizarin Red S stain.

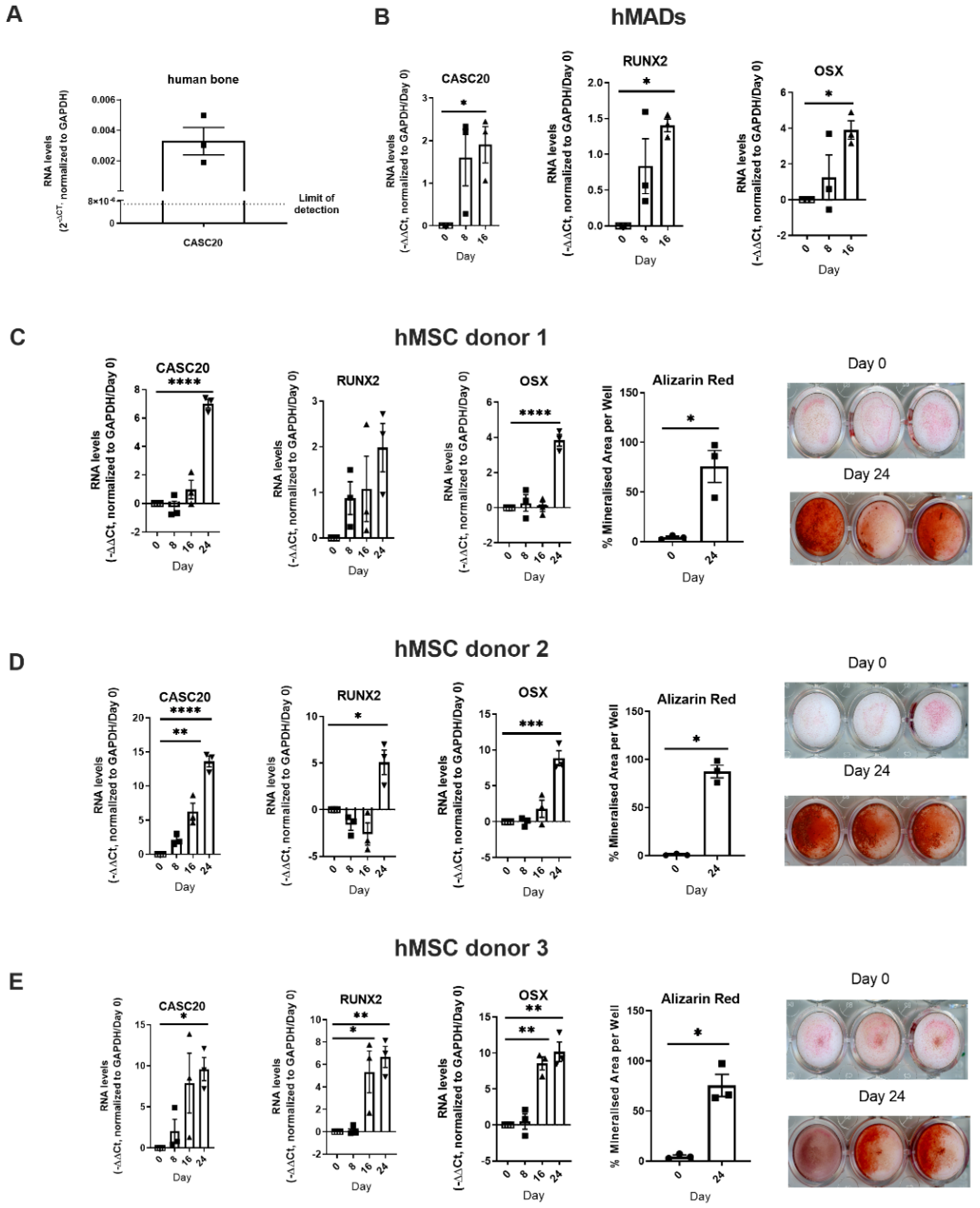


Figure 9.3. *CASC20* is expressed in human bone and are induced in mesenchymal stem cells by BMP2 *in vitro*. A) RT-qPCR was used to measure the expression of *CASC20* RNA in waste bone samples retrieved at joint replacement (n=3 subjects). B) *CASC20* is induced in human Multipotent Adipose-Derived Stem cells (hMADs). RT-qPCR was used to measure the expression of *CASC20*, *RUNX2* and *OSX* in hMADs at days 0, 8 and 16 of hMAD differentiation. Data were analysed using 2^(-ΔCt) by normalising to *GAPDH* (n=3 biological replicates). One-way ANOVA with Dunnett's multiple comparisons; data is plotted as mean ±SEM. **C-E)** *CASC20* expression and mineralisation are induced in BMP2-stimulated human Bone Marrow Derived Mesenchymal Stem Cells (hMSC). RT-qPCR was used to measure the expression of *CASC20*, *RUNX2* and *OSX* in hMSCs from 3 donors at days 0, 8, 16 and 24 of differentiation. Data were analysed using –ΔΔCt, normalising to day 0 and GAPDH (n=3 biological replicate cultures).

Percentage mineralised area per well at day 0 and day 24 were measured by percentage Alizarin Red S staining. Analyses are one-way ANOVA with Dunnett's multiple comparisons or Student's t-test (Alizarin Red S); data is plotted as mean ±SEM.

9.1.4 DISCUSSION

In this first exploration of the genetic architecture of post-traumatic HO, we identify a robust, replicating signal for HO susceptibility within the previously uncharacterised human-only lncRNA *CASC20*, with the lead signal at rs11699612. We show that although *CASC20*'s nearest neighbour is *BMP2*, variation within *CASC20* is independent of *BMP2* nor does it act as a cis-eQTL for *BMP2*. In functional analyses, we demonstrate that *CASC20* is expressed at very low levels, but is upregulated following an osteogenic stimulus.

Although both the discovery and replication sample sizes examined here were limited, the genotyped collections studied amass the largest available sample size globally to date. Larger sample sizes will be necessary to increase the number of HO risk loci robustly identified (Supplemental Figure 2). For example, in the discovery dataset we also identified genome-wide significant variation at rs59084763, 1kB downstream from the gene encoding Rho GTPase activating protein 18 that subsequently failed to replicate and was therefore not followed up in functional analyses, but may represent a further HO susceptibility locus. We demonstrate here that *CASC20* is expressed at low levels in unstimulated human adult musculoskeletal tissue and that its expression is upregulated in response to BMP2 as an osteogenic stimulus. Although BMP2 expression has been demonstrated in post-traumatic clinical HO tissue (640, 641), this may not represent the dominant mechanistic pathway in post-traumatic HO.

Here, we provide first insights into the genetic architecture of post-traumatic HO that indicate it is a common, complex disorder. Our data establish *CASC20* as the first robust locus for HO susceptibility. We present evidence in support of *CASC20* modulating HO susceptibility through its interaction with miRs that regulate osteo- and chondrogenesis. Further studies of *CASC20* in relevant human cell models will help clarify its role in human health and disease and the role of the rs11699612 variant in the pathogenesis of the disease.

AUTHOR CONTRIBUTIONS

EKT, EZ and JMW designed the experiments. KH, FFI, SD, GAEP, SD, MJC, KK, JS, SJM, MK, KB, DAY and AG carried out the experimental work; IB provided the hMSC lines. All authors were involved in the analysis and interpretation of the results. EKT and JMW wrote the manuscript, and all authors contributed to its editing.

9.2 R Scripts for miRSeq analysis

9.2.1 DESeq2

#Set up environment

Set working directory

```{r setup, include=FALSE}

#knitr::opts_knit$set(root.dir = "X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/miRNA_Seq_DESeq_miRNA_mature/")

#setwd("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/miRNA_Seq_DESeq_miRNA_mature/")

```

Load required packages

```{r}

#if (!requireNamespace("BiocManager", quietly = TRUE))

#    install.packages("BiocManager")

#BiocManager::install("DESeq2")

library("DESeq2")

library("ggplot2")

library("pheatmap")

library("dplyr")

library(RColorBrewer)

library(knitr)

library(ggforce)

```

Read ensembleIDs2miRID

```{r}
```

#ensemblIDs2miRID <-
read.table("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/miRNA_Seq_DESeq_miRNA_only/ensemblIDs2miRbaseID.tsv")

ensemblIDs2miRID2 <-
read.table("X:/sudlab1/General/mirror/miRBase/ensemblIDs2miRbaseID.tsv", header=T)

miRBase_hp2mature <-
read.table("X:/sudlab1/General/mirror/miRBase/miRBase_hp2mature.tsv", header=T)

```

Filter input

```{r}

input <-
read.table("X:/sudlab1/General/projects/Favour_EKT/miRNASeq_20230313/pipeline_seq_favour/counts_first.dir/Primary_resume_counts.txt",

                header = T, sep = "\t")

header <- colnames(input)

header <- sub("^Back", "", header)

colnames(input) <- header

```

Remove NAs and lowly expressed miRs

```{r}

# Remove rows with NAs from miRNA_filter_inputs

input <- input[complete.cases(input), ]


input <- input[rowSums(input[2:ncol(input)]) >= 1,]

input <- as.matrix(input)


```

Convert to miRNA name id
```

```{r}
# Create a mapping vector from miRBase_ID to Name
mapping <- with(miRBase_hp2mature, setNames(Name, miRBase_ID))
# Rename row names in input using the miRBase_ID2miRID_vector
rownames(input) <- mapping[rownames(input)]
```

```{r}
#rename header of input table
# Renaming the headers
colnames(input) <- gsub("^CTRLD", "Wild type-D", colnames(input))
colnames(input) <- gsub("^OED", "*CASC20* OE-D", colnames(input))
# Adding an underscore before the last digit in each header
# Get the column names
col_names <- colnames(input)
# Function to add an underscore based on the last digit in a string
add_underscore <- function(x) {
  gsub("(\\d)$", "_\\1", x)
}
# Apply the function to all column names and update them
colnames(input) <- sapply(col_names, add_underscore)
```

#1. Analyse osteo and chondro

Extract sample information

```{r}
header <- colnames(input)
samples <- data.frame(samples = header,
```

```
                condition = ifelse(grepl("^Wild type", header), "Wild type", "CASC20 OE"),
                timepoint = ifelse(grepl("D0", header), "D0",
                                   ifelse(grepl("D7", header), "D7",
                                          ifelse(grepl("D10", header), "D10",
                                                 ifelse(grepl("D20", header), "D20", NA)))),
                diff = ifelse(grepl("D0", header), "undifferentiated",
                              ifelse(grepl("D7", header), "chondrogenic",
                                     ifelse(grepl("D10", header), "osteogenic",
                                            ifelse(grepl("D20", header), "osteogenic", NA)))))

samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
merged <- paste(samples$timepoint, samples$diff, sep = "_")
samples <- samples %>%
  mutate(diff = merged) %>%
  dplyr::select(-timepoint)
#head(samples)
```

Convert columns to factors and make them unordered.

```{r}
samples$condition <- factor(samples$condition, levels = c("Wild type", "CASC20 OE"), ordered = FALSE)
samples$diff <- factor(samples$diff, levels = c("D0_undifferentiated", "D7_chondrogenic",
                                                "D10_osteogenic", "D20_osteogenic"), ordered = FALSE)
```

Create a deseq by specifying the design within the script itself.

```{r}
#base specify it
dds <- DESeqDataSetFromMatrix(input,
                              colData = samples,
                              design = ~ diff + condition + condition:diff)

dds <- DESeq(dds)
vsd <- vst(dds, blind=FALSE, nsub = sum( rowMeans( counts(dds, normalized=TRUE)) > 5 ))
```
Plot PCA graph
```{r}
pcaData <- plotPCA(vsd, intgroup=c("condition", "diff"), returnData=TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))
p <- ggplot(pcaData, aes(PC1, PC2, color=diff, shape=condition)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ",percentVar[1],"% variance")) +
  ylab(paste0("PC2: ",percentVar[2],"% variance")) +
  coord_fixed() +
  geom_mark_ellipse() +
theme_minimal()
# Manually increase the x and y axis limits by 10%
p <- p + coord_cartesian(xlim = range(pcaData$PC1) * 1.1, ylim = range(pcaData$PC2) * 1.1)
#dev.off()
#ggsave("merged.1.PlotPCA.pdf", p)
#ggsave("merged.1.PlotPCA.png", p)
P
```

```

Heatmap of the sample-to-sample distances

```{r}
sampleDists <- dist(t(assay(vsd)))
sampleDistMatrix <- as.matrix(sampleDists)
#rownames(sampleDistMatrix) <- paste(vsd$condition, vsd$samples, sep="-")
#colnames(sampleDistMatrix) <- paste(vsd$condition, vsd$samples, sep="-")
colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
#run pheatmap
p <- pheatmap(sampleDistMatrix,
      clustering_distance_rows=sampleDists,
      clustering_distance_cols=sampleDists,
      col=colors)
# Adjust row annotation font size
#p <- p + theme(axis.text.y = element_text(size = 8))
#ggsave("merged.1.h_sampledist.pdf", p, width = 6, height = 6)
#ggsave("merged.1.h_sampledist.png", p, width = 6, height = 6)

dev.off()

include_graphics("merged.1.h_sampledist.png")
```

Heatmap of the count matrix. To explore a count matrix, it is often

instructive to look at it as a heatmap

```{r}
ntd <- normTransform(dds)
```
```

```
select <- order(rowMeans(counts(dds,normalized=TRUE)),
                decreasing=TRUE)#[1:352]
df <- as.data.frame(colData(dds)[,c("condition","diff")])
normalized_data <- t(scale(t(assay(ntd))))
p <- pheatmap(normalized_data[select,], cluster_rows=TRUE, show_rownames=FALSE,
       cluster_cols=TRUE, annotation_col=df)
#ggsave("merged.1.h_count.pdf", p)
#ggsave("merged.1.h_count.png", p)
```

#2. Analyse osteo only

```{r}
#head(filter_inputs)
# identify columns to remove
cols_to_remove <- grep("D7", colnames(input))
# remove columns
filter_inputs_2 <- input[, -cols_to_remove]
colnames(filter_inputs_2)
header <- colnames(filter_inputs_2)
# Extract sample information
samples <- data.frame(samples = header,
                  condition = ifelse(grepl("^Wild type", header), "Wild type", "*CASC20* OE"),
                  timepoint = ifelse(grepl("D0", header), "D0",
                                     ifelse(grepl("D7", header), "D7",
                                            ifelse(grepl("D10", header), "D10",
                                                   ifelse(grepl("D20", header), "D20", NA)))),
                  diff = ifelse(grepl("D0", header), "undifferentiated",
```

```
                        ifelse(grepl("D7", header), "chondrogenic",
                               ifelse(grepl("D10", header), "osteogenic",
                                      ifelse(grepl("D20", header), "osteogenic", NA)))))
samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
head(samples)
merged <- paste(samples$timepoint, samples$diff, sep = "_")
samples <- samples %>%
  mutate(diff = merged) %>%
  select(-timepoint)
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("Wild type", "CASC20 OE"), ordered = FALSE)
samples$diff <- factor(samples$diff, levels = c("D0_undifferentiated",
                                                "D10_osteogenic", "D20_osteogenic"), ordered = FALSE)
#base specify it
dds_2 <- DESeqDataSetFromMatrix(filter_inputs_2,
                                colData = samples,
                                design = ~ diff + condition + condition:diff)


dds_2 <- DESeq(dds_2)
vsd_2 <- vst(dds_2, blind=FALSE, nsub = sum( rowMeans( counts(dds_2, normalized=TRUE)) > 5 ))
```

PCA

```{r}
pcaData_2 <- plotPCA(vsd_2, intgroup=c("condition", "diff"), returnData=TRUE)
```

```
percentVar_2 <- round(100 * attr(pcaData_2, "percentVar"))
p <- ggplot(pcaData_2, aes(PC1, PC2, color=diff, shape=condition)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ",percentVar_2[1],"% variance")) +
  ylab(paste0("PC2: ",percentVar_2[2],"% variance")) +
  coord_fixed() +
geom_mark_ellipse() +
theme_minimal()
# Manually increase the x and y axis limits by 10%
p <- p + coord_cartesian(xlim = range(pcaData$PC1) * 1.1, ylim = range(pcaData$PC2) * 1.1)

#dev.off()
#ggsave("Osteo.2.PlotPCA.png", p)
#ggsave("Osteo.2.PlotPCA.pdf", p)
p
#include_graphics("Osteo.2.PlotPCA.png")
```

Heatmap of the sample-to-sample distances

```{r}
sampleDists <- dist(t(assay(vsd_2)))
sampleDistMatrix <- as.matrix(sampleDists)
#rownames(sampleDistMatrix) <- paste(vsd_2$condition, vsd_2$samples, sep="-")
#colnames(sampleDistMatrix) <- paste(vsd_2$condition, vsd_2$samples, sep="-")
colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
p <- pheatmap(sampleDistMatrix,
        clustering_distance_rows=sampleDists,
```

```
        clustering_distance_cols=sampleDists,
        col=colors)
#ggsave("Osteo.2.h_sampledist.png", p)
#ggsave("Osteo.2.h_sampledist.pdf", p)
```

Heatmap of the count matrix

```{r}
ntd <- normTransform(dds_2)
select <- order(rowMeans(counts(dds_2,normalized=TRUE)),
                decreasing=TRUE) [1:530]
df <- as.data.frame(colData(dds_2)[,c("condition","diff")])
normalized_data <- t(scale(t(assay(ntd))))
p <- pheatmap(normalized_data[select,], cluster_rows=TRUE, show_rownames=FALSE,
        cluster_cols=TRUE, annotation_col=df,
 color = colorRampPalette(c("green", "white", "red"))(100),
  legend = FALSE
 )


#ggsave("Osteo.2.h_count.png", p)
#ggsave("Osteo.2.h_count.pdf", p)
```

D0 - Filter the ntd object by a specific level of the diff column and generate a heatmap for the filtered data

```{r}
# Generate heatmap for D0_undifferentiated condition only
select <- order(rowMeans(assay(ntd)), decreasing=TRUE)[1:530]
```

```
df <- as.data.frame(colData(ntd)[,c("condition","diff")])


normalized_data <- t(scale(t(assay(ntd))))
p <- pheatmap(normalized_data[select,colData(ntd)$diff == "D0_undifferentiated"],
         cluster_rows=TRUE, show_rownames=FALSE,
         cluster_cols=TRUE, annotation_col=df,
 color = colorRampPalette(c("green", "white", "red"))(100),
  legend = FALSE
 )


#ggsave("Osteo.2.D0_h_count.png", p)
#ggsave("Osteo.2.D0_h_count.pdf", p)
p
```

D10 - Filter the ntd object by a specific level of the diff column and generate a heatmap for the filtered data

```{r}
# Generate heatmap for D0_undifferentiated condition only
select <- order(rowMeans(assay(ntd)), decreasing=TRUE)[1:530]
df <- as.data.frame(colData(ntd)[,c("condition","diff")])


normalized_data <- t(scale(t(assay(ntd))))
p <- pheatmap(normalized_data[select,colData(ntd)$diff == "D10_osteogenic"],
         cluster_rows=TRUE, show_rownames=FALSE,
         cluster_cols=TRUE, annotation_col=df,
 color = colorRampPalette(c("green", "white", "red"))(100),
```

```
  legend = FALSE
 )
#ggsave("Osteo.2.D10_h_count.png", p)
#ggsave("Osteo.2.D10_h_count.pdf", p)
p
```

D20 - Filter the ntd object by a specific level of the diff column and generate a heatmap for the filtered data

```{r}
# Generate heatmap for D0_undifferentiated condition only
select <- order(rowMeans(assay(ntd)), decreasing=TRUE)[1:530]
df <- as.data.frame(colData(ntd)[,c("condition","diff")])
normalized_data <- t(scale(t(assay(ntd))))
p <- pheatmap(normalized_data[select,colData(ntd)$diff == "D20_osteogenic"],
         cluster_rows=TRUE, show_rownames=FALSE,
         cluster_cols=TRUE, annotation_col=df,
 color = colorRampPalette(c("green", "white", "red"))(100),
  legend = FALSE
 )
#ggsave("Osteo.2.D20_h_count.png", p)
#ggsave("Osteo.2.D20_h_count.pdf", p)
p
```

#3. Analyse chondro only ----

```{r}
#filter_inputs
```

```
cols_to_remove <- grep("D10|D20", colnames(input))
filter_inputs_3 <- input[, -cols_to_remove]
header <- colnames(filter_inputs_3)
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^Wild type", header), "Wild type", "CASC20 OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))

samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
head(samples)
merged <- paste(samples$timepoint, samples$diff, sep = "_")
samples <- samples %>%
  mutate(diff = merged) %>%
  select(-timepoint)
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("Wild type", "CASC20 OE"),
ordered = FALSE)
samples$diff <- factor(samples$diff, levels = c("D0_undifferentiated", "D7_chondrogenic"),
ordered = FALSE)
```

```
#base specify it
dds_3 <- DESeqDataSetFromMatrix(filter_inputs_3,
                        colData = samples,
                        design = ~ diff + condition + condition:diff)


dds_3 <- DESeq(dds_3)
vsd_3 <- vst(dds_3, blind=FALSE, nsub = sum( rowMeans( counts(dds_3, normalized=TRUE)) > 5 ))
```

PCA

```{r}
pcaData_3 <- plotPCA(vsd_3, intgroup=c("condition", "diff"), returnData=TRUE)
percentVar_3 <- round(100 * attr(pcaData_3, "percentVar"))
p <- ggplot(pcaData_3, aes(PC1, PC2, color=diff, shape=condition)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ",percentVar_3[1],"% variance")) +
  ylab(paste0("PC2: ",percentVar_3[2],"% variance")) +
  coord_fixed() +
geom_mark_ellipse() +
theme_minimal()
# Manually increase the x and y axis limits by 10%
p <- p + coord_cartesian(xlim = range(pcaData$PC1) * 1.6, ylim = range(pcaData$PC2) * 1.6)
#dev.off()
#ggsave("Chondro.3.PlotPCA.pdf", p)
#ggsave("Chondro.3.PlotPCA.png", p)
p
```

```

Heatmap of the sample-to-sample distances

```{r}
sampleDists <- dist(t(assay(vsd_3)))
sampleDistMatrix <- as.matrix(sampleDists)
#rownames(sampleDistMatrix) <- paste(vsd_3$condition, vsd_3$samples, sep="-")
#colnames(sampleDistMatrix) <- paste(vsd_3$condition, vsd_3$samples, sep="-")
colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
p <- pheatmap(sampleDistMatrix,
        clustering_distance_rows=sampleDists,
        clustering_distance_cols=sampleDists,
        col=colors)

#ggsave("Chondro.3.p_sampledist.pdf", p)
#ggsave("Chondro.3.p_sampledist.png", p)
p
```

Heatmap of the count matrix

```{r}
ntd <- normTransform(dds_3)
select <- order(rowMeans(counts(dds_3,normalized=TRUE)),
              decreasing=TRUE)[1:507]
df <- as.data.frame(colData(dds_3)[,c("condition","diff")])
normalized_data <- t(scale(t(assay(ntd))))
p <- pheatmap(normalized_data[select,], cluster_rows=TRUE, show_rownames=FALSE,
        cluster_cols=TRUE, annotation_col=df,
```
```

```
 color = colorRampPalette(c("green", "white", "red"))(100),
  legend = FALSE
 )
##ggsave("Chondro.3.p_count.pdf", p)
#ggsave("Chondro.3.p_count.png", p)
p
```

CASC20 miRs

```{r}
expression_data <- assay(ntd)
expression_data <- expression_data[intersect(tolower(rownames(expression_data)), tolower(*CASC20*miRs$V1)), ]
# Extract sample names
sample_names <- colnames(expression_data)
# Identify unique sample groups
sample_groups <- unique(sub(".$", "", sample_names))
# Calculate average of each group of three samples
averaged_data <- sapply(sample_groups, function(group) {
  group_samples <- sample_names[grep(group, sample_names)]
  rowMeans(expression_data[, group_samples])
})
# Create a new matrix with averaged column data
averaged_matrix <- as.matrix(averaged_data)
averaged_matrix <- averaged_matrix[rowSums(averaged_matrix) != 0, ]
# Get the column names
col_names <- colnames(averaged_matrix)
```

```
# Function to remove the last character from a string
remove_last_character <- function(x) {
  substr(x, 1, nchar(x) - 1)
}
# Apply the function to all column names and update them
colnames(averaged_matrix) <- sapply(col_names, remove_last_character)
# Set the color palette for the heatmap
color_palette <- colorRampPalette(c("green", "white", "red"))(100)
# Plot the heatmap using pheatmap
p <- pheatmap(t(scale(t(averaged_matrix))),
         color = color_palette,
         clustering_method = "complete",
         cluster_rows = TRUE,
         cluster_cols = TRUE,
         show_rownames = TRUE,
         show_colnames = TRUE,
         fontsize_row = 10)
#ggsave("Chondro.3.CASC20miRs.png", p)
#ggsave("Chondro.3.CASC20miRs.pdf", p)
```

D0 - Filter the ntd object by a specific level of the diff column and generate a heatmap for the filtered data

```{r}
# Generate heatmap for D0_undifferentiated condition only
select <- order(rowMeans(assay(ntd)), decreasing=TRUE)[1:507]
df <- as.data.frame(colData(ntd)[,c("condition","diff")])
```

```
normalized_data <- t(scale(t(assay(ntd))))

p <- pheatmap(normalized_data[select,colData(ntd)$diff == "D0_undifferentiated"],

        cluster_rows=TRUE, show_rownames=FALSE,

        cluster_cols=TRUE, annotation_col=df)

#ggsave("Chondro.3.D0_p_count.pdf", p)

#ggsave("Chondro.3.D0_p_count.png", p)

p
```

D7 - Filter the ntd object by a specific level of the diff column and generate a heatmap for the filtered data

```{r}
# Generate heatmap for D0_undifferentiated condition only

select <- order(rowMeans(assay(ntd)), decreasing=TRUE)[1:507]

df <- as.data.frame(colData(ntd)[,c("condition","diff")])

normalized_data <- t(scale(t(assay(ntd))))

p <- pheatmap(normalized_data[select,colData(ntd)$diff == "D7_chondrogenic"],

        cluster_rows=TRUE, show_rownames=FALSE,

        cluster_cols=TRUE, annotation_col=df)


#ggsave("Chondro.3.D7_p_count.pdf", p)

#ggsave("Chondro.3.D7_p_count.png", p)

p

#save(dds, dds_2, dds_3, vsd, vsd_2, vsd_3, ensemblIDs2miRID2, file = "dds.RData")
```

---

title: "Log2FC osteo miRNA-Seq FFI"

author: "Favour Felix-Ilemhenbhio"

date: "2023-05-10"

output: html_document

#Set up environment

Set working directory

```{r setup, include=FALSE}

knitr::opts_knit$set(root.dir = "X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/GO_KEGG/Log2FC/")

setwd("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/GO_KEGG/Log2FC/")

"X:/sudlab1/General/mirror/miRBase/ensemblIDs2miRbaseID.tsv"

```

#Library

```{r}

library(DESeq2)

library(ggplot2)

library(pheatmap)

library(dplyr)

library(RColorBrewer)

library(knitr)

library(ggforce)

library(tidyverse)

library(ashr)

library(apeglm)

library(EnhancedVolcano)
```

```

#Open files and run DESeqDataSetFromMatrix

```{r}

input <-
read.table("X:/sudlab1/General/projects/Favour_EKT/miRNASeq_20230313/pipeline_seq_favour/counts_first.dir/Primary_resume_counts.txt",

                header = T, sep = "\t")

miRBase_hp2mature <-
read.table("X:/sudlab1/General/mirror/miRBase/miRBase_hp2mature.tsv", header=T)

header <- colnames(input)

header <- sub("^Back", "", header)

colnames(input) <- header

# Remove rows with NAs from miRNA_filter_inputs

input <- input[complete.cases(input), ]

input <- input[rowSums(input[2:ncol(input)]) >= 1,]

input <- as.matrix(input)

# Create a mapping vector from miRBase_ID to Name

mapping <- with(miRBase_hp2mature, setNames(Name, miRBase_ID))

# Rename row names in input using the miRBase_ID2miRID_vector

rownames(input) <- mapping[rownames(input)]

# identify columns to remove

cols_to_remove <- grep("D7", colnames(input))

# remove columns

filter_inputs_2 <- input[, -cols_to_remove]

colnames(filter_inputs_2)

```

#D10 only
```

From here I am running the pairwise comparsions

```{r}
D10_inputs <- filter_inputs_2[, grep("^CTRLD10|^OED10", colnames(filter_inputs_2))]
header <- colnames(D10_inputs)
# Calculate row-wise sums of counts
row_sums <- rowSums(D10_inputs)
# Filter out rows with zero counts
D10_inputs <- D10_inputs[row_sums > 0, ]
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))
samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_10 <- DESeqDataSetFromMatrix(D10_inputs,
```

```
                        colData = samples,
                        design = ~ condition)
dds_10 <- DESeq(dds_10)
vsd_10 <- varianceStabilizingTransformation(dds_10)
```

```{r}
resultsNames(dds_10)
resLFC_D10_only <- results(dds_10, name="condition_OE_vs_CTRL", alpha = 0.05)
# Use the coefficient index in lfcShrink
resLFC_D10_only <- lfcShrink(dds_10, coef="condition_OE_vs_CTRL", res=resLFC_D10_only)
sig_vsd_10 <-vsd_10[is.finite(resLFC_D10_only$padj) & resLFC_D10_only$padj < 0.05,]
assay(sig_vsd_10) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_10))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05
# Convert DESeqResults object to data frame
```

```
resLFC_D10_only_df <- as.data.frame(resLFC_D10_only)

# Exclude missing values in padj column

complete_cases <- !is.na(resLFC_D10_only_df$padj)

# Create a subset of significantly differentially expressed miRNAs

significant_miRNAs_D10_only <- resLFC_D10_only_df[complete_cases &
resLFC_D10_only_df$padj < pvalue_threshold, ]

# Count total and significant differentially expressed miRNAs

total_miRs <- length(resLFC_D10_only$baseMean)  # Total miRNAs

downregulated_miRs <- sum(resLFC_D10_only$log2FoldChange < 0, na.rm = TRUE)  #
Downregulated miRNAs

upregulated_miRs <- sum(resLFC_D10_only$log2FoldChange > 0, na.rm = TRUE)  #
Upregulated miRNAs

# Count significant differentially expressed miRNAs

sig_downregulated_miRs <- sum(resLFC_D10_only$log2FoldChange < 0 &
resLFC_D10_only$padj < 0.05, na.rm = TRUE)  # Significant downregulated miRNAs

sig_upregulated_miRs <- sum(resLFC_D10_only$log2FoldChange > 0 &
resLFC_D10_only$padj < 0.05, na.rm = TRUE)  # Significant upregulated miRNAs

downregulated_miR_name <-
rownames(resLFC_D10_only[complete.cases(resLFC_D10_only) &
resLFC_D10_only$log2FoldChange < 0 & resLFC_D10_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D10_only[complete.cases(resLFC_D10_only)
& resLFC_D10_only$log2FoldChange > 0 & resLFC_D10_only$padj < 0.05, , drop = FALSE])

# Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")
```

```
#cat(upregulated_miR_name, sep = "\n")
```

```{r}
# Extract the expression data from the sig_vsd object
expression_data <- assay(sig_vsd_10)
# Extract sample names
sample_names <- colnames(expression_data)
# Identify unique sample groups
sample_groups <- unique(sub(".$", "", sample_names))
# Calculate average of each group of three samples
averaged_data <- sapply(sample_groups, function(group) {
  group_samples <- sample_names[grep(group, sample_names)]
  rowMeans(expression_data[, group_samples])
})
# Create a new matrix with averaged column data
averaged_matrix <- as.matrix(averaged_data)
# Set the color palette for the heatmap
color_palette <- colorRampPalette(c("green", "white", "red"))(100)
# Plot the heatmap using pheatmap
pheatmap(t(scale(t(averaged_matrix))),
         color = color_palette,
         clustering_method = "complete",
         cluster_rows = TRUE,
         cluster_cols = TRUE,
         show_rownames = FALSE,
         show_colnames = TRUE)
```

```

```{r fig.height=7}

miRNA_names <- rownames(resLFC_D10_only)[abs(resLFC_D10_only$log2FoldChange) > 1]

p <- EnhancedVolcano(resLFC_D10_only,

  lab = ifelse(rownames(resLFC_D10_only) %in% miRNA_names,
rownames(resLFC_D10_only), ''),

   x = 'log2FoldChange',

   y = 'padj',

   title = 'D10 - *CASC20* OE vs Wild type',

   pCutoff = 0.05,

   FCcutoff = 0,

   pointSize = 3.0,

   labSize = 6.0,

   col=c('black', 'black', 'black', 'red'),

   colAlpha = 1,

   legendPosition = 'right',

   legendLabSize = 16,

   legendIconSize = 5.0,

   legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),

  max.overlaps = Inf)

ggsave("D10_Osteo_OEvsCTRL.png", p, device = "png")

ggsave("D10_Osteo_OEvsCTRL.pdf", p, device = "pdf")

p

```

#D20

```{r}

```
D20_inputs <- filter_inputs_2[, grep("^CTRLD20|^OED20", colnames(filter_inputs_2))]
header <- colnames(D20_inputs)

# Calculate row-wise sums of counts
row_sums <- rowSums(D20_inputs)
# Filter out rows with zero counts
D20_inputs <- D20_inputs[row_sums > 0,]
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))
samples <- samples[complete.cases(samples),]
rownames(samples) <- NULL
Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_D20 <- DESeqDataSetFromMatrix(D20_inputs,
                                  colData = samples,
```

```
                        design = ~ condition)
dds_D20 <- DESeq(dds_D20)
vsd_D20 <- varianceStabilizingTransformation(dds_D20)
resultsNames(dds_D20)
resLFC_D20_only <- results(dds_D20, name="condition_OE_vs_CTRL", alpha = 0.05)
# Apply lfcShrink
#resLFC_D20_only <- lfcShrink(dds_D20, coef="condition_OE_vs_CTRL",
res=resLFC_D20_only)
sig_vsd_D20 <-vsd_D20[is.finite(resLFC_D20_only$padj) & resLFC_D20_only$padj < 0.05,]
assay(sig_vsd_D20) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_D20))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05
# Convert DESeqResults object to data frame
resLFC_D20_only_df <- as.data.frame(resLFC_D20_only)
# Exclude missing values in padj column
complete_cases <- !is.na(resLFC_D20_only_df$padj)
# Create a subset of significantly differentially expressed miRNAs
```

```
significant_miRNAs_D20_only <- resLFC_D20_only_df[complete_cases &
resLFC_D20_only_df$padj < pvalue_threshold, ]

# Count total and significant differentially expressed miRNAs

total_miRs <- length(resLFC_D20_only$baseMean)  # Total miRNAs

downregulated_miRs <- sum(resLFC_D20_only$log2FoldChange < 0, na.rm = TRUE)  #
Downregulated miRNAs

upregulated_miRs <- sum(resLFC_D20_only$log2FoldChange > 0, na.rm = TRUE)  #
Upregulated miRNAs

# Count significant differentially expressed miRNAs

sig_downregulated_miRs <- sum(resLFC_D20_only$log2FoldChange < 0 &
resLFC_D20_only$padj < 0.05, na.rm = TRUE)  # Significant downregulated miRNAs

sig_upregulated_miRs <- sum(resLFC_D20_only$log2FoldChange > 0 &
resLFC_D20_only$padj < 0.05, na.rm = TRUE)  # Significant upregulated miRNAs


downregulated_miR_name <-
rownames(resLFC_D20_only[complete.cases(resLFC_D20_only) &
resLFC_D20_only$log2FoldChange < 0 & resLFC_D20_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D20_only[complete.cases(resLFC_D20_only)
& resLFC_D20_only$log2FoldChange > 0 & resLFC_D20_only$padj < 0.05, , drop = FALSE])

# Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")

#cat(upregulated_miR_name, sep = "\n")
```

```{r}
```

```
miRNA_names <- rownames(resLFC_D20_only)[abs(resLFC_D20_only$log2FoldChange) > 1]
p <- EnhancedVolcano(resLFC_D20_only,
  lab = rownames(resLFC_D20_only),
   x = 'log2FoldChange',
   y = 'padj',
   title = 'D20 - CASC20 OE vs Wild type',
   pCutoff = 0.05,
   FCcutoff = 0,
   pointSize = 3.0,
   labSize = 6.0,
   col=c('black', 'black', 'black', 'red'),
   colAlpha = 1,
   legendPosition = 'right',
   legendLabSize = 16,
   legendIconSize = 5.0,
   legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
  max.overlaps = Inf)
ggsave("D20_Osteo_OEvsCTRL.png", p, device = "png")
ggsave("D20_Osteo_OEvsCTRL.pdf", p, device = "pdf")
p
```
#D0
```{r}
D0_inputs <- filter_inputs_2[, grep("^CTRLD0|^OED0", colnames(filter_inputs_2))]
header <- colnames(D0_inputs)
# Calculate row-wise sums of counts
```

```
row_sums <- rowSums(D0_inputs)
# Filter out rows with zero counts
D0_inputs <- D0_inputs[row_sums > 0, ]
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))
samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_D0 <- DESeqDataSetFromMatrix(D0_inputs,
                                 colData = samples,
                                 design = ~ condition)
dds_D0 <- DESeq(dds_D0)
vsd_D0 <- varianceStabilizingTransformation(dds_D0)
resultsNames(dds_D0)
```

```
resLFC_D0_only <- results(dds_D0, name="condition_OE_vs_CTRL", alpha = 0.05)
# Apply lfcShrink
resLFC_D0_only <- lfcShrink(dds_D0, coef="condition_OE_vs_CTRL", res=resLFC_D0_only)
sig_vsd_D0 <-vsd_D0[is.finite(resLFC_D0_only$padj) & resLFC_D0_only$padj < 0.05,]
assay(sig_vsd_D0) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_D0))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05

# Convert DESeqResults object to data frame
resLFC_D0_only_df <- as.data.frame(resLFC_D0_only)
# Exclude missing values in padj column
complete_cases <- !is.na(resLFC_D0_only_df$padj)
# Create a subset of significantly differentially expressed miRNAs
significant_miRNAs_D0_only <- resLFC_D0_only_df[complete_cases &
resLFC_D0_only_df$padj < pvalue_threshold, ]
# Count total and significant differentially expressed miRNAs
total_miRs <- length(resLFC_D0_only$baseMean)  # Total miRNAs
```

```
downregulated_miRs <- sum(resLFC_D0_only$log2FoldChange < 0, na.rm = TRUE)  # Downregulated miRNAs

upregulated_miRs <- sum(resLFC_D0_only$log2FoldChange > 0, na.rm = TRUE)  # Upregulated miRNAs

# Count significant differentially expressed miRNAs

sig_downregulated_miRs <- sum(resLFC_D0_only$log2FoldChange < 0 & resLFC_D0_only$padj < 0.05, na.rm = TRUE)  # Significant downregulated miRNAs

sig_upregulated_miRs <- sum(resLFC_D0_only$log2FoldChange > 0 & resLFC_D0_only$padj < 0.05, na.rm = TRUE)  # Significant upregulated miRNAs

downregulated_miR_name <- rownames(resLFC_D0_only[complete.cases(resLFC_D0_only) & resLFC_D0_only$log2FoldChange < 0 & resLFC_D0_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D0_only[complete.cases(resLFC_D0_only) & resLFC_D0_only$log2FoldChange > 0 & resLFC_D0_only$padj < 0.05, , drop = FALSE])

# Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")

#cat(upregulated_miR_name, sep = "\n")
```

```{r}

# Extract the expression data from the sig_vsd object

expression_data <- assay(sig_vsd_D0)

# Extract sample names

sample_names <- colnames(expression_data)

# Identify unique sample groups
```

```
sample_groups <- unique(sub(".$", "", sample_names))
# Calculate average of each group of three samples
averaged_data <- sapply(sample_groups, function(group) {
  group_samples <- sample_names[grep(group, sample_names)]
  rowMeans(expression_data[, group_samples])
})
# Create a new matrix with averaged column data
averaged_matrix <- as.matrix(averaged_data)

# Set the color palette for the heatmap
color_palette <- colorRampPalette(c("green", "white", "red"))(100)

# Plot the heatmap using pheatmap
pheatmap(t(scale(t(averaged_matrix))),
         color = color_palette,
         clustering_method = "complete",
         cluster_rows = TRUE,
         cluster_cols = TRUE,
         show_rownames = FALSE,
         show_colnames = TRUE)
```

```{r}
miRNA_names <- rownames(resLFC_D0_only)[abs(resLFC_D0_only$log2FoldChange) > 1]

p <- EnhancedVolcano(resLFC_D0_only,
```

```
  lab = ifelse(rownames(resLFC_D0_only) %in% miRNA_names,
rownames(resLFC_D10_only), ''),
   x = 'log2FoldChange',
   y = 'padj',
   title = 'D0 - CASC20 OE vs Wild type',
   pCutoff = 0.05,
   FCcutoff = 0,
   pointSize = 3.0,
   labSize = 6.0,
   col=c('black', 'black', 'black', 'red'),
   colAlpha = 1,
   legendPosition = 'right',
   legendLabSize = 16,
   legendIconSize = 5.0,
   legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
  max.overlaps = Inf)
ggsave("D0_OEvsCTRL.png", p, device = "png")
ggsave("D0_OEvsCTRL.pdf", p, device = "pdf")
p
```

#D7 Chondrogenic

```{r}
#filter_inputs
cols_to_remove <- grep("D10|D20", colnames(input))
filter_inputs_3 <- input[, -cols_to_remove]
header <- colnames(filter_inputs_3)
```

```
D7_inputs <- filter_inputs_3[, grep("^CTRLD7|^OED7", colnames(filter_inputs_3))]
header <- colnames(D7_inputs)
# Calculate row-wise sums of counts
row_sums <- rowSums(D7_inputs)
# Filter out rows with zero counts
D7_inputs <- D7_inputs[row_sums > 0, ]
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))

samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL

# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_D7 <- DESeqDataSetFromMatrix(D7_inputs,
```

```
                    colData = samples,
                    design = ~ condition)
dds_D7 <- DESeq(dds_D7)
vsd_D7 <- varianceStabilizingTransformation(dds_D7)
resultsNames(dds_D7)
resLFC_D7_only <- results(dds_D7, name="condition_OE_vs_CTRL", alpha = 0.05)
# Apply lfcShrink
resLFC_D7_only <- lfcShrink(dds_D7, coef="condition_OE_vs_CTRL", res=resLFC_D7_only)
sig_vsd_D7 <-vsd_D7[is.finite(resLFC_D7_only$padj) & resLFC_D7_only$padj < 0.05,]
assay(sig_vsd_D7) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_D7))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05
# Convert DESeqResults object to data frame
resLFC_D7_only_df <- as.data.frame(resLFC_D7_only)
# Exclude missing values in padj column
complete_cases <- !is.na(resLFC_D7_only_df$padj)
# Create a subset of significantly differentially expressed miRNAs
```

```
significant_miRNAs_D7_only <- resLFC_D7_only_df[complete_cases &
resLFC_D7_only_df$padj < pvalue_threshold, ]

# Count total and significant differentially expressed miRNAs

total_miRs <- length(resLFC_D7_only$baseMean)  # Total miRNAs

downregulated_miRs <- sum(resLFC_D7_only$log2FoldChange < 0, na.rm = TRUE)  #
Downregulated miRNAs

upregulated_miRs <- sum(resLFC_D7_only$log2FoldChange > 0, na.rm = TRUE)  #
Upregulated miRNAs


# Count significant differentially expressed miRNAs

sig_downregulated_miRs <- sum(resLFC_D7_only$log2FoldChange < 0 &
resLFC_D7_only$padj < 0.05, na.rm = TRUE)  # Significant downregulated miRNAs

sig_upregulated_miRs <- sum(resLFC_D7_only$log2FoldChange > 0 & resLFC_D7_only$padj
< 0.05, na.rm = TRUE)  # Significant upregulated miRNAs

downregulated_miR_name <- rownames(resLFC_D7_only[complete.cases(resLFC_D7_only)
& resLFC_D7_only$log2FoldChange < 0 & resLFC_D7_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D7_only[complete.cases(resLFC_D7_only) &
resLFC_D7_only$log2FoldChange > 0 & resLFC_D7_only$padj < 0.05, , drop = FALSE])

# Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")

#cat(upregulated_miR_name, sep = "\n")
```

```{r}

miRNA_names <- rownames(resLFC_D7_only)[abs(resLFC_D7_only$log2FoldChange) > 1]
```

```
p <- EnhancedVolcano(resLFC_D7_only,
  lab = rownames(resLFC_D7_only),
   x = 'log2FoldChange',
   y = 'padj',
   title = 'D7 - CASC20 OE vs Wild type',
   pCutoff = 0.05,
   FCcutoff = 0,
   pointSize = 3.0,
   labSize = 6.0,
   col=c('black', 'black', 'black', 'red'),
   colAlpha = 1,
   legendPosition = 'right',
   legendLabSize = 16,
   legendIconSize = 5.0,
   legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
  max.overlaps = Inf)

ggsave("D7_Chondro_OEvsCTRL.png", p, device = "png")
ggsave("D7_Chondro_OEvsCTRL.pdf", p, device = "pdf")
p
```

#Save

```{r}
write.table(resLFC_D0_only_df, file = "resLFC_D0_only.tsv", sep = "\t", row.names = TRUE)
```

```
write.table(resLFC_D10_only_df, file = "resLFC_D10_only.tsv", sep = "\t", row.names = TRUE)

write.table(resLFC_D20_only_df, file = "resLFC_D20_only.tsv", sep = "\t", row.names = TRUE)

write.table(resLFC_D7_only_df, file = "resLFC_D7_only.tsv", sep = "\t", row.names = TRUE)
```

9.2.2 Log2FC

#Library

```{r}

library(DESeq2)

library(ggplot2)

library(pheatmap)

library(dplyr)

library(RColorBrewer)

library(knitr)

library(ggforce)

library(tidyverse)

library(ashr)

library(apeglm)

library(EnhancedVolcano)

```

#Open files and run DESeqDataSetFromMatrix

```{r}

input <- read.table("X:/sudlab1/General/projects/Favour_EKT/miRNASeq_20230313/pipeline_seq_favour/counts_first.dir/Primary_resume_counts.txt",

header = T, sep = "\t")

miRBase_hp2mature <- read.table("X:/sudlab1/General/mirror/miRBase/miRBase_hp2mature.tsv", header=T)

header <- colnames(input)

header <- sub("^Back", "", header)
```

```
colnames(input) <- header
# Remove rows with NAs from miRNA_filter_inputs
input <- input[complete.cases(input), ]
input <- input[rowSums(input[2:ncol(input)]) >= 1,]
input <- as.matrix(input)
# Create a mapping vector from miRBase_ID to Name
mapping <- with(miRBase_hp2mature, setNames(Name, miRBase_ID))
# Rename row names in input using the miRBase_ID2miRID_vector
rownames(input) <- mapping[rownames(input)]
# identify columns to remove
cols_to_remove <- grep("D7", colnames(input))
# remove columns
filter_inputs_2 <- input[, -cols_to_remove]
colnames(filter_inputs_2)
```

#D10 only

From here I am running the pairwise comparsions

```{r}
D10_inputs <- filter_inputs_2[, grep("^CTRLD10|^OED10", colnames(filter_inputs_2))]
header <- colnames(D10_inputs)

# Calculate row-wise sums of counts
row_sums <- rowSums(D10_inputs)

# Filter out rows with zero counts
D10_inputs <- D10_inputs[row_sums > 0, ]
```

```
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))

samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_10 <- DESeqDataSetFromMatrix(D10_inputs,
                                 colData = samples,
                                 design = ~ condition)
dds_10 <- DESeq(dds_10)
vsd_10 <- varianceStabilizingTransformation(dds_10)
```

```{r}
```

```
resultsNames(dds_10)
resLFC_D10_only <- results(dds_10, name="condition_OE_vs_CTRL", alpha = 0.05)
# Use the coefficient index in lfcShrink
resLFC_D10_only <- lfcShrink(dds_10, coef="condition_OE_vs_CTRL", res=resLFC_D10_only)
sig_vsd_10 <-vsd_10[is.finite(resLFC_D10_only$padj) & resLFC_D10_only$padj < 0.05,]
assay(sig_vsd_10) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_10))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05
# Convert DESeqResults object to data frame
resLFC_D10_only_df <- as.data.frame(resLFC_D10_only)
# Exclude missing values in padj column
complete_cases <- !is.na(resLFC_D10_only_df$padj)
# Create a subset of significantly differentially expressed miRNAs
significant_miRNAs_D10_only <- resLFC_D10_only_df[complete_cases & resLFC_D10_only_df$padj < pvalue_threshold, ]
# Count total and significant differentially expressed miRNAs
total_miRs <- length(resLFC_D10_only$baseMean)  # Total miRNAs
```

```
downregulated_miRs <- sum(resLFC_D10_only$log2FoldChange < 0, na.rm = TRUE) # Downregulated miRNAs

upregulated_miRs <- sum(resLFC_D10_only$log2FoldChange > 0, na.rm = TRUE) # Upregulated miRNAs

# Count significant differentially expressed miRNAs

sig_downregulated_miRs <- sum(resLFC_D10_only$log2FoldChange < 0 & resLFC_D10_only$padj < 0.05, na.rm = TRUE) # Significant downregulated miRNAs

sig_upregulated_miRs <- sum(resLFC_D10_only$log2FoldChange > 0 & resLFC_D10_only$padj < 0.05, na.rm = TRUE) # Significant upregulated miRNAs


downregulated_miR_name <- rownames(resLFC_D10_only[complete.cases(resLFC_D10_only) & resLFC_D10_only$log2FoldChange < 0 & resLFC_D10_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D10_only[complete.cases(resLFC_D10_only) & resLFC_D10_only$log2FoldChange > 0 & resLFC_D10_only$padj < 0.05, , drop = FALSE])

# Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")

#cat(upregulated_miR_name, sep = "\n")

```

```{r}

# Extract the expression data from the sig_vsd object

expression_data <- assay(sig_vsd_10)

# Extract sample names

sample_names <- colnames(expression_data)
```

```
# Identify unique sample groups
sample_groups <- unique(sub(".$", "", sample_names))
# Calculate average of each group of three samples
averaged_data <- sapply(sample_groups, function(group) {
  group_samples <- sample_names[grep(group, sample_names)]
  rowMeans(expression_data[, group_samples])
})

# Create a new matrix with averaged column data
averaged_matrix <- as.matrix(averaged_data)
# Set the color palette for the heatmap
color_palette <- colorRampPalette(c("green", "white", "red"))(100)
# Plot the heatmap using pheatmap
pheatmap(t(scale(t(averaged_matrix))),
         color = color_palette,
         clustering_method = "complete",
         cluster_rows = TRUE,
         cluster_cols = TRUE,
         show_rownames = FALSE,
         show_colnames = TRUE)
```

```{r fig.height=7}
miRNA_names <- rownames(resLFC_D10_only)[abs(resLFC_D10_only$log2FoldChange) > 1]
p <- EnhancedVolcano(resLFC_D10_only,
  lab = ifelse(rownames(resLFC_D10_only) %in% miRNA_names, rownames(resLFC_D10_only), ''),
```

```
    x = 'log2FoldChange',
    y = 'padj',
    title = 'D10 - CASC20 OE vs Wild type',
    pCutoff = 0.05,
    FCcutoff = 0,
    pointSize = 3.0,
    labSize = 6.0,
    col=c('black', 'black', 'black', 'red'),
    colAlpha = 1,
    legendPosition = 'right',
    legendLabSize = 16,
    legendIconSize = 5.0,
    legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
   max.overlaps = Inf)
ggsave("D10_Osteo_OEvsCTRL.png", p, device = "png")
ggsave("D10_Osteo_OEvsCTRL.pdf", p, device = "pdf")
p
```

#D20

```{r}
D20_inputs <- filter_inputs_2[, grep("^CTRLD20|^OED20", colnames(filter_inputs_2))]
header <- colnames(D20_inputs)
# Calculate row-wise sums of counts
row_sums <- rowSums(D20_inputs)
# Filter out rows with zero counts
D20_inputs <- D20_inputs[row_sums > 0, ]
```

```
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
                                                       ifelse(grepl("D20", header), "D20", NA)))),
                      diff = ifelse(grepl("D0", header), "undifferentiated",
                                    ifelse(grepl("D7", header), "chondrogenic",
                                           ifelse(grepl("D10", header), "osteogenic",
                                                  ifelse(grepl("D20", header), "osteogenic", NA)))))

samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_D20 <- DESeqDataSetFromMatrix(D20_inputs,
                                  colData = samples,
                                  design = ~ condition)
dds_D20 <- DESeq(dds_D20)
vsd_D20 <- varianceStabilizingTransformation(dds_D20)
resultsNames(dds_D20)
resLFC_D20_only <- results(dds_D20, name="condition_OE_vs_CTRL", alpha = 0.05)
# Apply lfcShrink
```

```
#resLFC_D20_only <- lfcShrink(dds_D20, coef="condition_OE_vs_CTRL", res=resLFC_D20_only)

sig_vsd_D20 <-vsd_D20[is.finite(resLFC_D20_only$padj) & resLFC_D20_only$padj < 0.05,]

assay(sig_vsd_D20) %>% as.data.frame() %>%

  rownames_to_column("gene_id") %>%

  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%

  inner_join(as.data.frame(colData(vsd_D20))) %>%

  ggplot() + aes(x=condition, y=expr, col=condition) +

  geom_point() +

  facet_wrap(~gene_id, scales = "free_y") +

  theme(axis.text.x = element_text(angle=90))
```

```{r}

# Set the p-value threshold for significance

pvalue_threshold <- 0.05

# Convert DESeqResults object to data frame

resLFC_D20_only_df <- as.data.frame(resLFC_D20_only)

# Exclude missing values in padj column

complete_cases <- !is.na(resLFC_D20_only_df$padj)

# Create a subset of significantly differentially expressed miRNAs

significant_miRNAs_D20_only <- resLFC_D20_only_df[complete_cases & resLFC_D20_only_df$padj < pvalue_threshold, ]

# Count total and significant differentially expressed miRNAs

total_miRs <- length(resLFC_D20_only$baseMean)  # Total miRNAs

downregulated_miRs <- sum(resLFC_D20_only$log2FoldChange < 0, na.rm = TRUE)  # Downregulated miRNAs
```

upregulated_miRs <- sum(resLFC_D20_only$log2FoldChange > 0, na.rm = TRUE) # Upregulated miRNAs

Count significant differentially expressed miRNAs

sig_downregulated_miRs <- sum(resLFC_D20_only$log2FoldChange < 0 & resLFC_D20_only$padj < 0.05, na.rm = TRUE) # Significant downregulated miRNAs

sig_upregulated_miRs <- sum(resLFC_D20_only$log2FoldChange > 0 & resLFC_D20_only$padj < 0.05, na.rm = TRUE) # Significant upregulated miRNAs

downregulated_miR_name <- rownames(resLFC_D20_only[complete.cases(resLFC_D20_only) & resLFC_D20_only$log2FoldChange < 0 & resLFC_D20_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D20_only[complete.cases(resLFC_D20_only) & resLFC_D20_only$log2FoldChange > 0 & resLFC_D20_only$padj < 0.05, , drop = FALSE])

Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")

#cat(upregulated_miR_name, sep = "\n")

```

```{r}

miRNA_names <- rownames(resLFC_D20_only)[abs(resLFC_D20_only$log2FoldChange) > 1]

p <- EnhancedVolcano(resLFC_D20_only,

  lab = rownames(resLFC_D20_only),

   x = 'log2FoldChange',

   y = 'padj',

   title = 'D20 - *CASC20* OE vs Wild type',

```
    pCutoff = 0.05,
    FCcutoff = 0,
    pointSize = 3.0,
    labSize = 6.0,
    col=c('black', 'black', 'black', 'red'),
    colAlpha = 1,
    legendPosition = 'right',
    legendLabSize = 16,
    legendIconSize = 5.0,
    legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
   max.overlaps = Inf)
ggsave("D20_Osteo_OEvsCTRL.png", p, device = "png")
ggsave("D20_Osteo_OEvsCTRL.pdf", p, device = "pdf")
p
```

#D0

```{r}
D0_inputs <- filter_inputs_2[, grep("^CTRLD0|^OED0", colnames(filter_inputs_2))]
header <- colnames(D0_inputs)
# Calculate row-wise sums of counts
row_sums <- rowSums(D0_inputs)
# Filter out rows with zero counts
D0_inputs <- D0_inputs[row_sums > 0, ]
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
```

```
                timepoint = ifelse(grepl("D0", header), "D0",
                                   ifelse(grepl("D7", header), "D7",
                                          ifelse(grepl("D10", header), "D10",
                                                 ifelse(grepl("D20", header), "D20", NA)))),
                diff = ifelse(grepl("D0", header), "undifferentiated",
                              ifelse(grepl("D7", header), "chondrogenic",
                                     ifelse(grepl("D10", header), "osteogenic",
                                            ifelse(grepl("D20", header), "osteogenic", NA)))))
samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_D0 <- DESeqDataSetFromMatrix(D0_inputs,
                         colData = samples,
                         design = ~ condition)
dds_D0 <- DESeq(dds_D0)
vsd_D0 <- varianceStabilizingTransformation(dds_D0)
resultsNames(dds_D0)
resLFC_D0_only <- results(dds_D0, name="condition_OE_vs_CTRL", alpha = 0.05)
# Apply lfcShrink
resLFC_D0_only <- lfcShrink(dds_D0, coef="condition_OE_vs_CTRL", res=resLFC_D0_only)
sig_vsd_D0 <-vsd_D0[is.finite(resLFC_D0_only$padj) & resLFC_D0_only$padj < 0.05,]
assay(sig_vsd_D0) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
```

```
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_D0))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05
# Convert DESeqResults object to data frame
resLFC_D0_only_df <- as.data.frame(resLFC_D0_only)
# Exclude missing values in padj column
complete_cases <- !is.na(resLFC_D0_only_df$padj)
# Create a subset of significantly differentially expressed miRNAs
significant_miRNAs_D0_only <- resLFC_D0_only_df[complete_cases & resLFC_D0_only_df$padj < pvalue_threshold, ]
# Count total and significant differentially expressed miRNAs
total_miRs <- length(resLFC_D0_only$baseMean)  # Total miRNAs
downregulated_miRs <- sum(resLFC_D0_only$log2FoldChange < 0, na.rm = TRUE)  # Downregulated miRNAs
upregulated_miRs <- sum(resLFC_D0_only$log2FoldChange > 0, na.rm = TRUE)  # Upregulated miRNAs
# Count significant differentially expressed miRNAs
sig_downregulated_miRs <- sum(resLFC_D0_only$log2FoldChange < 0 & resLFC_D0_only$padj < 0.05, na.rm = TRUE)  # Significant downregulated miRNAs
sig_upregulated_miRs <- sum(resLFC_D0_only$log2FoldChange > 0 & resLFC_D0_only$padj < 0.05, na.rm = TRUE)  # Significant upregulated miRNAs
```

```
downregulated_miR_name <- rownames(resLFC_D0_only[complete.cases(resLFC_D0_only) & resLFC_D0_only$log2FoldChange < 0 & resLFC_D0_only$padj < 0.05, , drop = FALSE])

upregulated_miR_name <- rownames(resLFC_D0_only[complete.cases(resLFC_D0_only) & resLFC_D0_only$log2FoldChange > 0 & resLFC_D0_only$padj < 0.05, , drop = FALSE])

# Print the summary

cat("Total miRNAs:", total_miRs, "\n")

cat("Downregulated miRNAs:", downregulated_miRs, "\n")

cat("Upregulated miRNAs:", upregulated_miRs, "\n")

cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")

cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")

#cat(downregulated_miR_name, sep = "\n")

#cat(upregulated_miR_name, sep = "\n")
```

```{r}
# Extract the expression data from the sig_vsd object

expression_data <- assay(sig_vsd_D0)

# Extract sample names

sample_names <- colnames(expression_data)

# Identify unique sample groups

sample_groups <- unique(sub(".$", "", sample_names))

# Calculate average of each group of three samples

averaged_data <- sapply(sample_groups, function(group) {

  group_samples <- sample_names[grep(group, sample_names)]

  rowMeans(expression_data[, group_samples])

})

# Create a new matrix with averaged column data

averaged_matrix <- as.matrix(averaged_data)
```

```
# Set the color palette for the heatmap
color_palette <- colorRampPalette(c("green", "white", "red"))(100)
# Plot the heatmap using pheatmap
pheatmap(t(scale(t(averaged_matrix))),
         color = color_palette,
         clustering_method = "complete",
         cluster_rows = TRUE,
         cluster_cols = TRUE,
         show_rownames = FALSE,
         show_colnames = TRUE)
```

```{r}
miRNA_names <- rownames(resLFC_D0_only)[abs(resLFC_D0_only$log2FoldChange) > 1]
p <- EnhancedVolcano(resLFC_D0_only,
  lab = ifelse(rownames(resLFC_D0_only) %in% miRNA_names, rownames(resLFC_D10_only), ''),
   x = 'log2FoldChange',
   y = 'padj',
   title = 'D0 - *CASC20* OE vs Wild type',
   pCutoff = 0.05,
   FCcutoff = 0,
   pointSize = 3.0,
   labSize = 6.0,
   col=c('black', 'black', 'black', 'red'),
   colAlpha = 1,
   legendPosition = 'right',


```

```
   legendLabSize = 16,
   legendIconSize = 5.0,
   legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
  max.overlaps = Inf)
ggsave("D0_OEvsCTRL.png", p, device = "png")
ggsave("D0_OEvsCTRL.pdf", p, device = "pdf")
p
```

#D7 Chondrogenic

```{r}
#filter_inputs
cols_to_remove <- grep("D10|D20", colnames(input))
filter_inputs_3 <- input[, -cols_to_remove]
header <- colnames(filter_inputs_3)
D7_inputs <- filter_inputs_3[, grep("^CTRLD7|^OED7", colnames(filter_inputs_3))]
header <- colnames(D7_inputs)
# Calculate row-wise sums of counts
row_sums <- rowSums(D7_inputs)
# Filter out rows with zero counts
D7_inputs <- D7_inputs[row_sums > 0, ]
# Extract sample information
samples <- data.frame(samples = header,
                      condition = ifelse(grepl("^CTR", header), "CTRL", "OE"),
                      timepoint = ifelse(grepl("D0", header), "D0",
                                         ifelse(grepl("D7", header), "D7",
                                                ifelse(grepl("D10", header), "D10",
```

```
                                    ifelse(grepl("D20", header), "D20", NA))),
                diff = ifelse(grepl("D0", header), "undifferentiated",
                              ifelse(grepl("D7", header), "chondrogenic",
                                     ifelse(grepl("D10", header), "osteogenic",
                                            ifelse(grepl("D20", header), "osteogenic", NA)))))
samples <- samples[complete.cases(samples), ]
rownames(samples) <- NULL
# Convert columns to factors and make them unordered
samples$condition <- factor(samples$condition, levels = c("CTRL", "OE"), ordered = FALSE)
samples <- samples %>% select(-c(timepoint, diff))
#base specify it
dds_D7 <- DESeqDataSetFromMatrix(D7_inputs,
                                 colData = samples,
                                 design = ~ condition)
dds_D7 <- DESeq(dds_D7)
vsd_D7 <- varianceStabilizingTransformation(dds_D7)
resultsNames(dds_D7)
resLFC_D7_only <- results(dds_D7, name="condition_OE_vs_CTRL", alpha = 0.05)
# Apply lfcShrink
resLFC_D7_only <- lfcShrink(dds_D7, coef="condition_OE_vs_CTRL", res=resLFC_D7_only)
sig_vsd_D7 <-vsd_D7[is.finite(resLFC_D7_only$padj) & resLFC_D7_only$padj < 0.05,]
assay(sig_vsd_D7) %>% as.data.frame() %>%
  rownames_to_column("gene_id") %>%
  pivot_longer(-gene_id, names_to="samples", values_to="expr") %>%
  inner_join(as.data.frame(colData(vsd_D7))) %>%
  ggplot() + aes(x=condition, y=expr, col=condition) +
```

```
  geom_point() +
  facet_wrap(~gene_id, scales = "free_y") +
  theme(axis.text.x = element_text(angle=90))
```

```{r}
# Set the p-value threshold for significance
pvalue_threshold <- 0.05
# Convert DESeqResults object to data frame
resLFC_D7_only_df <- as.data.frame(resLFC_D7_only)
# Exclude missing values in padj column
complete_cases <- !is.na(resLFC_D7_only_df$padj)
# Create a subset of significantly differentially expressed miRNAs
significant_miRNAs_D7_only <- resLFC_D7_only_df[complete_cases & resLFC_D7_only_df$padj < pvalue_threshold, ]
# Count total and significant differentially expressed miRNAs
total_miRs <- length(resLFC_D7_only$baseMean)  # Total miRNAs
downregulated_miRs <- sum(resLFC_D7_only$log2FoldChange < 0, na.rm = TRUE)  # Downregulated miRNAs
upregulated_miRs <- sum(resLFC_D7_only$log2FoldChange > 0, na.rm = TRUE)  # Upregulated miRNAs
# Count significant differentially expressed miRNAs
sig_downregulated_miRs <- sum(resLFC_D7_only$log2FoldChange < 0 & resLFC_D7_only$padj < 0.05, na.rm = TRUE)  # Significant downregulated miRNAs
sig_upregulated_miRs <- sum(resLFC_D7_only$log2FoldChange > 0 & resLFC_D7_only$padj < 0.05, na.rm = TRUE)  # Significant upregulated miRNAs
downregulated_miR_name <- rownames(resLFC_D7_only[complete.cases(resLFC_D7_only) & resLFC_D7_only$log2FoldChange < 0 & resLFC_D7_only$padj < 0.05, , drop = FALSE])
upregulated_miR_name <- rownames(resLFC_D7_only[complete.cases(resLFC_D7_only) & resLFC_D7_only$log2FoldChange > 0 & resLFC_D7_only$padj < 0.05, , drop = FALSE])
```

```
# Print the summary
cat("Total miRNAs:", total_miRs, "\n")
cat("Downregulated miRNAs:", downregulated_miRs, "\n")
cat("Upregulated miRNAs:", upregulated_miRs, "\n")
cat("Significant downregulated miRNAs:", sig_downregulated_miRs, "\n")
cat("Significant upregulated miRNAs:", sig_upregulated_miRs, "\n")
#cat(downregulated_miR_name, sep = "\n")
#cat(upregulated_miR_name, sep = "\n")
```

```{r}
miRNA_names <- rownames(resLFC_D7_only)[abs(resLFC_D7_only$log2FoldChange) > 1]
p <- EnhancedVolcano(resLFC_D7_only,
  lab = rownames(resLFC_D7_only),
   x = 'log2FoldChange',
   y = 'padj',
   title = 'D7 - *CASC20* OE vs Wild type',
   pCutoff = 0.05,
   FCcutoff = 0,
   pointSize = 3.0,
   labSize = 6.0,
   col=c('black', 'black', 'black', 'red'),
   colAlpha = 1,
   legendPosition = 'right',
   legendLabSize = 16,
   legendIconSize = 5.0,
   legendLabels=c('Not sig.','Log2FC','padj','padj & Log2FC'),
```

```
  max.overlaps = Inf)
ggsave("D7_Chondro_OEvsCTRL.png", p, device = "png")
ggsave("D7_Chondro_OEvsCTRL.pdf", p, device = "pdf")
p
```

#Save

```{r}
write.table(resLFC_D0_only_df, file = "resLFC_D0_only.tsv", sep = "\t", row.names = TRUE)
write.table(resLFC_D10_only_df, file = "resLFC_D10_only.tsv", sep = "\t", row.names = TRUE)
write.table(resLFC_D20_only_df, file = "resLFC_D20_only.tsv", sep = "\t", row.names = TRUE)
write.table(resLFC_D7_only_df, file = "resLFC_D7_only.tsv", sep = "\t", row.names = TRUE)
```

9.2.3 RISmed

```{r}
# Run RISmed analysis ----
library(plyr)
library(dplyr)
library (RISmed)
#call dataframe
first_column <- c(Gene$V1)
search = c("bone", "cartilage")
pubmed = function(x){
 d = as.data.frame(x)
 colnames(d) = "geneName"
 name = as.character(d$geneName)
 print (name)
 res1 <- EUtilsSummary(name,
          type = "esearch",
          db = "pubmed",
          datetype = "pdat",
          mindate = 2000,
          maxdate = 2023)
 fetch <- EUtilsGet(res1, type = "efetch", db = "pubmed")
 abstracts <- data.frame(title = fetch@ArticleTitle,
           abstract = fetch@AbstractText,
           journal = fetch@Title,
           DOI = fetch@PMID,
           year = fetch@YearPubmed)
```

```
 abstracts <- abstracts %>% mutate(abstract = as.character(abstract))
 g <- abstracts[grepl(paste(search, collapse="|"), abstracts$abstract),]
 g1 <- g$DOI %>% list
 abstracts <- abstracts %>%
  mutate(DOI = as.character(DOI))
 mirnas = as.data.frame(name)
 write.csv (abstracts,paste0(name,".csv"))
 mirnas$count = length(g$title)
 mirnas$DOI = toString(g$DOI)
 date_time<-Sys.time()
 while((as.numeric(Sys.time()) - as.numeric(date_time))<2.5){}
 return (mirnas)
}
m = lapply(first_column, pubmed)
df_m = do.call(rbind,m)
```

9.2.4 GOSeq

This notebook performs GOseq analysis from DEseq2 LFC results.

Getting UTR length and GC, IDs and LFC results.

```{r input, include=FALSE}

library(org.Hs.eg.db)

library(goseq)

library(tidyverse)

library(dplyr)

#Loading table of UTR features I generated previously

utr_features <- read.table("X:/sudlab1/General/projects/SynthUTR_hepG2_a549/lasso_with_utr_features/hg38_longestcds_and_detected_3utr_features.tsv",

                           header = T)

#Grouping multiple entries per transcripts

utr_features <- utr_features %>%

  dplyr::select(transcript_id, utr_length, C.G) %>%

  group_by(transcript_id) %>%

  summarize(group_GC = mean(C.G),

            utr_length = sum(utr_length))

#Getting gene_IDs

gene2transcript <- read.table("X:/sudlab1/General/mirror/ensembl/hg38_ensembl93/Homo_sapiens.GRCh38.93.IDs.tsv",

                              header = T)

utr_length <- utr_features %>%

  inner_join(gene2transcript, by = c("transcript_id" = "transcript_stable_id"))

#LFC input
```

```
resLFC_D0 <- read.table("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/GO_KEGG/resLFC_D0_only.tsv",
                        header = T, stringsAsFactors = FALSE)

resLFC_D10 <- read.table("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/GO_KEGG/resLFC_D10_only.tsv",
                         header = T, stringsAsFactors = FALSE)

resLFC_D20 <- read.table("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/GO_KEGG/resLFC_D20_only.tsv",
                         header = T, stringsAsFactors = FALSE)

resLFC_D7 <- read.table("X:/sudlab1/General/projects/Favour_EKT/r_scripts/Working_directory/GO_KEGG/resLFC_D7_only.tsv",
                        header = T, stringsAsFactors = FALSE)

#Loading miR_targets

file_path <- "X:/sudlab1/General/mirror/bindingSites_predictions/mRNA-miRNA_bindingSites/starBaseV3_mRNA_miRNA_hg19_allinfo.txt"

miR_targets <- read.table(file_path,
                          header = T, comment.char = "#") %>%
  mutate(miRNAname = str_replace_all(miRNAname, "R", "r"))

#miR_targets <- read.table("X:/sudlab1/General/mirror/bindingSites_predictions/mRNA-miRNA_bindingSites/starBaseV3_mRNA_miRNA_hg19_allinfo.txt",
#                          header = T, comment.char = "#") %>%
  #mutate(miRNAname = str_replace_all(miRNAname, "R", "r"))

#IDs miR

miR_ids <- read.table("X:/sudlab1/General/mirror/miRBase/miRBase_hp2mature.tsv",
                      header = T)

#KEGG IDs
```

KEGG <- read.table("X:/sudlab1/General/mirror/gene_sets/KEGG/KEGG_MAP_IDs.tsv", sep = "\t", col.names = c("id", "term"))

```

Select DB of targets :

```{r}

#miR_targets <- #filter

```

Creating list of significantly DE genes according to your thresholds.

```{r get_targets, echo=FALSE}

#First removing NAs in padj because filter doesn't handle that well

resLFC = resLFC_D0

NO_NA <- resLFC %>% filter(!(is.na(padj)))

#Filter

UP <- NO_NA %>% filter(padj <= 0.05,

                       log2FoldChange > 0)   %>%

  row.names()

DOWN <- NO_NA %>% filter(padj <= 0.05,

                         log2FoldChange < 0)  %>%

  row.names()

ALL <-resLFC  %>%  row.names()

#Are all you miRs in the miR target DB ?

all(ALL %in% miR_targets$miRNAname)

#Which aren't ?

get_accession <- ALL[!(ALL %in% miR_targets$miRNAname)]

#Let's get miRBase accession ID's

all(get_accession %in% miR_ids$Name)

```
get_accession <- miR_ids %>%
  dplyr::select(miRBase_ID, Name) %>%
  filter(Name %in% get_accession)
#Still have some of them not there
all(get_accession$miRBase_ID %in% miR_targets$miRNAid)
#Cross-checking manually some of them
#get_accession[!(get_accession$miRBase_ID %in% miR_targets$miRNAid),]
#miR_targets[grepl("hsa-mir-548",miR_targets$miRNAname),]
```

Still `r nrow(get_accession[!(get_accession$miRBase_ID %in% miR_targets$miRNAid),])`` miR without known targets.

Checking if any are in our UP and DOWN lists:

```{r no_target, echo=FALSE}
#Checking the ones without targets in DOWN and UP
ALL_notargets <- get_accession[!(get_accession$miRBase_ID %in% miR_targets$miRNAid),]
UP[UP %in% ALL_notargets$Name]
DOWN[DOWN %in% ALL_notargets$Name]
```

Creating target lists and GO vectors

```{r targets, include=FALSE}
#In the end switching to accession IDs for simplicity
ALL_ids <- miR_ids %>%
  dplyr::select(miRBase_ID, Name) %>%
  filter(Name %in% ALL)
UP_ids <- miR_ids %>%
  dplyr::select(miRBase_ID, Name) %>%
```

```
  filter(Name %in% UP)
DOWN_ids <- miR_ids %>%
  dplyr::select(miRBase_ID, Name) %>%
  filter(Name %in% DOWN)
ALL_targets <- miR_targets %>%
  filter(miRNAid %in% ALL_ids$miRBase_ID)
UP_targets <- miR_targets %>%
  filter(miRNAid %in% UP_ids$miRBase_ID)
DOWN_targets <- miR_targets %>%
  filter(miRNAid %in% DOWN_ids$miRBase_ID)
#Vectors
UP.vector <- c(t(unique(UP_targets$geneID)))
DOWN.vector <- c(t(unique(DOWN_targets$geneID)))
ALL.vector<-c(t(unique(ALL_targets$geneID)))
```

If you want to have a one miR target vector:

```{r one, echo = FALSE}
#One miRNA
ONE <- miR_ids %>%
  filter(Name == "")
ONE_targets <- miR_targets %>%
  filter(miRNAid %in% ONE$miRBase_ID)
ONE.vector <- c(t(ONE_targets$geneID))
```

Creating tables of length bias (3'UTR)

```{r lengths, echo=FALSE}
```

```
lengths <- utr_length %>%
  filter(gene_stable_id %in% ALL_targets$geneID) %>%
  dplyr::select(gene_stable_id, utr_length) %>%
  group_by(gene_stable_id) %>%
  summarize(utr_length = mean(utr_length)) %>%
  ungroup()
#Checking they all have a UTR length
all(unique(ALL_targets$geneID) %in% unique(lengths$gene_stable_id))
test_all <- unique(ALL_targets$geneID)
test_len <- unique(utr_length$gene_stable_id)
missing <- setdiff(test_all, test_len)
#No other transcripts for these genes
nrow(utr_length %>%
  filter(gene_stable_id %in% missing))
#Creating Fake UTR length = to mean of all others.
mean_utr_length = mean(lengths$utr_length)
missing.vector <- rep(mean_utr_length, times = length(missing))
names(missing.vector) <- missing
#I could also just get rid of them, most of them will likely not have a GO
#annotation anyway.
lengths.vector <- c(t(lengths$utr_length))
names(lengths.vector) <- lengths$gene_stable_id
lengths.vector <- c(lengths.vector, missing.vector)
test_all <- names(ALL.vector)
test_len <- names(lengths.vector)
print(c("This should return TRUE: ", all(test_all %in% test_len)))
```

```

```{r}
KEGG <- subset(KEGG, !is.na(term) & term != "")
```

Creating functions

```{r functions, include=FALSE}
GOseqFunction <- function(background_vector = ALL.vector,
                          DE_vector, bias_vector = lengths.vector,
                          test_cat = "GO:BP",
                          GOmethod = "Wallenius") {
gene.vector=as.integer(background_vector%in%DE_vector)
names(gene.vector)=background_vector
pwf=nullp(gene.vector,"hg19","ensGene", bias.data = lengths.vector)
GO.wall=goseq(pwf,"hg19","ensGene", test.cats=test_cat, method = GOmethod)
GO.wall$FDR <- p.adjust(GO.wall$over_represented_pvalue, method="BH")
if (test_cat == "KEGG") {
  GO.wall %>%
    left_join(KEGG, by = c("category" = "id"))
}
}
graphFunction <- function(table) {
table %>%
    top_n(10, wt=-FDR) %>%
    mutate(hitsPerc=numDEInCat*100/numInCat) %>%
    ggplot(aes(x=hitsPerc,
               y=term,
```
```

```
            colour=-log(FDR),
            size=numDEInCat)) +
      geom_point() + theme_minimal(base_size = 14) +
      expand_limits(x=0) +
      labs(x="Hits (%)", y="GO term", colour="-log(p-value)", size="Count")
}
```

Runing functions

```{r GO, echo=FALSE}
GO_UP_BP <- GOseqFunction(DE_vector = UP.vector, test_cat = "GO:BP" )
GO_DOWN_BP <- GOseqFunction(DE_vector = DOWN.vector, test_cat = "GO:BP" )
```

```{r}
library(dplyr)
library(ggplot2)
sorted_data <- GO_DOWN_BP %>%
  top_n(37, wt = -over_represented_pvalue) %>%
  mutate(hitsPerc = numDEInCat * 100 / numInCat) %>%
  arrange(desc(numDEInCat))  # Arrange by count in descending order
# Now you can visualize the sorted data using ggplot
p <- ggplot(sorted_data, aes(x = hitsPerc,
                             y = reorder(term, numDEInCat),  # Reorder terms based on numDEInCat
                             colour = over_represented_pvalue,
                             size = numDEInCat)) +
  geom_point() +
  expand_limits(x = 0) +
```

```
  labs(x = "Hits (%)", y = "GO term", colour = "p value", size = "Count")
ggsave(filename = "GO_BP_D10_down.png", plot = p)
ggsave(filename = "GO_BP_D10_down.pdf", plot = p)
p
```
```

Creating functions KEGG

Runing functions

```{r GO, echo=FALSE}
KEGG_UP <- GOseqFunction(DE_vector = UP.vector, test_cat = "KEGG" )
KEGG_DOWN <- GOseqFunction(DE_vector = DOWN.vector, test_cat = "KEGG" )
```

```{r}
library(dplyr)
library(ggplot2)
sorted_data <- KEGG_UP %>%
  top_n(37, wt = -over_represented_pvalue) %>%
  mutate(hitsPerc = numDEInCat * 100 / numInCat) %>%
  arrange(desc(numDEInCat))  # Arrange by count in descending order
# Now you can visualize the sorted data using ggplot
p <- ggplot(sorted_data, aes(x = hitsPerc,
                             y = reorder(term, numDEInCat),  # Reorder terms based on numDEInCat
                             colour = over_represented_pvalue,
                             size = numDEInCat)) +
  geom_point() +
  expand_limits(x = 0) +
```
```

```
  labs(x = "Hits (%)", y = "KEGG term", colour = "p value", size = "Gene count")
ggsave(filename = "KEGG_D0_up.png", plot = p, width = 6, height = 6)
ggsave(filename = "KEGG_D0_up.pdf", plot = p, width = 6, height = 6)
p
```

References

1. Dejerine A, Ceillier A. Paraosteoarthropathies of paraplegic patients by spinal cord lesion. Clinical and roentgenographic study. Clin Orthop Relat Res. 1991(263):3-12.
2. Pignolo RJ, Hsiao EC, Baujat G, Lapidus D, Sherman A, Kaplan FS. Prevalence of fibrodysplasia ossificans progressiva (FOP) in the United States: estimate from three treatment centers and a patient organization. Orphanet J Rare Dis. 2021;16(1):350.
3. Shore EM, Kaplan FS. Inherited human diseases of heterotopic bone formation. Nat Rev Rheumatol. 2010;6(9):518-27.
4. Kaplan FS, Craver R, MacEwen GD, Gannon FH, Finkel G, Hahn G, et al. Progressive osseous heteroplasia: a distinct developmental disorder of heterotopic ossification. Two new case reports and follow-up of three previously reported cases. J Bone Joint Surg Am. 1994;76(3):425-36.
5. Kaplan FS, Shore EM. Progressive osseous heteroplasia. J Bone Miner Res. 2000;15(11):2084-94.
6. Kaplan FS, Zasloff MA, Kitterman JA, Shore EM, Hong CC, Rocke DM. Early mortality and cardiorespiratory failure in patients with fibrodysplasia ossificans progressiva. J Bone Joint Surg Am. 2010;92(3):686-91.
7. Vanden Bossche L, Vanderstraeten G. Heterotopic ossification: a review. J Rehabil Med. 2005;37(3):129-36.
8. Cipriano CA, Pill SG, Keenan MA. Heterotopic ossification following traumatic brain injury and spinal cord injury. J Am Acad Orthop Surg. 2009;17(11):689-97.
9. Brooker AF, Bowerman JW, Robinson RA, Riley LH, Jr. Ectopic ossification following total hip arthroplasty. Incidence and method of classification. J Bone Joint Surg. 1973;55-A(8):1629-32.
10. Potter BK, Burns TC, Lacap AP, Granville RR, Gajewski DA. Heterotopic ossification following traumatic and combat-related amputations. Prevalence, risk factors, and preliminary results of excision. J Bone Joint Surg Am. 2007;89(3):476-86.
11. Berendsen AD, Olsen BR. Bone development. Bone. 2015;80:14-8.
12. Buck DW, 2nd, Dumanian GA. Bone biology and physiology: Part I. The fundamentals. Plast Reconstr Surg. 2012;129(6):1314-20.
13. Shehab D, Elgazzar AH, Collier BD. Heterotopic ossification. J Nucl Med. 2002;43(3):346-53.
14. Chalmers J, Gray DH, Rush J. Observations on the induction of bone in soft tissues. J Bone Joint Surg Br. 1975;57(1):36-45.
15. Kan L, Kessler JA. Evaluation of the cellular origins of heterotopic ossification. Orthopedics. 2014;37(5):329-40.
16. Lees-Shepard JB, Goldhamer DJ. Stem cells and heterotopic ossification: Lessons from animal models. Bone. 2018;109:178-86.
17. Friedenstein AY, Lalykina KS. Lymphoid cell populations are competent systems for induced osteogenesis. Calcif Tissue Res. 1970:Suppl:105-6.
18. Shafritz AB, Shore EM, Gannon FH, Zasloff MA, Taub R, Muenke M, et al. Overexpression of an osteogenic morphogen in fibrodysplasia ossificans progressiva. N Engl J Med. 1996;335(8):555-61.
19. Olmsted-Davis EA, Gugala Z, Camargo F, Gannon FH, Jackson K, Kienstra KA, et al. Primitive adult hematopoietic stem cells can function as osteoblast precursors. Proc Natl Acad Sci U S A. 2003;100(26):15877-82.

20. Gussoni E, Soneoka Y, Strickland CD, Buzney EA, Khan MK, Flint AF, et al. Dystrophin expression in the mdx mouse restored by stem cell transplantation. Nature. 1999;401(6751):390-4.
21. Jackson KA, Majka SM, Wang H, Pocius J, Hartley CJ, Majesky MW, et al. Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. J Clin Invest. 2001;107(11):1395-402.
22. Dominici M, Pritchard C, Garlits JE, Hofmann TJ, Persons DA, Horwitz EM. Hematopoietic cells and osteoblasts are derived from a common marrow progenitor after bone marrow transplantation. Proc Natl Acad Sci U S A. 2004;101(32):11761-6.
23. Kaplan FS, Glaser DL, Shore EM, Pignolo RJ, Xu M, Zhang Y, et al. Hematopoietic stem-cell contribution to ectopic skeletogenesis. J Bone Joint Surg Am. 2007;89(2):347-57.
24. Otsuru S, Tamai K, Yamazaki T, Yoshikawa H, Kaneda Y. Bone marrow-derived osteoblast progenitor cells in circulating blood contribute to ectopic bone formation in mice. Biochem Biophys Res Commun. 2007;354(2):453-8.
25. Otsuru S, Tamai K, Yamazaki T, Yoshikawa H, Kaneda Y. Circulating bone marrow-derived osteoblast progenitor cells are recruited to the bone-forming site by the CXCR4/stromal cell-derived factor-1 pathway. Stem Cells. 2008;26(1):223-34.
26. Egan KP, Duque G, Keenan MA, Pignolo RJ. Circulating osteogentic precursor cells in non-hereditary heterotopic ossification. Bone. 2018;109:61-4.
27. Lounev VY, Ramachandran R, Wosczyna MN, Yamamoto M, Maidment AD, Shore EM, et al. Identification of progenitor cells that contribute to heterotopic skeletogenesis. J Bone Joint Surg Am. 2009;91(3):652-63.
28. Medici D, Shore EM, Lounev VY, Kaplan FS, Kalluri R, Olsen BR. Conversion of vascular endothelial cells into multipotent stem-like cells. Nature medicine. 2010;16(12):1400-6.
29. De Angelis L, Berghella L, Coletta M, Lattanzi L, Zanchi M, Cusella-De Angelis MG, et al. Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. J Cell Biol. 1999;147(4):869-78.
30. Medici D, Kalluri R. Endothelial-mesenchymal transition and its contribution to the emergence of stem cell phenotype. Semin Cancer Biol. 2012;22(5-6):379-84.
31. Potenta S, Zeisberg E, Kalluri R. The role of endothelial-to-mesenchymal transition in cancer progression. Br J Cancer. 2008;99(9):1375-9.
32. Wosczyna MN, Biswas AA, Cogswell CA, Goldhamer DJ. Multipotent progenitors resident in the skeletal muscle interstitium exhibit robust BMP-dependent osteogenic activity and mediate heterotopic ossification. J Bone Miner Res. 2012;27(5):1004-17.
33. Agarwal S, Loder S, Cholok D, Peterson J, Li J, Fireman D, et al. Local and Circulating Endothelial Cells Undergo Endothelial to Mesenchymal Transition (EndMT) in Response to Musculoskeletal Injury. Sci Rep. 2016;6:32514.
34. Chu M, Li T, Shen B, Cao X, Zhong H, Zhang L, et al. Angiopoietin receptor Tie2 is required for vein specification and maintenance via regulating COUP-TFII. Elife. 2016;5.
35. Sato A, Iwama A, Takakura N, Nishio H, Yancopoulos GD, Suda T. Characterization of TEK receptor tyrosine kinase and its ligands, Angiopoietins, in human hematopoietic progenitor cells. Int Immunol. 1998;10(8):1217-27.
36. Yano M, Iwama A, Nishio H, Suda J, Takada G, Suda T. Expression and function of murine receptor tyrosine kinases, TIE and TEK, in hematopoietic stem cells. Blood. 1997;89(12):4317-26.

37. Joe AW, Yi L, Natarajan A, Le Grand F, So L, Wang J, et al. Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. Nat Cell Biol. 2010;12(2):153-63.
38. Uezumi A, Fukada S, Yamamoto N, Takeda S, Tsuchida K. Mesenchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. Nat Cell Biol. 2010;12(2):143-52.
39. Dey D, Bagarova J, Hatsell SJ, Armstrong KA, Huang L, Ermann J, et al. Two tissue-resident progenitor lineages drive distinct phenotypes of heterotopic ossification. Sci Transl Med. 2016;8(366):366ra163.
40. Lees-Shepard JB, Yamamoto M, Biswas AA, Stoessel SJ, Nicholas SE, Cogswell CA, et al. Activin-dependent signaling in fibro/adipogenic progenitors causes fibrodysplasia ossificans progressiva. Nat Commun. 2018;9(1):471.
41. Agarwal S, Loder S, Brownley C, Cholok D, Mangiavini L, Li J, et al. Inhibition of Hif1alpha prevents both trauma-induced and genetic heterotopic ossification. Proc Natl Acad Sci U S A. 2016;113(3):E338-47.
42. Agarwal S, Loder S, Cholok D, Li J, Breuler C, Drake J, et al. Surgical Excision of Heterotopic Ossification Leads to Re-Emergence of Mesenchymal Stem Cell Populations Responsible for Recurrence. Stem Cells Transl Med. 2017;6(3):799-806.
43. Agarwal S, Loder SJ, Cholok D, Peterson J, Li J, Breuler C, et al. Scleraxis-Lineage Cells Contribute to Ectopic Bone Formation in Muscle and Tendon. Stem Cells. 2017;35(3):705-10.
44. Eisner C, Cummings M, Johnston G, Tung LW, Groppa E, Chang C, et al. Murine Tissue-Resident PDGFR α+ Fibro-Adipogenic Progenitors Spontaneously Acquire Osteogenic Phenotype in an Altered Inflammatory Environment. J Bone Miner Res. 2020.
45. Yin H, Price F, Rudnicki MA. Satellite cells and the muscle stem cell niche. Physiol Rev. 2013;93(1):23-67.
46. Charge SB, Rudnicki MA. Cellular and molecular regulation of muscle regeneration. Physiol Rev. 2004;84(1):209-38.
47. Katagiri T, Yamaguchi A, Komaki M, Abe E, Takahashi N, Ikeda T, et al. Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage. J Cell Biol. 1994;127(6 Pt 1):1755-66.
48. Hashimoto N, Kiyono T, Wada MR, Umeda R, Goto Y, Nonaka I, et al. Osteogenic properties of human myogenic progenitor cells. Mech Dev. 2008;125(3-4):257-69.
49. Kan L, Liu Y, McGuire TL, Berger DM, Awatramani RB, Dymecki SM, et al. Dysregulation of local stem/progenitor cells as a common cellular mechanism for heterotopic ossification. Stem Cells. 2009;27(1):150-6.
50. Matthews BG, Torreggiani E, Roeder E, Matic I, Grcevic D, Kalajzic I. Osteogenic potential of alpha smooth muscle actin expressing muscle resident progenitor cells. Bone. 2016;84:69-77.
51. Shimono K, Tung WE, Macolino C, Chi AH, Didizian JH, Mundy C, et al. Potent inhibition of heterotopic ossification by nuclear retinoic acid receptor-gamma agonists. Nat Med. 2011;17(4):454-60.
52. Winbanks CE, Chen JL, Qian H, Liu Y, Bernardo BC, Beyer C, et al. The bone morphogenetic protein axis is a positive regulator of skeletal muscle mass. J Cell Biol. 2013;203(2):345-57.
53. Liu R, Ginn SL, Lek M, North KN, Alexander IE, Little DG, et al. Myoblast sensitivity and fibroblast insensitivity to osteogenic conversion by BMP-2 correlates with the expression of Bmpr-1a. BMC Musculoskelet Disord. 2009;10:51.

54. Wright V, Peng H, Usas A, Young B, Gearhart B, Cummins J, et al. BMP4-expressing muscle-derived stem cells differentiate into osteogenic lineage and improve bone healing in immunocompetent mice. Mol Ther. 2002;6(2):169-78.
55. Wu X, Rathbone CR. Satellite cell functional alterations following cutaneous burn in rats include an increase in their osteogenic potential. J Surg Res. 2013;184(2):e9-16.
56. Pagani CA, Huber AK, Hwang C, Marini S, Padmanabhan K, Livingston N, et al. Novel Lineage-Tracing System to Identify Site-Specific Ectopic Bone Precursor Cells. Stem Cell Reports. 2021;16(3):626-40.
57. Pineault KM, Song JY, Kozloff KM, Lucas D, Wellik DM. Hox11 expressing regional skeletal stem cells are progenitors for osteoblasts, chondrocytes and adipocytes throughout life. Nat Commun. 2019;10(1):3168.
58. Swinehart IT, Schlientz AJ, Quintanilla CA, Mortlock DP, Wellik DM. Hox11 genes are required for regional patterning and integration of muscle, tendon and bone. Development. 2013;140(22):4574-82.
59. Rux DR, Song JY, Swinehart IT, Pineault KM, Schlientz AJ, Trulik KG, et al. Regionally Restricted Hox Function in Adult Bone Marrow Multipotent Mesenchymal Stem/Stromal Cells. Dev Cell. 2016;39(6):653-66.
60. Suda RK, Billings PC, Egan KP, Kim JH, McCarrick-Walmsley R, Glaser DL, et al. Circulating osteogenic precursor cells in heterotopic bone formation. Stem Cells. 2009;27(9):2209-19.
61. Meyers C, Lisiecki J, Miller S, Levin A, Fayad L, Ding C, et al. Heterotopic Ossification: A Comprehensive Review. JBMR Plus. 2019;3(4):e10172.
62. Grcevic D, Pejda S, Matthews BG, Repic D, Wang L, Li H, et al. In vivo fate mapping identifies mesenchymal progenitor cells. Stem Cells. 2012;30(2):187-96.
63. Kan L, Peng CY, McGuire TL, Kessler JA. Glast-expressing progenitor cells contribute to heterotopic ossification. Bone. 2013;53(1):194-203.
64. Slezak M, Goritz C, Niemiec A, Frisen J, Chambon P, Metzger D, et al. Transgenic mice for conditional gene manipulation in astroglial cells. Glia. 2007;55(15):1565-76.
65. Howell K, Chien C, Bell R, Laudier D, Tufa SF, Keene DR, et al. Novel Model of Tendon Regeneration Reveals Distinct Cell Mechanisms Underlying Regenerative and Fibrotic Tendon Healing. Sci Rep. 2017;7:45238.
66. Kan L, Lounev VY, Pignolo RJ, Duan L, Liu Y, Stock SR, et al. Substance P signaling mediates BMP-dependent heterotopic ossification. J Cell Biochem. 2011;112(10):2759-72.
67. Lazard ZW, Olmsted-Davis EA, Salisbury EA, Gugala Z, Sonnet C, Davis EL, et al. Osteoblasts Have a Neural Origin in Heterotopic Ossification. Clin Orthop Relat Res. 2015;473(9):2790-806.
68. Salisbury E, Rodenberg E, Sonnet C, Hipp J, Gannon FH, Vadakkan TJ, et al. Sensory nerve induced inflammation contributes to heterotopic ossification. J Cell Biochem. 2011;112(10):2748-58.
69. Salisbury E, Sonnet C, Heggeness M, Davis AR, Olmsted-Davis E. Heterotopic ossification has some nerve. Crit Rev Eukaryot Gene Expr. 2010;20(4):313-24.
70. Olmsted-Davis E, Gannon FH, Ozen M, Ittmann MM, Gugala Z, Hipp JA, et al. Hypoxic adipocytes pattern early heterotopic bone formation. Am J Pathol. 2007;170(2):620-32.
71. Salisbury EA, Dickerson AR, Davis TA, Forsberg JA, Davis AR, Olmsted-Davis EA. Characterization of Brown Adipose-Like Tissue in Trauma-Induced Heterotopic Ossification in Humans. Am J Pathol. 2017;187(9):2071-9.
72. Urist MR. Bone: formation by autoinduction. Science. 1965;150(3698):893-9.

73. Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M, et al. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. Genes Dis. 2014;1(1):87-105.
74. Asharani PV, Keupp K, Semler O, Wang W, Li Y, Thiele H, et al. Attenuated BMP1 function compromises osteogenesis, leading to bone fragility in humans and zebrafish. Am J Hum Genet. 2012;90(4):661-74.
75. Chen D, Zhao M, Mundy GR. Bone morphogenetic proteins. Growth Factors. 2004;22(4):233-41.
76. Ma L, Lu MF, Schwartz RJ, Martin JF. Bmp2 is essential for cardiac cushion epithelial-mesenchymal transition and myocardial patterning. Development. 2005;132(24):5601-11.
77. Rahman MS, Akhtar N, Jamil HM, Banik RS, Asaduzzaman SM. TGF-beta/BMP signaling and other molecular events: regulation of osteoblastogenesis and bone formation. Bone Res. 2015;3:15005.
78. Zhang H, Bradley A. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. Development. 1996;122(10):2977-86.
79. Faucheux C, Ulysse F, Bareille R, Reddi AH, Amedee J. Opposing actions of BMP3 and TGF beta 1 in human bone marrow stromal cell growth and differentiation. Biochem Biophys Res Commun. 1997;241(3):787-93.
80. Cole AE, Murray SS, Xiao J. Bone Morphogenetic Protein 4 Signalling in Neural Stem and Progenitor Cells during Development and after Injury. Stem Cells Int. 2016;2016:9260592.
81. Lee TJ, Jang J, Kang S, Jin M, Shin H, Kim DW, et al. Enhancement of osteogenic and chondrogenic differentiation of human embryonic stem cells by mesodermal lineage induction with BMP-4 and FGF2 treatment. Biochem Biophys Res Commun. 2013;430(2):793-7.
82. Modica S, Wolfrum C. The dual role of BMP4 in adipogenesis and metabolism. Adipocyte. 2017;6(2):141-6.
83. Watanabe Y, Le Douarin NM. A role for BMP-4 in the development of subcutaneous cartilage. Mech Dev. 1996;57(1):69-78.
84. Bobinac D, Maric I, Zoricic S, Spanjol J, Dordevic G, Mustac E, et al. Expression of bone morphogenetic proteins in human metastatic prostate and breast cancer. Croat Med J. 2005;46(3):389-96.
85. Guenther CA, Wang Z, Li E, Tran MC, Logan CY, Nusse R, et al. A distinct regulatory region of the Bmp5 locus activates gene expression following adult bone fracture or soft tissue injury. Bone. 2015;77:31-41.
86. Wordinger RJ, Agarwal R, Talati M, Fuller J, Lambert W, Clark AF. Expression of bone morphogenetic proteins (BMP), BMP receptors, and BMP associated proteins in human trabecular meshwork and optic nerve head cells and tissues. Mol Vis. 2002;8:241-50.
87. Camaschella C. BMP6 orchestrates iron metabolism. Nat Genet. 2009;41(4):386-8.
88. Gitelman SE, Kobrin MS, Ye JQ, Lopez AR, Lee A, Derynck R. Recombinant Vgr-1/BMP-6-expressing tumors induce fibrosis and endochondral bone formation in vivo. J Cell Biol. 1994;126(6):1595-609.
89. Hahn GV, Cohen RB, Wozney JM, Levitz CL, Shore EM, Zasloff MA, et al. A bone morphogenetic protein subfamily: chromosomal localization of human genes for BMP5, BMP6, and BMP7. Genomics. 1992;14(3):759-62.
90. Dudley AT, Lyons KM, Robertson EJ. A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. Genes Dev. 1995;9(22):2795-807.

91. Itoh F, Asao H, Sugamura K, Heldin CH, ten Dijke P, Itoh S. Promoting bone morphogenetic protein signaling through negative regulation of inhibitory Smads. Embo j. 2001;20(15):4132-42.
92. Tseng YH, Kokkotou E, Schulz TJ, Huang TL, Winnay JN, Taniguchi CM, et al. New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. Nature. 2008;454(7207):1000-4.
93. DiLeone RJ, King JA, Storm EE, Copeland NG, Jenkins NA, Kingsley DM. The Bmp8 gene is expressed in developing skeletal tissue and maps near the Achondroplasia locus on mouse chromosome 4. Genomics. 1997;40(1):196-8.
94. Ying Y, Liu XM, Marble A, Lawson KA, Zhao GQ. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. Mol Endocrinol. 2000;14(7):1053-63.
95. Zhao GQ, Deng K, Labosky PA, Liaw L, Hogan BL. The gene encoding bone morphogenetic protein 8B is required for the initiation and maintenance of spermatogenesis in the mouse. Genes Dev. 1996;10(13):1657-69.
96. Zhao GQ, Liaw L, Hogan BL. Bone morphogenetic protein 8A plays a role in the maintenance of spermatogenesis and the integrity of the epididymis. Development. 1998;125(6):1103-12.
97. Kang Q, Sun MH, Cheng H, Peng Y, Montag AG, Deyrup AT, et al. Characterization of the distinct orthotopic bone-forming activity of 14 BMPs using recombinant adenovirus-mediated gene delivery. Gene Ther. 2004;11(17):1312-20.
98. Levet S, Ciais D, Merdzhanova G, Mallet C, Zimmers TA, Lee SJ, et al. Bone morphogenetic protein 9 (BMP9) controls lymphatic vessel maturation and valve formation. Blood. 2013;122(4):598-607.
99. Majumdar MK, Wang E, Morris EA. BMP-2 and BMP-9 promotes chondrogenic differentiation of human multipotential mesenchymal cells and overcomes the inhibitory effect of IL-1. J Cell Physiol. 2001;189(3):275-84.
100. Chen H, Shi S, Acosta L, Li W, Lu J, Bao S, et al. BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. Development. 2004;131(9):2219-31.
101. Mitrofan CG, Appleby SL, Nash GB, Mallat Z, Chilvers ER, Upton PD, et al. Bone morphogenetic protein 9 (BMP9) and BMP10 enhance tumor necrosis factor-alpha-induced monocyte recruitment to the vascular endothelium mainly via activin receptor-like kinase 2. J Biol Chem. 2017;292(33):13714-26.
102. Neuhaus H, Rosen V, Thies RS. Heart specific expression of mouse BMP-10 a novel member of the TGF-beta superfamily. Mech Dev. 1999;80(2):181-4.
103. Li Z, Zeng F, Mitchell AD, Kim YS, Wu Z, Yang J. Transgenic overexpression of bone morphogenetic protein 11 propeptide in skeleton enhances bone formation. Biochem Biophys Res Commun. 2011;416(3-4):289-92.
104. Zhang Y, Wei Y, Liu D, Liu F, Li X, Pan L, et al. Role of growth differentiation factor 11 in development, physiology and disease. Oncotarget. 2017;8(46):81604-16.
105. Berasi SP, Varadarajan U, Archambault J, Cain M, Souza TA, Abouzeid A, et al. Divergent activities of osteogenic BMP2, and tenogenic BMP12 and BMP13 independent of receptor binding affinities. Growth Factors. 2011;29(4):128-39.
106. Settle SH, Jr., Rountree RB, Sinha A, Thacker A, Higgins K, Kingsley DM. Multiple joint and skeletal patterning defects caused by single and double mutations in the mouse Gdf6 and Gdf5 genes. Dev Biol. 2003;254(1):116-30.

107. Chhabra A, Zijerdi D, Zhang J, Kline A, Balian G, Hurwitz S. BMP-14 deficiency inhibits long bone fracture healing: a biochemical, histologic, and radiographic assessment. J Orthop Trauma. 2005;19(9):629-34.
108. Coleman CM, Scheremeta BH, Boyce AT, Mauck RL, Tuan RS. Delayed fracture healing in growth differentiation factor 5-deficient mice: a pilot study. Clin Orthop Relat Res. 2011;469(10):2915-24.
109. Di Pasquale E, Beck-Peccoz P, Persani L. Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. Am J Hum Genet. 2004;75(1):106-11.
110. Yan C, Wang P, DeMayo J, DeMayo FJ, Elvin JA, Carino C, et al. Synergistic roles of bone morphogenetic protein 15 and growth differentiation factor 9 in ovarian function. Mol Endocrinol. 2001;15(6):854-66.
111. Wang M, Jin H, Tang D, Huang S, Zuscik MJ, Chen D. Smad1 plays an essential role in bone development and postnatal bone formation. Osteoarthritis Cartilage. 2011;19(6):751-62.
112. Iwasaki S, Iguchi M, Watanabe K, Hoshino R, Tsujimoto M, Kohno M. Specific activation of the p38 mitogen-activated protein kinase signaling pathway and induction of neurite outgrowth in PC12 cells by bone morphogenetic protein-2. J Biol Chem. 1999;274(37):26503-10.
113. Lou J, Tu Y, Li S, Manske PR. Involvement of ERK in BMP-2 induced osteoblastic differentiation of mesenchymal progenitor cell line C3H10T1/2. Biochem Biophys Res Commun. 2000;268(3):757-62.
114. Nakamura K, Shirai T, Morishita S, Uchida S, Saeki-Miura K, Makishima F. p38 mitogen-activated protein kinase functionally contributes to chondrogenesis induced by growth/differentiation factor-5 in ATDC5 cells. Exp Cell Res. 1999;250(2):351-63.
115. Yamaguchi K, Shirakabe K, Shibuya H, Irie K, Oishi I, Ueno N, et al. Identification of a member of the MAPKKK family as a potential mediator of TGF-beta signal transduction. Science. 1995;270(5244):2008-11.
116. Bloise E, Ciarmela P, Dela Cruz C, Luisi S, Petraglia F, Reis FM. Activin A in Mammalian Physiology. Physiol Rev. 2019;99(1):739-80.
117. de Caestecker MP, Parks WT, Frank CJ, Castagnino P, Bottaro DP, Roberts AB, et al. Smad2 transduces common signals from receptor serine-threonine and tyrosine kinases. Genes Dev. 1998;12(11):1587-92.
118. Gingery A, Bradley EW, Pederson L, Ruan M, Horwood NJ, Oursler MJ. TGF-beta coordinately activates TAK1/MEK/AKT/NFkB and SMAD pathways to promote osteoclast survival. Exp Cell Res. 2008;314(15):2725-38.
119. Hoffmann A, Preobrazhenska O, Wodarczyk C, Medler Y, Winkel A, Shahab S, et al. Transforming growth factor-beta-activated kinase-1 (TAK1), a MAP3K, interacts with Smad proteins and interferes with osteogenesis in murine mesenchymal progenitors. J Biol Chem. 2005;280(29):27271-83.
120. Nickel J, Mueller TD. Specification of BMP Signaling. Cells. 2019;8(12).
121. Evans KN, Potter BK, Brown TS, Davis TA, Elster EA, Forsberg JA. Osteogenic gene expression correlates with development of heterotopic ossification in war wounds. Clin Orthop Relat Res. 2014;472(2):396-404.
122. Ju C, Lv Z, Zhang C, Jiao Y. Regulatory effect of miR-421 on humeral fracture and heterotopic ossification in elderly patients. Exp Ther Med. 2019;17(3):1903-11.

123. Hannallah D, Peng H, Young B, Usas A, Gearhart B, Huard J. Retroviral delivery of Noggin inhibits the formation of heterotopic ossification induced by BMP-4, demineralized bone matrix, and trauma in an animal model. J Bone Joint Surg Am. 2004;86(1):80-91.
124. Yu PB, Deng DY, Lai CS, Hong CC, Cuny GD, Bouxsein ML, et al. BMP type I receptor inhibition reduces heterotopic [corrected] ossification. Nat Med. 2008;14(12):1363-9.
125. Kan L, Hu M, Gomes WA, Kessler JA. Transgenic mice overexpressing BMP4 develop a fibrodysplasia ossificans progressiva (FOP)-like phenotype. Am J Pathol. 2004;165(4):1107-15.
126. Scott MA, Levi B, Askarinam A, Nguyen A, Rackohn T, Ting K, et al. Brief review of models of ectopic bone formation. Stem Cells Dev. 2012;21(5):655-67.
127. Ranganathan K, Loder S, Agarwal S, Wong VW, Forsberg J, Davis TA, et al. Heterotopic Ossification: Basic-Science Principles and Clinical Correlates. J Bone Joint Surg Am. 2015;97(13):1101-11.
128. Chen J, Long F. mTOR signaling in skeletal development and disease. Bone Res. 2018;6:1.
129. Phornphutkul C, Wu KY, Auyeung V, Chen Q, Gruppuso PA. mTOR signaling contributes to chondrocyte differentiation. Dev Dyn. 2008;237(3):702-12.
130. Hino K, Horigome K, Nishio M, Komura S, Nagata S, Zhao C, et al. Activin-A enhances mTOR signaling to promote aberrant chondrogenesis in fibrodysplasia ossificans progressiva. J Clin Invest. 2017;127(9):3339-52.
131. Qureshi AT, Dey D, Sanders EM, Seavey JG, Tomasino AM, Moss K, et al. Inhibition of Mammalian Target of Rapamycin Signaling with Rapamycin Prevents Trauma-Induced Heterotopic Ossification. Am J Pathol. 2017;187(11):2536-45.
132. Jiang H, Chen Y, Chen G, Tian X, Tang J, Luo L, et al. Leptin accelerates the pathogenesis of heterotopic ossification in rat tendon tissues via mTORC1 signaling. J Cell Physiol. 2018;233(2):1017-28.
133. Meng D, Frank AR, Jewell JL. mTOR signaling in stem and progenitor cells. Development. 2018;145(1).
134. Karner CM, Lee SY, Long F. Bmp Induces Osteoblast Differentiation through both Smad4 and mTORC1 Signaling. Mol Cell Biol. 2017;37(4).
135. Chen J, Holguin N, Shi Y, Silva MJ, Long F. mTORC2 signaling promotes skeletal growth and bone formation in mice. J Bone Miner Res. 2015;30(2):369-78.
136. Sun W, Shi Y, Lee WC, Lee SY, Long F. Rictor is required for optimal bone accrual in response to anti-sclerostin therapy in the mouse. Bone. 2016;85:1-8.
137. Semenza GL. Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. Trends Mol Med. 2001;7(8):345-50.
138. Wang Y, Wan C, Deng L, Liu X, Cao X, Gilbert SR, et al. The hypoxia-inducible factor alpha pathway couples angiogenesis to osteogenesis during skeletal development. J Clin Invest. 2007;117(6):1616-26.
139. Galdones E, Hales BF. Retinoic acid receptor gamma-induced misregulation of chondrogenesis in the murine limb bud in vitro. Toxicol Sci. 2008;106(1):223-32.
140. Romand R, Hashino E, Dolle P, Vonesch JL, Chambon P, Ghyselinck NB. The retinoic acid receptors RARalpha and RARgamma are required for inner ear development. Mech Dev. 2002;119(2):213-23.
141. Weston AD, Chandraratna RA, Torchia J, Underhill TM. Requirement for RAR-mediated gene repression in skeletal progenitor differentiation. J Cell Biol. 2002;158(1):39-51.

142. Weston AD, Rosen V, Chandraratna RA, Underhill TM. Regulation of skeletal progenitor differentiation by the BMP and retinoid signaling pathways. J Cell Biol. 2000;148(4):679-90.
143. Cash DE, Bock CB, Schughart K, Linney E, Underhill TM. Retinoic acid receptor alpha function in vertebrate limb skeletogenesis: a modulator of chondrogenesis. J Cell Biol. 1997;136(2):445-57.
144. Kan C, Chen L, Hu Y, Ding N, Lu H, Li Y, et al. Conserved signaling pathways underlying heterotopic ossification. Bone. 2018;109:43-8.
145. Liu X, Qin J, Luo Q, Bi Y, Zhu G, Jiang W, et al. Cross-talk between EGF and BMP9 signalling pathways regulates the osteogenic differentiation of mesenchymal stem cells. J Cell Mol Med. 2013;17(9):1160-72.
146. Luo K. Signaling Cross Talk between TGF-beta/Smad and Other Signaling Pathways. Cold Spring Harb Perspect Biol. 2017;9(1).
147. Nakashima A, Katagiri T, Tamura M. Cross-talk between Wnt and bone morphogenetic protein 2 (BMP-2) signaling in differentiation pathway of C2C12 myoblasts. J Biol Chem. 2005;280(45):37660-8.
148. Zhang T, Wen F, Wu Y, Goh GS, Ge Z, Tan LP, et al. Cross-talk between TGF-beta/SMAD and integrin signaling pathways in regulating hypertrophy of mesenchymal stem cell chondrogenesis under deferral dynamic compression. Biomaterials. 2015;38:72-85.
149. Pavlou G, Kyrkos M, Tsialogiannis E, Korres N, Tsiridis E. Pharmacological treatment of heterotopic ossification following hip surgery: an update. Expert Opin Pharmacother. 2012;13(5):619-22.
150. Joice M, Vasileiadis GI, Amanatullah DF. Non-steroidal anti-inflammatory drugs for heterotopic ossification prophylaxis after total hip arthroplasty: a systematic review and meta-analysis. Bone Joint J. 2018;100-B(7):915-22.
151. Kan SL, Yang B, Ning GZ, Chen LX, Li YL, Gao SJ, et al. Nonsteroidal Anti-inflammatory Drugs as Prophylaxis for Heterotopic Ossification after Total Hip Arthroplasty: A Systematic Review and Meta-Analysis. Medicine (Baltimore). 2015;94(18):e828.
152. Migliorini F, Trivellas A, Eschweiler J, Driessen A, Tingart M, Maffulli N. NSAIDs for Prophylaxis for Heterotopic Ossification After Total Hip Arthroplasty: A Bayesian Network Meta-analysis. Calcif Tissue Int. 2021;108(2):196-206.
153. Hu ZH, Chen W, Sun JN, Zhang Y, Zhang Y, Chen XY, et al. Radiotherapy for the prophylaxis of heterotopic ossification after total hip arthroplasty: A systematic review and meta-analysis of randomized controlled trails. Med Dosim. 2021;46(1):65-73.
154. Popovic M, Agarwal A, Zhang L, Yip C, Kreder HJ, Nousiainen MT, et al. Radiotherapy for the prophylaxis of heterotopic ossification: a systematic review and meta-analysis of published data. Radiother Oncol. 2014;113(1):10-7.
155. Sheybani A, TenNapel MJ, Lack WD, Clerkin P, Hyer DE, Sun W, et al. Risk of radiation-induced malignancy with heterotopic ossification prophylaxis: a case-control analysis. Int J Radiat Oncol Biol Phys. 2014;89(3):584-9.
156. Balboni TA, Gobezie R, Mamon HJ. Heterotopic ossification: Pathophysiology, clinical features, and the role of radiotherapy for prophylaxis. Int J Radiat Oncol Biol Phys. 2006;65(5):1289-99.
157. Hamid N, Ashraf N, Bosse MJ, Connor PM, Kellam JF, Sims SH, et al. Radiation therapy for heterotopic ossification prophylaxis acutely after elbow trauma: a prospective randomized study. J Bone Joint Surg Am. 2010;92(11):2032-8.

158. Pavey GJ, Polfer EM, Nappo KE, Tintle SM, Forsberg JA, Potter BK. What Risk Factors Predict Recurrence of Heterotopic Ossification After Excision in Combat-related Amputations? Clin Orthop Relat Res. 2015;473(9):2814-24.
159. Meiners T, Abel R, Bohm V, Gerner HJ. Resection of heterotopic ossification of the hip in spinal cord injured patients. Spinal Cord. 1997;35(7):443-5.
160. Lee EK, Namdari S, Hosalkar HS, Keenan MA, Baldwin KD. Clinical results of the excision of heterotopic bone around the elbow: a systematic review. J Shoulder Elbow Surg. 2013;22(5):716-22.
161. Thomas BJ, Amstutz HC. Results of the administration of diphosphonate for the prevention of heterotopic ossification after total hip arthroplasty. J Bone Joint Surg Am. 1985;67(3):400-3.
162. Shafer DM, Bay C, Caruso DM, Foster KN. The use of eidronate disodium in the prevention of heterotopic ossification in burn patients. Burns. 2008;34(3):355-60.
163. Shimono K, Morrison TN, Tung WE, Chandraratna RA, Williams JA, Iwamoto M, et al. Inhibition of ectopic bone formation by a selective retinoic acid receptor alpha-agonist: a new therapy for heterotopic ossification? J Orthop Res. 2010;28(2):271-7.
164. Koyama E, Golden EB, Kirsch T, Adams SL, Chandraratna RA, Michaille JJ, et al. Retinoid signaling is required for chondrocyte maturation and endochondral bone formation during limb skeletogenesis. Dev Biol. 1999;208(2):375-91.
165. Weston AD, Hoffman LM, Underhill TM. Revisiting the role of retinoid signaling in skeletal development. Birth Defects Res C Embryo Today. 2003;69(2):156-73.
166. Chakkalakal SA, Uchibe K, Convente MR, Zhang D, Economides AN, Kaplan FS, et al. Palovarotene Inhibits Heterotopic Ossification and Maintains Limb Mobility and Growth in Mice With the Human ACVR1(R206H) Fibrodysplasia Ossificans Progressiva (FOP) Mutation. J Bone Miner Res. 2016;31(9):1666-75.
167. Lees-Shepard JB, Nicholas SE, Stoessel SJ, Devarakonda PM, Schneider MJ, Yamamoto M, et al. Palovarotene reduces heterotopic ossification in juvenile FOP mice but exhibits pronounced skeletal toxicity. Elife. 2018;7.
168. Wheatley BM, Cilwa KE, Dey D, Qureshi AT, Seavey JG, Tomasino AM, et al. Palovarotene inhibits connective tissue progenitor cell proliferation in a rat model of combat-related heterotopic ossification. J Orthop Res. 2018;36(4):1135-44.
169. Pavey GJ, Qureshi AT, Tomasino AM, Honnold CL, Bishop DK, Agarwal S, et al. Targeted stimulation of retinoic acid receptor-gamma mitigates the formation of heterotopic ossification in an established blast-related traumatic injury model. Bone. 2016;90:159-67.
170. Lebrun JJ, Takabe K, Chen Y, Vale W. Roles of pathway-specific and inhibitory Smads in activin receptor signaling. Mol Endocrinol. 1999;13(1):15-23.
171. Mueller TD, Nickel J. Promiscuity and specificity in BMP receptor activation. FEBS Lett. 2012;586(14):1846-59.
172. Massague J, Seoane J, Wotton D. Smad transcription factors. Genes Dev. 2005;19(23):2783-810.
173. Yang J, Jiang W. The Role of SMAD2/3 in Human Embryonic Stem Cells. Front Cell Dev Biol. 2020;8:653.
174. Yang J, Wahdan-Alaswad R, Danielpour D. Critical role of Smad2 in tumor suppression and transforming growth factor-beta-induced apoptosis of prostate epithelial cells. Cancer Res. 2009;69(6):2185-90.

175. Culbert AL, Chakkalakal SA, Theosmy EG, Brennan TA, Kaplan FS, Shore EM. Alk2 regulates early chondrogenic fate in fibrodysplasia ossificans progressiva heterotopic endochondral ossification. Stem Cells. 2014;32(5):1289-300.
176. van Dinther M, Visser N, de Gorter DJ, Doorn J, Goumans MJ, de Boer J, et al. ALK2 R206H mutation linked to fibrodysplasia ossificans progressiva confers constitutive activity to the BMP type I receptor and sensitizes mesenchymal cells to BMP-induced osteoblast differentiation and bone formation. J Bone Miner Res. 2010;25(6):1208-15.
177. Lyu H, Elkins CM, Pierce JL, Serezani CH, Perrien DS. MyD88 Is Not Required for Muscle Injury-Induced Endochondral Heterotopic Ossification in a Mouse Model of Fibrodysplasia Ossificans Progressiva. Biomedicines. 2021;9(6).
178. Lin H, Ying Y, Wang YY, Wang G, Jiang SS, Huang D, et al. AMPK downregulates ALK2 via increasing the interaction between Smurf1 and Smad6, leading to inhibition of osteogenic differentiation. Biochim Biophys Acta Mol Cell Res. 2017;1864(12):2369-77.
179. Hatsell SJ, Idone V, Wolken DM, Huang L, Kim HJ, Wang L, et al. ACVR1R206H receptor mutation causes fibrodysplasia ossificans progressiva by imparting responsiveness to activin A. Sci Transl Med. 2015;7(303):303ra137.
180. Hino K, Ikeya M, Horigome K, Matsumoto Y, Ebise H, Nishio M, et al. Neofunction of ACVR1 in fibrodysplasia ossificans progressiva. Proc Natl Acad Sci U S A. 2015;112(50):15438-43.
181. Song GA, Kim HJ, Woo KM, Baek JH, Kim GS, Choi JY, et al. Molecular consequences of the ACVR1(R206H) mutation of fibrodysplasia ossificans progressiva. J Biol Chem. 2010;285(29):22542-53.
182. Kaplan FS, Le Merrer M, Glaser DL, Pignolo RJ, Goldsby RE, Kitterman JA, et al. Fibrodysplasia ossificans progressiva. Best Pract Res Clin Rheumatol. 2008;22(1):191-205.
183. Lin H, Shi F, Gao J, Hua P. The role of Activin A in fibrodysplasia ossificans progressiva: a prominent mediator. Biosci Rep. 2019;39(8).
184. Ramirez DM, Ramirez MR, Reginato AM, Medici D. Molecular and cellular mechanisms of heterotopic ossification. Histol Histopathol. 2014;29(10):1281-5.
185. Aykul S, Corpina RA, Goebel EJ, Cunanan CJ, Dimitriou A, Kim HJ, et al. Activin A forms a non-signaling complex with ACVR1 and type II Activin/BMP receptors via its finger 2 tip loop. Elife. 2020;9.
186. Latres E, Mastaitis J, Fury W, Miloscio L, Trejos J, Pangilinan J, et al. Activin A more prominently regulates muscle mass in primates than does GDF8. Nat Commun. 2017;8:15153.
187. Vanhoutte F, Liang S, Ruddy M, Zhao A, Drewery T, Wang Y, et al. Pharmacokinetics and Pharmacodynamics of Garetosmab (Anti-Activin A): Results From a First-in-Human Phase 1 Study. J Clin Pharmacol. 2020;60(11):1424-31.
188. Yamamoto R, Matsushita M, Kitoh H, Masuda A, Ito M, Katagiri T, et al. Clinically applicable antianginal agents suppress osteoblastic transformation of myogenic cells and heterotopic ossifications in mice. J Bone Miner Metab. 2013;31(1):26-33.
189. Kitoh H, Achiwa M, Kaneko H, Mishima K, Matsushita M, Kadono I, et al. Perhexiline maleate in the treatment of fibrodysplasia ossificans progressiva: an open-labeled clinical trial. Orphanet J Rare Dis. 2013;8:163.
190. Molinuevo MS, Schurman L, McCarthy AD, Cortizo AM, Tolosa MJ, Gangoiti MV, et al. Effect of metformin on bone marrow progenitor cell differentiation: in vivo and in vitro studies. J Bone Miner Res. 2010;25(2):211-21.

191. Jang WG, Kim EJ, Bae IH, Lee KN, Kim YD, Kim DK, et al. Metformin induces osteoblast differentiation via orphan nuclear receptor SHP-mediated transactivation of Runx2. Bone. 2011;48(4):885-93.
192. Lin H, Shi F, Jiang S, Wang Y, Zou J, Ying Y, et al. Metformin attenuates trauma-induced heterotopic ossification via inhibition of Bone Morphogenetic Protein signalling. J Cell Mol Med. 2020;24(24):14491-501.
193. Yu PB, Hong CC, Sachidanandan C, Babitt JL, Deng DY, Hoyng SA, et al. Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. Nat Chem Biol. 2008;4(1):33-41.
194. Williams E, Bullock AN. Structural basis for the potent and selective binding of LDN-212854 to the BMP receptor kinase ALK2. Bone. 2018;109:251-8.
195. Strong AL, Spreadborough PJ, Dey D, Yang P, Li S, Lee A, et al. BMP Ligand Trap ALK3-Fc Attenuates Osteogenesis and Heterotopic Ossification in Blast-Related Lower Extremity Trauma. Stem Cells Dev. 2021;30(2):91-105.
196. Mohedas AH, Wang Y, Sanvitale CE, Canning P, Choi S, Xing X, et al. Structure-activity relationship of 3,5-diaryl-2-aminopyridine ALK2 inhibitors reveals unaltered binding affinity for fibrodysplasia ossificans progressiva causing mutants. J Med Chem. 2014;57(19):7900-15.
197. Hao J, Ho JN, Lewis JA, Karim KA, Daniels RN, Gentry PR, et al. In vivo structure-activity relationship study of dorsomorphin analogues identifies selective VEGF and BMP inhibitors. ACS Chem Biol. 2010;5(2):245-53.
198. Tsugawa D, Oya Y, Masuzaki R, Ray K, Engers DW, Dib M, et al. Specific activin receptor-like kinase 3 inhibitors enhance liver regeneration. J Pharmacol Exp Ther. 2014;351(3):549-58.
199. Hino K, Zhao C, Horigome K, Nishio M, Okanishi Y, Nagata S, et al. An mTOR Signaling Modulator Suppressed Heterotopic Ossification of Fibrodysplasia Ossificans Progressiva. Stem Cell Reports. 2018;11(5):1106-19.
200. Kitoh H. Clinical Aspects and Current Therapeutic Approaches for FOP. Biomedicines. 2020;8(9).
201. Hildebrandt S, Kampfrath B, Fischer K, Hildebrand L, Haupt J, Stachelscheid H, et al. ActivinA Induced SMAD1/5 Signaling in an iPSC Derived EC Model of Fibrodysplasia Ossificans Progressiva (FOP) Can Be Rescued by the Drug Candidate Saracatinib. Stem Cell Rev Rep. 2021;17(3):1039-52.
202. Williams E, Bagarova J, Kerr G, Xia DD, Place ES, Dey D, et al. Saracatinib is an efficacious clinical candidate for fibrodysplasia ossificans progressiva. JCI Insight. 2021;6(8).
203. Pang J, Zuo Y, Chen Y, Song L, Zhu Q, Yu J, et al. ACVR1-Fc suppresses BMP signaling and chondro-osseous differentiation in an in vitro model of Fibrodysplasia ossificans progressiva. Bone. 2016;92:29-36.
204. Kaplan FS, Pignolo RJ, Al Mukaddam MM, Shore EM. Hard targets for a second skeleton: therapeutic horizons for fibrodysplasia ossificans progressiva (FOP). Expert Opin Orphan Drugs. 2017;5(4):291-4.
205. Kaplan FS, Pignolo RJ, Shore EM. From mysteries to medicines: drug development for fibrodysplasia ossificans progressive. Expert Opin Orphan Drugs. 2013;1(8):637-49.
206. Cappato S, Tonachini L, Giacopelli F, Tirone M, Galietta LJ, Sormani M, et al. High-throughput screening for modulators of ACVR1 transcription: discovery of potential therapeutics for fibrodysplasia ossificans progressiva. Dis Model Mech. 2016;9(6):685-96.
207. Tirone M, Giovenzana A, Vallone A, Zordan P, Sormani M, Nicolosi PA, et al. Severe Heterotopic Ossification in the Skeletal Muscle and Endothelial Cells Recruitment to

Chondrogenesis Are Enhanced by Monocyte/Macrophage Depletion. Front Immunol. 2019;10:1640.
208. Wang ZQ, Lu YQ, Han JX. MicroRNAs: important mediators of ossification. Chin Med J (Engl). 2012;125(22):4111-6.
209. Mura M, Cappato S, Giacopelli F, Ravazzolo R, Bocciardi R. The role of the 3'UTR region in the regulation of the ACVR1/Alk-2 gene expression. PloS one. 2012;7(12):e50958.
210. Oishi T, Uezumi A, Kanaji A, Yamamoto N, Yamaguchi A, Yamada H, et al. Osteogenic differentiation capacity of human skeletal muscle-derived progenitor cells. PLoS One. 2013;8(2):e56641.
211. Guerit D, Philipot D, Chuchana P, Toupet K, Brondello JM, Mathieu M, et al. Sox9-regulated miRNA-574-3p inhibits chondrogenic differentiation of mesenchymal stem cells. PloS one. 2013;8(4):e62582.
212. Sun Y, Cai J, Yu S, Chen S, Li F, Fan C. MiR-630 Inhibits Endothelial-Mesenchymal Transition by Targeting Slug in Traumatic Heterotopic Ossification. Sci Rep. 2016;6:22729.
213. Qin X, Zhu B, Jiang T, Tan J, Wu Z, Yuan Z, et al. miR-17-5p Regulates Heterotopic Ossification by Targeting ANKH in Ankylosing Spondylitis. Mol Ther Nucleic Acids. 2019;18:696-707.
214. Liu N, Zhang Z, Li L, Shen X, Sun B, Wang R, et al. MicroRNA-181 regulates the development of Ossification of Posterior longitudinal ligament via Epigenetic Modulation by targeting PBX1. Theranostics. 2020;10(17):7492-509.
215. Aartsma-Rus A, Fokkema I, Verschuuren J, Ginjaar I, van Deutekom J, van Ommen GJ, et al. Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. Hum Mutat. 2009;30(3):293-9.
216. Shi S, Cai J, de Gorter DJ, Sanchez-Duffhues G, Kemaladewi DU, Hoogaars WM, et al. Antisense-oligonucleotide mediated exon skipping in activin-receptor-like kinase 2: inhibiting the receptor that is overactive in fibrodysplasia ossificans progressiva. PloS one. 2013;8(7):e69096.
217. Shi F, Gao J, Zou J, Ying Y, Lin H. Targeting heterotopic ossification by inhibiting activin receptorlike kinase 2 function (Review). Mol Med Rep. 2019;20(4):2979-89.
218. Kaplan J, Kaplan FS, Shore EM. Restoration of normal BMP signaling levels and osteogenic differentiation in FOP mesenchymal progenitor cells by mutant allele-specific targeting. Gene Ther. 2012;19(7):786-90.
219. Shrivats AR, Hsu E, Averick S, Klimak M, Watt AC, DeMaio M, et al. Cationic Nanogel-mediated Runx2 and Osterix siRNA Delivery Decreases Mineralization in MC3T3 Cells. Clin Orthop Relat Res. 2015;473(6):2139-49.
220. Patil S, Dang K, Zhao X, Gao Y, Qian A. Role of LncRNAs and CircRNAs in Bone Metabolism and Osteoporosis. Front Genet. 2020;11:584118.
221. Ghafouri-Fard S, Abak A, Tavakkoli Avval S, Rahmani S, Shoorei H, Taheri M, et al. Contribution of miRNAs and lncRNAs in osteogenesis and related disorders. Biomed Pharmacother. 2021;142:111942.
222. Liu L, Li Z, Chen S, Cui H, Li X, Dai G, et al. BRD4 promotes heterotopic ossification through upregulation of LncRNA MANCR. Bone Joint Res. 2021;10(10):668-76.
223. Hatzikotoulas K, Pickering GA, Clark MJ, Felix-Ilemhenbhio F, Kocsy K, Simpson J, et al. Genome-wide association and functional analyses identify CASC20 and KIF26B as target loci in heterotopic ossification. bioRxiv. 2019:845958.

224. Pickering GAE, Felix-Ilemhenbhio F, Clark MJ, Kocsy K, Simpson J, Bellantuono I, et al. The Kinesin Gene KIF26B Modulates the Severity of Post-Traumatic Heterotopic Ossification. Int J Mol Sci. 2022;23(16).
225. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641-9.
226. Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, et al. An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet. 2019;51(2):258-66.
227. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8.
228. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. Nature. 2019;571(7766):489-99.
229. Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity (Edinb). 2010;105(1):4-13.
230. Greally JM. A user's guide to the ambiguous word 'epigenetics'. Nat Rev Mol Cell Biol. 2018;19(4):207-8.
231. Miller JL, Grant PA. The role of DNA methylation and histone modifications in transcriptional regulation in humans. Subcell Biochem. 2013;61:289-317.
232. Moosavi A, Motevalizadeh Ardekani A. Role of Epigenetics in Biology and Human Diseases. Iran Biomed J. 2016;20(5):246-58.
233. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011;3(3):267-77.
234. Breton CV, Landon R, Kahn LG, Enlow MB, Peterson AK, Bastain T, et al. Exploring the evidence for epigenetic regulation of environmental influences on child health across generations. Commun Biol. 2021;4(1):769.
235. Tiffon C. The Impact of Nutrition and Environmental Epigenetics on Human Health and Disease. Int J Mol Sci. 2018;19(11).
236. Meda F, Folci M, Baccarelli A, Selmi C. The epigenetics of autoimmunity. Cell Mol Immunol. 2011;8(3):226-36.
237. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27-36.
238. Kiefer JC. Epigenetics in development. Dev Dyn. 2007;236(4):1144-56.
239. Pal S, Tyler JK. Epigenetics and aging. Sci Adv. 2016;2(7):e1600584.
240. Trerotola M, Relli V, Simeone P, Alberti S. Epigenetic inheritance and the missing heritability. Hum Genomics. 2015;9(1):17.
241. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. Cell. 2014;157(1):95-109.
242. Feng L, Chen X. Epigenetic regulation of germ cells-remember or forget? Curr Opin Genet Dev. 2015;31:20-7.
243. Cunningham AM, Walker DM, Ramakrishnan A, Doyle MA, Bagot RC, Cates HM, et al. Sperm transcriptional state associated with paternal transmission of stress phenotypes. J Neurosci. 2021;41(29):6202-16.
244. Casas E, Vavouri T. Mechanisms of epigenetic inheritance of variable traits through the germline. Reproduction. 2020;159(6):R251-r63.
245. Rozek LS, Dolinoy DC, Sartor MA, Omenn GS. Epigenetics: relevance and implications for public health. Annu Rev Public Health. 2014;35:105-22.

246. Wang K, Liu H, Hu Q, Wang L, Liu J, Zheng Z, et al. Epigenetic regulation of aging: implications for interventions of aging and diseases. Signal Transduct Target Ther. 2022;7(1):374.
247. Day JJ. New approaches to manipulating the epigenome. Dialogues Clin Neurosci. 2014;16(3):345-57.
248. Jurkowska RZ, Jeltsch A. Silencing of gene expression by targeted DNA methylation: concepts and approaches. Methods Mol Biol. 2010;649:149-61.
249. Wolffe AP, Kurumizaka H. The nucleosome: a powerful regulator of transcription. Prog Nucleic Acid Res Mol Biol. 1998;61:379-422.
250. Sahu RK, Singh S, Tomar RS. The mechanisms of action of chromatin remodelers and implications in development and disease. Biochem Pharmacol. 2020;180:114200.
251. Zhang P, Torres K, Liu X, Liu CG, Pollock RE. An Overview of Chromatin-Regulating Proteins in Cells. Curr Protein Pept Sci. 2016;17(5):401-10.
252. Sadakierska-Chudy A, Filip M. A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. Neurotox Res. 2015;27(2):172-97.
253. Han P, Chang CP. Long non-coding RNA and chromatin remodeling. RNA Biol. 2015;12(10):1094-8.
254. Grossi E, Raimondi I, Goñi E, González J, Marchese FP, Chapaprieta V, et al. A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions. Nat Commun. 2020;11(1):936.
255. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689-93.
256. Brockdorff N. Noncoding RNA and Polycomb recruitment. Rna. 2013;19(4):429-42.
257. Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature. 2013;503(7476):371-6.
258. Arab K, Park YJ, Lindroth AM, Schäfer A, Oakes C, Weichenhan D, et al. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. Mol Cell. 2014;55(4):604-14.
259. Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. Science. 2003;301(5634):798-802.
260. Tamaru H. Confining euchromatin/heterochromatin territory: jumonji crosses the line. Genes Dev. 2010;24(14):1465-78.
261. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974;184(4139):868-71.
262. Simpson B, Tupper C, Al Aboud NM. Genetics, DNA Packaging. StatPearls. Treasure Island (FL): StatPearls Publishing

263. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251-60.
264. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol. 2002;319(5):1097-113.
265. Makde RD, England JR, Yennawar HP, Tan S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. Nature. 2010;467(7315):562-6.
266. Kornberg RD, Lorch Y. Primary Role of the Nucleosome. Mol Cell. 2020;79(3):371-5.

267. Arya G, Schlick T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. Proc Natl Acad Sci U S A. 2006;103(44):16236-41.
268. Iwasaki W, Miya Y, Horikoshi N, Osakabe A, Taguchi H, Tachiwana H, et al. Contribution of histone N-terminal tails to the structure and stability of nucleosomes. FEBS Open Bio. 2013;3:363-9.
269. Lorch Y, Kornberg RD, Maier-Davis B. Role of the histone tails in histone octamer transfer. Nucleic Acids Res. 2023;51(8):3671-8.
270. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381-95.
271. Vettese-Dadey M, Grant PA, Hebbes TR, Crane- Robinson C, Allis CD, Workman JL. Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. Embo j. 1996;15(10):2508-18.
272. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science. 2006;311(5762):844-7.
273. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. Annu Rev Biochem. 2001;70:81-120.
274. Hodawadekar SC, Marmorstein R. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. Oncogene. 2007;26(37):5528-40.
275. Seto E, Yoshida M. Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol. 2014;6(4):a018713.
276. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, et al. Active genes are tri-methylated at K4 of histone H3. Nature. 2002;419(6905):407-11.
277. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311-8.
278. Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, et al. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A. 2002;99(13):8695-700.
279. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet. 2012;13(5):343-57.
280. Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science. 2006;312(5774):748-51.
281. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693-705.
282. Sims RJ, 3rd, Chen CF, Santos-Rosa H, Kouzarides T, Patel SS, Reinberg D. Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem. 2005;280(51):41789-92.
283. Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, 3rd, Grant PA. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature. 2005;433(7024):434-8.
284. Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. Genes Dev. 2013;27(12):1318-38.
285. Grandy RA, Whitfield TW, Wu H, Fitzgerald MP, VanOudenhove JJ, Zaidi SK, et al. Genome-Wide Studies Reveal that H3K4me3 Modification in Bivalent Genes Is Dynamically Regulated during the Pluripotent Cell Cycle and Stabilized upon Differentiation. Mol Cell Biol. 2016;36(4):615-27.

286. Rojas A, Aguilar R, Henriquez B, Lian JB, Stein JL, Stein GS, et al. Epigenetic Control of the Bone-master Runx2 Gene during Osteoblast-lineage Commitment by the Histone Demethylase JARID1B/KDM5B. J Biol Chem. 2015;290(47):28329-42.
287. Sepulveda H, Villagra A, Montecino M. Tet-Mediated DNA Demethylation Is Required for SWI/SNF-Dependent Chromatin Remodeling and Histone-Modifying Activities That Trigger Expression of the Sp7 Osteoblast Master Gene during Mesenchymal Lineage Commitment. Mol Cell Biol. 2017;37(20).
288. Byvoet P, Shepherd GR, Hardin JM, Noland BJ. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. Arch Biochem Biophys. 1972;148(2):558-67.
289. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell. 2004;119(7):941-53.
290. Rossetto D, Avvakumov N, Côté J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics. 2012;7(10):1098-108.
291. Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ, Jackson SP. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell. 2005;123(7):1213-26.
292. Wei Y, Mizzen CA, Cook RG, Gorovsky MA, Allis CD. Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proc Natl Acad Sci U S A. 1998;95(13):7480-4.
293. Rossetto D, Truman AW, Kron SJ, Côté J. Epigenetic modifications in double-strand break DNA damage signaling and repair. Clin Cancer Res. 2010;16(18):4543-52.
294. van Attikum H, Gasser SM. The histone code at DNA breaks: a guide to repair? Nat Rev Mol Cell Biol. 2005;6(10):757-65.
295. Komar D, Juszczynski P. Rebelled epigenome: histone H3S10 phosphorylation and H3S10 kinases in cancer biology and therapy. Clin Epigenetics. 2020;12(1):147.
296. Jeong YS, Cho S, Park JS, Ko Y, Kang YK. Phosphorylation of serine-10 of histone H3 shields modified lysine-9 selectively during mitosis. Genes Cells. 2010;15(3):181-92.
297. Castellano-Pozo M, Santos-Pereira JM, Rondón AG, Barroso S, Andújar E, Pérez-Alegre M, et al. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell. 2013;52(4):583-90.
298. Choi HS, Choi BY, Cho YY, Mizuno H, Kang BS, Bode AM, et al. Phosphorylation of histone H3 at serine 10 is indispensable for neoplastic cell transformation. Cancer Res. 2005;65(13):5818-27.
299. Chadee DN, Hendzel MJ, Tylipski CP, Allis CD, Bazett-Jones DP, Wright JA, et al. Increased Ser-10 phosphorylation of histone H3 in mitogen-stimulated and oncogene-transformed mouse fibroblasts. J Biol Chem. 1999;274(35):24914-20.
300. Li B, Huang G, Zhang X, Li R, Wang J, Dong Z, et al. Increased phosphorylation of histone H3 at serine 10 is involved in Epstein-Barr virus latent membrane protein-1-induced carcinogenesis of nasopharyngeal carcinoma. BMC Cancer. 2013;13:124.
301. Philip S, Kumarasiri M, Teo T, Yu M, Wang S. Cyclin-Dependent Kinase 8: A New Hope in Targeted Cancer Therapy? J Med Chem. 2018;61(12):5073-92.
302. Tang A, Gao K, Chu L, Zhang R, Yang J, Zheng J. Aurora kinases: novel therapy targets in cancers. Oncotarget. 2017;8(14):23937-54.
303. Mattiroli F, Penengo L. Histone Ubiquitination: An Integrative Signaling Platform in Genome Stability. Trends Genet. 2021;37(6):566-81.

304. Ryu HY, Hochstrasser M. Histone sumoylation and chromatin dynamics. Nucleic Acids Res. 2021;49(11):6043-52.
305. Liu YC, Penninger J, Karin M. Immunity by ubiquitylation: a reversible process of modification. Nat Rev Immunol. 2005;5(12):941-52.
306. Cao J, Yan Q. Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. Front Oncol. 2012;2:26.
307. Li J, Chai QY, Liu CH. The ubiquitin system: a critical regulator of innate immunity and pathogen-host interactions. Cell Mol Immunol. 2016;13(5):560-76.
308. Huen MS, Grant R, Manke I, Minn K, Yu X, Yaffe MB, et al. RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell. 2007;131(5):901-14.
309. Kolas NK, Chapman JR, Nakada S, Ylanko J, Chahwan R, Sweeney FD, et al. Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science. 2007;318(5856):1637-40.
310. Shanbhag NM, Rafalska-Metcalf IU, Balane-Bolivar C, Janicki SM, Greenberg RA. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. Cell. 2010;141(6):970-81.
311. Zhu B, Zheng Y, Pham AD, Mandal SS, Erdjument-Bromage H, Tempst P, et al. Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell. 2005;20(4):601-11.
312. Joo HY, Zhai L, Yang C, Nie S, Erdjument-Bromage H, Tempst P, et al. Regulation of cell cycle progression and gene expression by H2A deubiquitination. Nature. 2007;449(7165):1068-72.
313. Matsui SI, Seon BK, Sandberg AA. Disappearance of a structural chromatin protein A24 in mitosis: implications for molecular basis of chromatin condensation. Proc Natl Acad Sci U S A. 1979;76(12):6386-90.
314. Goldknopf IL, Taylor CW, Baum RM, Yeoman LC, Olson MO, Prestayko AW, et al. Isolation and characterization of protein A24, a "histone-like" non-histone chromosomal protein. J Biol Chem. 1975;250(18):7182-7.
315. West MH, Bonner WM. Histone 2B can be modified by the attachment of ubiquitin. Nucleic Acids Res. 1980;8(20):4671-80.
316. Cao R, Tsukada Y, Zhang Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol Cell. 2005;20(6):845-54.
317. Gearhart MD, Corcoran CM, Wamstad JA, Bardwell VJ. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol Cell Biol. 2006;26(18):6880-9.
318. Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, et al. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell. 2008;29(1):69-80.
319. Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol. 2008;10(4):483-8.
320. Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, Johnston M, et al. Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem. 2002;277(32):28368-71.
321. Yu J, Qin B, Lou Z. Ubiquitin and ubiquitin-like molecules in DNA double strand break repair. Cell Biosci. 2020;10:13.

322. Wang Y, Ladunga I, Miller AR, Horken KM, Plucinak T, Weeks DP, et al. The small ubiquitin-like modifier (SUMO) and SUMO-conjugating system of Chlamydomonas reinhardtii. Genetics. 2008;179(1):177-92.
323. Chen WT, Alpert A, Leiter C, Gong F, Jackson SP, Miller KM. Systematic identification of functional residues in mammalian histone H2AX. Mol Cell Biol. 2013;33(1):111-26.
324. Shiio Y, Eisenman RN. Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A. 2003;100(23):13225-30.
325. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim Biophys Acta. 2014;1839(12):1362-72.
326. Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, et al. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature. 2005;438(7071):1116-22.
327. Nelson CJ, Santos-Rosa H, Kouzarides T. Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell. 2006;126(5):905-16.
328. Song JJ, Garlick JD, Kingston RE. Structural basis of histone H4 recognition by p55. Genes Dev. 2008;22(10):1313-8.
329. Wapenaar H, Dekker FJ. Histone acetyltransferases: challenges in targeting bi-substrate enzymes. Clin Epigenetics. 2016;8:59.
330. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Exp Mol Med. 2017;49(4):e324.
331. Ding Y, Yao Y, Gong X, Zhuo Q, Chen J, Tian M, et al. JMJD3: a critical epigenetic regulator in stem cell fate. Cell Commun Signal. 2021;19(1):72.
332. Zhang J, Jing L, Li M, He L, Guo Z. Regulation of histone arginine methylation/demethylation by methylase and demethylase (Review). Mol Med Rep. 2019;19(5):3963-71.
333. So CC, Ramachandran S, Martin A. E3 Ubiquitin Ligases RNF20 and RNF40 Are Required for Double-Stranded Break (DSB) Repair: Evidence for Monoubiquitination of Histone H2B Lysine 120 as a Novel Axis of DSB Signaling and Repair. Mol Cell Biol. 2019;39(8).
334. Roman-Trufero M, Dillon N. The UBE2D ubiquitin conjugating enzymes: Potential regulatory hubs in development, disease and evolution. Front Cell Dev Biol. 2022;10:1058751.
335. Gold M, Gefter M, Hausmann R, Hurwitz J. Methylation of DNA. J Gen Physiol. 1966;49(6):5-28.
336. Dhar GA, Saha S, Mitra P, Nag Chaudhuri R. DNA methylation and regulation of gene expression: Guardian of our health. Nucleus (Calcutta). 2021;64(3):259-70.
337. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. 2014;28(8):812-28.
338. Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol Cell. 2012;47(4):633-47.
339. Suelves M, Carrió E, Núñez-Álvarez Y, Peinado MA. DNA methylation dynamics in cellular commitment and differentiation. Brief Funct Genomics. 2016;15(6):443-53.
340. Zhou J, Sears RL, Xing X, Zhang B, Li D, Rockweiler NB, et al. Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. BMC Genomics. 2017;18(1):724.
341. Gensous N, Bacalini MG, Franceschi C, Meskers CGM, Maier AB, Garagnani P. Age-Related DNA Methylation Changes: Potential Impact on Skeletal Muscle Aging in Humans. Front Physiol. 2019;10:996.

342. Ehrlich M. DNA hypermethylation in disease: mechanisms and clinical relevance. Epigenetics. 2019;14(12):1141-63.
343. Nishiyama A, Nakanishi M. Navigating the DNA methylation landscape of cancer. Trends Genet. 2021;37(11):1012-27.
344. Reynard LN. Analysis of genetics and DNA methylation in osteoarthritis: What have we learnt about the disease? Semin Cell Dev Biol. 2017;62:57-66.
345. Rasmi Y, Shokati A, Hassan A, Aziz SG, Bastani S, Jalali L, et al. The role of DNA methylation in progression of neurological disorders and neurodegenerative diseases as well as the prospect of using DNA methylation inhibitors as therapeutic agents for such disorders. IBRO Neurosci Rep. 2023;14:28-37.
346. Barres R, Zierath JR. DNA methylation in metabolic disorders. Am J Clin Nutr. 2011;93(4):897s-900.
347. Lam D, Clark S, Stirzaker C, Pidsley R. Advances in Prognostic Methylation Biomarkers for Prostate Cancer. Cancers (Basel). 2020;12(10).
348. Levenson VV. DNA methylation as a universal biomarker. Expert Rev Mol Diagn. 2010;10(4):481-8.
349. Tavares NT, Gumauskaitė S, Lobo J, Jerónimo C, Henrique R. DNA Methylation Biomarkers for Prediction of Response to Platinum-Based Chemotherapy: Where Do We Stand? Cancers (Basel). 2022;14(12).
350. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer. 2011;2(6):607-17.
351. Petryk N, Bultmann S, Bartke T, Defossez PA. Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. Nucleic Acids Res. 2021;49(6):3020-32.
352. Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S, et al. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun. 2020;11(1):3355.
353. Baccarelli A, Bollati V. Epigenetics and environmental chemicals. Curr Opin Pediatr. 2009;21(2):243-51.
354. Costantino S, Libby P, Kishore R, Tardif JC, El-Osta A, Paneni F. Epigenetics and precision medicine in cardiovascular patients: from basic concepts to the clinical arena. Eur Heart J. 2018;39(47):4150-8.
355. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife. 2013;2:e00523.
356. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472-9.
357. Tan L, Shi YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development. 2012;139(11):1895-902.
358. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300-3.
359. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 2016;30(7):733-50.
360. Kim M, Costello J. DNA methylation: an epigenetic mark of cellular memory. Exp Mol Med. 2017;49(4):e322.
361. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23-38.

362. Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet. 2009;10(11):805-11.
363. Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H. The role of DNA methylation in setting up chromatin structure during development. Nat Genet. 2003;34(2):187-92.
364. Bogdanović O, Veenstra GJ. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. Chromosoma. 2009;118(5):549-65.
365. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MS, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014;15:119.
366. Wolf SF, Jolly DJ, Lunnen KD, Friedmann T, Migeon BR. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci U S A. 1984;81(9):2806-10.
367. Sharifi-Zarchi A, Gerovska D, Adachi K, Totonchi M, Pezeshk H, Taft RJ, et al. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. BMC Genomics. 2017;18(1):964.
368. Rauluseviciute I, Drabløs F, Rye MB. DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. BMC Med Genomics. 2020;13(1):6.
369. Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, et al. CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc Natl Acad Sci U S A. 2010;107(47):20311-6.
370. Hershey JW, Sonenberg N, Mathews MB. Principles of translational control: an overview. Cold Spring Harb Perspect Biol. 2012;4(12).
371. Lacal I, Ventura R. Epigenetic Inheritance: Concepts, Mechanisms and Perspectives. Front Mol Neurosci. 2018;11:292.
372. Aristizabal MJ, Anreiter I, Halldorsdottir T, Odgers CL, McDade TW, Goldenberg A, et al. Biological embedding of experience: A primer on epigenetics. Proc Natl Acad Sci U S A. 2020;117(38):23261-9.
373. Leibovitch M, Topisirovic I. Dysregulation of mRNA translation and energy metabolism in cancer. Adv Biol Regul. 2018;67:30-9.
374. Woodward K, Shirokikh NE. Translational control in cell ageing: an update. Biochem Soc Trans. 2021;49(6):2853-69.
375. Jishi A, Qi X, Miranda HC. Implications of mRNA translation dysregulation for neurological disorders. Semin Cell Dev Biol. 2021;114:11-9.
376. Gebauer F, Hentze MW. Molecular mechanisms of translational control. Nat Rev Mol Cell Biol. 2004;5(10):827-35.
377. Aiso T, Yoshida H, Wada A, Ohki R. Modulation of mRNA stability participates in stationary-phase-specific expression of ribosome modulation factor. J Bacteriol. 2005;187(6):1951-8.
378. Kiss DL, Vasudevan D, Ho CK, Caliskan N. Editorial: mRNA Translational Control as a Mechanism of Post-transcriptional Gene Regulation. Front Mol Biosci. 2022;9:947516.
379. Roy B, von Arnim AG. Translational Regulation of Cytoplasmic mRNAs. Arabidopsis Book. 2013;11:e0165.
380. Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 2013;9(8):e1003529.
381. Chen HH, Tarn WY. uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. RNA Biol. 2019;16(10):1327-38.

382. Bonnal S, Boutonnet C, Prado-Lourenço L, Vagner S. IRESdb: the Internal Ribosome Entry Site database. Nucleic Acids Res. 2003;31(1):427-8.
383. Komar AA, Hatzoglou M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. Cell Cycle. 2011;10(2):229-40.
384. Ratti M, Lampis A, Ghidini M, Salati M, Mirchev MB, Valeri N, et al. MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside. Target Oncol. 2020;15(3):261-78.
385. Leppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat Rev Mol Cell Biol. 2018;19(3):158-74.
386. Karakas D, Ozpolat B. The Role of LncRNAs in Translation. Noncoding RNA. 2021;7(1).
387. Ying SY, Chang DC, Lin SL. The microRNA (miRNA): overview of the RNA genes that modulate gene function. Mol Biotechnol. 2008;38(3):257-68.
388. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell. 2009;101(5):251-62.
389. Boo SH, Kim YK. The emerging role of RNA modifications in the regulation of mRNA stability. Exp Mol Med. 2020;52(3):400-8.
390. Gebauer F, Schwarzl T, Valcárcel J, Hentze MW. RNA-binding proteins in human genetic disease. Nat Rev Genet. 2021;22(3):185-98.
391. de la Parra C, Walters BA, Geter P, Schneider RJ. Translation initiation factors and their relevance in cancer. Curr Opin Genet Dev. 2018;48:82-8.
392. Mauro VP, Matsuda D. Translation regulation by ribosomes: Increased complexity and expanded scope. RNA Biol. 2016;13(9):748-55.
393. Zhang M, Zhao Y, Zhang Y, Wang D, Gu S, Feng W, et al. LncRNA UCA1 promotes migration and invasion in pancreatic cancer cells via the Hippo pathway. Biochim Biophys Acta Mol Basis Dis. 2018;1864(5 Pt A):1770-82.
394. Chen J, Guo J, Cui X, Dai Y, Tang Z, Qu J, et al. The Long Noncoding RNA LnRPT Is Regulated by PDGF-BB and Modulates the Proliferation of Pulmonary Artery Smooth Muscle Cells. Am J Respir Cell Mol Biol. 2018;58(2):181-93.
395. Schlosser K, Hanson J, Villeneuve PJ, Dimitroulakos J, McIntyre L, Pilote L, et al. Assessment of Circulating LncRNAs Under Physiologic and Pathologic Conditions in Humans Reveals Potential Limitations as Biomarkers. Sci Rep. 2016;6:36596.
396. Standaert L, Adriaens C, Radaelli E, Van Keymeulen A, Blanpain C, Hirose T, et al. The long noncoding RNA Neat1 is required for mammary gland development and lactation. Rna. 2014;20(12):1844-9.
397. Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. Cancer Cell. 2016;29(4):452-63.
398. Zhang J, Liu SC, Luo XH, Tao GX, Guan M, Yuan H, et al. Exosomal Long Noncoding RNAs are Differentially Expressed in the Cervicovaginal Lavage Samples of Cervical Cancer Patients. J Clin Lab Anal. 2016;30(6):1116-21.
399. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013;10(6):925-33.
400. Dong P, Xiong Y, Yue J, Hanley SJB, Kobayashi N, Todo Y, et al. Long Non-coding RNA NEAT1: A Novel Target for Diagnosis and Therapy in Human Tumors. Front Genet. 2018;9:471.
401. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353-8.
402. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10(3):155-9.

403. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011;30(16):1956-62.
404. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010;28(5):503-10.
405. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S, et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. 2004;14(3):331-42.
406. Zhang X, Rice K, Wang Y, Chen W, Zhong Y, Nakayama Y, et al. Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. Endocrinology. 2010;151(3):939-47.
407. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775-89.
408. Leygue E. Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. Nucl Recept Signal. 2007;5:e006.
409. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. Genetics. 2013;193(3):651-69.
410. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26-46.
411. St Laurent G, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. Trends Genet. 2015;31(5):239-51.
412. Sartorelli V, Lauberth SM. Enhancer RNAs are an important regulatory layer of the epigenome. Nat Struct Mol Biol. 2020;27(6):521-8.
413. Li Y, Syed J, Sugiyama H. RNA-DNA Triplex Formation by Long Noncoding RNAs. Cell Chem Biol. 2016;23(11):1325-33.
414. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell. 2014;55(5):791-802.
415. Chen LL. Linking Long Noncoding RNA Localization and Function. Trends Biochem Sci. 2016;41(9):761-72.
416. Chujo T, Hirose T. Nuclear Bodies Built on Architectural Long Noncoding RNAs: Unifying Principles of Their Construction and Function. Mol Cells. 2017;40(12):889-96.
417. Gourvest M, Brousset P, Bousquet M. Long Noncoding RNAs in Acute Myeloid Leukemia: Functional Characterization and Clinical Relevance. Cancers (Basel). 2019;11(11).
418. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. 2012;22(3):577-91.
419. Erdmann VA, Szymanski M, Hochberg A, de Groot N, Barciszewski J. Collection of mRNA-like non-coding RNAs. Nucleic Acids Res. 1999;27(1):192-5.
420. Lottin S, Vercoutter-Edouart AS, Adriaenssens E, Czeszak X, Lemoine J, Roudbaraki M, et al. Thioredoxin post-transcriptional regulation by H19 provides a new function to mRNA-like non-coding RNA. Oncogene. 2002;21(10):1625-31.
421. Berteaux N, Lottin S, Adriaenssens E, Van Coppenolle F, Leroy X, Coll J, et al. Hormonal regulation of H19 gene expression in prostate epithelial cells. J Endocrinol. 2004;183(1):69-78.

422. Chen R, Wang SK, Belk JA, Amaya L, Li Z, Cardenas A, et al. Engineering circular RNA for enhanced protein production. Nat Biotechnol. 2023;41(2):262-72.
423. Wu H, Yang L, Chen LL. The Diversity of Long Noncoding RNAs and Their Generation. Trends Genet. 2017;33(8):540-52.
424. Ranganathan K, Sivasankar V. MicroRNAs - Biology and clinical applications. J Oral Maxillofac Pathol. 2014;18(2):229-34.
425. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.
426. Zhiyanov A, Nersisyan S, Tonevitsky A. Hairpin sequence and structure is associated with features of isomiR biogenesis. RNA Biol. 2021;18(sup1):430-8.
427. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281-97.
428. Perron MP, Provost P. Protein interactions and complexes in human microRNA biogenesis and function. Front Biosci. 2008;13:2537-47.
429. Havens MA, Reich AA, Hastings ML. Drosha promotes splicing of a pre-microRNA-like alternative exon. PLoS Genet. 2014;10(5):e1004312.
430. Roth BM, Ishimaru D, Hennig M. The core microprocessor component DiGeorge syndrome critical region 8 (DGCR8) is a nonspecific RNA-binding protein. J Biol Chem. 2013;288(37):26785-99.
431. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 2004;18(24):3016-27.
432. Westholm JO, Lai EC. Mirtrons: microRNA biogenesis via splicing. Biochimie. 2011;93(11):1897-904.
433. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne). 2018;9:402.
434. de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat Biotechnol. 2017;35(9):872-8.
435. Kim YK, Kim VN. Processing of intronic microRNAs. Embo j. 2007;26(3):775-83.
436. Tanzer A, Stadler PF. Molecular evolution of a microRNA cluster. J Mol Biol. 2004;339(2):327-35.
437. Wu K, He J, Pu W, Peng Y. The Role of Exportin-5 in MicroRNA Biogenesis and Cancer. Genomics Proteomics Bioinformatics. 2018;16(2):120-6.
438. Saito K, Ishizuka A, Siomi H, Siomi MC. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. PLoS Biol. 2005;3(7):e235.
439. Sheng P, Fields C, Aadland K, Wei T, Kolaczkowski O, Gu T, et al. Dicer cleaves 5'-extended microRNA precursors originating from RNA polymerase II transcription start sites. Nucleic Acids Res. 2018;46(11):5737-52.
440. Takimoto K, Wakiyama M, Yokoyama S. Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. Rna. 2009;15(6):1078-89.
441. Braun JE, Huntzinger E, Izaurralde E. The role of GW182 proteins in miRNA-mediated gene silencing. Adv Exp Med Biol. 2013;768:147-63.
442. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 2010;465(7298):584-9.
443. Sala L, Chandrasekhar S, Vidigal JA. AGO unchained: Canonical and non-canonical roles of Argonaute proteins in mammals. Front Biosci (Landmark Ed). 2020;25(1):1-42.

444. Dueck A, Meister G. MicroRNA processing without Dicer. Genome Biol. 2010;11(6):123.
445. Tat TT, Maroney PA, Chamnongpol S, Coller J, Nilsen TW. Cotranslational microRNA mediated messenger RNA destabilization. Elife. 2016;5.
446. Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. Mol Cell. 2016;64(2):320-33.
447. Iftikhar H, Carney GE. Evidence and potential in vivo functions for biofluid miRNAs: From expression profiling to functional testing: Potential roles of extracellular miRNAs as indicators of physiological change and as agents of intercellular information exchange. Bioessays. 2016;38(4):367-78.
448. Bartel DP. Metazoan MicroRNAs. Cell. 2018;173(1):20-51.
449. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet. 2015;16(7):421-33.
450. Ørom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell. 2008;30(4):460-71.
451. Pulik Ł, Mierzejewski B, Sibilska A, Grabowska I, Ciemerych MA, Łęgosz P, et al. The role of miRNA and lncRNA in heterotopic ossification pathogenesis. Stem Cell Res Ther. 2022;13(1):523.
452. Sadakierska-Chudy A. MicroRNAs: Diverse Mechanisms of Action and Their Potential Applications as Cancer Epi-Therapeutics. Biomolecules. 2020;10(9).
453. Truesdell SS, Mortensen RD, Seo M, Schroeder JC, Lee JH, LeTonqueze O, et al. MicroRNA-mediated mRNA translation activation in quiescent cells and oocytes involves recruitment of a nuclear microRNP. Sci Rep. 2012;2:842.
454. Bukhari SIA, Truesdell SS, Lee S, Kollu S, Classon A, Boukhali M, et al. A Specialized Mechanism of Translation Mediated by FXR1a-Associated MicroRNP in Cellular Quiescence. Mol Cell. 2016;61(5):760-73.
455. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science. 2010;328(5986):1694-8.
456. Lewis AM, Jr., Thomas R, Breen M, Peden K, Teferedegne B, Foseh G, et al. The AGMK1-9T7 cell model of neoplasia: Evolution of DNA copy-number aberrations and miRNA expression during transition from normal to metastatic cancer cells. PLoS One. 2022;17(10):e0275394.
457. Kim K, Lee HC, Park JL, Kim M, Kim SY, Noh SM, et al. Epigenetic regulation of microRNA-10b and targeting of oncogenic MAPRE1 in gastric cancer. Epigenetics. 2011;6(6):740-51.
458. Li Z, Hassan MQ, Volinia S, van Wijnen AJ, Stein JL, Croce CM, et al. A microRNA signature for a BMP2-induced osteoblast lineage commitment program. Proc Natl Acad Sci U S A. 2008;105(37):13906-11.
459. Misiewicz-Krzeminska I, Krzeminski P, Corchete LA, Quwaider D, Rojas EA, Herrero AB, et al. Factors Regulating microRNA Expression and Function in Multiple Myeloma. Noncoding RNA. 2019;5(1).
460. Bak RO, Mikkelsen JG. miRNA sponges: soaking up miRNAs for regulation of gene expression. Wiley Interdiscip Rev RNA. 2014;5(3):317-33.
461. Pasquinelli AE. Molecular biology. Paring miRNAs through pairing. Science. 2010;328(5985):1494-5.

462. Cazalla D, Yario T, Steitz JA. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. Science. 2010;328(5985):1563-6.
463. Ghini F, Rubolino C, Climent M, Simeone I, Marzi MJ, Nicassio F. Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. Nat Commun. 2018;9(1):3119.
464. Kato M. Target RNA-directed microRNA degradation; which controls which? Noncoding RNA Investig. 2018;2.
465. Sikora M, Marycz K, Smieszek A. Small and Long Non-coding RNAs as Functional Regulators of Bone Homeostasis, Acting Alone or Cooperatively. Mol Ther Nucleic Acids. 2020;21:792-803.
466. Wu J, Zhao J, Sun L, Pan Y, Wang H, Zhang WB. Long non-coding RNA H19 mediates mechanical tension-induced osteogenesis of bone marrow mesenchymal stem cells via FAK by sponging miR-138. Bone. 2018;108:62-70.
467. Tang Z, Gong Z, Sun X. LncRNA DANCR involved osteolysis after total hip arthroplasty by regulating FOXO1 expression to inhibit osteoblast differentiation. J Biomed Sci. 2018;25(1):4.
468. Wang Q, Li Y, Zhang Y, Ma L, Lin L, Meng J, et al. LncRNA MEG3 inhibited osteogenic differentiation of bone marrow mesenchymal stem cells from postmenopausal osteoporosis by targeting miR-133a-3p. Biomed Pharmacother. 2017;89:1178-86.
469. Xiao X, Zhou T, Guo S, Guo C, Zhang Q, Dong N, et al. LncRNA MALAT1 sponges miR-204 to promote osteoblast differentiation of human aortic valve interstitial cells through up-regulating Smad4. Int J Cardiol. 2017;243:404-12.
470. Wei B, Wei W, Zhao B, Guo X, Liu S. Long non-coding RNA HOTAIR inhibits miR-17-5p to regulate osteogenic differentiation and proliferation in non-traumatic osteonecrosis of femoral head. PLoS One. 2017;12(2):e0169097.
471. Zhang J, Hao X, Yin M, Xu T, Guo F. Long non-coding RNA in osteogenesis: A new world to be explored. Bone Joint Res. 2019;8(2):73-80.
472. Huang Y, Zheng Y, Jia L, Li W. Long Noncoding RNA H19 Promotes Osteoblast Differentiation Via TGF-beta1/Smad3/HDAC Signaling Pathway by Deriving miR-675. Stem Cells. 2015;33(12):3481-92.
473. Jin C, Zheng Y, Huang Y, Liu Y, Jia L, Zhou Y. Long non-coding RNA MIAT knockdown promotes osteogenic differentiation of human adipose-derived stem cells. Cell Biol Int. 2017;41(1):33-41.
474. Tang S, Xie Z, Wang P, Li J, Wang S, Liu W, et al. LncRNA-OG Promotes the Osteogenic Differentiation of Bone Marrow-Derived Mesenchymal Stem Cells Under the Regulation of hnRNPK. Stem Cells. 2019;37(2):270-83.
475. Zhu XX, Yan YW, Chen D, Ai CZ, Lu X, Xu SS, et al. Long non-coding RNA HoxA-AS3 interacts with EZH2 to regulate lineage commitment of mesenchymal stem cells. Oncotarget. 2016;7(39):63561-70.
476. Misawa A, Orimo H. lncRNA HOTAIR Inhibits Mineralization in Osteoblastic Osteosarcoma Cells by Epigenetically Repressing ALPL. Calcif Tissue Int. 2018;103(4):422-30.
477. Chen H, Yang S, Shao R. Long non-coding XIST raises methylation of TIMP-3 promoter to regulate collagen degradation in osteoarthritic chondrocytes after tibial plateau fracture. Arthritis Res Ther. 2019;21(1):271.
478. Xu Y, Wang S, Tang C, Chen W. Upregulation of long non-coding RNA HIF 1alpha-anti-sense 1 induced by transforming growth factor-beta-mediated targeting of sirtuin 1 promotes

osteoblastic differentiation of human bone marrow stromal cells. Mol Med Rep. 2015;12(5):7233-8.
479. Zhu L, Xu PC. Downregulated LncRNA-ANCR promotes osteoblast differentiation by targeting EZH2 and regulating Runx2 expression. Biochem Biophys Res Commun. 2013;432(4):612-7.
480. Jia Q, Jiang W, Ni L. Down-regulated non-coding RNA (lncRNA-ANCR) promotes osteogenic differentiation of periodontal ligament stem cells. Arch Oral Biol. 2015;60(2):234-41.
481. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription? Sci Adv. 2017;3(9):eaao2110.
482. Zhuang W, Ge X, Yang S, Huang M, Zhuang W, Chen P, et al. Upregulation of lncRNA MEG3 Promotes Osteogenic Differentiation of Mesenchymal Stem Cells From Multiple Myeloma Patients By Targeting BMP4 Transcription. Stem Cells. 2015;33(6):1985-97.
483. Chen S, Jia L, Zhang S, Zheng Y, Zhou Y. DEPTOR regulates osteogenic differentiation via inhibiting MEG3-mediated activation of BMP4 signaling and is involved in osteoporosis. Stem Cell Research & Therapy. 2018;9(1):185.
484. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature. 2016;539(7629):452-5.
485. Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, et al. Long noncoding RNAs with snoRNA ends. Mol Cell. 2012;48(2):219-30.
486. Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R, et al. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev. 2008;22(6):756-69.
487. Munroe SH, Lazar MA. Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. J Biol Chem. 1991;266(33):22083-6.
488. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med. 2008;14(7):723-30.
489. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol. 2010;11(5):R56.
490. Nogalska A, Engel WK, Askanas V. Increased BACE1 mRNA and noncoding BACE1-antisense transcript in sporadic inclusion-body myositis muscle fibers--possibly caused by endoplasmic reticulum stress. Neurosci Lett. 2010;474(3):140-3.
491. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature. 2011;470(7333):284-8.
492. Wu Y, Jiang Y, Liu Q, Liu C-Z. lncRNA H19 promotes matrix mineralization through up-regulating IGF1 by sponging miR-185-5p in osteoblasts. BMC Molecular and Cell Biology. 2019;20(1):48.
493. Jia B, Wang Z, Sun X, Chen J, Zhao J, Qiu X. Long noncoding RNA LINC00707 sponges miR-370-3p to promote osteogenesis of human bone marrow-derived mesenchymal stem cells through upregulating WNT2B. Stem Cell Research & Therapy. 2019;10(1):67.
494. Li L, Liu B, Wapinski OL, Tsai MC, Qu K, Zhang J, et al. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. Cell Rep. 2013;5(1):3-12.

495. Wang L, Wu F, Song Y, Li X, Wu Q, Duan Y, et al. Long noncoding RNA related to periodontitis interacts with miR-182 to upregulate osteogenic differentiation in periodontal mesenchymal stem cells of periodontitis patients. Cell Death Dis. 2016;7(8):e2327.
496. Weng J, Peng W, Zhu S, Chen S. Long Noncoding RNA Sponges miR-454 to Promote Osteogenic Differentiation in Maxillary Sinus Membrane Stem Cells. Implant Dent. 2017;26(2):178-86.
497. Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, et al. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol. 2012;14(7):659-65.
498. Savary G, Dewaeles E, Diazzi S, Buscot M, Nottet N, Fassy J, et al. The Long Noncoding RNA DNM3OS Is a Reservoir of FibromiRs with Major Functions in Lung Fibroblast Response to TGF-β and Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019;200(2):184-98.
499. Liang WC, Fu WM, Wang YB, Sun YX, Xu LL, Wong CW, et al. H19 activates Wnt signaling and promotes osteoblast differentiation by functioning as a competing endogenous RNA. Sci Rep. 2016;6:20121.
500. He J, Tu C, Liu Y. Role of lncRNAs in aging and age-related diseases. Aging Med (Milton). 2018;1(2):158-75.
501. Jung HJ, Lee KP, Kwon KS, Suh Y. MicroRNAs in Skeletal Muscle Aging: Current Issues and Perspectives. J Gerontol A Biol Sci Med Sci. 2019;74(7):1008-14.
502. Harfe BD, McManus MT, Mansfield JH, Hornstein E, Tabin CJ. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. Proc Natl Acad Sci U S A. 2005;102(31):10898-903.
503. Kobayashi T, Lu J, Cobb BS, Rodda SJ, McMahon AP, Schipani E, et al. Dicer-dependent pathways regulate chondrocyte proliferation and differentiation. Proc Natl Acad Sci U S A. 2008;105(6):1949-54.
504. Hensley AP, McAlinden A. The role of microRNAs in bone development. Bone. 2021;143:115760.
505. McCarthy JJ. The MyomiR network in skeletal muscle plasticity. Exerc Sport Sci Rev. 2011;39(3):150-4.
506. Xu C, Zhang Z, Liu N, Li L, Zhong H, Wang R, et al. Small extracellular vesicle-mediated miR-320e transmission promotes osteogenesis in OPLL by targeting TAK1. Nat Commun. 2022;13(1):2467.
507. Han Z, Shi F, Chen Y, Dong X, Zhang B, Li M. Relationship between miRNA-433 and SPP1 in the presence of fracture and traumatic brain injury. Exp Ther Med. 2021;22(3):928.
508. Tang Y, Sun Y, Zeng J, Yuan B, Zhao Y, Geng X, et al. Exosomal miR-140-5p inhibits osteogenesis by targeting IGF1R and regulating the mTOR pathway in ossification of the posterior longitudinal ligament. J Nanobiotechnology. 2022;20(1):452.
509. Han Y, Zhang K, Hong Y, Wang J, Liu Q, Zhang Z, et al. miR-342-3p promotes osteogenic differentiation via targeting ATF3. FEBS Lett. 2018;592(24):4051-65.
510. Zhang J, Yu X, Yu Y, Gong Y. MicroRNA expression analysis during FK506-induced osteogenic differentiation in rat bone marrow stromal cells. Mol Med Rep. 2017;16(1):581-90.
511. Zhang SY, Gao F, Peng CG, Zheng CJ, Wu MF. miR-485-5p promotes osteoporosis via targeting Osterix. Eur Rev Med Pharmacol Sci. 2018;22(15):4792-9.
512. Rodriguez AM, Elabd C, Amri EZ, Ailhaud G, Dani C. The human adipose tissue is a source of multipotent stem cells. Biochimie. 2005;87(1):125-8.

513. Sun Q, Hao Q, Prasanth KV. Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. Trends Genet. 2018;34(2):142-57.
514. Rashid F, Shah A, Shan G. Long Non-coding RNAs in the Cytoplasm. Genomics Proteomics Bioinformatics. 2016;14(2):73-80.
515. Paraskevopoulou MD, Hatzigeorgiou AG. Analyzing MiRNA-LncRNA Interactions. Methods Mol Biol. 2016;1402:271-86.
516. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. Database (Oxford). 2010;2010:baq020.
517. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS Biol. 2004;2(11):e363.
518. Gouy M, Milleret F, Mugnier C, Jacobzone M, Gautier C. ACNUC: a nucleic acid sequence data base and analysis system. Nucleic Acids Res. 1984;12(1 Pt 1):121-7.
519. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155-D62.
520. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733-45.
521. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014;42(Database issue):D92-7.
522. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.
523. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one. 2011;6(7):e21800.
524. Chang CC, Veno MT, Chen L, Ditzel N, Le DQS, Dillschneider P, et al. Global MicroRNA Profiling in Human Bone Marrow Skeletal-Stromal or Mesenchymal-Stem Cells Identified Candidates for Bone Regeneration. Mol Ther. 2018;26(2):593-605.
525. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
526. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40(10):4288-97.
527. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
528. Barter MJ, Tselepi M, Gomez R, Woods S, Hui W, Smith GR, et al. Genome-Wide MicroRNA and Gene Analysis of Mesenchymal Stem Cell Chondrogenesis Identifies an Essential Role and Multiple Targets for miR-140-5p. Stem Cells. 2015;33(11):3266-80.
529. Wu M, Chen G, Li YP. TGF-beta and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. Bone Res. 2016;4:16009.
530. Crowgey EL, Wyffels JT, Osborn PM, Wood TT, Edsberg LE. A Systems Biology Approach for Studying Heterotopic Ossification: Proteomic Analysis of Clinical Serum and Tissue Samples. Genomics Proteomics Bioinformatics. 2018;16(3):212-20.
531. Huynh NPT, Zhang B, Guilak F. High-depth transcriptomic profiling reveals the temporal gene signature of human mesenchymal stem cells during chondrogenesis. FASEB J. 2019;33(1):358-72.
532. Wu Y, Jiang Y, Liu Q, Liu CZ. lncRNA H19 promotes matrix mineralization through up-regulating IGF1 by sponging miR-185-5p in osteoblasts. BMC Mol Cell Biol. 2019;20(1):48.

533. Huang MJ, Zhao JY, Xu JJ, Li J, Zhuang YF, Zhang XL. lncRNA ADAMTS9-AS2 Controls Human Mesenchymal Stem Cell Chondrogenic Differentiation and Functions as a ceRNA. Mol Ther Nucleic Acids. 2019;18:533-45.
534. Kurata H, Guillot PV, Chan J, Fisk NM. Osterix induces osteogenic gene expression but not differentiation in primary human fetal mesenchymal stem cells. Tissue Eng. 2007;13(7):1513-23.
535. Piek E, Sleumer LS, van Someren EP, Heuver L, de Haan JR, de Grijs I, et al. Osteo-transcriptomics of human mesenchymal stem cells: accelerated gene expression and osteoblast differentiation induced by vitamin D reveals c-MYC as an enhancer of BMP2-induced osteogenesis. Bone. 2010;46(3):613-27.
536. Gong Y, Qian Y, Yang F, Wang H, Yu Y. Lentiviral-mediated expression of SATB2 promotes osteogenic differentiation of bone marrow stromal cells in vitro and in vivo. Eur J Oral Sci. 2014;122(3):190-7.
537. Wang Q, Steigelman MB, Walker JA, Chen S, Hornsby PJ, Bohnenblust ME, et al. In vitro osteogenic differentiation of adipose stem cells after lentiviral transduction with green fluorescent protein. J Craniofac Surg. 2009;20(6):2193-9.
538. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9(1):e78644.
539. Bernar A, Gebetsberger JV, Bauer M, Streif W, Schirmer M. Optimization of the Alizarin Red S Assay by Enhancing Mineralization of Osteoblasts. Int J Mol Sci. 2022;24(1).
540. Kannan S, Ghosh J, Dhara SK. Osteogenic differentiation potential of porcine bone marrow mesenchymal stem cell subpopulations selected in different basal media. Biol Open. 2020;9(10).
541. Wang XF, Song Y, Liu YS, Sun YC, Wang YG, Wang Y, et al. Osteogenic Differentiation of Three-Dimensional Bioprinted Constructs Consisting of Human Adipose-Derived Stem Cells In Vitro and In Vivo. PLoS One. 2016;11(6):e0157214.
542. Hanna H, Mir LM, Andre FM. In vitro osteoblastic differentiation of mesenchymal stem cells generates cell layers with distinct properties. Stem Cell Res Ther. 2018;9(1):203.
543. Amarasekara DS, Kim S, Rho J. Regulation of Osteoblast Differentiation by Cytokine Networks. Int J Mol Sci. 2021;22(6).
544. Schroeder TM, Jensen ED, Westendorf JJ. Runx2: a master organizer of gene transcription in developing and maturing osteoblasts. Birth Defects Res C Embryo Today. 2005;75(3):213-25.
545. Sinha KM, Zhou X. Genetic and molecular control of osterix in skeletal formation. J Cell Biochem. 2013;114(5):975-84.
546. Liu Q, Li M, Wang S, Xiao Z, Xiong Y, Wang G. Recent Advances of Osterix Transcription Factor in Osteoblast Differentiation and Bone Formation. Front Cell Dev Biol. 2020;8:601224.
547. Vimalraj S. Alkaline phosphatase: Structure, expression and its function in bone mineralization. Gene. 2020;754:144855.
548. Lin X, Patil S, Gao YG, Qian A. The Bone Extracellular Matrix in Bone Formation and Regeneration. Front Pharmacol. 2020;11:757.
549. Rutkovskiy A, Stensløkken KO, Vaage IJ. Osteoblast Differentiation at a Glance. Med Sci Monit Basic Res. 2016;22:95-106.
550. Caron MMJ, Janssen MPF, Peeters L, Haudenschild DR, Cremers A, Surtel DAM, et al. Aggrecan and COMP Improve Periosteal Chondrogenesis by Delaying Chondrocyte Hypertrophic Maturation. Front Bioeng Biotechnol. 2020;8:1036.

551. Dateki S. ACAN mutations as a cause of familial short stature. Clin Pediatr Endocrinol. 2017;26(3):119-25.
552. Nenna R, Turchetti A, Mastrogiorgio G, Midulla F. COL2A1 Gene Mutations: Mechanisms of Spondyloepiphyseal Dysplasia Congenita. Appl Clin Genet. 2019;12:235-8.
553. Cui J, Zhang J. Cartilage Oligomeric Matrix Protein, Diseases, and Therapeutic Opportunities. Int J Mol Sci. 2022;23(16).
554. Lefebvre V, Dvir-Ginzberg M. SOX9 and the many facets of its regulation in the chondrocyte lineage. Connect Tissue Res. 2017;58(1):2-14.
555. Long JT, Leinroth A, Liao Y, Ren Y, Mirando AJ, Nguyen T, et al. Hypertrophic chondrocytes serve as a reservoir for marrow-associated skeletal stem and progenitor cells, osteoblasts, and adipocytes during skeletal development. Elife. 2022;11.
556. Eslaminejad MB, Fani N, Shahhoseini M. Epigenetic regulation of osteogenic and chondrogenic differentiation of mesenchymal stem cells in culture. Cell J. 2013;15(1):1-10.
557. Robert AW, Marcon BH, Dallagiovanna B, Shigunov P. Adipogenesis, Osteogenesis, and Chondrogenesis of Human Mesenchymal Stem/Stromal Cells: A Comparative Transcriptome Approach. Front Cell Dev Biol. 2020;8:561.
558. Pinnell SR. Regulation of collagen biosynthesis by ascorbic acid: a review. Yale J Biol Med. 1985;58(6):553-9.
559. Convente MR, Wang H, Pignolo RJ, Kaplan FS, Shore EM. The immunological contribution to heterotopic ossification disorders. Curr Osteoporos Rep. 2015;13(2):116-24.
560. Shore EM. Osteoinductive signals and heterotopic ossification. J Bone Miner Res. 2011;26(6):1163-5.
561. Leblanc E, Trensz F, Haroun S, Drouin G, Bergeron E, Penton CM, et al. BMP-9-induced muscle heterotopic ossification requires changes to the skeletal muscle microenvironment. J Bone Miner Res. 2011;26(6):1166-77.
562. Huang Y, Wang X, Lin H. The hypoxic microenvironment: a driving force for heterotopic ossification progression. Cell Commun Signal. 2020;18(1):20.
563. Fakhry M, Hamade E, Badran B, Buchet R, Magne D. Molecular mechanisms of mesenchymal stem cell differentiation towards osteoblasts. World J Stem Cells. 2013;5(4):136-48.
564. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
565. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
566. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35(12):2084-92.
567. Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB. Common features of microRNA target prediction tools. Front Genet. 2014;5:23.
568. Riolo G, Cantara S, Marzocchi C, Ricci C. miRNA Targets: From Prediction Tools to Experimental Validation. Methods Protoc. 2020;4(1).
569. Min H, Yoon S. Got target? Computational methods for microRNA target prediction and their extension. Exp Mol Med. 2010;42(4):233-44.
570. Shan KS, Li WW, Ren W, Kong S, Peng LP, Zhuo HQ, et al. LncRNA cancer susceptibility 20 regulates the metastasis of human gastric cancer cells via the miR-143-5p/MEMO1 molecular axis. World J Gastroenterol. 2022;28(16):1656-70.
571. Zhang Y, Pizzute T, Pei M. A review of crosstalk between MAPK and Wnt signals and its impact on cartilage regeneration. Cell Tissue Res. 2014;358(3):633-49.

572. Nickien M, Heuijerjans A, Ito K, van Donkelaar CC. Comparison between in vitro and in vivo cartilage overloading studies based on a systematic literature review. J Orthop Res. 2018;36(8):2076-86.
573. Hulsart-Billström G, Dawson JI, Hofmann S, Müller R, Stoddart MJ, Alini M, et al. A surprisingly poor correlation between in vitro and in vivo testing of biomaterials for bone regeneration: results of a multicentre analysis. Eur Cell Mater. 2016;31:312-22.
574. Ma B, Wang S, Wu W, Shan P, Chen Y, Meng J, et al. Mechanisms of circRNA/lncRNA-miRNA interactions and applications in disease and drug research. Biomed Pharmacother. 2023;162:114672.
575. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet. 2016;17(5):272-83.
576. Peng T, Qiao M, Liu H, Teotia S, Zhang Z, Zhao Y, et al. A Resource for Inactivation of MicroRNAs Using Short Tandem Target Mimic Technology in Model and Crop Plants. Mol Plant. 2018;11(11):1400-17.
577. Lefebvre V, Angelozzi M, Haseeb A. SOX9 in cartilage development and disease. Curr Opin Cell Biol. 2019;61:39-47.
578. Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, Stevanović M, et al. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature. 1994;372(6506):525-30.
579. Shi S, Wang C, Acton AJ, Eckert GJ, Trippel SB. Role of sox9 in growth factor regulation of articular chondrocytes. J Cell Biochem. 2015;116(7):1391-400.
580. Lefebvre V, Huang W, Harley VR, Goodfellow PN, de Crombrugghe B. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol. 1997;17(4):2336-46.
581. Csukasi F, Duran I, Zhang W, Martin JH, Barad M, Bamshad M, et al. Dominant-negative SOX9 mutations in campomelic dysplasia. Hum Mutat. 2019;40(12):2344-52.
582. Sodhi H, Panitch A. Glycosaminoglycans in Tissue Engineering: A Review. Biomolecules. 2020;11(1).
583. Kittlick PD, Hadházy C, Oláh EH. Studies on cartilage formation. XVIII. Changes of the composition of glycosaminoglycans in the regenerating articular surface. Acta Biol Acad Sci Hung. 1975;26(3-4):165-74.
584. Kikukawa K, Suzuki K. Histochemical and immunohistochemical distribution of glycosaminoglycans, type II collagen, and fibronectin in developing fetal cartilage of congenital osteochondrodysplasia rat (ocd/ocd). Teratology. 1992;46(5):509-23.
585. Sophia Fox AJ, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. Sports Health. 2009;1(6):461-8.
586. Bełdowski P, Mazurkiewicz A, Topoliński T, Małek T. Hydrogen and Water Bonding between Glycosaminoglycans and Phospholipids in the Synovial Fluid: Molecular Dynamics Study. Materials (Basel). 2019;12(13).
587. Chen J, Sun T, You Y, Wu B, Wang X, Wu J. Proteoglycans and Glycosaminoglycans in Stem Cell Homeostasis and Bone Tissue Regeneration. Front Cell Dev Biol. 2021;9:760532.
588. Templeton DM. The basis and applicability of the dimethylmethylene blue binding assay for sulfated glycosaminoglycans. Connect Tissue Res. 1988;17(1):23-32.
589. Zhou P, Shi JM, Song JE, Han Y, Li HJ, Song YM, et al. Establishing a deeper understanding of the osteogenic differentiation of monolayer cultured human pluripotent stem cells using novel and detailed analyses. Stem Cell Res Ther. 2021;12(1):41.

590. Wang W, Rigueur D, Lyons KM. TGF β signaling in cartilage development and maintenance. Birth Defects Res C Embryo Today. 2014;102(1):37-51.
591. Baugé C, Girard N, Lhuissier E, Bazille C, Boumediene K. Regulation and Role of TGF β Signaling Pathway in Aging and Osteoarthritis Joints. Aging Dis. 2014;5(6):394-405.
592. Akkiraju H, Nohe A. Role of Chondrocytes in Cartilage Formation, Progression of Osteoarthritis and Cartilage Regeneration. J Dev Biol. 2015;3(4):177-92.
593. Haleem-Smith H, Calderon R, Song Y, Tuan RS, Chen FH. Cartilage oligomeric matrix protein enhances matrix assembly during chondrogenesis of human mesenchymal stem cells. J Cell Biochem. 2012;113(4):1245-52.
594. Jing Y, Jing J, Ye L, Liu X, Harris SE, Hinton RJ, et al. Chondrogenesis and osteogenesis are one continuous developmental and lineage defined biological process. Sci Rep. 2017;7(1):10020.
595. Yu DA, Han J, Kim BS. Stimulation of chondrogenic differentiation of mesenchymal stem cells. Int J Stem Cells. 2012;5(1):16-22.
596. Wang D, Canaff L, Davidson D, Corluka A, Liu H, Hendy GN, et al. Alterations in the sensing and transport of phosphate and calcium by differentiating chondrocytes. J Biol Chem. 2001;276(36):33995-4005.
597. Mueller MB, Tuan RS. Functional characterization of hypertrophy in chondrogenesis of human mesenchymal stem cells. Arthritis Rheum. 2008;58(5):1377-88.
598. Takács R, Vágó J, Póliska S, Pushparaj PN, Ducza L, Kovács P, et al. The temporal transcriptomic signature of cartilage formation. Nucleic Acids Res. 2023;51(8):3590-617.
599. Ding J, Wu W, Yang J, Wu M. Long non-coding RNA MIF-AS1 promotes breast cancer cell proliferation, migration and EMT process through regulating miR-1249-3p/HOXB8 axis. Pathol Res Pract. 2019;215(7):152376.
600. Shu H, Hu J, Deng H. miR-1249-3p accelerates the malignancy phenotype of hepatocellular carcinoma by directly targeting HNRNPK. Mol Genet Genomic Med. 2019;7(10):e00867.
601. Ai N, Yu Z, Xu X, Liufu S, Wang K, Huang S, et al. Circular Intronic RNA circTTN Inhibits Host Gene Transcription and Myogenesis by Recruiting PURB Proteins to form Heterotypic Complexes. Int J Mol Sci. 2023;24(12).
602. Tamkeen N, AlOmar SY, Alqahtani SAM, Al-Jurayyan A, Farooqui A, Tazyeen S, et al. Identification of the Key Regulators of Spina Bifida Through Graph-Theoretical Approach. Front Genet. 2021;12:597983.
603. Wang Y, Liu W, Liu Y, Cui J, Zhao Z, Cao H, et al. Long noncoding RNA H19 mediates LCoR to impact the osteogenic and adipogenic differentiation of mBMSCs in mice through sponging miR-188. J Cell Physiol. 2018;233(9):7435-46.
604. Goveas N, Waskow C, Arndt K, Heuberger J, Zhang Q, Alexopoulou D, et al. MLL1 is required for maintenance of intestinal stem cells. PLoS Genet. 2021;17(12):e1009250.
605. Shih YR, Tseng KF, Lai HY, Lin CH, Lee OK. Matrix stiffness regulation of integrin-mediated mechanotransduction during osteogenic differentiation of human mesenchymal stem cells. J Bone Miner Res. 2011;26(4):730-8.
606. Novakov V, Novakova O, Churnosova M, Sorokina I, Aristova I, Polonikov A, et al. Intergenic Interactions of SBNO1, NFAT5 and GLT8D1 Determine the Susceptibility to Knee Osteoarthritis among Europeans of Russia. Life (Basel). 2023;13(2).
607. Sanglard LP, Nascimento M, Moriel P, Sommer J, Ashwell M, Poore MH, et al. Impact of energy restriction during late gestation on the muscle and blood transcriptome of beef calves after preconditioning. BMC Genomics. 2018;19(1):702.

608. Li D, Wang TW, Aratani S, Omori S, Tamatani M, Johmura Y, et al. Transcriptomic characterization of Lonrf1 at the single-cell level under pathophysiological conditions. J Biochem. 2023;173(6):459-69.
609. Kuçi S, Kuçi Z, Schäfer R, Spohn G, Winter S, Schwab M, et al. Molecular signature of human bone marrow-derived mesenchymal stromal cell subsets. Sci Rep. 2019;9(1):1774.
610. Oberbauer AM, Peng R. Fractionation of growth plate chondrocytes: differential expression of IGF-I and growth hormone and IGF-I receptor mRNA in purified populations. Connect Tissue Res. 1995;31(3):179-87.
611. Greenbaum J, Wu K, Zhang L, Shen H, Zhang J, Deng HW. Increased detection of genetic loci associated with risk predictors of osteoporotic fracture using a pleiotropic cFDR method. Bone. 2017;99:62-8.
612. Wang S, Ren Q, Li G, Zhao X, Zhao X, Zhang Z. The Targeted Therapies for Osteosarcoma via Six Major Pathways. Curr Mol Pharmacol. 2023.
613. Fiori JL, Billings PC, de la Peña LS, Kaplan FS, Shore EM. Dysregulation of the BMP-p38 MAPK signaling pathway in cells from patients with fibrodysplasia ossificans progressiva (FOP). J Bone Miner Res. 2006;21(6):902-9.
614. Huang Y, Wang X, Zhou D, Zhou W, Dai F, Lin H. Macrophages in heterotopic ossification: from mechanisms to therapy. NPJ Regen Med. 2021;6(1):70.
615. Sorkin A, von Zastrow M. Endocytosis and signalling: intertwining molecular networks. Nat Rev Mol Cell Biol. 2009;10(9):609-22.
616. Ehrlich M. Endocytosis and trafficking of BMP receptors: Regulatory mechanisms for fine-tuning the signaling response in different cellular contexts. Cytokine Growth Factor Rev. 2016;27:35-42.
617. Tseng WP, Yang SN, Lai CH, Tang CH. Hypoxia induces BMP-2 expression via ILK, Akt, mTOR, and HIF-1 pathways in osteoblasts. J Cell Physiol. 2010;223(3):810-8.
618. Dizon ML, Maa T, Kessler JA. The bone morphogenetic protein antagonist noggin protects white matter after perinatal hypoxia-ischemia. Neurobiol Dis. 2011;42(3):318-26.
619. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell. 2013;153(3):654-65.
620. Prelich G. Gene overexpression: uses, mechanisms, and interpretation. Genetics. 2012;190(3):841-54.
621. Uhl M, Houwaart T, Corrado G, Wright PR, Backofen R. Computational analysis of CLIP-seq data. Methods. 2017;118-119:60-72.
622. Kan L, Kessler JA. Animal models of typical heterotopic ossification. J Biomed Biotechnol. 2011;2011:309287.
623. Mascetti VL, Pedersen RA. Human-Mouse Chimerism Validates Human Stem Cell Pluripotency. Cell Stem Cell. 2016;18(1):67-72.
624. Iimori Y, Morioka M, Koyamatsu S, Tsumaki N. Implantation of Human-Induced Pluripotent Stem Cell-Derived Cartilage in Bone Defects of Mice. Tissue Eng Part A. 2021;27(21-22):1355-67.
625. Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, et al. Considerations when investigating lncRNA function in vivo. Elife. 2014;3:e03058.
626. Morita E, Arii J, Christensen D, Votteler J, Sundquist WI. Attenuated protein expression vectors for use in siRNA rescue experiments. Biotechniques. 2012;0(0):1-5.
627. Zhao R, Fu J, Zhu L, Chen Y, Liu B. Designing strategies of small-molecule compounds for modulating non-coding RNAs in cancer therapy. J Hematol Oncol. 2022;15(1):14.

628. Rohban MH, Fuller AM, Tan C, Goldstein JT, Syangtan D, Gutnick A, et al. Virtual screening for small-molecule pathway regulators by image-profile matching. Cell Syst. 2022;13(9):724-36.e9.
629. Forsberg JA, Pepek JM, Wagner S, Wilson K, Flint J, Andersen RC, et al. Heterotopic ossification in high-energy wartime extremity injuries: prevalence and risk factors. J Bone Joint Surg Am. 2009;91(5):1084-91.
630. DeLee J, Ferrari A, Charnley J. Ectopic bone formation following low friction arthroplasty of the hip. Clin Orthop Relat Res. 1976(121):53-9.
631. Neal B, Gray H, MacMahon S, Dunn L. Incidence of heterotopic bone formation after major hip surgery. ANZJ Surg. 2002;72(11):808-21.
632. Garland DE. Clinical observations on fractures and heterotopic ossification in the spinal cord and traumatic brain injured populations. Clin Orthop Relat Res. 1988(233):86-101.
633. Richards AM, Klaassen MF. Heterotopic ossification after severe burns: a report of three cases and review of the literature. Burns. 1997;23(1):64-8.
634. Zhu Y, Zhang F, Chen W, Zhang Q, Liu S, Zhang Y. Incidence and risk factors for heterotopic ossification after total hip arthroplasty: a meta-analysis. Arch Orthop Trauma Surg. 2015;135(9):1307-14.
635. Davis G, Patel RP, Tan TL, Alijanipour P, Naik TU, Parvizi J. Ethnic differences in heterotopic ossification following total hip arthroplasty. Bone Joint J. 2016;98-B(6):761-6.
636. Shore EM, Ahn J, Jan de Beur S, Li M, Xu M, Gardner RJ, et al. Paternally inherited inactivating mutations of the GNAS1 gene in progressive osseous heteroplasia. N Engl J Med. 2002;346(2):99-106.
637. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
638. Maeda M, Hasegawa H, Hyodo T, Ito S, Asano E, Yuang H, et al. ARHGAP18, a GTPase-activating protein for RhoA, controls cell shape, spreading, and motility. Mol Biol Cell. 2011;22(20):3840-52.
639. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016;32(10):1479-85.
640. Toom A, Arend A, Gunnarsson D, Ulfsparre R, Suutre S, Haviko T, et al. Bone formation zones in heterotopic ossifications: histologic findings and increased expression of bone morphogenetic protein 2 and transforming growth factors beta2 and beta3. Calcif Tissue Int. 2007;80(4):259-67.
641. Suutre S, Toom A, Arend A, Selstam G. Bone tissue content of TGF-beta2 changes with time in human heterotopic ossification after total hip arthroplasty. Growth Factors. 2009;27(2):114-20.
642. MacInnes SJ, Hatzikotoulas K, Fenstad AM, Shah K, Southam L, Tachmazidou I, et al. The 2018 Otto Aufranc Award: How Does Genome-wide Variation Affect Osteolysis Risk After THA? Clin Orthop Relat Res. 2019;477(2):297-309.
643. arcOGEN Consortium; arcOGEN Collaborators ZE, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, Boraska V, Esko T, Evangelou E, Hoffman A, Houwing-Duistermaat JJ, Ingvarsson T, Jonsdottir I, Jonnson H, Kerkhof HJ, Kloppenburg M, Bos SD, Mangino M, Metrustry S, Slagboom PE, Thorleifsson G, Raine EV, Ratnayake M, Ricketts M, Beazley C, Blackburn H, Bumpstead S, Elliott KS, Hunt SE, Potter SC, Shin SY, Yadav VK, Zhai G, Sherburn K, Dixon K, Arden E, Aslam N, Battley PK, Carluke I, Doherty S, Gordon A, Joseph J, Keen R, Koller NC, Mitchell S, O'Neill F, Paling E, Reed MR, Rivadeneira F, Swift D,

Walker K, Watkins B, Wheeler M, Birrell F, Ioannidis JP, Meulenbelt I, Metspalu A, Rai A, Salter D, Stefansson K, Stykarsdottir U, Uitterlinden AG, van Meurs JB, Chapman K, Deloukas P, Ollier WE, Wallis GA, Arden N, Carr A, Doherty M, McCaskie A, Willkinson JM, Ralston SH, Valdes AM, Spector TD, Loughlin J. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet. 2012;380(9844):815-23.
644. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007;39(7):906-13.
645. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
646. Pavlou G, Salhab M, Murugesan L, Jallad S, Petsatodis G, West R, et al. Risk factors for heterotopic ossification in primary total hip arthroplasty. Hip Int. 2012;22(1):50-5.
647. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
648. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis [Internet]. 2016. Available from: http://biorxiv.org/content/early/2016/11/30/038067.