The University Of Sheffield.

MRC DiMeN Doctoral Training Partnership

AN INVESTIGATION INTO THE MECHANISM OF ACTION OF THE LNCRNA *CASC20* ON OSTEOGENESIS AND CHONDROGENESIS

Phoebe Tamblin-Hopper

University of Sheffield

MPhil

July 2025

Contents

Abstract

CASC20 encodes a long noncoding RNA (lncRNA) and is hypothesised to be a novel regulator of bone formation through promotion of osteogenesis and a reduction in chondrogenesis. Understanding the role *CASC20* plays could help uncover novel therapeutic approaches for treatment and/or prevention of diseases such as heterotopic ossification (HO), where GWAS has shown *CASC20* is associated with susceptibility. This thesis describes the work done to investigate the role of *CASC20* in musculoskeletal development and its mechanism of action. lncRNAs can act via decoy, scaffold or guide mechanisms and evidence was gathered to determine which of these is responsible for the actions of *CASC20*.

An RNA-sequencing analysis of murine cells expressing *CAS20* during osteogenic differentiation was performed to observe changes in gene expression and identify *CASC20* micro (mi)RNA targets. This work was then repeated in primary, adipose-derived human mesenchymal stromal cells (hMSCs), isolated from the stromal vascular fraction (SVF) in knee fat pads obtained from patients undergoing joint replacement surgery. The hMSCs were transduced with a *CASC20* overexpressing (OE) or control (CTRL) lentivirus and differentiated to observe the effects of *CASC20* on osteogenesis and chondrogenesis. However, whilst these cells were able to differentiate, the process of lentiviral transduction with *CASC20* caused them to lose their differentiative capacity. Pilots with adenoviral transduction demonstrated this was a viable alternative to be used going forward for *CASC20* overexpression in primary hMSCs.

Future work is ongoing to investigate whether the action of *CASC20* is through *BMP2*, a neighbouring gene critical to osteogenesis, by clustered regularly interspaced palindromic repeats (CRISPR) deletion of *CASC20* regions. This work alongside that described in this thesis helps provide further insight into the action of *CASC20* and its role as a potential novel regulator of bone formation.

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Bone formation and regulation

Bone tissue is made up of various cell types: osteoblasts, bone lining cells, osteocytes and osteoclasts (Florencio-Silva et al., 2015) (figure 1.1). Osteoblasts are cuboidal cells responsible for bone formation and make up 4-6% of bone resident cells (Capulli et al., 2014). Bone lining cells are thin, flat, inactive osteoblasts that play roles in linking bone resorption and formation (Miller et al., 1989). Osteocytes are dendritic cells, making up the majority of total bone cells (90-95%) (Capulli et al., 2014). They act as overall mechanosensors of the bone remodelling process (Florencio-Silva et al., 2015). Finally, osteoclasts are multinucleated, terminally differentiated cells, responsible for bone resorption (Florencio-Silva et al., 2015). Bone cells and matrix are formed through the differentiation of various stem cells, notably mesenchymal stromal cells (MSCs), which are multi lineage cells capable of becoming osteoblasts, chondrocytes or adipocytes (Pittenger et al., 1999), and hematopoietic stem cells (HSCs) that differentiate into osteoclasts (figure 1.1). The bone matrix, the extracellular mineralised component of bone, is mainly made up of collagen and hydroxyapatite, and supports bone cells whilst regulating their activity (Florencio-Silva et al., 2015, Ma et al., 2022a).

Transforming growth factor-beta (TGFβ) and bone morphogenic protein 2 (BMP2) play critical roles in regulating MSC differentiation during skeletal development, bone formation and homeostasis (Wu et al., 2016) (figure 1.2). TGFβ and BMP2 both transduce signals to the canonical Smad-dependent and non-canonical p38 Mitogen-activated protein kinase (MAPK) signalling pathways, to promote expression of various transcription factors and crosstalk with cytokine signalling pathways (Wu et al., 2016).

Bone formation and regulation is controlled by local factors, such as growth factors and cytokines, as well as systemic factors, including calcitonin and oestrogens (Florencio-Silva et al., 2015). Bone is highly dynamic, and bone resorption (the breakdown of existing bone tissue) by osteoclasts, is balanced with bone formation by osteoblasts, in healthy individuals (Florencio-Silva et al., 2015). Bone remodelling, a process critical to fracture healing and skeleton adaptation, involves osteoclasts initiating bone resorption, a transition from

resorption to formation of new bone, then osteoblasts forming new bone (Sims and Gooi, 2008). Imbalances between resorption and formation result in disease, such as osteoporosis when resorption outweighs formation and osteopetrosis caused by the opposite (Florencio-Silva et al., 2015).

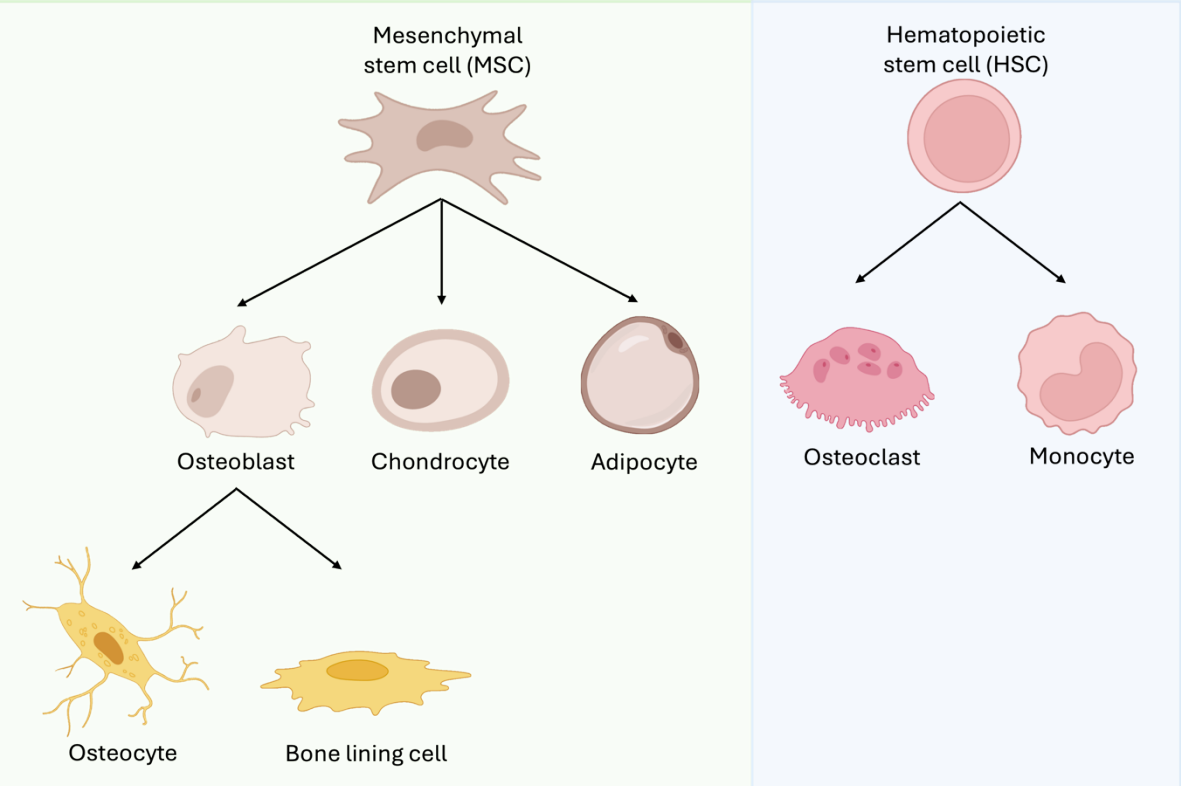


Figure 1.1: Lineage of bone cells. *Mesenchymal stem cells (MSCs) and haematopoietic stem cells (HSCs) are both important progenitor cells in the formation of various cells important to bone formation and function. MSCs differentiate into osteoblasts, chondrocytes and adipocytes. Osteoblasts then can further differentiate to osteocytes, or form bone lining cells. HSCs differentiate into osteoclasts or monocytes, which then become macrophages or dendritic cells.*

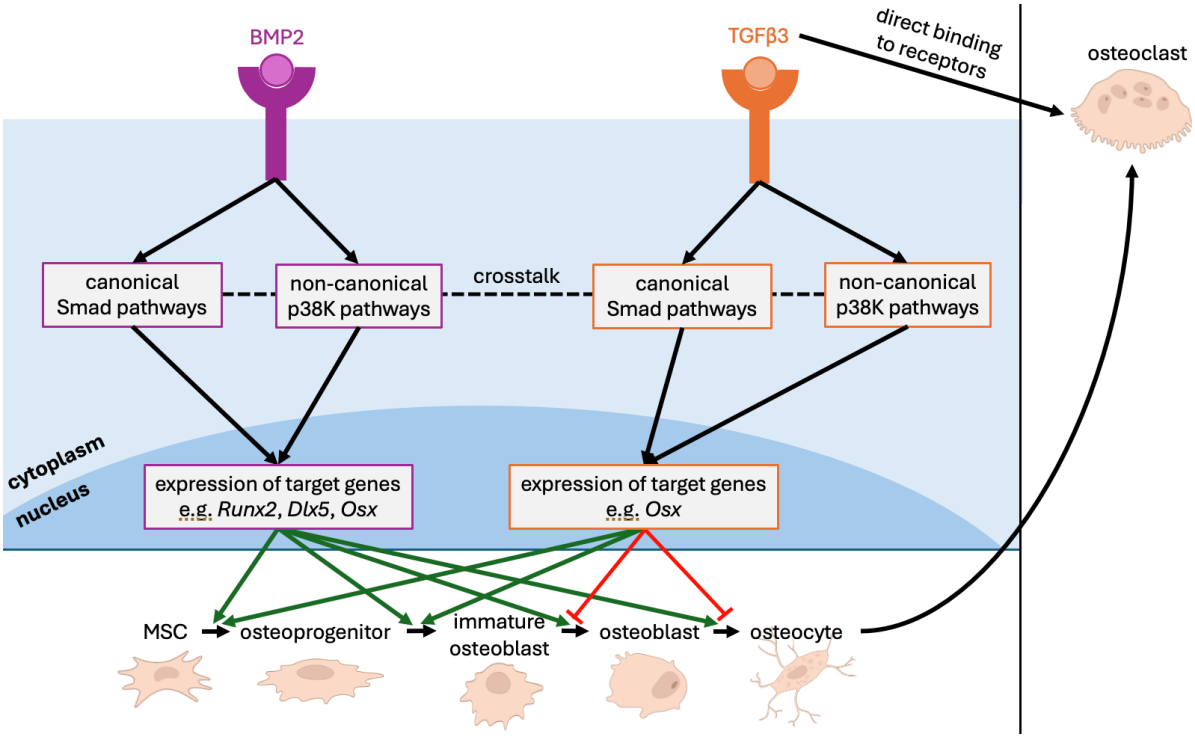


Figure 1.2: Key pathways of bone formation and turnover. *BMP2 and TGFβ3 bind to receptors in the cell membrane, activating canonical Smad pathways and non-canonical p38 kinase pathways. All pathways lead to the expression of respective target genes in the nucleus, which act on different parts of the differentiation process of an MSC to an osteocyte. The green arrows indicate promotion of specific parts of this process and red lines show inhibition of the pathways. TGFβ3 can also act directly on osteoclasts through binding to their receptors. Osteocytes can also influence osteoclast activity through signalling molecules which may promote their differentiation.*

Chondrogenesis is another important process in musculoskeletal development and involves the formation of cartilage (Jing et al., 2017). This involves expression of cartilage-specific collagens during the differentiation of MSCs, to form chondrocytes, which form the cartilage at the end of bones in limb development (Jing et al., 2017). Furthermore, osteogenesis and chondrogenesis are both critical processes interlinked in endochondral bone formation in fetal development (Jing et al., 2017). Chondrocytes contribute to bone modelling, through transformation into bone cells, to determine the skeletons shape and size (Jing et al., 2017)., or proliferate until they become hypertrophic and undergo terminal differentiation, then this hypertrophic cartilage is resorbed and replaced by bone (Jing et al., 2017).

1.2 The role of *CASC20*

One disease resulting from problems during bone formation and regulation is heterotopic ossification (HO), defined as the presence of bone in soft tissues where it should not be present and caused by pathological bone formation (Meyers et al., 2019). HO is commonly observed as a result of trauma or surgery (trauma-induced HO), or more rarely can have genetic causes (Meyers et al., 2019). Genetic HO occurs in fibrodysplasia ossificans progressiva (FOP), where heterotopic bone forms sporadically or in response to external trauma (Dey et al., 2017). In both trauma-induced and genetic HO the bone forms at sites where connective tissue cells and entheses are prevalent (Dey et al., 2017). The wider clinical features and biological features of HO are reviewed in depth in (Meyers et al., 2019). A genome-wide association study (GWAS) of HO patients after total hip arthroplasty (THA) found that cancer susceptibility 20 (*CASC20*), which encodes a long noncoding RNA (lncRNA), is strongly associated with HO susceptibility (Konstantinos Hatzikotoulas et al., 2019). The *CASC20* gene located on chromosome 20p12.3, is 286,079 bases in length (GFCh38/hg38) and has its highest level of expression in prostate and testis tissue, although its role here is unknown (Health, 2023). At least four alternate transcripts of *CASC20* have been identified, ranging from 1484 to 2242 nucleotides in length, and with varying states of polyadenylation and splicing. The most frequently identified transcript, URS000075ACC6_9606, was used experimentally in this project.

CASC20 is also expressed in human bone and is upregulated in BMP2-induced osteogenic differentiation (Felix-Ilemhenbhio, 2023). Whilst the neighbouring gene of *CASC20* is *BMP2*, there are no HO associating variants within *BMP2* or any *BMP2* variant in significant linkage disequilibrium with the *CASC20* signal, suggesting *CASC20* is the origin of the signal (Felix-Ilemhenbhio, 2023).

Furthermore, *CASC20* has been shown to play an important role in the early stages of osteogenesis and chondrogenesis (Felix-Ilemhenbhio, 2023). In osteogenic differentiation *CASC20* is upregulated, resulting in increased mineralisation and promotion of osteogenic gene expression (Felix-Ilemhenbhio, 2023). This was tested in human multipotent adipose-derived stem cells (hMADs), murine MSCs and ASC52telo human Telomerase Reverse

Transcriptase (hTERT) immortalised adipose-derived human MSCs. Conversely, in chondrogenic differentiation *CASC20* is a negative regulator and slows chondrogenesis (Felix-Ilemhenbhio, 2023). This was evidenced by the inhibition of glycosaminoglycan (GAG) deposition in differentiated ASC52telo cells overexpressing *SOX9* and murine MSCs (Felix-Ilemhenbhio, 2023).

lncRNAs, such as *CASC20*, can act through a variety of mechanisms, including as a scaffold, a guide or as a decoy to modulate micro RNAs (miRNAs). Decoy lncRNAs are also known as competing endogenous RNAs (ceRNAs) or 'sponges', which function by competing with other RNAs for a limited pool of miRNAs, thus co-regulating each other's binding; lncRNAs in particular have been observed to function in this manner (Salmena et al., 2011, Tay et al., 2014). These mechanisms are discussed in further detail in chapter 2. The *CASC20* transcript has a 3' poly-A tail, so has the ability to undergo cytoplasmic export, thus meaning it could act as a ceRNA to modulate osteo- and chondrogenesis (Felix-Ilemhenbhio, 2023). In gastric cancer cells, *CASC20* acts as a ceRNA to bind miR-143-5p and regulate the expression of *Mediator Of Cell Motility 1* (*MEMO1*). Here *CASC20* promotes tumorigenesis by regulating metastasis through the miR-143-5p/MEMO1 axis, with its high expression in gastric cancer tumours and cells correlating with lymphatic metastasis and a poor prognosis. This indicates the potential to target *CASC20* in gastric cancer, as its silencing reduced proliferation, migration and invasion of gastric cancer cells. It is possible that further research could elucidate a similarly targetable role for *CASC20* in stromal cells. However, lncRNAs can act through multiple mechanisms and further evidence is required to elucidate *CASC20*s exact mechanism of action. RNA-seq will help elucidate any interacting miRNAs, whilst knockdown and overexpression studies will provide further information into how it acts. Although it falls beyond the experimental scope of this thesis there are further methodologies that could provide evidence for the mechanism of action of *CASC20*, which will be discussed in the future plans. Thirteen genes were found that are commonly targeted by *CASC20*-interacting miRNAs, which included established osteogenic and chondrogenic markers, such as MAPK1 (Felix-Ilemhenbhio, 2023). This evidence suggests *CASC20* could play a key role in the process of bone formation and thus could be a potential target for treatment of HO and other related diseases.

1.3 miRNAs

As one putative mechanism of action of *CASC20* is to act on miRNAs, it is important to understand how the interactions of miRNAs with their gene targets and non-coding (nc)RNAs could occur. miRNAs are a class of small non-coding RNAs, of 18-24 nucleotides in length, which act on genes at post-transcriptional and translational levels to perform various regulatory roles in physiology and disease pathologies (Baek et al., 2008, Miska, 2005). Precursor miRNAs have two arms: miR-5p and miR-3p, which produce two different mature miRNAs when processed (Kozomara and Griffiths-Jones, 2014). Whilst it was originally believed that only one arm was incorporated into the RNA induced silencing complex (RISC) to become functional and the other was degraded, it has since been found that in certain cell types both arms can be functional (Mitra et al., 2015). miRNAs typically act by associating with the Argonaute (AGO) effector protein, which is directed to target mRNAs with complimentary sequences to the miRNA. Here, miRNAs bind to the 3' untranslated regions (UTRs) of mRNAs, and the AGO-miRNA complex leads to translational repression and/or destabilisation of the mRNA, causing downregulation; this mechanism is known as miRNA-mediated target repression (Lee et al., 1993, Wightman et al., 1993, Wilczynska and Bushell, 2015). Without AGO, the miRNA is rapidly degraded (Han et al., 2020). For interaction with their target mRNAs, miRNAs have ~8 nucleotide seed sequences at their 5' ends, which are important for binding (Grimson et al., 2007). These nucleotides can have between 6 and 8 bps of similarity with their target, determining how the miRNA binding is categorised (figure 1.3). miRNA binding sites are often predicted by these seed sequences, but seed sequences are not solely responsible for effective targeting; the sequence context around the seed is also of importance (Grimson et al., 2007). Also, whilst canonical seed pairing is one of the most important factors in miRNA target recognition, it is not present in the majority of miRNA target sites (Wang, 2014). This however, is just in locations where miRNAs have been shown to bind, and not necessarily where they may be having a regulatory role.

There is a high level of conservation between human and murine miRNAs, both in the sequence of the miRNAs and of their binding sites (Altuvia et al., 2005, Friedman et al., 2009). This has been seen generally and more specifically in small RNA-sequencing of cell

lines overexpressing *CASC20*, where >90% of miRNA reads were found to map to multiple species genomes. Furthermore, the names of miRNAs are consistent between species, with highly similar miRNAs (i.e. one differing base) that are found in different species given the same name (Ambros et al., 2003). However, conservation is dependent on the interaction between a miRNA and mRNA, so even if miRNAs have the same names between species, they may regulate different genes and thus not be truly conserved. Within species, identical or similar miRNA sequences can be given the same number with letters or numbers added as suffixes to distinguish them, according to the relevant organism convention.

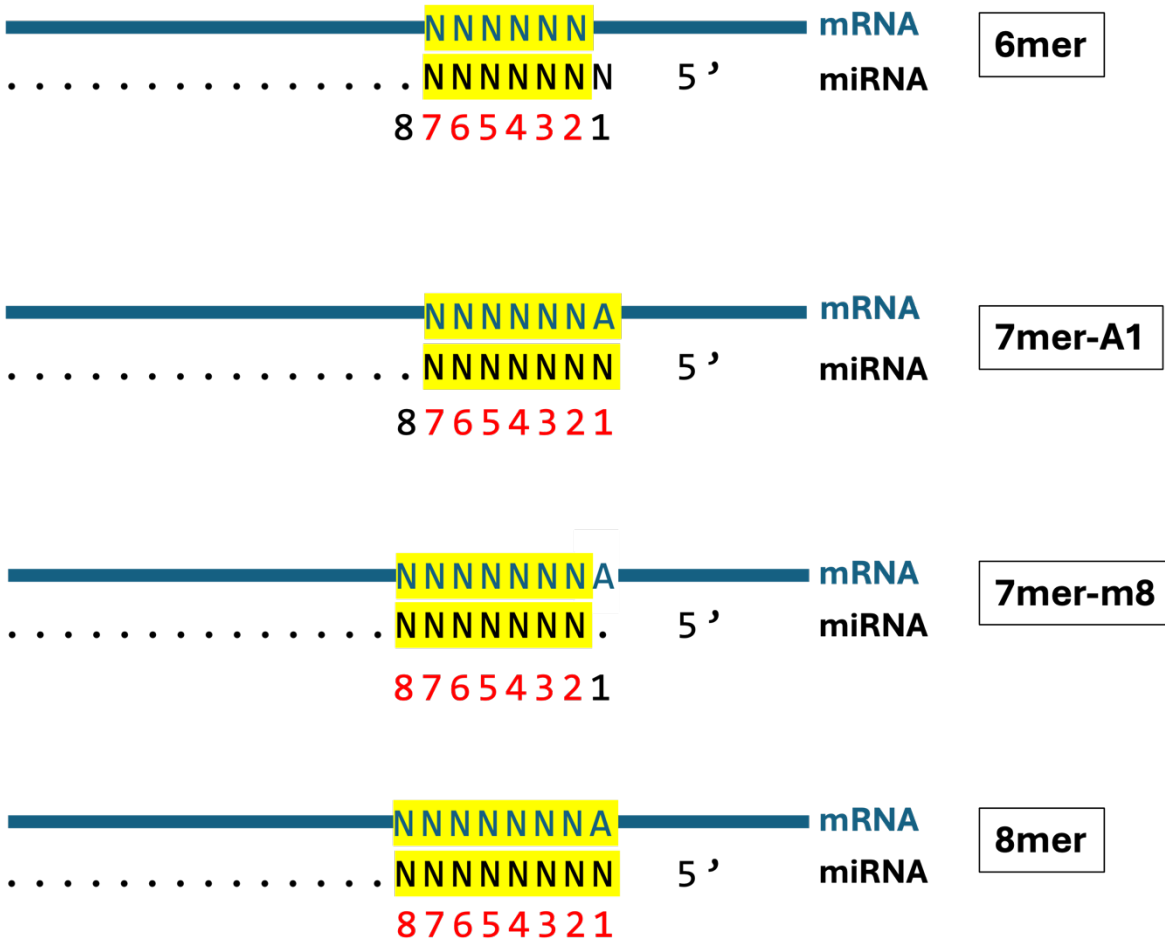


Figure 1.3: Different seed sequences of miRNAs. *The number of complimentary bases that a miRNA has with its target mRNA and the context of this sequence determines the level of binding. The number of bases that are complimentary defines whether the seed sequence is known as 6mer, 7mer or 8mer, with a greater number signifying greater binding. N represents G/C/A/T, with yellow highlights showing the bases that are binding. The numbering shows which base of the miRNA is binding to the mRNA, with those positions where binding is taking place in red. The mRNA is shown with the polyA tail to the right-hand side and the open reading frame to the left.*

1.3 Project hypothesis and aims

Hypothesis: *CASC20* is a novel regulator of bone formation acting independently of *BMP2*.

Aims:

1) **To discuss the potential of targeting lncRNAs such as *CASC20* therapeutically through a review publication.**

Other ncRNAs, such as miRNAs, have been successfully targeted therapeutically and are currently undergoing clinical trials. However, there are currently no treatments on the market targeting lncRNAs. The likelihood of this becoming possible in the near future was assessed through evaluation of the current literature on lncRNA mechanisms in disease.

2) **To identify potential miRNA targets of *CASC20*.**

The work described in this thesis aimed to identify potential miRNAs of interest through analysis of RNA-seq data for murine MSCs, where cells were transduced with a lentivirus containing either *CASC20* or an empty control vector. Analysis of mRNA- and miRNA-sequencing data was carried out to observe changes in gene expression between the control and *CASC20* expression. This was used to identify any genes and miRNAs differentially expressed (DE) with *CASC20* expression and provide evidence of potential miRNA targets of *CASC20*. DE miRNAs were compared to a list of miRNAs predicted to have a binding sequence within *CASC20* to determine any miRNAs of interest.

This work can then be validated in human MSCs, isolated from the stromal vascular fraction (SVF) of patient samples. Through analysis of genes and miRNAs DE with *CASC20* overexpression, the miRNAs that are enriched within the differentially expressed genes (DEGs) can be uncovered. Identification of miRNAs of interest and looking at how many binding sites exist for them within *CASC20* provides *in silico* evidence for a potential mechanism of action of *CASC20*.

3) **To develop and optimise a methodology to validate findings that *CASC20* upregulates osteogenesis and downregulates chondrogenesis in human mesenchymal stromal cells (hMSCs).**

This required overexpression of *CASC20* in hMSCs with a *CASC20* expressing vector. Alizarin Red Staining (ARS) for calcification and a 1,9-Dimethyl-Methylene Blue Dye (DMMB) assay for GAG were used as endpoints for osteogenesis and chondrogenesis, respectively. The original observations of these effects of *CASC20* were made in human multipotent adipose-derived stem cells (hMADs) and ASC52telo human Telomerase Reverse Transcriptase (hTERT) immortalised adipose-derived hMSCs, as well as murine MSCs, but required validation in primary human cells.

Chapter 2: The Potential Therapeutic Applications of long non-coding RNAs

Whilst lncRNAs, such as *CASC20* have not yet been clinically targeted, it is a rapidly evolving field and is becoming an increasingly possible likelihood. The therapeutic potential of lncRNAs is reviewed in a paper entitled *'The potential therapeutic applications of long non-coding RNAs'*, originally written to fulfil the requirements of my literature review and further adapted for publication in the Journal of Translational Genetics and Genomics.

The potential therapeutic applications of long noncoding RNAs

Phoebe Tamblin-Hopper[1], Endre Kiss-Toth[1], Ian Sudbery[2], David Young[3], J. Mark Wilkinson[1]

[1]Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield S10 2RX, UK.
[2]School of Biosciences, University of Sheffield, Sheffield S10 2TN, UK.
[3]Biosciences Institute, Newcastle University, Newcastle-Upon-Tyne NE1 3BZ, UK.

Correspondence to: Prof. J. Mark Wilkinson, Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, The Medical School, Beech Hill Road, Sheffield S10 2RX, UK. E-mail: j.m.wilkinson@sheffield.ac.uk

How to cite this article: Tamblin-Hopper P, Kiss-Toth E, Sudbery I, Young D, Wilkinson JM. The potential therapeutic applications of long non-coding RNAs. J Transl Genet Genom 2024;8:225-43. https://dx.doi.org/10.20517/jtgg.2024.12

Received: 14 Mar 2024 **First Decision:** 25 Apr 2024 **Revised:** 12 May 2024 **Accepted:** 7 Jun 2024 Published: 27 Jun 2024

Academic Editor: Andrea L. Gropman **Copy Editor:** Fangling Lan **Production Editor:** Fangling Lan

2.1 Abstract

The field of RNA-based therapeutics is rapidly evolving and targeting non-coding RNAs (ncRNAs) associated with disease is becoming increasingly feasible. MicroRNAs (miRNAs) are a class of small ncRNAs (sncRNAs) and the first anti-miRNA drugs, e.g., Miravirsen and Cobomarsen, have successfully completed phase II clinical trials. Long ncRNAs (lncRNAs) are another class of ncRNAs that are commonly dysregulated in disease. Thus, they hold potential as putative therapeutic targets or agents. LncRNAs can function through a variety of mechanisms, including as guide, scaffold or decoy molecules, and understanding of these actions is critical to devising effective targeting strategies. LncRNA expression can be modulated with small interfering RNAs (siRNAs), antisense oligonucleotides (ASOs), CRISPR-Cas9, or small molecule inhibitors. These approaches have been employed to target a number of lncRNAs and tested in animal models of disease, including targeting ANRIL for non-small cell lung cancer and H19 for pancreatitis. However, there are currently no clinical trials registered in the ClinicalTrials.gov database that target lncRNAs as a therapeutic intervention. In order to translate lncRNA targeting into clinical use, several limitations must be overcome, such as potential toxicity and off-target effects. Overall, while significant progress has been made in the field, further development is required before the clinical application of the first therapeutics targeting lncRNAs. In this review, we discuss recent advances in our understanding of the mechanisms of action of lncRNAs that present avenues for clinical therapeutic targeting and consider off-target effects as a limiting factor in their application.

Keywords: lncRNAs, mechanism of action, therapeutics, ASO, siRNA, CRISPR-Cas9

2.2 Introduction

Non-coding RNAs (ncRNAs) are RNAs that do not encode protein but may possess important regulatory functions (Derrien et al., 2012, Mattick and Makunin, 2006). In this review, we discuss the potential application of long non-coding RNAs (lncRNAs) as therapeutic targets. Examples of lncRNAs that are dysregulated and causative in disease pathologies will be discussed in relation to their mechanism of action, together with how their roles in disease can make them suitable therapeutic targets, as well as examples of successfully targeting them. Targeting lncRNAs therapeutically has been hailed as an exciting field for some time

and is rapidly expanding, but to date, we have found no registered clinical trials for this application. In this review, we explore the real potential of lncRNA targeting and where they provide a unique therapeutic opportunity. This will be achieved by looking at some of the best examples where there is genuine evidence for modulation of lncRNA expression in treating disease. Furthermore, while studies have linked many lncRNAs to a variety of diseases, not all of them are robust and many of their applications have limitations that are also considered in the review.

2.2.1 Non-coding RNAs

Regulatory ncRNAs are largely split into two classes: small non-coding (sncRNAs) RNAs of < 200 nucleotides in length and lncRNAs of > 200 nucleotides [Figure 2.1] (Kapranov et al., 2007, Yao et al., 2019). Some classes of ncRNAs are of variable length, so they can belong to either classification. These include enhancer RNAs (eRNAs) (Arnold et al., 2019) and circular RNAs (circRNAs) (Lasda and Parker, 2014).

There are numerous lncRNAs, with the GENCODE database annotating over 20,000 to date in humans (GENCODE, 2024), but other resources estimate over 90,000 (Iyer et al., 2015, Li et al., 2023). lncRNAs are found throughout the cell, but are most commonly in the nucleus (Cabili et al., 2015) or cytoplasm (Benoit Bouvrette et al., 2018, Carlevaro-Fita et al., 2016). The subcellular localisation of lncRNAs is highly regulated and is important in deciphering their functions (Bridges et al., 2021). Nuclear lncRNAs are generally less stable (Clark et al., 2012), but only a minority of lncRNAs are deemed unstable enough to make them unsuitable for therapeutic targeting (Clark et al., 2012). In the nucleus, lncRNAs function in chromatin remodelling, transcriptional regulation, and as scaffolds for spatial organisation, whereas cytoplasmic lncRNAs are involved in translational control, post-transcriptional control of gene expression, and protein localisation (Bridges et al., 2021, Ma et al., 2013). The factors controlling their localisation have been reviewed in depth elsewhere (Bridges et al., 2021).

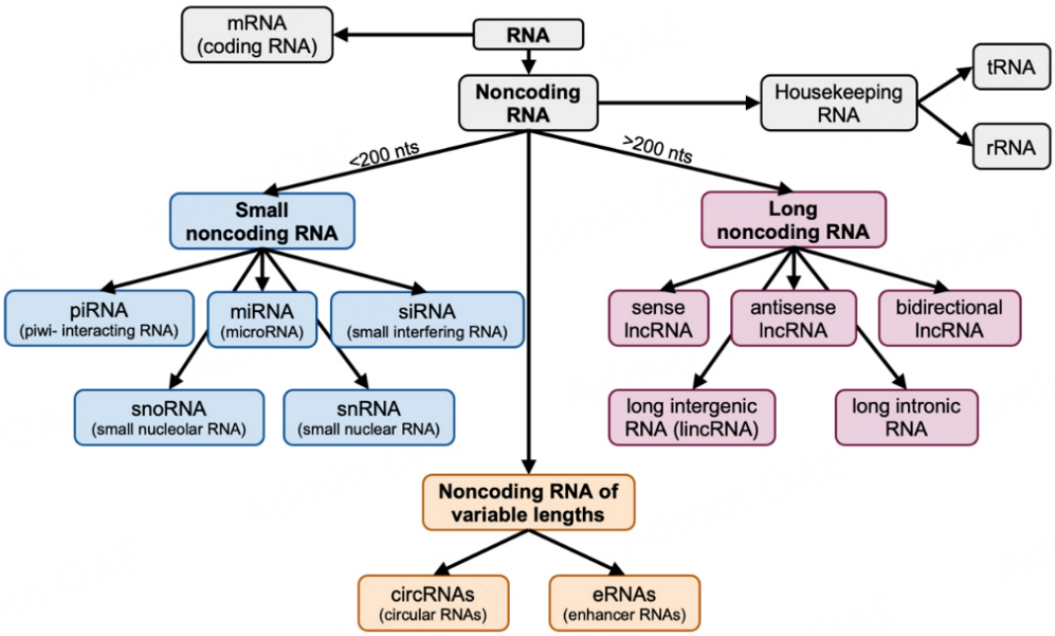


Figure 2.1. Classification of RNAs. RNA can be classified as coding or non-coding RNA. *Non-coding RNA can be separated into housekeeping ncRNAs and regulatory ncRNAs, which are the subject of this review. Regulatory ncRNAs can be separated according to their length, with small non-coding RNAs being less than 200 nucleotides and long non-coding RNAs greater than 200 nucleotides. snoRNAs and snRNAs are generally classified as sncRNAs but possess housekeeping functions. lncRNAs can be classified according to their relative location to protein-coding genes. The main types of small non-coding RNAs are also further classified here according to their function. Additionally, circRNAs and eRNAs are ncRNAs that can be of variable length, either greater or less than 200 nucleotides, so they do not fit into either length classification.*

2.2.2 Current therapeutics targeting ncRNAs

Due to their ability to regulate gene expression, ncRNA targeting offers exciting opportunities for disease treatment. To date, there are 85 trials listed on the Clinical Trials database (ClinicalTrials.gov) that contain “lncRNA” when searched in other terms (ClinicalTrials.gov, 2024). Of these, the majority are testing or have tested the use of lncRNAs as biomarkers for monitoring disease progression/severity or as diagnostic tools. The few that are not looking at lncRNAs as biomarkers are concerned with identifying their mechanisms of action and role in a particular disease, for example, by investigating their relationship with other elements of the functional pathway (examples include www.clinicaltrials.gov NCT04937855, NCT06213493, and NCT04767750). None of these trials test drugs/treatments to target lncRNAs and all are listed as observational studies. This

lack of interventional trials suggests an underexploited area or, alternatively, that there are significant barriers to overcome before treatments targeting lncRNAs can successfully reach the clinical trial stage.

However, although the field of lncRNA targeting in clinical trials is undeveloped, the therapeutic targeting of microRNAs (miRNAs) is more advanced. Several anti-miRNA drugs are currently undergoing clinical trials, including Miravirsen, which has successfully completed phase II clinical trials for hepatitis C (Janssen et al., 2013, van der Ree et al., 2016). Miravirsen is a locked nucleic acid (LNA)-modified antisense oligonucleotide (ASO) that binds miR-122, inhibiting its action in stabilising hepatitis C RNA (Janssen et al., 2013). This was one of the first drugs developed to specifically target an ncRNA, but it was discontinued due to the availability of other effective treatments.

However, its development demonstrates the potential of this therapeutic approach. Other methods have also been developed and miRNA mimics, synthetic copies of endogenous miRNAs, such as MRX34 to augment in vivo miR-34a levels, are being tested clinically (Hong et al., 2020). Despite immune-related severe adverse events causing early termination of this study, it demonstrates a method of exploration for ncRNA therapeutics (Hong et al., 2020). These effects included sepsis, hypoxia, cytokine release syndrome and hepatic failure, ultimately resulting in four patient deaths (Hong et al., 2020). Remlarsen, another miRNA mimic, is undergoing clinical trials for restricting fibrous scar tissue formation (ClinicalTrials.gov, 2021). For a comprehensive overview of therapeutics that have completed or are currently undergoing clinical trials for targeting miRNAs and other ncRNAs, see Winkle et al. (2021) (Winkle et al., 2021).

ncRNA-targeted therapeutics offer significant potential, but they do have several limitations, many of which are also relevant to the development of therapeutics targeting lncRNAs. One problem is successfully delivering RNA-based therapeutics to target tissues other than the liver. To this end, chemical modification of various parts of the nucleotide has enabled the successful delivery of ASOs to multiple tissues without a delivery agent (Zhu et al., 2022). Furthermore, novel polymer and peptide-based nanoparticle delivery systems have reduced issues with charge and electrostatic interactions to improve miRNA delivery (Babar et al.,

2012, Woodrow et al., 2009). The addition of surface peptides also improves cellular uptake (Babar et al., 2012). These methods are safe and biodegradable and, thus, could be applicable to developing other ncRNA therapeutics (Adams et al., 2017). However, other significant issues include sequence and tissue specificity, leading to off-target binding (Kamola et al., 2015), as well as tolerability, leading to toxicity, particularly hepatotoxicity (Grimm et al., 2006). The Phase I MRX34 study encountered immunity-related toxicity, despite no problems being observed in animal studies (Hong et al., 2020), indicating that preclinical models do not consistently predict human responses. These issues must be thoroughly assessed before further clinical studies can be carried out safely and successfully.

2.2.3 The rationale for targeting lncRNAs

lncRNAs are present in a wide range of animals, plants, prokaryotes, yeast, and viruses, but their sequence is not well conserved between species (Ma et al., 2013). They face less selection pressure than mRNAs, since they do not need to maintain a specific open reading frame (Chodroff et al., 2010). Some lncRNAs, for example, X-inactive Specific Transcript (XIST), possess short regions with high conservation, suggesting a lack of evolutionary selection pressure on other regions, such that only the target binding sequences are conserved (Pang et al., 2006, Nesterova et al., 2001). Thus, lncRNAs can evolve rapidly. lncRNAs have specific spatio-temporal expression, i.e. they are activated in specific tissues at specific times during development, and this expression often affects the expression of nearby protein-coding genes (PCGs) and contributes to the lineage-specific expression of PCGs (Kutter et al., 2012). Furthermore, even conserved lncRNAs often function differently between species due to alternate processing and localisation (Guo et al., 2020). This initially led to questions over their importance and functionality, but subsequent research has uncovered diverse functions in the regulation of transcription, splicing, translation, differentiation, the cell cycle, nuclear bodies, and chromatin (Bridges et al., 2021, Ma et al., 2013). In fact, lncRNA promoters are conserved at a similar rate to PCG promoters (Derrien et al., 2012), suggesting lncRNA expression is important for fitness, and even those with rapidly changing sequences often have orthologous functions between species and are expressed from syntenic locations (Ulitsky et al., 2012, Hezroni et al., 2015).

The ubiquity of ncRNAs and their ability to target multiple genes within a pathway makes them excellent therapeutic targets (Winkle et al., 2021). A transcriptome-wide association study (TWAS) found that of 14,100 lncRNA genes, expression of 800 was associated with genetic traits of disease where the association was not due to any effects of neighbouring PCGs, making it less likely that the effects were due to alterations of cis-regulatory sequences overlapping the lncRNA (de Goede et al., 2021). This represents a large number of potential therapeutic targeting opportunities by modulating lncRNA expression. However, their ability to target multiple genes does raise the issue of off-target effects in other genes, as has been seen in therapeutics targeting other ncRNAs. lncRNAs, in particular, have high organ, tissue, and cell type specificity (Derrien et al., 2012, Sarropoulos et al., 2019). This spatial and temporal expression makes them excellent targets for lineage-specific gene therapy (Nguyen and Carninci, 2015). lncRNA dysregulation has been linked to cancer (Huarte, 2015), e.g., MEG3 downregulation in multiple cancers (Zhou et al., 2007), cardiovascular diseases (Boon et al., 2016), e.g., MALAT1 in diabetic retinopathy and angiogenesis (Liu et al., 2014), neurological disorders (Bhan and Mandal, 2014), e.g., BASE1-AS in Alzheimer's Disease (Faghihi et al., 2008), musculoskeletal disorders (Mishra et al., 2023), e.g., ANRIL in osteoarthritis (Li et al., 2019), and many other diseases. In cancer, lncRNAs have been identified as oncogenes, e.g., HOTAIR (Gupta et al., 2010) and MALAT1 (Ji et al., 2003) or tumour suppressors, e.g., MEG3 (Chen et al., 2021a); therefore, they are ideal targets for new cancer therapeutics (Huarte, 2015). lncRNAs also play key roles in tumour microenvironments. For example, in influencing immune cell function, LNMAT1 recruits macrophages into tumour cells to enhance lymphatic metastasis (Chen et al., 2018) and LINC00301 increases levels of regulatory T cells while decreasing CD8+ T cells, in non-small cell lung cancer (NSCLC) (Sun et al., 2020).

2.3 Strategies to target lncRNAs

There have been several well-documented methods used to target lncRNAs and modulate their expression, including siRNAs, ASOs, and CRISPR-Cas9. ASOs and siRNAs have been used in many studies of lncRNA knockdown. Typically, siRNA silencing is most effective for cytoplasmic lncRNAs, whereas ASOs are considered most effective for nuclear lncRNAs, but can also act in the cytoplasm (Chen et al., 2021b, Liang et al., 2017).

siRNAs are short oligonucleotides complementary to target ncRNAs and work by recruiting the RNA induced silencing complex (RISC) to degrade lncRNAs [Figure 2.2A] (Berezhna et al., 2006). To date, six siRNA-based drugs are approved by the Food and Drug Administration (FDA) and/or the European Medicines Agency (EMA), and all target mRNA in the liver [Table 1] (EMA, 2023, FDA, 2024), demonstrating the efficacy of this approach for targeting RNA. This method has also been successful in several lncRNA preclinical models, but concerns remain over the potential adverse effects of targeting molecules other than the intended lncRNA (Chen et al., 2021b).

Table 1. Currently approved siRNA/ASO-based drugs (as of May 1, 2024). *This shows the type of sequence targeted by these drugs and which condition they are approved to treat. Two ASO drugs have recently been discontinued by the FDA and are highlighted in grey.*

	Active substance (drug name)	**Brand name**	**Target sequence**	**Condition treated**	**Licensed by (EMA and/or FDA)**
siRNA	Givosiran	Givlaari	mRNA	Acute hepatic porphyria	EMA, FDA
	Inclisiran	Leqvio	mRNA	High cholesterol (primary hypercholesterolaemia, mixed dyslipidaemia)	EMA, FDA
	Lumasiran	Oxlumo	mRNA	Primary hyperoxaluria type 1	EMA, FDA
	Nedosiran	Rivfloza	mRNA	Primary hyperoxaluriatype 1	EMA, FDA
	Patisiran	Onpattro	mRNA	Polyneuropathy in hereditary tranthyretin-mediated amyloidosis	EMA, FDA
	Vutrisiran	Amvuttra	mRNA	Polyneuropathy in hereditary tranthyretin-mediated amyloidosis	EMA, FDA
ASO	Aganirsen	Olisens	mRNA	Ocular neovascularisation	EMA
	Casimersen	Amondys 45	exon	Duchenne muscular dystrophy	FDA
	Eteplirsen	Exondys 51	exon	Duchenne muscular dystrophy	FDA
	Fomivirsen	Vitravene	mRNA	CMV infection	FDA **(now discontinued)**
	Golodirsen	Vyondys 53	pre-mRNA	Duchenne muscular dystrophy	FDA
	Inotersen	Tegsedi	mRNA	Homozygous familial hypercholesterolemia	EMA, FDA
	Mipomersen	Kynamro	mRNA	Homozygous familial hypercholesterolemia	FDA **(now discontinued)**

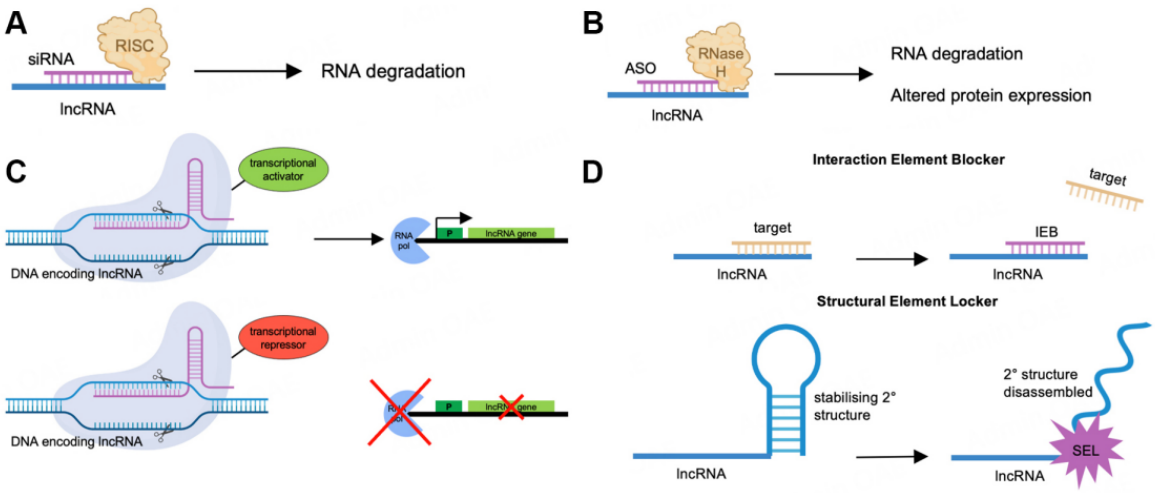


Figure 2.2. Schematic illustration of strategies used to target lncRNA. *(A) siRNAs: siRNAs bind to lncRNA and recruit RISC, resulting in degradation of the lncRNA. (B) ASOs: ASOs bind to lncRNA and recruit RNaseH, resulting in degradation of the lncRNA and altered downstream protein expression. (C) CRISPR-Cas9: The inactive Cas9 domain is bound to a transcriptional activator domain or a transcriptional stop signal, so that when it binds to the complementary DNA that encodes the lncRNA gene, it results in either transcriptional activation at the promoter, or repression through blocking RNA polymerase, respectively. (D) Small molecules: the first small molecules designed for modulating lncRNA expression can be classified as interaction element blockers (IEBs) or structural element lockers (SELs). IEBs block the binding of lncRNA to its target, which in some cases can be used to increase expression levels of lncRNAs that would normally undergo nonsense-mediated decay due to their normal binding. SELs work by binding to lncRNAs and disrupting secondary (2°) structures which stabilise the lncRNA, thus resulting in destabilisation and reduced expression.*

Nusinersen	Spinraza	pre-mRNA	Spinal muscular atrophy	EMA, FDA
Vitolarsen	Viltepso	exon	Duchenne muscular dystrophy	FDA

ASOs are 15-25 bp oligonucleotides that can bind complementary lncRNA, and commonly recruit RNase H to promote RNA degradation and alter downstream protein expression when coding elements are targeted [Figure 2.2B] (Chen et al., 2021b). ASOs can also act by binding mRNA to alter splicing that results in exon inclusion where mutations have led to exon skipping (Hua and Krainer, 2012), or they can be used to cause exon skipping in diseases such as Duchenne muscular dystrophy where deletion mutations shift the reading frame and generate premature stop codons (Matsuo, 2021). Additionally, ASOs can alter the

site of polyadenylation to destabilise RNA (Vickers et al., 2001). To date, seven ASO-based drugs currently have marketing authorisation, with many more currently undergoing clinical trials [Table 1] (EMA, 2023, FDA, 2024, Collotta et al., 2023). However, some ASO-based drugs have been discontinued from the market after their authorisation due to hepatotoxicity problems (Hawes et al.). The ASO drug Inotersen also contains an FDA warning for hepatotoxicity on its label (Hawes et al.). These mechanisms, however, are less applicable to lncRNA targeting. Although these ASOs do not target lncRNAs, their success in reaching the market demonstrates that they can be effective RNA-targeted drugs. Through their ongoing development, some initial problems have been overcome - the latest ASOs designed have high affinity and stability (Chen et al., 2021b). However, toxicity issues remain, including hepatotoxicity (Chen et al., 2021b) and renal toxicity that can lead to potentially fatal glomerulonephritis (Wu et al., 2022). Furthermore, the fact that they cannot be administered orally is a limitation (Chen et al., 2021b). Both ASOs and siRNAs can be used in conjunction with LNAs to increase potency, but this can also increase hepatotoxicity (Swayze et al., 2007).

CRISPR/Cas9 is another tool that can be used to target nuclear or cytoplasmic lncRNAs, but in one study, it was only effective in 38% of ~16,000 lncRNA loci tested (Goyal et al., 2017). CRISPR/Cas9 may upregulate lncRNA expression by activating the promoter with a fusion protein of inactive Cas9 with a transcriptional activator domain (Maeder et al., 2013), or disable the lncRNA gene with a transcriptional stop signal to block RNA polymerase [Figure 2.2C] (Boon et al., 2016, Bassett et al., 2014). Some loci are not readily targeted with CRISPR, due to their bidirectional, internal, or proximal promoter, or due to off-target effects in neighbouring genes (Goyal et al., 2017).

More recently, strategies have been developed to target ncRNAs with small molecules specifically targeted to their secondary structures [Figure 2.2D] (Winkle et al., 2021). The ability to predict structures and virtually screen compounds accelerates this process (Winkle et al., 2021, Zhao et al., 2022). This method has been used for lncRNAs; AC1NOD4Q, a compound targeting the lncRNA homeobox antisense intergenic RNA (HOTAIR), has been developed to selectively interfere with HOTAIR-EZH2 binding, thus blocking its activity (Ren et al., 2019). This was achieved by developing 3D models to predict hairpin loop structures

that could be targeted with small molecules, followed by virtual screening of potential molecules (Ren et al., 2019). This methodology can be applied to the design of small molecules to target other lncRNAs, but requires the high-resolution 3D structure of the respective lncRNAs (Zhao et al., 2022).

2.4 Targeting lncRNAs therapeutically according to their mechanism of action

The mechanisms of action for lncRNAs can be broadly separated into scaffold, guide, or decoy RNAs (Ma et al., 2013, Dong et al., 2018, Wang and Chang, 2011). These mechanisms are distinct, but many lncRNAs can act via multiple mechanisms (Wang and Chang, 2011) and thus there may be multiple ways to exploit their actions therapeutically.

2.4.1 lncRNAs as scaffold molecules

When functioning as scaffolds, lncRNAs act as a platform to assemble different regulatory proteins together to perform a specific function (Spitale et al., 2011). This is possible through the presence of different domains that simultaneously bind various effector molecules, such as transcriptional activators or repressors, which have specific effects when brought together both spatially and temporally [Figure 2.3] (Wang and Chang, 2011). lncRNAs can act in a cis manner on neighbouring genes, or a trans manner on distant genes (Guttman and Rinn, 2012). In transcription, scaffold lncRNAs can activate or silence specific genes by binding different subunits of chromatin-modifying complexes to facilitate their assembly, such as the polycomb repressive complex (PRC) 1 and PRC2 (Dong et al., 2018). Therefore, knockdown of scaffold lncRNAs would inhibit the effector molecules from interacting with their target and double knockdown of lncRNAs with the effectors should exacerbate these effects (Wang and Chang, 2011). Understanding how lncRNAs assemble and regulate these effector molecules is thus crucial to targeting them effectively.

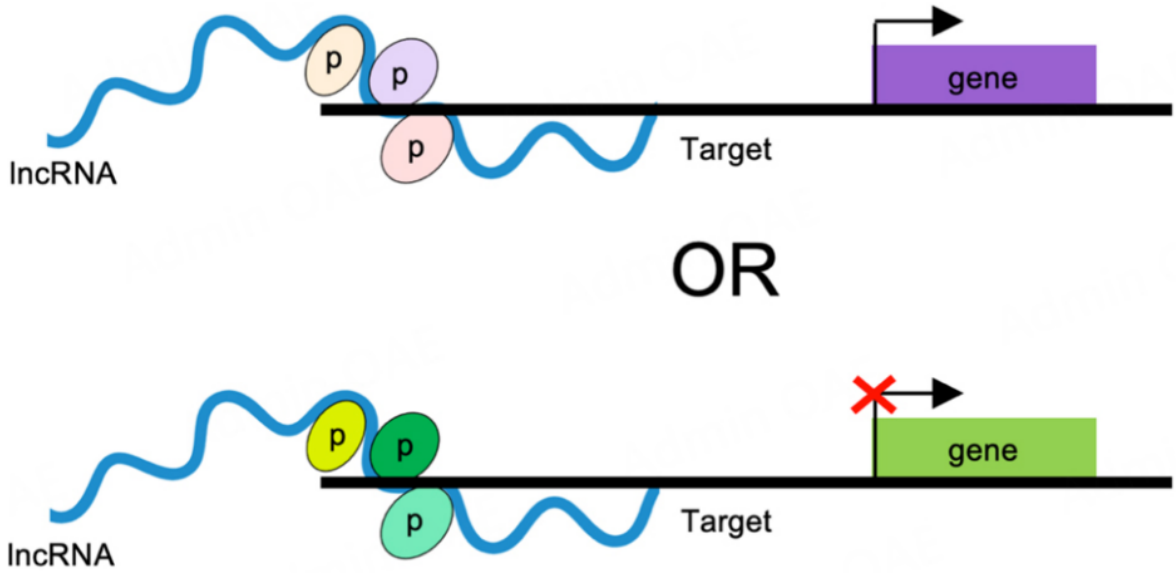


Figure 2.3. Schematic illustration of the functional mechanism of scaffold lncRNAs. *Scaffold lncRNAs act by using different modules to bring together different proteins (labelled "p"), such as transcriptional activators and repressors, in time and space to cause specific effects on target molecules.*

The lncRNA HOTAIR can act through a scaffold mechanism. HOTAIR binds PCR2 and LSD1 chromatin remodelling complexes and acts as a platform to target these to HOX loci, to influence cell epigenetic states (Rinn et al., 2007, Tsai et al., 2010). HOTAIR has four independently folded domains, with two of them recruiting and interacting with various transcription factors, nine of which are involved in pan-cancer processes (Wang et al., 2017). Increased HOTAIR expression in primary tumours strongly predicts metastases and death, for example, in breast cancer, where it is commonly highly expressed (Gupta et al., 2010, Sorensen et al., 2013). Consequently, enforced expression in multiple breast cancer cell lines led to increased cancer invasiveness and PRC2-dependent metastasis, while knockdown by siRNAs inhibited cancer invasiveness, with a bias for cells with high PRC2 activity (Gupta et al., 2010). Furthermore, grafting HOTAIR-expressing cells into murine fat pads accelerated primary tumour growth and promoted lung metastasis (Gupta et al., 2010). HOTAIR led to selective retargeting of PRC2 across the genome by aiding the localisation of its subunits to 854 genes, which gain PRC2 occupancy and are consequently downregulated in the most aggressive breast cancer tumours (Gupta et al., 2010). This identifies HOTAIR as a relevant target for new cancer therapeutics, particularly by exploiting its interactions with PRC2. However, as HOTAIR acts in trans, validating its role would be more robust if HOTAIR could be knocked out and then the phenotype rescued by expressing it from an independent

transgene (Bassett et al., 2014). Furthermore, the study of human HOTAIR in vivo is challenging, as there is poor sequence conservation between human and murine HOTAIR, but the orthologs do have similar functions and conserved RNA structures (Somarowthu et al., 2015, Li et al., 2013). Ma et al. (2022) created a transgenic murine model with inducible expression of human HOTAIR to study the role of HOTAIR in breast cancer progression (Ma et al., 2022b). Mice overexpressing HOTAIR were crossed with MMTV-PyMT mice, a commonly used model of breast cancer. Overexpression of HOTAIR for several months increased the invasiveness of breast cancer cells, promoting their migration and metastasis to the lungs (Ma et al., 2022b). Removal of this overexpression abrogated these effects. Mechanistically, HOTAIR alters chromatin states and the transcriptome to cause changes that result in the promotion of metastatic pathways, by influencing both repressive and activatory modifications (Ma et al., 2022b). This demonstrates that HOTAIR could be a target for downregulation in breast cancer treatment. Although this model did not test potential therapeutics to knock down HOTAIR, it does represent an important tool that could be used to test siRNAs, ASOs, or other therapeutics targeting HOTAIR.

Another example of a scaffold lncRNA is the antisense non-coding RNA in the INK4 Locus (ANRIL) that also recruits and binds to PRC1/2 to modify transcription (Kotake et al., 2011). High levels of ANRIL expression are associated with increased metastases and tumour size, leading to poor prognosis in NSCLC patients (Nie et al., 2015). Knockdown of ANRIL through RNAi in six human NSCLC cell lines resulted in impaired cell proliferation in five of the cell lines and induced apoptosis, as well as inhibiting tumour growth in vivo when one cell line was transfected into mice (Nie et al., 2015). The mechanism for this involves ANRIL binding to EZH2, a core subunit of PRC2, to silence KLF2 and P21 (Nie et al., 2015). Both KLF2 and P21 play significant roles in cancer, where KLF2 expression is typically reduced and is also associated with NSCLC cell apoptosis (Nie et al., 2015). Thus, investigating methods to downregulate ANRIL in NSCLC, through its scaffolding role in binding PRC2, could provide potential therapeutic options for its treatment.

2.4.2 lncRNAs as guide molecules

lncRNAs can act as guides by binding proteins and guiding their localisation to specific targets, causing alterations in gene expression [Figure 2.4] (Khalil et al., 2009). This can affect transcription by guiding the recruitment of transcriptional activators, e.g., Trithorax group proteins, or suppressors, e.g., Polycomb group proteins, in a site-specific manner (Wang and Chang, 2011). Guide lncRNAs can act in cis or trans to their protein-coding targets (Guttman and Rinn, 2012). Similar to scaffold lncRNAs, the knockdown of guide lncRNAs inhibits the localisation of the effector molecule to its target, resembling a loss of function phenotype, while a double knockdown with the effector molecule should augment this effect (Wang and Chang, 2011).

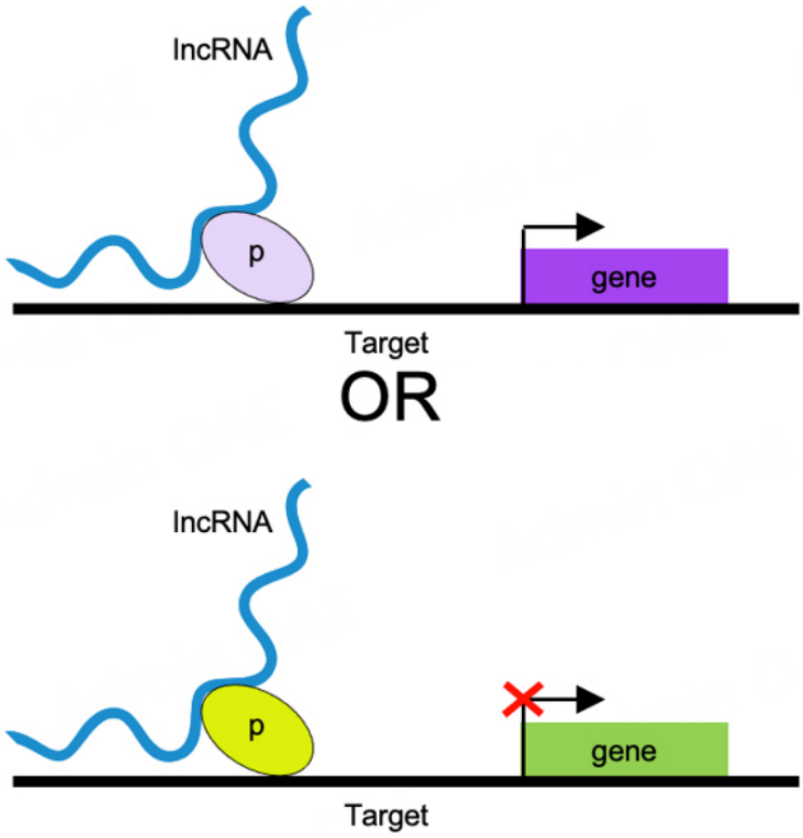


Figure 2.4. Schematic illustration of the functional mechanism of guide lncRNAs. *Guide lncRNAs function by guiding proteins (labelled "p") to localise to specific targets in cis or trans, resulting in changes in gene expression. These effects can include activation or repression of genes depending on the specific protein.*

XIST is one of the first lncRNAs to be characterised and acts via a cis guide mechanism to initiate X chromosome inactivation (XCI). XIST assists in recruiting the inactive X chromosome to the nuclear lamina and binds across it, with greater affinity for the most gene-dense regions (Engreitz et al., 2013, Chen et al., 2016). XIST recruits PRC2 and promotes repressive chromatin modifications that cause transcriptional silencing (Engreitz et al., 2013, Chaumeil et al., 2006). XIST is thus differentially expressed in X-linked diseases, and plays a critical role in many sex-biased diseases including autoimmune diseases such as rheumatoid arthritis (Youness et al., 2021, Bost et al., 2022), neurological disorders such as Alzheimer's Disease (AD) (Yue et al., 2020, Yan et al., 2022), pulmonary arterial hypertension (Batton et al., 2018, Qin et al., 2021), and sex-biased cancers, as reviewed in (Li et al., 2022). In cancer, the role of XIST is complex and can have conflicting effects in protecting against or promoting cancer progression (Li et al., 2022). In breast cancer, which predominantly affects females, XIST is abnormally downregulated relative to normal female breast tissue (Zheng et al., 2018). Additionally, studies in vitro and in mice have successfully altered XIST expression to slow bladder (Xu et al., 2018), colorectal (Sun et al., 2018), and lung cancer progression (Fang et al., 2016). In AD, higher levels of XIST are observed, alongside increased inflammatory cytokines (Yan et al., 2022). In murine models of AD and in vitro, XIST promotes Aβ protein accumulation and neuroinflammation by epigenetically silencing neprilysin (NEP), an Aβ degrading enzyme, through recruitment of EZH2 (Yan et al., 2022). Knockdown of XIST in the same murine AD model reduced cell injury and neuronal inflammation, as NEP levels were increased, suggesting a potential route for therapeutics could be to downregulate XIST (Yan et al., 2022). However, it is important to consider that attempts to target XIST for specific diseases could have wider consequences for other diseases influenced by XCI. Thus, XIST targeting must be approached cautiously and will require a thorough understanding of its mechanisms. Furthermore, attempts have also been made to utilise XIST's chromosome silencing function to target Down's Syndrome, by inserting an inducible XIST transgene into chromosome 21 of pluripotent stem cells with trisomy 21 (Jiang et al., 2013, Chiang et al., 2018). This successfully led to chromosome-wide silencing and methylation to create a Barr body with the additional chromosome 21 (Jiang et al., 2013). This approach to chromosome silencing could also be investigated for the treatment of other trisomy conditions.

Maternally expressed gene 3 (MEG3) is a lncRNA that also interacts with PRC2, guiding it to modulate the activity of TGF-β-regulated genes by binding to chromatin in a trans manner, and is recruited to loci by the formation of RNA-DNA triplex structures (Mondal et al., 2015). MEG3 is downregulated in many cancers (Zhang et al., 2017b, Xia et al., 2015, Zhang et al., 2017a), for example, in non-functioning pituitary adenomas (NFA), where MEG3 is silenced (Chunharojrith et al., 2015). Restoring MEG3 expression in cells derived from human pituitary tumours significantly slowed tumour growth when grafted into mice in vivo, by inducing G1 cell cycle arrest. This tumour suppression required the presence of functional p53. This presents a potential opportunity for therapeutic targeting and could be relevant to many cancer types if similar mechanisms are found to be at play.

2.4.3 lncRNAs as decoy molecules

lncRNAs can act as molecular decoys by binding proteins, such as transcription factors, chromatin modifiers, or other regulatory factors, to up- or downregulate transcription (Wang and Chang, 2011). lncRNAs can sequester these molecules to inhibit their target binding (frequently to chromatin), thus interfering with transcription [Figure 2.5A] (Dong et al., 2018). Knockdown or knockout of decoy lncRNAs may increase the expression of their targeted molecule, thus mimicking the gain of function for the protein (Wang and Chang, 2011).

Similarly, lncRNAs commonly contain miRNA binding sites and act as “molecular sponges” sequestering miRNAs away from their mRNA targets [Figure 2.5B] (Dong et al., 2018). miRNAs control the activity of PCGs by binding mRNA transcripts and recruiting protein complexes to repress translation and/or decrease mRNA stability (Filipowicz et al., 2008). As miRNAs are also dysregulated in many diseases, understanding these miRNA-lncRNA interactions could have far-reaching therapeutic potential by altering lncRNA expression.

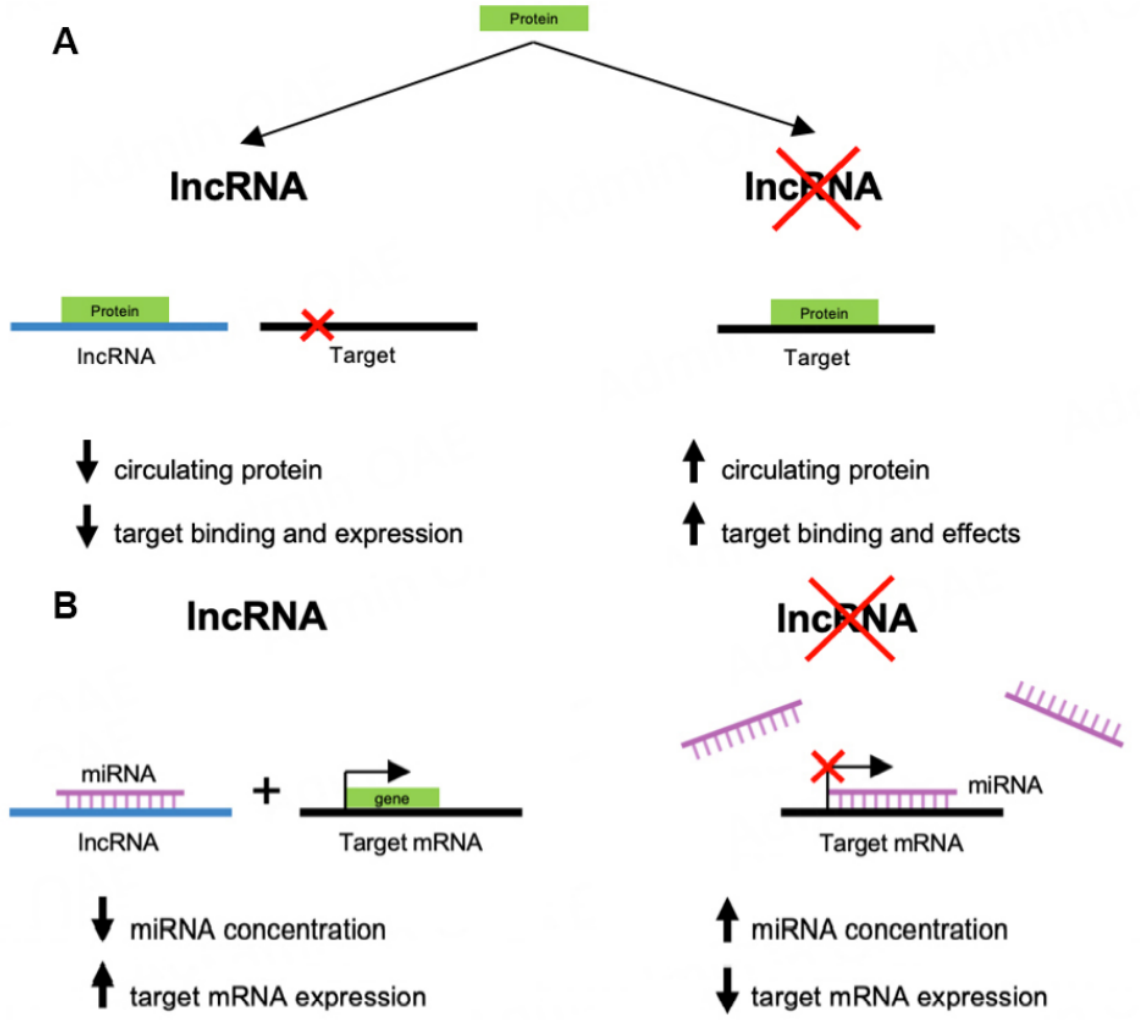


Figure 2.5. Schematic illustration of the functional mechanism of decoy lncRNAs. *LncRNAs are shown in blue, target mRNA/molecules in black, miRNA in purple, and genes/proteins in green. (A) LncRNAs can act as molecular decoys for proteins, such as those involved in transcription, by binding the protein and stopping binding to their target molecule, e.g., chromatin. This stops any effects that are a result of the protein binding to its target. Without lncRNA presence, e.g., through knockout, there is an increase in circulating protein, so it can bind its target and produce the intended effects. (B) Here, lncRNAs act as competing endogenous RNAs to bind and sequester miRNAs, making them unable to bind to their target mRNA. Typically, miRNAs inhibit their target mRNA sequence, so this sequestering results in an increase in the expression of their target mRNA (Karagkouni et al., 2021). In the absence of lncRNAs, e.g., through knockout, miRNAs are free to bind to their target mRNAs, inhibiting their expression.*

One lncRNA that acts through a decoy mechanism is the maternally expressed H19, which plays a role in the imprinted gene network during embryonic growth (Monnier et al., 2013). H19 is upregulated in cardiac, pulmonary, hepatic, and renal fibrosis, and can act by

sponging multiple miRNAs, as reviewed in (Jiang and Ning, 2020). In the pancreas, H19 acts as a competing endogenous RNA (ceRNA) and sponges miR-138-5p and miR-141-3p (Song et al., 2020). Knockdown of miR-138-5p and miR-141-3p suppresses autophagy by increasing the activity of the focal adhesion kinase (FAK) pathway and promoting cell proliferation by increasing β-catenin levels, respectively (Song et al., 2020). However, in severe acute pancreatitis (SAP), H19 expression is suppressed, suggesting it could be a potential target for upregulation in SAP treatment. Rats with SAP were treated with mesenchymal stem cells (MSCs) transfected with a H19 overexpression plasmid, which increased the MSCs anti-inflammatory properties, promoted FAK-associated pathways, and increased cell proliferation (Song et al., 2020). Overall, this use of MSCs demonstrates an effective route to target and modulate H19 expression for SAP therapeutics. Furthermore, H19 plays a role in triple-negative breast cancer (TNBC), where it is upregulated and its expression levels are inversely correlated with lncRNA PTCSC3 (Wang et al., 2019). Overexpression of PTCSC3 inhibits TNBC cell proliferation by downregulating H19, while overexpressing H19 has no effect on PTCSC3 and promotes TNBC cell proliferation (Wang et al., 2019). However, no evidence has been presented to indicate a physical interaction between H19 and PTCSC3, and further evidence is required to establish whether this interaction is causal in TNBC. If such evidence did come to light, then it is possible that downregulating H19 could provide therapeutic benefit, and this could be achieved through overexpression of PTCSC3 (Wang et al., 2019).

In addition to its more well-known roles in transcriptional regulation, HOTAIR also has sponging functions. In diabetic cardiomyopathy (DCM), HOTAIR is downregulated, while its target, miR-34a, is upregulated (Gao et al., 2019). Observation of HOTAIR knockdown in a mouse model of DCM showed increased inflammation, oxidative stress, and cell death in the heart, which represent the hallmarks of DCM (Gao et al., 2019). Increasing HOTAIR levels in DCM mice alleviated cardiac dysfunction and inhibited cardiac fibrosis (Gao et al., 2019). The putative mechanism for this is HOTAIR acting as ceRNA and sponging miR-34a, stopping its inhibition of SIRT1, an important gene involved in DCM regulation (Gao et al., 2019, Karbasforooshan and Karimi, 2017). This was evidenced through the knockout of SIRT1 in mice, which removed the benefits of HOTAIR expression and resembled the effects of HOTAIR knockout (Gao et al., 2019). Furthermore, only mature miR-34a (and not the primary

or precursor transcripts) was upregulated in HOTAIR knockout cells, suggesting HOTAIR downregulates miR-34a post transcriptionally, rather than regulating its transcription (Gao et al., 2019). Alongside evidence from RNA pull-down assays and luciferase reporter assays, this strongly supports HOTAIR directly targeting miR-34a (Gao et al., 2019). This evidence suggests HOTAIR could be targeted for DCM through methods to increase its expression.

Aberrant expression of ANRIL has been identified in osteoarthritis (OA) tissue, as it is significantly upregulated in synoviocytes of the OA-affected joint (Li et al., 2019). Downregulation of ANRIL halts cell cycle progression and promotes apoptosis in synoviocytes, possibly by sponging miR-122-5p, which regulates DUSP4 (Li et al., 2019). As synoviocyte proliferation is a common component of OA pathology, this suggests ANRIL downregulation could be a potential therapeutic pathway for OA by influencing the miR-122-5p/DUSP4 axis. However, this research is limited to an in vitro study of patient OA tissue and has not examined ANRIL mechanisms in vivo. There may also be potential issues with reducing DUSP4 expression, as it can have roles as a tumour suppressor, for example, in breast (Jung et al., 2016) and colorectal cancer (Saigusa et al., 2013) .

The lncRNA dishevelled binding antagonist of beta catenin3 antisense1 (DACT3-AS1) is downregulated in gastric cancer (GC) and plays a role in its chemoresistance (Qu et al., 2023). DACT3-AS1 aids suppression of cell proliferation, migration, and invasion, as identified through in vitro and in vivo experiments in a xenograft tumour mouse model (Qu et al., 2023). This is achieved by targeting the miR-181a-5p/sirtuin 1 axis, which may operate through a sponging mechanism (Qu et al., 2023). miR-181a-5p levels are increased in GC and negatively correlated with DACT3-AS1 (Qu et al., 2023). DACT3-AS1 directly targets miR-181a-5p and inhibits its levels in GC cell lines, whereas DACT3-AS1 silencing enhances miR-181a-5p levels (Qu et al., 2023). This negative regulation suggests that DACT3-AS1 may play a role in the transcription of miR-181a-5p, thereby affecting its expression levels, but the location of DACT3-AS1 in the cytoplasm does support a sponging mechanism (Qu et al., 2023). miR-181a-50 negatively regulates sirtuin 1 and this DACT3-AS1/ miR-181a-5p/ sirtuin 1 axis was demonstrated to suppress malignant characteristics of GC cells (Qu et al., 2023) .

2.4.4 Other mechanisms

Acting as signalling molecules has been suggested as an additional mechanism of action, where lncRNAs act as molecular signals through their spatial and temporal expression (Wang and Chang, 2011). The initiation, elongation, or termination of these lncRNAs is in itself regulatory, or they can also harbour additional regulatory functions (Wang and Chang, 2011). One example is lincRNA-p21, activated by p53, which acts as a transcriptional repressor and promotes apoptosis (Huarte et al., 2010). Its expression is downregulated in coronary artery disease patients and in murine atherosclerotic plaques in vitro, as it represses proliferation and is pro-apoptotic in vascular smooth muscle cells and murine mononuclear macrophages in vitro (Wu et al., 2014). Furthermore, in vivo silencing of lincRNA-p21 caused neointimal hyperplasia after endothelial injury (Wu et al., 2014). lincRNA-p21 acts by binding MDM2, to enhance p53 activity (Wu et al., 2014). This presents an opportunity for therapeutic targeting. However, lincRNA-p21 knockdown caused upregulation of 331 genes and downregulation of 274 genes, many of which are p53 target genes (Wu et al., 2014). Thus, altering lincRNA-p21 expression could have serious problems if off-target effects are produced by targeting it in unintended locations.

MALAT1 can act via a decoy mechanism as previously discussed but has additional functions in cancer and other diseases. There are several suggested mechanisms, including acting as a scaffold for nuclear speckles and chromatin, as a guide RNA in signalling pathways, or as a sponge for miR-200, but the exact mechanism remains unclear (Dhamija and Diederichs, 2016). MALAT1 plays a role in diabetic retinopathy and is upregulated in animal diabetes models (Liu et al., 2014). Its knockdown improves retinal function in diabetic rats by alleviating retinal vessel impairment and inflammation (Liu et al., 2014). Elevated MALAT1 levels have also been detected in human retinal endothelial cells (Biswas et al., 2018), as well as in association with other diabetic complications (Ashjari et al., 2022, Zhou et al., 2020, Zhang et al., 2018). MALAT1 acts by reducing levels of phosphorylated p38, part of the MAPK signalling pathway, to regulate endothelial cell function (Liu et al., 2014). Inhibiting MALAT1 presents a potential therapeutic target for diabetic retinopathy, although p38 MAPK signalling pathways act in many physiological processes, with the risk of off-target effects in other locations. Specific knockdown of MALAT1 in retinal tissues would be required by any

potential therapeutics to avoid impacting MALAT1 throughout the body and, importantly, in nuclear speckles. MALAT1 is also implicated in angiogenesis, where siRNA silencing of MALAT1 in mice causes endothelial cell migration instead of proliferation (Michalik et al., 2014). This implicates MALAT1 in angiogenesis regulation and in controlling the expression of cell cycle regulators. Thus, inhibiting MALAT1 could be of therapeutic benefit to induce antiangiogenic effects within tumour environments (Michalik et al., 2014). MALAT1 is frequently upregulated in cancers and has been successfully knocked down in mice using ASOs, resulting in differentiation of the primary tumour and a reduction in lung metastasis (Arun et al., 2016). Knockdown in breast cancer organoids also inhibited branching morphogenesis and altered expression of pro-tumourigenic and differentiation-related genes (Arun et al., 2016). This demonstrates the potential of ASO-based therapeutics in breast cancer, and, as MALAT1 plays a role in a large number of cancers, the potential for therapeutic targeting, although further studies with physiologically relevant in vivo models will be necessary.

The lncRNA nuclear enriched abundant transcript 1 (NEAT1) also has multiple suggested mechanisms of action. Importantly, it plays an essential role in the structure of nuclear paraspeckles by interacting with EZH2 and acting as a scaffold (Clemson et al., 2009). However, there is also evidence that it could act through sponging mechanisms to contribute to fibrosis development (Jiang, 2023). Knockdown of NEAT1 successfully reduces fibrosis in vitro and in various in vivo mouse models (Jiang, 2023). This identifies NEAT1 as a potential therapeutic target to prevent fibrosis of the liver (Yu et al., 2017), kidney (Chen et al., 2023a), heart (Ge et al., 2022), and lung (Fukushima et al., 2020) that are implicated in the progression of the diseases (Jiang, 2023). Evidence suggests that the mechanisms may differ somewhat between tissues. In the heart, NEAT1 recruits the PRC2 subunit EZH2 to Smad7, appearing to act as a scaffold, resulting in Smad7 inhibition and accelerating cardiac fibrosis (Ge et al., 2022). NEAT1 knockdown reduced cardiac fibrosis and dysfunction in mice (Ge et al., 2022). In the liver, NEAT1 upregulation with subsequent downregulation of miR-506 is associated with nonalcoholic fatty liver disease (NAFLD) (Jin et al., 2019). NEAT1 knockdown increases miR-506 expression, and inhibits GLI3, a miR-506 target, resulting in reduced fibrosis and inflammatory response (Jin et al., 2019). As luciferase assays identified miR-506 binding NEAT1 and GLI3, it is possible that this interaction is a sponging mechanism, but

further evidence would be required to rule out NEAT1 regulating miR-506 transcription as a mechanism. As NEAT1 is known to act via interactions with PRC2 in other tissues, the latter may be more likely. Regardless of the exact mechanism, it is possible that NEAT1 knockdown could be used therapeutically to treat fibrosis in various tissues.

LINC00301 is upregulated in NSCLC tumours and promotes cell proliferation, invasion, and tumourigenesis, while suppressing cell cycle arrest and apoptosis (Sun et al., 2020). This was demonstrated both in vitro and in vivo in a mouse model of human disease (Sun et al., 2020). LINC00301 is regulated by the transcription factor FOXC1, but there is evidence of it displaying multiple mechanisms of action. In the nucleus, it was found to interact with the PRC2 component, EZH2, suggesting either a guide or scaffold mechanism to affect transcription (Sun et al., 2020). However, LINC00301 is also present in the cytoplasm, and here, it may act as a ceRNA to sponge miR-1276 (Sun et al., 2020). Regardless of which mechanism is most significant to the role of LINC00301 in NSCLC pathogenesis, this represents an opportunity for therapeutic targeting.

2.5 Challenges and future directions

Several unresolved issues limit the clinical introduction of lncRNA therapeutic targeting. The low conservation of lncRNAs between species is a barrier for many research models. Humanised models or organoid cultures may be required to produce clinically translatable findings (Boon et al., 2016). Many studies discussed here have used mouse or rat models for knockout, knockdown, or overexpression studies. Although many have yielded promising results, it will be a new challenge to replicate these findings in human disease. Furthermore, human embryonic stem cells (ESCs) and mouse ESCs have different subcellular localisations of their lncRNAs, and as localisation is closely linked to function, this may affect how findings in other organisms relate to humans (Guo et al., 2020). lncRNA stability also varies between mice and humans; for example, MALAT1 and NEAT1 are both highly stable in humans, but unstable in mice (Clark et al., 2012) . Again, this can mean that findings in other organisms are not clinically relevant in humans. Another consideration is that lncRNAs are commonly expressed in multiple isoforms, as they undergo extensive alternative splicing (Bridges et al., 2021). These variants may have different functions, making mechanistic studies more challenging and further complicated by the fact that splicing varies between species (Boon

et al., 2016). Additionally, orthologs for human lncRNAs are only found for 38% and 35% of transcripts in mice and rats, respectively (Washietl et al., 2014). Many lncRNAs can also be modified, for example, through methylation, which may further affect their functions (Boon et al., 2016). Furthermore, the ability of lncRNAs to target multiple genes means attempts to target them could produce off-target effects and this risk has been highlighted in several of the examples of lncRNAs provided here. Careful consideration must be given to the likelihood and extent of these effects so that they can be most safely mitigated. The method of targeting lncRNAs is also a challenge and potential issues must be noted, particularly the ongoing problems with toxicity, which may not always be fully understood through animal studies.

Although much of the lncRNA research is promising, many assumptions are made regarding their association with particular diseases, as discussed in (Ponting and Haerty, 2022). This includes assuming that differential expression of a lncRNA in disease is causal and that interactions between a lncRNA and a protein implicated in disease indicate the lncRNA is responsible for modulating the disease risk (Ponting and Haerty, 2022). Moreover, the presence of disease associated single nucleotide polymorphisms (SNPs) within a lncRNA locus does not inherently imply a causal relationship with the disease (Ponting and Haerty, 2022). It is important to remember that although some lncRNAs have essential functions or play significant roles in disease phenotypes, not all of them present opportunities for therapeutic targeting. Going forward, a useful strategy to employ could be TWAS, which was developed in recent years to complement genome-wide association studies, to detect genes associated with traits (such as disease) and determine the regulatory relationship between them. TWAS offers improved gene interpretability, particularly for non-coding regions, and enables investigation of diseases on a tissue specific basis (Mai et al., 2023). For lncRNAs, the genetic association signals for transcript abundance in a specific tissue can be compared with signals for a particular disease, and if colocalisation is seen, then there is evidence of a causal role in the disease (Ponting and Haerty, 2022). Through such methods, the lncRNAs most relevant for experimental study can be identified and then tested, ideally with humanised models, to better understand their mechanism of action.

Overall, it can be challenging to characterise the functionally-relevant mechanisms of lncRNAs, particularly in discerning the difference between lncRNAs that sponge miRNAs and those that affect the transcription of their miRNA targets. However, if there is good evidence supporting the causal role of lncRNAs in a disease, precise characterisation of the mechanism may be of less importance and therapeutics targeting the lncRNA can still be developed. This makes it easier to develop therapeutics, but there are still other problems to overcome, including toxicity and off-target effects. Off-target effects can be a problem in several ways: either by affecting the intended target in an unintended location and causing undesirable effects, by targeting the correct molecule but producing undesired effects in its other downstream effectors that are unrelated to the disease, or by affecting the expression of molecules other than the intended target. The small molecule method of targeting lncRNAs offers an exciting avenue, as this reduces toxicity problems. The development of RNA-based therapeutics and particularly RNA vaccines in recent years has resulted in much research into the safety and efficacy of RNA-based therapeutics. There is also a trend for improvement in the development of ASOs (and other ncRNA-targeting therapeutics) for targeting PCGs and other ncRNAs. This knowledge is transferable to targeting lncRNAs, and ideally should make the journey to successful therapeutics somewhat easier.

2.6 Conclusion

There is excellent potential for developing therapeutics targeting lncRNAs for a wide range of diseases, but these are only currently in early-stage development. There is a long journey ahead to successfully target these lncRNAs therapeutically, overcome problems such as delivery efficiency, toxicity, and off-target effects through preclinical testing, then perform clinical trials, and eventually get approval to bring these treatments to market.

DECLARATIONS

Authors' contributions

Review design: Tamblin-Hopper P, Kiss-Toth E, Sudbery I, Young D, Wilkinson JM

Literature research, manuscript drafting: Tamblin-Hopper P

Manuscript editing and revisions: Tamblin-Hopper P, Kiss-Toth E, Sudbery I, Young D, Wilkinson JM

Availability of data and materials

Not applicable.

Financial support and sponsorship

Tamblin-Hopper P is supported by a studentship from the MRC Discovery Medicine North (DiMeN) Doctoral Training Partnership (MR/W006944/1).

Conflicts of interest

All authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Chapter 3: RNA-seq analysis of murine MSCs during osteogenic differentiation

3.1 Introduction

Analysis of previously collected RNA-seq data from murine MSCs with a *CASC20* knock-in (Felix-Ilemhenbhio, 2023) was performed. This experiment involved transduction of murine MSCs with either a *CASC20*-expressing lentivirus (referred to as OE) or a control green fluorescent protein (GFP)-expressing lentivirus (referred to as CTRL). Cells were cultured for 0, 10 or 20 days in Osteogenic Medium (OM), plus BMP2, to induce osteogenic differentiation. Robust CASC20 expression was confirmed in the OE samples at all time points through RT-qPCR, with no detectable expression in CTRL cells. After that cells were harvested to obtain RNA libraries and conduct RNA-sequencing on three biological replicates for each condition, Control (CTRL) and Over-Expression (OE), and time points Day (D)0, D10 and D20. The analysis aimed to elucidate evidence for the mechanism of action of *CASC20* by identifying DEGs and miRNAs. The identification of DEGs which are also targeted by miRNAs that bind to *CASC20* could provide evidence of a ceRNA mechanism.

3.2 Methods

3.2.1 RNA-sequencing Analysis

Sequencing output data, for both mRNA and miRNAs, (Fastq files) was handled in the University of Sheffield's High Performance Computing System 'Stanage'. Pipelines were run using Python (version 3.8.12). Reads were quality control checked using FastQC (version 0.12.0) and visualised with MultiQC (version 1.11), which looked at the GC content, presence of adapters and of any duplicated reads, which could indicate sequencing errors or contamination (Philip Ewels, 2016). Reads were trimmed with Cutadapt (version 3.7), to remove unwanted sequences such as adapter sequences and primers (Martin, 2011). For mRNA-seq, Salmon (version 1.7.0) (Patro et al., 2017) was used to quantify mouse transcripts from Ensembl Version 111 (GRCmm39). The percentage of mapped reads was checked to ensure sequence accuracy and check for any contaminating sequences. Differential expression analysis was performed in R (version 2023.09.1+494), with the DESeq2 package (version 1.42.1) (Love et al., 2014), to obtain the results of genes that have

a condition specific effect over time. The script for the analysis is shown in section 7.2.1. p-values were calculated using the Wald test. Principle Component Analysis (PCA) plots and volcano plots were made in R with the 'ggplot2' package (version 3.5.1) and heatmaps with 'pheatmap' package (version 1.0.12).

For miRNAs STAR (version 2.7.11a) was used to align sequencing reads to the GRCmm39 version of the mouse genome, with a 16bp minimum match length, allowing for a maximum of one mismatch. featureCounts (Liao et al., 2014) was used to quantify transcript expression. The annotations used were miRNA precursor sequences from miRbase and an 80% overlap between the read and precursor location was required. Targets of miRNAs were found with the dataset miRDB from Molecular Signatures Database (MSigDB) from GSEA (Chen and Wang, 2020, Liu and Wang, 2019). Analysis was performed in R Studio, using the script in section 7.2.2.

3.2.2 Pathway Enrichment Analysis

The R package GOseq (version 1.54.0) was used to identify overrepresented gene ontology (GO) terms amongst the DE gene and miRNA data sets (Young et al., 2010). GOseq uses a probability weighting function (PWF) based on gene lengths to calculate enrichment without being influenced by gene length bias. Gene targets for miRNAs were found with the miRDB dataset from GSEA. Graphs for GOseq were plotted with the ggplot2 package in R. Gene set enrichment analysis (GSEA) was run in R using the package 'fgsea' (version 1.28.0) and was used to determine miRNA binding sites that were enriched within the DEGs. The scripts for GOseq analyses are shown in sections 7.2.3 and 7.2.4 for mRNA and miRNA, respectively. For GSEA analyses the script in section 7.2.5 was used.

3.2.3 Binding analyses

MiRanda (Enright AJ, 2003) was used to predict miRNA target sites within *CASC20* and how many binding sites for these miRNAs were present in *CASC20*. The list of miRNAs predicted to bind to *CASC20* was compared to the list of miRNAs enriched with GSEA.

Sylamer was used to look for over and under-represented 6-mer, 7-mer and 8-mer miRNA seed sequences present in genes from RNA-seq (van Dongen et al., 2008), using the R script

in section 7.2.6. From an inputted ranked gene list of the most up- to most down-regulated 3' UTRs, it calculates if a sequence is significantly more or less common for each part of the list. It then plots the most enriched and depleted sequences using hypergeometric p-values for all six to eight base pair RNA sequences (van Dongen et al., 2008).

3.3 Results

3.3.1 mRNA-sequencing Analysis

To investigate the mechanism by which *CASC20* might affect the differentiation of MSCs, a PCA plot was created to visualise the patterns within the data by reducing it to two dimensions. This showed that the first component, and greatest variance, of 54% was between timepoints, i.e. genes up- or downregulated at D10 or 20 compared to D0 (figure 3.1A). Additionally, separate clustering was observed between the control and *CASC20*-expressing samples. This second component accounted for 18% of the variance on the PCA plot. The scaled heatmap showed similar overall patterns of expression between replicates, with some key differences between timepoints and conditions (figure 3.1B). The difference in expression between D0 and D10 in the OE samples is greater than in the CTRL, with greater expression of many genes occurring in OE D10. This shows the largest change in expression, as many genes had their lowest relative expression in OE D0 and then their highest relative expression in OE D10. There are also some genes in the bottom section of the heatmap for which the opposite is true, i.e. highest to lowest relative expression as differentiation progresses. The most significantly DEGs, when compared to D0, were plotted in a heatmap and had the greatest difference between D0 and D10 in the OE samples (figure 3.1C). OE D10 shows the clearest differences in expression from other conditions and timepoints. The top three genes are upregulated, whilst they are downregulated in CTRL samples and less strongly upregulated in OE D20 (figure 3.1C). The remaining genes for OE D10 are strongly downregulated, whilst showing lesser differential expression and even upregulation in other timepoints and conditions. The volcano plot shows the most significantly up- and downregulated genes according to the LFC and p-value (figure 3.1D). There were 139 genes that showed overall (i.e. at all time points) differential expression between CTRL and OE conditions. When looking at specific time points 10, 379 and 11 DEGs were identified for OE versus CTRL on D0, D10 and D20, respectively.

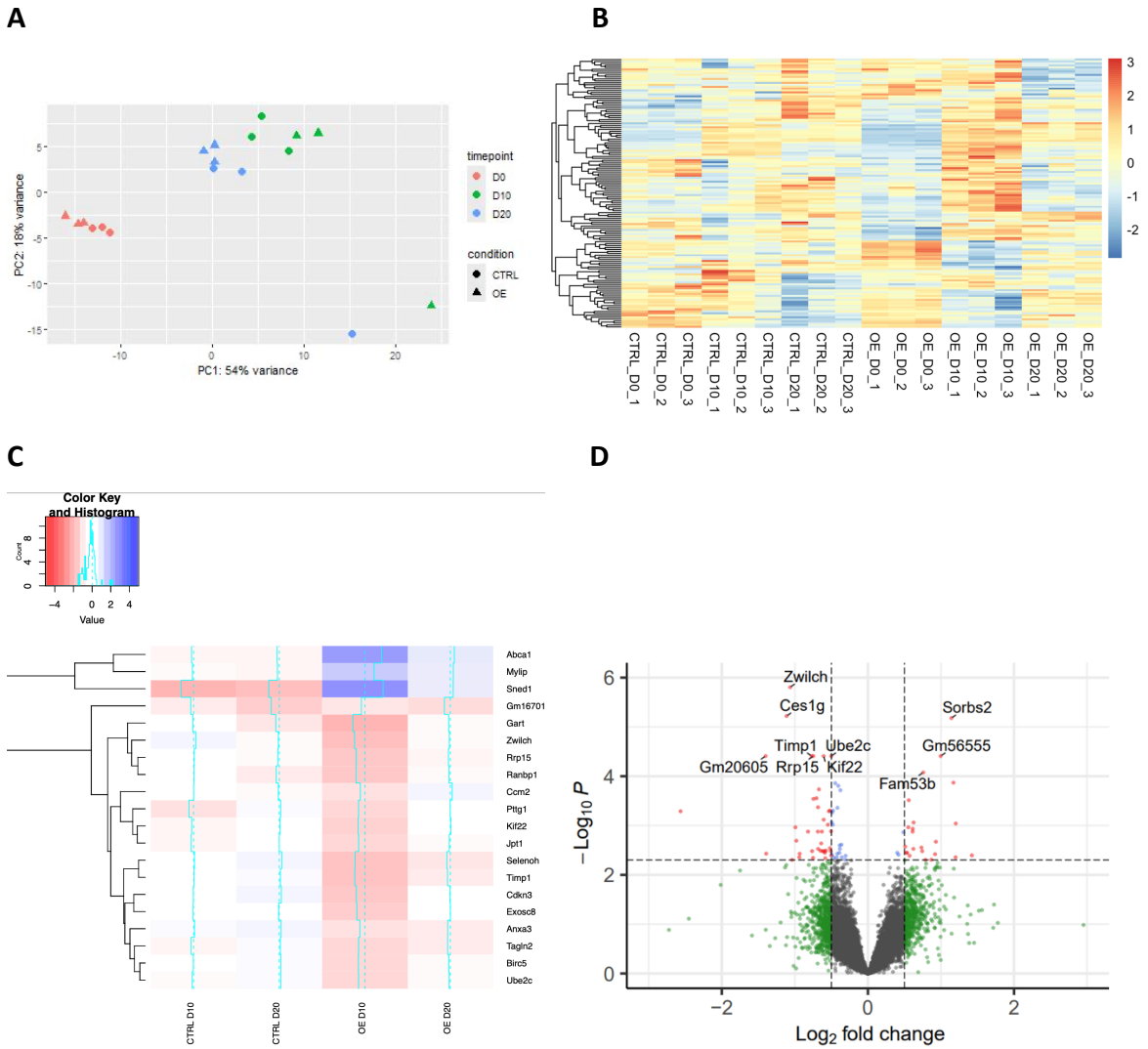


Figure 3.1: Differential expression analysis of mRNA throughout osteogenesis. (A) *PCA plot showing variance between samples. Each replicate is plotted as a separate point on the graph, with each timepoint represented with a different colour (day 0 = red, day 10 = green and day 20 = blue), whilst conditions are represented by different shapes.* ***(B)*** *Overall scaled heatmap of log2 Fold Changes observed in all genes, showing the relative expression of genes compared between each timepoint/condition. Variance stabilising transformation (VST) function was performed on data prior to plotting. The right-hand side legend shows the colour key for the log2 Fold Changes from -1 to +2.5, this is a relative number which relates to the relative expression of the gene between different samples, i.e. blue means the expression of the gene is lower in relation to its expression in other samples. Clustering of genes is displayed according to their Euclidean distance.* ***(C)*** *Heatmap of top 20 DEGs at each condition/timepoint. Each column represents a condition and timepoint when compared to D0. Clustering of these genes is shown according to their Euclidean distance.* ***(D)*** *Volcano plot of DEGs at D10, when CTRL compared to OE. The dotted lines show the cut offs for DEGs designated to be significant, with a LFC > +/- 0.5 and a p-value < 0.05. DEGs that reached significance are red and the top 10 most significant according to their p-value have been labelled.*

GO enrichment was performed to identify biological processes impacted in OE samples versus CTRL samples over time (figure 3.2). The categories identified in the set of overall differential genes seemed unrelated to differentiation and bone development (figure 3.2A). On day 10, the GO analysis suggested a role for *CASC20* in splicing (figure 3.2B). The GO categories 'U2 snRNP', 'U2-type spliceosomal complex' and 'spliceosomal complex' were among the top 10 most significant, with p-adjusted values of 1.3×10^{-4}, 5.7×10^{-6} and 5.7×10^{-6}, respectively. The percentage of hits in these categories were 33.3%, 31.0% and 10.8%, respectively, indicating these categories were the most upregulated with *CASC20* expression (figure 3.2B).

Overall, these analyses showed that there was greater variance between timepoints than between the control and *CASC20* expressing cells. However, variance between conditions was also observed and over 100 genes were DE, with the most DEGs observed at D10, where they were linked to a role in splicing.

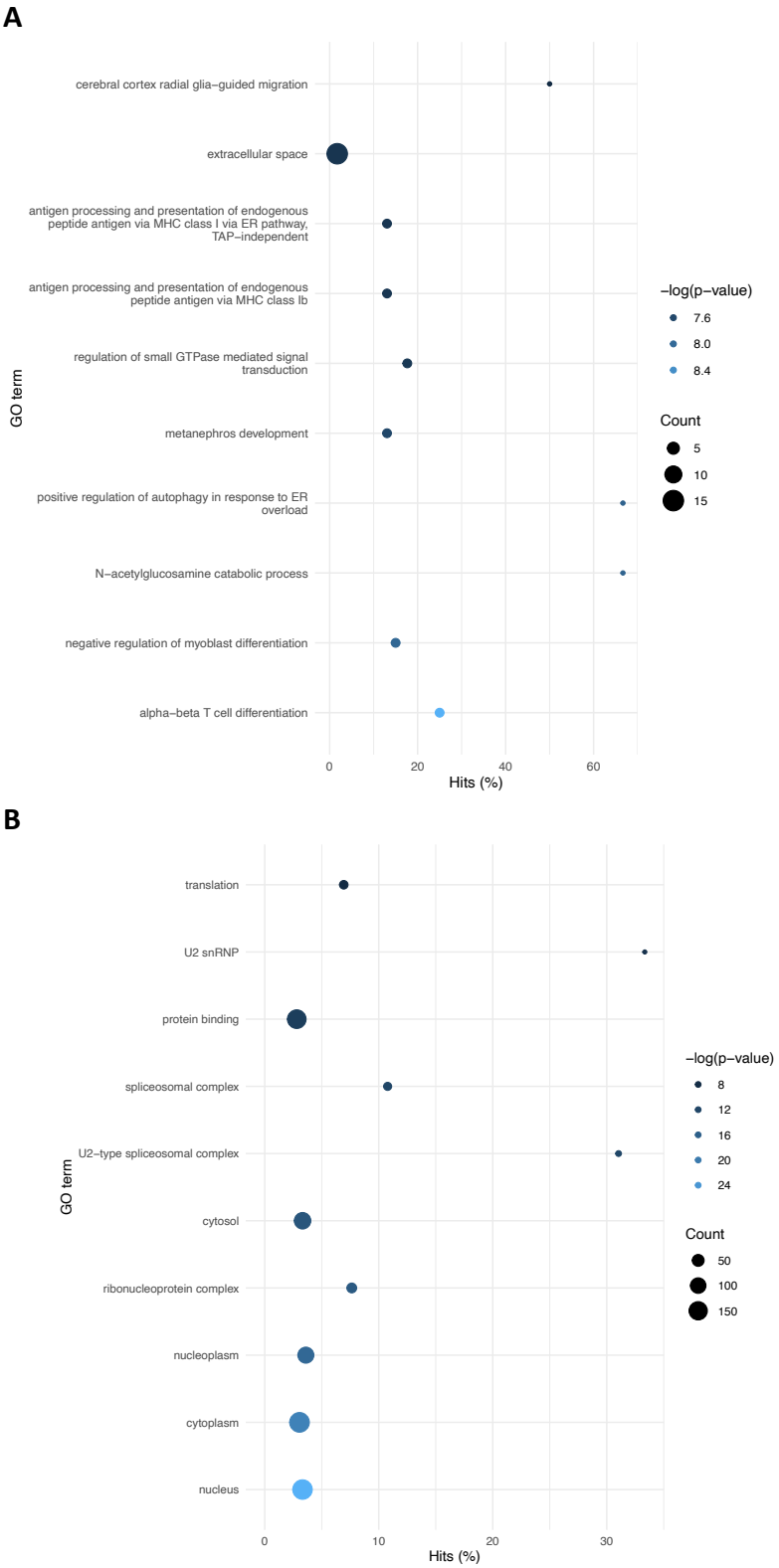


Figure 3.2: Gene Ontology (GO) terms enrichment analysis of differentially expressed genes (DEGs) in osteogenesis of CTRL vs OE samples. (A) *Overall CTRL vs OE and* ***(B)*** *CTRL vs OE at D10. GO terms are displayed in order of the most significant adjusted p-value (Benjamini-Hochberg method), with colours corresponding to negative log 10 of adjusted p value. The size of the dot corresponds to the number of DEGs in the sample that belong to a particular GO category and the x axis shows the percentage of DEGs in the GO category.*

3.3.2 miRNA Differential Expression Analysis

A similar analysis to above was performed for miRNA-seq data, in order to examine how miRNAs were affected by *CASC20* expression. This analysis aimed to identify miRNAs which could be involved in the mechanism of *CASC20*, however it may act. A PCA plot showed the greatest variance of 60% was between timepoints, with a lesser variance of 17% seen between conditions (figure 3.3A). The 50 most significantly DE miRNAs were plotted in a heatmap and the greatest difference was observed between CTRL D10 and OE D10 (figure 3.3B). Many of the miRNAs identified here were upregulated at D10 with overexpression of *CASC20*, with this reducing by D20. Several miRNAs were also downregulated here, again with lesser effects seen at D20 (figure 3.3B). Volcano plots were created for DE miRNAs overall and at D10, with the most significant p-values identified at D10 (figure 3.4). A number of miRNAs reached significance for both timepoints.

The differential expression of miRNAs overall between the conditions was studied and 8 miRNAs were upregulated, whilst 6 were downregulated (table 3.1). Of these 14 miRNAs identified, the gene targets of 7 were identified, providing 1522 gene targets in total. There was some overlap between the gene targets of these miRNAs, but 1290 of the genes were unique for an individual miRNA. 15 of these were also identified in the overall list of DEGs identified in the prior analysis (table 7.1, see appendix). For these gene targets, GOseq analysis identified several GO terms that were associated with development, including system development, anatomical structure morphogenesis, cell morphogenesis, developmental process, multicellular organism development and anatomical structure development (figure 3.5A). At day 10, 21 miRNAs were upregulated and 9 were downregulated (table 3.2). Of these 30 miRNAs, gene targets were available for 21 and 6292 gene targets for these were identified, of which 4289 were unique to an individual miRNA. 113 of these were also in the list of DEGs at D10 from the prior analysis, with the majority of these upregulated (table 7.1). Again, developmental processes were identified in the top GO categories, all of which were also in the top 10 GO categories overall (figure 3.5B). However, there was a much higher percent of hits (>30%) in each of the categories, as more of the genes in these categories were targeted by the DE miRNAs (figure 3.5B). The relationship between the developmental terms identified and more specific bone related

terms were investigated (figure 3.5C). Skeletal system development was a child terms of system development, which was identified within the top 10 GO terms, and bone development was a child term of this (figure 3.5C). Both skeletal system development and bone development were found to be significant in the GOseq analysis with 337 ($p=1.8 \times 10^{-12}$) and 142 ($p = 4.7 \times 10^{-5}$) genes identified, respectively (figure 3.5C).

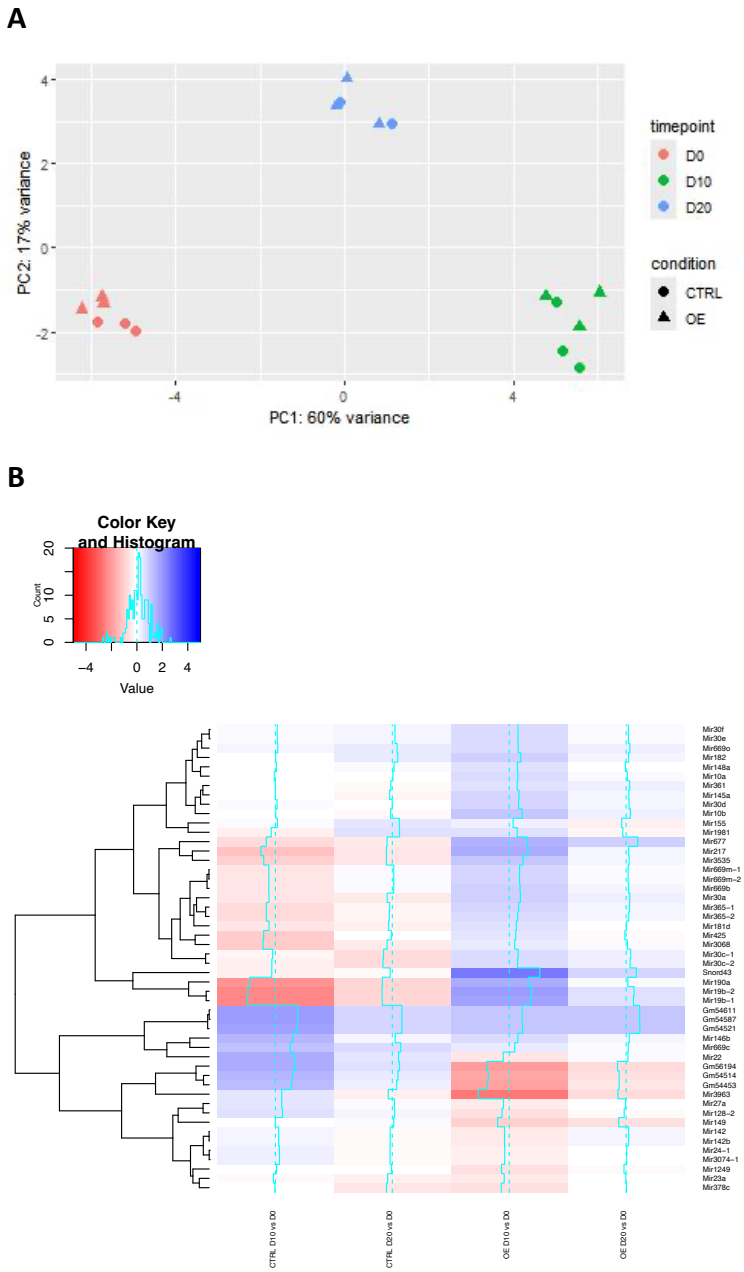


Figure 3.3: Differential expression analysis of miRNA throughout osteogenesis. (A) *PCA plot showing variance between samples. Each replicate is plotted as a separate point on the graph, with each timepoint represented with a different colour, whilst conditions are represented by different shapes.* ***(B)*** *Heatmap of top 50 DEGs at each condition/timepoint. Each column represents a condition and timepoint when compared to D0. Clustering of these genes is shown according to their Euclidean distance.*

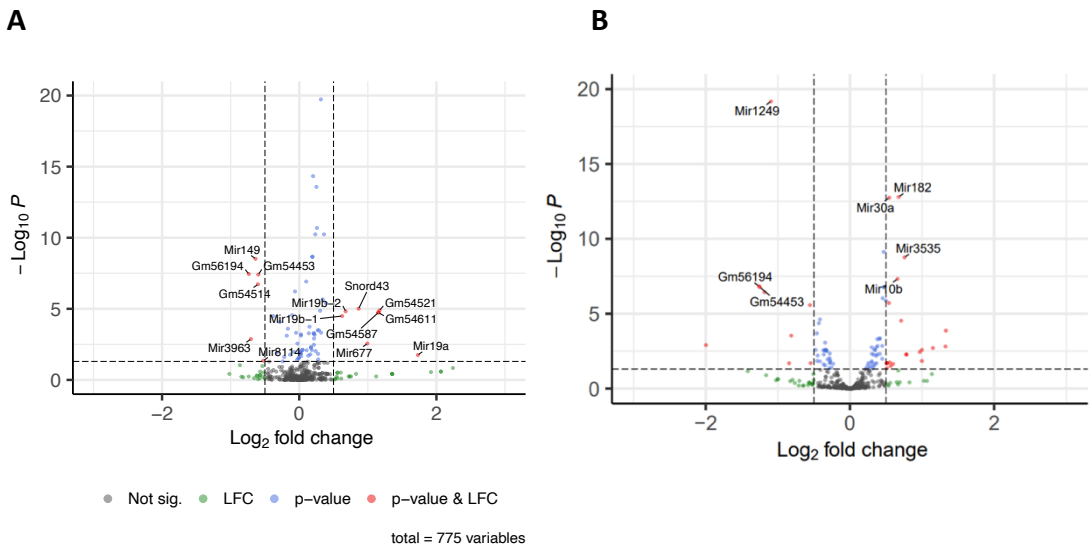


Figure 3.4: Volcano plots of differentially expressed miRNAs throughout osteogenesis. ***(A)*** *miRNAs DE overall and* ***(B)*** *at day 10. The dotted lines show the cut offs for DEGs designated to be significant, with a LFC > +/- 0.5 and a p-value < 0.05. DEGs that reached significance are red and the most significant according to their p-value have been labelled.*

Table 3.1: Differentially expressed miRNAs which reached significance overall from miRNA-seq analysis, when comparing CTRL vs OE. Significance was defined as a log2 fold change >0.5 or <-0.5 and a p-adjusted value of < 0.05. Upregulated miRNAs are highlighted in green and downregulated miRNAs in red.

MiRNA name	Ensembl gene ID	Base Mean	LFC	pvalue	p-adjusted
Mir-149	ENSMUSG00000065470	7437	-0.637	7.38E-11	3.03E-09
Gm56194	ENSMUSG00002075454	127.7	-0.735	9.67E-10	3.57E-08
Gm54453	ENSMUSG00002076138	133.4	-0.595	1.16E-09	3.90E-08
Gm54514	ENSMUSG00002076655	138.7	-0.602	6.65E-09	1.89E-07
Snord43	ENSMUSG00000105167	9274	0.866	4.48E-07	9.71E-06
Gm54521	ENSMUSG00002075625	269.9	1.171	6.25E-07	1.28E-05
Mir19b-2	ENSMUSG00000065473	131.1	0.677	8.18E-07	1.51E-05
Gm54587	ENSMUSG00002076451	272.7	1.154	1.09E-06	1.82E-05
Gm54611	ENSMUSG00002075983	272.7	1.154	1.09E-06	1.82E-05
Mir19b-1	ENSMUSG00000076256	125.2	0.626	2.08E-06	3.20E-05
Mir3963	ENSMUSG00000092830	16.20	-0.705	1.50E-04	1.35E-03
Mir677	ENSMUSG00000104755	63.46	0.994	3.16E-04	2.76E-03
Mir19a	ENSMUSG00000065416	20.79	1.730	2.94E-03	1.78E-02
Mir-8114	ENSMUSG00000099227	100.1	-0.523	8.29E-03	4.43E-02

Table 3.2: Differentially expressed miRNAs which reached significance at D10 from miRNA-seq analysis, when comparing CTRL vs OE. Significance was defined as a log2 fold change >0.5 or <-0.5 and a p-adjusted value of < 0.05. Upregulated miRNAs are highlighted in green and downregulated genes in red.

MiRNA name	Ensembl gene ID	Base Mean	LFC	pvalue	p-adjusted
Mir1249	ENSMUSG00000080441	5.209E+02	-1.095	1.671E-22	6.669E-20
Mir182	ENSMUSG00000076361	8.581E+04	0.677	8.040E-16	1.604E-13
Mir30a	ENSMUSG00000065405	2.207E+05	0.544	1.360E-15	1.808E-13
Mir3535	ENSMUSG00000104627	2.578E+03	0.757	2.180E-11	1.740E-09
Mir10b	ENSMUSG00000065500	3.610E+05	0.660	7.156E-10	4.758E-08
Gm56194	ENSMUSG00002075454	1.277E+02	-1.262	2.522E-09	1.437E-07
Gm54453	ENSMUSG00002076138	1.334E+02	-1.251	3.658E-09	1.670E-07
Gm54514	ENSMUSG00002076655	1.387E+02	-1.185	1.002E-08	3.633E-07
Mir148a	ENSMUSG00000065505	1.215E+06	0.540	6.735E-08	1.919E-06
Mir149	ENSMUSG00000065470	7.437E+03	-0.555	1.001E-07	2.663E-06
Mir1981	ENSMUSG00000088559	2.832E+02	0.709	1.254E-06	2.942E-05
Mir217	ENSMUSG00000065415	1.255E+02	1.332	6.406E-06	1.345E-04
Mir6236	ENSMUSG00000098973	1.791E+02	-0.814	1.549E-05	2.943E-04
Mir3963	ENSMUSG00000092830	1.620E+01	-1.997	8.173E-05	1.254E-03
Snord43	ENSMUSG00000105167	9.274E+03	1.326	1.119E-04	1.540E-03
Mir190a	ENSMUSG00000076379	6.255E+01	1.151	1.583E-04	1.974E-03
Mir19b-2	ENSMUSG00000065473	1.311E+02	0.998	2.427E-04	2.617E-03
Mir19b-1	ENSMUSG00000076256	1.252E+02	0.969	3.865E-04	3.586E-03
Mir7-1	ENSMUSG00000065434	6.858E+04	0.784	6.266E-04	5.240E-03
Mir7-2	ENSMUSG00000065609	6.671E+04	0.781	6.508E-04	5.252E-03
Mir704	ENSMUSG00000076060	4.403E+01	0.997	2.135E-03	1.444E-02
Mir295	ENSMUSG00000077886	1.430E+02	0.552	2.848E-03	1.804E-02
Mir5114	ENSMUSG00000093315	1.320E+02	-0.547	3.353E-03	1.985E-02
Mir669b	ENSMUSG00000076126	1.829E+02	0.516	3.344E-03	1.985E-02
Mir669m-1	ENSMUSG00000089570	1.892E+02	0.512	3.383E-03	1.985E-02
Mir669m-2	ENSMUSG00000088980	1.894E+02	0.508	3.584E-03	2.023E-02
Mir7680	ENSMUSG00000098400	4.555E+01	-0.844	3.666E-03	2.032E-02
Mir294	ENSMUSG00000077903	1.065E+02	0.596	4.260E-03	2.266E-02
Mir1947	ENSMUSG00000088860	9.725E+01	0.565	6.384E-03	3.184E-02
Mir466h	ENSMUSG00000077941	1.146E+02	0.512	9.110E-03	4.327E-02

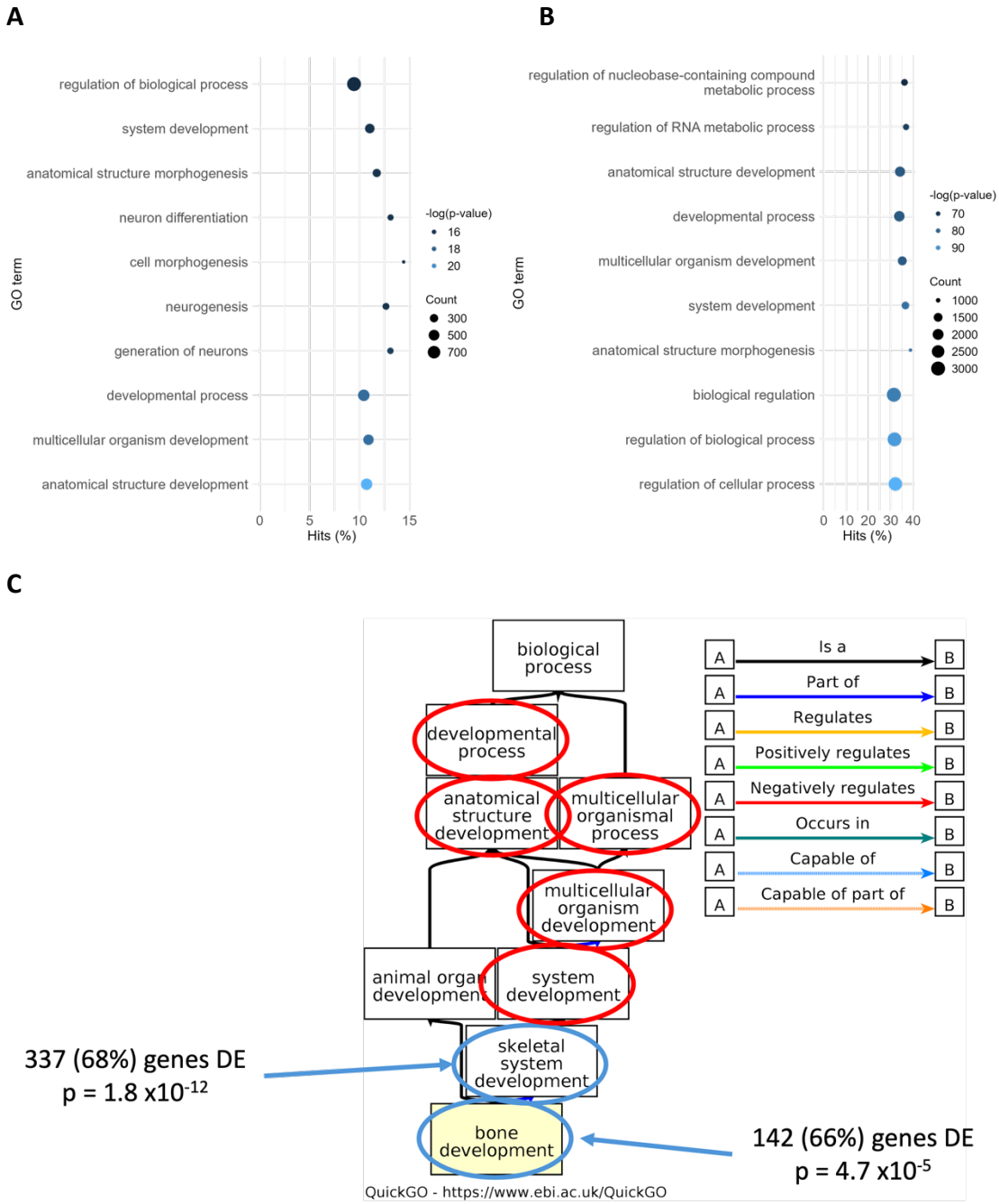


Figure 3.5: Gene Ontology (GO) terms enrichment analysis of differentially expressed miRNAs in osteogenesis of CTRL vs OE samples. (A) *Overall CTRL vs OE and* ***(B)*** *CTRL vs OE at D10. GO terms are displayed in order of the most significant adjusted p-value (Benjamini-Hochberg method), with colours corresponding to negative log 10 of adjusted p value. The size of the dot corresponds to the number of DEGs in the sample that belong to a particular GO category and the x axis shows the percentage of DEGs in the GO category.*

3.3.3 Enriched miRNAs in Differentially Expressed Genes

GSEA was performed to look at miRNA binding sites enriched in the DEGs identified from mRNA-seq. The aim of this analysis was to look for whether the gene targets of each miRNA were collectively enriched in up- or downregulated genes, to give an insight into the regulatory activity of the miRNAs.

Genes with target sites for 331 miRNAs were significantly enriched in the DEGs identified at day 10. When this list of miRNAs was compared with the list of DE miRNAs at day 10, nine miRNAs were identified in both, of which eight were upregulated (mi19b-1-5p, mir19b-2-5p, mir30a-3p, mir30a-5p, mir669b-5p, mir669b-3p, mir1947-3p, mir148-3p) and one was downregulated (mir7680-5p) in the miRNA-seq analysis (figure 3.6). These nine miRNAs were plotted to show how overrepresented the set of target genes for each miRNA were within the total dataset and where these genes appear in terms of up- or downregulation (figure 3.6A-I). These plots show mixed patterns but have a general pattern of upregulation of the gene targets (figure 3.6). Mir-30a-5p, mir-669b-3p and mir-669b-5p had targets in a large number of genes, but these were somewhat spread out across the x-axis, suggesting there is not a clearcut pattern of upregulation (figure 3.6D, F, G). For mir-19b-1-5p and mir-19b-2-5p only a small number of genes were identified, so whilst the peak of the graph is to the left suggesting upregulation, there are still some genes identified at higher ranks towards the right of the x-axis, so less confidence can be placed in these having clear upregulation (figure 3.6A, B). Mir-1947-3p produced the clearest pattern of upregulation with all the genes identified at low ranks on the left of the x-axis (figure 3.6H). Mir-7680-5p (the only downregulated miRNA) showed a clearer pattern of upregulation of enriched genes.

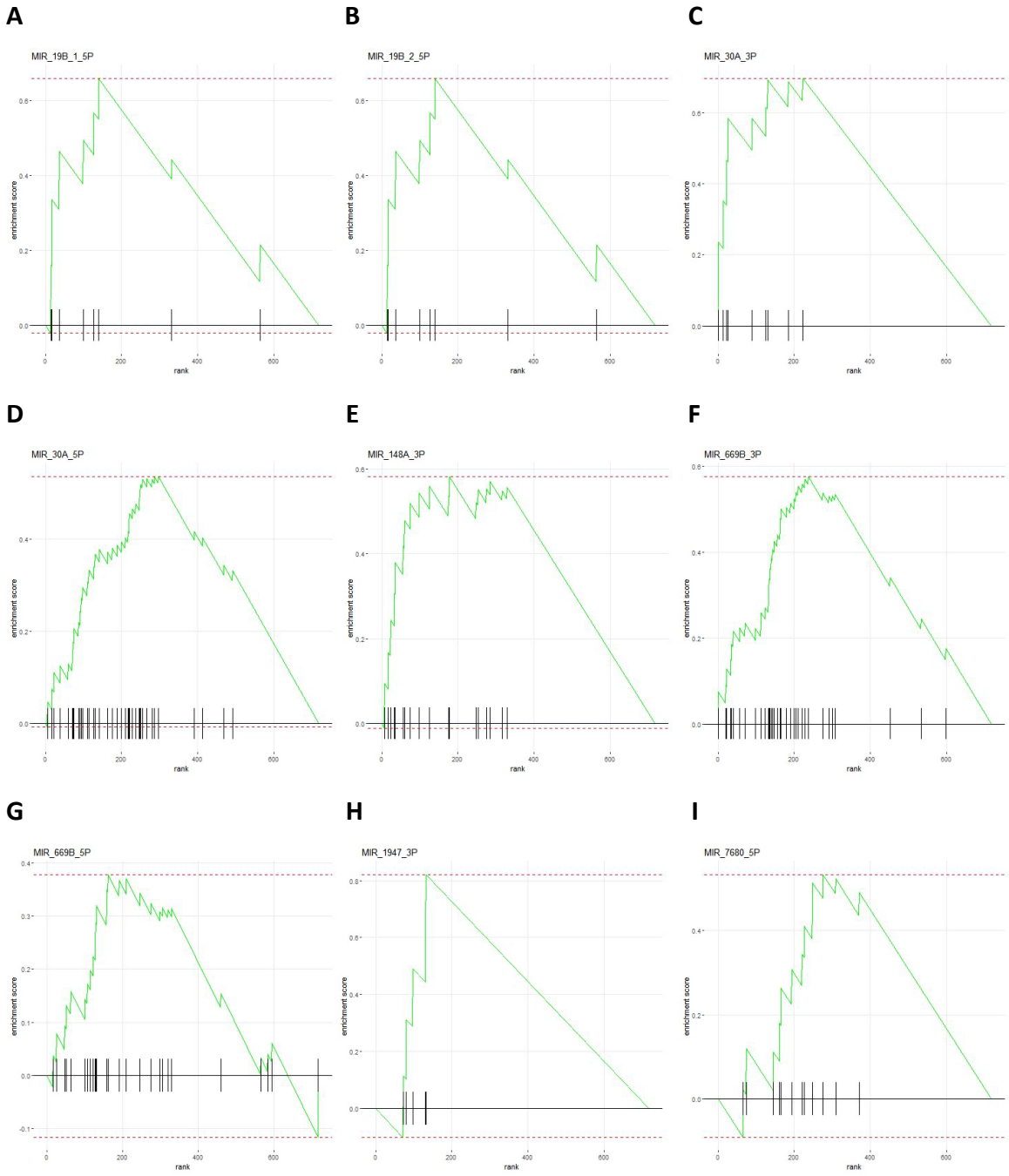


Figure 3.6: Gene set enrichment analysis (GSEA) of differentially expressed miRNAs during osteogenesis. (A) *mi19b-1-5p,* ***(B)*** *mir19b-2-5p,* ***(C)*** *mir30a-3p,* ***(D)*** *mir30a-5p,* ***(E)*** *mir148A-3p,* ***(F)*** *mir669b-3p,* ***(G)*** *mir669b-5p,* ***(H)*** *mir1947-3p and* ***(I)*** *mir7680-5p. The y-axis represents the enrichment score or how overrepresented this set of genes is within the total dataset. Along the x-axis are the genes that are significantly enriched for these miRNAs binding, where the left of the axis represents the most upregulated and the right of the axis the most downregulated.*

3.3.4 miRNAs targeting *CASC20*

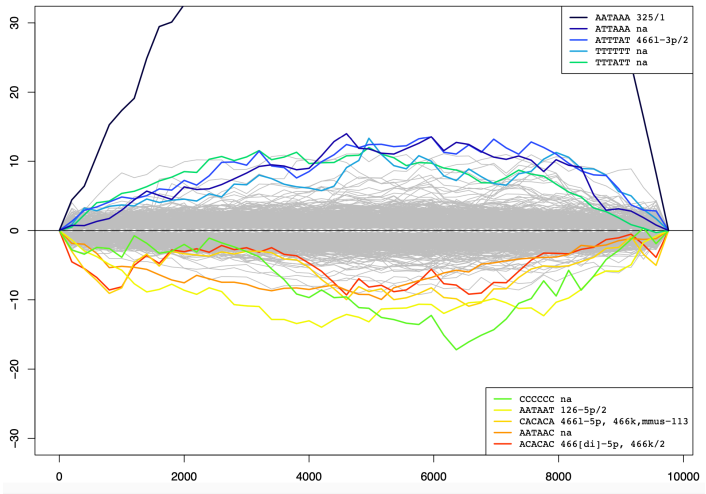
MiRanda was used to predict miRNAs that may have binding sites within *CASC20* and thus could have regulatory effects or be involved in the mechanisms. MiRanda identified 176 mouse miRNAs predicted to target *CASC20* (see appendix, table 7.1). Of these most had only one target site in *CASC20*, except mmu-miR-694 which had 3 predicted target sites, and 15 others with two sites. When this list was compared to the miRNAs enriched within the DEG list six miRNAs were identified: mir181d, mir190a, mir3074-2, mir33, mir466h and mir467d (table 3.3). None of these six miRNAs were ones that reached significance in the differential expression analysis (table 3.1, 3.2).

Table 3.3: miRNAs enriched within DEGs predicted to have a binding sequence in CASC20. *miRNAs identified by their name and corresponding ensemble gene ID. No. hits indicates the number of times the binding sequence of the miRNA is found within the CASC20 sequence.*

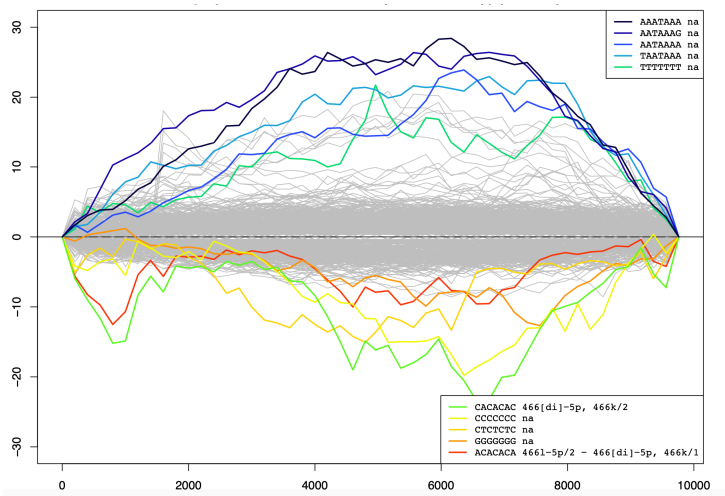
miRNA	No. Hits	Ensembl Gene ID	Base Mean	LFC	p-value	p-adjusted
Mir181d	1	ENSMUSG00000076338	2000.566	0.026	0.000	0.001
Mir190a	1	ENSMUSG00000076379	62.551	0.140	0.000	0.001
Mir3074-2	1	ENSMUSG00000105458	7265.324	0.024	0.002	0.014
Mir33	1	ENSMUSG00000065465	271.655	-0.242	0.009	0.050
Mir466h	1	ENSMUSG00000077941	114.597	0.133	0.003	0.020
Mir467d	1	ENSMUSG00000077021	545.525	0.277	0.007	0.036

Sylamer was used to predict the 6mer, 7mer and 8mer nucleotide sequences most over- and under-enriched in the DEGs, but most of these did not correspond to known miRNA seed sequences (figure 3.7A-C). Furthermore, many of the sequences identified were highly repetitive and very few of them were found within *CASC20* (table 3.4).

A



B



C

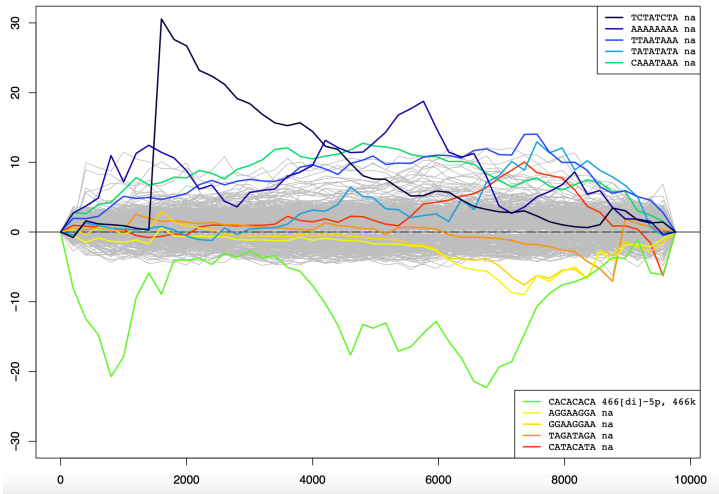


Figure 3.7: Sylamer outputs for n length sequences. (A) *6-mer sequences* ***(B)*** *7-mer sequences and* ***(C)*** *8-mer sequences. The x-axis represents the ranked gene list, and the y-axis shows log p-values, indicating enrichment or depletion of a particular motif.*

Table 3.4: Sequences detected by Sylamer and their presence in CASC20.

Sequence length	Seed sequence	Corresponding miRNA	Over- or Under-represented	No. times in *CASC20* sequence
6mer	AATAAA	325/1	Over	4
	ATTAAA	n/a	Over	5
	ATTTAT	4661-3p/2	Over	7
	TTTTTT	n/a	Over	8
	TTTATT	n/a	Over	4
	CCCCCC	n/a	Under	1
	AATAAT	126-5p/2	Under	5
	CACACA	4661-5p, 466k, mmus-113	Under	2
	AATAAC	n/a	Under	3
	ACACAC	466[di]-5p, 466k/2	Under	0
7mer	AAATAAA	n/a	Over	2
	AATAAAG	n/a	Over	0
	AATAAAA	n/a	Over	3
	TAATAAA	n/a	Over	1
	TTTTTTT	n/a	Over	4
	CACACAC	466[di]-5p, 466k/2	Under	0
	CCCCCCC	n/a	Under	0
	CTCTCTC	n/a	Under	0
	GGGGGGG	n/a	Under	0
	ACACACA	4661-5p/2 – 466[di]-5p, 466k/1	Under	0
8mer	TCTATCTA	n/a	Over	0
	AAAAAAAA	n/a	Over	2
	TTAATAAA	n/a	Over	1
	TATATATA	n/a	Over	0
	CAAATAAA	n/a	Over	0
	CACACACA	466[di]-5p, 466k	Under	0
	AGGAAGGA	n/a	Under	0
	GGAAGGAA	n/a	Under	0
	TAGATAGA	n/a	Under	0
	CATACATA	n/a	Under	0

3.4 Discussion

It was expected that timepoints would show the greatest levels of variance between samples for both mRNA- and miRNA-seq, due to the large numbers of genes that only become expressed at specific stages during differentiation in order to drive the process, regardless of *CASC20* expression (figure 3.1A, 3.3A). However, it was interesting to note there was also a visible difference between conditions, particularly in mRNA, showing that *CASC20* overexpression did cause a difference in gene expression during osteogenesis (figure 3.1A). This is supported by evidence that *CASC20* expression speeds up the rate of osteogenic differentiation, so genes critical to differentiation will be expressed at different

times between conditions (Felix-Ilemhenbhio, 2023). The two outliers, D20 CTRL and D10 OE, in the PCA plot could have been due to natural variation in gene expression between replicates, or a problem with sequencing affecting the quality of these samples (figure 3.1A). However, there was no visible difference in terms of the number of reads obtained for these samples, so there was no justification to exclude them based on the sequencing.

Whilst the overall scaled heatmap shows similar patterns of expression, there are areas where DEGs are seen between conditions and timepoints (figure 3.1B). The difference in OE D10 from the other samples and the fact that the patterns seen in D0 for OE and CTRL look more similar to each other than at other time points, then deviate more as differentiation begins, supports the idea that *CASC20* is affecting the rate of osteogenic differentiation. However, it is also possible that *CASC20* is driving the expression of genes unrelated to the differentiation processes. There are no published studies looking at the interactome of *CASC20* in MSCs or similar cells and whilst this analysis is interesting to get a general picture of expression patterns, it does not allow any specific genes or details to be studied. The heatmap of the most DEGs gave a more specific insight into the genes that are most significantly up- or downregulated (figure 3.1C). As OE D10 samples displayed the highest log2 fold changes in these genes, the combined effect of the OE condition and D10 time point compared to CTRL D0 produces the greatest effects. Whilst most of the top 20 genes are downregulated in OE D10, there are three that are upregulated: *Abca1*, *Mylip* and *Sned1* (figure 3.1C). These genes have roles in transporting ATP-binding cassettes (He et al., 2020), regulating low-density lipoprotein (LDL) receptor degradation (Lindholm et al., 2009) and mediating cell adhesion (Pally et al., 2025), respectively. The most downregulated genes include *Gart*, which encodes a trifunctional polypeptide required for de novo purine biosynthesis (NIH, 2024a) and *Selenoh*, which encodes a nucleolar oxidoreductase protein (NIH, 2024b) (figure 3.1C).

D10 showed the highest number of DEGs, compared to D0, possibly due to the enhanced rate of osteogenic differentiation caused by *CASC20*. By D20, it would be expected that cells under both conditions would have undergone osteogenic differentiation, and consistent with this, there were fewer genes showing differential expression. Some of these genes have roles in osteogenesis, with *ABCA1* forming part of a pathway with *GLP-1R,* which reduces

osteoblast apoptosis (Zhong et al., 2024). *PTTG1* has been found to be DE in OA and enriched in MAPK and Wnt signalling pathways (Wang et al., 2015). *TIMP1* has been shown to suppress growth and differentiation of osteoblasts by targeting the PI3K/AKT pathway (Xi et al., 2020). *EXOSC8* was identified as a key gene essential to osteogenic differentiation (Yang et al., 2019). *UBE2C* is critical to osteogenic differentiation of BMSCs, through stabilisation of SMAD1/5 and knock out of *Ube2c* impairs bone formation and regeneration (Zhang et al., 2024). Furthermore, *UBE2C* has also been shown to act via the PI3K/AKT pathway to target miR-140-3p and promote the progression of osteosarcoma (Huang et al., 2024).

However, GO enrichment analysis of these DEGs overall showed a variety of roles that appear unrelated to differentiation, such as those that are kidney related, and others that are very general, large categories, e.g. extracellular space. As there are such a large number of genes in categories like this, it is difficult to form any conclusions from the results (figure 3.2A). For the D10 comparison, the splicing related categories suggest this may be part of *CASC20*s mechanism or role, but there is overlap between the genes identified in the splicing related categories, i.e. many of the genes identified are the same between all of these categories (figure 3.2B). Furthermore, whilst a high percentage of hits were seen with some of these categories this is partly because there are only a very small number of genes in these GO categories.

When looking at the miRNA-seq, as with the mRNA-seq, the biggest difference being seen at D10 with *CASC20* overexpression supports *CASC20* playing a role in osteogenesis through modulation of miRNAs, as it is at D10 that the greatest difference would be expected between conditions (figure 3.3B).

It was a limitation that gene targets were only available for 7 of 14 miRNAs and for 21 of 30 miRNAs DE overall and at D10, respectively. If targets had been available for all miRNAs, then it may have influenced the top processes identified by GOseq. The presence of developmental related categories indicates the involvement in osteogenesis, as many of the terms identified within the top 10 were parent terms to bone development (figure 3.5). Furthermore, 142 mRNAs, targeted by the DE miRNAs, were from the bone development

category, which strongly supports *CASC20*s role here, as this forms 66% of the genes in this category (figure 3.5C). This suggests that as differentiation progresses *CASC20* may act on or through miRNAs that effect more genes involved in the developmental process.

The GSEA analysis primarily aimed to identify if any miRNAs were targeting many of the DEGs, as this could suggests they might be regulators driving the observed changes in gene expression. If the gene targets of a particular miRNA were significantly enriched in down-regulated DEGs then this could suggest the miRNA is functionally active and may suppress these genes. Conversely, if miRNA target genes are upregulated then it could indicate loss of miRNA function or ceRNA effects. The GSEA analysis for most of the identified miRNAs showed a general trend towards upregulation of gene targets, whilst the miRNAs themselves (except Mir-7680-5p) were also upregulated. Whilst miRNAs typically downregulate genes, there have been specific cases where they have been noted to lead to upregulation of genes. Furthermore, differential expression of miRNAs can result from a lncRNAs with a ceRNA mechanism in specialised cases where target-directed mRNA degradation (TDMD) has been observed (Hiers et al., 2024). Here, extensive base-pairing between miRNAs and “trigger” RNAs can cause destabilisation of the miRNA by destruction of its AGO protein, without repressing the trigger RNA (Hiers et al., 2024). However, this has only been observed in specific cases, for example Zinc Finger SWIM-Type Containing 8 (ZSWIM8), a ubiquitin ligase, causes polyubiquitination of AGO, resulting in its degradation and that of the accompanying miRNA (Hiers et al., 2024, Han et al., 2020). ZSWIM8-mediated degradation has currently been observed in 10 human miRNAs and over 70 murine miRNAs, with each of these likely to have their decay induced by a trigger (Hiers et al., 2024). The trigger RNAs for most of these are currently unknown, and whilst it has only been observed in limited cases, it is still an important mechanism to consider in analysis of these results (Hiers et al., 2024).

This could suggest the gene targets are not actually strongly regulated by the miRNA in these cells or that there may be other factors, such as transcription factors with a greater regulatory effect. Mir-7680-5p (the only downregulated miRNA) showed upregulation of enriched genes, suggesting it could act as a regulator for the expression of these genes, and thus its downregulation increases the levels of expression in these genes (figure 3.6I).

From the MiRanda analysis of miRNAs that target CASC20, the six miRNAs identified that were also in the DEG list were of interest, but as they were not significant in the differential expression analysis it suggests they are not relevant to the action of *CASC20* (table 3.1). However, they are worth noting as if they were identified in RNA-seq following the overexpression of *CASC20* in human cells then it may be valuable to investigate their interactions with *CASC20* further. The fact each of these miRNAs only had one site in *CASC20* suggests a ceRNA mechanism is unlikely, as they would need multiple sites in order for *CASC20* to sponge them effectively (Kartha and Subramanian, 2014). Although, it's possible they might not have a direct relationship and instead sponge or otherwise interact with other mRNA/miRNAs that in turn interact with *CASC20*. miRNAs can regulate hundreds of mRNA targets and multiple miRNAs can act on the same mRNA (Kartha and Subramanian, 2014). In future it would be useful to expand this analysis to look at other databases. Whilst MiRanda uses an algorithm to predict miRNAs binding, there are databases of experimentally validated miRNA binding that could be used, such as miRTarBase and starBase (Kariuki et al., 2023). In this case it was more complicated as a human-only gene was being investigated in mouse cells, rather than looking at the human miRNAs. *CASC20* being a ncRNA also limits the databases that can be used, as some such as Target Scan do not predict against ncRNAs (McGeary et al., 2019). Furthermore, it was a limitation that this was looking at mouse miRNAs, as although many are conserved there are still differences between species and it's possible that there are other interacting miRNAs in humans that were not seen in this analysis in murine MSCs.

The Sylamer analysis did not identify many miRNAs of note (figure 3.7). Most of the sequences were likely present due to their repetitive nature and low sequence complexity, as A and/or T strings are common in mRNA and most were not the seed sequences for any miRNAs. Some of these were found within *CASC20*, but only because of their highly repetitive nature and the fact that a 6 base sequence is highly likely to be present in a large gene by random chance. AATAAA was the top enriched 6mer sequence, and although this did correspond to an miRNA seed mir-325-1, it is also the poly-A possessing sequence and found at the 3' ends of mRNA, so it's unlikely this miRNA has any role with *CASC20* (Proudfoot, 2011). Mir-466 was identified multiple times but as the seed sequence for this is a series of repeating 'CACA' it is likely just a repetitive sequence found by chance. mir-4661

was more likely to be a genuine finding as both the 3p and 5p forms of these were identified, which have different seed sequences. The 3p form was enriched whilst the 5p was depleted. There is very little existing literature for miR-4661, but interestingly in humans it has been shown to promote osteogenic differentiation in bone marrow MSCs by suppressing the expression of fibroblast growth factor 23 (FGF23) (Zhang and Xu, 2024). High expression of miR-4661-3p led to increased expression of RUNX2, Col1, osteocalcin, osterix and dentin matrix protein 1 (DMP1), plus enhanced alkaline phosphatase (ALP) activity and calcium deposition (Zhang and Xu, 2024).

Chapter 4: Differentiation of hMSCs following lentiviral transduction

4.1 Introduction

The aim of this experiment was to isolate hMSCs from SVF taken from patient knee fat pad samples and overexpress *CASC20* via lentiviral transduction. Following this osteogenic or chondrogenic differentiation was induced in these cells. This method was chosen as it had previously been used to successfully transduce and differentiate other cell lines and semi-immortalised cells into osteoblasts and chondrocytes, whilst overexpressing *CASC20* (Felix-Ilemhenbhio, 2023). The aim was to perform successful differentiation of these hMSCs, then compare this to the RNA-seq analysis of murine MSCs described in chapter 3.

4.2 Methods

4.2.1 Isolation of hMSCs

Adipose-derived human hMSCs were isolated from the SVF of knee fat pads, donated by patients undergoing knee replacement surgery. These samples were collected as part of an ethically approved musculoskeletal biobank (REC reference number 15/SC/0132). An overview of the whole process is shown in figure 4.1. Upon receipt the tissue was homogenised with a scalpel and tweezers to maximise the surface area for digestion, then transferred to a 50 mL falcon tube. The tube was inverted several times and the amount of floating tissue was recorded to calculate the volume of collagenase required for digestion. The homogenised tissue was centrifuged in 50 mL phosphate buffered saline (PBS) for 8 minutes at 75 g, at room temperature. The floating fat tissue was transferred to a clean falcon tube and the PBS wash and centrifugation repeated. A 0.015% collagenase I (Gibco, Thermo Fisher Scientific, 17018029, USA) solution was prepared according to the volume of fat tissue recorded earlier (0.015 mg per 1 g of tissue), diluted in Hank's Balanced Salt Solution (HBSS) (Gibco, Thermo Fisher Scientific, 14025092, USA) and vortexed to mix. The collagenase solution was sterile filtered through a 0.45 µM syringe filter, then transferred to tubes (Miltenyi Biotec, 130-093-237, Germany) with up to 10 mL in each. The fat tissue was distributed between the Miltenyi tubes and incubated at 37° on a shaker at 180 rpm for 60 minutes. At the end of the incubation the Miltenyi tubes were placed into a Miltenyi gentleMACS dissociator machine to aid digestion of the tissue mechanically and the adipose

tissue programme (30 seconds of homogenisation) was run twice to aid tissue homogenisation. The digested tissue was then filtered through a 40 µM cell strainer into clean falcon tubes. 5 mL of Dulbecco's Modified Eagle Medium (DMEM) was added, then made up to 50 mL with PBS and centrifuged for 10 minutes at 300 xg, room temperature. The supernatant was discarded and the pellet resuspended in 50 mL PBS, then the centrifugation repeated. The supernatant was discarded and the cell pellet was resuspended in Dulbecco's Modified Eagle Medium (DMEM) + GlutaMAX (Gibco, 61965059, USA), with 15% fetal bovine serum (FBS) (Gibco, 1600044, USA) and 0.8% Penicillin-Streptomycin (P-S) (Gibco, 1570063, USA). The cell number was counted with 0.2% Trypan Blue solution and seeded into appropriate flasks, based on their number.

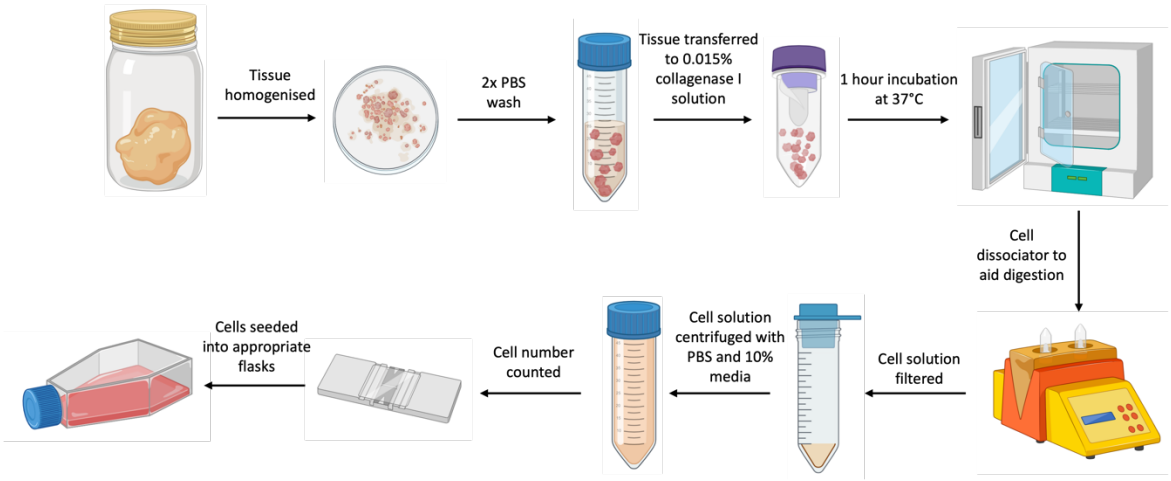


Figure 4.1: Methodology for isolation of human mesenchymal stromal cells (hMSCs) from stromal vascular fraction (SVF). *Flow chart describing the process from receiving the tissue sample, to seeding of the isolated hMSCs into flasks for culture. Created with BioRender.com.*

4.2.2 Culturing of hMSCs

hMSCs isolated from SVF were cultured in DMEM + GlutaMAX with 15% FBS and 0.8% P-S, and the media changed every 48 – 72 hours, until confluency. Cells were passaged at 70 – 80% confluency with a solution of 0.1% collagenase IV (Gibco, 17104019, USA) prepared in HBSS. The collagenase solution was sterile filtered through a 0.45 µM filter, as before. The media was removed from the cells, washed with PBS, then the collagenase solution added. The cells were incubated at 37° on a shaker for at least 10 minutes, to allow them to detach.

The digestion was stopped with a 10% volume of warm media, then the solution centrifuged for 5 minutes at 300 xg. The supernatant was removed and then the cell pellet resuspended in 1mL warm media. The cells were counted as before and seeded into appropriate flasks, or frozen. For freezing cells, the cell pellet was suspended in FBS with 10% dimethyl sulfoxide (DMSO) (Sigma, D4540, USA), then placed at -80° in a cryogenic freezing container overnight, before transfer to the liquid nitrogen. To thaw the cells, they were added to 1 mL of DMEM + GlutaMAX, centrifuged to form a pellet, then counted and seeded as previously described.

4.2.3 Lentiviral Transduction

Prior to lentiviral transduction cells were passaged as previously described and seeded into flasks or plates as appropriate. Cells were incubated for 24 hours in DMEM + Glutamax, with 15% FBS and 0.8% P-S. After 24 hours the media was replaced with fresh media containing 8 µg/mL (1x stock solution) of Polybrene (Sigma, TR-1003, USA), a commonly used media additive to increase transduction efficiency (Denning et al., 2013). The appropriate lentivirus, either *CASC20* overexpression (OE) or GFP control (CTRL), was added according to the cell number. Lentivirus was added at a concentration of 1µL per 2000 cells (MOI = ~150,000 viral particles/cell). The flask was mixed for at least 10 seconds by movement from side to side to ensure an even spread of lentivirus. The cells were then incubated for 24 hours to allow them to recover from transduction (ideally this allowed them to reach ~70% confluency) before beginning the appropriate differentiation protocol. For the CTRL lentivirus the expression of GFP was checked under a microscope (Bio-Rad, ZOE Fluorescent Cell Imager, USA) to confirm successful transduction. Images taken at 20x magnification unless otherwise specified.

4.2.4 Osteogenic Differentiation

OM was prepared as follows: DMEM with 0.8% P-S, 10mM β-glycerophosphate (Sigma-Aldrich, G9422, USA) in sterile water, 0.01 µM Dexamethasone (Sigma-Aldrich, D8893, USA) in PBS, 50 ng/mL L-Ascorbic acid 2-phosphate (AA2P) (Sigma-Aldrich, A4544, USA) in sterile water and 300 ng/mL BMP2 (GenScript, Z02913, USA). β-glycerophosphate, Dexamethasone and AA2P were prepared fresh once per month (or for each differentiation). Cells were passaged as described previously then seeded into appropriate wells. Unless otherwise

specified this was 62,800 cells per well in a 48 well plate (57,000 cells per cm^2). A plate was prepared for each time point, generally day 0 (D0), day 10 (D10) and day 20 (D20), with n=3 samples for each condition. The cells were then incubated for 24 hours, or until confluent. On the day the differentiation begun the D0 plate was fixed. To fix the cells, media was removed and each well was washed three times with PBS, then 4% paraformaldehyde (PFA) (Sigma-Aldrich, 158127, USA) added, incubated for 30 minutes at room temperature, then removed and the three PBS washes repeated. PBS was added to each well for storage at 4°C. For the remaining plates, the media was replaced with OM. Media was changed every 48 – 72 hours, with fresh OM prepared once per week. Upon reaching the other timepoints the plates were fixed for staining as described for D0.

4.2.5 Chondrogenic Differentiation

Chondrogenic medium was prepared as follows: DMEM with l-glutamine (Gibco, Thermo Fisher Scientific, 11995073, USA) and 5% P-S (no FBS), 100 nM Dexamethasone, 10 ng/mL TGFβ3, 50 µg/mL AA2P 40 ng/mL L-proline and 1x concentration of ITS+L premix (insulin, transferrin, selenous acid, linoleic acid) (Gibco, Thermo Fisher Scientific, 41400045, USA). Dexamethasone, AA2P and L-proline were prepared fresh for each differentiation. Cells were passaged as described previously then 50,000 cells/well were seeded into UV-irradiated 96 V well microplates (Greiner, 651101, UK). A plate was prepared for each time point, day 7 (D7) and day 14 (D14) unless otherwise specified. For D0 samples, cells were transferred to 1.5 mL tubes and spun for five minutes at 300 xg, then the supernatant was removed and pellets were stored on ice immediately before transfer to -80°C for storage. To seed cells for other time points, tubes were centrifuged and the pellet resuspended in the appropriate volume of chondrogenic media. 150 µL of cell suspension was added to each well. Plates were centrifuged for five minutes at 500 xg then stored in the incubator at 37°. Media was changed every 72 - 96 hours, and fresh TGFβ3, AA2P and L-proline were added to the media stock once per week, as they are subject to degradation. Upon reaching the other timepoints, the media was removed from the pellets and the plates stored at -80°C.

4.2.6 Assay for calcium deposition

The assay for calcium deposition as a marker of bone formation by mineralisation was performed by ARS. PBS was removed from each well and 500 µL of 40 mM Alizarin Red S

(Sigma-Aldrich, A5533, USA) was added. Alizarin Red S was made up in dH_2O to pH 4.2(+/-0.1). Plates were left for 30 minutes on an orbital shaker at room temperature before removing the dye from the well. Each well was washed at least 5x with PBS by dispensing and immediately aspirating, to remove unbound stain, then the plates were left to dry before imaging on a microscope (Leica, CTR4000, Germany) at 20x magnification, unless otherwise specified.

4.2.7 Assay for GAG release

The assay for GAG release as a marker of chondrogenic differentiation was carried out with the DMMB assay. GAGs are a class of polysaccharide critical in chondrocyte differentiation and the regulation of endochondral ossification, as they form an essential component of the cartilage extracellular matrix (Chen et al., 2023b). Thus, their release is a common marker for measuring successful chondrogenic differentiation. For cell pellets, cartilage digestion was required prior to performing the DMMB Assay. Phosphate buffer (0.1 M, pH 6.5) was prepared by combining 0.1 M NaH_2PO_4 (Sigma-Aldrich, S3139, USA) with 0.1 M NaH_2PO_4 (Sigma- Aldrich, 567547) in a ratio of 137:63, respectively. Papain solution was prepared by combining 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. 70 µL of papain solution and 40 µL of phosphate buffer were added to each well, and then transferred with the pellet to an Eppendorf tube. Tubes were briefly vortexed and centrifuged, then incubated for 4 hours at 65°C, with a brief vortex and centrifuge once per hour. For the DMMB assay on digested pellets or media, DMMB solution was prepared with 2.6g/mL DMMB (Sigma-Aldrich, 341088), 3.4g/mL NaCl, and 9.5mM HCl in dH2O, then made up to 1 L with dH_2O to pH 3. The solution was kept in the dark and prepared fresh at least every 3 months. Chondroitin Sulphate was used to prepare a standard curve between 0 and 40 µg/ml, by combining 1 mg/ml Chondroitin sulphate with phosphate buffer or chondrogenic media, when testing cell pellets or media, respectively. 25 µL of standards or samples were added to each well, followed by 200 µL DMMB and mixed by pipetting up and down several times. The plate was then read on a Varioskan Flash plate reader (Thermo Fisher Scientific, USA) at 530 nm.

4.2.8 RNA Extraction

RNA extraction was carried out with the ReliaPrep RNA Miniprep System (Promega, Z6011, USA), as per the standard protocol. RNA concentration was measured by the optical density reading on a NanoPhotometer (Geneflow, N60, UK).

4.2.9 RT-qPCR

Primers were designed for reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) when required with the NCBI Primer Blast tool (NCBI) and selected based upon those that create a product <100 base pairs (bp) in length, have a length of approximately 20bp, very close melting temperatures and GC contents between 45-55% (table 4.1).

Table 4.1: qPCR Primers. *All primers ordered from Sigma as custom qPCR probes.*

Target Gene	Forward (FW)/ Reverse (RV)	qPCR Primer Sequence (5' -> 3')
Lentivirus RRE	FW	AATGACGCTGACGGTACAGG
Lentivirus RRE	RV	GCCTCAATAGCCCTCAGCAA
Lentivirus RRE	FW	GACGGTACAGGCCAGACAAT
Lentivirus RRE	RV	AGATGCTGTTGCGCCTCAAT
eGFP	FW	AGGACGACGGCAACTACAAG
eGFP	RV	AAGTCGATGCCCTTCAGCTC
eGFP	FW	AAGGACGACGGCAACTACAA
eGFP	RV	TCCTTGAAGTCGATGCCCTT

For copy (c)DNA synthesis the volume of RNA loaded was calculated by: $V_{RNA} = m_{cDNA} / \rho_{RNA}$ and the iScript cDNA Synthesis Kit (BioRad, 1708890, USA) was used, according to the standard reaction protocol (Table 4.2). RNA was added to 5 µL of mastermix (4 µL iScript Reaction mix and 1 µL iScript Enzyme) and the remaining volume made up to 20 µL with nuclease free water. cDNA synthesis was run on the GeneAmp PCR System 9700 thermocycler (Thermo Fisher Scientific, USA).

Table 4.2: cDNA synthesis protocol for use with iScript cDNA Synthesis Kit.

Step	Time & Temperature
Priming	5 min at 25°C
Reverse Transcription (RT)	20 min at 46°C
RT inactivation	1 min at 95°C
Optional Step	Hold at 4°C

For the quantitative polymerase chain reaction (qPCR) 1ng of each sample was added to each well. Per sample a mastermix of 5 µL SYBR mix (Bio-Rad, 1725150, USA), 0.3 µL 10 µM forward and 0.3 µL 10 µM reverse primers was prepared, alongside a mix of 1ng cDNA and made up to 5 µL with nuclease free water. These were added to a plate to a total volume of 10.6 µL and a standard cycling protocol was run 40x on QuantStudio 5 PCR machine (table 4.3).

Table 4.3: Protocol for qPCR.

Step	Time & Temperature
Hold	2 mins at 95°C, rate 1.6°C/s
PCR: step 1	10 seconds at 95°C, rate 1.6°C/s
PCR: step 2	1 min at 60°C, rate 1.6°C/s, capture image
Melt Curve: step 1	5 seconds at 60°C, rate 0.5°C/s
Melt Curve: step 2 (dissociation)	5 seconds at 95°C, rate 0.5°C/s, capture image

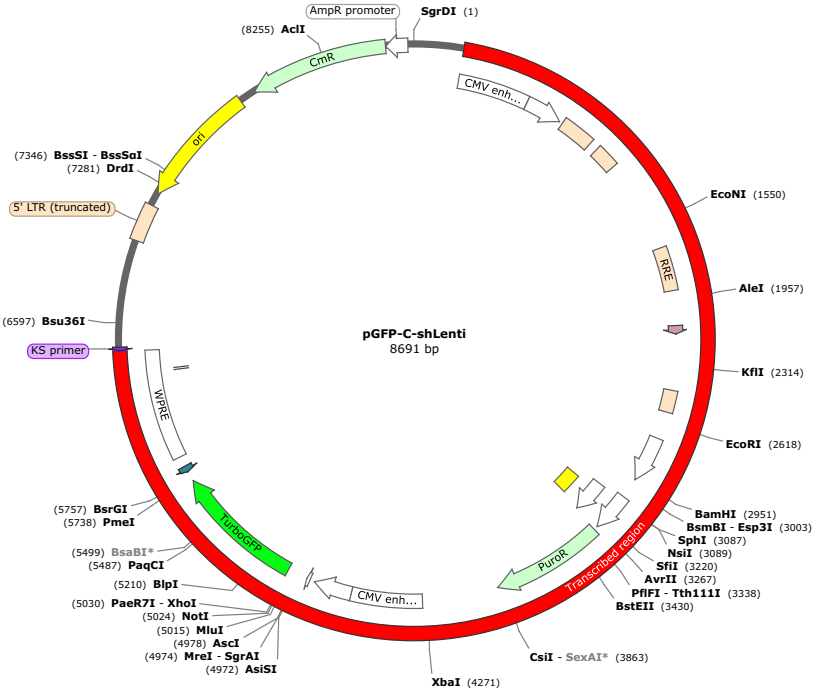
4.2.10 Statistical analyses

For statistical analysis GraphPad Prism 9 (Dotmatics, Boston, USA) was used to present and analyse quantitative data. Data presented as mean of replicates with standard deviation (SD), where applicable. One-sample t-test was used to compare the mean of single samples to a known mean, independent t-tests to compare the means of two independent groups and paired two-tailed t-tests when two related groups (such as OE vs CTRL) were compared. These methods were used to assess significance ($p < 0.05$), as appropriate for each data set.

4.2.11 Optimisation of lentivirus MOI

qPCR was performed to determine the multiplicity of infection (MOI) of both the lentiviruses (CTRL and OE) used to transduce hMSCs. This qPCR was also used to compare the new aliquots of lentiviruses used in the differentiations described in section 4.3. to old stocks used by a previous researcher to successfully transduce other hMSC cell types. This was achieved by first carrying out an RNA extraction of the lentivirus stocks, then performing cDNA synthesis to obtain a final DNA concentration of 5 ng/µL. The plasmid vectors used to produce the lentiviruses (figure 4.2) were used to prepare a dilution series that was run as part of the qPCR to create a standard curve. Lentiviral concentrations were interpolated from the standard curve using GraphPad Prism 9. Then the number of molecules corresponding to these concentrations was calculated with Avogadro's constant to create equivalent standard curves showing the number of plasmid molecules and interpolating the number of lentiviral particles in the respective qPCR reactions (figure 4.9). The MOI of both old and new lentiviruses was calculated from the number of lentiviral particles (table 4.3).

A



B

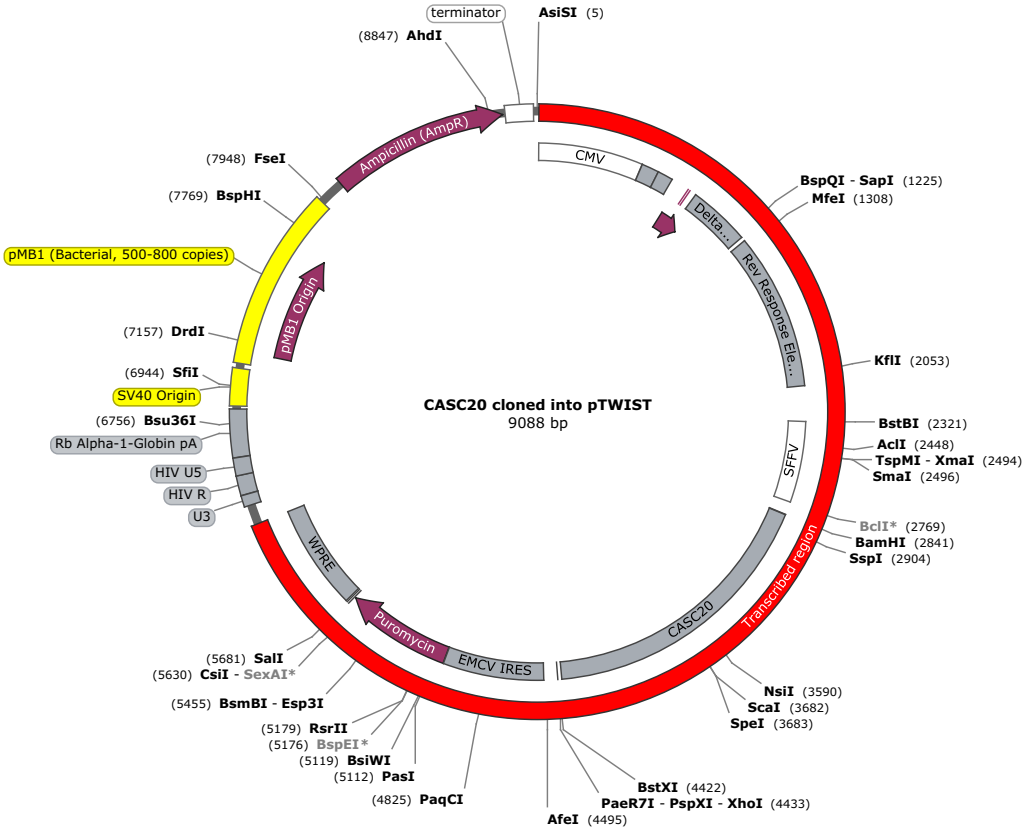


Figure 4.2: Plasmid maps for lentiviral vectors. (A) pGFP-C-shLenti *(CTRL lentivirus) and* ***(B) pTWIST with CASC20 insert*** *(OE lentivirus). Plasmid maps created in Snapgene. The transcribed region that is integrated into the lentivirus is shown in red.*

Table 4.3 Steps of calculation for MOIs of lentiviruses. *This describes the steps and exact calculations made for each lentivirus, old and new.*

Row no.	Step	*CASC20* OE - Old	*CASC20* OE - New	GFP CTRL - Old	GFP CTRL - New	Calculation made
1	Viral particle no. in 1 in 25 qPCR reaction	1.24 x10^{7}	*1.24 x10^{7}*	1.13 x10^{7}	1.13 x10^{7}	n/a
2	Viral particle no. in undiluted qPCR reaction (2µl cDNA)	3.10 x10^{8}	*3.10 x10^{8}*	2.83 x10^{8}	2.83 x10^{8}	Row 1 x 25
3	Viral particle no. in total cDNA reaction (20µl)	3.10 x10^{9}	*3.10 x10^{9}*	2.83 x10^{9}	2.83 x10^{9}	Row 2 x 10
4	Volume of RNA added to cDNA reaction (µl)	15	11	15	9	n/a
5	Viral particle no. in extracted RNA vol (30µl)	6.20 x10^{9}	*8.45 x10^{9}*	5.85 x10^{9}	9.11 x10^{9}	Row 4 x 30 / row 3
6	Viral particle no. in RNA vol used for extraction (60µl)	1.24 x10^{10}	*1.69 x10^{10}*	1.17 x10^{10}	1.82 x10^{10}	Row 5 x 2
7	Viral particle no. per µl of RNA	2.07 x10^{8}	*2.82 x10^{8}*	1.95 x10^{8}	3.04 x10^{8}	Row 6 / 60
8	MOI if 200µl lentivirus added to 400,000 cells	1.03 x10^{5}	*1.41 x10^{5}*	9.74 x10^{4}	1.51 x10^{5}	(200 x row 7) / 400,000

4.3 Results

4.3.1 Osteogenesis of hMSCs

A pilot differentiation experiment with four pooled hMSC isolates, termed ‘hMSC mix-1’, was performed. The isolates were pooled in an attempt to reduce donor variability and to provide sufficient overall cell number for the experimental design. The isolates were cultured individually and pooled at the start of passage (P)3, then transduced with either CTRL or OE lentivirus. Successful transduction was confirmed with microscopy to visualise GFP expression in the control cells (figure 4.3). The cells were induced to begin differentiation at the start of P4. After differentiation ARS was performed to visualise calcium deposits, however little to no staining was observed, with only very small amounts of staining seen at D20 (figure 4.4). Overall, this experiment showed no evidence of differentiation into osteoblasts.

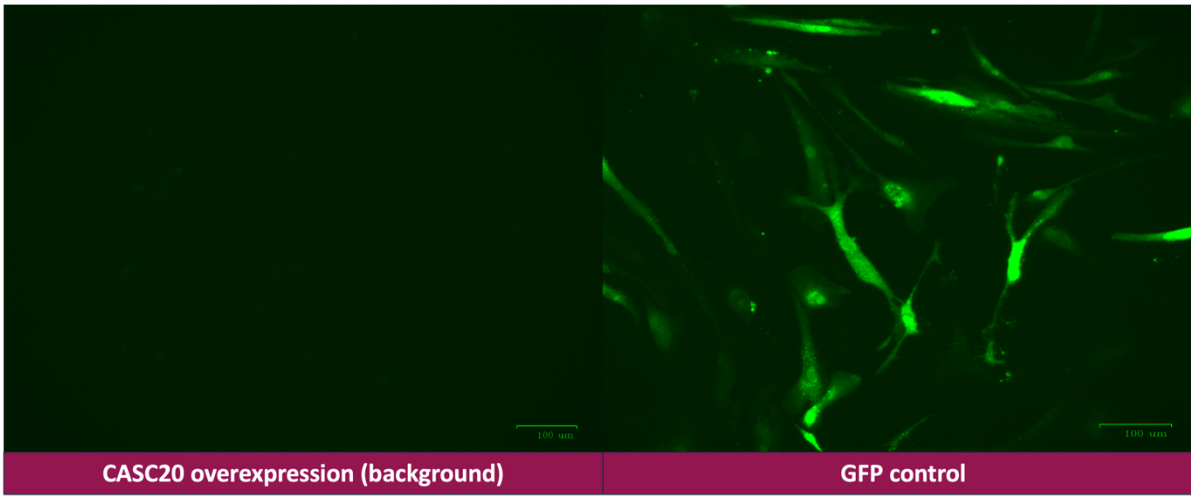
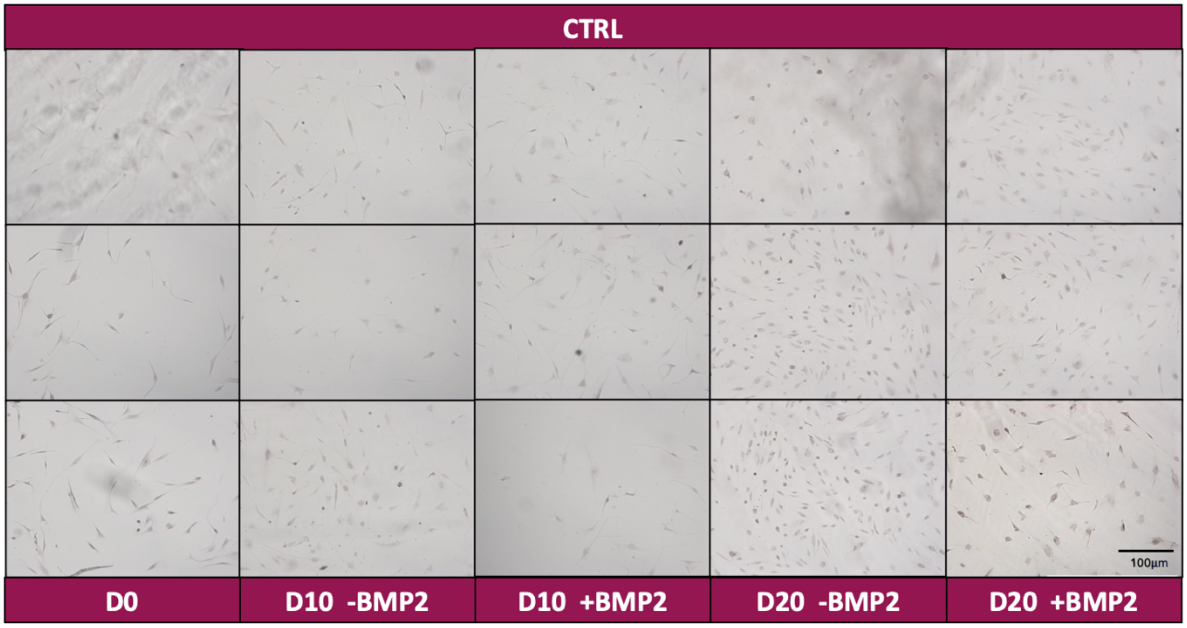


Figure 4.3: Fluorescence microscopy of transduced hMSCs. *Left panel: cells transduced with CASC20 OE lentivirus, which has no fluorescence so was used as background. Right panel: cells transduced with the GFP CTRL lentivirus.*

A



B

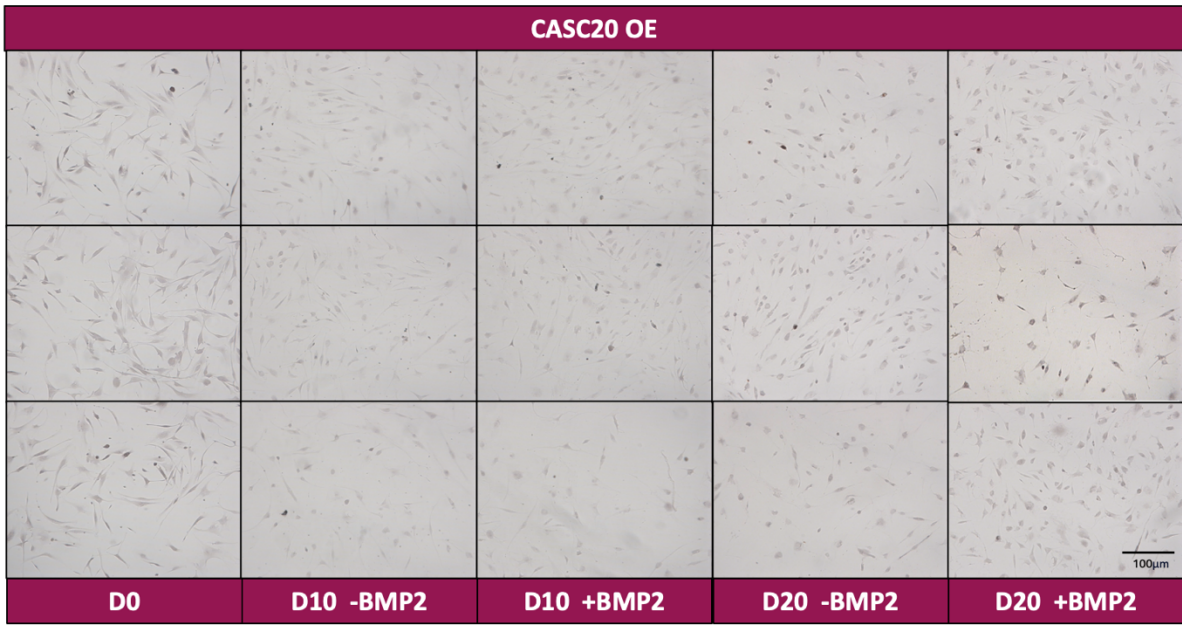
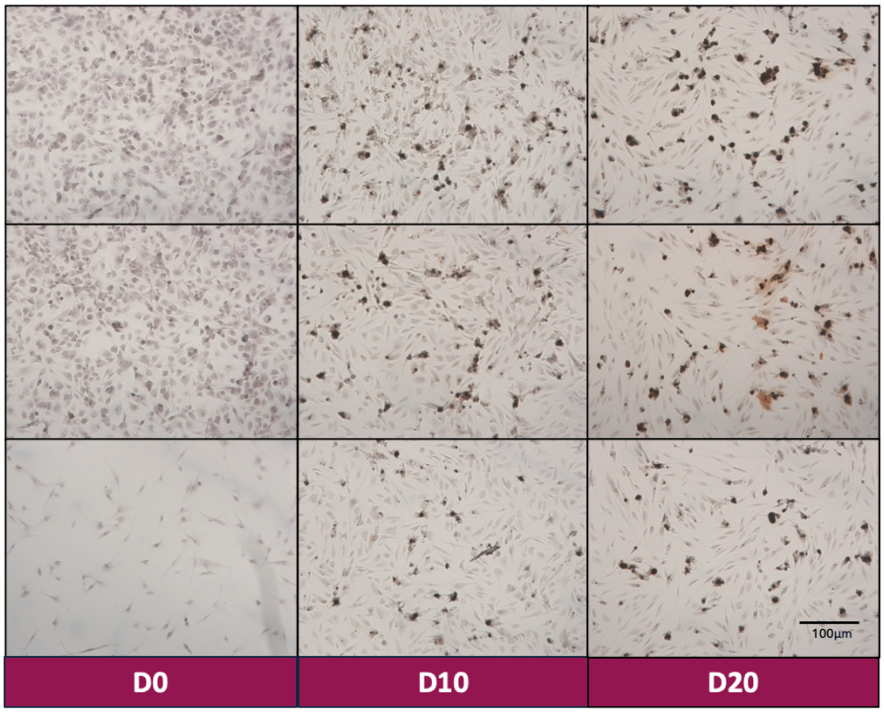


Figure 4.4: Microscopy images of hMSCs during osteodifferentiation following Alizarin Red Staining. *hMSC mix-1 was transduced with either* ***(A)*** *CTRL or* ***(B)*** *CASC20 OE lentivirus, then differentiated in OM, +/- BMP2, for 0, 10 or 20 days. Rows show different replicates. Scale bar on the bottom right image is the same for all images.*

A second mix of pooled cells ‘hMSC mix-2’ was differentiated, where half were non-transduced and the remaining cells were transduced with either CTRL or OE lentivirus. There was evidence of successful osteogenic differentiation in the non-transduced cells, with staining observed as well as the formation of characteristic bone nodules, both of which were observed to increase over time (i.e. more in D20 than D10) (figure 4.5A). In the transduced cells there was no evidence of osteogenic differentiation and more cell death was observed, with little to no cells remaining by D10 (figure 4.5B). Therefore, the decision was made to halt the experiment at this point to examine why the transduced cells were not differentiating as expected.

A



B

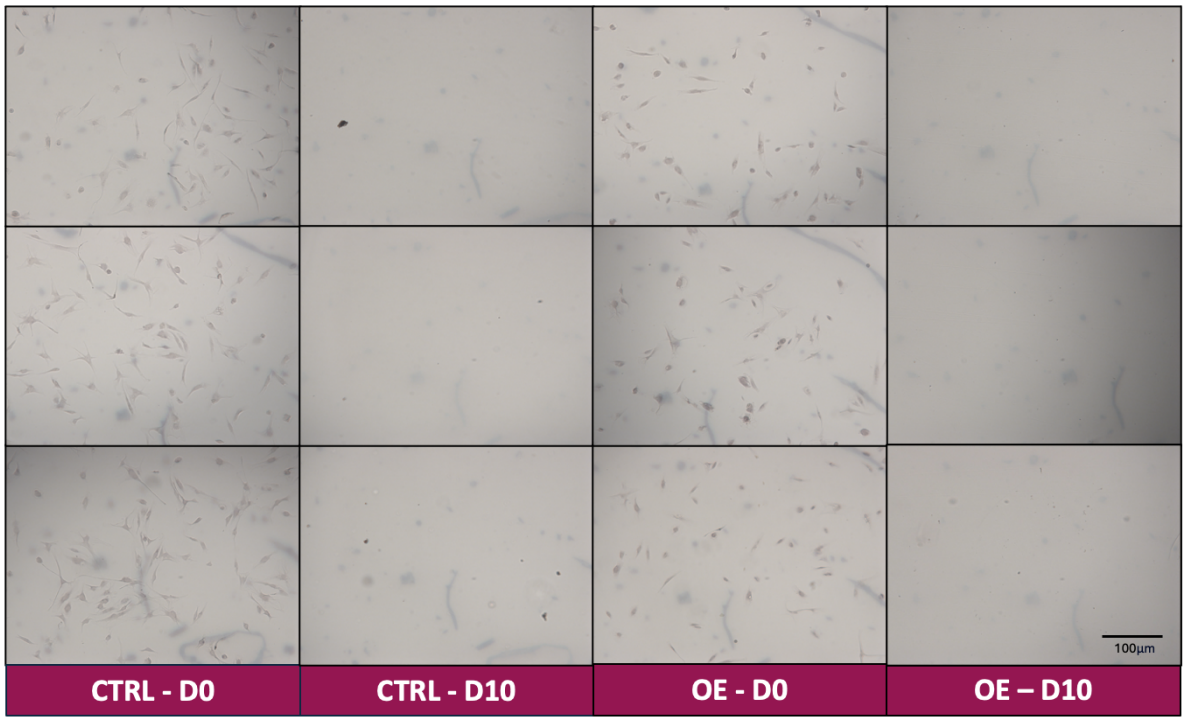
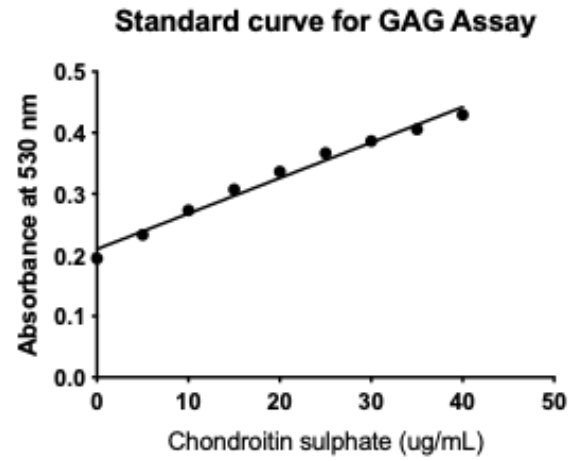


Figure 4.5: Microscopy images hMSCs during osteodifferentiation after Alizarin Red Staining. (A) *Non-transduced and* ***(B)*** *transduced cells. Images taken at D0, 10 and 20. Rows show different replicates. Scale bar on the bottom right image is the same for all images.*

4.3.2 Chondrogenesis of hMSCs

In parallel, the same hMSC mix-1 was induced to differentiate into chondrocytes by culturing in chondrogenic media for 0, 7 or 14 days. An assay for GAG release was performed to quantify GAG in the culture media collected throughout the experiment, and in cell pellets collected at each time point. A standard curve was created with Chondroitin Sulphate, a GAG, to interpolate the concentration of GAG in the experiment (figure 4.6A). However, no GAG was detected in either the media or pellet, indicating the hMSCs had not differentiated into chondrocytes, as was seen by the lack of colour change (figure 4.6B).

A



B

Figure 4.6: Standard Curve for DMMB Assay. (A) *Standard curve created with serial dilution of Chondroitin Sulphate and absorbance read at 530 nm, alongside samples.* ***(B)*** *Image of plate with standard curve in row A and samples in rows B-E.*

4.3.3 Optimisation of lentivirus MOI

Two sets of primers were designed to target the reverse response element (RRE) region of the lentiviral plasmids, which would form part of the sequence integrated into the viruses (figure 4.2). Both primers were compared for suitability, and both successfully amplified the gene of interest and produced similar CT values (figure 4.7). The cDNA was diluted 1 in 25 to analyse results from the linear part of the qPCR curve. No significant difference was found between old and new stocks of the lentiviruses with either primer pair. For primer pairs 1 and 2 the p-values were 0.2 and 0.1 for OE and CTRL viruses, respectively.

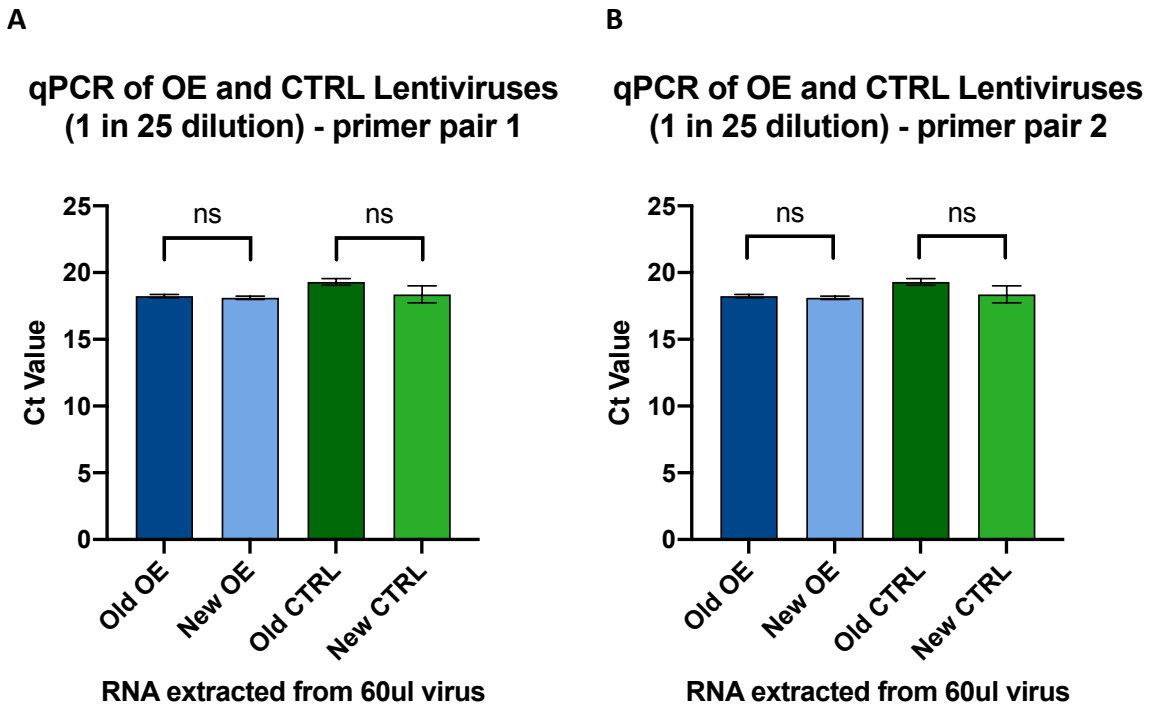


Figure 4.7: qPCR of CASC20 OE and CTRL lentiviruses with primer pairs (A) *1 and* ***(B)*** *2. Significance was determined by performing Mann-Whitney test.*

The dilution series of each plasmid was run with both sets of primers, and a standard curve for each plasmid produced (figure 4.8A/B). This standard curve was initially used to interpolate the concentration of the lentiviruses (figure 4.9C). Here the focus was on the 'new' lentiviruses, as these were used in the experiments described within in this chapter and thus the ones where it was most important for the MOI to be known. Then the number of molecules corresponding to these concentrations was calculated, showing a difference in

the numbers of viral particles between the OE and control lentiviruses, with the OE lentivirus having almost double the number of molecules. (figure 4.9). From this the MOI of both lentiviruses used in the previous differentiations was calculated to be approximately 140,000 and 150,000 for the *CASC20* OE and CTRL lentiviruses, respectively (table 4.1). The new lentiviruses both had slightly higher MOIs compared to the old lentiviruses, but all exhibited very high MOIs.

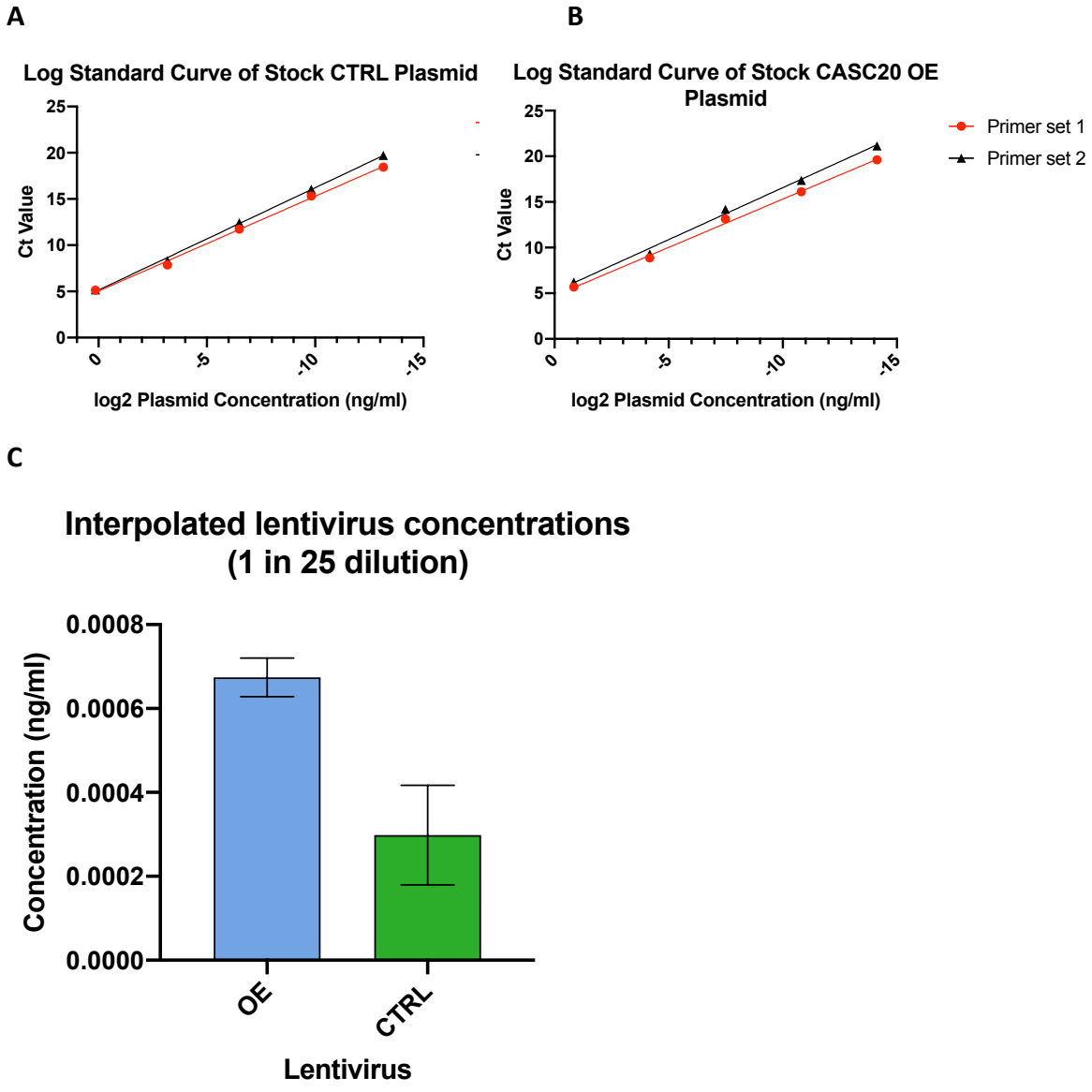


Figure 4.8: Interpolation of lentiviral concentrations from plasmid standard curves. (A) *Log standard curves for qPCR of CTRL and* ***(B)*** *OE plasmids.* ***(C)*** *Interpolated lentiviral concentrations.*

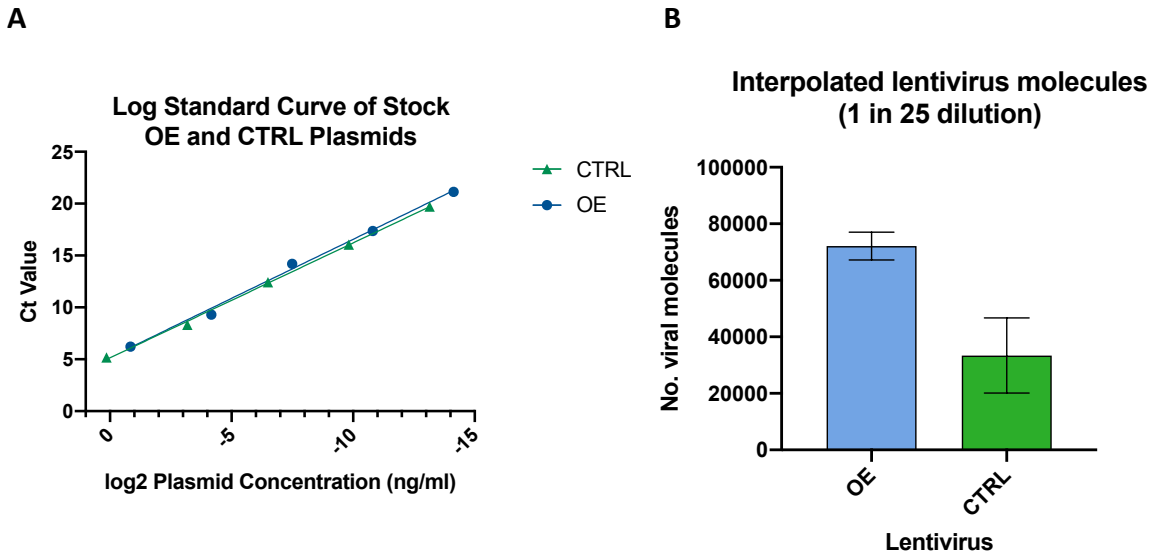


Figure 4.9: Interpolation of lentiviral concentrations from plasmid standard curves. (A) *Log standard curve for qPCR of CTRL and OE plasmids. Primer set 2 used.* ***(B)*** *Interpolated lentiviral concentrations.*

To optimise the MOI used for transductions and attempt to lower this a single hMSC isolate was cultured and transduced with CTRL lentivirus at five MOIs: 150,000, 30,000, 6000, 1200, 240. The cells were monitored for fluorescence for the next 72 hours, and GFP expression was visible after 48 hours in the cells transduced with an MOI of 150,000 (figure 4.10). There was almost no visible fluorescence at other concentrations, with only very small circles of fluorescence seen in some cells at an MOI of 30,000.

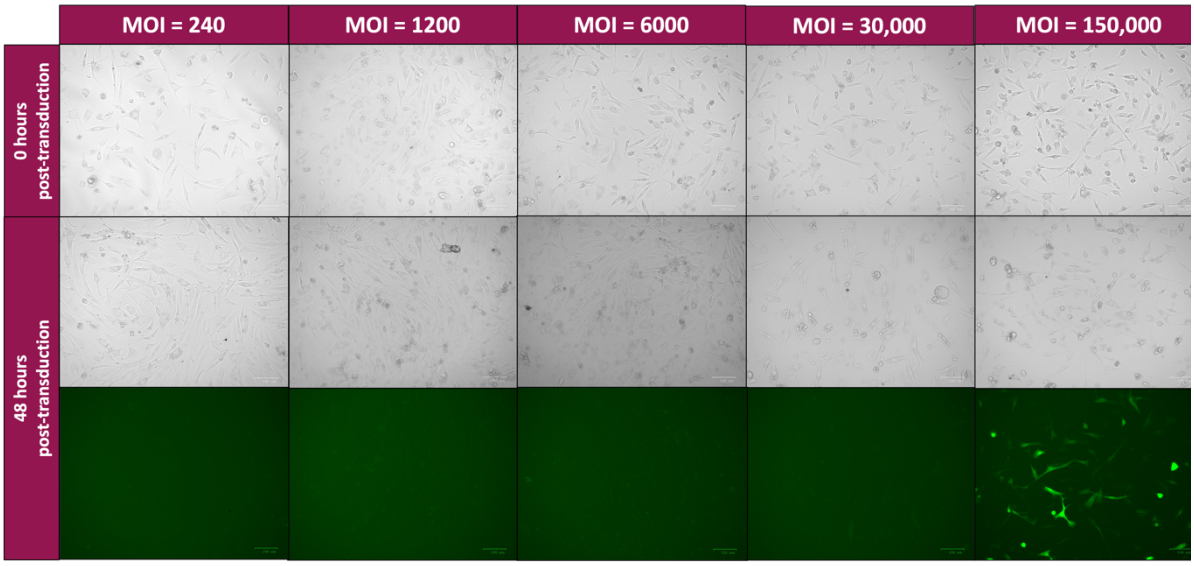


Figure 4.10: Microscopy of hMSCs transduction with increasing MOIs of GFP control lentivirus. *Initial images taken immediately after lentiviral transduction with CTRL lentivirus (0 hours) and then 48 hours after.*

Overall, these experiments showed that lentiviral transduction of primary hMSCs stopped their ability to differentiate into osteoblasts and chondrocytes.

4.4 Discussion

Cells were pooled for differentiation, as this has been shown to help with donor variability whilst increasing proliferation and differential capacity (Widholz et al., 2019). It has been demonstrated that variance between technical replicates is significantly less than variance seen between cells originating from different donors, when looking at cell proliferative and differential capacity (Widholz et al., 2019). Thus, donor variability is something that was of concern in these experiments, as many factors such as age (Kume et al., 2005, Zaim et al., 2012), sex (Bragdon et al., 2015), BMI (Oliva-Olivera et al., 2015) and co-morbidities (Ye and Zhang, 2017) can affect cell proliferation, vitality and differentiation (Widholz et al., 2019). However, no successful differentiation of hMSC mix-1 into osteoblasts or chondrocytes was observed. The lack of ARS in the cells cultured for osteogenesis indicates there were no calcium phosphate deposits, which is indicative of osteogenic differentiation (figure 4.4). Ideally a positive control would have also been included for ARS to ensure the methodology was working correctly. However, there was also no observation of characteristic morphological changes, such as the formation of bone nodules. In future a nuclear stain would have been included to confirm the presence of cells in the control, although cells were present on visual examination of the microscopy images. There was also no evidence of successful chondrogenesis, due to the lack of GAG present in either the cells pellets or media. As all hMSCs in this experiment had been transduced it was impossible to determine if the lack of successful differentiation was due to the characteristics of the primary hMSCs, or due to the process of transduction itself affecting the cell's ability to differentiate.

Thus, it was necessary to test the ability of non-transduced cells to differentiate into osteoblasts and chondrocytes. As non-transduced cells did undergo osteogenic differentiation, indicated by the presence of ARS at D20 and bone nodules at D10 and D20, this suggests the process of transduction is responsible for the hMSCs losing their capacity to differentiate into osteoblasts. Furthermore, the increase of bone nodules over time showed the cells proceeding down the osteogenic lineage, and the lack of uniformity in cells

may be partly due to the pooling of samples. As these are primary cells it is expected they will not differentiate at the same rate, so cells originating from certain isolates may successfully differentiate by D10, whilst others differentiate by D20 and others may take even longer. The amount of ARS is lesser than has been observed in cell lines under the same conditions (Felix-Ilemhenbhio, 2023), but the nature of these primary cells and their origin from unhealthy tissue, could affect their ability to differentiate to the same extent or at the same rate. Ideally this experiment would have carried with chondrogenesis in parallel, but was limited by cell numbers. However, successful osteogenesis is a good indicator that these cells are able to differentiate successfully, and thus would also be able to proceed down the chondrogenic lineage as well. The apparent ability of the non-transduced cells to differentiate, whilst the transduced cells still could not, suggests there is a problem in getting primary hMSCs to differentiate following transduction with a lentiviral vector.

There are several possibilities for why transducing hMSCs had this effect on differentiation, including that the reagent polybrene, added to increase the transduction efficiency, is known to be harsh to cells and affect their proliferation and differentiation at the working concentration used (Lin et al., 2011). Polybrene is a standard reagent used in transduction of human cells (Elegheert et al., 2018, Denning et al., 2013), but testing with an alternative additive would be useful in the future to see how this affects hMSC differentiation. Research has shown that using protamine sulphate as an alternative can help hMSCs retain their proliferative and differential capacity (Lin et al., 2012). There is also evidence that lentiviral transduction of primary human stem cells can affect their differentiative capacity, as the ability of lentiviruses to integrate into the genome means they can potentially disrupt genes through insertional mutagenesis (Schlimgen et al., 2016). This ability to affect gene expression means they could influence genes necessary for differentiation.

The new stocks of lentivirus were compared to the old stocks, to determine if there was a significant difference between these that could be responsible for the problems with differentiation. These old stocks had been successfully used to overexpress *CASC20* whilst cells were still able to differentiate, but had not been tested with primary human cells (Felix-Ilemhenbhio, 2023). However, as no significant difference was observed, it suggests there is not a problem with the new lentiviral stocks, but that the problem lies with the lentiviral

transduction as a whole. The calculation of the MOI used to be ~150,000 suggests the lentiviral transduction is very inefficient. A typical MOI value for hMSCs should be around 10,000-fold less (Lin et al., 2012). It is possible that such a high amount of lentivirus is toxic to the cells and could affect their ability to differentiate or function normally. As lower MOIs were incapable of producing fluorescence this is further evidence for the inefficiency of the current lentiviral transduction process. The very limited fluorescence seen at the MOI of 30,000 was likely autofluorescence from dying cells (Bertolo et al., 2019). From these results, it appears that transduction of hMSCs with lower concentrations of the lentivirus is not possible under the current conditions. Thus, lentiviral transduction is not a viable method to proceed with for overexpression of *CASC20* in hMSCs.

Chapter 5: Testing adenovirus as an alternative transduction method

5.1 Introduction

Adenoviruses offer an alternative method to lentiviruses for transducing cells, so this was tested to confirm if it was possible to transduce hMSCs whilst allowing them to retain their differentiative capacity. Adenoviral transduction is considered a more promising alternative, because adenoviruses do not integrate into the genome, reducing the risk of insertional mutagenesis (Bulcha et al., 2021). Furthermore, it is considered the most effective method of gene delivery, with high transgene expression which should be beneficial in achieving gene expression with lower MOIs (Crystal, 2014).

5.2 Methods

5.2.1 Isolation of hMSCs from SVF

hMSCs were isolated from SVF as described in chapter 4.

5.2.2 Culturing of hMSCs

hMSCs were cultured as described in chapter 4.

5.2.3 Adenoviral Transduction

Prior to adenoviral transduction cells were passaged as previously described and seeded into flasks or plates as appropriate. Cells were incubated for 24 hours in DMEM + Glutamax, with 15% FBS and 0.8% P-S. After this period, the media was changed and Ad-CMV-GFP vector (figure 5.1) was added at the desired multiplicity of infection (MOI) according to the experiment. The flask or plate was mixed for at least 10 seconds to ensure even spread of the adenovirus. The cells were incubated for 24 hours, then the media was changed to remove the adenovirus. The cells were cultured for ~48 hours to allow them to recover from transduction. Expression of GFP was checked under the microscope (Bio-Rad, ZOE Fluorescent Cell Imager, USA) after 48 hours to confirm successful transduction. Images taken at 20x magnification unless otherwise specified.

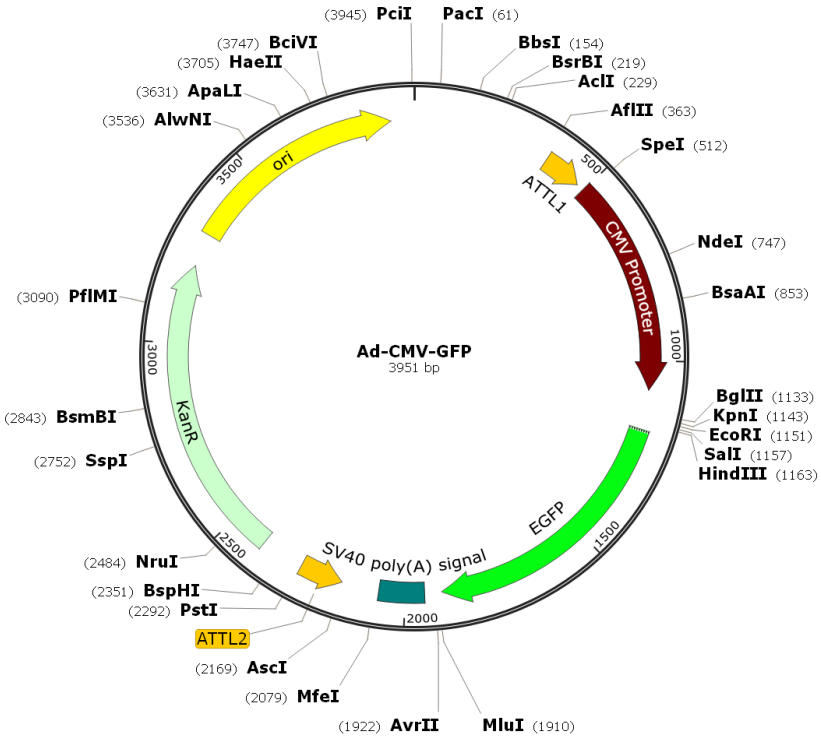


Figure 5.1: Plasmid map of Ad-CMV-GFP. *Vector kindly donated from Professor Munitta Muthana. Plasmid map created in Snapgene.*

5.2.4 Osteogenic Differentiation

hMSCs were seeded into plates at the start of P3, alongside non-transduced controls and induced to begin differentiation as in section 4. Cells were cultured in OM with BMP2, for 0 or 24 days. Osteogenesis carried out as in chapter 4.

5.2.5 Chondrogenic Differentiation

As for osteogenesis, hMSCs were seeded into plates at the start of P3 and cultured in chondrogenic media as cell pellets, as in section 4. Cells were cultured for 0 or 14 days. Chondrogenesis carried out as in chapter 4.

5.2.6 Assay for calcium deposition

ARS performed as per chapter 4.

5.2.7 Assay for GAG release

DMMB assay for GAG performed as per section 4.

5.3 Results

5.3.1 Optimisation of MOI

One hMSC isolate (P800) was cultured and transduced with Ad-CMV-GFP at three MOIs: 125, 250, 500. The cells were monitored for fluorescence over the next 72 hours, and microscopy images were taken after 0 and 48 hours (figure 5.2). Immediately following transduction there was no expression of GFP (figure 5.2A) After 48 hours the cells clearly expressed GFP at all MOIs, with an increasing amount of fluorescence seen as the MOI increased (figure 5.2B). However, some cell death was observed, particularly at the higher MOIs of adenovirus (figure 5.2B). Cellular death was determined through the visible structure of the cells in the microscopy images but was not quantified.

A

B

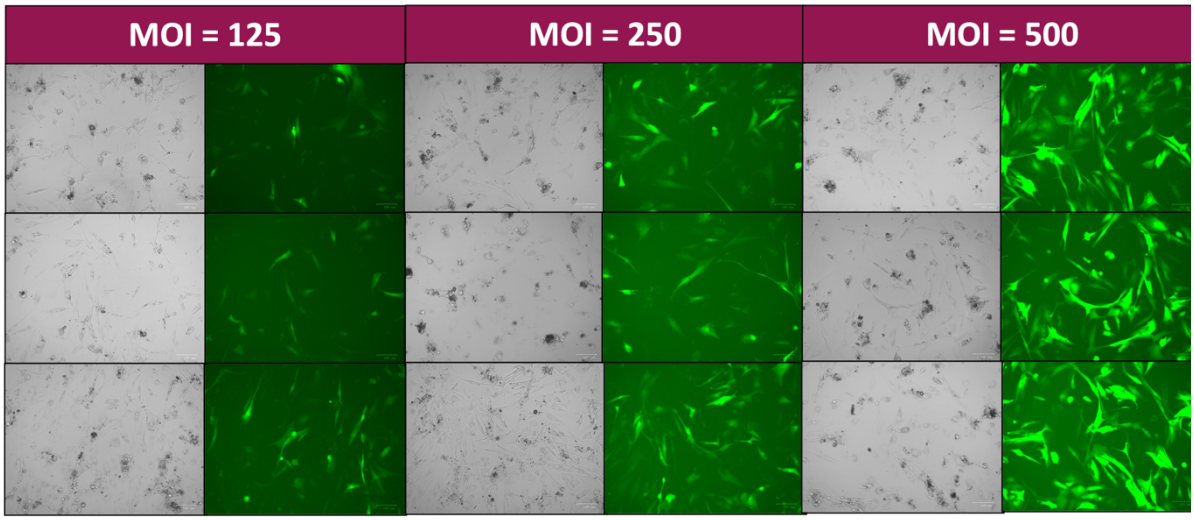


Figure 5.2: Microscopy of hMSCs after transduction with various MOIs of Ad-CMV-GFP. *P800.* ***(A)*** *Images taken 0 hours after transduction.* ***(B)*** *Images taken 48 hours after transduction.*

5.3.2 Osteogenesis of hMSCs

The hMSCs were successfully transduced with Ad-CMV-GFP, as microscopy showed their fluorescence prior to differentiation (figure 5.3). After culturing for 24 days some minor ARS was observed, indicating the presence of calcium deposits (figure 5.4), alongside the presence of bone nodules. This suggests osteogenic differentiation occurred, but at a slower rate than seen in pooled hMSC-mix 2. There was no observational difference between transduced and non-transduced cells in the rate or ability to differentiate (figure 5.4).

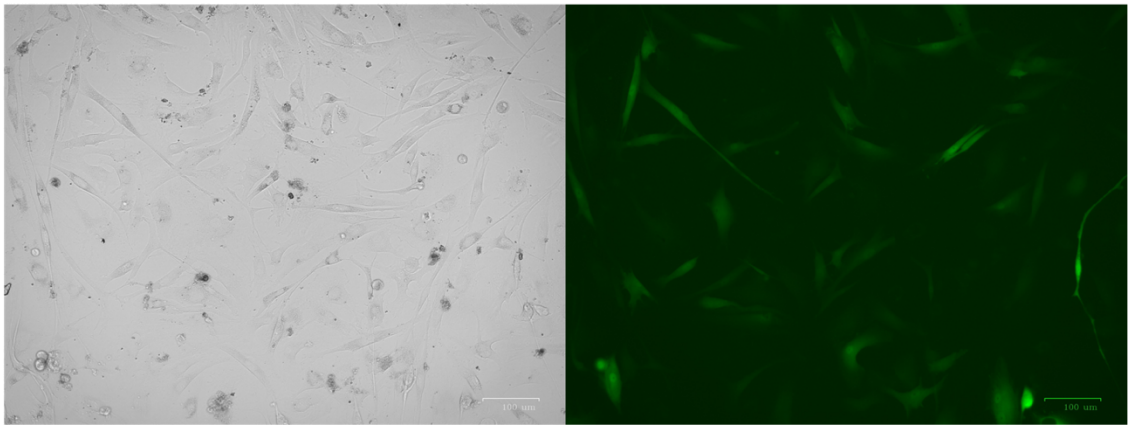


Figure 5.3: Microscopy of hMSCs after transduction with Ad-CMV-GFP, prior to differentiation (D0). *Left panel shows brightfield and right panel shows fluorescence (GFP).*

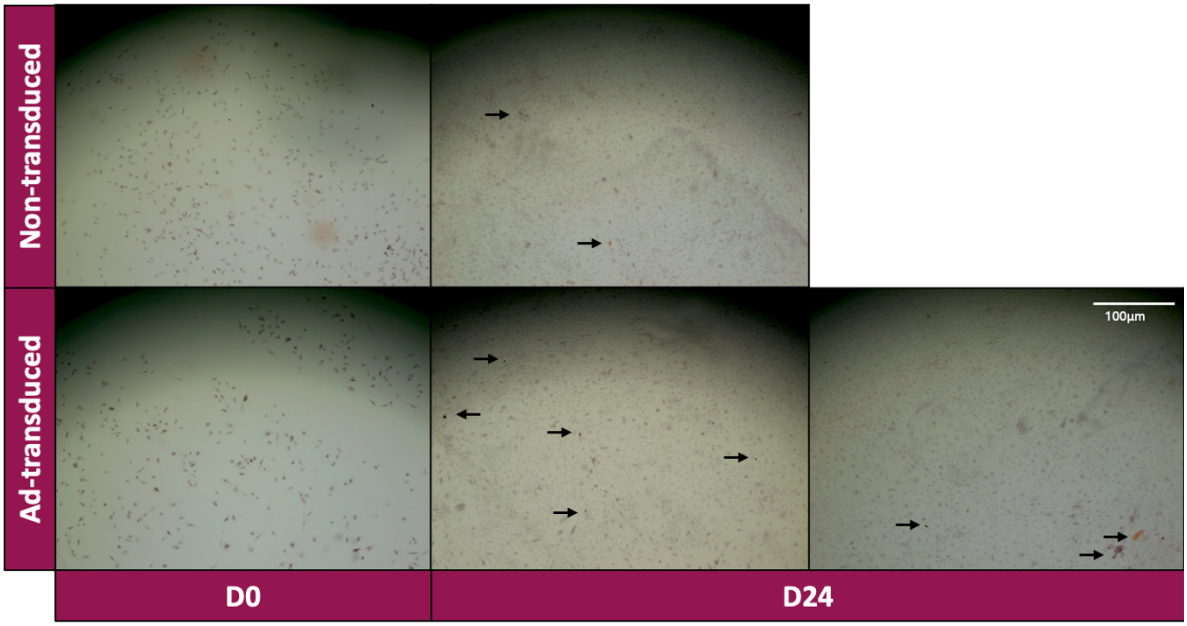


Figure 5.4 Alizarin red staining of hMSCs before osteogenic differentiation (D0) and at D24. *Cells were non-transduced or transduced with Ad-CMV-GFP prior to differentiation.*

Areas of staining and/or bone nodule formation are marked with black arrows. Scale bar on the bottom right image is the same for all images.

5.3.3 Chondrogenesis of hMSCs

In the culture media GAG was found in the D0-3 media for both non-transduced and transduced hMSCs, when compared to D0 (figure 5.5). In the transduced hMSCs GAG was also identified in media from D3-8 and D11-14. GAG release occurred initially during the differentiation, but the concentration decreased as the differentiation continued (figure 5.5). No GAG was identified in the cell pellets. The presence of GAG in the media is indicative of successful chondrogenic differentiation, but this is not supported by the lack of GAG in the pellet.

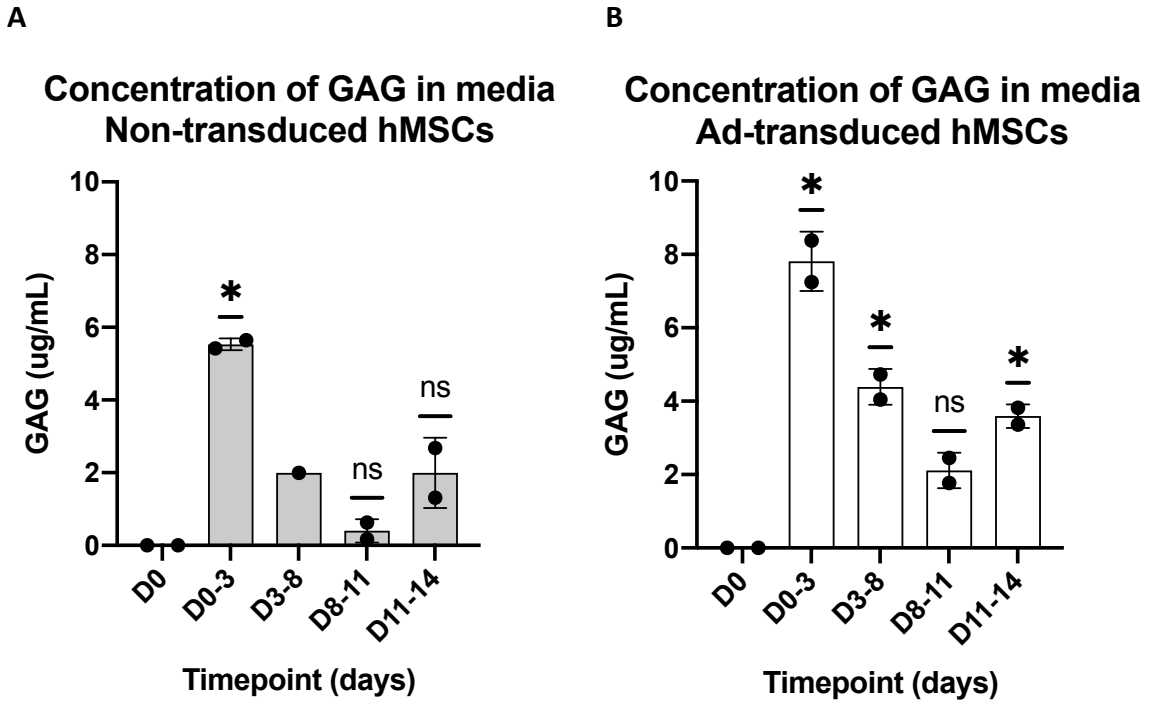


Figure 5.5: DMMB assay on media and pellet. *The waste culture media of non-transduced or Ad-CMV-GFP transduced hMSCs was tested for the release of GAG.* ***(A)*** *Interpolated concentrations of GAG in non-transduced and* ***(B)*** *Ad-CMV-GFP transduced cells. One-Sample Wilcoxon tests were performed to compare each timepoint to D0 to statistically test for significance. Significance is displayed as * = P< 0.05.*

5.4 Discussion

After testing transduction of hMSCs with Ad-CMV-GFP an MOI of 125 was chosen as this still resulted in a reasonable amount of adenovirus expression but with less cellular death and stress. Typical MOIs for infection with adenoviruses vary in the literature, depending on the cell type, but 125 sits within the standard range of 10 -1000 (Podvin et al., 2011, Ugai et al., 2012). Ideally the levels of cell death in cultures would have been quantified with trypan blue, but a visual check was deemed sufficient in this case to judge the best MOI to proceed with. Whilst this study was limited to only one isolate, the next differentiation with Ad-CMV-GFP was performed with a different isolate and showed similar levels of GFP expression when transduced with an MOI of 125, providing further supportive evidence.

With regards to the osteogenic differentiation with Ad-CMV-GFP, the morphological changes and formation of some bone nodules were indicative of successful osteogenesis, but there was less staining and nodules than in hMSC-mix 2. Ideally genes that are markers of successful osteogenesis, *ALP*, *Runt related transcription factor 2* (*RUNX2*) and *osteocalcin*, would have been checked for through qPCR (Okamura et al., 2020). It is expected that primary cells will all differentiate at different rates, so it is possible this particular isolate may not be representative of the average differentiation time. Alternatively, the act of pooling in itself may increase the rate of differentiation (Widholz et al., 2019). Whilst the mechanism for this is uncertain, it suggests that cells with better proliferative and differential potential contribute relatively more to the pooled population (Widholz et al., 2019). Furthermore, the apparent seeding density at D0 appeared sparser than in previous differentiations, despite the same number of cells aimed to be seeded. It is therefore possible that the cells took longer to differentiate because of the lower confluency, so they were more biased towards proliferation than differentiation. Mathematical modelling has shown there is a choice between a higher number of less differentiated cells or a lower number of more differentiated cells, as cells must direct their energy towards metabolic pathways that promote optimal growth, or those for differentiation (Győrgy et al., 2019). In future it will be important that cells are given time to reach confluency after seeding and before beginning differentiation, to somewhat standardise the confluency and cell number. ARS was performed at day 24 for the final time point, rather than the usual D20 because the

differentiation appeared to be proceeding at a slower rate. This is still in line with typical differentiation rates seen in the literature, where differentiation is often tested for up to 20-28 days (Tsai et al., 2009, Hanna et al., 2018, Kim et al., 2023). Furthermore, alternative endpoints of differentiation could be tested to validate what is seen with observational ARS testing. This could include testing for the presence/activity of *ALP* or *osteocalcin* (Sabokbar et al., 1994). It would also be beneficial to confirm the findings for osteogenesis in further isolates. This would aid understanding of the usual rate of osteogenesis and any possible effects of pooling versus individual isolates.

For the chondrogenic differentiation the highest concentration of GAG was seen in the media collected from D0-3 of culturing, which is mostly likely due to the initial release of GAGs to stabilise the extracellular matrix (Florencio-Silva et al., 2015). GAG release in the pellet would have been expected in D14 samples, and with small amounts likely at D7 too, as pellets were visible in all samples. The lack of GAG release is likely due to a problem with the papain used, as this was an old aliquot stored at -20°, which likely caused it to lose its activity. Storage of papain at -20° has been shown to cause structural rearrangements and aggregation, and multiple freeze-thaw cycles can cause significant activity loss (Raskovic et al., 2015). Therefore, based on the GAG release in the media, and explanation for the lack of GAG seen in the pellet, it is likely that chondrogenic differentiation did occur in these cells. However, to be sure the hMSCs are capable of successfully becoming chondrocytes after adenoviral transduction, further isolates need to be tested. Again, it may be beneficial to test an alternative endpoint, such as using qPCR to look for the expression of chondrogenic genes to further confirm successful differentiation. Staining methods, such as toluidine blue to visualise proteoglycans could also be useful to provide further confirmation (Bergholt et al., 2019).

However, this study was limited by only having two replicates, due to limited cell numbers and adenoviral supply. Ideally at least 3 replicates would have been used in order to perform proper statistical analysis.

Chapter 6: General Discussion & Conclusion

6.1 Key Findings

Assess the possibility of targeting lncRNAs such as *CASC20* therapeutically through a literature review.

There are currently no registered clinical trials targeting lncRNAs, due to a number of limitations for their usage. However, there is potential for this to change as there are treatments targeting other ncRNAs. If *CASC20* role was fully understood and could be targeted for disease treatment then this is a distinct possibility, but this would take many years of research before even reaching clinical trials. As very little is known about *CASC20*s effects throughout the body and in other cell types, further research would be required to evaluate the risk and likelihood of any off-target effects occurring if this lncRNA was targeted for treatment.

Identify potential miRNA targets of *CASC20*.

Analysis carried out on murine MSCs following lentiviral transduction with a *CASC20* expression vector or a control vector showed the greatest variance between timepoints, then conditions. Overall, D10 showed the greatest difference caused by *CASC20* expression, in both mRNA and miRNA-seq, with ~30x more DEGs than at D0/20. The mRNA analysis further identified a potential role for *CASC20* in splicing, as splicing related categories were found within the top 10 GO categories for the DEGs. As expected, the DE miRNAs showed a role in development from the GOseq analysis, suggesting *CASC20* causes differential expression of miRNAs during osteogenesis. This would be expected if *CASC20*s mechanism is through the modulation of miRNAs. In GSEA, nine miRNAs had binding sites enriched in DEGs at D10 and were DE at D10. Of these eight were upregulated, suggesting *CASC20* does not sponge miRNAs, as it would be expected that most would be downregulated. When looking at miRNAs that actually bind to *CASC20* there were six miRNAs that were predicted to target *CASC20* and also had enriched binding sites in the DEGs. Overall, there was no evidence of *CASC20* acting as a ceRNA. The miRNAs identified as binding to *CASC20* mostly had just one site present in the sequence, meaning this mechanism is unlikely. Of interest for further research is miR-4661-3p that was identified from Sylamer and has been found to

have a role in upregulating osteogenesis. However, this research is limited by being carried out in murine cells, where *CASC20* is not endogenously present, and results would need to be repeated in primary human cells to draw firm conclusions from this.

Validate findings that *CASC20* upregulates osteogenesis and downregulates chondrogenesis in human mesenchymal stromal cells (hMSCs).

Overall, the research presented in this thesis has shown that lentiviral transduction was not a viable method to induce overexpression of *CASC20* in primary hMSCs collected from SVF of patients. However, adenoviral transduction is a promising alternative and can be used at much lower MOIs than lentivirus. After adenoviral transduction of hMSCs to induce *CASC20* overexpression, a visible increase in osteogenesis was seen, but this was in a limited sample number. Ideally, this would be replicated in a larger pool of samples and for chondrogenic differentiation.

6.2 Future Work

One key avenue for future experimental work will be mRNA- and miRNA-sequencing of primary adipose-derived hMSCs during osteogenic and chondrogenic differentiation, to investigate DEGs as a result of *CASC20* overexpression. Cells would be transduced with *CASC20*-expressing or control GFP-expressing adenovirus. It will be important to observe if expression patterns in murine MSCs and the effects on differentiation, i.e. upregulation of osteogenesis and downregulation of chondrogenesis, are replicated. This data would be analysed as it was for the murine MSCs, including differential expression analysis and investigation of the interactions between miRNAs and mRNA. The high levels of similarity between the mouse and human genome mean it is likely similar results will be seen, but as *CASC20* is a human only gene this cannot be assumed.

Once miRNA targets of *CASC20* have been identified, the next steps would involve RNA pulldown and CLIP-seq experiments to identify interacting RNA-binding proteins, which may provide evidence for a particular mechanism. CLIP-seq could identify the miRNA(s) that bind to *CASC20* and the location(s) this occurs. CLIP utilises UV light to irradiate cells, causing nearby proteins (or miRNAs) to irreversibly cross link to RNA by covalent bonds (Hafner et

al., 2021, Clark et al., 2014). This cross-linking allows the RNA-protein/miRNA complex to be purified then the interactions between the protein/miRNA and the whole transcriptome can be characterised (Hafner et al., 2021). An understanding of the miRNAs that bind to *CASC20* and where this occurs is critical to understanding the mechanism. As discussed in chapter 2, there are three main types of lncRNA mechanism: decoy, scaffold and guide. *CASC20* may act by one or multiple of these mechanisms. As well as overexpression studies, knockdown studies will also be important to further characterise the mechanism. This could be achieved with small interfering RNAs (siRNAs) to target *CASC20*. Evidence of a decoy/sponging mechanism could include increased circulating levels of the interacting miRNA(s) following knockdown of *CASC20*. Overexpression would be expected to have the opposite effect and decrease circulating levels of the miRNA(s), as they would be bound to *CASC20*. Understanding the other binding partners of the miRNA(s) will also be critical to identify any decoy mechanisms, as how the levels of expression of these target genes change will provide further evidence. Evidence of a scaffold or guide mechanism will likely involve evidence of a role for *CASC20* in transcription, such as through identifying polycomb repressive complexes (PRCs) as binding partners. Knockdown of *CASC20* and its miRNA binding partners would inhibit target binding. Another important element to understand if *CASC20* acts through a scaffold or guide mechanism is whether it acts in cis on neighbouring genes, or in trans on distant genes. Uncovering the locations of binding as described earlier will provide evidence of this. Furthermore, RNA fluorescent in-situ hybridisation (FISH) could be used to understand where *CASC20* is acting within osteoblasts and chondrocytes, as cytoplasmic lncRNAs are more likely to act as ceRNAs whereas nuclear lncRNAs typically act via scaffold or guide mechanisms. Whilst it is known that *CASC20* is capable of reaching the cytoplasm it has not been observed whether this is the case in differentiating cells. This could be measured by performing qPCR on a sample from the cytoplasm to test for the presence of *CASC20*. Overall, this should help to create a picture of how *CASC20* is acting and how this mechanism causes the observed effects on osteogenesis and chondrogenesis.

Another key avenue of experimental work will include a CRISPR strategy to dissect out whether *CASC20* acts via *BMP2* or independently. As the *CASC20* and *BMP2* genes are in close proximity to each other and BMP2 is a critical protein to osteogenesis, it is possible that *CASC20* may act as an enhancer of *BMP2*. Thus, the observed effects of *CASC20* on

osteogenesis and chondrogenesis may be through *BMP2*. Expression quantitative trait loci (eQTLs) are genomic loci that regulate the expression levels of mRNA or proteins. When examining the association between single nucleotide polymorphisms (SNPs) within *CASC20* and *BMP2* expression in unstimulated primary human chondrocytes, no *BMP2*-acting eQTLs within *CASC20* were found (Tachmazidou et al., 2019). However, there are SNPs present in the same open chromatin regions of both genes that are associated with enhancer marks. This means there are potential enhancer regulatory regions in the *CASC20* locus. The SNPs in *CASC20* are in a region within an enhancer for *BMP2*, so could enhance *BMP2* activity. CRISPR will be used to investigate this, by studying the effects of deleting regions of *CASC20*, including the transcriptional start site (TSS) and the SNP region. For this an osteoblastic cell line, Saos-2, will be used due to the technical difficulties of performing CRISPR in primary human cells. After CRISPR has successfully deleted the desired *CASC20* region the cells will be differentiated down the osteogenic route to investigate if the same effects on osteogenesis are observed as when *CASC20* was present, as this would suggest that *BMP2* is responsible for the action rather than *CASC20*.

6.3 Concluding Remarks

Overall whilst the knock in model of *CASC20* expression in murine MSCs was a useful tool to better understand the genes and miRNAs involved with *CASC20*, it ultimately requires validation in human cells. The evidence presented in this thesis does not clearly point to one mechanism of action for *CASC20* over another but does suggest that it acting as a ceRNA is unlikely. Further research is needed to elucidate exactly how *CASC20* impacts on osteogenesis and chondrogenesis and whether it could be a possible target for disease treatment.

Chapter 7: Appendix

7.1 Predicted miRNAs in *CASC20*

Table 7.1: Mus musculus miRNAs predicted by MiRanda to bind to CASC20. The number of hits indicates the number of times the miRNA is predicted to bind within the whole sequence of *CASC20*.

miRNA	Target	Number of Hits
mmu-miR-694	*CASC20*	3
mmu-miR-150-5p	*CASC20*	2
mmu-miR-1897-5p	*CASC20*	2
mmu-miR-302b-5p	*CASC20*	2
mmu-miR-302c-5p	*CASC20*	2
mmu-miR-302d-5p	*CASC20*	2
mmu-miR-3113-3p	*CASC20*	2
mmu-miR-3113-5p	*CASC20*	2
mmu-miR-5127	*CASC20*	2
mmu-miR-544-3p	*CASC20*	2
mmu-miR-5619-5p	*CASC20*	2
mmu-miR-664-3p	*CASC20*	2
mmu-miR-6992-3p	*CASC20*	2
mmu-miR-7015-3p	*CASC20*	2
mmu-miR-7119-3p	*CASC20*	2
mmu-miR-7650-5p	*CASC20*	2
mmu-miR-1187	*CASC20*	1
mmu-miR-1192	*CASC20*	1
mmu-miR-1194	*CASC20*	1
mmu-miR-1198-3p	*CASC20*	1
mmu-miR-12188-3p	*CASC20*	1
mmu-miR-12200-5p	*CASC20*	1
mmu-miR-12205-3p	*CASC20*	1
mmu-miR-143-5p	*CASC20*	1
mmu-miR-153-5p	*CASC20*	1
mmu-miR-154-3p	*CASC20*	1
mmu-miR-181a-5p	*CASC20*	1
mmu-miR-181b-5p	*CASC20*	1
mmu-miR-181c-5p	*CASC20*	1
mmu-miR-181d-5p	*CASC20*	1
mmu-miR-190a-3p	*CASC20*	1
mmu-miR-1930-3p	*CASC20*	1

mmu-miR-1930-5p	*CASC20*	1
mmu-miR-1950	*CASC20*	1
mmu-miR-1955-5p	*CASC20*	1
mmu-miR-1956	*CASC20*	1
mmu-miR-1982-5p	*CASC20*	1
mmu-miR-203b-3p	*CASC20*	1
mmu-miR-205-3p	*CASC20*	1
mmu-miR-218-1-3p	*CASC20*	1
mmu-miR-218-2-3p	*CASC20*	1
mmu-miR-224-5p	*CASC20*	1
mmu-miR-292b-5p	*CASC20*	1
mmu-miR-294-5p	*CASC20*	1
mmu-miR-298-3p	*CASC20*	1
mmu-miR-29a-3p	*CASC20*	1
mmu-miR-29b-3p	*CASC20*	1
mmu-miR-29c-3p	*CASC20*	1
mmu-miR-3065-5p	*CASC20*	1
mmu-miR-3074-2-3p	*CASC20*	1
mmu-miR-3082-5p	*CASC20*	1
mmu-miR-3083-5p	*CASC20*	1
mmu-miR-3083b-5p	*CASC20*	1
mmu-miR-3094-3p	*CASC20*	1
mmu-miR-32-3p	*CASC20*	1
mmu-miR-320-3p	*CASC20*	1
mmu-miR-33-3p	*CASC20*	1
mmu-miR-338-5p	*CASC20*	1
mmu-miR-344b-3p	*CASC20*	1
mmu-miR-345-3p	*CASC20*	1
mmu-miR-3470a	*CASC20*	1
mmu-miR-3473f	*CASC20*	1
mmu-miR-3547-3p	*CASC20*	1
mmu-miR-369-3p	*CASC20*	1
mmu-miR-374b-5p	*CASC20*	1
mmu-miR-376c-3p	*CASC20*	1
mmu-miR-378d	*CASC20*	1
mmu-miR-381-5p	*CASC20*	1
mmu-miR-450a-1-3p	*CASC20*	1
mmu-miR-450b-3p	*CASC20*	1
mmu-miR-466a-5p	*CASC20*	1
mmu-miR-466e-5p	*CASC20*	1
mmu-miR-466h-5p	*CASC20*	1
mmu-miR-466j	*CASC20*	1

mmu-miR-466m-5p	*CASC20*	1
mmu-miR-466n-3p	*CASC20*	1
mmu-miR-466n-5p	*CASC20*	1
mmu-miR-466p-5p	*CASC20*	1
mmu-miR-466q	*CASC20*	1
mmu-miR-467b-3p	*CASC20*	1
mmu-miR-467c-3p	*CASC20*	1
mmu-miR-467d-3p	*CASC20*	1
mmu-miR-467e-3p	*CASC20*	1
mmu-miR-487b-5p	*CASC20*	1
mmu-miR-490-3p	*CASC20*	1
mmu-miR-495-3p	*CASC20*	1
mmu-miR-496b	*CASC20*	1
mmu-miR-5101	*CASC20*	1
mmu-miR-5107-3p	*CASC20*	1
mmu-miR-5110	*CASC20*	1
mmu-miR-5118	*CASC20*	1
mmu-miR-532-3p	*CASC20*	1
mmu-miR-543-3p	*CASC20*	1
mmu-miR-551b-5p	*CASC20*	1
mmu-miR-5627-5p	*CASC20*	1
mmu-miR-568	*CASC20*	1
mmu-miR-582-3p	*CASC20*	1
mmu-miR-592-5p	*CASC20*	1
mmu-miR-6340	*CASC20*	1
mmu-miR-6352	*CASC20*	1
mmu-miR-6370	*CASC20*	1
mmu-miR-6377	*CASC20*	1
mmu-miR-6481	*CASC20*	1
mmu-miR-6541	*CASC20*	1
mmu-miR-669d-2-3p	*CASC20*	1
mmu-miR-669h-5p	*CASC20*	1
mmu-miR-669l-3p	*CASC20*	1
mmu-miR-669m-3p	*CASC20*	1
mmu-miR-669m-5p	*CASC20*	1
mmu-miR-669n	*CASC20*	1
mmu-miR-669p-3p	*CASC20*	1
mmu-miR-6900-5p	*CASC20*	1
mmu-miR-6903-5p	*CASC20*	1
mmu-miR-6908-5p	*CASC20*	1
mmu-miR-691	*CASC20*	1
mmu-miR-6917-3p	*CASC20*	1

mmu-miR-6918-5p	*CASC20*	1
mmu-miR-6919-3p	*CASC20*	1
mmu-miR-6920-5p	*CASC20*	1
mmu-miR-6923-3p	*CASC20*	1
mmu-miR-6924-5p	*CASC20*	1
mmu-miR-6925-5p	*CASC20*	1
mmu-miR-6930-3p	*CASC20*	1
mmu-miR-6938-3p	*CASC20*	1
mmu-miR-6940-5p	*CASC20*	1
mmu-miR-6946-3p	*CASC20*	1
mmu-miR-6960-5p	*CASC20*	1
mmu-miR-6969-5p	*CASC20*	1
mmu-miR-6979-3p	*CASC20*	1
mmu-miR-6979-5p	*CASC20*	1
mmu-miR-698-5p	*CASC20*	1
mmu-miR-6981-5p	*CASC20*	1
mmu-miR-6982-5p	*CASC20*	1
mmu-miR-6988-5p	*CASC20*	1
mmu-miR-7001-3p	*CASC20*	1
mmu-miR-7004-3p	*CASC20*	1
mmu-miR-7014-5p	*CASC20*	1
mmu-miR-7021-5p	*CASC20*	1
mmu-miR-7030-5p	*CASC20*	1
mmu-miR-7037-5p	*CASC20*	1
mmu-miR-7054-3p	*CASC20*	1
mmu-miR-7056-5p	*CASC20*	1
mmu-miR-7062-5p	*CASC20*	1
mmu-miR-7072-5p	*CASC20*	1
mmu-miR-7075-5p	*CASC20*	1
mmu-miR-7076-5p	*CASC20*	1
mmu-miR-7078-3p	*CASC20*	1
mmu-miR-7078-5p	*CASC20*	1
mmu-miR-7081-5p	*CASC20*	1
mmu-miR-7084-5p	*CASC20*	1
mmu-miR-7091-5p	*CASC20*	1
mmu-miR-7093-5p	*CASC20*	1
mmu-miR-7220-3p	*CASC20*	1
mmu-miR-7227-3p	*CASC20*	1
mmu-miR-7227-5p	*CASC20*	1
mmu-miR-7232-5p	*CASC20*	1
mmu-miR-7239-5p	*CASC20*	1
mmu-miR-762	*CASC20*	1

mmu-miR-763	*CASC20*	1
mmu-miR-764-3p	*CASC20*	1
mmu-miR-764-5p	*CASC20*	1
mmu-miR-7648-3p	*CASC20*	1
mmu-miR-7656-3p	*CASC20*	1
mmu-miR-7660-5p	*CASC20*	1
mmu-miR-7664-3p	*CASC20*	1
mmu-miR-7668-3p	*CASC20*	1
mmu-miR-767	*CASC20*	1
mmu-miR-7675-3p	*CASC20*	1
mmu-miR-7679-5p	*CASC20*	1
mmu-miR-8100	*CASC20*	1
mmu-miR-8104	*CASC20*	1
mmu-miR-8118	*CASC20*	1
mmu-miR-873a-5p	*CASC20*	1
mmu-miR-875-3p	*CASC20*	1
mmu-miR-880-5p	*CASC20*	1
mmu-miR-881-3p	*CASC20*	1

7.2 R scripts

7.2.1 DEseq2 mRNA analysis

```
---
title: "DEseq2 mRNA mouse analysis"
output:
  word_document: default
  html_notebook: default
editor_options:
  chunk_output_type: inline
---
This notebook will perform DEseq2 analysis between CTRL and OE condition in the
mouse dataset.
Run DESeq and reduced LRT analysis.
```{r setup, include=FALSE}
library("tximport")
library("readr")
library("tximportData")
library("tidyverse")
library("DESeq2")
library("ggplot2")
library("biomaRt")
library("gplots")
library("tibble")
library("pheatmap")
library("EnhancedVolcano")
#load("X:/sudlab1/General/projects/Phoebe/mouse_analysis/RNAseq/salmonquant_results_mm11/ge
ne_tximport.RData")
load("/Users/phoebetamblinhopper/Downloads/gene_tximport.RData")
```
```

```
Create the required inputs for DESeq to run.
```{r coldata}
# Create the colData df and filter summarised_to_gene columns
vsamples <- colnames(summarised_to_gene$counts)
#remove D7 CTRL samples, as these are not for osteo analysis
vsamples <- vsamples[!(str_detect(vsamples, "D7"))]
vcondition <- factor(rep(c("CTRL","OE"), times = c(9, 9)))
vtimepoint <- factor(rep(c("D0", "D10", "D20", "D0", "D10", "D20"),
                         times = rep(3, 6)))
samples <- data.frame(
  condition = vcondition,
  timepoint = vtimepoint,
  row.names = vsamples)
#Remove D7 counts data from summarised_to_gene
filtered_summarised_to_gene <- lapply(summarised_to_gene,
  function(ma) {
    if (is.matrix(ma)) {
      ma <- ma[,vsamples]
      return(ma)
    } else {
return(ma) }
}) ```
Import data for analysis.
```{r import data in DESeq format}
ddsTxi <- DESeqDataSetFromTximport(filtered_summarised_to_gene,
                                   colData = samples,
                                   design = ~ condition + timepoint + condition:timepoint)
#Pre-filtering to remove any rows which add up to 10 or less
ddsTxi <- ddsTxi[rowSums(counts(ddsTxi)) >= 10,]
```
Run DESeq2
```{r run DEseq}
# Run DESeq and perform differential expression analysis.# Generate list of significant DE genes
between conditions and/or time points.
# Obtain results of genes that have a condition specific effect over time
ddsTxi <- DESeq(ddsTxi, test="LRT", reduced = ~ condition + timepoint)
resultsNames(ddsTxi)
# Obtain Results for genes that have a condition specific effect at Day 0
res <- results(ddsTxi)
res <- as.data.frame(res)
res <- tibble::rownames_to_column(res, "ensembl_gene_id")
# Look at table
head(res[order(res$padj),], 4)
```
Create Principle Component Analysis (PCA) plot
```{r PCA}
vsd <- vst(ddsTxi, blind=FALSE)
vsd_data <- assay(vsd) %>%
as.data.frame() %>%
rownames_to_column("ensembl_gene_id")
# Look at table
head(assay(vsd), 3)
vsd_PCA <- vst(ddsTxi, blind=FALSE,
               nsub = sum(rowMeans(counts(ddsTxi, normalized=TRUE)) > 5 ))
pcaData <- plotPCA(vsd, intgroup=c("condition", "timepoint"), returnData=TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))
jpeg("X:/sudlab1/General/projects/Phoebe/R outputs/plotPCA.png")
```

```
ggplot(pcaData, aes(PC1, PC2, color=timepoint, shape=condition, length=18)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ",percentVar[1],"% variance")) +
  ylab(paste0("PC2: ",percentVar[2],"% variance")) +
  coord_fixed()
dev.off()
```
```{r filtered PCA}
# Remove 2 outliers and replot
# By filtering VSD table
colnames(assay(vsd_PCA))
#Removes data of outliers
merged_matrix <- assay(vsd_PCA)[, - c(7,15)]
#head(merged_matrix)
pca <- prcomp(t(merged_matrix))
pca_df <- pca$x %>% data.frame() %>%
  rownames_to_column("Sample") %>%
  separate(Sample, c("Condition", "Timepoint"), sep = "_", remove = FALSE) %>%
  mutate(Condition = as.factor(Condition),
         Timepoint = as.factor(Timepoint)) %>%
  mutate(Sample = str_remove(Sample, "trimmed\\-"),
         Condition = str_remove(Condition, "trimmed\\-"))
#Proportional variance
percent <- round(100 * pca$sdev^2/sum(pca$sdev^2), 1)
set.seed(44)
jpeg("X:/sudlab1/General/projects/Phoebe/R outputs/plotPCA_filtered.png")
ggplot(pca_df, aes(PC1, PC2)) +
  geom_point(aes(col = Timepoint, shape = Condition), size = 3) +
  coord_fixed() +
  labs(x = paste0("PC1: ", percent[1], "%"),
       y = paste0("PC2: ", percent[2], "%")) +
  theme_light()
  #scale_color_manual(values=sample(palette.colors(palette = "Okabe-Ito", 3)))
dev.off()
```
Create gene ID table
```{r add ensembl IDs}
# Create table with ensembl gene IDs and names
# This table will be used later to join with res and add gene names
listEnsembl()
ensembl <- useEnsembl(biomart = "genes")
searchDatasets(mart = ensembl, pattern = "musculus")
ensembl <- useDataset(dataset = "mmusculus_gene_ensembl", mart = ensembl)
listAttributes(ensembl)
gene_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = res$ensembl_gene_id,
      mart = ensembl)
```
Create heatmap of DEGs
```{r heatmap}
#Create beta table for overall change over time between conditions
betas <- coef(ddsTxi)   #extract a matrix of the log2 fold changes
colnames(betas)
#add column with gene id to betas table (which is used to label heatmap)
betas <- as.data.frame(betas)
betas$ensembl_gene_id <- rownames(betas)
betas_genes <- inner_join(betas,
```

```
        gene_id_table)
#make gene name into row name for heatmap to use
betas_genes <- betas_genes %>%
  filter(!duplicated(external_gene_name))
betas_genes <- betas_genes %>%
  remove_rownames %>%
  column_to_rownames(var="external_gene_name")
betas_genes$ensembl_gene_id <- NULL #remove ensembl column
#plot log2 fold changes in heatmap
topGenes <- head(order(res$padj),100)
#Remove intercept and condition_OE_vs_CTRL ?
mat <- betas_genes[topGenes, -c(1,2)]
mat[is.na(mat)] <- 0 #to change NA values to 0
thr <- 5
mat[mat < -thr] <- -thr
mat[mat > thr] <- thr
```
```{r plotHMlfc, echo=FALSE}
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/OE_vs_CTRL.pdf") #to pdf heatmap need to
create file then
write to it
heatmap <- heatmap.2(as.matrix(mat), breaks=seq(from=-thr, to=thr, length=101),
        col = redblue(100),
        cexRow = 0.5,
        cexCol = 0.5,
        Colv = "FALSE",
        labCol = c("CTRL D10 vs D0", "CTRL D20 vs D0", "OE D10 vs D0", "OE D20 vs D0"),
        dendrogram = "row"
) dev.off()
#save unfiltered results as RData object
save(res, file = "X:/sudlab1/General/projects/Phoebe/R outputs/res_OE_vs_CTRL.RData")
```
Create table for GOSeq to use
```{r GO table}
#summary(res)
res_GOtable <- res %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.5, 1, 0))
summary(res_GOtable)
head(res_GOtable[order(res_GOtable$padj),])
#save filtered results as RData object
save(res_GOtable, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/resGO_OE_vs_CTRL.RData")
save(res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/nofilt_resGO_OE_vs_CTRL.RData")
```
Create other heatmaps
```{r sampleDists heatmap}
sampleDists <- dist(t(assay(vsd)))
names(as.matrix(sampleDists))
sampleDistMatrix <- as.matrix(sampleDists)
rownames(sampleDistMatrix) <- paste(vsd$condition, vsd$timepoint, sep="_")
colnames(sampleDistMatrix) <- NULL
#colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
jpeg("X:/sudlab1/General/projects/Phoebe/R outputs/plotDist.png")
pheatmap(sampleDistMatrix,
        clustering_distance_rows=sampleDists,
        clustering_distance_cols=sampleDists)
        #col=colors)
```

```
dev.off()
```
```{r vsd heatmap}
scaled_vsd <- t(scale(t(assay(vsd))))
# Filter only for genes that are significant
top_scaled <- res_GOtable %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
jpeg("X:/sudlab1/General/projects/Phoebe/R outputs/scaled_heatmap.png")
pheatmap(scaled_vsd[top_scaled,],
         clustering_distance_rows="euclidean",
         cluster_cols = FALSE,
         border_color = FALSE,
         show_rownames = FALSE,
         labels_col = c("CTRL_D0_1", "CTRL_D0_2", "CTRL_D0_3", "CTRL_D10_1", "CTRL_D10_2",
"CTRL_D10_3",
"CTRL_D20_1", "CTRL_D20_2", "CTRL_D20_3", "OE_D0_1", "OE_D0_2", "OE_D0_3", "OE_D10_1",
"OE_D10_2",
"OE_D10_3", "OE_D20_1", "OE_D20_2", "OE_D20_3"))
         #col=colors)
dev.off()
```
Look at genes with condition specific effect at D10
```{r}
#Obtain results for genes that have a condition specific effect at time D10 (or D20)
#Firstly D10
resultsNames(ddsTxi)
#At D10, difference between CTRL to OE
resD10 <- results(ddsTxi,
                  contrast = list(c("condition_OE_vs_CTRL", "conditionOE.timepointD10")),
                  test="Wald")
resD10 <- as.data.frame(resD10)
resD10 <- tibble::rownames_to_column(resD10, "ensembl_gene_id")
#Most significant DE gene
resD10[which.min(resD10$padj),]
#add column with gene id to res (which is used to label heatmap)
resD10_genes <- inner_join(resD10, gene_id_table)
#make gene name into row name for heatmap to use
resD10_genes <- resD10_genes %>%
  filter(!duplicated(external_gene_name))
resD10_genes <- resD10_genes %>%
  #remove_rownames %>%
  column_to_rownames(var="external_gene_name")
resD10_genes$ensembl_gene_id <- NULL #remove ensembl column
#save unfiltered results as RData object
save(resD10, file = "X:/sudlab1/General/projects/Phoebe/R outputs/resD10_OE_vs_CTRL.RData")
```
```{r}
#make new results table that is filtered for genes that have padj <0.05 and a log2fold change of
greater than
+/- 0.5 (this is a good starting point in terms of threshold)
summary(resD10)
resD10_GOtable <- resD10 %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.5, 1, 0))
summary(resD10_GOtable)
resD10_GOtable[order(resD10_GOtable$padj),]
#save filtered results as RData object
```

```
save(resD10_GOtable, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/resD10GO_OE_vs_CTRL.RData")
```
```{r}
#make vector of top 10 DEGs to label plot, as cannot fit labeling for all
#those that are DE
resD10_genes <- resD10_genes[order(resD10_genes$padj),]
resD10_genes_top10 <- rownames(resD10_genes)[1:10]
labels <- c(resD10_genes_top10, rep("", nrow(resD10_genes)-10))
#Make volcano plot for D10 DEGs
dev.off()
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/mRNA_osteo/volcano_D10_mRNA.pdf")
EnhancedVolcano(resD10_genes,
    lab = labels,
    x = 'log2FoldChange',
    y = 'padj',
    xlim = c(-3, 3),
    ylim = c(0, 6),
    axisLabSize = 18,
    #title = "Differentially expressed miRNAs with CASC20 overexpression at D10",
    titleLabSize = 10,
    subtitle = "",
    pCutoff = 0.05e-1,
    FCcutoff = 0.5,
    pointSize = 1,
    labSize = 5,
    legendLabels=c('Not sig.','LFC','p-value','p-value & LFC'),
    legendPosition = 'bottom',
    legendLabSize = 15,
    legendIconSize = 3,
    drawConnectors = TRUE,
    widthConnectors = 0.5,
    arrowheads = FALSE,
    max.overlaps = Inf)
#dev.off()

```
```

7.2.2 DESeq2 miRNA analysis

```
---
title: "DEseq2 miRNA mouse analysis"
output: html_notebook
editor_options:
  chunk_output_type: inline
---
```{r loadlibs, include=FALSE, message=FALSE}
library("tidyverse")
library("DESeq2")
library("ggplot2")
library("biomaRt")
library("tibble")
library("pheatmap")
library("purrr")
library("gplots")
library("EnhancedVolcano")
reading_dir <-
```

```
"X:/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/miRNA_pipeline/counts.dir/"
writing_dir <-
"X:/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/miRNA_pipeline"
#reading_dir <- "/Users/phoebetamblinhopper/Downloads/counts.dir"
#writing_dir <- "/Users/phoebetamblinhopper/Downloads"
```
```{r importCounts}
file_names <- list.files(reading_dir, "counts.tsv$", full.names = TRUE)
list_files <- lapply(file_names,
                     function(file) {
                     #Read file
                     df <-  read_delim(file,
                                       comment = "#",
                                       show_col_types = FALSE)
                     #Rename count column
                     name <- str_extract(file, "(?<=trimmed-).+(?=_counts\\.tsv)")
                     colnames(df)[ncol(df)] <- name
                     #Get rid of transcript in id
                     df <- df %>%
                       mutate(Geneid = str_remove(Geneid, "transcript\\:"))
                     #round up counts
                     df[, name] <- round(df[, name])
                     return(df)
}
                     )
names(list_files) <- str_extract(file_names, "(?<=trimmed-).+(?=_counts\\.tsv)")
#Merge_list
count_df <- purrr::reduce(list_files, full_join)
```
```{r}
#Filter count_df to remove extra columns not needed by DESeq2
count_df_filtered <- subset(count_df, select = -c(2,3,4,5,6))
count_df_filtered$Geneid <- str_remove(count_df_filtered$Geneid, "gene:")
count_df_filtered <- column_to_rownames(count_df_filtered, "Geneid")
``````{r coldata}
# Create the coldata df
vsamples <- names(list_files)
vcondition <- factor(rep(c("CTRL","OE"), times = c(9, 9)))
vtimepoint <- factor(rep(c("D0", "D10", "D20", "D0", "D10", "D20"),
                         times = rep(3, 6)))
samples <- data.frame(
  condition = vcondition,
  timepoint = vtimepoint,
  row.names = vsamples)
vsamples <- str_replace(vsamples, "__", "_")
```
```{r import data in DESeq format}
ddsMat <- DESeqDataSetFromMatrix(countData = count_df_filtered,
                                 colData = samples,
                                 design = ~ condition + timepoint + condition:timepoint)
#Pre-filtering to remove any rows which add up to 10 or less
ddsMat <- ddsMat[rowSums(counts(ddsMat)) >= 10,]
```
```{r run DESeq2}
# Run DESeq and perform differential expression analysis.
# Generate list of significant DE genes between conditions and/or time points.
# Obtain results of genes that have a condition specific effect over time
ddsMat <- DESeq(ddsMat, test="LRT", reduced = ~ condition + timepoint)
```
```

```
resultsNames(ddsMat)
```
######
Overall
######
```{r}
# Obtain Results for genes that have a condition specific effect at Day 0
res <- results(ddsMat)
res <- as.data.frame(res)
res <- tibble::rownames_to_column(res, "ensembl_gene_id")
head(res[order(res$padj),], 4)
```
```{r}
#for vst function default nsub is 1000, but less genes than that here
#sum( rowMeans( counts(ddsTxi, normalized=TRUE)) > 5 ) gives 462
#so set nsub to this
vsd <- vst(ddsMat, blind=FALSE, nsub=462)
vsd_data <- assay(vsd) %>%
as.data.frame() %>%
rownames_to_column("ensembl_gene_id")
head(assay(vsd), 3)
vsd_PCA <- vst(ddsMat, blind=FALSE,
               nsub = sum(rowMeans(counts(ddsMat, normalized=TRUE)) > 5 ))
pcaData <- plotPCA(vsd, intgroup=c("condition", "timepoint"), returnData=TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))
jpeg("X:/sudlab1/General/projects/Phoebe/R outputs/miRNA_osteo/miRNA_osteo_plotPCA.png")
ggplot(pcaData, aes(PC1, PC2, color=timepoint, shape=condition, length=18)) +
  geom_point(size=3) +
  xlab(paste0("PC1: ",percentVar[1],"% variance")) +
  ylab(paste0("PC2: ",percentVar[2],"% variance")) +
  coord_fixed()
dev.off()
```
```{r add ensembl IDs}
# Create table with ensembl gene IDs and names
# This table will be used later to join with res and add gene names
listEnsembl()
ensembl <- useEnsembl(biomart = "genes")
searchDatasets(mart = ensembl, pattern = "musculus")
ensembl <- useDataset(dataset = "mmusculus_gene_ensembl", mart = ensembl)
listAttributes(ensembl)
miR_gene_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = res$ensembl_gene_id,
      mart = ensembl)
```
```{r heatmap}
#Create beta table for overall change over time between conditions
betas <- coef(ddsMat)   #extract a matrix of the log2 fold changes
colnames(betas)
#add column with gene id to betas table (which is used to label heatmap)
betas <- as.data.frame(betas)
betas$ensembl_gene_id <- rownames(betas)
betas_miRNAs <- inner_join(betas,
          miR_gene_id_table)
#make gene name into row name for heatmap to use
betas_miRNAs <- betas_miRNAs %>%
  filter(!duplicated(external_gene_name))
```

```
betas_miRNAs <- betas_miRNAs %>%
  remove_rownames %>%
  column_to_rownames(var="external_gene_name")
betas_miRNAs$ensembl_gene_id <- NULL #remove ensembl column
#plot log2 fold changes in heatmap
topGenes <- head(order(res$padj),50)
#Remove intercept and condition_OE_vs_CTRL ?
mat <- betas_miRNAs[topGenes, -c(1,2)]
mat[is.na(mat)] <- 0 #to change NA values to 0
thr <- 5
mat[mat < -thr] <- -thr
mat[mat > thr] <- thr
```
```{r}
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/miRNA_osteo/OE_vs_CTRL.pdf",
width=210, height=297)
heatmap <- heatmap.2(as.matrix(mat), breaks=seq(from=-thr, to=thr, length=101),
        col = redblue(100),
        cexRow = 0.5,
        cexCol = 0.5,
        Colv = "FALSE",
        labCol = c("CTRL D10 vs D0", "CTRL D20 vs D0", "OE D10 vs D0", "OE D20 vs D0"),
        dendrogram = "row"
) dev.off()
```
```{r}
(res[order(res$padj),])
res_GOtable_miRNAs <- res %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.5,
1, 0))
summary(res_GOtable_miRNAs)
head(res_GOtable_miRNAs[order(res_GOtable_miRNAs$padj),])
#save filtered results as RData object
save(res_GOtable_miRNAs, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL.RData")
res_GOtable_miRNAs[order(res_GOtable_miRNAs$padj),]
#Same but with more stringest filter for LFC - 1
res_GOtable_miRNAs_LFC1 <- res %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 1,
1, 0))
save(res_GOtable_miRNAs_LFC1, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL_LFC1.RData")
#Same but with less stringest filter for LFC - 0.1
res_GOtable_miRNAs_LFC0.1 <- res %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.1,
1, 0))
save(res_GOtable_miRNAs_LFC0.1, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL_LFC0.1.RData")
```
######
Day 10
######
```{r}
#Obtain results for genes that have a condition specific effect at time D10
resultsNames(ddsMat)
#At D10, difference between CTRL to OE
resD10 <- results(ddsMat,
                  contrast = list(c("condition_OE_vs_CTRL",


```

```
"conditionOE.timepointD10")),
                test="Wald")
resD10 <- as.data.frame(resD10)
resD10 <- tibble::rownames_to_column(resD10, "ensembl_gene_id")
#Most significant DE gene
#resD10[which.min(resD10$padj),]
#Make gene ID table for D10
miR_gene_id_table_D10 <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = resD10$ensembl_gene_id,
      mart = ensembl)
#add column with gene id to res (which is used to label heatmap)
resD10_miRNAs <- inner_join(resD10, miR_gene_id_table_D10)
#make gene name into row name for heatmap to use
resD10_miRNAs <- resD10_miRNAs %>%
  filter(!duplicated(external_gene_name))
resD10_miRNAs <- resD10_miRNAs %>%
  #remove_rownames %>%
  column_to_rownames(var="external_gene_name")
resD10_miRNAs$ensembl_gene_id <- NULL #remove ensembl column
#save unfiltered results as RData object
save(resD10, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10_miR_OE_vs_CTRL.RData")
```
```{r}
#plot log2 fold changes in heatmap
topGenes <- head(order(resD10$padj),20)
#Remove intercept and condition_OE_vs_CTRL ?
mat_D10 <- betas_miRNAs[topGenes, -c(1,2)]
mat_D10[is.na(mat_D10)] <- 0 #to change NA values to 0
thr <- 5
mat_D10[mat_D10 < -thr] <- -thr
mat_D10[mat_D10 > thr] <- thr
#to pdf heatmap need to create file then write to it
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/miRNA_osteo/OE_vs_CTRL_D10_top20.pdf")
heatmap <- heatmap.2(as.matrix(mat_D10), breaks=seq(from=-thr, to=thr, length=101),
        col = redblue(100),
        cexRow = 0.5,
        cexCol = 0.5,
        Colv = "FALSE",
        labCol = c("CTRL D10 vs D0", "CTRL D20 vs D0", "OE D10 vs D0", "OE D20 vs D0"),
        dendrogram = "row"
) dev.off()
```
```{r}
summary(resD10)
head(resD10[order(resD10$padj),])
resD10_GOtable <- resD10 %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.5,
1, 0))
#test$Significant == resD10_GOtable$Significant
summary(resD10_GOtable)
resD10_GOtable[order(resD10_GOtable$padj),]
#save filtered results as RData object
save(resD10_GOtable, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_miR_OE_vs_CTRL.RData")
#Same but with more stringest filter for LFC - 1
resD10_GOtable_LFC1 <- resD10 %>%
```

```
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 1,
1, 0))
save(resD10_GOtable_LFC1, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_OE_vs_CTRL_LFC1.RData")
#Same but with less stringest filter for LFC - 0.1
resD10_GOtable_LFC0.1 <- resD10 %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.1,
1, 0))
save(resD10_GOtable_LFC0.1, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_OE_vs_CTRL_LFC0.1.RData")
```
######
Day 20
######
```{r D20 results}
#Obtain results for genes that have a condition specific effect at time D20
resultsNames(ddsMat)
#At D20, difference between CTRL to OE
resD20 <- results(ddsMat,
                  contrast = list(c("condition_OE_vs_CTRL",
"conditionOE.timepointD20")),
                  test="Wald")
resD20 <- as.data.frame(resD20)
resD20 <- tibble::rownames_to_column(resD20, "ensembl_gene_id")
#Most significant DE gene
#resD20[which.min(resD20$padj),]
#Make gene ID table for D20
miR_gene_id_table_D20 <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = resD20$ensembl_gene_id,
      mart = ensembl)
#add column with gene id to res (which is used to label heatmap)
resD20_miRNAs <- inner_join(resD20, miR_gene_id_table)
#make gene name into row name for heatmap to use
resD20_miRNAs <- resD20_miRNAs %>%
  filter(!duplicated(external_gene_name))
resD20_miRNAs <- resD20_miRNAs %>%
  #remove_rownames %>%
  column_to_rownames(var="external_gene_name")
resD20_miRNAs$ensembl_gene_id <- NULL #remove ensembl column
#save unfiltered results as RData object
save(resD20, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD20_miR_OE_vs_CTRL.RData")
```
```{r D20 heatmap}
#plot log2 fold changes in heatmap
topGenes <- head(order(resD20$padj),20)
#Remove intercept and condition_OE_vs_CTRL ?
mat_D20 <- betas_miRNAs[topGenes, -c(1,2)]
mat_D20[is.na(mat_D20)] <- 0 #to change NA values to 0
thr <- 5
mat_D20[mat_D20 < -thr] <- -thr
mat_D20[mat_D20 > thr] <- thr
#to pdf heatmap need to create file then write to it
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/miRNA_osteo/OE_vs_CTRL_D20_top20.pdf")
heatmap <- heatmap.2(as.matrix(mat_D20), breaks=seq(from=-thr, to=thr, length=101),
D0"),
col = redblue(100),
```

```
cexRow = 0.5,
cexCol = 0.5,
Colv = "FALSE",
#labCol = c("CTRL D10 vs D0", "CTRL D20 vs D0", "OE D10 vs D0", "OE D20 vs
dendrogram = "row"
) dev.off()
```
```{r D20 create GO tables}
summary(resD20)
head(resD20[order(resD20$padj),])
resD20_GOtable <- resD20 %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.5,
1, 0))
summary(resD20_GOtable)
resD20_GOtable[order(resD20_GOtable$padj),]
#save filtered results as RData object
save(resD20_GOtable, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD20GO_miR_OE_vs_CTRL.RData")
#Same but with more stringest filter for LFC - 1
resD20_GOtable_LFC1 <- resD20 %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 1,
1, 0))
save(resD20_GOtable_LFC1, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD20GO_OE_vs_CTRL_LFC1.RData")
#Same but with less stringest filter for LFC - 0.1
resD20_GOtable_LFC0.1 <- resD20 %>%
  mutate(Significant = if_else((padj < 0.05 & !is.na(padj)) & abs(log2FoldChange) > 0.1,
1, 0))
save(resD20_GOtable_LFC0.1, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD20GO_OE_vs_CTRL_LFC0.1.RData")
```
Create volcano plots
```{r volcano overall}
#Make volcano plot for overall
res_miRNAs <- inner_join(res, miR_gene_id_table)
#dev.off()
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/miRNA_osteo/volcano_overall_miRNAs.pdf")
EnhancedVolcano(res_miRNAs,
    lab = res_miRNAs$external_gene_name,
    x = 'log2FoldChange',
    y = 'padj',
    xlim = c(-3, 3),
    ylim = c(0, 20),
    axisLabSize = 18,
    #title = "Differentially expressed miRNAs with CASC20 overexpression at D10",
    titleLabSize = 10,
    subtitle = "",
    pCutoff = 0.05,
    FCcutoff = 0.5,
    pointSize = 1,
    labSize = 4,
    legendLabels=c('Not sig.','LFC','p-value','p-value & LFC'),
    legendPosition = 'bottom',
    legendLabSize = 15,
    legendIconSize = 3,
    drawConnectors = TRUE,
    widthConnectors = 0.5,
    arrowheads = FALSE,
```

```
    max.overlaps = Inf)
```
```{r volcano D10}
#make vector of top 10 DE miRNAs to label plot, as cannot fit labelling for all
#those that are DE
resD10_miRNAs <- resD10_miRNAs[order(resD10_miRNAs$padj),]
resD10_miRNAs_top10 <- rownames(resD10_miRNAs)[1:10]
labels <- c(resD10_miRNAs_top10, rep("", nrow(resD10_miRNAs)-10))
#Make volcano plot for D10
#dev.off()
pdf("X:/sudlab1/General/projects/Phoebe/R outputs/miRNA_osteo/volcano_D10_miRNAs.pdf")
EnhancedVolcano(resD10_miRNAs,
    lab = labels,
    x = 'log2FoldChange',
    y = 'padj',
    xlim = c(-3, 3),
    ylim = c(0, 20),
    axisLabSize = 18,
    #title = "Differentially expressed miRNAs with CASC20 overexpression at D10",
    titleLabSize = 10,
    subtitle = "",
    pCutoff = 0.05,
    FCcutoff = 0.5,
    pointSize = 1,
    labSize = 4,
    legendLabels=c('Not sig.','LFC','p-value','p-value & LFC'),
    legendPosition = 'bottom',
    legendLabSize = 15,
    legendIconSize = 3,
    drawConnectors = TRUE,
    widthConnectors = 0.5,
    arrowheads = FALSE,
    max.overlaps = Inf)
#dev.off()

```
```

7.2.3 GOSeq mRNA analysis

```
---
title: "GOSeq Mouse mRNA Analysis"
output: html_notebook
---
```{r, include=FALSE, message=FALSE}
library(tidyverse)
library(DESeq2)
library("goseq")
#library("org.Mm.eg.db")
library("TxDb.Mmusculus.UCSC.mm39.knownGene")
library("biomaRt")
```
```{r GOseq, echo=TRUE}
#For overall diff OE vs CTRL
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/resGO_OE_vs_CTRL.RData")
#load(file = "/Users/phoebetamblinhopper/Downloads/resGO_OE_vs_CTRL_mRNA.RData")
```

```
#res_GOtable <- subset(res, log2FoldChange > 1 | log2FoldChange < -1)
ALL <- res_GOtable$ensembl_gene_id
DE <- res_GOtable %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
gene.vector <- as.integer(ALL %in% DE)
names(gene.vector) <- ALL
#using base mean for bias instead of length, as saw no pattern with length
#create vector to use as bias.data
gene.basemean <- subset(res_GOtable, select = c(1,2))
gene.basemean <- setNames(as.numeric(gene.basemean$baseMean),
gene.basemean$ensembl_gene_id)
#nullp fits a model to account for gene length biases in data
#essentially the probablity a gene will be DE based on length alone
#Mind warning here
pwf = nullp(gene.vector,
          bias.data = gene.basemean)
plotPWF(pwf)
```
Fetching GO from biomart as it seems DB of Goseq is not as complete (almost 5000 genes without annotations)
```{r}
#Create a list of all possible GO terms
ensembl = useMart("ensembl",dataset="mmusculus_gene_ensembl")
pre.gene.data <- getBM(attributes=c('ensembl_gene_id', 'go_id'),
                 filters = 'ensembl_gene_id',
                 values = ALL,
                 mart = ensembl)
``````{r}
#default wallenius method
GO.wall = goseq(pwf,"mm39","ensGene",
              gene2cat = pre.gene.data)
head(GO.wall)
#Correction for multiple testing using BH method
GO.wall$over_represented_adj_pvalue <- p.adjust(GO.wall$over_represented_pvalue,
                                    method="BH")
save(GO.wall, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/GO_wall_overall.Rdata")
```
```{r}
#For sampling method, as pwf plot showed negative correlation with gene length,
#when plotted without filtering for log2FC
GO.samp = goseq(pwf,"mm39","ensGene",
              gene2cat = pre.gene.data,
              method = "Sampling",
              repcnt = 20000)
head(GO.samp)
#Correction for multiple testing using BH method
GO.samp$over_represented_adj_pvalue <- p.adjust(GO.wall$over_represented_pvalue,
                                    method="BH")
save(GO.samp, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/GO_samp20000_overall.Rdata")
```
```{r}
GO.hyp = goseq(pwf,"mm39","ensGene",
              gene2cat = pre.gene.data,
              method = "Hypergeometric")
head(GO.hyp)
```
```

```
#Correction for multiple testing using BH method
GO.hyp$over_represented_adj_pvalue <- p.adjust(GO.wall$over_represented_pvalue,
```
```{r GOplot, echo=FALSE}
go_res_plot <- GO.wall %>%
    arrange(over_represented_adj_pvalue) %>%
  head(10) %>%
  mutate(log_p_adj = -log10(over_represented_adj_pvalue)) %>%
    mutate(hitsPerc=numDEInCat*100/numInCat)
GO_overall <- go_res_plot %>%
   dplyr::arrange(log_p_adj) %>%
  mutate(term=factor(term , levels= go_res_plot$term)) %>%
    ggplot(aes(x=hitsPerc,
method="BH")
              y=term,
              colour=-log(over_represented_adj_pvalue),
              size=numDEInCat)) +
       geom_point() +
       expand_limits(x=0) +
      labs(x="Hits (%)", y="GO term", colour="-log(p-value)", size="Count")    +
       guides(colour = guide_legend(order = 1),
            size = guide_legend(order = 1)) +
      theme_minimal(base_size = 10) +
       scale_y_discrete(labels = str_wrap(GO.wall$term, width = 50))
ggsave("X:/sudlab1/General/projects/Phoebe/R outputs/Overall_GOSeq_Sampling_20000.pdf")
```
```{r GOseq D10}
#Create GO plot in same way for D10, OE vs CTRL
#For overall diff OE vs CTRL
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/resD10GO_OE_vs_CTRL.RData")
#load(file = "/Users/phoebetamblinhopper/Downloads/resD10GO_OE_vs_CTRL.RData")
ALL <- resD10_GOtable$ensembl_gene_id
DE <- resD10_GOtable %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#To look at how many genes are DE
DE_LFC <- resD10_GOtable %>%
  filter(Significant == 1) %>%
  pull(log2FoldChange, ensembl_gene_id)
DE_LFC <- as.data.frame(DE_LFC)
#count upregulated
sum(DE_LFC > 0.5)
#count downregulated
sum(DE_LFC < -0.5)
gene.vector <- as.integer(ALL %in% DE)
names(gene.vector) <- ALL
#nullp fits a model to account for gene length biases in data
pwf = nullp(gene.vector,"mm39","ensGene")
plotPWF(pwf)
GO.wall = goseq(pwf,"mm39","ensGene",
               gene2cat = pre.gene.data)
head(GO.wall)
#Correction for multiple testing using BH method
GO.wall$over_represented_adj_pvalue <- p.adjust(GO.wall$over_represented_pvalue,
                                        method="BH")
```
```{r GOplot D10, echo=FALSE}
```

```
#Create plot for D10
go_res_plot <- GO.wall %>%
   arrange(over_represented_adj_pvalue) %>%
  head(10) %>%
  mutate(log_p_adj = -log10(over_represented_adj_pvalue)) %>%
   mutate(hitsPerc=numDEInCat*100/numInCat)
GO_D10 <- go_res_plot %>%
   dplyr::arrange(log_p_adj) %>%
  mutate(term=factor(term , levels= go_res_plot$term)) %>%
   ggplot(aes(x=hitsPerc,
              y=term,
              colour=log_p_adj,
              size=numDEInCat)) +
      geom_point() +
      expand_limits(x=0) +
    labs(x="Hits (%)", y="GO term", colour="-log(adj. p-value)", size="Count")    +
      guides(colour = guide_legend(order = 1),
           size = guide_legend(order = 1)) +
    theme_minimal(base_size = 10) +
      scale_y_discrete(labels = str_wrap(go_res_plot$term, width = 50)) +
    theme(axis.text.x=element_text(size=14),
          axis.text.y=element_text(size=14),
          axis.title.x=element_text(size=14),
          axis.title.y=element_text(size=14),
          legend.title = element_text(size=12),
          legend.text = element_text(size=12))
      #+
    #scale_color_gradient(low="black", high="red")
ggsave("X:/sudlab1/General/projects/Phoebe/R outputs/mRNA_osteo/D10_GOSeq.png")
```
```{r}
#For sampling method
GO.samp = goseq(pwf,"mm39","ensGene",
              gene2cat = pre.gene.data,
              method = "Sampling",
              repcnt = 20000)
head(GO.samp)
#Correction for multiple testing using BH method
GO.samp$over_represented_adj_pvalue <- p.adjust(GO.wall$over_represented_pvalue,
                                      method="BH")
save(GO.samp, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/GO_samp20000_overall.Rdata")
#find gene IDs DE in cat
getGenes <- function(pwf, goterm, genome, ids){
   gene2cat <-  getgo(rownames(pwf), genome, ids)
   cat2gene <- split(rep(names(gene2cat), sapply(gene2cat, length)),
                  unlist(gene2cat, use.names = FALSE))
   out <- pwf[cat2gene[[goterm]],]
   out <- out[out$DEgenes > 0,]
out }
example(goseq)
goterm <- pvals$category[1]
getGenes(pwf, "GO:0005684", "mm39","ensGene")

```

7.2.4 GOSeq miRNA analysis

````
---
title: "GOSeq analysis miRNA"
output: html_notebook
editor_options:
  chunk_output_type: inline
---
```{r loadlibs, include=FALSE, message=FALSE}
library(tidyverse)
library(DESeq2)
library("goseq")
library("TxDb.Mmusculus.UCSC.mm39.knownGene")
library("biomaRt")
library("readr")
library("clusterProfiler")
#miRDB gene sets downloaded from GSEA
#GSEA <- read.gmt("X:/sudlab1/General/projects/Phoebe/m3.mirdb.v2024.1.Mm.symbols.gmt")
GSEA <- read.gmt("/Users/phoebetamblinhopper/Downloads/m3.mirdb.v2024.1.Mm.symbols.gmt")
```
```{r}
#load DESeq results file
#load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL.RData")
load(file = "/Users/phoebetamblinhopper/Downloads/resGO_OE_vs_CTRL.RData")
ALL <- res_GOtable_miRNAs$ensembl_gene_id
DE_miRNAs <- res_GOtable_miRNAs %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#Loading GSEA with ensembl IDs
#Code below commented was used to create this file
#Can load file instead of running code
#load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GSEA_db_miRNAs_targets.RData")
load(file = "/Users/phoebetamblinhopper/Downloads/GSEA_db_miRNAs_targets.RData")
```
```{r}
#Code commented as not required if load GSEA db file above
#Make gene ID table for miRNAs to convert gene names to ensembl IDs
# listEnsembl()
# ensembl <- useEnsembl(biomart = "genes")
# searchDatasets(mart = ensembl, pattern = "musculus")
# ensembl <- useDataset(dataset = "mmusculus_gene_ensembl", mart = ensembl)
# listAttributes(ensembl)
#
# miR_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
#       filters = "ensembl_gene_id",
#       values = res_GOtable_miRNAs$ensembl_gene_id,
#       mart = ensembl)
#
#
# #Same for gene IDs of miR targets
## gene_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
#       filters = "external_gene_name",
#       values = GSEA$gene,
#       mart = ensembl)
```
````

```{r}
#Code commented as not required if load GSEA db file above
#Edit naming of miRNAs to match miR ID table
# GSEA_db <- GSEA %>%
#   mutate(
#     term = str_replace_all(term, "_", "-")
#   )  %>%
# #Separate miRNAs that are together onto separate lines
#
#
#
# separate_longer_delim(term, "SPLIT") %>%
mutate(
  term = str_replace_all(term, "-LET-", "SPLITLET")
)%>%
#
#
#
#
#
#
# separate_longer_delim(term, "SPLIT") %>%
mutate(
  term = str_replace_all(term, "LET-", "LET")
)%>% mutate(
term = str_replace_all(term, "-MIR-", "SPLITMIR") )%>%
#
#
#
# #Edit naming to include mir as in gene ID table
mutate(
  term = str_replace_all(term, "MIR-", "MIR")
)%>%
#
#
#
# #Remove P and number before
#   mutate(
mutate(
  term = str_replace_all(term, "LET", "MIRLET")
)%>%
#     term = str_replace(term, "-.P", "")
# )%>%
#   #Make all names lowercase to match
# mutate(term = tolower(term))
#
# #Same to lower case
# miR_id_table$external_gene_name <- tolower(miR_id_table$external_gene_name)
#
# #Rename column so matches gene ID table
# GSEA_db <- GSEA_db %>%
# rename(
# external_gene_name = term
#)
#
# #Save table
# save(GSEA_db, miR_id_table, gene_id_table, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GSEA_db_miRNAs_targets.RData")
```

```
```{r}
#Code commented as not required if load GSEA db file above
#Join GSEA db with correct naming to miR and gene ID tables
GSEA_db_ID <- inner_join(GSEA_db,
miR_id_table)
GSEA_db_ID <- inner_join(GSEA_db_ID,
          gene_id_table, by = join_by("gene" == "external_gene_name")) %>%
distinct()
#Rename columns to be more explanatory
colnames(GSEA_db_ID) <- c("miRNA", "gene_target", "miRNA_ensembl_id",
                          "gene_ensembl_id")
#Add base mean to table
filtered_res_GOtable_miRNAs <- subset(res_GOtable_miRNAs,
                                      select = c("ensembl_gene_id", "baseMean"))
GSEA_db_ID <- inner_join(GSEA_db_ID,
                         filtered_res_GOtable_miRNAs,
                         by = join_by("miRNA_ensembl_id" == "ensembl_gene_id"))
#Summarise table to check miRNAs all there and joined correctly
#Counts number of unique values in column
n_distinct(GSEA_db$external_gene_name)
n_distinct(miR_id_table$external_gene_name)
n_distinct(GSEA_db_ID$miRNA)
#775-623
#152 missing
all(miR_id_table$external_gene_name %in% GSEA_db$external_gene_name)
#Look at missing miRNAs
subset(miR_id_table, !(external_gene_name %in% GSEA_db_ID$miRNA))
```
```{r}
nrow(GSEA_db_ID) == nrow(distinct(GSEA_db_ID))
all(res_GOtable_miRNAs$ensembl_gene_id %in% GSEA_db_ID$miRNA_ensembl_id)
all(GSEA_db_ID$miRNA_ensembl_id %in% res_GOtable_miRNAs$ensembl_gene_id)
all(GSEA_db_ID$miRNA_ensembl_id %in% ALL)
all(DE_miRNAs %in% GSEA_db_ID$miRNA_ensembl_id)
#Filter miRNA with no targets
DE_miRNAs <- DE_miRNAs[DE_miRNAs %in%  GSEA_db_ID$miRNA_ensembl_id]
GSEA_db_ID %>%
   filter(miRNA_ensembl_id %in% DE_miRNAs)
```
```{r functions, echo=FALSE}
runGOseq <- function(miRNA, df_ids) {
  "Run GOseq analysis for a miRNA or a vector list of miRNA"
  #Extract data for miRNA
  df_filtered <- df_ids[df_ids$miRNA_ensembl_id %in% miRNA,]
  #Create DE vector
  DE.vector <- unique(df_filtered$gene_ensembl_id)
  #Create the ALL vector with 0...11 if DE or not
  all.vector <- unique(df_ids$gene_ensembl_id)
  #Gene vector
  gene.vector <- as.integer(all.vector %in% DE.vector)
  names(gene.vector) <- all.vector
    #Create the bias vector
  bias.vector <- df_ids %>%
    group_by(gene_ensembl_id) %>%
    summarise(meanCount = mean(baseMean))
  print(all(names(gene.vector) %in% bias.vector$gene_ensembl_id))
  gene.basemean <- as.vector(bias.vector$meanCount)
```
```

```
  names(gene.basemean) <- bias.vector$gene_ensembl_id
  #Run the pwf and GO
  pwf = nullp(gene.vector,
          bias.data = gene.basemean)
  GO.hyp = goseq(pwf,"mm39","ensGene",
              method = "Hypergeometric")
  #BH correction
  GO.hyp$over_represented_adj_pvalue <- p.adjust(GO.hyp$over_represented_pvalue,
                                       method="BH")
  #make plot
  go_res_plot <- GO.hyp %>%
    arrange(over_represented_adj_pvalue) %>%
  head(10) %>%
  mutate(log_p_adj = -log10(over_represented_adj_pvalue)) %>%
    mutate(hitsPerc=numDEInCat*100/numInCat)
  GO_plot <- go_res_plot %>%
   dplyr::arrange(log_p_adj) %>%
  mutate(term=factor(term , levels= go_res_plot$term)) %>%
    ggplot(aes(x=hitsPerc,
               y=term,
               colour=-log(over_represented_adj_pvalue),
               size=numDEInCat)) +
       geom_point() +
       expand_limits(x=0) +
     labs(x="Hits (%)", y="GO term", colour="-log(p-value)", size="Count")    +
       guides(colour = guide_legend(order = 1),
            size = guide_legend(order = 1)) +
     theme_minimal(base_size = 10) +
       scale_y_discrete(labels = str_wrap(GO.hyp$term, width = 50)) +
     theme(axis.text.x=element_text(size=14),
           axis.text.y=element_text(size=14),
           axis.title.x=element_text(size=14),
           axis.title.y=element_text(size=14),
           legend.title = element_text(size=12),
           legend.text = element_text(size=12))
   return(list(GO.hyp, GO_plot))
}
```

```{r, overall LFC 0.5}
#run using function created
#for overall GOseq, LFC threshold of 0.5
miRNA_overall_LFC0.5 <- runGOseq(DE_miRNAs,
        GSEA_db_ID)
#Save outputs
miRNA_overall_LFC0.5_res <- miRNA_overall_LFC0.5[[1]]
save(miRNA_overall_LFC0.5_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_overall_LFC0.5.RData")
miRNA_overall_LFC0.5[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_overall_LFC0.5.png")
  #Create DE vector
  DE.vector <- unique(df_filtered$gene_ensembl_id)
#Check how many unique gene targets there are
check <- GSEA_db_ID[GSEA_db_ID$miRNA_ensembl_id %in% DE_miRNAs,]
n_distinct(check$gene_target)
#############
#ADDED LATER
#check if any of the gene targets were also identified as significant DEGs
```

```
#load list of DEGs from DESeq2 analysis
load(file = "/Users/phoebetamblinhopper/Downloads/resGO_OE_vs_CTRL.RData")
#filter list only for significant DEGs
sig_res <- subset.data.frame(res_GOtable, res_GOtable$Significant == 1)
#turn this into a list of gene IDs
DEG_IDs <- sig_res$ensembl_gene_id
#turn list of gene targets into list of gene IDs
miR_targets <- check$gene_ensembl_id
#look for gene IDs in common between 2 lists
DEGs_targeted_by_DE_miRNAs_overall <- intersect(DEG_IDs, miR_targets)
DEGs_targeted_by_DE_miRNAs_overall <- as.data.frame(DEGs_targeted_by_DE_miRNAs_overall)
colnames(DEGs_targeted_by_DE_miRNAs_overall) <- c("ensembl_gene_id")
#join with gene ID table to get gene names
DEGs_targeted_by_DE_miRNAs_overall_ID <- inner_join(
                        DEGs_targeted_by_DE_miRNAs_overall, gene_id_table
                          %>% distinct)
#join with GO table for other info from analysis
DEGs_targeted_by_DE_miRNAs_overall_full <-
inner_join(DEGs_targeted_by_DE_miRNAs_overall_ID,
res_GOtable)
```
```{r, D10 LFC 0.5}
#for D10 GOSeq, LFC threshold of 0.5
#load DESeq results file
#load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_miR_OE_vs_CTRL.RData")
load(file = "/Users/phoebetamblinhopper/Downloads/resD10GO_miR_OE_vs_CTRL.RData")
DE_miRNAs_D10 <- resD10_GOtable %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#Filter miRNA with no targets
DE_miRNAs_D10 <- DE_miRNAs_D10[DE_miRNAs_D10 %in%  GSEA_db_ID$miRNA_ensembl_id]
GSEA_db_ID %>%
    filter(miRNA_ensembl_id %in% DE_miRNAs_D10)
#run function
miRNA_D10_LFC0.5 <- runGOseq(DE_miRNAs_D10,
        GSEA_db_ID)
#Save outputs
miRNA_D10_LFC0.5_res <- miRNA_D10_LFC0.5[[1]]
save(miRNA_D10_LFC0.5_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_D10_LFC0.5.RData")
miRNA_D10_LFC0.5[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_D10_LFC0.5.png")
#Check how many unique gene targets there are
check_D10 <- GSEA_db_ID[GSEA_db_ID$miRNA_ensembl_id %in% DE_miRNAs_D10,]
n_distinct(check_D10$gene_target)
#############
#ADDED LATER
#check if any of the gene targets were also identified as significant DEGs
#load list of DEGs from DESeq2 analysis
load(file = "/Users/phoebetamblinhopper/Downloads/resD10GO_OE_vs_CTRL.RData")
#filter list only for significant DEGs
sig_resD10 <- subset.data.frame(resD10_GOtable, resD10_GOtable$Significant == 1)
#turn this into a list of gene IDs
DEG_IDs_D10 <- sig_resD10$ensembl_gene_id
#turn list of gene targets into list of gene IDs
miR_targets_D10 <- check_D10$gene_ensembl_id
```

```
#look for gene IDs in common between 2 lists
DEGs_targeted_by_DE_miRNAs <- intersect(DEG_IDs_D10, miR_targets_D10)
DEGs_targeted_by_DE_miRNAs <- as.data.frame(DEGs_targeted_by_DE_miRNAs)
colnames(DEGs_targeted_by_DE_miRNAs) <- c("ensembl_gene_id")
#join with gene ID table to get gene names
DEGs_targeted_by_DE_miRNAs_ID <- inner_join(DEGs_targeted_by_DE_miRNAs, gene_id_table
                    %>% distinct)
#join with GO table for other info from analysis
DEGs_targeted_by_DE_miRNAs_full <- inner_join(DEGs_targeted_by_DE_miRNAs_ID,
resD10_GOtable)
```
```{r, D20 LFC 0.5}
#for D20 GOSeq, LFC threshold of 0.5
#load DESeq results file
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD20GO_miR_OE_vs_CTRL.RData")
load(file = "/Users/phoebetamblinhopper/Downloads/resD10GO_miR_OE_vs_CTRL.RData")
DE_miRNAs_D20 <- resD20_GOtable %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#Filter miRNA with no targets
DE_miRNAs_D20 <- DE_miRNAs_D20[DE_miRNAs_D20 %in%  GSEA_db_ID$miRNA_ensembl_id]
GSEA_db_ID %>%
    filter(miRNA_ensembl_id %in% DE_miRNAs_D20)
#run function
miRNA_D20_LFC0.5 <- runGOseq(DE_miRNAs_D20,
        GSEA_db_ID)
#Save outputs
miRNA_D20_LFC0.5_res <- miRNA_D20_LFC0.5[[1]]
save(miRNA_D20_LFC0.5_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_D20_LFC0.5.RData")
miRNA_D20_LFC0.5[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_D20_LFC0.5.pdf")
  #ggsave(file = "/Users/phoebetamblinhopper/Downloads/GOfig_miRNA_D20_LFC0.5.pdf")
#Check how many unique gene targets there are
check <- GSEA_db_ID[GSEA_db_ID$miRNA_ensembl_id %in% DE_miRNAs_D20,]
n_distinct(check$gene_target)
```
```{r, overall LFC 1}
#for overall GOSeq, LFC threshold of 1
#load DESeq results file
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL_LFC1.RData")
#load(file = "/Users/phoebetamblinhopper/Downloads/resGO_OE_vs_CTRL_LFC1.RData")
DE_miRNAs_LFC1 <- res_GOtable_miRNAs_LFC1 %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#run function
miRNA_overall_LFC1 <- runGOseq(DE_miRNAs_LFC1,
        GSEA_db_ID)
#Save outputs
miRNA_overall_LFC1_res <- miRNA_overall_LFC1[[1]]
save(miRNA_overall_LFC1_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_overall_LFC1.RData")
miRNA_overall_LFC1[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_overall_LFC1.pdf")
```

```
```{r, D10 LFC 1}
#for D10 GOSeq, LFC threshold of 1
#load DESeq results file
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_OE_vs_CTRL_LFC1.RData")
#load(file = "/Users/phoebetamblinhopper/Downloads/resD10GO_OE_vs_CTRL_LFC1.RData")
DE_miRNAs_D10_LFC1 <- resD10_GOtable_LFC1 %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#run function
miRNA_D10_LFC1 <- runGOseq(DE_miRNAs_D10_LFC1,
        GSEA_db_ID)
#Save outputs
miRNA_D10_LFC1_res <- miRNA_D10_LFC1[[1]]
save(miRNA_D10_LFC1_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_D10_LFC1.RData")
miRNA_D10_LFC1[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_D10_LFC1.pdf")
```
```{r, overall LFC 0.1}
# For overall, with LFC of 0.1
#load DESeq results file
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL_LFC0.1.RData")
#load(file = "/Users/phoebetamblinhopper/Downloads/resGO_OE_vs_CTRL_LFC0.1.RData")
DE_miRNAs_LFC0.1 <- res_GOtable_miRNAs_LFC0.1 %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#run function
miRNA_overall_LFC0.1 <- runGOseq(DE_miRNAs_LFC0.1,
        GSEA_db_ID)
#Save outputs
miRNA_overall_LFC0.1_res <- miRNA_overall_LFC0.1[[1]]
save (miRNA_overall_LFC0.1_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_overall_LFC0.1.RData")
miRNA_overall_LFC0.1[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_overall_LFC0.1.pdf")
```
```{r, D10 LFC 0.1}
#for D10 GOSeq, LFC threshold of 0.1
#load DESeq results file
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_OE_vs_CTRL_LFC0.1.RData")
#load(file = "/Users/phoebetamblinhopper/Downloads/resD10GO_OE_vs_CTRL_LFC0.1.RData")
DE_miRNAs_D10_LFC0.1 <- resD10_GOtable_LFC0.1 %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
#run function
miRNA_D10_LFC0.1 <- runGOseq(DE_miRNAs_D10_LFC0.1,
        GSEA_db_ID)
#Save outputs
miRNA_D10_LFC0.1_res <- miRNA_D10_LFC0.1[[1]]
save(miRNA_D10_LFC0.1_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA_D10_LFC0.1.RData")
miRNA_D10_LFC0.1[[2]] %>%
```
```

```
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA_D10_LFC0.1.pdf")
#Just having a look at some specific terms that come under system development
#(in top categories) and are related to osteogenesis
#to see if these are significant, which they are as become more general
#ossification involved in bone maturation
miRNA_D10_LFC0.1_res %>% filter(category == "GO:0043931")
#bone maturation
miRNA_D10_LFC0.1_res %>% filter(category == "GO:0070977")
#bone development
miRNA_D10_LFC0.1_res %>% filter(category == "GO:0060348")
#skeletal system development
miRNA_D10_LFC0.1_res %>% filter(category == "GO:0001501")
#ossification
miRNA_D10_LFC0.1_res %>% filter(category == "GO:0001503")
#cartilage development
miRNA_D10_LFC0.1_res %>% filter(category == "GO:0051216")
```
```{r, miR30, smallest padj}
#Look at miR with smallest padj in overall DESEq res
#ENSMUSG00000065405 mir30a
DE_mir30a <- setNames(c("ENSMUSG00000065405"), c("1"))
#run function
miRNA_overall_padj_mir30a <- runGOseq(DE_mir30a,
        GSEA_db_ID)
#Save outputs
miRNA_overall_padj_mir30a_res <- miRNA_overall_padj_mir30a[[1]]
save(miRNA_overall_padj_mir30a_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA30a_lowestpadj_overall.RData")
miRNA_overall_padj_mir30a[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA30a_lowestpadj_overall.pdf")
```
```{r, miR149, smallest padj LFC>0.5}
#For miR with LFC over 0.5 and smallest padj
#ENSMUSG00000065470
DE_mir149 <- setNames(c("ENSMUSG00000065470"), c("1"))
#run function
miRNA_overall_padjLFC_mir149 <- runGOseq(DE_mir149,
GSEA_db_ID)
#Save outputs
miRNA_overall_padjLFC_mir149_res <- miRNA_overall_padjLFC_mir149[[1]]
save(miRNA_overall_padjLFC_mir149_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA149_lowestpadj_LFC0.5_overall.RData")
miRNA_overall_padjLFC_mir149[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA149_lowestpadj_LFC0.5_overall.pdf")
```
```{r, miR1249, smallest padj at D10}
#Look at miR with smallest padj at D10 (also has LFC of over 1)
#ENSMUSG00000080441 mir1249
DE_mir1249 <- setNames(c("ENSMUSG00000080441"), c("1"))
#run function
miRNA_D10_padj_mir1249 <- runGOseq(DE_mir1249,
        GSEA_db_ID)
#Save outputs
miRNA_D10_padj_mir1249_res <- miRNA_D10_padj_mir1249[[1]]
save(miRNA_D10_padj_mir1249_res, file = "X:/sudlab1/General/projects/Phoebe/R
```

```
outputs/miRNA_osteo/GOres_miRNA1249_lowestpadj_D10.RData")
miRNA_D10_padj_mir1249[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA1249_lowestpadj_D10.pdf")
```
```{r, miR182, 2nd smallest padj at D10}
#Look at miR with 2nd smallest padj at D10 (also has LFC of over 0.5)
#ENSMUSG00000076361 mir182
DE_mir182 <- setNames(c("ENSMUSG00000076361"), c("1"))
#run function
miRNA_D10_padj_2nd_mir182 <- runGOseq(DE_mir182,
        GSEA_db_ID)
#Save outputs
miRNA_D10_padj_2nd_mir182_res <- miRNA_D10_padj_2nd_mir182[[1]]
save(miRNA_D10_padj_2nd_mir182_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA182_2ndlowestpadj_D10.RData")
miRNA_D10_padj_2nd_mir182[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA182_2ndlowestpadj_D10.pdf")
```
```{r, miR30a, 3rd smallest padj at D10}
#Look at miR with 3rd smallest padj at D10 (also has LFC of over 0.5)
#ENSMUSG00000065405 mir30a
DE_mir30a <- setNames(c("ENSMUSG00000065405"), c("1"))
#run function
miRNA_D10_padj_3rd_mir30a <- runGOseq(DE_mir30a,
        GSEA_db_ID)
#Save outputs
miRNA_D10_padj_3rd_mir30a_res <- miRNA_D10_padj_3rd_mir30a[[1]]
save(miRNA_D10_padj_3rd_mir30a_res, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOres_miRNA30a_3rdlowestpadj_D10.RData")
miRNA_D10_padj_3rd_mir30a[[2]] %>%
  ggsave (file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/GOfig_miRNA30a_3rdlowestpadj_D10.pdf")

```
```

7.2.5 GSEA analysis

```
---
title: "GSEA mRNA Analysis of miR binding"
output: html_notebook
editor_options:
  chunk_output_type: inline
---
```{r, include=FALSE, message=FALSE}
library(tidyverse)
library(DESeq2)
library(fgsea)
library(data.table)
library(ggplot2)
library("biomaRt")
GO_mouse <-
gmtPathways("/shared/sudlab1/General/projects/Phoebe/m3.mirdb.v2024.1.Mm.symbols.gmt")
```
######
```
```

Overall
######
```{r overall}
#for overall CTRL vs OE
#look at format data needs to be in
#data("examplePathways")
#head(examplePathways)
#data("exampleRanks")
#load res table and filter for significant padj
load(file = "/shared/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/res_OE_vs_CTRL.RData")
filtered_res <- subset(res, padj < 0.05)
#filtered_res <- subset(res, padj < 0.05, abs(log2FoldChange) > 0.1)
#can more stringently filter res like this and should lower padj values
#replace gene names with ensembl IDs
#listEnsembl()
ensembl <- useEnsembl(biomart = "genes", mirror = "useast")
#searchDatasets(mart = ensembl, pattern = "musculus")
ensembl <- useDataset(dataset = "mmusculus_gene_ensembl", mart = ensembl)
#listAttributes(ensembl)
gene_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = res$ensembl_gene_id,
      mart = ensembl)
res_genes <- inner_join(filtered_res,
          gene_id_table)
#create vector of gene name and log2FC
res_genes_filtered <- subset(res_genes, select = c(3,8))
#remove duplicate gene names
res_genes_filtered <- res_genes_filtered %>%
  filter(!duplicated(external_gene_name))
vector_genes_filtered <- as.numeric(res_genes_filtered$log2FoldChange)
names(vector_genes_filtered) <- res_genes_filtered$external_gene_name
```
```{r overall run gsea}
#this can be set to any numberset.seed(1)
#run gsea analysis
fgseaRes <- fgsea(pathways = GO_mouse,
        stats = vector_genes_filtered)
head(fgseaRes[order(padj), ])
#just look for upreg genes as this is what ceRNA would do
#look for binding sites of these miRNAs in CASC20
```
```{r overall plots}
plotEnrichment(GO_mouse[["MIR_3094_3P"]],
              res_GOtable_genes_filtered) + labs(title="MIR_3094_3P")
```
```{r overall table}
topmiRNAsUp <- fgseaRes[ES > 0][head(order(pval), n=10), pathway]
topmiRNAsDown <- fgseaRes[ES < 0][head(order(pval), n=10), pathway]
topmiRNAs <- c(topmiRNAsUp, rev(topmiRNAsDown))
plotGseaTable(GO_mouse[topmiRNAs], res_GOtable_genes_filtered, fgseaRes,
```
######
Day 10
######
gseaParam=0.5)
```{r D10}
```

```
#for D10, CTRL vs OE
load(file = "/shared/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/resD10GO_OE_vs_CTRL.rdata")
#load res table and filter for significant padj
load(file = "/shared/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/resD10_OE_vs_CTRL.RData")
filtered_resD10 <- subset(resD10, padj < 0.05)
#merge with gene id table to get gene names
resD10_genes <- inner_join(filtered_resD10,
          gene_id_table)
#create vector of gene name and log2FC
resD10_genes_filtered <- subset(resD10_genes, select = c(3,8))
#remove duplicate gene names
resD10_genes_filtered <- resD10_genes_filtered %>%
  filter(!duplicated(external_gene_name))
vector_D10_genes_filtered <- as.numeric(resD10_genes_filtered$log2FoldChange)
names(vector_D10_genes_filtered) <- resD10_genes_filtered$external_gene_name
```
```{r}
set.seed(1)
#run gsea analysis
fgseaResD10 <- fgsea(pathways = GO_mouse,
stats = vector_D10_genes_filtered)
head(fgseaResD10[order(padj), ], n=10)
save(fgseaResD10, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/mRNA_osteo/GSEA_D10_miR_binding_enrichment.rdata")
```
how many enriched target sets?
```{r}
subset(fgseaResD10, padj<0.05)
```
```{r}
#look if any of these miRNAs were in DE set from DESeq analysis at D10
#filter fgsea table for sig (p < 0.05)
filtered_fgseaResD10 <- subset(fgseaResD10, fgseaResD10$padj < 0.05)
#load resD10 table
load(file = "/shared/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10_miR_OE_vs_CTRL.RData")
#filter resD10 for sig (p <0.05) and LFC
#as without filtering for LFC was too many miRNAs to look at individually
filtered_resD10 <- subset(resD10, resD10$padj < 0.05 & abs(log2FoldChange) > 0.5)
#make id table to convert miR names to IDs
miR_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = resD10$ensembl_gene_id,
      mart = ensembl)
#make lowercase to match naming
miR_id_table$external_gene_name <- tolower(miR_id_table$external_gene_name)
#alter naming of miRNAs in fgsea table to match miR id table
#add new column first so keep original naming also
filtered_fgseaResD10 <- filtered_fgseaResD10 %>%
  mutate(ensembl_pathway = pathway)
filtered_fgseaResD10_db <- filtered_fgseaResD10 %>%
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "_", "-")
  )  %>%
#Separate miRNAs that are together onto separate lines
  mutate(
```

```
    ensembl_pathway = str_replace_all(ensembl_pathway, "-LET-", "SPLITLET")
  ) %>%
separate_longer_delim(ensembl_pathway, "SPLIT") %>%
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "LET-", "LET")
  ) %>%
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "-MIR-", "SPLITMIR")
  ) %>%
separate_longer_delim(ensembl_pathway, "SPLIT") %>%
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "MIR-", "MIR")
  ) %>%
#Edit naming to include mir as in ID table
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "LET", "MIRLET")
  ) %>%
#Remove P and number before
  mutate(
    ensembl_pathway = str_replace(ensembl_pathway, "-.P", "")
) %>%
  #Make all names lowercase to match
mutate(ensembl_pathway = tolower(ensembl_pathway))
#join miR id table with fgsea table to get gene ids
fgseaResD10_ID <- inner_join(filtered_fgseaResD10_db, miR_id_table, by =
join_by(ensembl_pathway == external_gene_name))
#are any of DE miRNAs in list of enriched miR sites from DEGs?
DE_enriched_miRNAs <- subset(fgseaResD10_ID, (ensembl_gene_id %in%
filtered_resD10$ensembl_gene_id))
save(DE_enriched_miRNAs, file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_enriched_miRNAs_in_DEGs")
```
Normalise miR names to match mirbase IDs
```{r}
filtered_fgseaResD10 <- filtered_fgseaResD10 %>%
  mutate(ensembl_pathway = pathway)
filtered_fgseaResD10_db <- filtered_fgseaResD10 %>%
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "_", "-")
  )  %>%
#Separate miRNAs that are together onto separate lines
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "-LET-", "SPLITLET-")
  ) %>%
separate_longer_delim(ensembl_pathway, "SPLIT") %>%
  mutate(
    ensembl_pathway = str_replace_all(ensembl_pathway, "-MIR-", "SPLITMIR-")
  ) %>%
separate_longer_delim(ensembl_pathway, "SPLIT") %>%
  mutate(ensembl_pathway = paste0("mmu-", ensembl_pathway)) %>%
  #Make all names lowercase to match
mutate(ensembl_pathway = tolower(ensembl_pathway),
       ensembl_pathway = sub("mir", "miR", ensembl_pathway))
casc20_miRNA <-
read.delim("/shared/sudlab1/General/projects/Phoebe/mouse_analysis/targets_miranda_counts.tsv")
inner_join(filtered_fgseaResD10_db, casc20_miRNA, by=c("ensembl_pathway"="miRNA")) %>%
  rowwise() %>%
  mutate(leadingEdge = paste(leadingEdge, collapse=",")) %>%
```

```
  write.table("miRNAs_targeting_casc20_and_DE_mRNAs.tsv", sep="\t")
```
Look which of these are DE miRNAs
```{r}
load(file = "/shared/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resGO_OE_vs_CTRL.RData")
gene_id_table <- getBM(attributes = c("external_gene_name", "ensembl_gene_id"),
      filters = "ensembl_gene_id",
      values = res_GOtable_miRNAs$ensembl_gene_id,
      mart = ensembl)
res_GOtable_miRNAs_ID <- inner_join(res_GOtable_miRNAs,
          gene_id_table)
casc20_miRNA %>%
  filter(grepl("mmu", miRNA)) %>%
  mutate(miRNA = sub("mmu", "", miRNA),
         miRNA = sub("-.p", "", miRNA),
         miRNA = sub("-miR-", "Mir", miRNA)) %>%
  group_by(miRNA) %>%
  summarise(Num.Hits = sum(Num.Hits)) %>%
  inner_join(res_GOtable_miRNAs_ID, by=c("miRNA"="external_gene_name")) %>%
  filter(padj<0.05) %>%
  write.table("DE_miRNAs_targeting_casc20.tsv", sep="\t", row.names = FALSE)
```
```{r plots for DE miRNAs}
#Look at graphs for each of miRNAs that had enriched sites in DEGs, that are also
#significantly DE from DESeq analysis
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_148A_3P.png")
plotEnrichment(GO_mouse[["MIR_148A_3P_MIR_148B_3P_MIR_152_3P"]],
               vector_D10_genes_filtered) + labs(title="MIR_148A_3P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_1947_3P.png")
plotEnrichment(GO_mouse[["MIR_1947_3P"]],
               vector_D10_genes_filtered) + labs(title="MIR_1947_3P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_19B_1_5P.png")
plotEnrichment(GO_mouse[["MIR_19B_1_5P"]],
               vector_D10_genes_filtered) + labs(title="MIR_19B_1_5P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_19B_2_5P.png")
plotEnrichment(GO_mouse[["MIR_19B_2_5P"]],
               vector_D10_genes_filtered) + labs(title="MIR_19B_2_5P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_30A_3P.png")
plotEnrichment(GO_mouse[["MIR_30A_3P_MIR_30E_3P"]],
               vector_D10_genes_filtered) + labs(title="MIR_30A_3P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_30A_5P.png")
plotEnrichment(GO_mouse[["MIR_30A_5P_MIR_30E_5P"]],
               vector_D10_genes_filtered) + labs(title="MIR_30A_5P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_669B_3P.png")
```

```
plotEnrichment(GO_mouse[["MIR_669B_3P"]],
               vector_D10_genes_filtered) + labs(title="MIR_669B_3P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_669B_5P.png")
plotEnrichment(GO_mouse[["MIR_669B_5P"]],
               vector_D10_genes_filtered) + labs(title="MIR_669B_5P")
dev.off()
jpeg("X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/DE_miRNAs_enriched_in_DEGs/MIR_7680_5P.png")
plotEnrichment(GO_mouse[["MIR_7680_5P"]],
               vector_D10_genes_filtered) + labs(title="MIR_7680_5P")
dev.off()
#look at highest ranking genes for a particular miR
MIR_7680_5P <- fgseaResD10 %>%
  filter(pathway == "MIR_7680_5P")
MIR_7680_5P$leadingEdge[1]
```
```{r}
#Look at whether these DE miRNAs enriched in DEGs are up or down reg
#by looking at LFC from resD10
DE_enriched_miRNAs_LFC <- inner_join(DE_enriched_miRNAs, resD10, by = join_by
("ensembl_gene_id"
== "ensembl_gene_id"))
```
```{r}
topmiRNAsUp <- fgseaResD10[ES > 0][head(order(pval), n=10), pathway]
topmiRNAsDown <- fgseaResD10[ES < 0][head(order(pval), n=10), pathway]
topmiRNAs <- c(topmiRNAsUp, rev(topmiRNAsDown))
GSEA_tab_D10 <- plotGseaTable(GO_mouse[topmiRNAs], resD10_GOtable_genes_filtered,
fgseaRes,
                              gseaParam=1)

```
```

7.2.6 Sylamer analysis

```
---
title: "Sylamer"
output: html_notebook
editor_options:
  chunk_output_type: inline
---
```{r, include=FALSE, message=FALSE}
library(tidyverse)
library("biomaRt")
library(Biostrings)
```
#Create input files for sylamer
```{r}
#use biomart to find 3' UTRS of DE mRNAs
#load table of mRNAs
load("/shared/sudlab1/General/projects/Phoebe/R outputs/resD10GO_OE_vs_CTRL.RData")
#load("/Users/phoebetamblinhopper/Downloads/resD10GO_OE_vs_CTRL.RData")
#pull DE miRNAs
DE <- resD10_GOtable %>%
```

```
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
ensembl = useMart("ensembl",dataset="mmusculus_gene_ensembl")
UTRs <- getBM(attributes=c('ensembl_gene_id', '3utr'),
                  filters = 'ensembl_gene_id',
                  values = DE,
                  mart = ensembl)
#this method failed to find UTRs for many genes
#Get the mouse GTF
#Get annotations for 3'UTRs
#/mnt/parscratch/users/bo1cv/public/annotations/mm_GRCm39_ensembl111/ensembl.dir/geneset_all
.gtf
cds_fasta <-
"/shared/sudlab1/General/projects/Phoebe/mirror/genomes/Mus_musculus.GRCm39.cds.all.fa.gz"
#cds_fasta <- "/Users/phoebetamblinhopper/Downloads/Mus_musculus.GRCm39.cds.all.fa.gz"
GRCm39 <- readDNAStringSet(cds_fasta,
                      format = "fasta")
#each gene will have multiple transcripts, choose longest to take UTR of
#find longest transcript
GRCm39_cf <- oligonucleotideFrequency(GRCm39, width = 3, step = 3)
#Transcript_id
GRCm39_id <- GRCm39@ranges@NAMES
row.names(GRCm39_cf) <- str_extract(GRCm39_id, "ENSMUST[:digit:]+")
#Summing number of codons
cds_lengths = data.frame(transcript_id = GRCm39_id, cds_length = rowSums(GRCm39_cf))
#make table of transcript IDs corresponding to gene IDs of DE genes
transcript_ID <- getBM(attributes = c("ensembl_gene_id", "ensembl_transcript_id"),
     mart = ensembl)
#Merging DE transcripts with codon length df, then select for each gene, the transcript with longest
CDS
cds_lengths <- rownames_to_column(cds_lengths, "ensembl_transcript_id")
cds_lengths <- merge(transcript_ID, cds_lengths) %>% distinct()
selected_transcript_long <- cds_lengths %>% group_by(ensembl_gene_id) %>%
  slice_max(order_by=cds_length, n=1, with_ties = FALSE) %>%
  dplyr::select(ensembl_transcript_id, ensembl_gene_id) %>%
  ungroup()
utr_sequences <- getBM(ensembl, attributes = c("ensembl_gene_id", "3utr"),
                    filters = "ensembl_transcript_id",```
Run sylamer
values = selected_transcript_long$ensembl_transcript_id)
utr_sequence_vector <- utr_sequences$`3utr`
names(utr_sequence_vector) <- utr_sequences$ensembl_gene_id
utr_sequence_vector <- utr_sequence_vector[utr_sequence_vector != "Sequence unavailable"]
UTR_ss = DNAStringSet(x=utr_sequence_vector)
writeXStringSet(UTR_ss,
"/shared/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/pipeline_sylamer/utrs.fasta")
```
```{r}
# to create ranked list make .tsv with LFCs
all_LFCs <- resD10_GOtable %>%
  #filter(Significant == 1) %>%
  arrange(log2FoldChange/lfcSE) %>%
  na.omit() %>%
  mutate(t=log2FoldChange/lfcSE) %>%
  dplyr::select(ensembl_gene_id, t,log2FoldChange)
write.table(dplyr::select(all_LFCs, ensembl_gene_id, t),
file="/shared/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/pipeline_sylamer/ranked_
t.tsv",
```

```
sep="\t",
          quote=F, col.names = F, row.names = F)
write.table(arrange(all_LFCs, log2FoldChange) %>% dplyr::select(ensembl_gene_id,
log2FoldChange),
file="/shared/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/pipeline_sylamer/ranked_
LFC.tsv",
sep="\t",
quote=F, col.names = F, row.names = F)
```{bash, engine.opts="-l"}
source /home/mb1ims/.bashrc
syl.master.sh -r
/shared/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/pipeline_sylamer/ranked_t.tsv \
          -S mmu \
          -d sylamar_output_t \
          -f
/shared/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/pipeline_sylamer/utrs.fasta \
          -t
/shared/sudlab1/General/projects/Phoebe/mouse_analysis/miRNAseq/pipeline_sylamer/mouse.table
```
```{r}
#miRbase to find seeds of miRNAs of interest
#load DE miRNAs
load(file = "X:/sudlab1/General/projects/Phoebe/R
outputs/miRNA_osteo/resD10GO_miR_OE_vs_CTRL.RData")
DE_miRNAs <- resD10_GOtable %>%
  filter(Significant == 1) %>%
  pull(ensembl_gene_id)
```
```

References

ADAMS, B. D., PARSONS, C., WALKER, L., ZHANG, W. C. & SLACK, F. J. 2017. Targeting noncoding RNAs in disease. *J Clin Invest,* 127**,** 761-771.

ALTUVIA, Y., LANDGRAF, P., LITHWICK, G., ELEFANT, N., PFEFFER, S., ARAVIN, A., BROWNSTEIN, M. J., TUSCHL, T. & MARGALIT, H. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res,* 33**,** 2697-706.

AMBROS, V., BARTEL, B., BARTEL, D. P., BURGE, C. B., CARRINGTON, J. C., CHEN, X., DREYFUSS, G., EDDY, S. R., GRIFFITHS-JONES, S., MARSHALL, M., MATZKE, M., RUVKUN, G. & TUSCHL, T. 2003. A uniform system for microRNA annotation. *RNA,* 9**,** 277-9.

ARNOLD, P. R., WELLS, A. D. & LI, X. C. 2019. Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front Cell Dev Biol,* 7**,** 377.

ARUN, G., DIERMEIER, S., AKERMAN, M., CHANG, K. C., WILKINSON, J. E., HEARN, S., KIM, Y., MACLEOD, A. R., KRAINER, A. R., NORTON, L., BROGI, E., EGEBLAD, M. & SPECTOR, D. L. 2016. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev,* 30**,** 34-51.

ASHJARI, D., KARAMALI, N., RAJABINEJAD, M., HASSANI, S. S., AFSHAR HEZARKHANI, L., AFSHARI, D., GORGIN KARAJI, A., SALARI, F. & REZAIEMANESH, A. 2022. The axis of long non-coding RNA MALAT1/miR-1-3p/CXCR4 is dysregulated in patients with diabetic neuropathy. *Heliyon,* 8**,** e09178.

BABAR, I. A., CHENG, C. J., BOOTH, C. J., LIANG, X., WEIDHAAS, J. B., SALTZMAN, W. M. & SLACK, F. J. 2012. Nanoparticle-based therapy in an in vivo microRNA-155 (miR-155)-dependent mouse model of lymphoma. *Proc Natl Acad Sci U S A,* 109**,** E1695-1704.

BAEK, D., VILLEN, J., SHIN, C., CAMARGO, F. D., GYGI, S. P. & BARTEL, D. P. 2008. The impact of microRNAs on protein output. *Nature,* 455**,** 64-71.

BASSETT, A. R., AKHTAR, A., BARLOW, D. P., BIRD, A. P., BROCKDORFF, N., DUBOULE, D., EPHRUSSI, A., FERGUSON-SMITH, A. C., GINGERAS, T. R., HAERTY, W., HIGGS, D. R., MISKA, E. A. & PONTING, C. P. 2014. Considerations when investigating lncRNA function in vivo. *Elife,* 3**,** e03058.

BATTON, K. A., AUSTIN, C. O., BRUNO, K. A., BURGER, C. D., SHAPIRO, B. P. & FAIRWEATHER, D. 2018. Sex differences in pulmonary arterial hypertension: role of infection and autoimmunity in the pathogenesis of disease. *Biol Sex Differ,* 9**,** 15.

BENOIT BOUVRETTE, L. P., CODY, N. A. L., BERGALET, J., LEFEBVRE, F. A., DIOT, C., WANG, X., BLANCHETTE, M. & LECUYER, E. 2018. CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells. *RNA,* 24**,** 98-113.

BEREZHNA, S. Y., SUPEKOVA, L., SUPEK, F., SCHULTZ, P. G. & DENIZ, A. A. 2006. siRNA in human cells selectively localizes to target RNA sites. *Proc Natl Acad Sci U S A,* 103**,** 7682-87.

BERGHOLT, N. L., LYSDAHL, H., LIND, M. & FOLDAGER, C. B. 2019. A Standardized Method of Applying Toluidine Blue Metachromatic Staining for Assessment of Chondrogenesis. *Cartilage,* 10**,** 370-374.

BERTOLO, A., BAUR, M., GUERRERO, J., POTZEL, T. & STOYANOV, J. 2019. Autofluorescence is a Reliable in vitro Marker of Cellular Senescence in Human Mesenchymal Stromal Cells. *Sci Rep,* 9**,** 2074.

BHAN, A. & MANDAL, S. S. 2014. Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. *ChemMedChem,* 9**,** 1932-56.
BISWAS, S., THOMAS, A. A., CHEN, S., AREF-ESHGHI, E., FENG, B., GONDER, J., SADIKOVIC, B. & CHAKRABARTI, S. 2018. MALAT1: An Epigenetic Regulator of Inflammation in Diabetic Retinopathy. *Sci Rep,* 8**,** 6526.
BOON, R. A., JAE, N., HOLDT, L. & DIMMELER, S. 2016. Long Noncoding RNAs: From Clinical Genetics to Therapeutic Targets? *J Am Coll Cardiol,* 67**,** 1214-1226.
BOST, C., ARLEEVSKAYA, M. I., BROOKS, W. H., PLAZA, S., GUERY, J. C. & RENAUDINEAU, Y. 2022. Long non-coding RNA Xist contribution in systemic lupus erythematosus and rheumatoid arthritis. *Clin Immunol,* 236**,** 108937.
BRAGDON, B., BURNS, R., BAKER, A. H., BELKINA, A. C., MORGAN, E. F., DENIS, G. V., GERSTENFELD, L. C. & SCHLEZINGER, J. J. 2015. Intrinsic Sex-Linked Variations in Osteogenic and Adipogenic Differentiation Potential of Bone Marrow Multipotent Stromal Cells. *J Cell Physiol,* 230**,** 296-307.
BRIDGES, M. C., DAULAGALA, A. C. & KOURTIDIS, A. 2021. LNCcation: lncRNA localization and function. *J Cell Biol,* 220**,** e202009045.
BULCHA, J. T., WANG, Y., MA, H., TAI, P. W. L. & GAO, G. 2021. Viral vector platforms within the gene therapy landscape. *Signal Transduct Target Ther,* 6**,** 53.
CABILI, M. N., DUNAGIN, M. C., MCCLANAHAN, P. D., BIAESCH, A., PADOVAN-MERHAR, O., REGEV, A., RINN, J. L. & RAJ, A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol,* 16**,** 20.
CAPULLI, M., PAONE, R. & RUCCI, N. 2014. Osteoblast and osteocyte: games without frontiers. *Arch Biochem Biophys,* 561**,** 3-12.
CARLEVARO-FITA, J., RAHIM, A., GUIGO, R., VARDY, L. A. & JOHNSON, R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA,* 22**,** 867-82.
CHAUMEIL, J., LE BACCON, P., WUTZ, A. & HEARD, E. 2006. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev,* 20**,** 2223-37.
CHEN, C., HE, W., HUANG, J., WANG, B., LI, H., CAI, Q., SU, F., BI, J., LIU, H., ZHANG, B., JIANG, N., ZHONG, G., ZHAO, Y., DONG, W. & LIN, T. 2018. LNMAT1 promotes lymphatic metastasis of bladder cancer via CCL2 dependent macrophage recruitment. *Nat Commun,* 9**,** 3826.
CHEN, C. K., BLANCO, M., JACKSON, C., AZNAURYAN, E., OLLIKAINEN, N., SURKA, C., CHOW, A., CERASE, A., MCDONEL, P. & GUTTMAN, M. 2016. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science,* 354**,** 468-472.
CHEN, P. Y., HSIEH, P. L., PENG, C. Y., LIAO, Y. W., YU, C. H. & YU, C. C. 2021a. LncRNA MEG3 inhibits self-renewal and invasion abilities of oral cancer stem cells by sponging miR-421. *J Formos Med Assoc,* 120**,** 1137-1142.
CHEN, Y., HUANG, C., DUAN, Z. B., CHEN, Y. X. & XU, C. Y. 2023a. LncRNA NEAT1 accelerates renal fibrosis progression via targeting miR-31 and modulating RhoA/ROCK signal pathway. *Am J Physiol Cell Physiol,* 324**,** C292-C306.
CHEN, Y., LI, Z., CHEN, X. & ZHANG, S. 2021b. Long non-coding RNAs: From disease code to drug role. *Acta Pharm Sin B,* 11**,** 340-354.
CHEN, Y., MEHMOOD, K., CHANG, Y. F., TANG, Z., LI, Y. & ZHANG, H. 2023b. The molecular mechanisms of glycosaminoglycan biosynthesis regulating chondrogenesis and endochondral ossification. *Life Sci,* 335**,** 122243.

CHEN, Y. & WANG, X. 2020. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res,* 48**,** D127-D131.

CHIANG, J. C., JIANG, J., NEWBURGER, P. E. & LAWRENCE, J. B. 2018. Trisomy silencing by XIST normalizes Down syndrome cell pathogenesis demonstrated for hematopoietic defects in vitro. *Nat Commun,* 9**,** 5180.

CHODROFF, R. A., GOODSTADT, L., SIREY, T. M., OLIVER, P. L., DAVIES, K. E., GREEN, E. D., MOLNAR, Z. & PONTING, C. P. 2010. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol,* 11**,** R72.

CHUNHAROJRITH, P., NAKAYAMA, Y., JIANG, X., KERY, R. E., MA, J., DE LA HOZ ULLOA, C. S., ZHANG, X., ZHOU, Y. & KLIBANSKI, A. 2015. Tumor suppression by MEG3 lncRNA in a human pituitary tumor derived cell line. *Mol Cell Endocrinol,* 416**,** 27-35.

CLARK, M. B., JOHNSTON, R. L., INOSTROZA-PONTA, M., FOX, A. H., FORTINI, E., MOSCATO, P., DINGER, M. E. & MATTICK, J. S. 2012. Genome-wide analysis of long noncoding RNA stability. *Genome Res,* 22**,** 885-98.

CLEMSON, C. M., HUTCHINSON, J. N., SARA, S. A., ENSMINGER, A. W., FOX, A. H., CHESS, A. & LAWRENCE, J. B. 2009. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell,* 33**,** 717-26.

CLINICALTRIALS.GOV. 2021. *Efficacy, Safety, and Tolerability of Remlarsen (MRG-201) Following Intradermal Injection in Subjects With a History of Keloids (NCT03601052)* [Online]. ClinicalTrials.gov. Available: https://clinicaltrials.gov/study/NCT03601052 [Accessed 23/01/24 2024].

CLINICALTRIALS.GOV. 2024. *ClinicalTrials.gov* [Online]. National Library of Medicine. Available: https://clinicaltrials.gov/ [Accessed 02/05/2024 2024].

COLLOTTA, D., BERTOCCHI, I., CHIAPELLO, E. & COLLINO, M. 2023. Antisense oligonucleotides: a novel Frontier in pharmacological strategy. *Front Pharmacol,* 14**,** 1304342.

CRYSTAL, R. G. 2014. Adenovirus: the first effective in vivo gene delivery vector. *Hum Gene Ther,* 25**,** 3-11.

DE GOEDE, O. M., NACHUN, D. C., FERRARO, N. M., GLOUDEMANS, M. J., RAO, A. S., SMAIL, C., EULALIO, T. Y., AGUET, F., NG, B., XU, J., BARBEIRA, A. N., CASTEL, S. E., KIM-HELLMUTH, S., PARK, Y., SCOTT, A. J., STROBER, B. J., CONSORTIUM, G. T., BROWN, C. D., WEN, X., HALL, I. M., BATTLE, A., LAPPALAINEN, T., IM, H. K., ARDLIE, K. G., MOSTAFAVI, S., QUERTERMOUS, T., KIRKEGAARD, K. & MONTGOMERY, S. B. 2021. Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell,* 184**,** 2633-2648 e19.

DENNING, W., DAS, S., GUO, S., XU, J., KAPPES, J. C. & HEL, Z. 2013. Optimization of the transductional efficiency of lentiviral vectors: effect of sera and polycations. *Mol Biotechnol,* 53**,** 308-14.

DERRIEN, T., JOHNSON, R., BUSSOTTI, G., TANZER, A., DJEBALI, S., TILGNER, H., GUERNEC, G., MARTIN, D., MERKEL, A., KNOWLES, D. G., LAGARDE, J., VEERAVALLI, L., RUAN, X., RUAN, Y., LASSMANN, T., CARNINCI, P., BROWN, J. B., LIPOVICH, L., GONZALEZ, J. M., THOMAS, M., DAVIS, C. A., SHIEKHATTAR, R., GINGERAS, T. R., HUBBARD, T. J., NOTREDAME, C., HARROW, J. & GUIGO, R. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res,* 22**,** 1775-89.

DEY, D., WHEATLEY, B. M., CHOLOK, D., AGARWAL, S., YU, P. B., LEVI, B. & DAVIS, T. A. 2017. The traumatic bone: trauma-induced heterotopic ossification. *Transl Res,* 186**,** 95-111.

DHAMIJA, S. & DIEDERICHS, S. 2016. From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis. *Int J Cancer,* 139**,** 269-80.

DONG, P., XIONG, Y., YUE, J., HANLEY, S. J. B., KOBAYASHI, N., TODO, Y. & WATARI, H. 2018. Long Non-coding RNA NEAT1: A Novel Target for Diagnosis and Therapy in Human Tumors. *Front Genet,* 9**,** 471.

ELEGHEERT, J., BEHIELS, E., BISHOP, B., SCOTT, S., WOOLLEY, R. E., GRIFFITHS, S. C., BYRNE, E. F. X., CHANG, V. T., STUART, D. I., JONES, E. Y., SIEBOLD, C. & ARICESCU, A. R. 2018. Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat Protoc,* 13**,** 2991-3017.

EMA. 2023. *Article 57 product data* [Online]. European Medicines Agency. Available: https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/data-medicines-iso-idmp-standards-post-authorisation/public-data-article-57-database [Accessed 05/01/2024 2024].

ENGREITZ, J. M., PANDYA-JONES, A., MCDONEL, P., SHISHKIN, A., SIROKMAN, K., SURKA, C., KADRI, S., XING, J., GOREN, A., LANDER, E. S., PLATH, K. & GUTTMAN, M. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science,* 341**,** 1237973.

ENRIGHT AJ, J. B., GAUL U, TUSCHL T, SANDER C AND MARKS DS 2003. MiRanda. *Genome Biology,* 5.

FAGHIHI, M. A., MODARRESI, F., KHALIL, A. M., WOOD, D. E., SAHAGAN, B. G., MORGAN, T. E., FINCH, C. E., ST LAURENT, G., 3RD, KENNY, P. J. & WAHLESTEDT, C. 2008. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med,* 14**,** 723-30.

FANG, J., SUN, C. C. & GONG, C. 2016. Long noncoding RNA XIST acts as an oncogene in non-small cell lung cancer by epigenetically repressing KLF2 expression. *Biochem Biophys Res Commun,* 478**,** 811-7.

FDA. 2024. *Drugs@FDA: FDA-Approved Drugs* [Online]. Food and Drug Administration. Available: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm [Accessed 22/01/24 2024].

FELIX-ILEMHENBHIO, F. 2023. *Investigating the functional role of lncRNA CASC20 in heterotopic bone formation*

. University of Sheffield.

FILIPOWICZ, W., BHATTACHARYYA, S. N. & SONENBERG, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet,* 9**,** 102-14.

FLORENCIO-SILVA, R., SASSO, G. R., SASSO-CERRI, E., SIMOES, M. J. & CERRI, P. S. 2015. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. *Biomed Res Int,* 2015**,** 421746.

FRIEDMAN, R. C., FARH, K. K., BURGE, C. B. & BARTEL, D. P. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res,* 19**,** 92-105.

FUKUSHIMA, K., SATOH, T., SUGIHARA, F., SATO, Y., OKAMOTO, T., MITSUI, Y., YOSHIO, S., LI, S., NOJIMA, S., MOTOOKA, D., NAKAMURA, S., KIDA, H., STANDLEY, D. M., MORII, E., KANTO, T., YANAGITA, M., MATSUURA, Y., NAGASAWA, T., KUMANOGOH, A. & AKIRA, S. 2020. Dysregulated Expression of the Nuclear Exosome Targeting Complex

Component Rbm7 in Nonhematopoietic Cells Licenses the Development of Fibrosis. *Immunity,* 52**,** 542-556.

GAO, L., WANG, X., GUO, S., XIAO, L., LIANG, C., WANG, Z., LI, Y., LIU, Y., YAO, R., LIU, Y. & ZHANG, Y. 2019. LncRNA HOTAIR functions as a competing endogenous RNA to upregulate SIRT1 by sponging miR-34a in diabetic cardiomyopathy. *J Cell Physiol,* 234**,** 4944-4958.

GE, Z., YIN, C., LI, Y., TIAN, D., XIANG, Y., LI, Q., TANG, Y. & ZHANG, Y. 2022. Long noncoding RNA NEAT1 promotes cardiac fibrosis in heart failure through increased recruitment of EZH2 to the Smad7 promoter region. *J Transl Med,* 20**,** 7.

GENCODE. 2024. *GENCODE Release 46 (GRCh38.p14).* [Online]. Available: https://www.gencodegenes.org/human/ [Accessed].

GOYAL, A., MYACHEVA, K., GROSS, M., KLINGENBERG, M., DURAN ARQUE, B. & DIEDERICHS, S. 2017. Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic Acids Res,* 45**,** e12.

GRIMM, D., STREETZ, K. L., JOPLING, C. L., STORM, T. A., PANDEY, K., DAVIS, C. R., MARION, P., SALAZAR, F. & KAY, M. A. 2006. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature,* 441**,** 537-41.

GRIMSON, A., FARH, K. K., JOHNSTON, W. K., GARRETT-ENGELE, P., LIM, L. P. & BARTEL, D. P. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell,* 27**,** 91-105.

GUO, C. J., MA, X. K., XING, Y. H., ZHENG, C. C., XU, Y. F., SHAN, L., ZHANG, J., WANG, S., WANG, Y., CARMICHAEL, G. G., YANG, L. & CHEN, L. L. 2020. Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell,* 181**,** 621-636.

GUPTA, R. A., SHAH, N., WANG, K. C., KIM, J., HORLINGS, H. M., WONG, D. J., TSAI, M. C., HUNG, T., ARGANI, P., RINN, J. L., WANG, Y., BRZOSKA, P., KONG, B., LI, R., WEST, R. B., VAN DE VIJVER, M. J., SUKUMAR, S. & CHANG, H. Y. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature,* 464**,** 1071-76.

GUTTMAN, M. & RINN, J. L. 2012. Modular regulatory principles of large non-coding RNAs. *Nature,* 482**,** 339-46.

GYŐRGY, R., KLONTZAS, M. E., KOSTOGLOU, M., PANOSKALTSIS, N., MANTALARIS, A. & GEORGIADIS, M. C. 2019. Capturing Mesenchymal Stem Cell Heterogeneity during Osteogenic Differentiation: An Experimental–Modeling Approach. *Industrial & Engineering Chemistry Research,* 58**,** 13900-13909.

HAN, J., LAVIGNE, C. A., JONES, B. T., ZHANG, H., GILLETT, F. & MENDELL, J. T. 2020. A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. *Science,* 370.

HANNA, H., MIR, L. M. & ANDRE, F. M. 2018. In vitro osteoblastic differentiation of mesenchymal stem cells generates cell layers with distinct properties. *Stem Cell Res Ther,* 9**,** 203.

HAWES, J., SHI, Q., REN, L., SCHNACKENBERG, L. & YANG, K. *Toxicity of three antisense oligonucleotide drugs and eighteen of their impurities in primary human hepatocytes* [Online]. FDA. Available: https://www.fda.gov/science-research/fda-science-forum/toxicity-three-antisense-oligonucleotide-drugs-and-eighteen-their-impurities-primary-human [Accessed 03/05/2024 2024].

HE, P., GELISSEN, I. C. & AMMIT, A. J. 2020. Regulation of ATP binding cassette transporter A1 (ABCA1) expression: cholesterol-dependent and - independent signaling pathways with relevance to inflammatory lung disease. *Respir Res,* 21**,** 250.

HEALTH, N. I. O. 2023. *CASC20 cancer susceptibility 20 [Homo sapiens (human)]* [Online]. National Library of Medicine. Available: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=101929244 [Accessed 30/05/2024 2024].

HEZRONI, H., KOPPSTEIN, D., SCHWARTZ, M. G., AVRUTIN, A., BARTEL, D. P. & ULITSKY, I. 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep,* 11**,** 1110-22.

HIERS, N. M., LI, T., TRAUGOT, C. M. & XIE, M. 2024. Target-directed microRNA degradation: Mechanisms, significance, and functional implications. *Wiley Interdiscip Rev RNA,* 15**,** e1832.

HONG, D. S., KANG, Y. K., BORAD, M., SACHDEV, J., EJADI, S., LIM, H. Y., BRENNER, A. J., PARK, K., LEE, J. L., KIM, T. Y., SHIN, S., BECERRA, C. R., FALCHOOK, G., STOUDEMIRE, J., MARTIN, D., KELNAR, K., PELTIER, H., BONATO, V., BADER, A. G., SMITH, S., KIM, S., O'NEILL, V. & BEG, M. S. 2020. Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours. *Br J Cancer,* 122**,** 1630-1637.

HUA, Y. & KRAINER, A. R. 2012. Antisense-mediated exon inclusion. *Methods Mol Biol,* 867**,** 307-23.

HUANG, W., ZHOU, Y., REN, J. & CHEN, C. 2024. UBE2C, targeted by miR-140-3p, promotes the progression of osteosarcoma via PI3K/AKT signaling pathway. *J Recept Signal Transduct Res,* 44**,** 107-114.

HUARTE, M. 2015. The emerging role of lncRNAs in cancer. *Nat Med,* 21**,** 1253-61.

HUARTE, M., GUTTMAN, M., FELDSER, D., GARBER, M., KOZIOL, M. J., KENZELMANN-BROZ, D., KHALIL, A. M., ZUK, O., AMIT, I., RABANI, M., ATTARDI, L. D., REGEV, A., LANDER, E. S., JACKS, T. & RINN, J. L. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell,* 142**,** 409-19.

IYER, M. K., NIKNAFS, Y. S., MALIK, R., SINGHAL, U., SAHU, A., HOSONO, Y., BARRETTE, T. R., PRENSNER, J. R., EVANS, J. R., ZHAO, S., POLIAKOV, A., CAO, X., DHANASEKARAN, S. M., WU, Y. M., ROBINSON, D. R., BEER, D. G., FENG, F. Y., IYER, H. K. & CHINNAIYAN, A. M. 2015. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet,* 47**,** 199-208.

JANSSEN, H. L., REESINK, H. W., LAWITZ, E. J., ZEUZEM, S., RODRIGUEZ-TORRES, M., PATEL, K., VAN DER MEER, A. J., PATICK, A. K., CHEN, A., ZHOU, Y., PERSSON, R., KING, B. D., KAUPPINEN, S., LEVIN, A. A. & HODGES, M. R. 2013. Treatment of HCV infection by targeting microRNA. *N Engl J Med,* 368**,** 1685-94.

JI, P., DIEDERICHS, S., WANG, W., BOING, S., METZGER, R., SCHNEIDER, P. M., TIDOW, N., BRANDT, B., BUERGER, H., BULK, E., THOMAS, M., BERDEL, W. E., SERVE, H. & MULLER-TIDOW, C. 2003. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene,* 22**,** 8031-41.

JIANG, J., JING, Y., COST, G. J., CHIANG, J. C., KOLPA, H. J., COTTON, A. M., CARONE, D. M., CARONE, B. R., SHIVAK, D. A., GUSCHIN, D. Y., PEARL, J. R., REBAR, E. J., BYRON, M., GREGORY, P. D., BROWN, C. J., URNOV, F. D., HALL, L. L. & LAWRENCE, J. B. 2013. Translating dosage compensation to trisomy 21. *Nature,* 500**,** 296-300.

JIANG, X. 2023. The mechanisms and therapeutic potential of long noncoding RNA NEAT1 in fibrosis. *Clin Exp Med,* 23**,** 3339-3347.
JIANG, X. & NING, Q. 2020. The mechanism of lncRNA H19 in fibrosis and its potential as novel therapeutic target. *Mech Ageing Dev,* 188**,** 111243.
JIN, S. S., LIN, X. F., ZHENG, J. Z., WANG, Q. & GUAN, H. Q. 2019. lncRNA NEAT1 regulates fibrosis and inflammatory response induced by nonalcoholic fatty liver by regulating miR-506/GLI3. *Eur Cytokine Netw,* 30**,** 98-106.
JING, Y., JING, J., YE, L., LIU, X., HARRIS, S. E., HINTON, R. J. & FENG, J. Q. 2017. Chondrogenesis and osteogenesis are one continuous developmental and lineage defined biological process. *Sci Rep,* 7**,** 10020.
JUNG, H. H., LEE, S. H., KIM, J. Y., AHN, J. S., PARK, Y. H. & IM, Y. H. 2016. Statins affect ETS1-overexpressing triple-negative breast cancer cells by restoring DUSP4 deficiency. *Sci Rep,* 6**,** 33035.
KAMOLA, P. J., NAKANO, Y., TAKAHASHI, T., WILSON, P. A. & UI-TEI, K. 2015. The siRNA Non-seed Region and Its Target Sequences Are Auxiliary Determinants of Off-Target Effects. *PLoS Comput Biol,* 11**,** e1004656.
KAPRANOV, P., CHENG, J., DIKE, S., NIX, D. A., DUTTAGUPTA, R., WILLINGHAM, A. T., STADLER, P. F., HERTEL, J., HACKERMULLER, J., HOFACKER, I. L., BELL, I., CHEUNG, E., DRENKOW, J., DUMAIS, E., PATEL, S., HELT, G., GANESH, M., GHOSH, S., PICCOLBONI, A., SEMENTCHENKO, V., TAMMANA, H. & GINGERAS, T. R. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science,* 316**,** 1484-88.
KARAGKOUNI, D., KARAVANGELI, A., PARASKEVOPOULOU, M. D. & HATZIGEORGIOU, A. G. 2021. Characterizing miRNA-lncRNA Interplay. *Methods Mol Biol,* 2372**,** 243-262.
KARBASFOROOSHAN, H. & KARIMI, G. 2017. The role of SIRT1 in diabetic cardiomyopathy. *Biomed Pharmacother,* 90**,** 386-392.
KARIUKI, D., ASAM, K., AOUIZERAT, B. E., LEWIS, K. A., FLOREZ, J. C. & FLOWERS, E. 2023. Review of databases for experimentally validated human microRNA-mRNA interactions. *Database (Oxford),* 2023.
KARTHA, R. V. & SUBRAMANIAN, S. 2014. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front Genet,* 5**,** 8.
KHALIL, A. M., GUTTMAN, M., HUARTE, M., GARBER, M., RAJ, A., RIVEA MORALES, D., THOMAS, K., PRESSER, A., BERNSTEIN, B. E., VAN OUDENAARDEN, A., REGEV, A., LANDER, E. S. & RINN, J. L. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A,* 106**,** 11667-72.
KIM, M., JANG, H. J., BAEK, S. Y., CHOI, K. J., HAN, D. H. & SUNG, J. S. 2023. Regulation of base excision repair during adipogenesis and osteogenesis of bone marrow-derived mesenchymal stem cells. *Sci Rep,* 13**,** 16384.
KONSTANTINOS HATZIKOTOULAS, G. A. P., MATTHEW J CLARK, FAVOUR, FELIX-ILEMHENBHIO, K. K., JONATHAN SIMPSON, SCOTT J MACINNES, MINE, KOPRULU, L. S., ILARIA BELLANTUONO, KAJUS BAIDŽAJEVAS, DAVID A & YOUNG, A. G., ELEFTHERIA ZEGGINI, ENDRE KISS-TOTH, J MARK WILKINSON 2019. Genome-wide association and functional analyses identify CASC20 and KIF26B as target loci in heterotopic ossification. *bioRxiv*.

KOTAKE, Y., NAKAGAWA, T., KITAGAWA, K., SUZUKI, S., LIU, N., KITAGAWA, M. & XIONG, Y. 2011. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene,* 30**,** 1956-62.

KOZOMARA, A. & GRIFFITHS-JONES, S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res,* 42**,** D68-73.

KUME, S., KATO, S., YAMAGISHI, S., INAGAKI, Y., UEDA, S., ARIMA, N., OKAWA, T., KOJIRO, M. & NAGATA, K. 2005. Advanced glycation end-products attenuate human mesenchymal stem cells and prevent cognate differentiation into adipose tissue, cartilage, and bone. *J Bone Miner Res,* 20**,** 1647-58.

KUTTER, C., WATT, S., STEFFLOVA, K., WILSON, M. D., GONCALVES, A., PONTING, C. P., ODOM, D. T. & MARQUES, A. C. 2012. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet,* 8**,** e1002841.

LASDA, E. & PARKER, R. 2014. Circular RNAs: diversity of form and function. *RNA,* 20**,** 1829-42.

LEE, R. C., FEINBAUM, R. L. & AMBROS, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75**,** 843-54.

LI, J., MING, Z., YANG, L., WANG, T., LIU, G. & MA, Q. 2022. Long noncoding RNA XIST: Mechanisms for X chromosome inactivation, roles in sex-biased diseases, and therapeutic opportunities. *Genes Dis,* 9**,** 1478-1492.

LI, L., LIU, B., WAPINSKI, O. L., TSAI, M. C., QU, K., ZHANG, J., CARLSON, J. C., LIN, M., FANG, F., GUPTA, R. A., HELMS, J. A. & CHANG, H. Y. 2013. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell Rep,* 5**,** 3-12.

LI, X., HUANG, T. L., ZHANG, G. D., JIANG, J. T. & GUO, P. Y. 2019. LncRNA ANRIL impacts the progress of osteoarthritis via regulating proliferation and apoptosis of osteoarthritis synoviocytes. *Eur Rev Med Pharmacol Sci,* 23**,** 9729-9737.

LI, Z., LIU, L., FENG, C., QIN, Y., XIAO, J., ZHANG, Z. & MA, L. 2023. LncBook 2.0: integrating human long non-coding RNAs with multi-omics annotations. *Nucleic Acids Res,* 51**,** D186-D191.

LIANG, X. H., SUN, H., NICHOLS, J. G. & CROOKE, S. T. 2017. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Mol Ther,* 25**,** 2075-2092.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LIN, P., CORREA, D., LIN, Y. & CAPLAN, A. I. 2011. Polybrene inhibits human mesenchymal stem cell proliferation during lentiviral transduction. *PLoS One,* 6**,** e23891.

LIN, P., LIN, Y., LENNON, D. P., CORREA, D., SCHLUCHTER, M. & CAPLAN, A. I. 2012. Efficient lentiviral transduction of human mesenchymal stem cells that preserves proliferation and differentiation capabilities. *Stem Cells Transl Med,* 1**,** 886-97.

LINDHOLM, D., BORNHAUSER, B. C. & KORHONEN, L. 2009. Mylip makes an Idol turn into regulation of LDL receptor. *Cell Mol Life Sci,* 66**,** 3399-402.

LIU, J. Y., YAO, J., LI, X. M., SONG, Y. C., WANG, X. Q., LI, Y. J., YAN, B. & JIANG, Q. 2014. Pathogenic role of lncRNA-MALAT1 in endothelial cell dysfunction in diabetes mellitus. *Cell Death Dis,* 5**,** e1506.

LIU, W. & WANG, X. 2019. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol,* 20**,** 18.

MA, C., DU, T., NIU, X. & FAN, Y. 2022a. Biomechanics and mechanobiology of the bone matrix. *Bone Res,* 10**,** 59.

MA, L., BAJIC, V. B. & ZHANG, Z. 2013. On the classification of long non-coding RNAs. *RNA Biol,* 10**,** 925-33.
MA, Q., YANG, L., TOLENTINO, K., WANG, G., ZHAO, Y., LITZENBURGER, U. M., SHI, Q., ZHU, L., YANG, C., JIAO, H., ZHANG, F., LI, R., TSAI, M. C., CHEN, J. A., LAI, I., ZENG, H., LI, L. & CHANG, H. Y. 2022b. Inducible lncRNA transgenic mice reveal continual role of HOTAIR in promoting breast cancer metastasis. *Elife,* 11.
MAEDER, M. L., LINDER, S. J., CASCIO, V. M., FU, Y., HO, Q. H. & JOUNG, J. K. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods,* 10**,** 977-9.
MAI, J., LU, M., GAO, Q., ZENG, J. & XIAO, J. 2023. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol,* 6**,** 899.
MATSUO, M. 2021. Antisense Oligonucleotide-Mediated Exon-skipping Therapies: Precision Medicine Spreading from Duchenne Muscular Dystrophy. *JMA J,* 4**,** 232-240.
MATTICK, J. S. & MAKUNIN, I. V. 2006. Non-coding RNA. *Hum Mol Genet,* 15 Spec No 1**,** R17-29.
MCGEARY, S. E., LIN, K. S., SHI, C. Y., PHAM, T. M., BISARIA, N., KELLEY, G. M. & BARTEL, D. P. 2019. The biochemical basis of microRNA targeting efficacy. *Science,* 366.
MEYERS, C., LISIECKI, J., MILLER, S., LEVIN, A., FAYAD, L., DING, C., SONO, T., MCCARTHY, E., LEVI, B. & JAMES, A. W. 2019. Heterotopic Ossification: A Comprehensive Review. *JBMR Plus,* 3**,** e10172.
MICHALIK, K. M., YOU, X., MANAVSKI, Y., DODDABALLAPUR, A., ZORNIG, M., BRAUN, T., JOHN, D., PONOMAREVA, Y., CHEN, W., UCHIDA, S., BOON, R. A. & DIMMELER, S. 2014. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ Res,* 114**,** 1389-97.
MILLER, S. C., DE SAINT-GEORGES, L., BOWMAN, B. M. & JEE, W. S. 1989. Bone lining cells: structure and function. *Scanning Microsc,* 3**,** 953-60; discussion 960-1.
MISHRA, A., KUMAR, R., MISHRA, S. N., VIJAYARAGHAVALU, S., TIWARI, N. K., SHUKLA, G. C., GURUSAMY, N. & KUMAR, M. 2023. Differential Expression of Non-Coding RNAs in Stem Cell Development and Therapeutics of Bone Disorders. *Cells 2023,* 12**,** 1159.
MISKA, E. A. 2005. How microRNAs control cell division, differentiation and death. *Curr Opin Genet Dev,* 15**,** 563-8.
MITRA, R., LIN, C. C., EISCHEN, C. M., BANDYOPADHYAY, S. & ZHAO, Z. 2015. Concordant dysregulation of miR-5p and miR-3p arms of the same precursor microRNA may be a mechanism in inducing cell proliferation and tumorigenesis: a lung cancer study. *RNA,* 21**,** 1055-65.
MONDAL, T., SUBHASH, S., VAID, R., ENROTH, S., UDAY, S., REINIUS, B., MITRA, S., MOHAMMED, A., JAMES, A. R., HOBERG, E., MOUSTAKAS, A., GYLLENSTEN, U., JONES, S. J., GUSTAFSSON, C. M., SIMS, A. H., WESTERLUND, F., GORAB, E. & KANDURI, C. 2015. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat Commun,* 6**,** 7743.
MONNIER, P., MARTINET, C., PONTIS, J., STANCHEVA, I., AIT-SI-ALI, S. & DANDOLO, L. 2013. H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proc Natl Acad Sci U S A,* 110**,** 20693-98.
NCBI. *Primer-BLAST* [Online]. National Library of Medicine. Available: https://www.ncbi.nlm.nih.gov/tools/primer-blast/ [Accessed 08/05/2025 2025].
NESTEROVA, T. B., SLOBODYANYUK, S. Y., ELISAPHENKO, E. A., SHEVCHENKO, A. I., JOHNSTON, C., PAVLOVA, M. E., ROGOZIN, I. B., KOLESNIKOV, N. N., BROCKDORFF, N.

& ZAKIAN, S. M. 2001. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res,* 11**,** 833-49.

NGUYEN, Q. & CARNINCI, P. 2015. Expression Specificity of Disease-Associated lncRNAs: Toward Personalized Medicine. *Curr Top Microbiol and Immunol,* 394**,** 237-58.

NIE, F. Q., SUN, M., YANG, J. S., XIE, M., XU, T. P., XIA, R., LIU, Y. W., LIU, X. H., ZHANG, E. B., LU, K. H. & SHU, Y. Q. 2015. Long noncoding RNA ANRIL promotes non-small cell lung cancer cell proliferation and inhibits apoptosis by silencing KLF2 and P21 expression. *Mol Cancer Ther,* 14**,** 268-77.

NIH. 2024a. *GART phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase [Homo sapiens (human)]* [Online]. National Library of Medicine. Available: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2618 [Accessed 29/07/2024 2024].

NIH. 2024b. *SELENOH selenoprotein H [Homo sapiens (human)]* [Online]. National Library of Medicine. Available: https://www.ncbi.nlm.nih.gov/gene/280636 [Accessed 29/07/2024 2024].

OKAMURA, K., INAGAKI, Y., MATSUI, T. K., MATSUBAYASHI, M., KOMEDA, T., OGAWA, M., MORI, E. & TANAKA, Y. 2020. RT-qPCR analyses on the osteogenic differentiation from human iPS cells: an investigation of reference genes. *Sci Rep,* 10**,** 11748.

OLIVA-OLIVERA, W., LEIVA GEA, A., LHAMYANI, S., COIN-ARAGUEZ, L., ALCAIDE TORRES, J., BERNAL-LOPEZ, M. R., GARCIA-LUNA, P. P., MORALES CONDE, S., FERNANDEZ-VELEDO, S., EL BEKAY, R. & TINAHONES, F. J. 2015. Differences in the Osteogenic Differentiation Capacity of Omental Adipose-Derived Stem Cells in Obese Patients With and Without Metabolic Syndrome. *Endocrinology,* 156**,** 4492-501.

PALLY, D., KAPOOR, N. & NABA, A. 2025. The novel ECM protein SNED1 mediates cell adhesion via the RGD-binding integrins alpha5beta1 and alphavbeta3. *J Cell Sci,* 138.

PANG, K. C., FRITH, M. C. & MATTICK, J. S. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet,* 22**,** 1-5.

PITTENGER, M. F., MACKAY, A. M., BECK, S. C., JAISWAL, R. K., DOUGLAS, R., MOSCA, J. D., MOORMAN, M. A., SIMONETTI, D. W., CRAIG, S. & MARSHAK, D. R. 1999. Multilineage potential of adult human mesenchymal stem cells. *Science,* 284**,** 143-7.

PODVIN, S., GONZALEZ, A. M., MILLER, M. C., DANG, X., BOTFIELD, H., DONAHUE, J. E., KURABI, A., BOISSAUD-COOKE, M., ROSSI, R., LEADBEATER, W. E., JOHANSON, C. E., COIMBRA, R., STOPA, E. G., ELICEIRI, B. P. & BAIRD, A. 2011. Esophageal cancer related gene-4 is a choroid plexus-derived injury response gene: evidence for a biphasic response in early and late brain injury. *PLoS One,* 6**,** e24609.

PONTING, C. P. & HAERTY, W. 2022. Genome-Wide Analysis of Human Long Noncoding RNAs: A Provocative Review. *Annu Rev Genomics Hum Genet,* 23**,** 153-172.

PROUDFOOT, N. J. 2011. Ending the message: poly(A) signals then and now. *Genes Dev,* 25**,** 1770-82.

QIN, S., PREDESCU, D., CARMAN, B., PATEL, P., CHEN, J., KIM, M., LAHM, T., GERACI, M. & PREDESCU, S. A. 2021. Up-Regulation of the Long Noncoding RNA X-Inactive-Specific Transcript and the Sex Bias in Pulmonary Arterial Hypertension. *Am J Pathol,* 191**,** 1135-1150.

QU, X., LIU, B., WANG, L., LIU, L., ZHAO, W., LIU, C., DING, J., ZHAO, S., XU, B., YU, H., ZHANG, X. & CHAI, J. 2023. Loss of cancer-associated fibroblast-derived exosomal DACT3-AS1

promotes malignant transformation and ferroptosis-mediated oxaliplatin resistance in gastric cancer. *Drug Resist Updat,* 68**,** 100936.
RASKOVIC, B., POPOVIC, M., OSTOJIC, S., ANDELKOVIC, B., TESEVIC, V. & POLOVIC, N. 2015. Fourier transform infrared spectroscopy provides an evidence of papain denaturation and aggregation during cold storage. *Spectrochim Acta A Mol Biomol Spectrosc,* 150**,** 238-46.
REN, Y., WANG, Y. F., ZHANG, J., WANG, Q. X., HAN, L., MEI, M. & KANG, C. S. 2019. Targeted design and identification of AC1NOD4Q to block activity of HOTAIR by abrogating the scaffold interaction with EZH2. *Clin Epigenetics,* 11**,** 29.
RINN, J. L., KERTESZ, M., WANG, J. K., SQUAZZO, S. L., XU, X., BRUGMANN, S. A., GOODNOUGH, L. H., HELMS, J. A., FARNHAM, P. J., SEGAL, E. & CHANG, H. Y. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell,* 129**,** 1311-23.
SABOKBAR, A., MILLETT, P. J., MYER, B. & RUSHTON, N. 1994. A rapid, quantitative assay for measuring alkaline phosphatase activity in osteoblastic cells in vitro. *Bone Miner,* 27**,** 57-67.
SAIGUSA, S., INOUE, Y., TANAKA, K., TOIYAMA, Y., OKUGAWA, Y., SHIMURA, T., HIRO, J., UCHIDA, K., MOHRI, Y. & KUSUNOKI, M. 2013. Decreased expression of DUSP4 is associated with liver and lung metastases in colorectal cancer. *Med Oncol,* 30**,** 620.
SALMENA, L., POLISENO, L., TAY, Y., KATS, L. & PANDOLFI, P. P. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell,* 146**,** 353-8.
SARROPOULOS, I., MARIN, R., CARDOSO-MOREIRA, M. & KAESSMANN, H. 2019. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature,* 571**,** 510-514.
SCHLIMGEN, R., HOWARD, J., WOOLEY, D., THOMPSON, M., BADEN, L. R., YANG, O. O., CHRISTIANI, D. C., MOSTOSLAVSKY, G., DIAMOND, D. V., DUANE, E. G., BYERS, K., WINTERS, T., GELFAND, J. A., FUJIMOTO, G., HUDSON, T. W. & VYAS, J. M. 2016. Risks Associated With Lentiviral Vector Exposures and Prevention Strategies. *J Occup Environ Med,* 58**,** 1159-1166.
SIMS, N. A. & GOOI, J. H. 2008. Bone remodeling: Multiple cellular interactions required for coupling of bone formation and resorption. *Semin Cell Dev Biol,* 19**,** 444-51.
SOMAROWTHU, S., LEGIEWICZ, M., CHILLON, I., MARCIA, M., LIU, F. & PYLE, A. M. 2015. HOTAIR forms an intricate and modular secondary structure. *Mol Cell,* 58**,** 353-61.
SONG, G., ZHOU, J., SONG, R., LIU, D., YU, W., XIE, W., MA, Z., GONG, J., MENG, H., YANG, T. & SONG, Z. 2020. Long noncoding RNA H19 regulates the therapeutic efficacy of mesenchymal stem cells in rats with severe acute pancreatitis by sponging miR-138-5p and miR-141-3p. *Stem Cell Res Ther,* 11**,** 420.
SORENSEN, K. P., THOMASSEN, M., TAN, Q., BAK, M., COLD, S., BURTON, M., LARSEN, M. J. & KRUSE, T. A. 2013. Long non-coding RNA HOTAIR is an independent prognostic marker of metastasis in estrogen receptor-positive primary breast cancer. *Breast Cancer Res Treat,* 142**,** 529-36.
SPITALE, R. C., TSAI, M. C. & CHANG, H. Y. 2011. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics,* 6**,** 539-43.
SUN, C. C., ZHU, W., LI, S. J., HU, W., ZHANG, J., ZHUO, Y., ZHANG, H., WANG, J., ZHANG, Y., HUANG, S. X., HE, Q. Q. & LI, D. J. 2020. FOXC1-mediated LINC00301 facilitates tumor progression and triggers an immune-suppressing microenvironment in non-small cell lung cancer by regulating the HIF1alpha pathway. *Genome Med,* 12**,** 77.

SUN, N., ZHANG, G. & LIU, Y. 2018. Long non-coding RNA XIST sponges miR-34a to promotes colon cancer progression via Wnt/beta-catenin signaling pathway. *Gene,* 665**,** 141-148.

SWAYZE, E. E., SIWKOWSKI, A. M., WANCEWICZ, E. V., MIGAWA, M. T., WYRZYKIEWICZ, T. K., HUNG, G., MONIA, B. P. & BENNETT, C. F. 2007. Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals. *Nucleic Acids Res,* 35**,** 687-700.

TAY, Y., RINN, J. & PANDOLFI, P. P. 2014. The multilayered complexity of ceRNA crosstalk and competition. *Nature,* 505**,** 344-52.

TSAI, M. C., MANOR, O., WAN, Y., MOSAMMAPARAST, N., WANG, J. K., LAN, F., SHI, Y., SEGAL, E. & CHANG, H. Y. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science,* 329**,** 689-93.

TSAI, M. T., LI, W. J., TUAN, R. S. & CHANG, W. H. 2009. Modulation of osteogenesis in human mesenchymal stem cells by specific pulsed electromagnetic field stimulation. *J Orthop Res,* 27**,** 1169-74.

UGAI, H., DOBBINS, G. C., WANG, M., LE, L. P., MATTHEWS, D. A. & CURIEL, D. T. 2012. Adenoviral protein V promotes a process of viral assembly through nucleophosmin 1. *Virology,* 432**,** 283-95.

ULITSKY, I., SHKUMATAVA, A., JAN, CALVIN H., SIVE, H. & BARTEL, DAVID P. 2012. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell,* 151**,** 684-686.

VAN DER REE, M. H., VAN DER MEER, A. J., VAN NUENEN, A. C., DE BRUIJNE, J., OTTOSEN, S., JANSSEN, H. L., KOOTSTRA, N. A. & REESINK, H. W. 2016. Miravirsen dosing in chronic hepatitis C patients results in decreased microRNA-122 levels without affecting other microRNAs in plasma. *Aliment Pharmacol Ther,* 43**,** 102-13.

VAN DONGEN, S., ABREU-GOODGER, C. & ENRIGHT, A. J. 2008. Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods,* 5**,** 1023-5.

VICKERS, T. A., WYATT, J. R., BURCKIN, T., BENNETT, C. F. & FREIER, S. M. 2001. Fully modified 2' MOE oligonucleotides redirect polyadenylation. *Nucleic Acids Res,* 29**,** 1293-9.

WANG, H., ZHENG, H., WANG, C., LU, X., ZHAO, X. & LI, X. 2017. Insight into HOTAIR structural features and functions as landing pads for transcription regulation proteins. *Biochem Biophys Res Commun,* 485**,** 679-685.

WANG, K. C. & CHANG, H. Y. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell,* 43**,** 904-14.

WANG, N., HOU, M., ZHAN, Y. & SHENG, X. 2019. LncRNA PTCSC3 inhibits triple-negative breast cancer cell proliferation by downregulating lncRNA H19. *J Cell Biochem,* 120**,** 15083-15088.

WANG, Q., LI, Y., ZHANG, Z., FANG, Y., LI, X., SUN, Y., XIONG, C., YAN, L. & ZHAO, J. 2015. Bioinformatics analysis of gene expression profiles of osteoarthritis. *Acta Histochem,* 117**,** 40-6.

WANG, X. 2014. Composition of seed sequence is a major determinant of microRNA targeting patterns. *Bioinformatics,* 30**,** 1377-83.

WASHIETL, S., KELLIS, M. & GARBER, M. 2014. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res,* 24**,** 616-28.

WIDHOLZ, B., TSITLAKIDIS, S., REIBLE, B., MOGHADDAM, A. & WESTHAUSER, F. 2019. Pooling of Patient-Derived Mesenchymal Stromal Cells Reduces Inter-Individual Confounder-

Associated Variation without Negative Impact on Cell Viability, Proliferation and Osteogenic Differentiation. *Cells,* 8.

WIGHTMAN, B., HA, I. & RUVKUN, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell,* 75**,** 855-62.

WILCZYNSKA, A. & BUSHELL, M. 2015. The complexity of miRNA-mediated repression. *Cell Death Differ,* 22**,** 22-33.

WINKLE, M., EL-DALY, S. M., FABBRI, M. & CALIN, G. A. 2021. Noncoding RNA therapeutics - challenges and potential solutions. *Nat Rev Drug Discov,* 20**,** 629-651.

WOODROW, K. A., CU, Y., BOOTH, C. J., SAUCIER-SAWYER, J. K., WOOD, M. J. & SALTZMAN, W. M. 2009. Intravaginal gene silencing using biodegradable polymer nanoparticles densely loaded with small-interfering RNA. *Nat Mater,* 8**,** 526-33.

WU, G., CAI, J., HAN, Y., CHEN, J., HUANG, Z. P., CHEN, C., CAI, Y., HUANG, H., YANG, Y., LIU, Y., XU, Z., HE, D., ZHANG, X., HU, X., PINELLO, L., ZHONG, D., HE, F., YUAN, G. C., WANG, D. Z. & ZENG, C. 2014. LincRNA-p21 regulates neointima formation, vascular smooth muscle cell proliferation, apoptosis, and atherosclerosis by enhancing p53 activity. *Circulation,* 130**,** 1452-1465.

WU, H., WAHANE, A., ALHAMADANI, F., ZHANG, K., PARIKH, R., LEE, S., MCCABE, E. M., RASMUSSEN, T. P., BAHAL, R., ZHONG, X. B. & MANAUTOU, J. E. 2022. Nephrotoxicity of marketed antisense oligonucleotide drugs. *Curr Opin Toxicol,* 32.

WU, M., CHEN, G. & LI, Y. P. 2016. TGF-beta and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res,* 4**,** 16009.

XI, Y., HUANG, H., ZHAO, Z., MA, J. & CHEN, Y. 2020. Tissue inhibitor of metalloproteinase 1 suppresses growth and differentiation of osteoblasts and differentiation of osteoclasts by targeting the AKT pathway. *Exp Cell Res,* 389**,** 111930.

XIA, Y., HE, Z., LIU, B., WANG, P. & CHEN, Y. 2015. Downregulation of Meg3 enhances cisplatin resistance of lung cancer cells through activation of the WNT/beta-catenin signaling pathway. *Mol Med Rep,* 12**,** 4530-4537.

XU, R., ZHU, X., CHEN, F., HUANG, C., AI, K., WU, H., ZHANG, L. & ZHAO, X. 2018. LncRNA XIST/miR-200c regulates the stemness properties and tumourigenicity of human bladder cancer stem cell-like cells. *Cancer Cell Int,* 18**,** 41.

YAN, X. W., LIU, H. J., HONG, Y. X., MENG, T., DU, J. & CHANG, C. 2022. lncRNA XIST induces Abeta accumulation and neuroinflammation by the epigenetic repression of NEP in Alzheimer's disease. *J Neurogenet,* 36**,** 11-20.

YANG, W., XIA, Y., QIAN, X., WANG, M., ZHANG, X., LI, Y. & LI, L. 2019. Co-expression network analysis identified key genes in association with mesenchymal stem cell osteogenic differentiation. *Cell Tissue Res,* 378**,** 513-529.

YAO, R. W., WANG, Y. & CHEN, L. L. 2019. Cellular functions of long noncoding RNAs. *Nat Cell Biol,* 21**,** 542-551.

YE, X. & ZHANG, C. 2017. Effects of Hyperlipidemia and Cardiovascular Diseases on Proliferation, Differentiation and Homing of Mesenchymal Stem Cells. *Curr Stem Cell Res Ther,* 12**,** 377-387.

YOUNESS, A., MIQUEL, C. H. & GUERY, J. C. 2021. Escape from X Chromosome Inactivation and the Female Predominance in Autoimmune Diseases. *Int J Mol Sci,* 22.

YOUNG, M. D., WAKEFIELD, M. J., SMYTH, G. K. & OSHLACK, A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol,* 11**,** R14.

YU, F., JIANG, Z., CHEN, B., DONG, P. & ZHENG, J. 2017. NEAT1 accelerates the progression of liver fibrosis via regulation of microRNA-122 and Kruppel-like factor 6. *J Mol Med (Berl),* 95**,** 1191-1202.

YUE, D., GUANQUN, G., JINGXIN, L., SEN, S., SHUANG, L., YAN, S., MINXUE, Z., PING, Y., CHONG, L., ZHUOBO, Z. & YAFEN, W. 2020. Silencing of long noncoding RNA XIST attenuated Alzheimer's disease-related BACE1 alteration through miR-124. *Cell Biol Int,* 44**,** 630-636.

ZAIM, M., KARAMAN, S., CETIN, G. & ISIK, S. 2012. Donor age and long-term culture affect differentiation and proliferation of human bone marrow mesenchymal stem cells. *Ann Hematol,* 91**,** 1175-86.

ZHANG, H., DU, Y., LU, D., WANG, X., LI, Y., QING, J., ZHANG, Y., LIU, H., LV, L., ZHANG, X., LIU, Y., ZHOU, Y. & ZHANG, P. 2024. UBE2C orchestrates bone formation through stabilization of SMAD1/5. *Bone,* 187**,** 117175.

ZHANG, J., LIN, Z., GAO, Y. & YAO, T. 2017a. Downregulation of long noncoding RNA MEG3 is associated with poor prognosis and promoter hypermethylation in cervical cancer. *J Exp Clin Cancer Res,* 36**,** 5.

ZHANG, W., SHI, S., JIANG, J., LI, X., LU, H. & REN, F. 2017b. LncRNA MEG3 inhibits cell epithelial-mesenchymal transition by sponging miR-421 targeting E-cadherin in breast cancer. *Biomed Pharmacother,* 91**,** 312-319.

ZHANG, X. & XU, J. 2024. A novel miR-466l-3p/FGF23 axis promotes osteogenic differentiation of human bone marrow mesenchymal stem cells. *Bone,* 185**,** 117123.

ZHANG, Y., WU, H., WANG, F., YE, M., ZHU, H. & BU, S. 2018. Long non-coding RNA MALAT1 expression in patients with gestational diabetes mellitus. *Int J Gynaecol Obstet,* 140**,** 164-169.

ZHAO, R., FU, J., ZHU, L., CHEN, Y. & LIU, B. 2022. Designing strategies of small-molecule compounds for modulating non-coding RNAs in cancer therapy. *J Hematol Oncol,* 15**,** 14.

ZHENG, R., LIN, S., GUAN, L., YUAN, H., LIU, K., LIU, C., YE, W., LIAO, Y., JIA, J. & ZHANG, R. 2018. Long non-coding RNA XIST inhibited breast cancer cell growth, migration, and invasion via miR-155/CDX1 axis. *Biochem Biophys Res Commun,* 498**,** 1002-1008.

ZHONG, M., WU, Z., CHEN, Z., WU, L. & ZHOU, J. 2024. Geniposide alleviates cholesterol-induced endoplasmic reticulum stress and apoptosis in osteoblasts by mediating the GLP-1R/ABCA1 pathway. *J Orthop Surg Res,* 19**,** 179.

ZHOU, L. J., YANG, D. W., OU, L. N., GUO, X. R. & WU, B. L. 2020. Circulating Expression Level of LncRNA Malat1 in Diabetic Kidney Disease Patients and Its Clinical Significance. *J Diabetes Res,* 2020**,** 4729019.

ZHOU, Y., ZHONG, Y., WANG, Y., ZHANG, X., BATISTA, D. L., GEJMAN, R., ANSELL, P. J., ZHAO, J., WENG, C. & KLIBANSKI, A. 2007. Activation of p53 by MEG3 non-coding RNA. *J Biol Chem,* 282**,** 24731-42.

ZHU, Y., ZHU, L., WANG, X. & JIN, H. 2022. RNA-based therapeutics: an overview and prospectus. *Cell Death Dis,* 13**,** 644.