Universidade do Algarve

Characterization of 5' and 3' UTRs from Bone morphogenetic protein receptor type 1A (Bmpr1A) transcripts under the context of bone metabolism

Joel Mourato da Silva

Master's in molecular and Microbial Biology

Work under the supervision of:

Professor Leonor Cancela

Professor Márcio Simão

FCT-Faculdade de Ciências e Tecnologia

2021

Universidade do Algarve

Characterization of 5' and 3' UTRs from Bone morphogenetic protein receptor type 1A (Bmpr1A) transcripts under the context of bone metabolism

Joel Mourato da Silva

Master's in molecular and Microbial Biology

Work under the supervision of:

Professor Leonor Cancela

Professor Márcio Simão

FCT-Faculdade de Ciências e Tecnologia

2021

Characterization of 5' and 3' UTRs from Bone morphogenetic protein receptor type 1A (Bmpr1A) transcripts under the context of bone metabolism

Authorship Statement

I hereby declare to be the author of this work, which is original and unpublished. Authors and papers consulted are duly cited in the text and are listed in the included references.

Acknowledgments/Agradecimentos

Antes de mais queria agradecer aos meus 2 orientadores, professora Doutora Leonor Cancela e ao professor Márcio Simão por todo o tempo despendido comigo.

Queria também agradecer a todos os colegas de laboratório que me ajudaram para que esta tese fosse possível particularmente à Débora e Tatiana Varela que me aturaram muitas vezes e explicaram o que lhes perguntava mesmo se tivessem bastante ocupadas, à Helena Caiado que também me ajudou muito sempre que tinha algumas dúvidas. Queria ainda agradecer ao Gil Martins por todas as conversas sobre impressão 3D e sempre disponível para ajudar no que fosse preciso e possível. Por último queria agradecer também ao Sunil Poudel por todas as científicas e tentativas de ver outras possibilidades para o trabalho que estava a desenvolver.

Queria agradecer ao Dr. Vigo Heissmeyer por prontamente nos ter enviado uns plasmídeos que vamos usar para desenvolver este trabalho futuramente.

Queria agradecer à professora Filomena Fonseca por me ter ajudado num momento muito difícil do meu mestrado.

Queria agradecer aos meus amigos particularmente ao Diogo Teixeira, Alexandra Pereira, Pedro Souza, Tiago Rosário e Rita João por sempre me aturarem quando saía do laboratório mais chateado ou quando as coisas não estavam a correr bem, ou quando falamos sobre ciência.

Por último, mas não menos importante queria agradecer aos meus pais por serem compreensivos enquanto tirava o mestrado e por me apoiarem em todos os momentos.

Abstract

During our life our bones are in a constant balance between bone formation by osteoblasts and bone resorption by osteoclasts. Bone morphogenetic proteins are a group of cytokines from the transforming growth factor-β family involved in several processes. Their signal is transduced through two types of serine/threonine kinase receptors, BMPRI and BMPRII. BMPRIA is a type one receptor involved in osteoblast/osteoclast communication and therefore affecting bone metabolism. Upon binding of different ligands, type II phosphorylates the type I and activates one of the signal pathways. Although their function is known, the mechanisms of regulation of expression are still unclear.

The objective of this work was to characterize some of the *Mus musculus Bmpr1a* molecular regulatory mechanisms including upstream open reading frames (uORFs), polyadenylation sites, microRNA binding sites and constitutive decay elements (CDE).

Bioinformatically, we were able to identify 11 *Mus musculus Bmpr1a* transcripts on available databases, 11 polyadenylations and a constitutive decay element on the 3'UTR, while on 5'UTR we were able to identify 3 uORFs. Experimentally, we isolated a new and shorter *Bmpr1a* 3'UTR with an alternative polyadenylation site from MC3T3-E1 but without the CDE We tried to isolate different fragments from different tissues, but we were capable of isolate a longer transcript only from mouse liver that contains the CDE loop. Within the 5'UTRs we isolated 4 different fragments containing 2 or 3 uORFs and insert them on reporter plasmids for functional analysis. The last step was mutating all the AUG from uORFs and validating their effect on regulation of luciferase expression. We measured a decrease in the expression of luciferase on fragments containing no mutations and an increase in the expression of fragments with mutations, except in one case where the mutation was inserted 7bp after the start of the fragment. Results indicate that there were no different transcripts from proliferating and differentiated cells, and in the 5' UTR , the uORFs could possibly contribute to regulate the expression of Bmpr1a. However further studies are needed to confirm and expand our data.

Keywords: Bone, BMP, BMPR1a, uORFs, CDE

Resumo

As *Bone morphogenetic proteins* (BMP) são um grupo de citocinas pertencentes à família de proteínas do *Transforming growth factor β* e estão envolvidas em vários processos celulares importantes como o regulamento da divisão celular e de diferenciação de células. Estas proteínas transmitem o sinal para dentro das células através de dois tipos de recetores de serina/treonina diferentes, os *Bone morphogenetic proteins receptors* (BMPR) tipo I e tipo II. Estes recetores formam homo e hétero dímeros entre si para formar um recetor totalmente funcional; quando isso acontece e após a ligação de uma molécula aos receptores, o tipo II fosforila o tipo I levando a uma cascata de fosforilação no interior da célula podendo ativar diferentes vias se sinalização dependendo da molécula que se liga ao recetor e qual os dímeros de recetores que se formaram. Alguns dos exemplos de vias de sinalização que se ativam aquando da ligação de moléculas a estes recetores são as vias *Smad, MAPK, PI3K/Akt, Wnt* ou *ERK1/2* que são importantes reguladores da expressão de genes e que atuam em processos celulares importantes como diferenciação celular e formação de tecidos. As BMPRs estão envolvidas também nos processos de metabolismo do osso associados à deposição de hidroxiapatite pelos osteoblastos e reabsorção óssea pelos osteoclastos que degradam a matriz extracelular. Apesar de se conhecer a função da BMPR1A, nem todos os mecanismos de regulação pós-transcrição da sua expressão estão estudados. O objetivo deste trabalho é contribuir para a caracterização dos mecanismos de regulação dos diferentes transcritos de *Bmpr1a*, e determinar quais os elementos regulatórios pós-transcrição presentes nas sequências identificadas, tanto na região 5' *untraslated region (*UTR), como na região 3'UTR.

Começou-se por identificar os diferentes transcritos de *bmpr1a* sendo possível identificar 11 diferentes transcritos de *Bmpr1a* de *mus musculus* (ratinho) e 5 diferentes transcritos de *BMPR1A* de *homo sapiens* (humano) em diferentes bases de dados. Procurou-se também identificar sequencias associadas a locais de poliadenilação sendo que no ratinho foi possível encontrar 11 diferentes locais de poliadenilação, dos quais 6 contêm a sequencia canónica de poliadenilação, enquanto os outros 5 foram obtidos através do uso de ferramentas bioinformáticas como o RegRNA2.0. Nos transcritos de *homo sapiens* identificou se 8 locais de poliadenilação. De seguida identificou se a presença de um *loop* de degradação constitutivo (*Constitutive Decay Element* - CDE). Este *loop* está presente em 8 das 11 sequencias de ratinho identificadas enquanto no humano está presente em todos os transcritos. Verificou-se igualmente a conservação deste *loop* em 24 espécies diferentes e foi possível verificar que se encontra totalmente conservado em 18 dessas 24

espécies e em 22 delas contem a sequencia necessária para que seja funcional. Outro dos elementos regulatórios que se identificou foram locais ricos em Adeninas/Uracilos que são associados a um aumento de afinidade de ligação de proteínas reguladoras. Foi possível identificar a sequencia mínima para a ligação dessas proteínas, mas não foi possível encontrar um dos 5 diferentes conjuntos de sequências que estão descritas na literatura. Por último construiu-se uma árvore filogenética da região 3'UTR dos transcritos de *Bmpr1a* para perceber as relações filogenéticas desta região não codificante entre diferentes espécies de vertebrados. Na região 5' identificou-se a presença de *upstream Open Reading frames (uORFs)* tendo sido verificada a presença de 3 uORFs conservadas na maior parte dos transcritos analisados e conservados em várias espécies de mamíferos exceto nos marsupiais.

Experimentalmente, foi possível isolar um transcrito da região 3'UTR de *Bmpr1a* de ratinho com 250 pares de bases tanto em células em proliferação como em células diferenciadas em osteoblastos. Este transcrito contém um local de poliadenilação alternativo antes da cauda de poliadeninas e não contém o *loop* de degradação constitutivo. Foi-se então tentar isolar transcritos de outros tecidos de ratinho. Começou se por tentar isolar transcritos de *Bmpr1a* em tecidos de joelho e de articulações não sendo possível isolar transcritos de outros tamanhos além do transcrito de 250 pares de bases acima indicado. No entanto foi possível identificar a partir de fígado um transcrito com 1500 pares de base o qual já contem o *loop* de degradação constitutivo. Na região 5' isolamos 4 fragmentos diferentes contendo 2 ou 3 das uORFs e com diferentes distâncias entre estas e o início dos fragmentos. Esses 4 fragmentos foram de seguida inseridos com a orientação correta num vetor de expressão para a luciferase e transfectados em células Hek293 (do inglês *Human Embrionic Kidney*) e os valores de luminescência registados. Observou-se nos resultados preliminares uma diminuição da expressão da luciferase associada à presença das uORFs dos transcritos sem mutações. Apos a mutação dos AUGs para AAG por SDM verificou-se que no fragmento mais longo (fragmento A) parece haver uma tendência para o aumento da luciferase ao longo das mutações sequenciais dos AUG, confirmado pelas mutações individuais. No segundo fragmento (fragmento B) analisado ocorre um aumento da expressão da luciferase quando se mutou o AUG da uORF1 em comparação com o controlo, contudo na mutação da uORF2 ocorre uma diminuição dramática da expressão da luciferase que é confirmada pela mutação isolada da uORF. Uma das possíveis explicações para esta diminuição da expressão da luciferase é a localização desta uORF no fragmento, pois esta uORF está localizada no início do fragmento, a 7 pares de base

do início da mesma, podendo levar a uma destabilização deste fragmento e consequentemente levando ao resultado observado, ou seja, à diminuição da expressão da luciferase. No último fragmento (fragmento D) que não contem a uORF2 é possível observar uma diminuição da expressão da luciferase no fragmento sem nenhuma mutação dos AUG, enquanto um aumento da expressão da luciferase nos fragmentos que contem as mutações sequenciais. Sendo que apenas duas experiências foram efetuadas com este fragmento, mais estudos são necessários para determinar se a ausência da uORF2 tem algum efeito na regulação da expressão deste fragmento

Podemos então com este trabalho concluir que na região 3'UTR se encontra um *loop* de degradação constitutivo que está bastante conservado, podendo assim significar que aquela região pode ser importante para a regulação da expressão do gene. Contudo experimentalmente não foi possível isolar das células MC3T3-E1 um fragmento dessa região, podendo ser devido à necessidade destas células de ratinho expressarem esta proteína que está envolvida na comunicação entre osteoblastos e osteoclastos. Experimentalmente podemos também verificar que existem na região 5'UTR do *Bmpr1a* 3 uORFs que podem, portanto, vir a ter efeito na regulação da expressão do gene, mas mais estudos precisam de ser feitos para confirmar ou não esta atividade.

Palavras-chave: BMP, *Bmpr1a*, CDE, uORFs.

Table of contents

List of Figures

List of Tables

Abbreviations

A20- Tumor necrosis factor, alpha-induced protein 3

ActR - Nuclear Receptor Coactivator

ALK- Activin Receptor-Like Kinase

ALP - Alkaline Phosphatase

ARE- A/U rich Elements

BMP- Bone Morphogenetic Protein

BMPR- Bone Morphogenetic protein Receptors

CAF-1- Chromatin Assembly Factor-1

Ccr4- Motif Chemokine Receptor 4

CDE- Constitutive Decay Element

CTLA-4- Cytotoxic T-Lymphocyte Associated Protein 4

CSDC2- Cold Shock Domain Containing C2

DCP2- Decapping MRNA 2

DDX6/RCK- DEAD-Box Helicase 6

Dkk- Dickkopf

DNA- Deoxyribonucleic acid

eIF- eukaryotic Initiation Factors

ESEs- Exonic Splicing Enhancers

ESSs- Exonic Splicing Silencers

ERK- Extracellular Signal-regulated Kinase

GPI- Glycosylphosphatidylinositol

Icos- Inducible T Cell Costimulator

IHH- Indian Hedgehog

IL- Interlukines

Irf4- Interferon Regulatory Factor 4

ISSs- Intronic Splicing Silencers

ITCH - Itchy E3 Ubiquitin Protein Ligase

JNK- c-Jun NH2-terminal Kinase

MAPK- Mitogen Activated Protein Kinase

MEF- Mouse Embryonic Fibroblast

MSC- Mesenchymal Stem Cells

mTOR- mechanistic Target Of Rapamycin

Nfkbid- Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor delta

NMD -Nonsense Mediated Decay

NOT- Negative Regulator Of Transcription Subunit

Ocn- Osteocalcin

OSX- Osterix

PI3K/Akt- Phosphoinositide 3-kinase/ Protein kinase B

PTEN- Phosphatase and Tensin Homolog

RANKL- Receptor Activator of Nuclear factor-κB ligand

RBP- RNA Binding Proteins

RGM- Repulsive Guidance Molecules

RISC- RNA Induced Silencing complex

RNA-Ribonucleic acid

RUNX2- Runt-related Transcription Factor 2

SHH- Sonic Hedgehog

snRNP- small nuclear Ribonucleoprotein

Smad- Gene products of the *Drosophila* gene 'mothers against decapentaplegic' (Mad) and the *C. elegans* gene Sma (Smad)

Sost- Sclerostin

SPC- Furin and/or Subtilisin/like Proprotein Convertases

TGF-β- Transforming Growth Factor

TNF- Tumor Necrosis Factor

uORF- upstream Open Reading Frame

Wnt- Wingless-related integration site

1 Introduction

1.1. Gene expression

The term gene was used for the first time by Wilhelm Johannsen in 1909 and since then it had different definitions. The current one says that a gene is "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions", and this term was reported by the Sequence Ontology Consortium[1,2]. A gene is composed by several elements including a promoter in the 5´flanking region, as well as enhancer and inhibitor sequences usually located in the 5' side of a given gene, commonly within the first introns. However, in some cases the enhancers/repressors can also be found at large distances from the promoter, either downstream or upstream[3,4]. The transcription of a given gene starts by the ligation of a complex of proteins to the promoter that, in combination with those enhancers/repressors, are responsible for determining which genes are transcribed in the different cell types and in specific developmental stages[3]. The levels of expression are determined by the combination of transcription factors and those enhancers/repressors. Following the promoter, the gene initiates in the transcription start site and, when encoding a protein, it contains a coding sequence called open reading frame (ORF), located between the start and stop codon which code for the final protein. In its 3' there is a terminator sequence marking the end of transcription and the release of the RNA polymerase[4].

After transcription, regulation of expression can occur at different stages of post-transcriptional processing by adding or removing some sequences from the transcript as the introns or the poly-(A) and cap addition, nuclear export, RNA processing[5] and finally post-translational modifications as methylation or phosphorylation[6,7].

At the messenger RNA (mRNA) level, post-transcriptional regulation occurs in different forms, some are well described as transcripts with alternative lengths due to polyadenylation sites (PAS) or alternative splicing (AS), which regulates the presence in the transcripts of binding signature sequences for RNA-binding proteins (RBP) or micro-RNA's (miRNA) that affect mRNA stability and therefore the final expression of a given gene. These regulatory elements present in the mRNA are called *cis-elements* and act upstream or downstream of the gene known as 5' untranslated region (UTR) and 3' UTR respectively. 5'UTR and 3'UTR are also important in mRNA stability and

regulatory protein localization[8]. After the transcription of a certain gene, its mRNA passes through several maturation processes regulated by the presence of several cis-acting elements. First Cap is added to the 5'end, pre-mRNA is spliced by the spliceosome and in the 3'end a poly-(A) tail is added.[9] Simultaneously, the spliceosome cleaves the introns present on the mRNA. Some of the regulatory elements are removed from the transcript while others are kept in place contributing for the regulation of mRNA stability and gene expression at the post-transcription level[10,11]. When the mRNA becomes mature, the 5'UTR cap-binding complex interacts with the 3'UTR associated with Poly-(A) binding proteins stabilizing the mRNA in an almost circular structure. When this happens the 3' regulatory elements are closer to the translation initiation site and can also have a role in the mRNA stability and intracellular localization[12] .

1.1.1. Splicing and Alternative splicing

Alternative splicing is a crucial mechanism for gene regulation and it has been estimated that almost 95% of mammalian genes contain alternative splicing[13]. Several genes important for developmental processes have alternative splicing events demonstrating its role in organisms development[14]. Through splicing and alternative splicing it is possible to obtain more than one mRNA from a single gene, allowing the synthesis of different mRNAs that can lead to the expression of several protein isoforms promoting alternative combinations of regulatory elements in mRNA sequence and affecting its stability, localization and translation.

Splicing of pre-mRNAs allows different combinations of exons to form distinct mature mRNAs. The ligation of the spliceosome is determined by the presence of specific *cis*-regulatory elements present in the sequence and those elements are classified depending on their function and position. They are known as exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs).[13,15] The spliceosome functions with the help of multiple auxiliary proteins[16] and this complex is composed of five small nuclear ribonucleoprotein particles (snRNP) U1, U2, U4, U5, and U6 and more than a hundred other proteins.[17,18] Splicing starts with the assembly of the spliceosome on the 5'splice site by the small nuclear ribonucleoprotein (snRNP) U1 and the binding of splicing factor 1 (SF1) to the branch point, then auxiliary factor heterodimer U2 (U2AF) binds to terminal AG. After the ligation of SF1

it is replaced by U2. [15] The snRNP U4/U6–U5 are then recruited completing the B complex that is converted to its catalytic active form (C complex) after some remodelling and conformational changes. Although the ligation of those early factors is fundamental for the assembly of the complete spliceosome to the pre-mRNA, the decision of which intron is removed can be made at different stages of spliceosome assembly[15].

Alternative splicing leads to insertion or deletion of exons that could contain regulatory elements which were responsible for regulation of gene expression.[14,15,17]Another way of regulation by alternative splicing is through the splicing events leading to an introduction of a premature stop codon and to the activation of Nonsense-mediated Decay(NMD).[13]

Competition between splicing factors, cleavage and polyadenylation factors can influence splicing and polyadenylation.

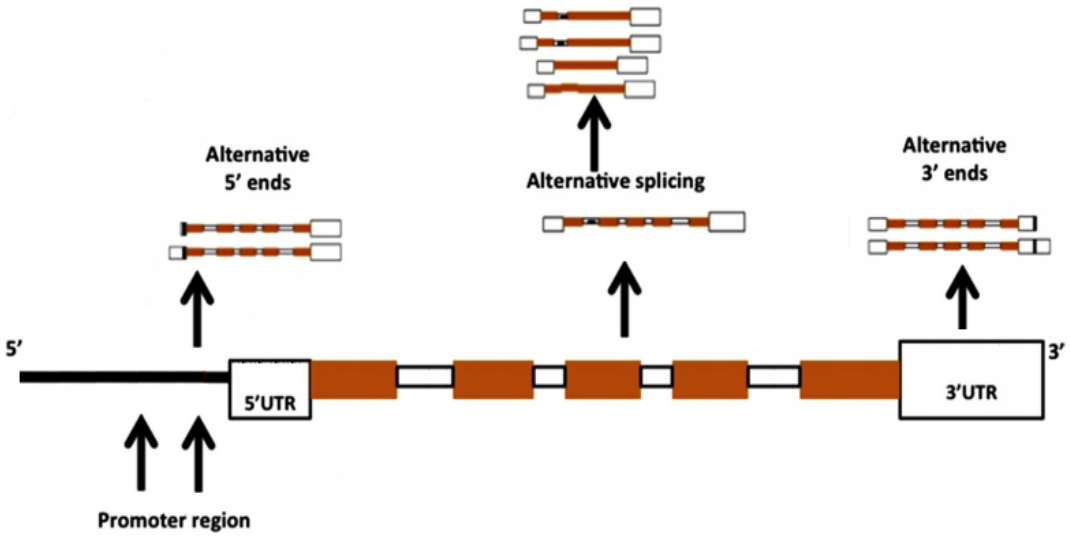

Figure 1- Schematic representation of synthesis of different transcripts from a single gene due to changes in promoter recognition, alternative splicing and alternative polyadenylation. Shorter or longer 5'UTRs results from different promoter regions while alternative splicing leads to transcript with different introns and therefore longer or shorter fragments. Finally, in the 3'UTR the alternative polyadenylation sites leads to variance on the size of transcripts. Adapted from Hegde and Crowley, 2019[3]

1.1.2. Polyadenylation and Alternative polyadenylation

Polyadenylation signalling is one of the RNA maturation process. When the polyadenylation site is recognized, a complex of proteins is recruited and cleave the pre-mRNA and add a number of adenines to the 3'UTR[19]. Alternative polyadenylation of 3'UTR of mRNA is an important regulatory mechanism of gene expression and protein localization[20,21] where different 3'UTR sizes result from different polyadenylation sites which maintain or reduce the presence of regulatory elements in the mRNA 3' UTR[19].

It has been shown that when cells are in active proliferation, shorter 3'UTRs are more prevalent, while in differentiated cells the mRNAs have their transcripts with longer 3'UTR, and this enables the cell to optimize the regulation of protein expression. The proliferation stage demands the expression of high levels of proteins, which favours the expression of transcripts subject to lower degree of posttranscription regulation[22].

Several hypothesis were raised about how the polyadenylation site is chosen. One was that the proximal sites are preferred over the distal ones due to the temporal advantage during transcription, which was previously demonstrated by Denome and Cole in 1998[23], where a 3'UTR construct with 2 similar polyadenylation sites at different distances between proximal and the distal and when this difference increased the preference of the proximal site increased as well, showing the temporal advantage of the first one[24]. Additional evidences showed that the elongation rates affect the polyadenylation site usage where slower elongation rates increased proximal site usage due to more temporal advantage over the distal sites. Indeed, in human and mouse transcriptome, more abundant genes have smaller 3'UTR than their less expressed counterparts[25]. In addition, the presence of several factors like cis elements near the polyadenylation site or the interaction of cleavage factors in the cell can contribute for the recognition of the proximal or distal sites depending on the cellular conditions. The polyadenylation signal (PAS) located before the cleavage site by approximately 40 nucleotides is an hexamer of AAUAAA nucleotides[26]. Additional evidences showed that derivations of this hexamer are also recognized and used as PAS but with a lower binding affinity and therefore being less recognized[24]. The availability of the cleavage sites is also taken in consideration in genes with multiple PASs, the higher the affinity for a site recognition, the lower nucleosome occupancy will occur in association with that specific site and

a nucleosome enrichment downstream[27] leading the polyadenylation complex to cleave the RNA on the site with high affinity instead of the others sites due to less occupancy.

The process of polyadenylation is carried out by a complex of 4 subunits, composed by cleavage and polyadenylation specificity factor (CPSF), cleavage factors I_m and II_m (CFI_m and $CFII_m$), a poly(A)polymerase (PAP) and a cleavage stimulation factor (CstF)[20,28,29]. Recognition and cleavage requires 3 cis elements present in the 3' UTR mRNA, the AAUAAA hexamer, the GU-Rich element and the UGUA element[28,29]. The first element is recognized by the large subunit of CPSF (160kDa) then the subunit CPSF73 binds the cleavage site and is responsible for the cut of the 3'UTR. For the recognition and cleavage of a certain polyadenylation site is also necessary a U/GU-rich element where the CstF complex binds located 15-30bp downstream of the hexamer, the last element the UGUA is usually 40-100 bp upstream of the cleavage site and is bound by CFI_m.

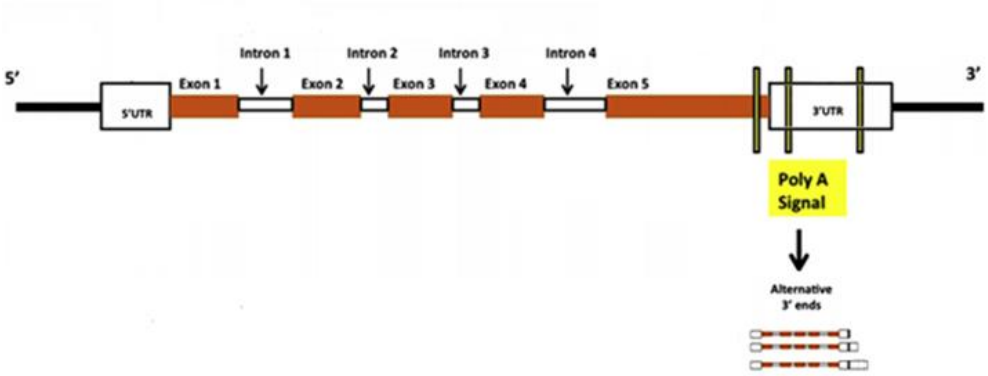

Figure 2- Structure of mRNA and possible localization of alternative polyadenylation sites. Alternative PAS sitest originate alternative transcripts with distinct 3'UTRs. Adapted from Hegde and Crowley, 2019[3]

1.1.3. AU/GU-rich elements

AU-rich elements (ARE) are sequences responsible for the interaction of proteins with RNA (ARE-RBPs). There are several classes of AREs (I to V) [30], and they are determined by the number of repeats of the minimal sequence AUUUA considered necessary to destabilize mRNA[31]

At the beginning, AREs were associated with mRNA decay and translation repression[30], but nowadays it is known that AREs can also induce protein expression. For example in mouse, deletion of AREs from tumour necrosis factor (TNF)-α mRNA lead to inflammatory arthritis and

bowel disease, because usually in unstimulated hematopoietic cells, TNF-α AREs are responsible for mRNA degradation, but in the presence of lipopolysaccharides they stabilize mRNA and increase translation at an early stage, becoming repressive after that[32].

The GU elements are repeats of the consensus sequence "UGUUUGUUUGU" and was first identified in the 3'UTRs of transcripts expressed in primary human T-cells that originate from hematopoietic stem cells (HSCs) in the bone marrow [33,34,35]. When present in the 3'UTR of the transcripts, this consensus sequence leads to destabilization of those transcripts upon ligation of CUG-BP, Elav-like family (CELF) 1, a member of CELF family known to bind RNA and work as post-transcriptional regulator of mRNA decay, deadenylation translation and pre-mRNA processing[35]. The GU- rich elements are also involved in the process of alternative polyadenylation cleavage through the interaction with CstF, where the presence of GU-rich elements downstream of the cleavage site are recognized by the 64KDa subunit and affect the efficiency of recognition of the polyadenylation site[36]. The CstF complex also recognizes GU-rich sequences, but in this case the affinity is stronger to sequences with at least two consecutives Uracil nucleotides rather than with the consensus sequence.[36]

1.1.4. Roquin constitutive decay element

Cells require rapid modifications in expression of transcription factors, signalling compounds or immune response modulators[37] in rapid response to stimuli such as environmental changes. One way of rapidly change the expression of genes is through posttranslational mechanisms. These posttranscriptional modifications can be achieved by the binding of proteins to the RNA (RBPs). These proteins usually bind to an element present in the 3'UTR region of the transcripts. There are several RBP binding sites ARE-independent. One example of this is the Constitutive Decay element which is part of mechanisms that mediate rapid RNA degradation[38]. When a given conformation and sequence is present some proteins can bind to RNA and are able to alter its stability. Ring finger and CCCH-type domain proteins also known as ROQUIN are an example of a RBP that is capable of binding to specific RNAs that contain a loop with a given sequence and structure present. These proteins are scaffold proteins that help the ligation of a deadenylase complex responsible for rapid RNA degradation.[37,39–41] As an example, there are evidences that

Roquin and Roquin2 can bind to Bone Morphogenetic protein receptor 1a (*Bmpr1a*) 3'UTR due to a stem-loop present on this region.[37] When Roquin binds to the specific loop it works as a protein adaptor of a complex composed by C-C Motif Chemokine Receptor 4 (Ccr4), Chromatin assembly factor-1 (CAF-1) and Negative Regulator Of Transcription Subunit (NOT) which is called Ccr4-Caf1-Not deadenylase complex or the DEAD-box (from the amino acid motif D–E–A–D (Asp–Glu–Ala–Asp)) protein. As an example, DEAD-Box Helicase 6 (DDX6/RCK) protein associates with Decapping MRNA 2 (DCP2) that is known to decap and mark the mRNA for degradation by other posttranscriptional effectors[42].The NOT protein works as a chaperone that contributes to the ligation of CAF1 or CCR4-NOT Transcription Complex 8 (CNOT8), two mRNA deadenylases which will start removing adenines from the 3'end of target mRNAs. The NOT protein will also recruit and act as scaffold to the ligation of the DDX6/RCK that is a decapping activator and repressor of translation[41] (Fig. 3). The interaction of Roquin with the Ccr4-Not complex occurs through its C-terminal region that is conserved only among vertebrates while the DDX6/RCK interacts with roquin N-terminal region.[41] This mechanism of regulation by the Ccr4-Not complex recruited by roquin only functions when a 5'Cap and 3' poly-A-tail is present but has redundant functions[42]. Roquin-1 is a member of E3 ubiquitin ligase family encoded by the gene *Rc3hl* and plays a role in gene expression. Although being a member of E3 ubiquitin ligase family and containing a RING (from Really Interesting New Gene) finger domain neither the human nor the mouse Roquin-1 have described function on the ubiquitination pathway nor it has been shown to have a function as E3 ubiquitin ligase. In addition, Roquin is associated with posttranscriptional modifications acting as an RNA binding protein. The ROQ domain which can bind to the mRNA constitutive decay element (CDE) stem-loop has been also associated with the binding and ligation of mRNAs to stress granules[39] This constitutive decay loop recognized by roquin is present in the 3´UTR of 50 different genes including *Bmpr1a* that were bioinformatically described by Leppek (2013). Depletion of Roquin in an extract of a mouse embryonic fibroblast (MEF) cell line lead to a shift in Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor delta (Nfkbid) mRNA from monosomal to polysomal fractions of sucrose gradients, showing that roquin is also associated with posttranslational regulation of gene expression[40]. There are also evidences that Roquin regulates the expression of some receptors as Inducible T Cell Costimulator (Icos) and Cytotoxic T-Lymphocyte Associated Protein 4 (CTLA-4), some cytokines as tumor necrosis factor (TNF) and interleukin (IL)-64,7, some enzymes as Itchy E3 Ubiquitin Protein Ligase (ITCH) and

Tumor necrosis factor, alpha-induced protein 3 (A20) as well as some transcription factors as Interferon Regulatory Factor 4 (Irf4) and proto-oncogene NF-KB Subunit (c-Rel) that are part of central signals pathways in most cells. It is also described that mRNAs that contain the Roquin CDE loop have increased length compared with mRNAs that are nontargets. This can in part be explained by the numbers of binding sites present in the 3'UTRs of those specific mRNAs[42]

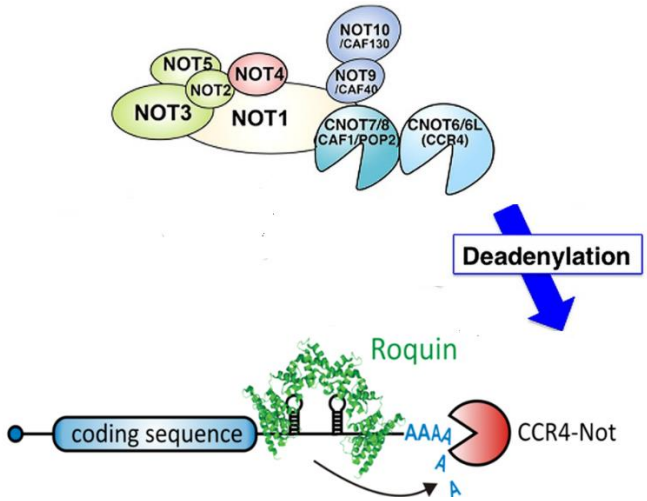

Figure 3-Mechanism of degradation by the CDE on mRNA. The ligation of Roquin which acts as a scaffold protein to the ligation of Ccr4-Caf1-Not complex marking that specific mRNA to degradation through deadenylation. Adapted from Inada and Makino [43] and Dr. Julia Weigand (https://weigand-lab.de/?page_id=503)

1.1.5. Micro-RNAs as post transcription regulators

Micro-RNAs are small sequences of 20-24 nucleotides produced from longer transcripts following a complex multi-step process of maturation. The expression of micro-RNAs is regulated by similar mechanisms as others RNAs. The genes encoding micro-RNA are often located in intergenic regions and contain their own promoters. Alternatively, micro-RNAs are encoded by introns of host genes and in this case their expression is regulated by the promoter of the host gene[44].

There are several mechanisms of repression mediated by micro-RNAs[44] acting as post transcriptional regulators. One example is the binding to a specific sequence usually located in the 3´UTR of target mRNAs and then either recruiting the RNA induced silencing complex (RISC) promoting the degradation of that specific mRNA or blocking the translation of that mRNA into protein[45,46] (Fig 4). By blocking translation of repressor proteins, micro-RNAs can also have

stimulatory effects on expression. Another example of their action is miR-373 that can increase expression of E-cadherin and Cold Shock Domain Containing C2 (CSDC2). There are some evidences that microRNAs can also work as tumour suppressors or oncogenes[39] having a great impact on cancer development[47,48,49].

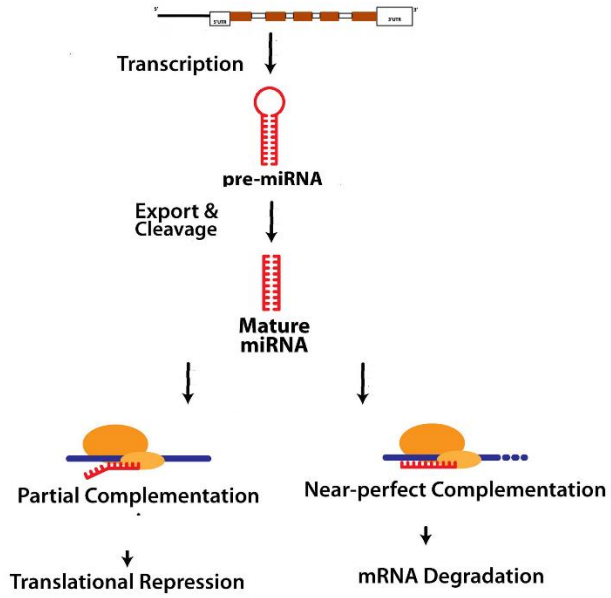

Figure 4- Synthesis of microRNA through its process of maturation and mechanism of post transcription regulation. Partial complementation of microRNA to the target RNA sequence leads to a decrease in the translational process while a perfect or near-perfect complementation leads to mRNA degradation. Adapted from Williams et al., 2015[50]

1.1.6. Mechanisms for regulation of transcripts translation

Translation initiation requires a complex of proteins that bind to mRNA in a cap-dependent mode. The cap binding Eukaryotic initiation factors complex (eIF4F complex) is composed of eIF4E which is able to bind to mRNA cap, eIF4G which is a scaffold protein that allows the ligation of initiation factors, and the eIF4A that removes the secondary structures of RNA because of its ATP-dependent RNA helicase activity. When the complex IF4F is linked to mRNA, the 40S subunit of ribosome, associated with additional initiation factors and the Met-tRNA, is recruited and starts to scan the mRNA from 5´to 3´ in search of an AUG codon. Once identified, the large subunit of ribosome, the 60S complex, bind to the 40S forming the complete 80S complex. It is

known that the AUGs located in the first ~10 nucleotides are usually overpassed and the ribosome complex continues its scan for the correct AUG, which is possible with the help from the RNA sequence and structure. When a purine(A/G) is in the position -3 (considering the A from AUG as the position+1) and a G is in position+4, known as kozak's sequence, GCC A/GCCAUGG, the ribosome complex changes his structure, stops scanning and starts translation.

1.1.7. Upstream Open Reading Frames (uORFs)

Upstream ORFs (uORF) are small sequences in the 5'UTR that contain an AUG in frame with a stop condon upstream the main AUG and a minimal length of 9 nucleotides[51]. It is estimated that over 40% of mammalian mRNAs contain at least one uORF on their 5' region[51,52]. They are one of the mechanisms for the regulation of translation[52] and some studies show that uORFs can regulate this process by activating the nonsense-mediated decay (NMD) in case of scanning stalling or by reducing the translation initiation efficiency if the ribosome dissociates from the mRNA[8,53]. The impact of uORFs on transcript translation efficiency depends from different conditions, like distance between the 5'cap and the start codon of the uORF, the strength of their AUG signature, number of uORFs, respective length and distance to the main ORF[52,53]. In stress conditions, certain genes need to increase their translation efficiency. One of the strategies promoted by stress response is to decrease the abundance of eIF2 initiator methionine tRNA ·(GTPMet-tRNAMet) ternary complex, leading to increase rate of translation reinitialization at the main ORF[53,54].

The activation of translational repression by a uORF requires that the start codon on this sequence must be recognized, at least at certain times, by the scanning 40S ribosomal subunit[7]. After recognition, the ribosome translates the uORF and then stall near the stop codon which inhibits binding and scanning of the transcripts by the additional ribosomal complexes available for translating the mRNA. After the recognition the ribosome complex is able to reinitiate or disassemble from the mRNA resulting in reinitialization or repression, respectively[5,51] (Fig 5).

The uORFs are classified as strong or weak depending on the Kozak sequence where a strong uORF contains a purine(A/G) on position -3 and a guanine on position +4 (GCC A/GCCAUGG). If these rules are not accomplished this uORF is considered weak [54]. Another aspect to look to is the position of the uORF relatively to the 5'cap where closer uORFs to the 5'cap are less likely to be recognized by the ribosome[5,54].

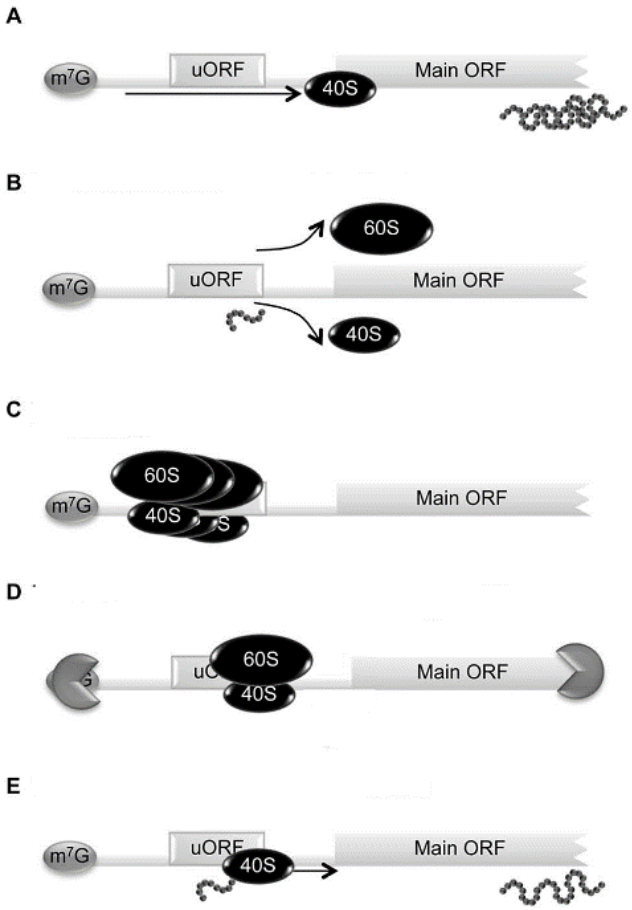

Figure 5-Mechanism of translational regulation through uORFs. A) The small subunit of ribosome (40S) skips the uORF and the translation of the main ORF occurs and therefore the protein is expressed. B) The 40S subunit recognizes the uORF and it is translated followed by dissociation of the ribosome through the recognition of the stop codon. C) The ribosome recognizes the uORF but it stalls on the AUG and the translation does not occur. D) The premature stop codon is recognized and the NMD is activated, and that specific mRNA is marked for degradation. E) The ribosome recognizes and translate the uORF but the 40S subunit stays and continues the scanning of mRNA and with enough space between the uORF and the main ORF the translation reinitiates and both the uORF and mainORF are translated. Adapted from Romão et al., 2013[54].

1.2. Skeletal development

The skeletal development starts with a condensation of mesenchymal stem cells (MSC) of mesoderm or neural crest and leads to mammalian embryonic skeleton patterning formation. After the skeleton patterning the new bone formation can be divided in two processes that occur during development, the endochondral and the intra-membranous ossification[55,56].

The majority of bones are formed by endochondral ossification where mesenchymal cells condensate and differentiate into chondrocytes which form a cartilage template that will be

replaced by bone during development[57]. Flat bones like cranial bones, hip bone, sternum, ribs, and scapulae are formed by intra-membranous ossification that is characterized by direct differentiation of mesenchymal cells to osteoblasts[58].

In endochondral ossification, after the cartilage template is formed, it suffers an invasion of different cell types as osteoblasts progenitors, osteoclasts, blood vessel endothelial cells and hematopoietic cells to that region[59]. The osteoblast progenitors near the hypertrophic cartilage, that is being resorbed, differentiate into trabecular bone-forming osteoblasts while hematopoietic and endothelial cells give risen to bone marrow. Osteoblast progenitors near the perichondrium differentiate into osteoblasts and deposit cortical bone[60]. In both of these processes bone formation through osteoblasts is identical and it starts with the construction of a type 1 collagen matrix which is responsible for the strength and elasticity of the bone and also works as scaffold for the deposition of other matrix compounds[61].(Fig 6).

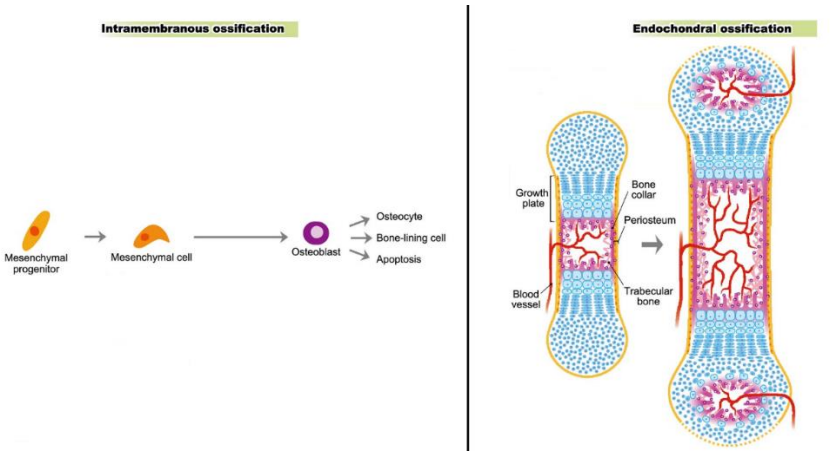

Figure 6- Cellular mechanism of intramembranous ossification. In the left mesenchymal progenitors differentiate into mesenchymal cells and through several processes to osteoblasts; at this point osteoblasts differentiate into either osteocytes or bone-lining cells, if differentiation is not achieved the cells enter in apoptosis. On the right endochondral ossification process where vascularization of the bone occurs during bone development.

The bone development is orchestrated by several molecules and pathways involved in bone formation as Wingless-related integration site (Wnt), Notch, sonic and Indian Hedgehog (shh and ihh respectively), transforming growth factor-β (TGF-β) and Bone morphogenetic proteins (BMP). Besides bone formation and development, TGF-β and Bmp's are also involved in bone metabolism.

1.2.1. Composition and metabolism of bone tissue

Bone tissue is an important constituent of vertebrate skeleton, with multiple functions as structural support, movement, protection of internal organs, reservoir for several molecules as growth factors and cytokines, and minerals like calcium, phosphorus, magnesium, sodium and potassium[62]. The bone tissue can be classified in two different types, cortical and trabecular. The shafts of long bones and the compact matrix in flat bones are constituted by cortical bone (Fig. 6) and is a dense tissue composed by osteocytes, rich in collagens and noncollagenous proteins as osteocalcin (Ocn). This tissue is highly vascularized, connecting their cellular processes [63,]. Trabecular bone, which can be found at the end of long bones, between flat bones compact matrix and close to joint surface (subchondral bone), is composed by thin trabecular plates that compose a network (Fig. 7). Both of these tissues have similar molecular and cellular composition but different functions and mechanical properties[64,65].

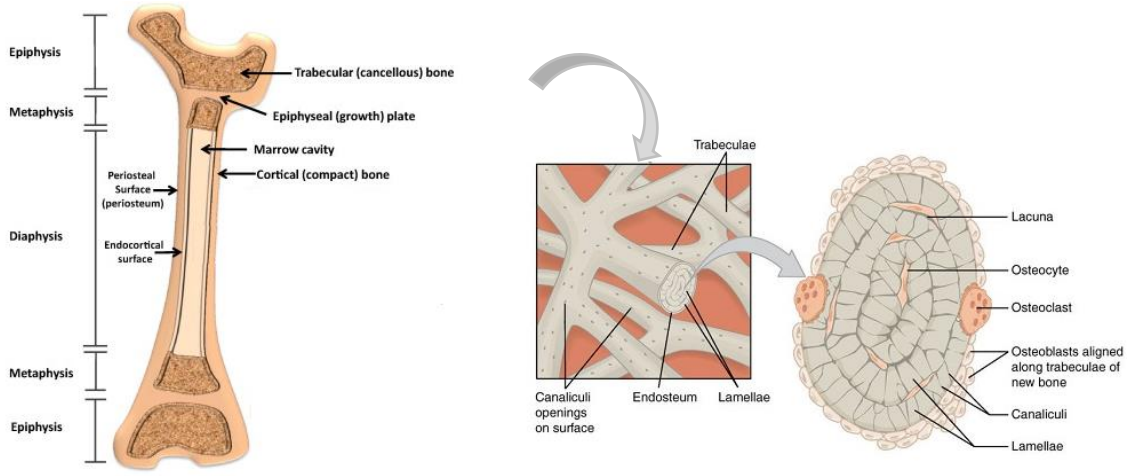

Figure 7- Cellular composition of trabecular bone. Trabecular bone is located on the epiphysis of the long bones and is composed of a network of plates called trabeculae with a layer of osteoblasts.

At the cellular level, bone is composed by two major cell types, osteoblasts, and osteoclasts. Osteoblasts differentiate from mesenchymal stem cells and are involved in bone formation processes like the deposition of mineral extracellular matrix and communicate with osteoclast precursors to regulate bone resorption rate[66]. Through bone formation osteoblast can go through a specialization and differentiation into either osteocytes or bone-lining cells. When osteoblasts start

to become isolated within the bone matrix and are subject to certain levels of mechanotransduction forces, they differentiate into osteocytes[67,68]. Osteoblasts that never progress to the inside of bone matrix will mature and differentiate into bone-lining cells that neither differentiate into osteocytes nor undergo apoptosis[69] and can be found on the surface of bone[70,71]. These cells are responsible for the removal of any bone matrix leftovers and formation of new layers of collagen deposition [71].

Osteoclasts are cells that derive from hemopoietic cell progenitors and are responsible for bone resorption[72]. They are recruited when bone formation cells start dying and are able to dissolve bone mineral matrix and recruit new osteoblast precursors to replace damage bone tissue[7333,69]. The osteoclast activity is also responsible for the cleavage of TGF-β precursor molecule which is composed by a mature TGF-β and a non-covalently bound Latency-associated protein.[74] Osteoclast activity leads to cleavage of this precursor molecule which releases TGF-β1 inducing the accumulation of osteoprogenitors in that region[74]. However, TGF-β1 alone is unable to induce osteoblastogenesis, but BMP2 (discussed below) has been demonstrated in some cell lines to induce osteoblast differentiation [55,75,76].

Balance between bone resorption made by osteoclasts and bone formation by osteoblasts is important for the maintenance of bone homeostasis and to maintain its strength and integrity (Fig 8). Unbalance in this mechanism can lead to several diseases during life as osteoporosis, osteopetrosis, Paget's disease of bone and renal osteodystrophy[73]. There are some mediators in bone metabolism such as prostaglandins that help in the differentiation of osteoblasts[77], interleukins that are capable of inducing the expression of Receptor Activator of Nuclear factor-κB ligand (RANKL)[78], chemokines, leukotrienes, growth factors, Wnt signalling and bone morphogenetic proteins (BMPs)[61].

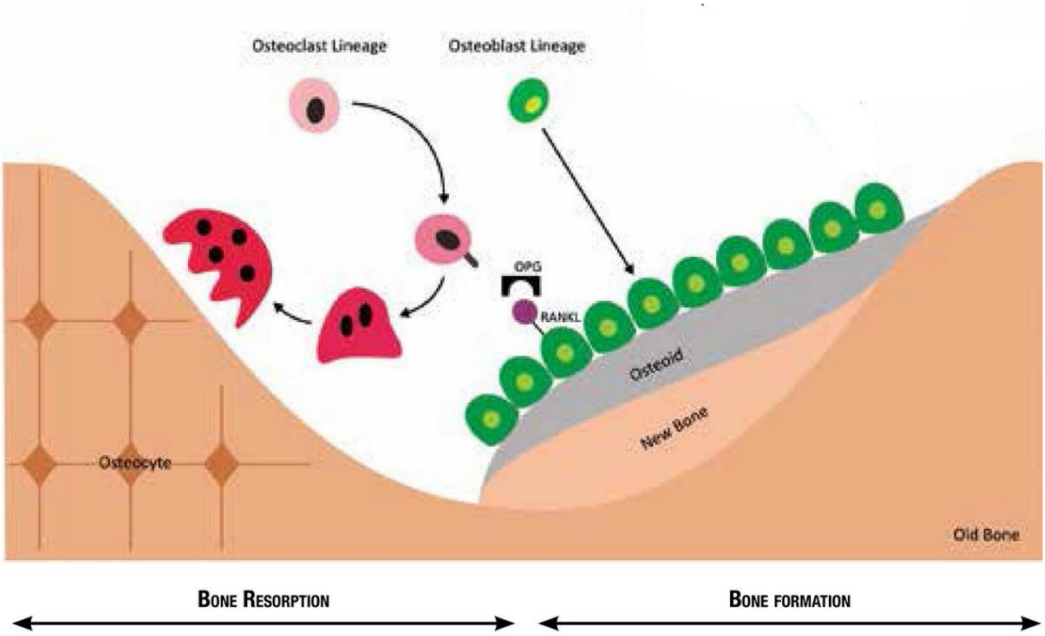

Figure 8- Cellular composition of bones. In Red osteoclasts that are responsible for bone resorption by degrading the extracellular matrix by other and in green osteoblasts that are responsible for new bone formation by depositing osteoid. Recruitment of osteoclasts occur due the levels of RANKL/OPG in circulation. Adapted from Francisco J. A. de Paula et al., 2014[79]

1.2.2. Bone morphogenetic proteins

Bone morphogenetic proteins (BMPs) were first discovered in 1965 but only characterized and cloned in 1980's[80,81]. They are a subgroup of transforming growth factor beta (TGF-β) family proteins [82] with over 30 different members and are involved in cell differentiation, proliferation and apoptosis[46]. It is known that not all BMPs have similar effect on bone metabolism, for example BMP3 is an inhibitor of bone formation by blocking SMA (Smad pathway) [83,84] while BMP 2, 4, 5, 6,7 and 9 are able to induce osteogenesis [83,85] by transducing their pathway through a family of proteins similar to the gene products of the *Drosophila* gene 'mothers against decapentaplegic' (Mad) and the *C. elegans* gene Sma (Smad), in this case, through SMAD 2/3 instead of transducing by the SMAD1/5/8[55,86] . There are studies that show that BMP2 is necessary and sufficient to induce bone formation, there are even medical treatments that use BMP2 as a osteogenic agent in treatment for various illnesses of bone as fractures[83]. Another example of a BMP acting in bone formation is BMP7 which induces the expression of osteoblast differentiation markers[55].

BMPs work as morphogens and in dose-dependent manners. A morphogen is a molecule that is synthesized in a location of the organism and diffuses to the surroundings inducing a regulatory

effect in a dose-dependent manner[87]. The BMPs ligands can form homo or heterodimers in the endoplasmic reticulum and in the Golgi complex, which are processed proteolytically by Furin and/or Subtilisin/like Proprotein Convertases (SPCs) and the full mature BMP dimers are released to the outside of the cell. This homo- or hetero-dimerization will affect the efficiency of signalling being done.[88] The BMP's signalling pathways is transduced through Bmp receptors type I and II (Fig 9).

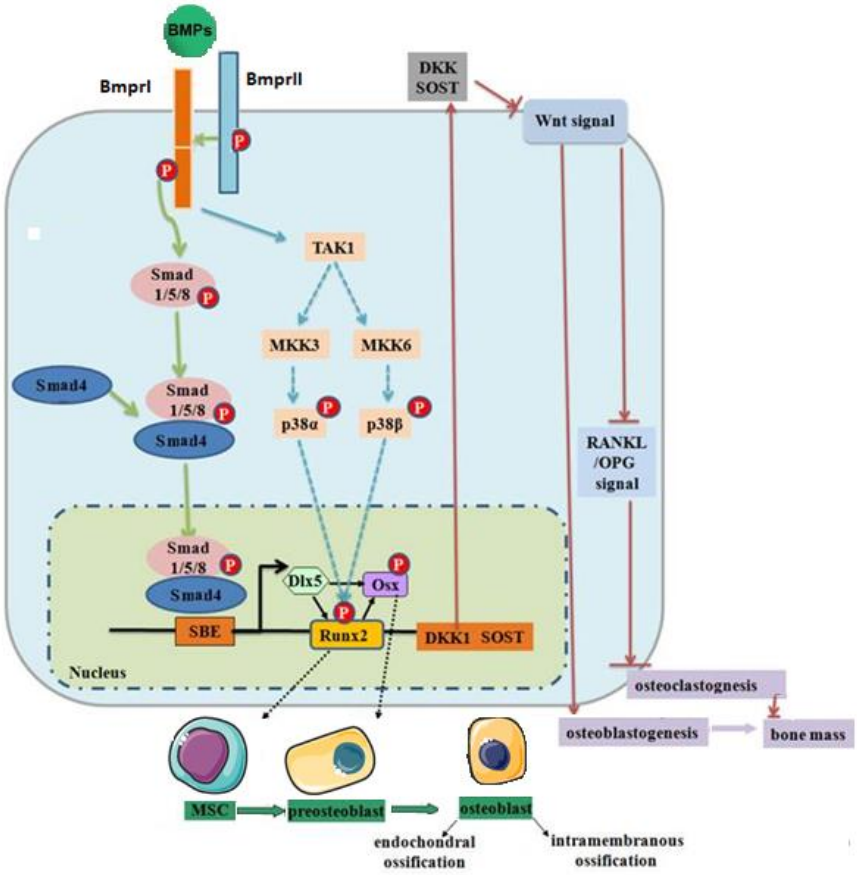

Figure 9- Regulation of expression of target genes of BMP signalling pathways. After the ligation of BMPs to it receptors the BMPR2 phosphorylates the BMPR1, and the signal is passed inside the cell through the Smad canonical pathway or through alternative pathways. One of the target gene is Runx2 which is expressed in the osteoblast's precursors, other genes that are a target for the BMP pathways are the Dickkopf (Dkk) and Sclerostin (Sost) which are known as inhibitors of the Wnt signal pathway involved in osteoclastogenesis through the RANKL/OPG signalling. Adapted from Yi-Ping Li et al., 2012[89]

1.2.3. Bone morphogenetic protein receptors (BMPRs)

Bone morphogenetic proteins interact with seven receptors type I, BMPR-IA, Activin Receptor-Like Kinase (ALK) 3, BMPR-IB, ALK-1, ALK-2, ALK-4, ALK-5, and ALK-7 [90]. There are also four type II receptors in which three of them interact with BMPs BMPRII, Nuclear receptor coactivator ACTR (ActR) IIA and ActRIIB [81]. The signalling pathway from the cell cytoplasm to the nucleus is made by a cascade of phosphorylation reactions[91,92,81]. The ligands specificity is determined primarily by the type I receptor[93] but the ligand affinity is determined through the combination of those receptors[94,91,92,81]. The pathway that is activated depends also on the type of complex formed, where BMP receptor activation occurs upon ligand binding leading to activation of the Smad pathway and the BMP-induced signalling complexes lead to the activation of Smad-independent pathway[95]. The activity of the BMPRs is also affected by co-receptors that regulate the interaction of the BMPs to the receptors. One of those examples of co-receptors are the Repulsive Guidance Molecules (RGMs). The RGM family is composed of three proteins (RGMa, RGMb, and RGMc) which contain a glycosylphosphatidylinositol (GPI)-anchor that link to the plasma membrane. The RGMs are responsible for enhancing the functions of BMP2, BMP4[96] and BMP12 as indicated in one study in which reduced expression of a single RGM protein reduces the BMP12 activity[97].

1.2.4. Bone morphogenetic protein receptor 1a

Bone morphogenetic protein receptor 1a (BMPR1A) is a type I serine/threonine kinase that is known to regulate bone and cartilage differentiation[82,98] and adipocyte formation[46,99,100]. It is also associated to the osteoblasts/osteoclasts communication, as shown in one study in which deletion of BMPR1A from mature osteoclasts lead to increase bone resorption while deletion of BMPR1A in osteoclasts of adult mice lead to increased bone formation[101]. In another study performed by Kamiya et al.2008, knockdown of *Bmpr1a* in osteoblasts lead also to increased bone mass. BMPR1A is also involved in the differentiation of adipocytes, when the BMPR1A forms a truncated protein or *bmpr1a* mRNA is blocked occurs the inhibition of differentiation of the mesenchymal progenitors into adipocytes.

Bmpr1a gene, which encodes the protein BMPR1A, is located on chromosome 14 in mouse and in chromosome 10 in humans. The mouse *Bmpr1a* has multiple isoform variants that result in two distinct proteins, one with 532 amino acids and with a molecular weight of 60kDa and a second smaller transcript with 500aa. Heterodimerization of type I with type II receptors is necessary to form a functional membrane receptor, with a kinase domain. Upon activation the phosphorylated type II receptor phosphorylates the short GS domain (from characteristic sequence of SGSGSG) in type I receptor and this phosphorylation activates the kinase activity on the type I receptor. Then a cascade of signal transduction is activated and the signal passes through different pathways as the canonical Smad pathway or the noncanonical Mitogen Activated Protein Kinase (MAPK) or the Phosphoinositide 3-kinase/ Protein kinase B (PI3K/Akt) pathway[55,82,98] depending on the ligand.

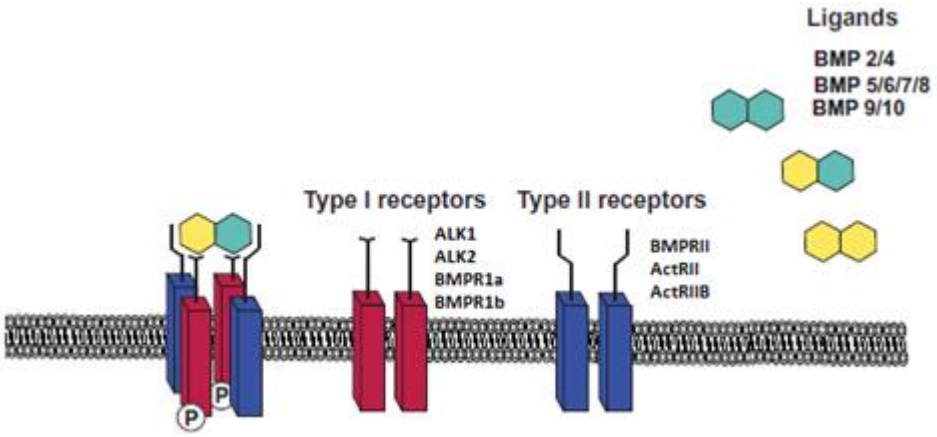

Figure 10- BMP receptors and their ligands. There are 7 receptors type I but only four interact with BMPs there is also 4 type II receptors but only 3 interact with BMPs[92]. Those receptors need to homo and heterodimerize in order to be functional. After ligand binding the type II receptors phosphorylate the type I. Adapted from Marie-Christine Ramel and Caroline S. Hill, 2012[92].

1.2.5. Bone morphogenetic proteins Signal Transduction

1.2.5.1. Smad Dependent Pathway

Until 2008 there were 8 different SMAD proteins identified in humans, divided in three sub classes: R-Smads and C-Smads, previously mentioned, and the inhibitory Smads (I-Smads). The members of the TGFβ family transduce their signal through the Smad pathway which is known to be involved in different processes[102]. BMP transduce their signal through a Heterodimer of BMPR

type I and Type II; the type II, which is always active, phosphorylates the type I and this last phosphorylates the R-Smads, that in turn phosphorylate the C-Smad, which moves to the nucleus and activates the transcription of the target genes[103] (Fig 11).

The Smad pathway is known to work as transcription regulator of the expression of Runt-related transcription factor 2 (RUNX2), that is a transcription factor osteoblast-specific[103] that has been demonstrated to be essential for osteoblast differentiation. It has also been shown that the Smad-4 protein interacts with Wnt pathway, which is known to induce bone formation[104]. Other members of TGF-β are responsible for inducing the transcription of I-Smads, which will act as a mechanism of negative feedback for this pathway. The smad pathway is also involved in the crosstalk between BMPs and MAPK through the phosphorylation of Smad-1.

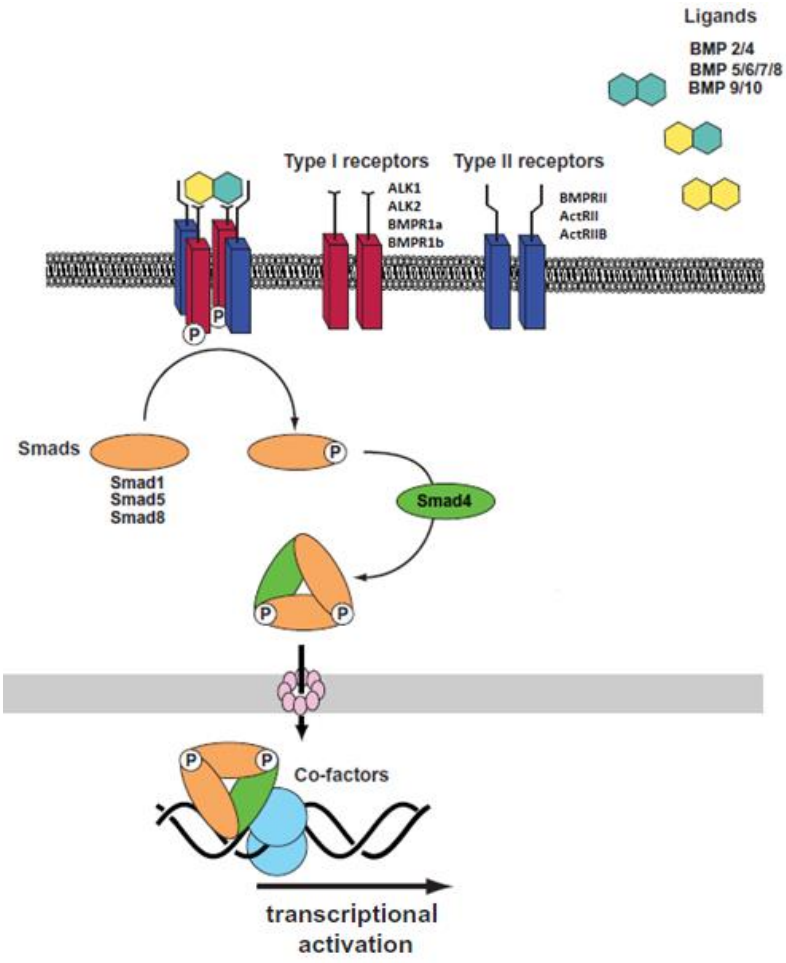

Figure 11- Smad pathway leading to the expression of BMPs target genes. After activation of BMPR phosphorylation of Smads1/5/8 occurs those Smads bind to Smad 4 which forms a complex that is able to accumulate in the nucleus of the cell and act as expression regulator. Adapted from Ramel and Hill, 2012[92].

1.2.5.2. Smad Independent Pathways

Besides the Smad pathway, BMPR signal is also transduced by other pathways as MAPK, PI3K/Akt[55,82,98] or Wnt[105,106].

On bone it has been proved that Wnt through it downstream signal molecule β-catenin stimulates the differentiation of mesenchymal stem cells into more mature osteoblast and blocking their differentiation into adipocytes or chondrocytes[61]. This was further proved through activation of the canonical Wnt signalling lead to expression of *osterix* (OSX) wich promotes osteoblast differentiation[107]. *In vivo* study on osteoblasts, the BMP induction pathway downregulates the Wnt signalling during postnatal and embryonic development and it is explained by the fact that Wnt inhibitors, *Sost* and *Dkk1*, are targets of BMP signalling[106,108]. Both the pathways, Smad-Dependent and Smad- independent, are known to contribute to the *Dkk1* expression where only Smad dependent signalling is required to expression of *Sost* showing a differential regulation by BMP signalling through BMPR1A[106]

The PI3K/Akt signal pathway is other pathway involved in bone formation and bone metabolism, which inactivation of this signal pathway led to a decrease in the expression of RUNX2 which is an essential transcription factor to the osteoblast differentiation. Tian, Q et al. 2005, proposed that interaction between the PI3K/Akt signal pathway and the Wnt pathway occurs by the activation of β-catenin through the Akt pathway. The Phosphatase and tensin homolog (PTEN), which is an inhibitor of the PI3K/Akt signal pathway, upon inactivation of PTEN the levels of β-catenin increase. BMPR1A blocks the phosphorylation of PTEN and therefore increasing the levels of the PI3K/AKT.

The mechanistic target of rapamycin (mTOR) pathway is a downstream target of the PI3K/AKT pathway[109,110,111]. mTOR is a serine/threonine kinase that can be found in two different complexes (mTORC1 and mTORC2) and control a variety of cellular processes. This pathway has been associated with osteoblasts differentiation where in some cases mTOR exerts a stimulatory effect while in other cases it has been linked to an inhibitory effect[111]. In one study conducted by Lim et al. 2017, deletion of Bmpr1a lead to a decrease in the mTOR signalling pathway and therefore leading to an increase in trabecular bone mass and a reduced periosteal bone growth[111]

Another non-canonical pathway involved in bone formation and bone maintenance is the mitogen-activated protein kinase (MAPK) cascade, as extracellular signal-regulated kinase 1/2 (ERK1/2), p38, and c-Jun NH2-terminal kinase 1/2 (JNK1/2)[112] (Fig 12).

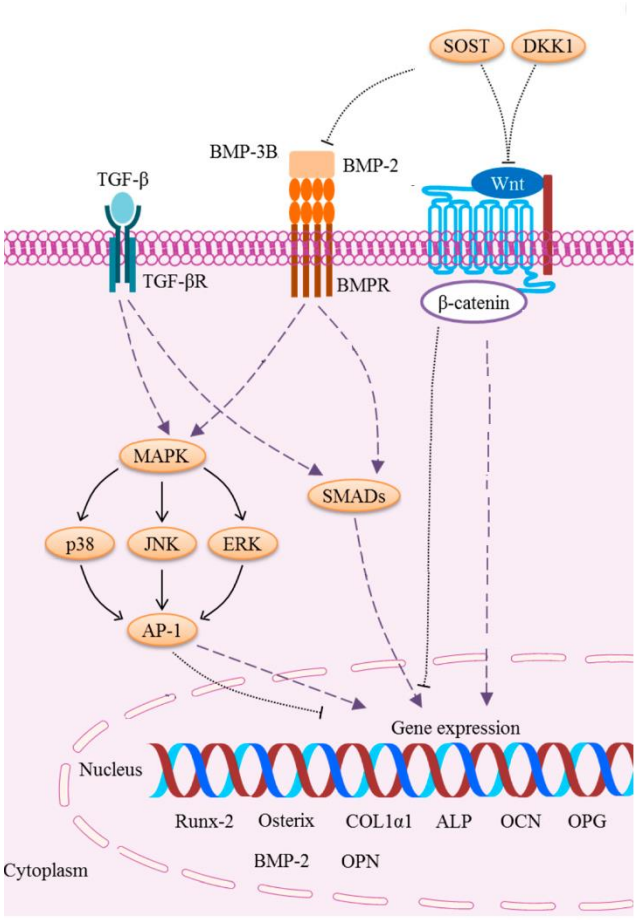

Figure 12- Smad independent pathway of BMPs. The full lines represent the inhibition of bone formation while the doted lines represent the activity of bone formation. After BMP binding, the activation of the MAPK cascade occurs followed by the activation of ERK/JNK. Adapted from Ima-Nirwana et al., 2019[113]

1.2.5.3. BMPRIA in bone formation and resorption

Throughout human life, bone is in constant formation by osteoblasts, and in constant resorption by osteoclasts. Differentiation of MSC into mature osteoblasts promote bone formation and during this process several genes are express in different stages of differentiation of MSC. In one study deletion of *Bmpr1a* led to a decrease in alkaline phosphatase (ALP) staining and a decrease in the mRNA levels of *Alp*, *Osx/SP7*, *Runx2*, and *Ocn*[114]. Those proteins are necessary for bone formation and are expressed in different stages of osteoblast differentiation.

RUNX2 is a member of a family of transcription factors that are responsible for several roles during normal development and neoplasias[103]. From this family RUNX2 is the only transcription regulator involved in the early osteoblast differentiation and bone formation.[103,115] but

inhibits late osteoblast differentiation and plays also crucial roles in cartilage formation and chondrocyte maturation[115]. Nishio et al. 2006, showed that in the Runx2-null mice OSX/SP7 was absent during osteoblast differentiation confirming that OSX/SP7 is a downstream target of Runx2. Similarly, to RUNX2, OSX/SP7 is also a transcription factor involved in skeletal development and osteoblast differentiation (Fig 13).

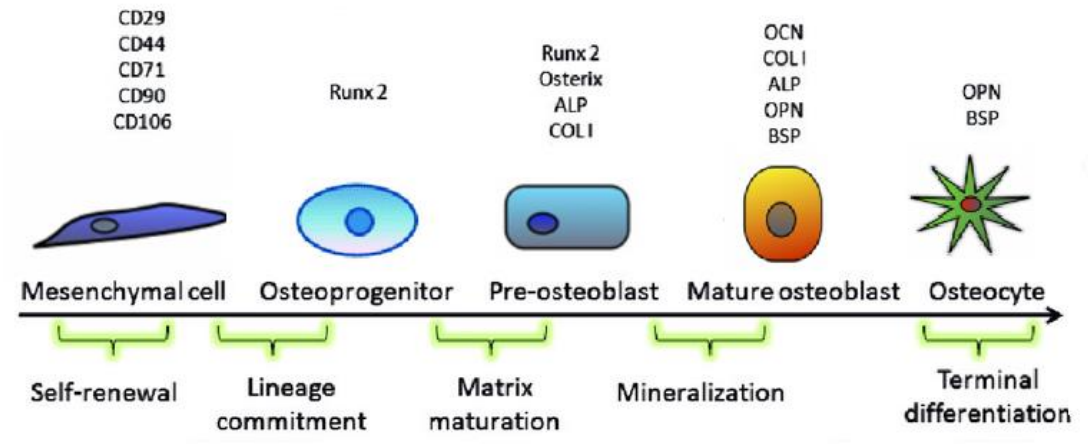

Figure 13- Osteoblast differentiation process with all the marker genes expressed in different time points during that process. Osteoblasts originate from mesenchymal stem cells and terminally differentiate to osteocytes. Adapted from Zhang et al., 2017116

The deletion of *bmpr1a* in osteoblasts also leads to a decrease in resorption activity of osteoclasts due to the decrease of RANKL/OPG ratio[117,114,91].

RANKL belongs to the tumour necrosis factor (TNF) family and acts as a signalling molecule that can transduce their signal in both directions producing intracellular reverse signaling[118]. RANKL is expressed by several bone-associated cell types but disruption of RANKL in an early stage of development leads to complete abolish of osteoclasts.

During bone remodelling osteocytes produce RANKL that will act as membrane-bound factor leading to osteoclastogenesis and inhibits osteoblastogenesis[117].

Mature osteoclasts secrete small extracellular vesicles that contain RANK which is a transmembrane protein that when presented to an osteoblast cell type will activate the expression of *Osx*/*Sp7* and *Runx2*[118] that are crucial osteoblast differentiation factors.

Deletion of Bmpr1a in bone marrow mesenchymal stromal cells leads to a decrease in expression of RANKL and therefore to a decrease in bone resorption. This occurs due to the response of

hematopoietic precursors that respond to the production of RANKL by the osteoblasts and stromal cells leading to the differentiation of those precursors into osteoclasts[114].

2 Methodology

2.1. Bioinformatic

We started by doing an *in silico* research of some regulatory elements present in the *Bmpr1a*, to work as guideline for us and facilitate the search for those regulatory elements experimentally. We searched for regulatory elements both in the 5' and 3' UTR of the Bmpra1a mRNA; on the 5' we searched for uORFs and for Iron responsive elements and on the 3'UTR we searched for different possible polyadenylation sites, AU-rich and GU-rich elements.

2.1.1. Prediction of alternative polyadenylation sites

The identification of alternative poly-A sites was done through the search of *Bmpr1a* gene sequence (ENSMUSG00000021796). The search started by looking for the canonical polyadenylations sequences, AAUAAA and AUUAAA, which were bioinformatically described, and recognized 70% and 15% of the times respectivelly[119]. After canonical sequences identification, , we decided to search for polyadenylation sites using an online bioinformatic tool RegRNA2.0[120] (http://regrna2.mbc.nctu.edu.tw/) where the default parameters were kept and only selected the polyadenylation sites.

2.1.2. Identification of AU-rich and GU-rich elements

Several bioinformatics databases and tools were used to predict AU-rich elements as ARED2[121] (http://rc.kfshrc.edu.sa/ared), RNAAnalyzer [122](http://rnaanalyzer.bioapps.biozentrum.uni-wuerzburg.de/) and RegRNA2.0[120] (http://regrna2.mbc.nctu.edu.tw/) where the tool default parameters were kept . The minimal sequence ("AUUUA", "UUAUUUAUU") and different clusters of AU-rich elements (I-V)[123] were manually searched on *Bmpr1a* transcripts

2.1.3. Identification of uORFs

All the 5'regions of Bmpr1a transcripts were inserted on CLC sequence Viewer 8.0 and performed an ORF search with the default parameters (Start codon "AUG", standard genetic code, stop codon included in annotation) maintained except the length of ORF detection was changed a to minimal 9 nucleotides long.

2.1.4. Phylogeny

To study the phylogenetic relations between the species we started by collecting sequences from different species and preform a separate multiple alignment of 3'UTR and 5'UTR from *Bmpr1a* using the software MAFFT (v.7.428) with the alignment algorithm L-INS-i [124]. The alignment was then submitted to the jmodeltest2[125,126] software to determine the best model of nucleotide substitution for those sequences (GTR+I+G). To evaluate the phylogenetic relations between chosen species we constructed a phylogenetic tree through two independent statistical approaches, Bayesian Inference (BI) and Maximum likelihood (ML). Using the software MrBayes (v. 3.2.7)[127] we obtain the best phylogenetic tree by Bayesian inference. The posterior probability of each ramification was obtained after250000 generations. A coinciding phylogenetic tree was obtained with maximum likelihood method [128] with software RaxMLGui (v2.0) [129] and confidence levels of ramification was obtained through 500 replicates applying bootstrap method[130]

2.1.5. Statistics

Statistical analysis represented on figures 26, 27,28 and 29 were performed using an ordinary one-way anova followed by Tukey post hoc test with p< 0 05 for statistical significance. For these experiments was assumed normal distribution for data obtained. The values are represented as mean ± SD. All the statistical analysis were performed using GraphPad Prism version 8.0.1 (GraphPad Software, La Jolla USA)

2.2. Experimental work

In our experimental work we started to differentiate the pre-osteoblast cell line into osteoblasts to identify alternative transcripts as consequence of alternative polyadenylation in 3'-UTR of *Bmpr1a* transcripts in a context of osteoblast differentiation. To Identify the presence of regulatory elements in the transcripts expressed by MC3T3-E1 cells we amplified by PCR cDNA fragments obtained from the mRNAs from terminal differentiated MC3T3-E1 cells.

2.2.1. MC3T3-E1 cell culture and differentiation

In order to differentiate MC3T3-E1 pre-osteoblasts cells into osteoblast the cells were initially maintained in α-MEM (Minimal Essential Medium) supplemented with 10% Fetal bovine Serum (FBS), 1% Pen/Strep (10000 units/mL of penicillin and 10000 µg/mL of streptomycin. Gibco) and 0.2% Amphotericin (250µg/mL. Gibco). The medium was changed every 3/4 days and cells incubated at 37º C in 5% CO_2.humidified chamber. To differentiate MC3T3-E1 into osteoblasts, cells were seeded in 6-wells plates with a density of $2.5x10^4$ cells per well. When the cells reached approximately 80% confluence, differentiation towards osteoblast stage was induced by treating the cells with mineralization induction medium (50 µg/ml Ascorbic acid (Sigma), 10mM β-glycerophosphate (Sigma) and 4mM of $CaCl_2$ (Sigma). The mineralization medium was also changed every 3/4 days and for 23 days. After this period, 3 wells subject to mineralization treatments, and 3 wells corresponding to control condition, were stained with alizarin red. An additional 6 well plate, subject to the same conditions, control, and mineralization triplicates wells, was used to isolate total RNA.

2.2.2. Alizarin Red Staining

To evaluate if the cells in wells subject to mineralization induction treatments differentiate into osteoblasts, the cells were stained with Alizarin Red, which allows the detection of calcium deposition by the cells as part of hydroxyapatite deposition by osteoblasts. After 23 days the medium of each of the 6-well plate was discarded and the cells were washed 3 times with phosphate

buffered saline (PBS) (137mM of NaCl (Sigma), 2,7mM of KCl(MERCK), 8,1mM of Na_2HPO_4(Sigma) and KH_2PO_4 (MERCK) adjust to pH7,4). After that, 400 µL of 4% formaldehyde (Sigma) solution was added in each well and incubated for 30 minutes at room temperature. The fixative was discarded, and the cells were washed 3 times with ddH_2O. To stain the cells, 1mL of 40mM alizarin red (Sigma™) stain, buffered with ammonium hydroxide (MERK) to pH 4.2, was added to each well and incubated for 15minutes with agitation. The cells were then rinsed in distilled water.

2.2.3. RNA extraction and purification

The protocol for total RNA extraction was adapted from nzytech™ protocol. Briefly, cells were washed in cold PBS once and 400µL of NZYOL (nzytech™) was added to each well to promote cell lysis, following by a scrapping process to detach the cells. After scrapping the cell solution was pipetted several times to ensure that most of the cells were removed from the plate and to homogenize and after that transferred to a 1,5mL microtube and incubated 5minutes at room temperature. For phase separation 210 µL of chloroform (Sigma™) was added to each sample and vigorously shaken for about 30 seconds and left to incubate 5minutes at room temperature. Samples were centrifuged at 12000x g for 15minutes at 4°C and then the upper aqueous phase was transferred to a new 1,5mL microtube with care not to disturb the layer between the aqueous phase and the organic phase when pipetting. RNA was precipitated by adding 500µL of isopropyl alcohol (Scharlau) incubated at -20°C for 20 min and centrifuged at 12000x g for 10min at 4°C. The supernatant was discarded. The pellet was resuspended in 1mL of ethanol 75% (Emsure) and then centrifuged at 12000x g for 5 minutes at 4°C, the supernatant was discarded, and the step repeated one more time, after this the pellet was air dried and the RNA was dissolved in 50 µL of RNAse and DNAse free water from Sigma™. And if it was observed that the pellet was not completely dissolved, more 50µL were added each time until no pellet was observed. The total quantity of RNA was determined using Thermo Scientific™ NanoDrop One™ and the RNA integrity was evaluated by running 3 µL of each sample in an agarose Gel, the remain RNA was stored at -80ºC for further studies.

2.2.4. DNAse treatment and Reverse transcriptase Polymerase Chain Reaction (RT-PCR)

To ensure that the samples only contained RNA, we performed a DNAse treatment before the RT-PCR. From each sample, 1µg of RNA was used and diluted in 1µL 10x DNAse buffer (Promega™), 1 µL of RQ1 DNAse (Promega™, 1U/ µL) and up to 10 µL of DNAse and RNAse free water from Sigma. The samples were incubated at 37ºC for 30 minutes, and after that 1 µL of RQ1 DNAse stop buffer (Promega™) was added and the solution incubated for 10 minutes at 65ºC to stop the DNAse activity.

After DNAse treatment, the samples were used to synthetize the first strand of DNA (cDNA) as follows: for each 1µg of RNA treated with DNAse, 1 µL of 10mM dNTPs mix and 1 µL (100µM) of dT adapter (check primers table 1) was added. The samples were then incubated at 65ºC for 5minutes and rapidly transfered to ice. 4 µL of 5x First strand buffer (Invitrogen), 2 µL of 0.1M DTT, 1 µL of RiboLock (Thermo scientific™, 40 U/µL) and 1 µL of M-MLV (200U/ µL) reverse transcriptase enzyme were added to the samples to start the reaction and incubated at 37ºC for 60 minutes. The reaction was stopped by incubating at 70ºC for 15 minutes and samples were stored at -20ºC.

2.2.5. Polymerase Chain Reaction

Each PCR reaction was prepared from between 50ng to 200ng of cDNA, 3mM of $MgCl_2$, 10x reaction buffer, 0.5mM of Primers, 0.5 mM dNTP's and 0.5µL of NZYTaq II (5 U/µL) polymerase (NzyTech™) used for the amplification of DNA for 35 cycles on a DNA thermal cycler (2720 thermal Cycler Applied biosystems™).

Table 1- Description of primers for amplification of the different fragments from different regions of Bmpr1a with the sequence, the number of base pairs

Primer name	Gene	Region	Orientation	Sequence	Size
3utr1	Bmpr1a	3'UTR	Fw	ATAATCCAGCCTCCAGACTC	20bp
3utr2	Bmpr1a	3'UTR	Fw	GAAGTTGACATACCCTTGAATACC	24bp
3utr3	Bmpr1a	3'UTR	Fw	CCAGTGACCCATCCTATGAG	20bp
3utr4	Bmpr1a	3'UTR	Fw	GGCCAATCGTGTCTAACCGCTG	22bp
3utr5	Bmpr1a	3'UTR	Rv	TGGGCTGTCTTGTATAGAACTGAGTG	26bp
3utr6	Bmpr1a	3'UTR	Rv	CAGAGGACAATGATGGGGTAGTTG	24bp
3utr7	Bmpr1a	3'UTR	Rv	GAGACTCACTAAACAAACCAGGCAAG	26bp
3utr8	Bmpr1a	3'UTR	Rv	CTCATTAAGGTGGGTTGGGCATCTAC	26bp
P1	Bmpr1a	5'UTR	Fw	CGCGAGACGACGACTGTACG	20bp
P2	Bmpr1a	5'UTR	Fw	CTGAGGCGGCAGAGATTGGAA	21bp
P3	Bmpr1a	5'UTR	Fw	CCGGAGGATGAGTTTCTCGGGAT	23bp
P4	Bmpr1a	5'UTR	Rv	TGTCTGATTCGCACGCGTCCTG	22bp
P5	Bmpr1a	5'UTR	Rv	GAACAGACAGGCTCCCAGTAATCT	24bp
Universal Adapter		3'UTR	Rv	ACGCGTCGACCTCGAGATCGATGT	24bp
dT Adapter		3'UTR	Rv	ACGCGTCGACCTCGAGATCGATGT TTTTTTTTTTTT	36bp

2.2.6. DNA fragment extraction from agarose gel

The DNA fragments of interest amplified by PCR were separated by electrophoresis in an agarose gel and extracted and purified using GeneJet Gel Extraction kit (Termo Scientific™). Briefly, DNA fragments in agarose were excised, and their weight was measured and stored in a 1,5ml microtube. For the purification and extraction, the GeneJet Gel Extraction kit (Termo Scientific™) was used where a volume of binding buffer equal to the mass of gel excised was added to the microtube and incubated for 10minutes at 55ºC. After incubation, not more than 800µL of sample was added to the column provided by the kit and centrifuged at top speed (13000rpm Eppendorf 5415D) for 1 minute, the flow through was discarded and 700µL of wash

buffer was added and centrifuged for 1min and the flow through discarded. The column was centrifuged again to remove the remain wash buffer which was transferred to a 1,5ml tube, between 30 to 50 µL of elution buffer was added and the tube centrifuged at top speed (13000rpm Eppendorf 5415D) for 1 minute, the flow through was then stored at -20ºC if not used right after.

2.2.7. DNA fragments cloning

Each different fragment obtained by PCR was inserted on TOPO TA PCRII (Thermo Fisher™) vector (Annexes Fig- 31) using the TOPO™ TA Cloning™ Kit. 4.5µL of DNA fragment of interest previously amplified by PCR were added to 0.5µL of TOPO vector and after 1µL of Salt solution was added in order to close the vector with the fragment insert; the mix was then incubated for 30 minutes at room temperature.

The transformation of DH5α competent bacteria started by incubating in ice for 20 minutes 100 µL suspension of competent bacteria with 4 µL of the ligation solution prepared before and then incubated exactly 45 seconds at 42ºC for heat shock treatment and placed in ice for 5 minutes. After, 300µL of S.O.C. medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose) was added to the bacteria. After the heat shock, each transformation sample was incubated for 1 hour in an orbital incubator and during that time, at 30 minutes, 40µL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), a substrate for the β-galactosidase, was spread on the agar plates and placed on the incubator, and after 50 minutes 5µL of Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to the bacteria in the incubator in order to express the genes controlled by *lac* operon in the bacteria. After incubation, 300µL of the bacteria solution was spread in the agar plated and incubated overnight at 37ºC. On the next day, the white colonies present on the agar plate were picked using a wood tip and placed on 2mL of liquid agar and incubated overnight on the orbital incubator (BIORAD) with 280 rpm at 37ºC to propagate the clones previously selected.

From each tube with bacteria ,1,5mL was transferred to a microtube and centrifuged 1 minute at top speed (13000rpm Eppendorf 5415D), the supernatant discarded and 100µL of P_1 solution (50 mM Tris (Sigma™) and 10 mM Ethylenediamine tetraacetic acid (EDTA, pH 8.0), (PanReac AppliChem)), was added and the pellet resuspended on vortex. Next, 100µL of P_2 solution (0.1 M NaOH (Merk) and 1% Sodium dodecyl sulphate (SDS) were added, and the samples agitated by

inversion several times and incubated 5 minutes at room temperature. After incubation, 100 µL of P_3 solution (1.5 M Potassium Acetate, pH 5.5) was added to the samples, agitated by inversion and placed on ice for 10 minutes. The samples were centrifuged during 3 minutes at top speed (13000rpm Eppendorf 5415D) at room temperature, the supernatant was transferred to a sterile microtube and 2 times (800µL) ethanol 100% added, the solution was centrifuged for 5 minutes, the supernatant was completely removed, and the pellet was washed with 70% ethanol and left to dry for 5 to 10 minutes. The DNA was dissolved with 30-40µL of DNAse and RNAse free water from Sigma and stored at -20ºC.

To confirm that the insertion worked, a restriction analysis was performed. 2µL of buffer H(10x) and 0.5 µL of Eco RI enzyme was added to 10 µL of sample and the final volume was adjusted to 20µL with H_2O DNAse and RNAse free water from Sigma™ The solution was incubated at 37ºC for 1h. The fragments were submitted to an electrophoresis in an 1.5 % agarose gel for separation of plasmid from cloned fragment, and if the correct size of the fragment was present, 5µL of sample not digested was sent to be sequenced.

Each different fragment previously obtained by PCR with the restriction enzyme cut site (table II) amplified from TOPO plasmid (See Annexes Fig. 31) was inserted on pGL3 control vector (See Annexes Fig. 32). Both the fragments and the vector were cut with the same restriction enzymes (NCoI and HindIII) for 2h at 37ºC. The ligation of the fragments to pGL3 occurred overnight with T4 ligase (Promega™) at 4ºC, where 100ng of pGL3 control were mixed with 15-20ng of the fragment, depending on fragment size in order to maintain 1:3 proportion, 0,5 µL of T4 Ligase (Promega™), 1 µL of T4 Buffer and with RNAse DNAse free water from Sigma up to 10 µL of total mix.

The transformation and pGL3 cloning were also done as described above, and the plasmid was extracted using ZR Plasmid Miniprep Classic Kit ™, where 1.5mL from each tube with bacteria was transferred to a microtube and centrifuged for 30 s at 13000rpm, (Eppendorf Mini Spin Plus 5453) the liquid LB medium was removed and the pellet was centrifuged one more time to remove the remaining medium. Then 200µL of P1 buffer was added and pipetted up and down to resuspend the pellet, next 200 µL of P2 was added and the samples mixed by inverting the tubes 2 to 4 times and incubated at room temperature until the solution appear purple. Then 400µL of P3 was added and the sample was mixed by inversion thoroughly but gently until it turned yellow and then incubated 2 minutes at room temperature. The samples were then centrifuged for 4minutes at

13000rpm (Eppendorf Mini Spin Plus 5453), the supernatant was collected and transferred to the column provided and centrifuged for 45seconds, the flow through was discarded, 200 µL of Endo-wash buffer added to the column and centrifuged for 45 seconds at top speed (13000rpm Eppendorf Mini Spin Plus 5453). 400 µL of Plasmid-wash buffer were added and centrifuged for 1minute at top speed (13000rpm Eppendorf Mini Spin Plus 5453), the flow through was discarded and the column transferred to a clean 1,5mL microtube. Next, 30µL of DNA Elution Buffer was added to the column and incubated for 1minute at room temperature. The samples were then centrifuged 45seconds at 13000rpm (Eppendorf Mini Spin Plus 5453) and stored at -20ºC if not used right away. To confirm that the insertion occurred, an enzymatic digestion in a 20µL reaction, with 2µL of buffer K (10x), 2µL of BSA (1%, Thermo fisher™), 0.5 µL of NCoI and HindIII (4–12 units/µL) enzymes added to each sample (5 µL) and adjusted to final reaction volume with H_2O DNAse and RNAse free water from Sigma™. The solution was incubated at 37ºC for 2h. The evaluation of the restriction reaction was done by agarose (1.5%, Sigma) electrophoresis and UV exposition in a transilluminator (Cleaver Scientific™).

2.2.8. Sequencing

The positive fragments with insertions of interest on TOPO-TA PCRII (Thermo Fisher™) and the positive fragments with insertions of interest on pGL3 were sent for sequencing in the CCMAR sequencing department.

The analysis of sequencing results were done with Blast VecScreen (https://www.ncbi.nlm.nih.gov/tools/vecscreen/) with default parameters to remove the plasmid sequence and a nBlast (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) with default parameters was performed to confirm that the sequences were the ones that we isolated.

2.2.9. HEK293 Cell culture

The HEK293 cell line was used to perform all the transformations for functional essays, more precisely the uORFs activity through the dual luciferase essay. Cells were maintained in Dulbecco's

Modified Eagle Medium (DMEM) supplemented with 10% FBS, 1% Pen/Strep (10000 units/mL of penicillin and 10000 µg/mL of streptomycin. Gibco) and 0,2% Amphotericin (250µg/mL. Gibco), the medium was changed every 3/4 days and incubated at 37º C in 5% CO_2.humidified chamber.

2.2.10. Site Directed mutagenesis

After analysing the results from the luciferase essay, all the ATG's from the uORFs were mutated to AAG through site directed mutagenesis (SDM). To achieve the mutation several primers (Table II) were designed using the QuickChange primer design website (https://www.agilent.com/store/primerDesignProgram.jsp).

A PCR was performed using KAPPA Hi-Fi polymerase (1 U/µL) using the primers containing the desired mutation. The PCR product was then treated with methylation-dependent endonuclease *DpnI* (10 U/µl) to remove the parent template that does not contain the desired mutation and the endonuclease was inactivated at 80ºC for 20 minutes. After inactivation, the product was cloned into DH5α bacteria following the protocol described above. The positive colonies were selected and sequenced to confirm for the presence of the desired mutation, and samples were stored for further studies.

2.2.11. Transfection

When HEK293 cells reached approximately 80% confluence were briefly washed with PBS, treated with trypsin (0.05%) and plated on a 24-well plate with a cell density of $5x10^4$. On the next day cells were transiently transfected with 50µL of a mix that contained 250ng of the plasmid construct, 5ng of Renilla plasmid and 1µL of X-tremeGene™ HP DNA Transfection Reagent on DMEM medium. 48H after transfection the cells were washed with cold PBS and 100 µL of lysis buffer was added to each well. Cells were removed from each well using a scrapper and pipetted several times up and down to detach the maximum possible cells from the plate and transferred to a 1,5mL microtube. The samples were incubated at room temperature for 15minutes and centrifuged for 30 seconds at 14000 rpm (VWR Hitach CT15RE) at 4ºC. The supernatant was then transferred to a new 1,5mL microtube and used to the luciferase essay.

2.2.12. Luciferase assay

In order to measure the expression of luciferase gene present in the transfected plasmids a dual luciferase assay was performed using the Biotium Firefly & Renilla Luciferase Single Tube Assay Kit. To the luminescence-specific plate (Nucleon Delta™), 10µL of cell lysate was transferred to each well and 50µL of Firefly working solution (0,2mg/mL of D-luciferin) was added to each well and luminescence was recorded in a multiplate reader (Biotek Synergy 4). In same well 50µL of Renilla working solution (0,04mg/mL of Aquaphile™ coelenterazine) was also added and luminescence recorded to normalize the data and remove background noise.

Table 2- Description of primers for site-directed mutagenesis and plasmid insertion with the sequence, the number of base pairs

Primer name	Gene	Region	Orientation	Sequence	Size
ORF1 SDM	Bmpr1a	5'UTR	Fw	TGATACTGTCTTGGAATTCATGAGAAGGAA GCATAGGTC	39bp
ORF SDM	Bmpr1a	5'UTR	Rw	GACCTATGCTTCCTTCTCATGAATTCCAAGA CAGTATCA	39bp
ORF2 SDM	Bmpr1a	5'UTR	Fw	GGGATCCCGCAGATTTATGAAAATAAGCAT CGCTTTGAT	39bp
ORF SDM2	Bmpr1a	5'UTR	Rv	ATCAAAGCGATGCTTATTTTCATAAATCTGC GGGATCCC	39bp
ORF3 SDM	Bmpr1a	5'UTR	Fw	GAGCGCCGGAGGAAGAGTTTCTCGGGA	27bp
ORF SDM3	Bmpr1a	5'UTR	Rv	TCCCGAGAAACTCTTCCTCCGGCGCTC	27bp
5utr1 HindIII	Bmpr1a	5'UTR	Fw	CCCAAGCTTACGCGTCGACTTCGAGATCGA TGT	33bp
5utr3 HindIII	Bmpr1a	5'UTR	Fw	CCCAAGCTTCCGGAGGATGAGTTTCTCGGG AT	32bp
5utr6 HindIII	Bmpr1a	5'UTR	Fw	CCCAAGCTTTGCATCGCTTTGATACTGTCTT GGA	34bp
5utr4 SmaI	Bmpr1a	5'UTR	Rv	TCCCCCGGGTGTCTGATTCGCACGCGTCTTG	31bp
5utr5 SmaI	Bmpr1a	5'UTR	Rv	TCCCCCGGGGAACAGACAGGCTCCCAGTAA TCT	33bp

3 Results

3.1. Bioinformatics

3.1.1. Identification of mouse and human *BMPR1A* Transcripts

It was possible by mining several online databases to find 11 different mouse *Bmpr1a* transcripts where the transcripts Ensembl 202 (ENSMUST00000165280.8) and Ensembl 203 (ENSMUST00000171343.9) are incomplete. All the bioinformatic regulatory elements as the polyadenylation sites, CDE loop, minimal Au-rich element and micro-RNA binding sites were annotated on these transcripts (Fig. 14 and 15)

Mouse *bmpr1a* transcripts:

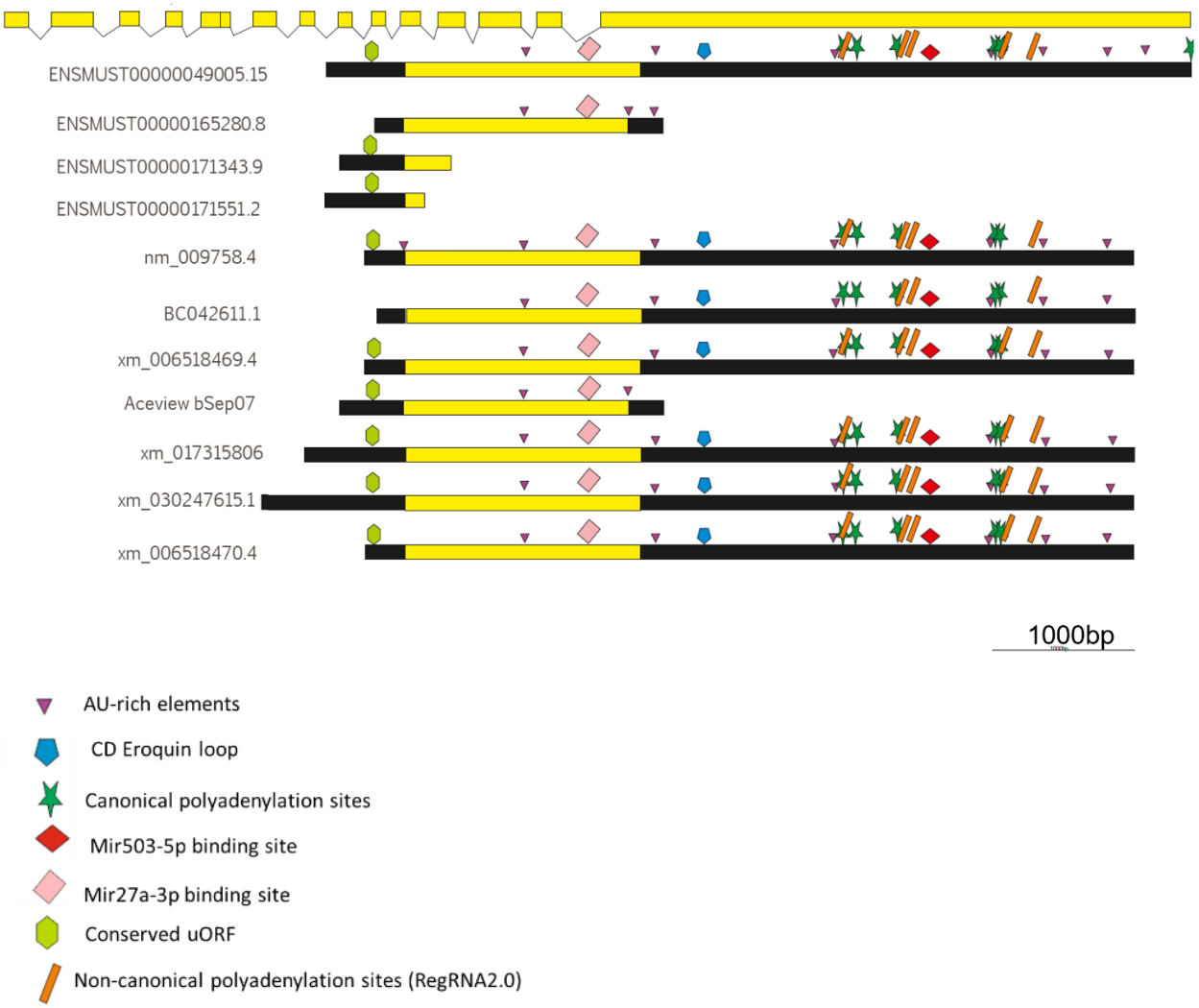

Figure 14- Identification of *Bmpr1a* mouse Transcripts. Scheme of online available mus musculus Bmpr1a transcripts with the representation of the different regulatory elements that were searched on this study. Represented in blue pentagon is the CDE loop sequence, the green hexagon represents the location of the uORF, the green star represents the canonical polyadenylation sites (AATAAA), the orange rectangle represents the polyadenylation sites obtained using the online bioinformatic tool RegRNA2.0, the red square represents the binding site of the microRNA 503-5p, the pink square represents the binding site of the microRNA 27a-3p and the purple triangle represents the minimal A/U sequence (AUUUA). On Top in yellow the gene structure

Human *BMPR1A* transcripts:

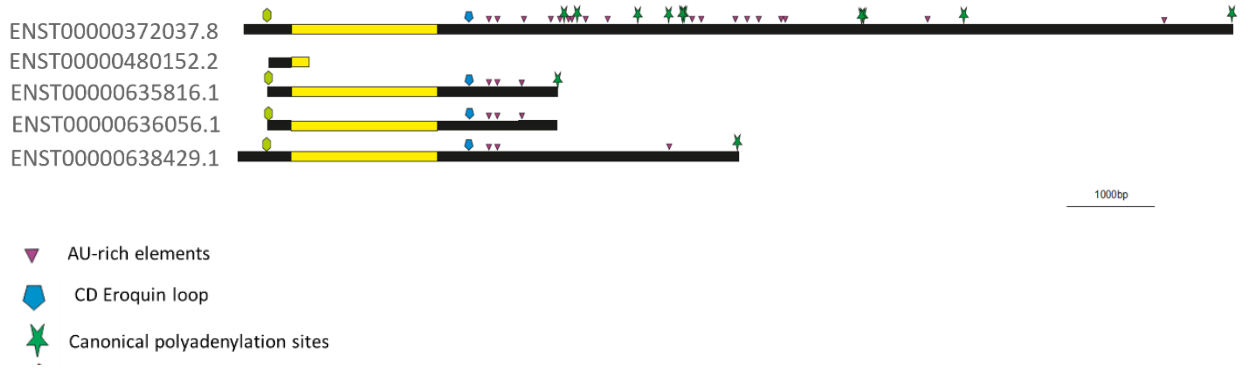

Figure 15-Identification of Human transcripts. Scheme of online available homo sapiens Bmpr1a transcripts with the representation of the different regulatory elements that were searched on this study. Represented in blue pentagon is the CDE loop sequence, the green hexagon represents the location of the uORF, the green star represents the canonical polyadenylation sites (AATAAA) and the purple triangle represents the minimal A/U sequence (AUUUA)

In order to analyse if regulatory elements were conserved among human and mouse sequences, we searched also for all elements that were present in annotated transcripts except for micro-RNA binding sites for which there was no conservations among the regulatory elements found in both species. All regulatory elements have the same colour scheme as used on the mouse.

3.1.2. Identification of upstreamORFs in *Mus musculus Bmpr1a*

Table 3-All the uORFs determined by the CLC sequence viewer. It is possible to observe the uORF selected from the transcripts due to conservation. Sequence of the 3 uORFs selected where the uORF1 is 66 bp long, the uORF2 is 12 bp long and the last one (uORF3) is 126 bp long

ENSMUSG00000021796 uORF ID	cDNA position	Type	uORF length	GC%	5' space	3' space	ICCu	ICCm
1.1	[274,339]	Type1	66	37.88%	273	208	AGGATGA	ACAATGA
6.1	[298,339]	Type1	42	33.33%	297	208	TTTATGA	ACAATGA
7.1	[306,431]	Type1	126	41.27%	305	116	AATATGC	ACAATGA
8.1	[336,431]	Type1	96	41.67%	335	116	TTCATGA	ACAATGA
9.1	[341,352]	Type1	12	41.67%	340	195	GAGATGG	ACAATGA
2.1	[115,201]	Type1	87	52.87%	114	324	AGGATGA	ACAATGA
3.1	[157,201]	Type1	45	53.33%	156	324	GAAATGC	ACAATGA
4.1	[161,241]	Type1	81	49.38%	160	284	TGCATGG	ACAATGA
5.1	[181,201]	Type1	21	42.86%	180	324	GGGATGT	ACAATGA

- uORF1:
- ATGGAAGCATAG
- uORF2:

ATGAGTTTCTCGGGATCCCGATTTATGAAAATATGCATCGCTTTGATACTGT
CTTGGAATTCATGA

- uORF3:

ATGCATCGCTTTGATACTGTCTTGGAATTCATGAGATGGAAGCATAGGTCA
AAGCTGTTCGGAGAAATTGGAACTACAGTTTTATCTAGCCACATCTCTGA
GAATTCTGAAGAAAGCAGCAGGTGA

For the human *BMPR1A* it was already described two uORFs on AceView (https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/), but none for the mouse transcripts. For the identifications of the uORFS we searched for AUGs in frame with a stop codon on CLC Sequence Viewer 8 (table 3). The uORFs with a length of 9 nucleotides were taken in consideration and annotated on the transcripts. All the transcripts, except Ensembl 202

(ENSMUST00000165280.8) which is a small and truncated transcript, contain the three conserved uORF, one with 12 nucleotides, another with 126 and the last with 66 nucleotides.

The cluster with the 3 different uORFs (Fig. 16) that we described above were aligned against several sequences of *Bmpr1a* from different species using MAFFT-G algorithm and it was showed that they are conserved in all mammalian species except marsupials. However, in the primate's species, we were able to find another small uORF that was not present in the mouse transcripts (Fig 16) and further studies about the possible function of this uORF are required. Also, in the bony fish species analysed, we found a different cluster of four uORFs (Fig. 16) that was not present also on the mouse transcripts. This cluster has different uORFs with different sizes from the ones found in mouse and human transcripts and inside the cluster but the structure of those uORF was similar between species. However this cluster can possibly also regulate the expression of *Bmpr1a* in those species but further studies are required.

Figure 16-Identification of possible uORFs in different species. The blue squares represent the conserved cluster among the mammals except marsupials. The red squares represent the possible cluster of uORFs in fish species. The green circles represent the small and conserved uORF on primates. This image was obtained using the ClC sequence viewer 8. The species analyzed were *Homo sapiens, Macaca mulatta, Gorilla gorilla, Mus musculus, Mus caroli, Gallus gallus, Odobenus rosmarus, Felis catus, Sarcophilus harrisii, Vombatus ursinus, Crocodylus porosus, Athene cunicularia, Zonotrichia albicollis, Salmo salar, Danio Rerio, Spaurus Aurata* and *Callorhinchus milii* respectively.

The strength of each uORF was then measured and the positions -3 and +4 were analysed. The smallest one with 12 nucleotides is considered a strong uORF, that contain a purine on position -3 and a guanine on position +4 (Fig. 17).

Figure 17- Alignment of the uORF to the sequences of different species. On this alignment is possible to observe high conservation of that specific region in A) in the red square the conservation of the uORF1, B) in the purple square the conservation of the uORF2 and C) in the yellow square the conservation of the uORF3. The sequences used for the alignment of the uORFs were Homo Sapiens, Cricetulus griseus, Ratus norvegicus, mus caroli, mus Pahari, macaca mulatta and mus musculus.

3.1.3. Identification of alternative polyadenylation sites

We found 6 canonical polyadenylation sites (AATAAA) in the Ensemble 201 *Bmpr1a* transcript (ENSMUST00000049005.15) and 5 different polyadenylation sites using an online bioinformatic tool RegRNA 2.0. Five of the 6 canonical polyadenylation sites were near or overlapped the ones found in RegRNA2.0 and the last one is at the end of the longer *Bmpr1a* transcript (ENSMUST00000049005.15). In the human transcripts it was possible to observe 8 canonical poly adenylations sites that were manually searched. (Fig. 15) The conservation of the annotated polyadenylation sites between human and mouse transcripts was observed after the alignment of both sequences using the software MAFFT with the algorithm L-INS-I, then the polyadenylation sites were manually searched, and it was only possibly to observe that the first one is conserved in human transcripts of *Bmpr1a* but none of the others are conserved among the two species (Fig.18).

Figure 18-Bmpr1a transcript (ENSMUST00000049005.15) scheme. The uORFs represented on the 5' and the polyadenylation sites on 3'UTR. In green are represented the canonical polyadenylation sites and in red are represented the polyadenylation sites obtained on RegRNA2.0. The purple arrow shows the conserved poly-A signal between the mouse (ENSMUST00000049005.15) and human transcript (ENST00000372037.8)

3.1.4. Identification of Constitutive Decay Elements (CDE)

Previously, Leppek et al. 2013[37] described a constitutive decay element in the 3'UTR of *mus musculus Bmpr1a* transcripts. To observe if CDE is conserved in *Bmpr1a* transcripts, 24 vertebrate species were aligned with MAFFT software. In 18 of 24 species the complete nucleotide sequence necessary for the binding of Roquin to RNA were conserved (Fig. 19B), however in the rest of those species, with the exception of *Antrostomus carolinensis* and *Echeneis naucrates*, the nucleotides necessary to the binding of ROQUIN are present with 53% and 66% homology for the

Drosophila sequences respectively. To confirm if three-dimensional RNA conformation was probable in order to be recognized by proteins Roquin and Roquin-2, the conserved sequences present in the transcripts were submitted to RNA Structure 6.1 software using the default parameters. The results obtained showed the formation of a stem loop for CDE sequence (Fig. 19A). with a Gibbs free energy of -2.2kcal/mol.

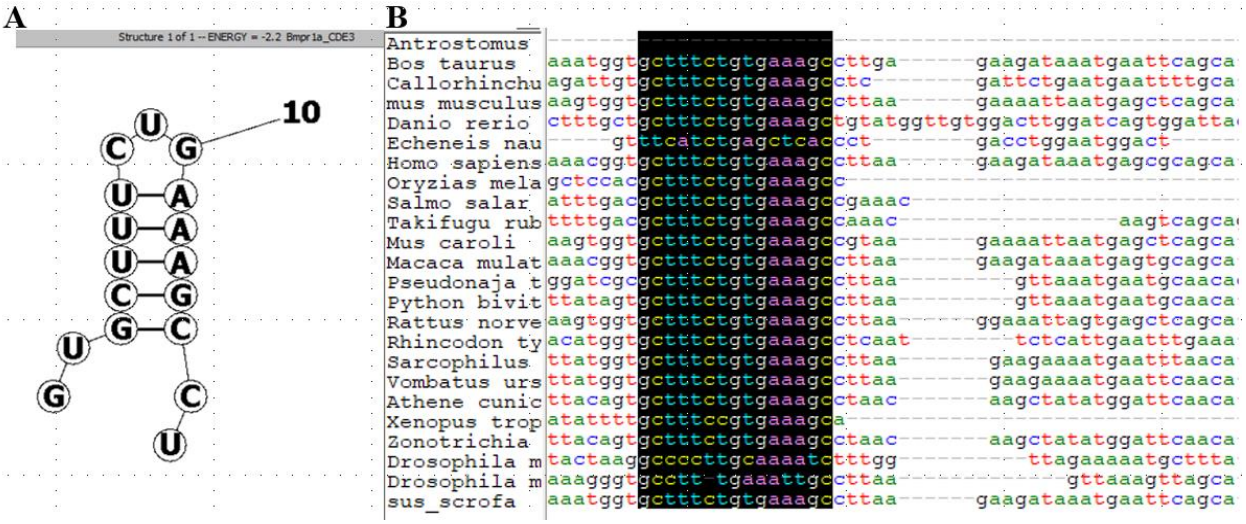

Figure 19- Sequence structure and conservation of the CDE loop present on 3'UTR of Bmpr1a. (A) -Nucleotide sequence and structure of Bmpr1a CDE loop. (B) Sequence alignment of Bmpr1a 3'UTR and drosophila melanogaster homologous gene (thickveins) where it is possible to observe the complete conservation of the CDE throughout the different species. The CDE is conserved among 18 of the 24 species analyzed.

3.1.5. AU-Rich elements

The sequences of alternative transcripts for mouse *Bmpr1a* were analysed and it was possible to find several sequences of the minimal AU-rich element (AUUUA) across the 3' UTR of the *Bmpr1a* l transcripts. All of them were annotated as shown on figure 15 for the different isoforms. Only the minimal sequence of AU-rich elements was found and none of the 5 classes of AU-rich (I-V) were present after the sequences analysis. The same procedure was done for the human *BMPR1A* where the minimal AU-rich element was annotated (Fig. 15) but none of the clusters have been found. Thus, it was not possible to confirm the conservation of the minimal sequence of A/U-rich elements between the human *BMPR1A* and mouse *Bmpr1a* after an alignment of both sequences, probably meaning they have different A/U-rich elements.

3.1.6. MicroRNA binding sites

To check for micro-RNA biding sites in *Bmpr1a* sequences, these were submitted to miRTarBase, (http://mirtarbase.mbc.nctu.edu.tw/php/search.php?org=mmu&kw=Bmpr1a&opt=target). Additionally, it was decided to annotate the sequences of micro-RNAs that were proven by bioinformatic and experimental approaches to bind to *Bmpr1a*. Two micro-RNAs were identified that fit in our search parameters, *mmu*-miR 27a-3p and *mmu*-miR 503-5p, which have been proved experimentally in MirTarBase online database.

3.1.7. Phylogeny

From our 3'UTR phylogeny it was possible to observe and confirm the evolution and relations between different species as Primates (H*omo sapiens* and *Macaca mulatta*), bony fishes (*Danio rerio*, *Echeneis naucrates* and *Labrus bergylta*), birds (*Athene cunicularia* and *Zonotrichia albicollis*), snakes (Python bivittatus and *Pseudonaja textilis*) rodents (*mus musculus* and *mus caroli*), marsupials (*Vombatus ursinus* and *Sarcophilus harrisii*) and cartilaginous fish (*Callorhinchus milii*). It was possible also to observe the similarity of *Bmpr1a* to its homologous in *Drosophila melanogaster Thickveins* and *Sax* genes (Fig. 20).

This phylogenetic study also showed us that the most related branch to the *mus musculus bmpr1a* is the one containing the *homo sapiens bmpr1a* and therefore that was probably a good model for the study of regulatory elements in this region due to less evolutionary distance between both species.

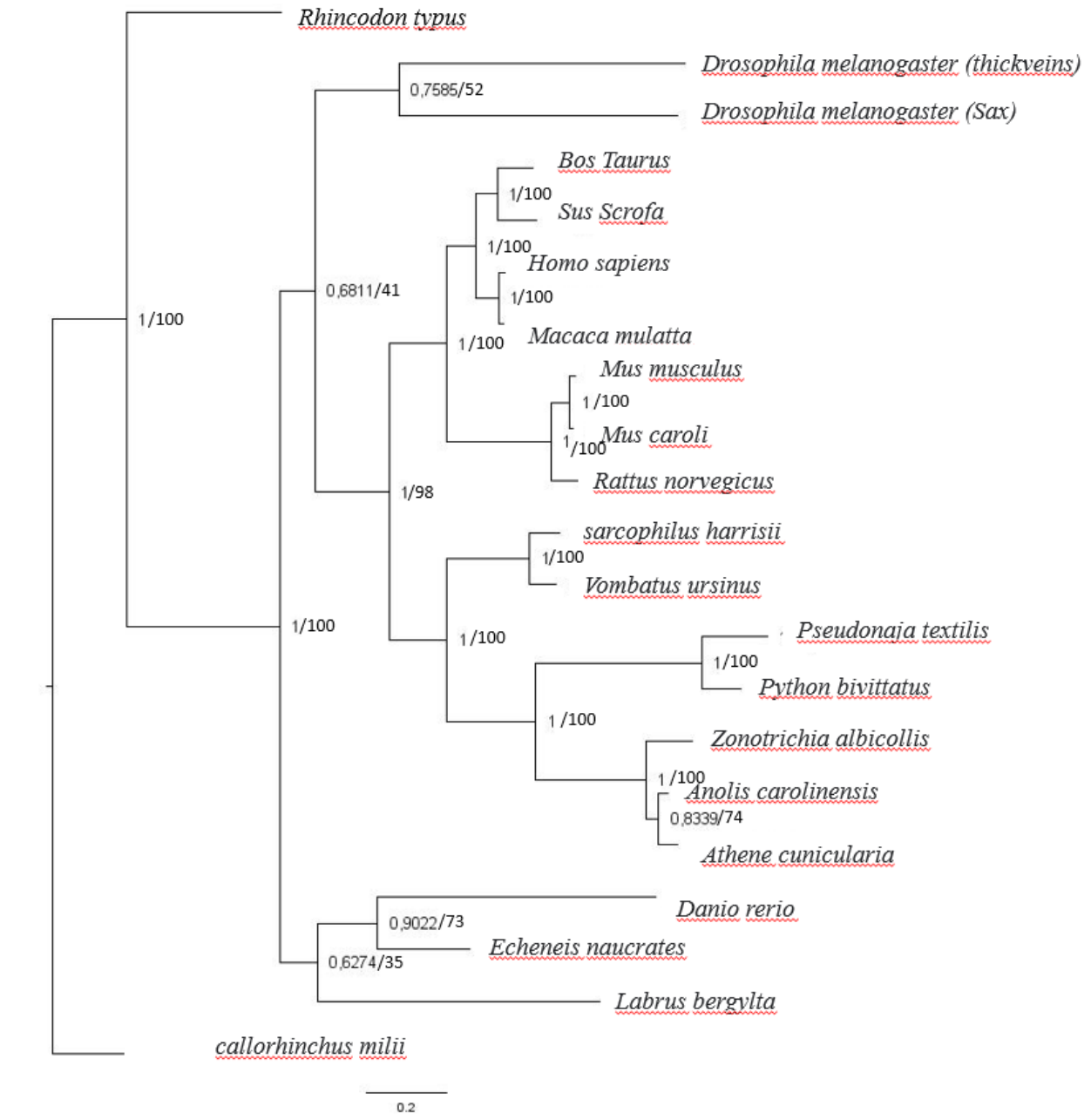

Figure 20- Phylogenetic analysis of Bmpr1a from 3'UTR of 20 different species. All the sequences were aligned using the software MAFFT using the algorithm L-INS-i. After the alignment the sequences were submitted to the software jModelTest to determine the best substitution model (GTR+I+G). Two threes were constructed. The first was constructed by submitting the alignment to the software MrBayes (v3.2.7) where through 25000 generations the posterior probability was calculated. The second tree was constructed by submitting the sequence to the software RaxmlGUI (v2.0) where through 500 replicates the bootstrap was calculated.

3.2. Experimental Results

3.2.1. Mineralization Induction of MC3T3-E1 cells

After 23 days MC3T3-E1 cells were stained with Alizarin red, which stains calcium deposits, and were then photographed and cells collected for further studies. It was possible to observe the formation of mineralization nodules in the induced mineralization cells but not on the control cells which have not received the mineralization cocktail as described before (Fig. 21).

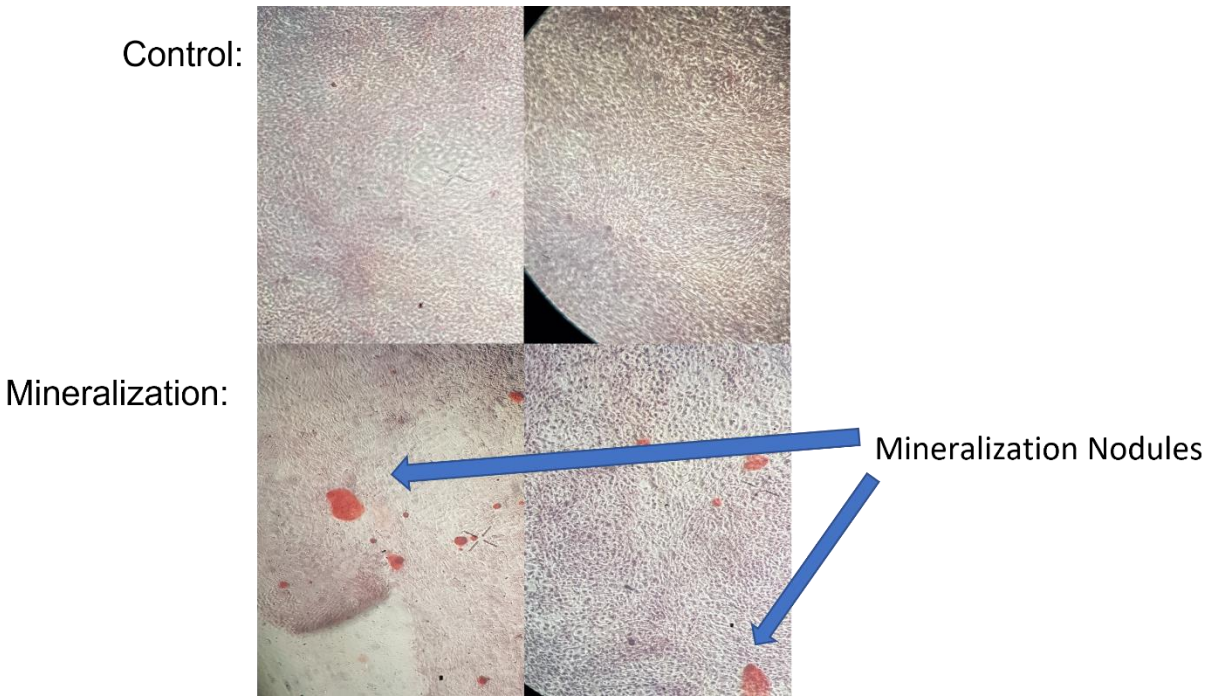

Figure 21- MC3T3-E1 mineralization induction. Photo of Alizarin red stained MC3T3 cells after 23days. The three photos on top are from the cells that were not treated with the mineralization cocktail, while the 3 photos on the bottom are from cell that were treated with the mineralization cocktail. It is possible to observe the formation of calcium deposits in the cells that were treated when compared to the control ones.

3.2.2. Amplification of *Bmpr1a* 3' UTR

We performed several different PCRs with different combination of primers to isolate different 3'UTR fragments and therefore identify different transcripts of mouse *Bmpr1a* from both differentiated and control cells to study if the differentiation would affect the size of *Bmpr1a* transcripts and therefore also affecting the number of regulatory elements present in the 3'UTR.

We were able to isolate a transcript with 250bp (Fig. 22) and a possible alternative polyadenylation site which is estimated to occur in 14.9% of 3'UTR cleavage and polyadenylation sites (Fig. 23).

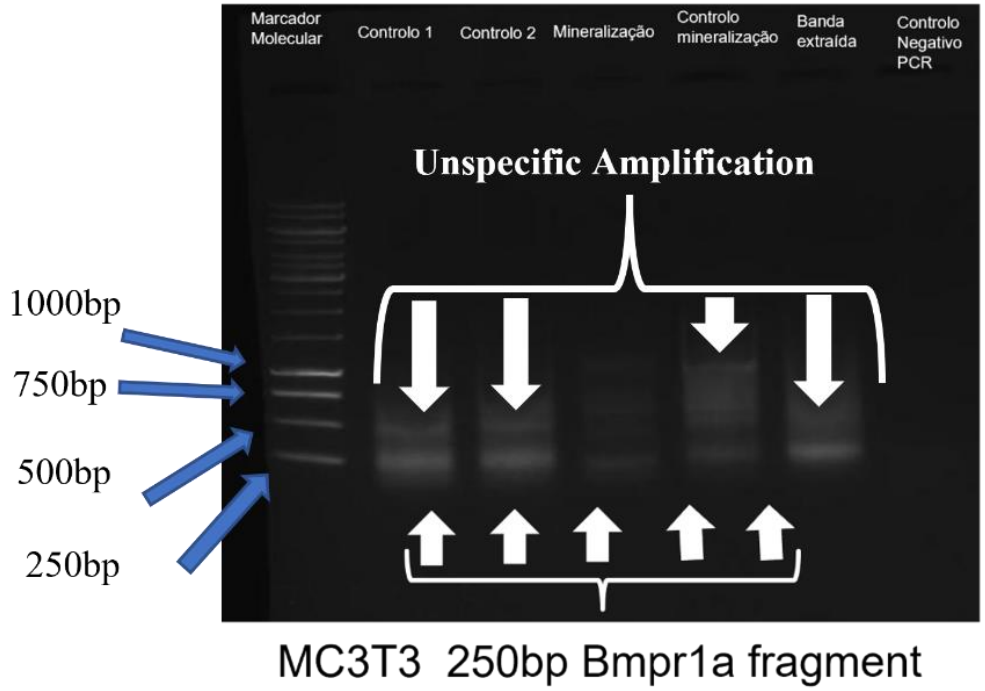

Figure 22-- Identification of Bmpr1a 3'UTR transcripts. Agarose Gel electrophoresis of a PCR for the isolation of 3'UTR. The bottom band is a 250bp fragment from bmpr1a while the top band is unspecific amplification.

We also performed isolation of 3' UTRs in different mouse tissues using two different approaches. In the first one we tried to isolate the 3'UTR by using as reverse primer an oligo dT amplifying transcript from the poly-A tail. The second approach was by using primers that could bind the sequence upstream the polyadenylation sites. With these two approaches we were not able to isolate any longer transcript from mouse knee and joint tissues and therefore we tried to isolate a Bmpr1a from liver mouse tissue. From that tissue (Liver) we found a larger Bmpr1a fragment with 1500bp (Fig. 24).

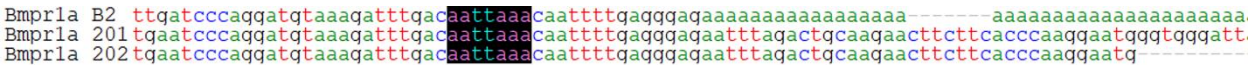

Figure 23-Alternative polyadenylation site on 250bp Bmpr1a fragment. Alignment of the Bmpr1a fragments isolated from MC3T3 cells versus the sequence of ensemble (ENST00000372037.8). Marked in black is an alternative polyadenylation site.

The larger fragment contains the CDE loop as regulatory element present in that sequence but no polyadenylation site because the reverse primers were binding the sequence before the first polyadenylation site annotated in this study.

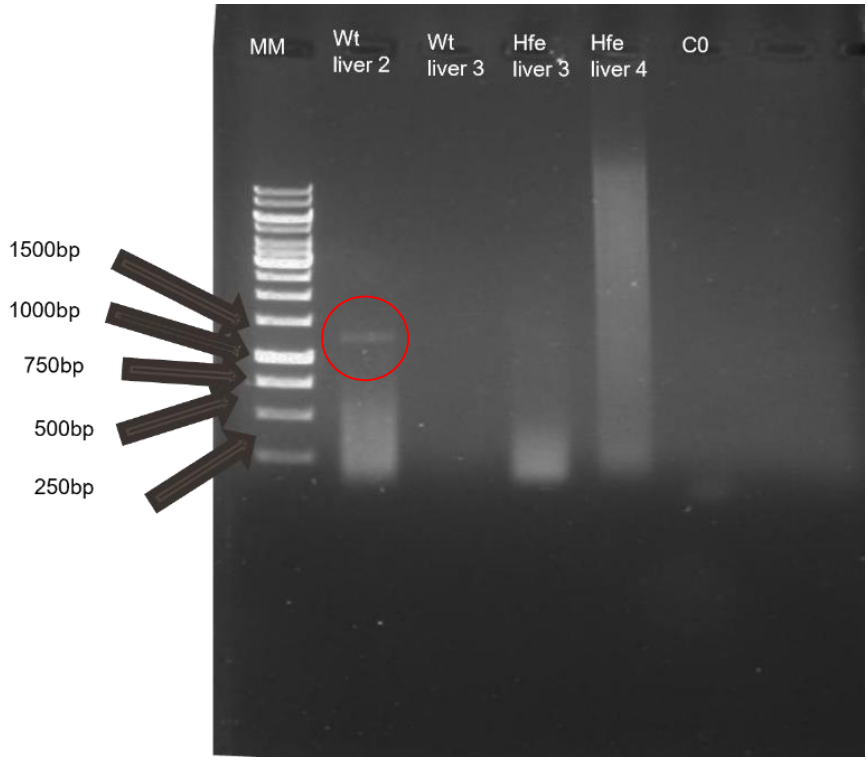

Figure 24- Identification of bmpr1a 1500bp liver transcript. Agarose Gel electrophoresis of a PCR from mouse liver cells. The red circle shows the Bmpr1a liver transcript obtained.

3.2.3. Amplification of *Bmpr1a* 5' UTR

With a combination of primers, we aimed to amplify 4 different fragments from the 5' UTR, the smallest fragment (**D**) was 244 bp long and included only two of the three uORFs identified previously on this study. The second fragment (**B**) was 277 bps long and included all the uORFs, however it starts 7 base pairs before the first uORF. The third fragment(**C**) was 329 bp long and starts in the same nucleotide as fragment **B** but also contains the AUG from the main ORF of *bmpr1a*. This AUG is considered a strong AUG and therefore will be used as a positive control for the expression regulation by the uORFs because it will be recognized by the translation machinery and alter the expression of luciferase. The last fragment and the longer one (**A**) contain 335 bp and

the three uORFs, representing increased distance from the start of the fragment to the first uORFs (Fig. 25).

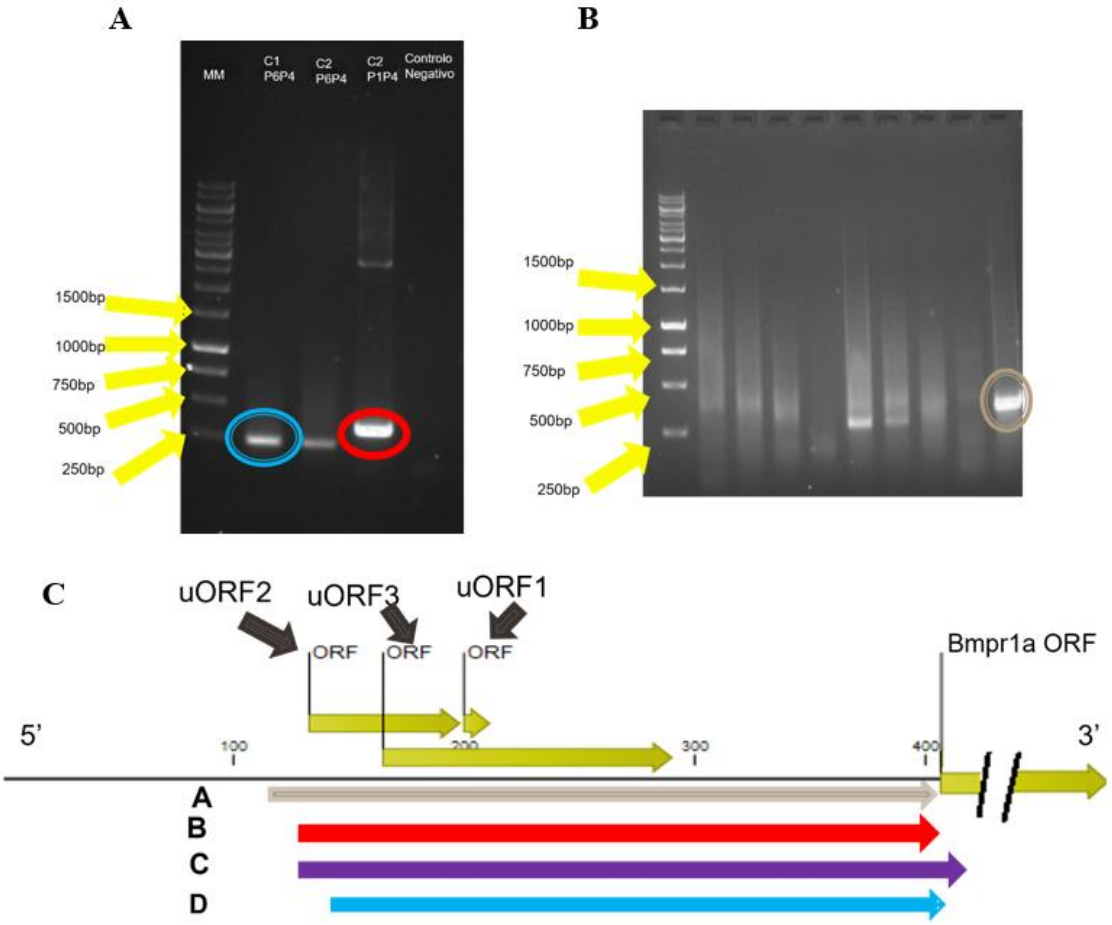

Figure 25- Isolatation of the 5'UTR transcripts. A) Agarose Gel electrophoresis of Bmpr1a 5'UTR transcript. In Blue the fragment D, in red the fragment B and B) in Brown the fragment A. C) Schematic representation of Bmpr1a 5'UTR with the localization and size of the uORF and the representation of the fragments isolated from this region.

3.2.4. Functional activity of uORFs present in *Bmpr1a* 5'UTR

The four fragments amplified and cloned previously (A, B, C and D) were inserted on pGL3 plasmid before the sequence that encodes the luciferase protein. The constructs, after being cloned, were transfected into HEK293 cells and after 48h cells were lysed, and the luminescence produced recorded. Our firsts transfections with all the fragments without any mutations showed a decrease in the expression of all four transcripts (Fig. 26). The first fragment analysed (fragment A) showed

a decrease to 0.7-fold; the second fragment (fragment B) it was possible to observe a significant 0.6-fold decrease in the expression of luciferase; on the fragment C that contains the AUG from *Bmpr1a* was also observed a significant decrease in the expression of about 0.2-fold and in fragment D that is the smallest one the expression of luciferase decreased by 0.3-fold when compared to the positive control

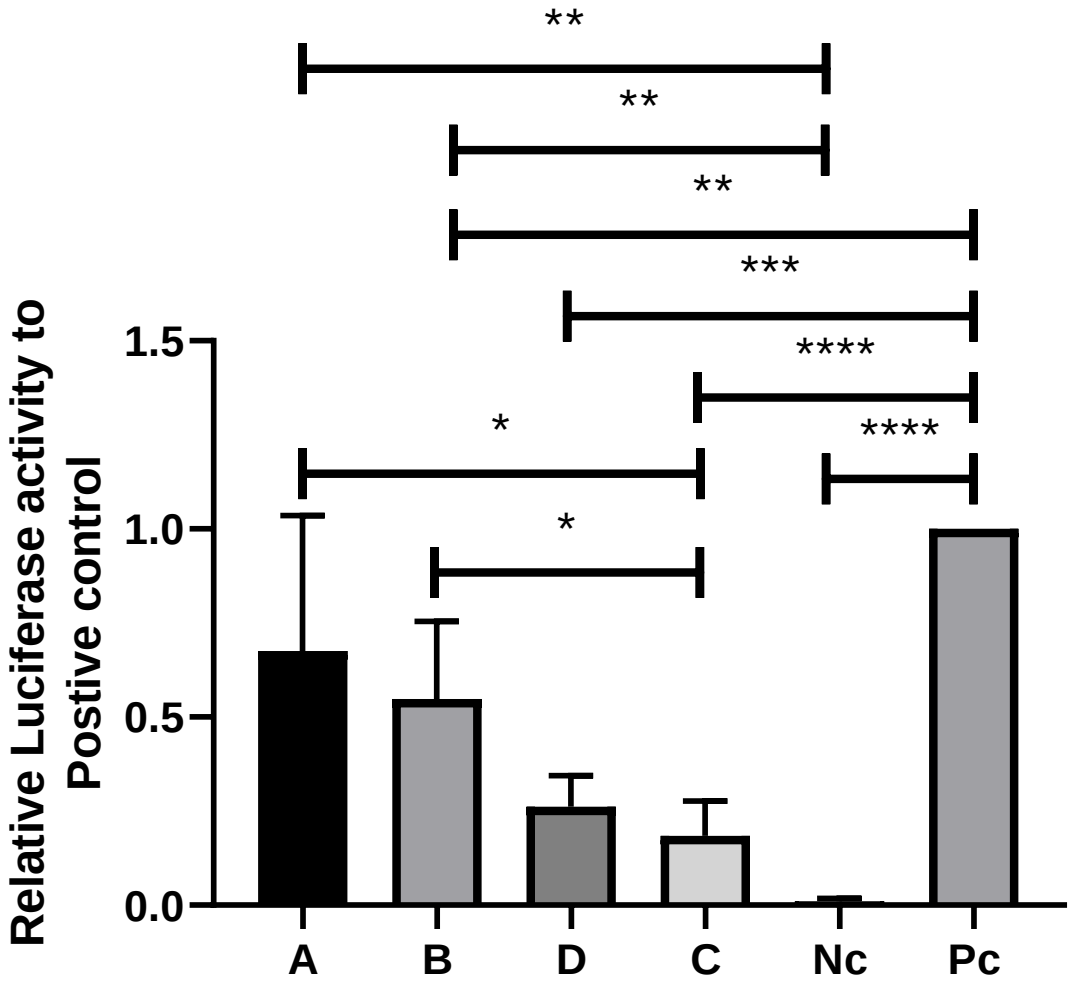

Figure 26- Relative fold of luminescence recorded by the expression of the luciferase gene of pGL3 with the different fragments isolated from Bmpr1a 5'UTR. The statistical approach was an ordinary one-way anova followed by Tukey post hoc test with p< 0 05 for statistical significance N=2. The error bar corresponds to the Standard derivation (SD)

To understand if the elimination of the AUGs present in uORFs annotated in 5'UTR of *Bmpr1a* can reverse their impact on luciferase expression we have promoted a direct mutagenesis assay

which leaded to the exchange of T by A in pGL3 plasmid, changing the start codon in uORFs to AAG.

The analysis of fragment B uORFs on the expression of luciferase when compared to the positive control showed a decrease to 0.5-fold in expression of luciferase relative to control, suggesting an inhibition of the uORFs in the expression of luciferase. The mutation of AUG on uORF1 in fragment B showed a significant increase to 1.5-fold (Fig. 27). The mutation of uORFs 1 and 2 (SDM2) showed a significant decrease to 0.1-fold and mutations of uORFs 1, 2 and 3 (SDM3) showed no significant differences to control but were increased relatively to fragment B without mutations (Fig. 27). The fragment C in this case was included and working as a proof of concept of expression of the uORFs, because it contains the AUG from *Bmpr1a* that could work as a strong uORF and therefore regulating the expression of luciferase. The negative control confirmed downregulation of luciferase expression (Fig. 27).

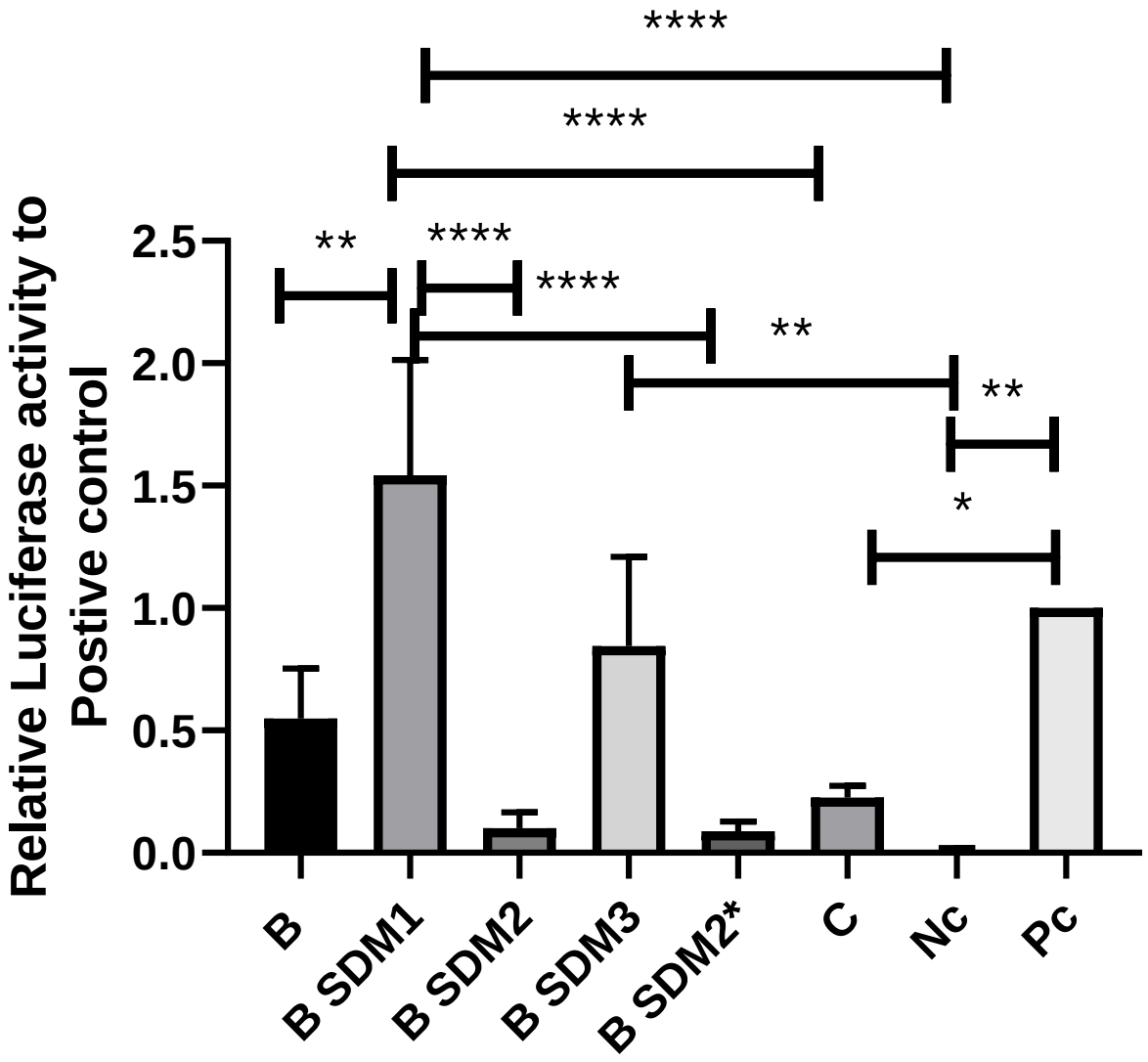

Figure 27-Relative fold of luminescence recorded by the expression of the luciferase gene of pGL3 with fragment B and the sequential mutations and the isolated mutations (marked with* on the name). The statistical approach was an ordinary one-way anova followed by Tukey post hoc test with p< 0 05 for statistical significance N=3. The error bar corresponds to the Standard derivation (SD)

The second fragment that we analysed was the fragment A, which is the longest one and it was possible to observe a decrease to 0.6-fold in the expression of luciferase in the fragment that does not contain any mutation, suggesting also an inhibition of the uORFs in the expression of luciferase. The expression of luciferase in the first mutation (SDM1) decreased to 0.8-fold when compared to

the control and the sequential mutations SDM2 and SDM3 increased expression of luciferase (1.5-fold and 1.3-fold respectively) compared to control (Fig. 28). The single mutations SDM1* and SDM2* confirm the impact in the expression of luciferase by SDM1 and SDM2 but at the time of writing of this work only 2 independent experiments were done with this fragment and further studies are required to validate the regulation of expression by these uORFs in 5´UTR longer than fragment B for example.

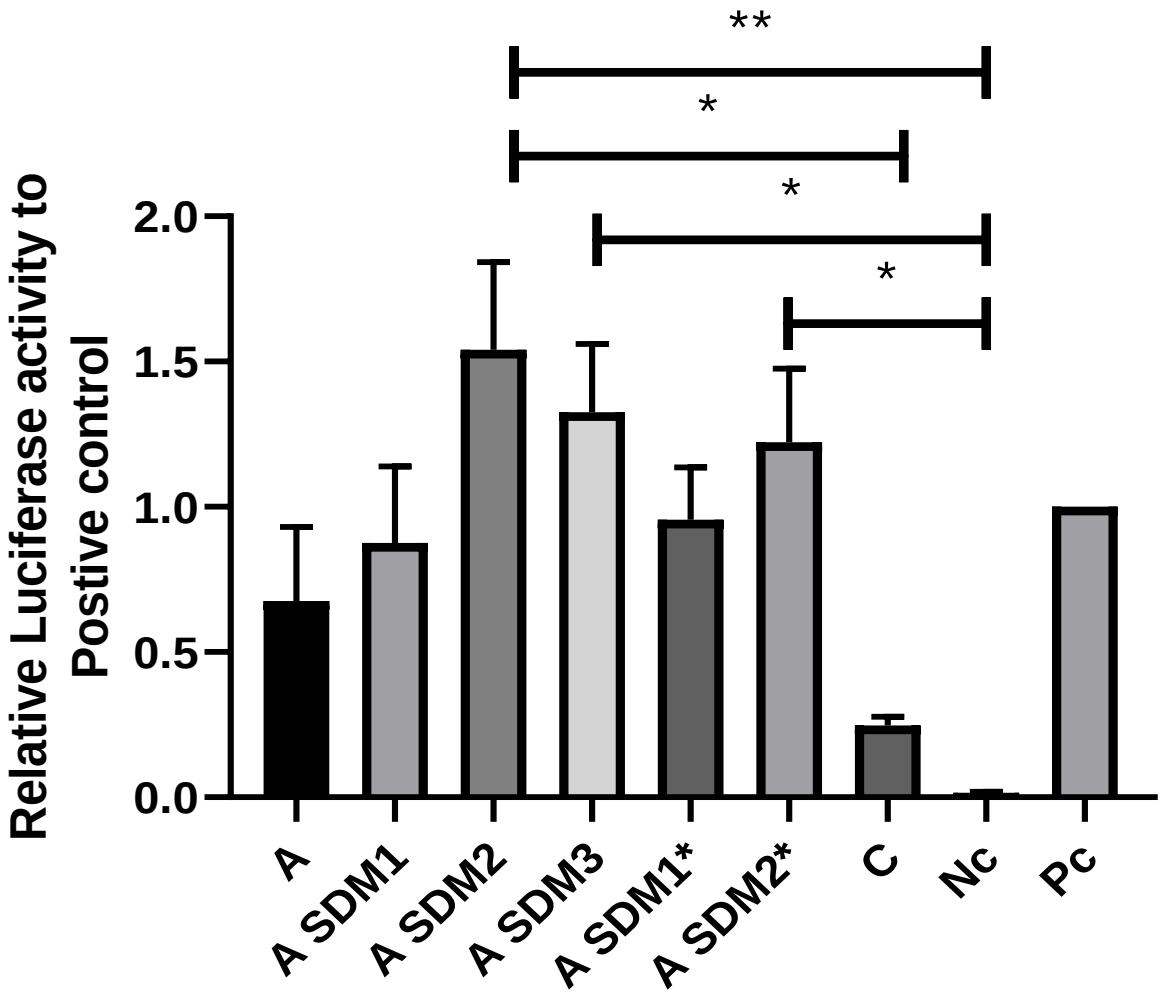

Figure 28- Relative fold of luminescence recorded by the expression of the luciferase gene of pGL3 with fragment A and the sequential mutations and the isolated mutations (marked with* on the name). The statistical approach was an ordinary one-way anova followed by Tukey post hoc test with p< 0 05for statistical significance. N=2. The error bar corresponds to the Standard derivation (SD)

The last fragment analyzed was the fragment D which is the smallest one and does not contain the uORF2. It was possible to observe a decrease to 0.6-fold in the expression of the fragment that does not contain any mutations. The expression of luciferase in the first mutation (SDM1) increased to 1.3-fold when compared to the control and an increase to 1.35-fold in the fragment that contains both mutations. For this fragment it was also possible to obtain 2 independent experiments. Statistical analysis showed no significant differences. One possible explanation is the lack of uORF2 that can be important for the regulation of expression by uORFS. Further studies are required to validate the regulation of *Bmpr1a* transcripts with shorter 5´UTRs and the impact of uORF2 (Fig. 29).

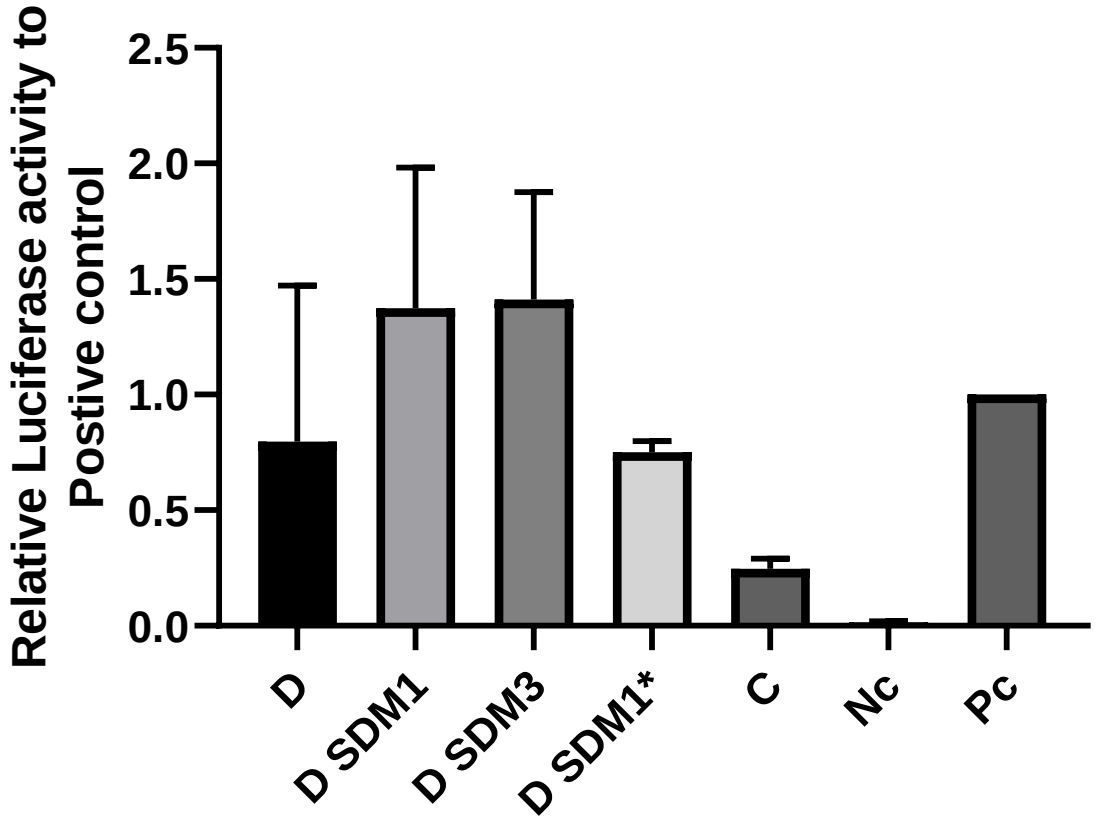

Figure 29- Relative fold of luminescence recorded by the expression of the luciferase gene of pGL3 with fragment D and the sequential mutations and the isolated mutations (marked with* on the name). The statistical approach was an ordinary one-way anova followed by Tukey post hoc test with p< 0 05for statistical significance. N=2. The error bar corresponds to the Standard derivation (SD)

3.3. Discussion

The main objective of this work was to identify and characterize the regulatory elements present in 5'UTR and 3'UTR of mouse *Bmpr1a* transcripts and compare with available human *BMPR1A* transcripts. We started by searching online databases for all the available *bmpr1a* transcripts from *mus musculus* and *Homo sapiens* and annotated the regulatory elements identified in the sequences.

Focusing on the identification of alternative polyadenylation sites, the analysis of *Bmpr1a* gene allowed the identification of 11 different polyadenylation sites (Fig. 18), 6 canonical polyadenylation sites and 5 using a bioinformatic tool RegRNA2.0. It was possible to observe that 5 of the canonical polyadenylation sites are close to the ones identified by the bioinformatic tool. The last canonical site is near the end of the ensembl 201 (ENSMUST00000049005.15) transcript (Fig. 14), and it is possible that the site can be recognized by the polyadenylation machinery and the RNA is cleaved on that site[130] In the human transcripts we found 8 canonical polyadenylation sites in the longer transcript (ENST00000372037.8) but only one near the end of the other 2 transcripts (ENST00000480152.2 and ENST00000638429.1).After seeking conservation between mouse and human polyadenylation sites it was only possible to observe conservation of one site. A previous study reported that 49.3% of the human genes have 3 or more polyadenylation sites[131] while in mouse the mean number of polyadenylation per gene is 2.53 [132]. Accordingly, our results do not show a big discrepancy from what is described in the literature.

On our experimental procedures we were able to isolate only a fragment from *Bmpr1a* 3'UTR in MC3T3-E1 cells from proliferating and differentiated cells. This fragment was 250bp and we could find an alternative polyadenylation site that is close to the poly-A tail and therefore be recognized by the polyadenylation complex, which occurs in 14.9% of the times [133]. This shorter fragment can be explained by the necessity of the osteoblasts to produce BMPR1A because a shorter 3'UTR means lower probability of containing regulatory elements that decrease expression[134,135,132] .

We have tried to amplify alternative 3´UTRs from other mouse tissues like knee and joints but it was only possible to amplify fragments with around 250bp (see Annexes Fig. 34), like what was obtained for MC3T3-E1 cells, suggesting predominance of these shorter 3´UTRs when *Bmpr1a* is expressed. However, due to unspecific nature of dT adaptor primer and combination of cells present in these tissues, it will be necessary to confirm these results. Attending to this fact we have used mouse liver tissue cDNA library to try to amplify longer *Bmpr1a* 3´UTRs. Results revealed the

amplification of a fragment with 1500bp, an alternative 3´UTR longer than the one previously obtained from MC3T3-E1 cells. From this tissue we also tried to amplify shorter and longer Bmpr1a 3´UTRs fragments but was only possible to amplify the 1500bp fragment, suggesting possible specificity of this *Bmpr1a* 3´UTR to this tissue. The RNA is cleaved on the first polyadenylation site in accordance with the literature that says that there is an temporal advantage of proximal sites[24]

In 2013 Leppek and colleagues found that the 3´UTR of *Bmpr1a* transcripts could possibly have a CDE sequence which was reported to be recognized by Roquin proteins and promote the rapid RNA degradation [136]. The analysis of the alternative transcripts previously published showed that CDE were present in 7 of the 11 alternative transcripts identified (Fig. 14). The analysis of 3´UTR *Bmpr1a* sequences of transcripts amplified from MC3T3-E1 cells showed that CDE was not present in these isoforms obtained from osteoblasts. Additionally, we have used other mouse tissues cDNA libraries (bone, articular cartilage, and liver) but it was possible to identify the CDE signature sequence only in liver tissue cells. After identification of the CDE loop on the 3'UTR, we look for conservation of the obtained sequence and compared with different vertebrate species and concluded that it was 100% conserved in 18 of the 24 species analysed (Fig. 6). Furthermore, in 22 of them, even in those that it was not conserved, the CDE could possibly also be functional by the nucleotide sequence requirements[40]

In this study we were able to identify 3 possible uORFs on the 5'UTR of *Bmpr1a;* our findings go in accordance with the bioinformatic tools (CLC sequence Viewer 8) used and we were able to identify 9 possible uORFs in which 3 of them where present in most of the transcripts. Taking that in consideration, we selected those 3 to develop further studies and confirmed that the smaller one was a strong uORF based on what is described by Romão et al., 2013[54], meaning that it contains a purine in the position -3 and an adenine on position+1 (A/GCCAUGG). This is required to make the translation machinery stop and recognize the premature AUG from the uORF[54]. After seeing conservation along the different transcripts described for *mus musculus,* we decided to confirm it in transcripts from other species. For that we have done a multiple alignment of *Bmpr1a* 5´UTRs of vertebrate species. It was possible to identify uORFs clusters highly conserved in all mammalian species except for marsupials (Fig. 16), suggesting high levels of conservation in the function of these uORFs for regulation of *Bmpr1a* transcript translation. In primates we discovered other uORFs downstream of the cluster discussed, still is not clear how important can be its location

before main ORF and further studies about this uORFs are required. The analysis of uORFs clusters organization in *Bmpr1a* 5´UTR of bird species like *Gallus gallus* showed a very similar organization but with some differences in the uORFs length and number (Fig. 16). For fish species we also found clusters of 4 uORF that appear to be similar to mammals in its relative position to main ORF (Fig. 16). However, their nucleotide composition is not conserved when compared with mammals, which suggests that the presence of uORFs clusters in *Bmpr1a* 5´UTR, leading to the eventual translation of small peptides, could be very important in the fine tuning of *Bmpr1a* transcript translation[5,12,136]. .

Experimentally we were able to isolate four different fragments from MC3T3 cells, where the fragment A, B and C contain all the 3 uORFs from *Bmpr1a* 5´UTR, while fragment D only contains the uORF 1 and 3. From our results the fragments B and D promoted a significant reduction of the expression of luciferase relatively to positive control (pGL3 plasmid without fragments) (Fig. 26). The results suggest that the uORFs clusters cloned affect translation efficiency of the luciferase transcript expressed by the plasmids[137,51] and possibly play a role on Bmpr1a expression. Fragment A also promoted a decrease in luciferase detection but without statistical significance. Further assays are needed to understand if it can have the same effect as fragment B and D. Fragment C was used as a negative control of luciferase translation efficiency because it contains the AUG from the *Bmpr1a* main ORF and was used as an artificial uORFs, in addition to the already cloned cluster. It was clear that it affected the expression of luciferase (Fig. 26)

With the objective of validating the impact of uORFs on the expression of luciferase, site directed mutagenesis was done to the 3 uORFs of fragments A, B and D. After mutation of the AUGs from the uORFs to AAG, it was possible to observe in fragment B a significant increase in the expression of luciferase on the fragment where only the uORF1 was mutated (Fig. 27), which suggests that it was possible to revert the impact of uORF1 on the efficiency of translation of the luciferase transcript. A significant decrease in the expression of the same fragment with the mutations of both uORF1 and uORF2 was observed. The decrease in luciferase detection was not expected with this combination of mutations. The negative impact of uORF2 mutation in fragment B was confirmed by the single mutation of uORF2 (Fig. 27). One of the possible explanation for this fact is that the AUG from uORF2 is near the start of the fragment B and the mutation is possibly causing a destabilization on that fragment (see fig) and therefore decrease the expression of luciferase. Fragment A promoted a decreased in the expression of luciferase relatively to positive

control. The accumulation of mutations in AUGs of uORF1, 2 and 3 leaded to an increase in luciferase detection relative to fragment A without mutations (Fig. 28). These results could represent a reversal impact of uORF1, 2 and 3 on luciferase expression. This fragment was longer towards 5´UTR, which could explain why it did not show the same effect on mutation of uORF2, allowing a stabilization of uORF2 region. For this fragment only two independent experiments were done, therefore further studies are needed to further understand this mechanism. In the fragment D, although no significance between the control and mutations in uORFs 1 and 3 was observed, the absence of the uORF2 could have an effect on the regulation of expression by the uORFs as one of the features affecting the expression of a gene by uORF is the number of uORFs[54,136] present in the sequences but further studies are required.

3.4. Conclusion and Future Perspectives

From our experimental work it was possible to observe an alternative polyadenylation site of *bmpr1a* expressed in MC3T3 cells. This site was not described in the literature, and its presence could possibly indicate the need to rapidly express a new shorter transcript that does not contain any of the 3'UTR regulatory elements searched on this study. For example, the presence of a CDE loop in the transcript of Bmpr1a isolated from mouse liver cells could possibly mean that, on that tissue, the regulation of expression of *Bmpr1a* occurs in a different manner through the expression of a different transcript isoform that contain other regulatory elements. This CDE loop was also conserved in several species that we compared and therefore it can possibly influence the expression of *Bmpr1a* across different taxa.

Bioinformatically, we identified the presence of several uORFs in the 5'UTR region of *bmpr1a* transcripts and 3 of those uORFs were conserved in almost all transcripts of this gene annotated from mouse. These uORF were selected for this study due to the conservation among the different mouse transcripts and to the conservation on mammals except marsupials.

From MC3T3 cells that we were able to differentiate into mature osteoblast cells, as confirmed by alizarin red staining of its calcium deposits in the extracellular matrix, we isolated different fragments containing either 2 or the 3 uORFs present in the sequences, depending on the fragments. With the luciferase assay we could observe alterations in the expression of those fragments in the presence or absence of mutations in the corresponding AUGs, In addition, the expression of reporter vector gene (luciferase) decreased in all fragments without the mutations indicating that the presence of those uORF negatively affected the expression of the luciferase. Through mutations of those AUGs to AAG we saw an increase in the expression of luciferase except on mutation of uORF2 on fragment B. In this case, we hypothesized that the decrease is due to the destabilization of the transcript because the mutation is located at the start of the transcript.

Our results suggest that, under certain cellular conditions, those uORF are useful to the cell machinery to regulate the expression of *Bmpr1a.* In this study it was possible to observe that the reduction of expression caused by the presence of those uORFs could be reverted upon mutation of the corresponding AUGs, however further studies are needed to confirm this finding.

In the future, and to extend and complement this work, it should be interesting to identify more 3'UTRs from different tissues from mouse or human using different cDNA libraries or through

RNAseq and then, with the 3'UTRs obtained, evaluate the function of the regulatory elements found. For example the function of the CDE loop could be further explored by other techniques such as ChIp or co-transfections of plasmids containing the CDE and *Roquin*.

More studies concerning the evaluation of levels of mRNA containing different uORFs are needed to confirm if, indeed, their expression are regulated by the translation machinery interacting at those different AUGs. It would also be important to investigate if human transcripts containing different uORFs are associated with its regulation in different tissues or times of development and perform functional assays to prove it.

The identification of regulatory elements both in 5'- and 3'-UTRs of gene transcripts is providing new insights into the mechanisms of regulation of those genes and could prove to be very relevant to identify new possibilities to interact with the regulation of genes of interest in pathological situations and identify eventual novel therapeutic targets.

3.5. References

1. Gerstein, M. B. *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17**, 669–681 (2007).
2. Keller, E. F. & Harel, D. Beyond the gene. *PLoS One* **2**, (2007).
3. Hegde, M. R. & Crowley, M. R. Genome and gene structure. *Emery Rimoin's Princ. Pract. Med. Genet. Genomics Found.* 53–77 (2018) doi:10.1016/B978-0-12-812537-3.00004-4.
4. Shafee, T. & Lowe, R. Eukaryotic and prokaryotic gene structure. *WikiJournal Med.* **4**, 0–3 (2017).
5. Morris, D. R. & Geballe, A. P. Upstream Open Reading Frames as Regulators of mRNA Translation. *Mol. Cell. Biol.* **20**, 8635–8642 (2000).
6. Wang, Y. C., Peterson, S. E. & Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **24**, 143–160 (2014).
7. Li, R., Wei, X. & Jiang, D. S. Protein methylation functions as the posttranslational modification switch to regulate autophagy. *Cell. Mol. Life Sci.* **76**, 3711–3722 (2019).
8. Stoecklin, G., Lu, M., Rattenbacher, B. & Moroni, C. A constitutive decay element promotes tumor necrosis factor alpha mRNA degradation via an AU-rich element-independent pathway. *Mol. Cell. Biol.* **23**, 3506–15 (2003).
9. Hocine, S., Singer, R. H. & Gru, D. RNA Processing and Export. 1–20 (2010).
10. Matoulkova, E., Sommerova, L., Pastorek, M., Vojtesek, B. & Hrstka, R. Regulation of AGR2 expression via 3'UTR shortening. *Exp. Cell Res.* **356**, 40–47 (2017).
11. Haneklaus, M. Analysis of Post-transcriptional Gene Regulation of Nod- Like Receptors via the 3′UTR. *Methods Mol. Biol.* **1390**, v (2016).
12. Leppek, Kathrin ; Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol.* (2018) doi:10.1038/nrm.2017.103.Functional.

13. Kornblihtt, A. R. *et al*. Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165 (2013).

14. Bush, S. J., Chen, L., Tovar-corona, J. M. & Urrutia, A. O. Alternative splicing and the evolution of phenotypic novelty. 1–7 (2017).

15. Chen, M. & Manley, J. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol.* **10**, 741–754 (2009).

16. Chen, M. & Manley, J. L. NIH Public Access. **10**, 741–754 (2010).

17. Ramanouskaya, T. V & Grinev, V. V. The determinants of alternative RNA splicing in human cells. *Mol. Genet. Genomics* **292**, 1175–1195 (2017).

18. Zhang, X. *et al.* Branch Point Identification and Sequence Requirements for Intron Splicing in Plasmodium falciparum □ †. **10**, 1422–1428 (2011).

19. Chen, W. *et al.* Alternative Polyadenylation : Methods , Findings , and Impacts. *Genomics. Proteomics Bioinformatics* **15**, 287–300 (2017).

20. Yang, Quin ; Dublié, S. Structural biology of poly(A) site definition. *Wiley Interdiscip Rev RNA* (2011).

21. Tian, B., Manley, J. L. & Genetics, M. HHS Public Access. **18**, 18–30 (2017).

22. Mayr, C. Regulation by 3 -Untranslated Regions. (2017).

23. Denome, R. M. & Cole, C. N. Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. *Mol. Cell. Biol.* **8**, 4829–4839 (1988).

24. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev* **25**, 6110–6112 (2011).

25. Elkon, R., Ugalde, A. P. & Agami, R. Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* **14**, 496–506 (2013).

26. Martin, G., Gruber, A. R., Keller, W. & Zavolan, M. Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. *Cell Rep.* **1**, 753–763 (2012).

27. Huang, H., Chen, J., Liu, H. & Sun, X. The nucleosome regulates the usage of polyadenylation sites in the human genome. *BMC Genomics* **14**, (2013).

28. Gruber, A. J. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* doi:10.1038/s41576-019-0145-z.

29. Turner, R. E., Pattison, A. D. & Beilharz, T. H. Alternative polyadenylation in the regulation and dysregulation of gene expression. *Semin. Cell Dev. Biol.* **75**, 61–69 (2018).

30. Barreau, C., Paillard, L. & Osborne, H. B. AU-rich elements and associated factors: Are there unifying principles? *Nucleic Acids Res.* **33**, 7138–7150 (2005).

31. Zubiaga, A. N. A. M., Belasco, J. G. & Greenberg, M. E. The Nonamer UUAUUUAUU Is the Key AU-Rich Sequence Motif That Mediates mRNA Degradation. **15**, 2219–2230 (1995).

32. Dean, J. L. E. *et al.* The 3′ Untranslated Region of Tumor Necrosis Factor Alpha mRNA Is a Target of the mRNA-Stabilizing Factor HuR. *Mol. Cell. Biol.* **21**, 721–730 (2001).

33. Wei, W. *et al.* Osteoclast progenitors reside in the peroxisome proliferator-activated receptor γ-expressing bone marrow cell population. *Mol. Cell. Biol.* **31**, 4692–4705 (2011).

34. Yang, Q., Jeremiah Bell, J. & Bhandoola, A. T-cell lineage determination. *Immunol. Rev.* **238**, 12–22 (2010).

35. Halees, A. S. *et al.* Global assessment of GU-rich regulatory content and function in the human transcriptome. *RNA Biol.* **8**, (2011).

36. Pérez Cañadillas, J. M. & Varani, G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.* **22**, 2821–2830 (2003).

37. Leppek, K. *et al.* Roquin Promotes Constitutive mRNA Decay via a Conserved Class of Stem-Loop Recognition Motifs. *Cell* **153**, 869–881 (2013).

38. Masiewicz, P., Za, J., Stoecklin, G. & Carlomagno, T. A Distinct , Sequence-Induced Conformation Is Required for Recognition of the Constitutive Decay Element RNA by Roquin Article A Distinct , Sequence-Induced Conformation Is Required for Recognition of the Constitutive Decay Element RNA by Roquin. 1437–1447 (2015)

doi:10.1016/j.str.2015.06.001.

39. Schaefer, J. S. & Klein, J. R. Roquin--a multifunctional regulator of immune homeostasis. *Genes Immun.* **17**, 79–84 (2016).

40. Essig, K. *et al.* Roquin targets mRNAs in a 3′-UTR-specific manner by different modes of regulation. *Nat. Commun.* **9**, (2018).

41. Sgromo, A. *et al*. A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nat. Commun.* **8**, (2017).

42. Essig, K. *et al.* Roquin targets mRNAs in a 3′-UTR-specific mannerby different modes of regulation. *Nat. Commun.* (2018) doi:10.1038/s41467-018-06184-3.

43. Inada, T. & Makino, S. Novel roles of the multi-functional CCR4-NOT complex in post-transcriptional regulation. *Front. Genet.* **5**, 1–7 (2014).

44. Mohr, A. M. & Mott, J. L. Overview of microRNA biology. *Semin. Liver Dis.* **35**, 3–11 (2015).

45. Baskara-Yhuellou, I. & Tost, J. Chapter Six - The impact of microRNAs on alterations of gene regulatory networks in allergic diseases. in *Inflammatory Disorders - Part B* (ed. Donev, R. B. T.-A. in P. C. and S. B.) vol. 120 237–312 (Academic Press, 2020).

46. Man, X. *et al.* MiR-503 inhibits adipogenesis by targeting bone morphogenetic protein receptor 1a. **8**, 2727–2737 (2016).

47. Calin, G. A. & Croce, C. M. MicroRNA-Cancer Connection: The Beginning of a New Tale. *Cancer Res.* **66**, 7390 LP – 7394 (2006).

48. Yi, M. *et al.* The role of cancer-derived microRNAs in cancer immune escape. *J. Hematol. Oncol.* **13**, 25 (2020).

49. He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828–833 (2005).

50. Ryan, B., Joilin, G. & Williams, J. M. Plasticity-related microRNA and their potential contribution to the maintenance of long-term potentiation. *Front. Mol. Neurosci.* **8**, 1–17 (2015).

51. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7507–7512 (2009).

52. Young, S. K. & Wek, R. C. Upstream open reading frames differentially regulate genespecific translation in the integrated stress response. *J. Biol. Chem.* **291**, 16927–16935 (2016).

53. Chen, H. & Tarn, W. uORF-mediated translational control : recently elucidated mechanisms and implications in cancer. *RNA Biol*. **0**, 1–12 (2019).

54. Barbosa, C. & Peixeiro, I. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. **9**, 1–12 (2013).

55. Wu, M., Chen, G. & Li, Y. TGF- β and BMP signaling in osteoblast , skeletal development , and bone formation , homeostasis and disease. (2016) doi:10.1038/boneres.2016.9.

56. Karsenty, G., Kronenberg, H. M. & Settembre, C. Genetic Control of Bone Formation. *Annu. Rev. Cell Dev. Biol.* **25**, 629–648 (2009).

57. Paiva, K. B. S. & Granjeiro, J. M. Bone tissue remodeling and development: Focus on matrix metalloproteinase functions. *Arch. Biochem. Biophys.* **561**, 74–87 (2014).

58. Stricker, S. & Mundlos, S. *FGF and ROR2 Receptor Tyrosine Kinase Signaling in Human Skeletal Development*. *Current Topics in Developmental Biology* vol. 97 (Elsevier Inc., 2011).

59. Long, F. & Ornitz, D. M. Development of the endochondral skeleton. *Cold Spring Harb. Perspect. Biol.* **5**, 1–20 (2013).

60. Berendsen, A. D. & Olsen, B. R. Bone development. 14–18 (2016) doi:10.1016/j.bone.2015.04.035.Bone.

61. Shahi, M., Peymani, A. & Sahmani, M. Regulation of Bone Metabolism. **5**, (2017).

62. Clarke, B. Normal bone anatomy and physiology. *Clin. J. Am. Soc. Nephrol.* **3 Suppl 3**, 131–139 (2008).

63. Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and

osteogenesis by a specific vessel subtype in bone. **507**, 323–328 (2016).

64. Grabowski, P. Physiology of Bone. *Endocr. Dev.* **28**, 33–55 (2015).

65. Oftadeh, R., Perez-Viloria, M., Villa-Camacho, J. C., Vaziri, A. & Nazarian, A. Biomechanics and Mechanobiology of Trabecular Bone: A Review. *J. Biomech. Eng.* **137**, 1–15 (2015).

66. Rutkovskiy, A., Stensløkken, K.-O. & Vaage, I. J. Osteoblast Differentiation at a Glance. *Med. Sci. Monit. Basic Res.* **22**, 95–106 (2016).

67. Akbari, S. Vitamin K and Bone Metabolism : A Review of the Latest Evidence in Preclinical Studies. **2018**, (2018).

68. Papachroni, K. K., Karatzas, D. N., Papavassiliou, K. A., Basdra, E. K. & Papavassiliou, A. G. Mechanotransduction in osteoblast regulation and bone disease. *Trends Mol. Med.* **15**, 208–216 (2009).

69. Marie, P. Physiology of bone tissue. *Immuno-analyse Biol. Spec.* **7**, 17–24 (1992).

70. Xiao, W., Wang, Y., Pacios, S., Li, S. & Graves, D. T. Cellular and Molecular Aspects of Bone. **18**, 9–16 (2016).

71. Everts, V. *et al.* The bone lining cell: Its role in cleaning Howship's lacunae and initiating bone formation. *J. Bone Miner. Res.* **17**, 77–90 (2002).

72. Robinson, P. N., Haendel, M. A., Liu, X. & Dai, M. Key Triggers of Osteoclast-Related Diseases and Available Strategies for Targeted Therapies : A Review. **4**, 1–10 (2017).

73. Xu Feng, J. M. M. Disorders of bone remodelling. 121–145 (2013) doi:10.1146/annurev-pathol-011110-130203.Disorders.

74. Janet, C., Lingling, X. & Xu, C. Role of TGF-β Signaling in Coupling Bone Remodeling. *Methods Mol. Biol.* **1344**, 183–191 (2016).

75. Kamiya, N. & Mishina, Y. New insights on the roles of BMP signaling in bone-A review of recent mouse genetic studies. *BioFactors* **37**, 75–82 (2011).

76. Nakashima, A., Katagiri, T. & Tamura, M. Cross-talk between Wnt and bone morphogenetic

protein 2 (BMP-2) signaling in differentiation pathway of C2C12 myoblasts. *J. Biol. Chem.* **280**, 37660–37668 (2005).

77. Choudhary, S. & Pilbeam, C. *Prostaglandins and bone metabolism*. *Principles of Bone Biology* (Elsevier Inc., 2019). doi:10.1016/B978-0-12-814841-9.00051-8.

78. Kelly Smith, J. Exercise, interleukins and bone homeostasis. *Integr. Mol. Med.* **3**, 802–804 (2016).

79. Cardoso, L. F., Maciel, L. M. Z. & de Paula, F. J. A. The multiple effects of thyroid disorders on bone and mineral metabolism. *Arq. Bras. Endocrinol. Metabol.* **58**, 452–463 (2014).

80. Wozney, J. M. *et al.* Novel regulators of bone formation: molecular clones and activities. *Science (80-.).* **242**, 1528 LP – 1534 (1988).

81. Wang, R. N. *et al.* Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* **1**, 87–105 (2014).

82. Jing, J., Hinton, R. J. & Feng, J. Q. *Bmpr1a Signaling in Cartilage Development and Endochondral Bone Formation*. *Bone Morphogenic Protein* (Elsevier Inc., 2015). doi:10.1016/bs.vh.2015.06.001.

83. Huntley, R., Jensen, E., Gopalakrishnan, R. & Mansky, K. C. Bone Reports Bone morphogenetic proteins : Their role in regulating osteoclast di ff erentiation. **10**, (2019).

84. Kokabu, S. *et al.* BMP3 Suppresses Osteoblast Differentiation of Bone. **26**, 87–94 (2012).

85. Luu, H. H. *et al.* Distinct Roles of Bone Morphogenetic Proteins in Osteogenic Differentiation of Mesenchymal Stem Cells. 665–677 (2007) doi:10.1002/jor.

86. Holtzhausen, A. *et al.* Novel bone morphogenetic protein signaling through Smad2 and Smad3 to regulate cancer progression and development. *FASEB J.* **28**, 1248–1267 (2014).

87. Nakato, H. & Li, J. P. Functions of Heparan Sulfate Proteoglycans in Development: Insights From Drosophila Models. *Int. Rev. Cell Mol. Biol.* **325**, 275–293 (2016).

88. Ramel, M. C. & Hill, C. S. Spatial regulation of BMP activity. *FEBS Lett.* **586**, 1929–1941 (2012).

89. Chen, G., Deng, C. & Li, Y. P. TGF-β and BMP signaling in osteoblast differentiation and bone formation. *Int. J. Biol. Sci.* **8**, 272–288 (2012).

90. Miyazono, K., Kamiya, Y. & Morikawa, M. Bone morphogenetic protein receptors and signal transduction. *J. Biochem.* **147**, 35–51 (2010).

91. Kamiya, N. *et al.* Targeted disruption of BMP signaling through type IA receptor (BMPR1A) in osteocyte suppresses SOST and RANKL, leading to dramatic increase in bone mass, bone mineral density and mechanical strength. *Bone* **91**, 53–63 (2016).

92. Gomez-Puerto, M. C., Iyengar, P. V., García de Vinuesa, A., ten Dijke, P. & Sanchez-Duffhues, G. Bone morphogenetic protein receptor signal transduction in human disease. *J. Pathol.* **247**, 9–20 (2019).

93. Cárcamo, J. *et al.* Type I receptors specify growth-inhibitory and transcriptional responses to transforming growth factor beta and activin. *Mol. Cell. Biol.* **14**, 3810–3821 (1994).

94. Massagué, J. TGFbeta signaling: receptors, transducers, and Mad proteins. *Cell* **85**, 947–950 (1996).

95. Hartung, A. *et al.* Different Routes of Bone Morphogenic Protein (BMP) Receptor Endocytosis Influence BMP Signaling. *Mol. Cell. Biol.* **26**, 7791–7805 (2006).

96. Xia, Y. *et al.* Repulsive Guidance Molecule RGMa Alters Utilization of Bone Morphogenetic Protein (BMP) Type II Receptors by BMP2. (2007) doi:10.1074/jbc.M701679200.

97. Halbrooks, P. J., Ding, R., Wozney, J. M. & Bain, G. Role of RGM coreceptors in bone morphogenetic protein signaling. **10**, 1–10 (2007).

98. Pan, H. *et al.* BmpR1A is a major type 1 BMP receptor for BMP-Smad signaling during skull development. **429**, 260–270 (2018).

99. Schulz, T. J. *et al.* Loss of BMP receptor type 1A in murine adipose tissue attenuates age-related onset of insulin resistance. *Diabetologia* 1769–1777 (2016) doi:10.1007/s00125-016-3990-8.

100. Manley, G. Public Access NIH Public Access. **71**, 233–236 (2013).

101. Increases, O. *et al.* J BMR Conditional Deletion of Bmpr1a in Differentiated Increasing Volume of Remodeling Bone in Mice. doi:10.1002/jbmr.477.

102. Chaudhury, A. & Hove, P. H. The Tale of Transforming Growth Factor-b (TGFb) Signaling: A Soigne Enigma. *Int. Union Biochem. Mol. Biol. Life* **61**, 929–939 (2010).

103. Lee, K. S., Hong, S. H. & Bae, S. C. Both the Smad and p38 MAPK pathways play a crucial role in Runx2 expression following induction by transforming growth factor-β and bone morphogenetic protein. *Oncogene* **21**, 7156–7163 (2002).

104. Liu, X. *et al.* Wnt signaling in bone formation and its therapeutic potential for bone diseases. *Ther. Adv. Musculoskelet. Dis.* **5**, 13–31 (2013).

105. Tian, Q., He, X. C., Hood, L. & Li, L. Bridging the BMP and Wnt pathways by PI3 kinase/Akt and 14-3-3ζ. *Cell Cycle* **4**, 215–216 (2005).

106. Kamiya, N. *et al.* Wnt inhibitors Dkk1 and sost are downstream targets of BMP signaling through the type IA receptor (BMPRIA) in osteoblasts. *J. Bone Miner. Res.* **25**, 200–210 (2010).

107. Felber, K., Elks, P. M., Lecca, M. & Roehl, H. H. Expression of osterix is regulated by FGF and Wnt/β-catenin signalling during osteoblast differentiation. *PLoS One* **10**, 1–17 (2015).

108. Kamiya, N. *et al.* BMP signaling negatively regulates bone mass through sclerostin by inhibiting the canonical Wnt pathway. *Development* **135**, 3801–3811 (2008).

109. Lee, K. W. *et al.* Rapamycin promotes the osteoblastic differentiation of human embryonic stem cells by blocking the mTOR pathway and stimulating the BMP/Smad pathway. *Stem Cells Dev.* **19**, 557–568 (2010).

110. Irelli, A. *et al*. mTOR links tumor immunity and bone metabolism: What are the clinical implications? *Int. J. Mol. Sci.* **20**, (2019).

111. Chen, J. & Long, F. MTORC1 signaling promotes osteoblast differentiation from preosteoblasts. *PLoS One* **10**, 1–10 (2015).

112. Huang, R. L., Yuan, Y., Tu, J., Zou, G. M. & Li, Q. Opposing TNF-/IL-1β- and BMP-2-activated MAPK signaling pathways converge on Runx2 to regulate BMP-2-induced

osteoblastic differentiation. *Cell Death Dis.* **5**, 1–11 (2014).

113. Wong, S. K. *et al.* The molecular mechanism of Vitamin E as a bone-protecting agent: A review on current evidence. *Int. J. Mol. Sci.* **20**, (2019).

114. Biswas, S. *et al.* BMPRIA is required for osteogenic differentiation and RANKL expression in adult bone marrow mesenchymal stromal cells. *Sci. Rep.* **8**, 1–14 (2018).

115. Komori, T. Runx2, a multifunctional transcription factor in skeletal development. *J. Cell. Biochem.* **87**, 1–8 (2002).

116. Tang, Z., Li, X., Tan, Y., Fan, H. & Zhang, X. The material and biological characteristics of osteoinductive calcium phosphate ceramics. *Regen. Biomater.* **5**, 43–59 (2018).

117. Nakashima, T. *et al.* Evidence for osteocyte regulation of bone homeostasis through RANKL expression. *Nat. Med.* **17**, 1231–1234 (2011).

118. Ikebuchi, Y. *et al.* Coupling of bone resorption and formation by RANKL reverse signalling. *Nature* **561**, 195–200 (2018).

119. Qin, Yang; Sylvie, D. Structural biology of poly(A) site definition. *PMC* (2012) doi:10.1038/jid.2014.371.

120. Chang, T. H. *et al.* An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* **14**, (2013).

121. Bakheet, T., Frevel, M., Williams, B. R. G., Greer, W. & Khabar, K. S. A. ARED: Human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. *Nucleic Acids Res.* **29**, 246–254 (2001).

122. Bengert, P. & Dandekar, T. A software tool-box for analysis of regulatory RNA elements. *Nucleic Acids Res.* **31**, 3441–3445 (2003).

123. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445 (1999).

124. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–

3066 (2002).

125. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

126. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* vol. 9 772 (2012).

127. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees . *Bioinformatics* **17**, 754–755 (2001).

128. Hartley, H. O. Maximum Likelihood Estimation from Incomplete. *Int. Biometric Soc.* **14**, 174–194 (1958).

129. Edler, D., Klein, J., Antonelli, A. & Silvestro, D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.* **12**, 373–377 (2021).

130. Efron, B., Halloran, E. & Holmes, S. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci.* **93**, 13429 LP – 13429 (1996).

131. Lin, H. H., Huang, L. F., Su, H. C. & Jeng, S. T. Effects of the multiple polyadenylation signal AAUAAA on mRNA 3́-end formation and gene expression. *Planta* **230**, 699–712 (2009).

132. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res.* **22**, 1173–1183 (2012).

133. Gendreau, K. L., Unruh, B. A., Zhou, C. & Kojima, S. Identification and characterization of transcripts regulated by circadian alternative polyadenylation in mouse liver. *G3 Genes, Genomes, Genet.* **8**, 3539–3548 (2018).

134. Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M. & Gautheret, D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* **10**, 1001–1010 (2000).

135. Mittleman, B. E. *et al.* Alternative polyadenylation mediates genetic regulation of gene expression. *Elife* **9**, 1–21 (2020).

136. Mariella, E., Marotta, F., Grassi, E., Gilotto, S. & Provero, P. The length of the expressed 3′ UTR is an intermediate molecular phenotype linking genetic variants to complex diseases.

Front. Genet. **10**, 1–20 (2019).

137. Chew, G. L., Pauli, A. & Schier, A. F. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.* **7**, 1–10 (2016).

138. Chu, Y. *et al.* An Upstream Open Reading Frame Represses Translation of Chicken PPARγ Transcript Variant 1. *Front. Genet.* **11**, 1–12 (2020).

3.6. Annexes

Anexo 1

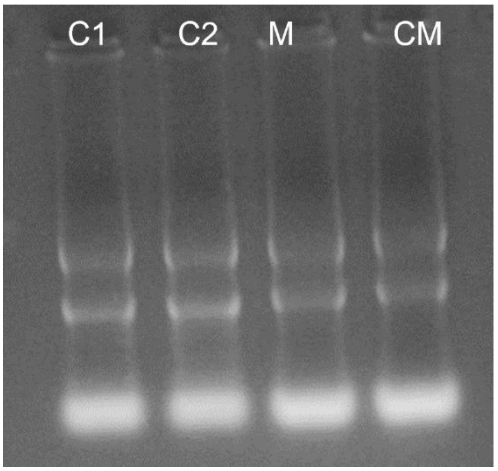

Figure 30- Evaluation of RNA Integrity. Electrophoresis gel of RNA extraction sample from MC3T3 cell to evaluate the integrity of the RNA. The samples were inserted on the wells by the following order, control 1, control 2, mineralization, and control of mineralization. The molecular marker used was 1kb from Invitrogen

Anexo 2

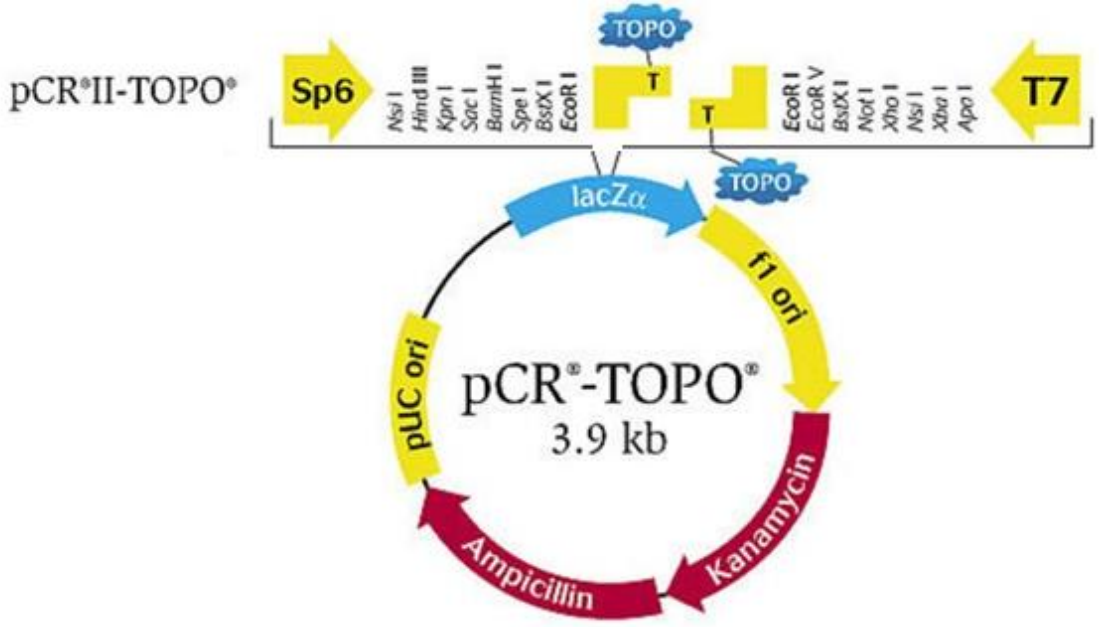

Figure 31-TOPO PCRII vector. Scheme with the restriction sites and with the SP6 phage promoter and T7 RNA Polymerase promoter in the flanking regions.

Anexo 3

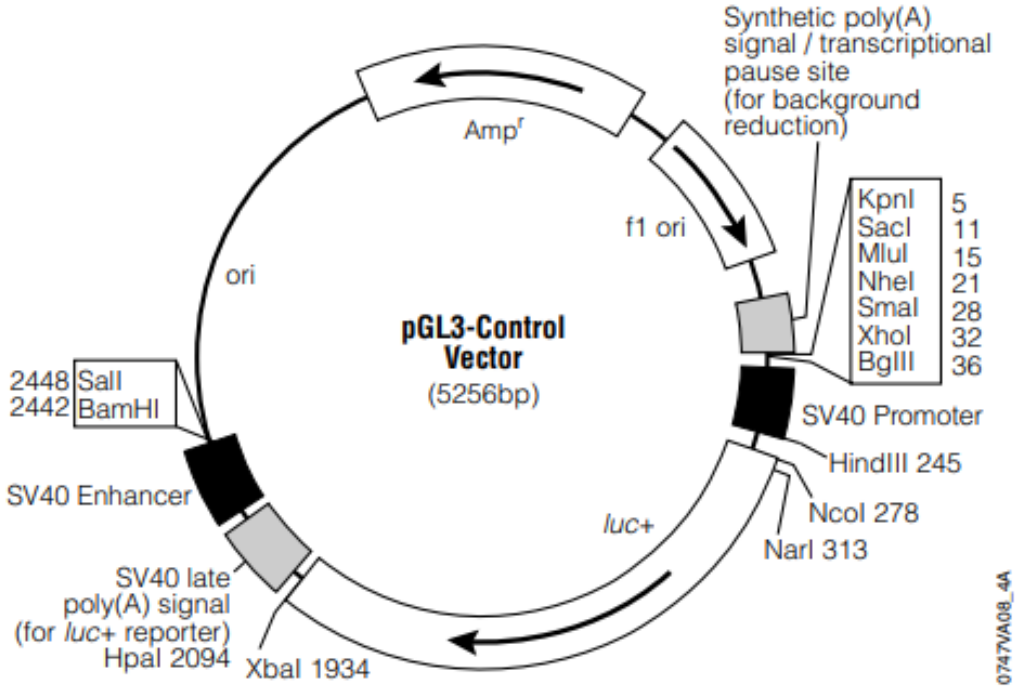

Figure 32- pGL3 Control vector. Scheme with the restriction sites, the SV40 promoter, the selection ORF (ampicillin), and the luciferase ORFs.

Anexo 4

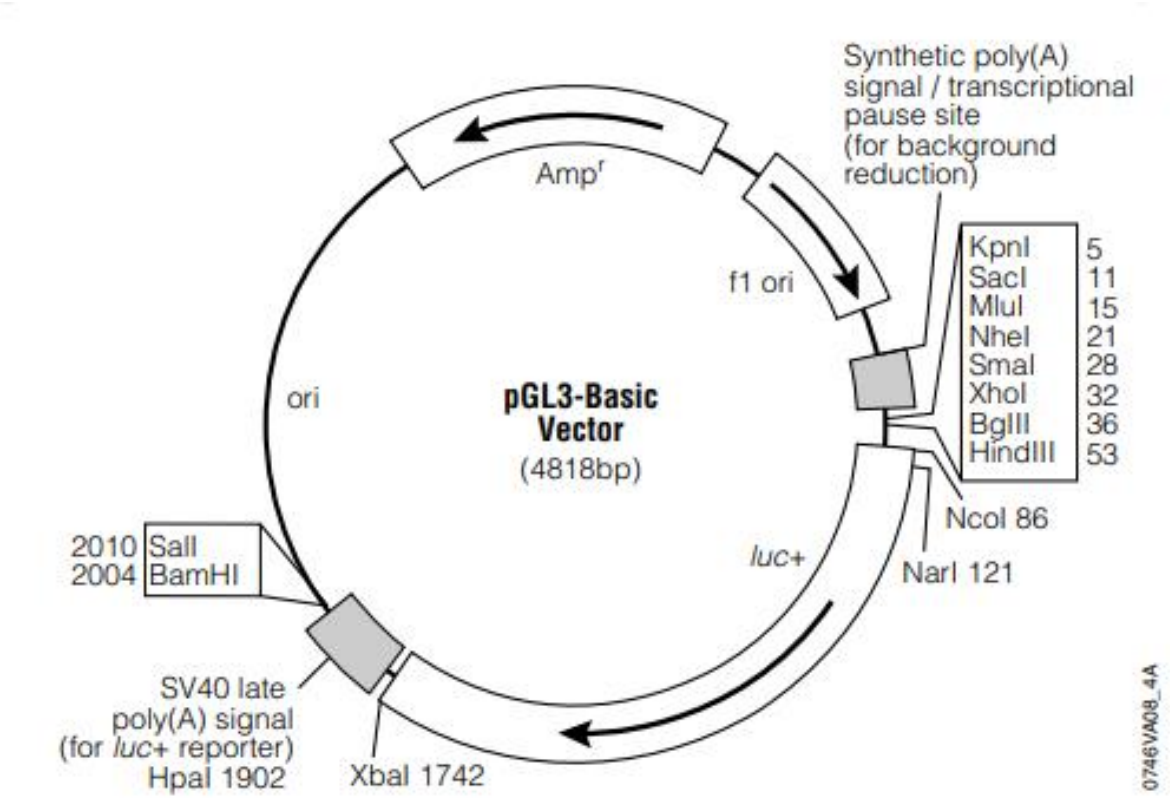

Figure 33- pGL3 Basic Vector. Scheme with the restriction sites, the selection ORF (ampicillin), and the luciferase ORFs

Anexo 5

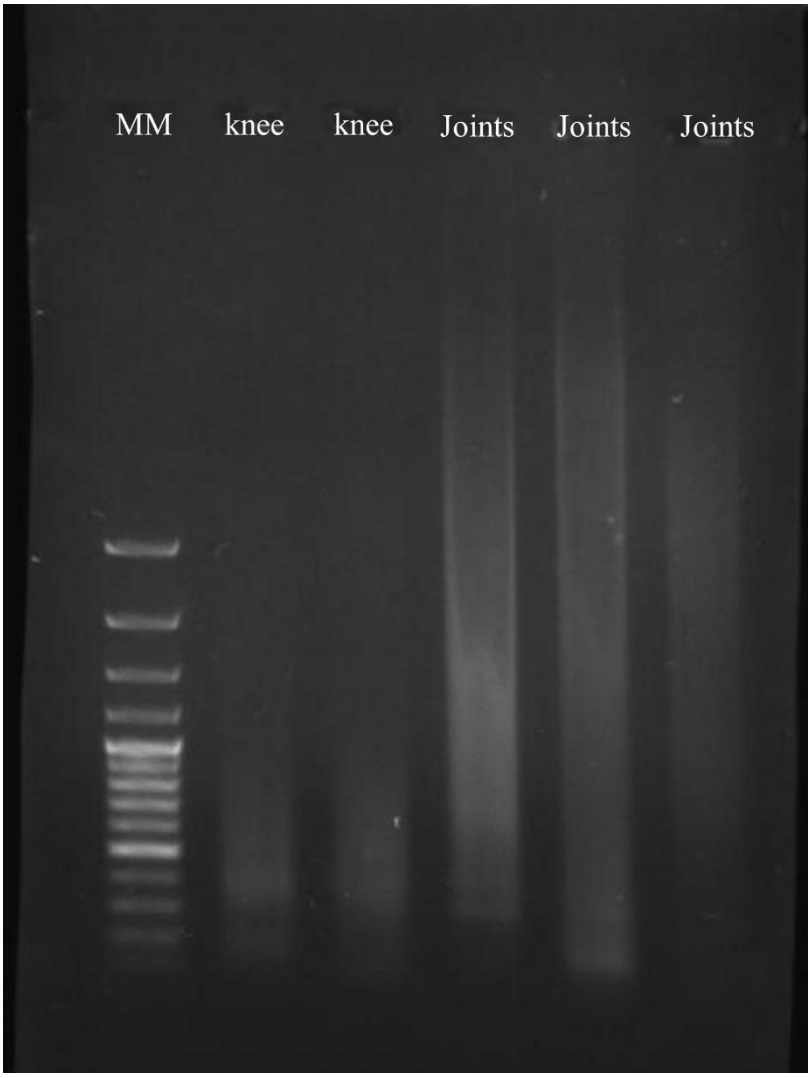

Figure 34- Amplification of 3'UTR from different mouse tissues. Electrophoresis agarose gel of the PCR amplification of 3'UTR of knee and joints tissues