Tatiana Marine da Silva Authier

Molecular and Functional Analysis of Mutations in the Transcription Factor ZNF687 Associated to Paget Disease of Bone

Universidade do Algarve

Faculdade de Ciências e Tecnologias

2018

Tatiana Marine da Silva Authier

Molecular and Functional Analysis of Mutations in the Transcription Factor ZNF687 Associated to Paget Disease of Bone

Master's Degree in Molecular and Microbial Biology

Under the supervision of:

Profª Drª Maria Leonor Cancela

Profª Drª Natércia Conceição

Universidade do Algarve

Faculdade de Ciências e Tecnologias

2018

Molecular and Functional Analysis of Mutations in the Transcription Factor ZNF687 Associated to Paget Disease of Bone

Declaração de autoria do trabalho

Declaro ser a autora deste trabalho, que é original e inédito.

Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

_____

(Tatiana Marine da Silva Authier)

ACKNOWLEDGEMENT

First, I would like to give my special thanks and appreciation to Professor Leonor Cancela and Professor Natércia Conceição for accepting me and giving me the opportunity to realize this amazing project. Thank you so much for your kindness, patience and understanding in difficult moments. But also, for the motivation and the courage that you gave me. For your availability, your time and your help. For answering all of my questions. For listening to me. For respecting me. I have learned and grown so much because of you. I just wanted to express my special gratitude and deep respect to Professor Leonor Cancela, whom I always admired. But also, to Professor Natércia Conceição for her concern and her heart, and for always supporting me. It really was a great honor for me to work with both of you.

I also would like to give a special thanks to Helena Caiado. For your kindness and help. Thank you for answering to all of my doubts, and for always being there for me when I needed. For stopping doing what you were doing in order to help me and answer my questions. For making me laugh in moment of despair. Thank you for being my colleague and my friend.

To Marco Tarasco for your kindness and for bringing so much joy to the lab. For helping me in my work, for all the shared laughers and funny moments. For making me feel better when I was grumpy and gloomy.

To Vincent Laizé for your kindness, help and orientation. For answering to all my crisis in the lab. For always being there when I needed. Thank you very much.

To Ana Alfaia. for your help when you saw me lost, for your kindness and for your friendship. Thank you for noticing when I was in difficulties and for reaching out to me.

Another special thanks to Bibiana Ferreira. For your big kindness and availability. For your share of knowledge. For helping me in my work. For always being there for me. For always answering my millions of questions. For all your help and generosity. For always treating me equally. For your simplicity and for always being so positive.

I also wanted to thank my parents for giving me the opportunity to pursue my studies. For giving me the opportunity to be someone, for giving me everything that I

have. For scarifying so much for me. For always believing in me, for loving me so much, for always being there for me, even when sometimes you are so far away. I wanted to specially thank my mother for everything. For always fighting for me and protecting me. For making me feel secure and loved. For giving me wings, for never letting me down. For taking so good care of me. For yelling at me when I feel depressed and pushing me to never give up and to always stay strong.

Thank you to my dear little ones. For always being there for me, for making me feel home and safe. For loving me for who I am. For comforting me when I am sad, for making me laugh, and for never letting me alone.

And finally, I would like to thank a certain person who is very special and important to me. Someone that I admire. Someone who always make me laugh and smile. Someone who share so much with me. Thank you for your help in my work, even if you think that you didn't do anything that deserved my appreciation. Thank you for being so kind to me.

ABSTRACT

The zinc finger protein ZNF687 is a transcription factor containing various Cys2-His2 zinc finger domains that has been related to Paget's disease of bone (PDB) associated with giant cell tumor of bone. Recently, four mutations have been independently identified in a Southern Italian population, and were determined as predisposal (p.Ser242Ile, p.Pro665Leu, and p.Gln784Glu,) or causal (p.Pro937Arg) for development of PDB. Moreover, it has been suggested that ZNF687 plays an important role in bone metabolism in both human and zebrafish, and that the function of this protein might be conserved throughout evolution. Nonetheless, the mechanism by which ZNF687 affects bone metabolism, and hence, PDB, remains to be elucidated. To contribute to respond to these questions we have proposed to perform *in silico* analysis in order to investigate how these four mutations could affect the protein conformation and function, and *in vitro* analyses to evaluate how mutant ZNF687 (p.Pro937Arg, mediated by site-directed mutagenesis), ZNF687 overexpression (mediated by transient transfection) and ZNF687 knock-down (mediated by CRISPR-Cas9), could affect the expression of genes involved in bone metabolism. We also assessed the mineralization process of knock-down clones and evaluated specific gene expression. Lastly, comparative *in silico* and *in vitro* analyses were performed in order to define the usefulness of zebrafish as a biological model for *ZNF687* study. Therefore, the mutation analysis suggested alterations in protein-protein and protein-acid nucleic interactions that might distort ZNF687 target genes expression and lead to PDB. Moreover, preliminary results suggested that ZNF687 might regulate genes involved in osteoblastogenesis and osteoclastogenesis, such as *RUNX2*, *OSX*, *RANK*, *SQSTM1* and *OPTN*, while its role in mineralization remained uncleared. Finally, by identifying significant similarities in genomic structure, protein domain, and molecular players affecting *znf687a*, *znf687b* and *ZNF687* transcription, together with the identification and characterization of promoter's regions that regulate the transcription of *znf687a* and human *ZNF687* genes, we have confirmed that zebrafish is a useful biological model system to study *ZNF687*.

Key words: *ZNF687*; Paget disease of bone; osteoblastogenesis; osteoclastogenesis; CRISPR-Cas9

RESUMO

A proteína dedo de zinco 687 (ZNF687) é um fator de transcrição que contem vários motivos Cis2-His2, sendo um dos motivos mais comuns de ligação ao DNA encontrados nos fatores de transcrições dos eucariotas. O gene *ZNF687* humano tem sido associado à várias doenças, nomeadamente à leucemia mielóide aguda, à doença óssea de Paget (DOP) severa associada ao tumor de células gigantes do osso, e ao carcinoma hepatocelular, advogando um putativo papel oncogénico. Além disso, foi demostrado que a expressão de RNA mensageiro do *ZNF687* era significativamente aumentada durante a osteoclastogénese e a osteoblastogénese, tanto no humano como no peixe zebra, sugerindo um papel importante no metabolismo do osso, e uma conservação quanto à função desta proteína ao longo da evolução, apesar da especiação. A doença óssea de Paget é um distúrbio metabólico crónico e raro, caracterizado por áreas focais com exagerada remodelação óssea. Com efeito, uma atividade anormal dos osteoclastos leva à reabsorção óssea excessiva, que é por sua vez, rapidamente compensada por uma hiperatividade dos osteoblastos, células responsáveis pela formação do osso. A doença pode ser monostótica ou poliostótica, mas em ambos os casos, a estrutura óssea resultante é desorganizada, desformada e frágil. A etiologia da DOP não é bem conhecida, mas um fator genético está fortemente associado à doença, existindo também um fator ambiental. Foram identificadas várias mutações em diferentes genes, nomeadamente no sequestossoma 1 (*SQSTM 1*), fator estimulante de colónia 1 (*CSF-1*), membro da superfamília do receptor do fator de necrose tumoral 11a (*TNFRSF11A*), superfamília 7 transmembrana 4 (*TM7SF4*), optineurina (*OPTN*), ras e intercatores rab 3 (*RIN3*), proteína contendo valosina (*VCP*), nucleoporina 205 (*NUP205*) e mais recentemente, *ZNF687* e ribonucleoproteína nuclear heterogênea A2 / B1 (*hnRNPA2B1*).

Recentemente, foram descritas, em estudos independentes, numa população do sul da Itália, quatro mutações associadas a DOP: três delas foram identificadas como não causais (p.Ser242Ile, p.Pro665Leu, e p.Gln784Glu) e uma delas (p.Pro937Arg) foi determinada como necessária e suficiente para desenvolver a DOP.

O mecanismo pelo qual a ZNF687 afeta o metabolismo ósseo e deste modo pode causar a DOP, ainda não foi esclarecido. Para responder a estas questões, propusemos fornecer dados, realizando vários estudos *in silico* e *in vitro*. Deste modo, para estudar o efeito das mutações descritas na proteína ZNF687, efetuaram-se análises de comparação

da estrutura da proteína normal *versus* a estrutura da proteína com as diferentes mutações. As mutações associadas a DOP não mostraram nenhuma alteração na estrutura secundária da proteína, no entanto, as diferenças nas propriedades entre os resíduos "wild-type" (WT) e os resíduos mutados podem levar às alterações, tanto na conformação local, como na interação de proteína-proteína ou de proteína-ácido nucleico. Contrariamente às mutações p.Ser242Ile e p.Pro665Leu, que não se encontram num domínio funcional, a mutação p.Glu784Gln ocorre na extremidade C-terminal de um domínio dedo de zinco (ZF), podendo afetar a interação aminoácido-DNA especifica deste local. A mutação p.Pro937Arg gera uma carga positiva adicional antes do local de sinalização nuclear (NLS). O resíduo mutado acaba por fazer parte deste NLS, tornando-o mais forte, o que pode levar a um aumento da translocação da proteína para o núcleo.

Após estes resultados preliminares, fomos avaliar a expressão de genes putativos alvos da ZNF687, que estão descritos como estando envolvidos no metabolismo do osso. Para isto, a região codificante do gene, inserida num vetor de expressão, foi submetida ao processo de mutagénese dirigida para se reproduzir a mutação p.Pro937Arg. Esta construção foi posteriormente utilizada na transfeção de células SaOS-2, em paralelo com o vetor contendo a forma normal, para induzir a sobre-expressão da ZNF687 mutada e normal. Outra estratégia do nosso trabalho consistiu na obtenção de clones de células SaOS-2, em que efetuámos a repressão do gene *ZNF687* através da ferramenta de edição génica, CRIPSR-Cas9. Tanto nas células submetidas a sobre-expressão como nos clones "knock-down", o RNA foi extraído e a expressão dos níveis de mRNA de vários genes de interesse foi analisada por PCR quantitativo. Em seguida, estudámos também o processo de mineralização nos clones "knock-down" e avaliámos a expressão dos genes alvos da ZNF687. Os nossos dados preliminares sugeriram que a ZNF687 poderia desempenhar um papel regulador de *RUNX2* e *OSX*, dois genes envolvidos na osteoblastogénese, e de *RANK*, *SQSTM1* e *OPTN*, genes envolvidos na osteoclastogénese através da via de sinalização do *NF-κB* induzida por RANKL. No entanto, não foi possível determinar se a regulação do *SQSTM1* e de *OPTN* pela ZNF687 dá-se de uma forma direta ou se é feita através da regulação de *RANK*. O mesmo pode acontecer no caso da regulação do *OSX*, podendo este ser diretamente regulado pelo RUNX2 e indiretamente pela ZNF687. O papel da ZNF687 no processo de mineralização também permanece incerto sendo que a diminuição dos níveis de expressão da osteocalcina (*OCN*) e da atividade da fosfatase alcalina (ALP), observados nos clones "knock-down", poderia

ser uma consequência da diminuição dos níveis de *RUNX2*, diminuição possivelmente causada pelo "knock-down" da ZNF687. Por outro lado, a expressão em células que sobre-expressam a forma mutada de ZNF687 demonstraram um padrão de expressão similar ao das células que sobre-expressam a forma WT. Anteriormente, foi descrito que a mutação p.Pro937Arg, mutação causal encontrada em pacientes com DOP, parece atuar como um ganho de função levando a um aumento da translocação de ZNF687 para o núcleo, onde a sua acumulação aumentará a expressão de genes alvos. Portanto, a alteração da expressão dos genes *RUNX2*, *OSX*, *RANK*, *OPTN* e *SQSTM1*, envolvidos na diferenciação dos osteoblastos e dos osteoclastos, poderá ser a causa do desenvolvimento da DOP em pacientes que apresentam essa mutação. Porém, após análises comparativas entre os genes *znf687a* e *znf687b* do peixe zebra com o gene *ZNF687*, foi possível sugerir que o gene *znf687b* era ortólogo do gene *ZNF687* humano sendo que é aquele que apresentava maiores semelhanças na estrutura genómica, domínio proteico e fatores moleculares afetando a transcrição *znf687a*, *znf687b* e *ZNF687*. Finalmente, através da análise funcional da região promotora dos genes do peixe zebra e do humano identificou-se uma região no *znf687a* do zebrafish e duas regiões no *ZNF687* humano como sendo responsáveis pela sua regulação. Além disso, o ensaio de co-transfecção demonstrou um efeito de repressão da transcrição da construção contendo o promotor do gene *zbf687a* devido ao fator de transcrição YY-1. Nos ensaios efetuados com a construção contendo o promotor do gene *ZNF687* observou-se uma repressão da sua transcrição devido ao fator de transcrição AP-1. Estes resultados parecem sugerir que os promotores do gene *ZNF687*, tanto em humano como em peixe zebra, são regulados negativamente por factores de transcrição, que regulam genes envolvidos no desenvolvimento do osso. Estes resultados juntamente com a nossa análise *in silico* comparativa do *locus* genómico, do gene e da proteína entre as duas espécies permite-nos sugerir a potencialidade de utilizar o peixe-zebra como modelo biológico para o estudo da função da ZNF687.

Palavras chaves: ZNF687; Doença ossea de Paget; osteoblastogenese; osteoclastogenese; CRISPR-Cas9

ABBREVIATIONS AND ACRONYMS

AR-S – Alizarin Red S

BSA – Bovine Serum Albumin

Cas9 – CRISPR associated protein 9 nickase

CDS – Coding DNA Sequence

CIAP – Calf Intestinal Alkaline Phosphatase

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

DNA – Deoxyribo Nucleic Acid

dNTPs – Deoxyribonucleotide Triphosphate

DTT - Dithiothreitol

E.coli – Escherichia coli

ECM – Extracellular Matrix

F1 – Fragment 1

F2 – Fragment 2

F3 – Fragment 3

F4 – Fragment 4

FA - Formarldehyde

Fwd – Forward

FBS – Foetal Bovine Serum

IPTG – Isopropyl β-D-1-thiogalactopyranoside

KO – Knockout

LB – Luria-Bertani

M-CSF – Macrophage-Colony Stimulating Factor

NCBI – National Center for Biotechnology Information

NLS – Nuclear Localization Signal

OD – Optical Density

OCN – Osteocalcin

OCP – Osteoclast Precursor

OSX – Osterix

P/S – Penicillin/Streptomycin

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PMA - Phorbol 12-myristate 13-acetate

p-NPP – p-Nitrophenyl phosphate

RANKL - Receptor Activator of NFKB Ligand

Rev – Reverse

RQ1 – RNA Qualified

RPMI 1640 – Roswell Park Memorial Institute 1640 Medium

RNA – Ribo Nucleic Acid

RUNX2 – Runt-related transcription factor 2

RANK – Receptor activator of nuclear factor

SQSTM1 – Sequestosome 1

S. pyogenes – Streptococcus pyogenes

sgRNA – single-guide Ribo Nucleic Acid

SOC – Super Optimal Broth with Catabolite Repressor

TAE – Tris-Acetete-EDTA

X-Gal – 5-bromo-4-cloro-3-indoxil-β-D-galactopiranoside

XTT - 2,3-Bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide salt

UV – Ultra violet

ZNF687 – Zinc Finger 687

LIST OF FIGURES

LIST OF TABLES

INDEX

CHAPTER I – INTRODUCTION

1.1 Normal Bone anatomy, physiology and metabolism

Bone, the basic component of the human skeletal tissue, is a strong, semi-rigid, dynamic and heterogeneous organ. The adult human skeleton is composed of 206 bones and organized in two systems (Figure 1.1): the axial skeleton (skull, laryngeal skeleton, vertebral column, and thoracic cage) and the appendicular skeleton (limbs and pelvis).[1] Together, these structures provide the framework for our body, protecting, supporting and housing our internal organs, such as brain, spinal cord, and hematopoietic tissue, but also serve as attachment for muscles, tendons and ligaments allowing mechanical movements and locomotion.[1-3]

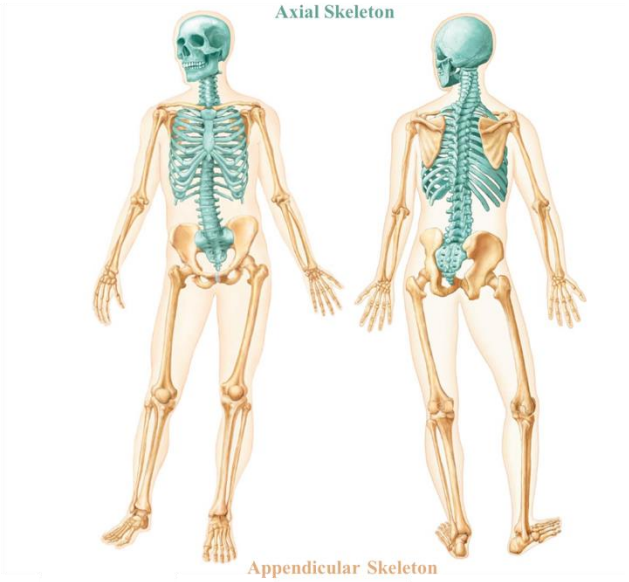


Figure 1.1 – The human skeleton systems. The human skeleton is divided in two systems: the axial skeleton (Caribbean blue), comprising the skull, hyoid bone, vertebral spine and rib cage; and the appendicular skeleton(yellow), including the limbs, shoulder gridle and pelvic girdle. Adapted from Marieb, Human Anatomy, 7th edition (2014).[4]

Despite its structural function, the skeletal tissue also performs metabolic and physiologic functions, being involved in mineral homeostasis, bone resorption and formation, mainly orchestrated by its own components.[1-3,5]

1.1.1 Bone composition

Bone is a complex organ composed of different types of cells, such as osteoblasts, osteoclasts, osteocytes and bone-lining cells, but is also constituted by a non-cellular component named extracellular matrix (ECM) (Figure 1.2).

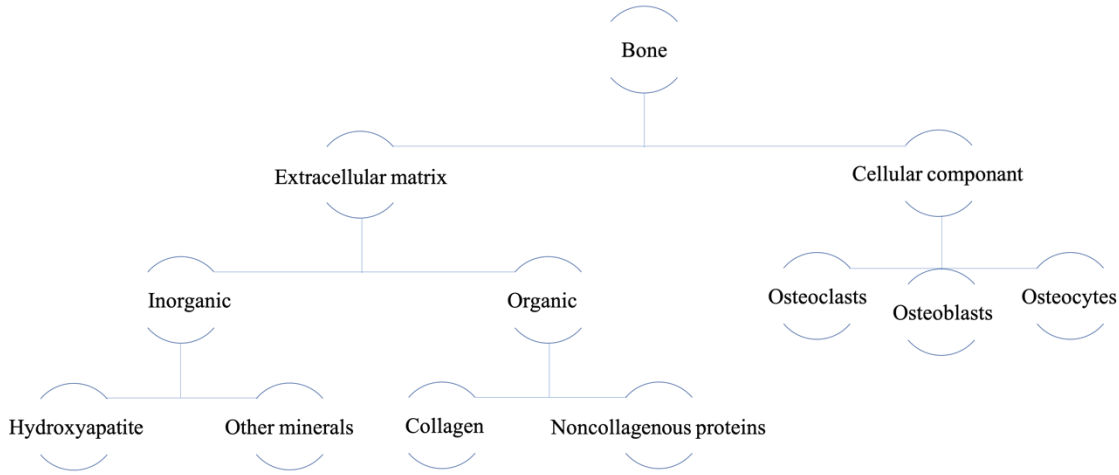


Figure 1.2 – Composition of bone. Bone is a complex organ composed of a cellular compartment, and mineral and organic phases.

1.1.1.1 Osteoblasts

The osteoblasts are bone-forming cells involved in the maintenance of bone architecture and located on the bone surface. They synthesize the ECM by secreting a complex mixture of bone matrix proteins and minerals and are also implicated in the regulation of osteoclasts through the NF- κB pathway.[6] Osteoblast progenitors (OBP) can arise from two different ways (Figure 1.3): directly derived from mesenchymal stem cells, where mature osteoblasts will direct squamous bone formation (craniofacial bones and clavicle), in a process called intramembranous ossification, or stem through an osteo/chondroprogenitor, where these mature osteoblasts will be responsible for the endochondral ossification of the rest of the skeleton.[6-9] The differentiation of an osteoblast is a multistep process and is regulated by several proteins, namely runt-related transcription factor 2 (RUNX2), osterix (OSX), distal-less homeobox 5 (DLX5), special AT-rich binding 2 (STAB2), β-catenin, sex determining region Y Box-9 (SOX-9), and bone morphogenic proteins (BMP) among others.[7-10] Once a certain amount of OBP expressing collagen I and RUNX2, a master gene involved in early stages of osteoblastogenesis, has been established, the multistep process of osteoblastogenesis changes to a proliferation phase. At this stage, OBP become pre-osteoblasts that gain alkaline phosphatase (ALP) and bone matrix protein secretion activities and suffer morphological changes by becoming larger and cuboidal.[10] A higher expression of bone matrix proteins such as osteocalcin (OCN), bone salioprotein (BSP) I and II and type I collagen illustrate the late stage of osteoblastogenesis where pre-osteoblasts turn into mature osteoblasts. These mature mononucleated osteoblasts form a single row of cells connected between them by tight-junctions and present an abundant rough endoplasmic

reticulum and a large Golgi apparatus that mirrors the protein synthesis and secretory activities of these cells.[2,6-7,10-11] Finally, mature osteoblasts undergo different pathways: they can either suffer apoptosis, differentiate in osteocytes or turn into bone-lining cells.

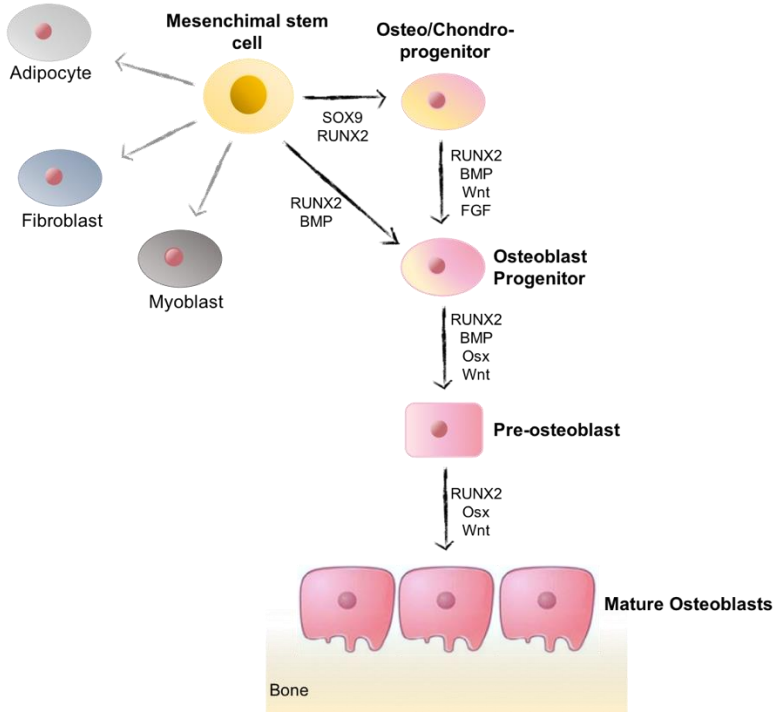


Figure 1.3 – Schematic representation of osteoblastogenesis. Mesenchymal stem cells can differentiate directly into an osteoblast progenitor, under runt-related transcription factor 2 (RUNX2) and bone morphogenic proteins (BMP) expression, or into an osteo/chondroprogenitor, under sex determining region y box-9 (SOX9) and RUNX2 expression. The latter will then suffer additional differentiation and stem as an osteoblast progenitor. Adapted from Arboleya *et al.* (2014).[12]

1.1.1.2 Osteocytes and bone-lining cells

Osteocytes are the most numerous cells in the bone, presenting a large star-shaped morphology, and are located inside the bone, completely embedded in the mineralized bone matrix, in a lacunocanicular system. They are tightly connected to each other by gap-junctions and communicate with bone surface cells through dendritic processes that traverse the bone.[10,13-14] Despite being able to sense mechanical signal and translate it into biochemical signal, osteocytes are also known to control and regulate the differentiation and activity of both osteoblasts and osteoclasts.

Bone-lining cells, on the other hand, are phenotypically flat shaped inactive osteoblasts covering the surface of the bone in order to prevent osteoclast to interact with the bone matrix during bone resting phase.[10] These cells are able to proliferate and are a nutritional and metabolic support for osteocytes.

1.1.1.3 Osteoclasts

On the other hand, osteoclasts are responsible for the degradation of the organic and inorganic matrices of bone, and are displayed in Howship lacunae, i.e small depressions located on the bone surface. Osteoclastogenesis occurs in various steps, including determination, proliferation, differentiation, multinucleation, polarization and construction of osteoclasts-specific structures (Figure 1.4A). Briefly, the osteoclast precursors (OCP), derived from myeloid lineage, proliferate and circulate into the blood stream.[15-17] When required, the OPC are recruited to the remodelling site of the bone, where a suitable environment allows them to undergo further differentiation and fusion in order to arise as multinucleated osteoclasts.[13,15-16,18-19] The differentiation of osteoclast is insured by two main cytokines: the macrophage-colony stimulating factor (M-CSF) and the receptor activator of NF-κB ligand (RANKL). The binding of these two cytokines to their respective receptors, M-CSFR (or c-fms, i.e. colony-stimulating factor 1 receptor) and RANK, located in the cell-surface of OPC, promote the activation of several transcription factors involved in osteoclastogenesis, such as PU.1, c-Fos, c-Jun, NF-κB and nuclear factor of activated T cells calcineurin-dependent 1 (NFATc1) (Figure 1.3A).[13,15-19] Expression of RANKL can also be induced by osteotropic factors such as 1,25-$(OH)_2D_3$ (vitamin D), IL-1, IL-11, and parathyroid hormone (PTH), promoting osteoclast formation. Furthermore, osteoblastic cells also regulate osteoclastogenesis, positively, by expressing M-CSF and RANKL in response to osteoclastogenic factors, and negatively, by secreting osteoprotegerin (OPG), in response to other factors, that will interact with RANKL and prevent it from binding to its receptor RANK (Figure 1.4B).[13] The multinucleation process allows coverage of a larger area in contact with the bone, and multinucleated cells contain 5 to 20 nuclei. These fused cells are then re-organized and polarized, forming a new cytoskeletal construction, namely ruffle borders and a sealing zone. The ruffle borders are responsible for the secretion of acid and protons, whereas the sealing zone maintains the acid environment necessary for the degradation of protein and mineral component of the bone.[13,15,18]

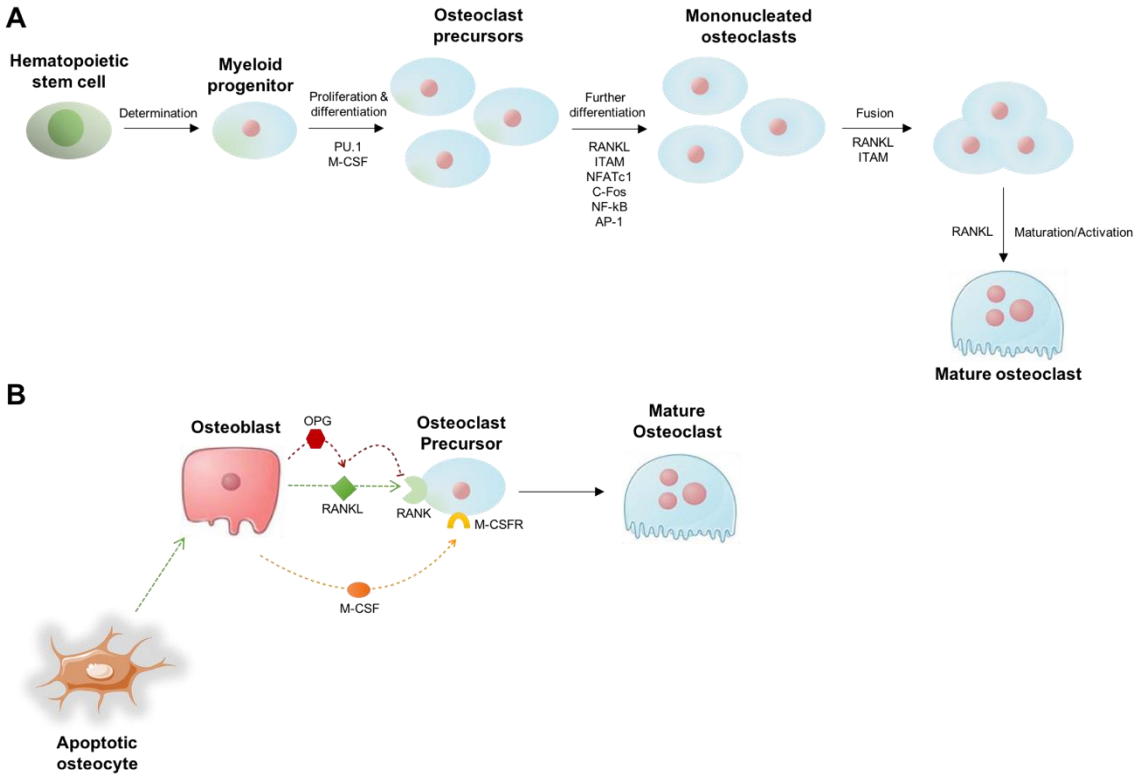


Figure 1.4 – Differentiation of osteoclasts. (A) PU.1 induces the proliferation of myeloid progenitors that undergo differentiation in order to arise as osteoclast precursors (OPC) under activation of macrophage-colony stimulating factor (M-CSF). The ligation of the receptor activator of NF-κB ligand (RANKL) to its receptor RANK, together with the activation and upregulation of several transcription factors and signalling pathway (NFTAC-1, AP-1, NF-κB etc), promote further differentiation of OPC. After, newly formed mononucleated osteoclasts fuse together and arise as multinuclear osteoclasts that will be polarized and re-organized. These osteoclasts will suffer final structural alterations and turn into mature osteoclasts. (B) Osteoclast differentiation is also regulated by osteoblast and osteocytes. The osteoblasts express M-CSF under osteoclastogenic factors that will bind to its respective receptor located on OPC surface and stimulates osteoclastogenesis. M-CSFR signaling promotes the expression of RANK on the cell surface and upon ligation of RANKL, secreted by osteoblast, generate further differentiation, fusion and maturation of osteoclasts. Osteoblasts also secrete osteoprotegerin (OPG) in response to other factors that act as an inhibitor of RANKL/RANK ligation. During bone resorption, apoptotic osteocytes secrete biochemical signals to osteoblast, instructing them to generate RANKL to enable osteoclastogenesis. Adapted from Zhao et al. (2011), and Boyce et al. (2012).[18,20]

1.1.1.4 Extracellular matrix

The ECM of bone is a dynamic network that serves as a scaffold for mineral deposition and plays an important role in regulating bone cell function like bone remodeling.[21-22] Its composition varies with its anatomical location, and with the age, health and diet of a person. Commonly, it is approximatively composed of 30 % of an organic matter, 60 % of a mineral matter, 7 % of water and less than 3 % of lipids.[2,21,23-24]

Most of the organic matrix is constituted by type 1 collagen that is synthesized by osteoblasts and deposited in layers, forming the osteoid. The specific orientation of this protein allows bone tissue to hold a very high density of collagenous filaments and

provide tensile strength to the bone. But the disorganization of the matrix in mature bone, called woven bone, leads to pathologies such as Paget disease of bone (PDB).[2] Indeed, woven bones are weaker and more flexible as they present a smaller amount of disorganized collagen fibres that were formed quickly by osteoblasts. The organic matter is also composed of growth factors and other non-collagenous proteins, namely osteopontin, osteonectin and osteocalcin, the latter being the most abundant. Synthesized by osteoblasts, this protein is involved in bone mineralization and formation. The mineral part of the ECM is mostly constituted by phosphate and calcium, forming small crystals, called hydroxyapatite, that provide rigidity and compressive strength to the bone.[2] These flat-shaped crystals possess some impurities such as carbonate, potassium, magnesium, and sodium among others, that are believed to play a role in mineral homeostasis.[2,25]

1.1.2 Bone remodelling

During life, bones suffer modelling (construction) to respond and adapt to physiological influences and mechanical forces, but also undergo constant remodelling (reconstruction) to maintain bone strength, mineral homeostasis and repair bone damages. Bone remodelling is a complex sequence of bone resorption and formation, orchestrated by osteoclast and osteoblasts respectively, but also by osteocytes (Figure 1.5).[1-3,5,23] The first step in bone remodelling is bone resorption. Osteocytes can sense bone damage or deformation and send biochemical signals to bone cells to inform them for the need of remodelling. Therefore, deforming or dying osteocytes emit signals to osteoclasts instructing them to form, where to go and how much bone is to be resorb.[26] After binding to the inculcated area, osteoclasts proceed to resorption by creating an acidic environment that will degrade the mineral part of the bone and expose the organic matter. The latter will then be degraded by proteases, secreted by osteoclasts. [2,13,22,26] Once the determined amount of bone has been resorbed, a reversal phase occurs where osteoclasts enter in apoptosis after receiving signals, most probably from osteocytes and osteoblasts. The surface of the resorbed bone is cleaned up by mononuclear cells, providing space so osteoblasts can initiate bone formation and mineralization. [2,13,26] Osteoblasts start to reconstruct the bone by producing the unmineralized organic compound, named osteoid, and then by laying down mineral that begins to crystalize around the collagen scaffold and form hydroxyapatite. In the end, the bone surface is covered with bone-lining cells and the resting period begins and lasts until a new remodelling cycle is triggered.

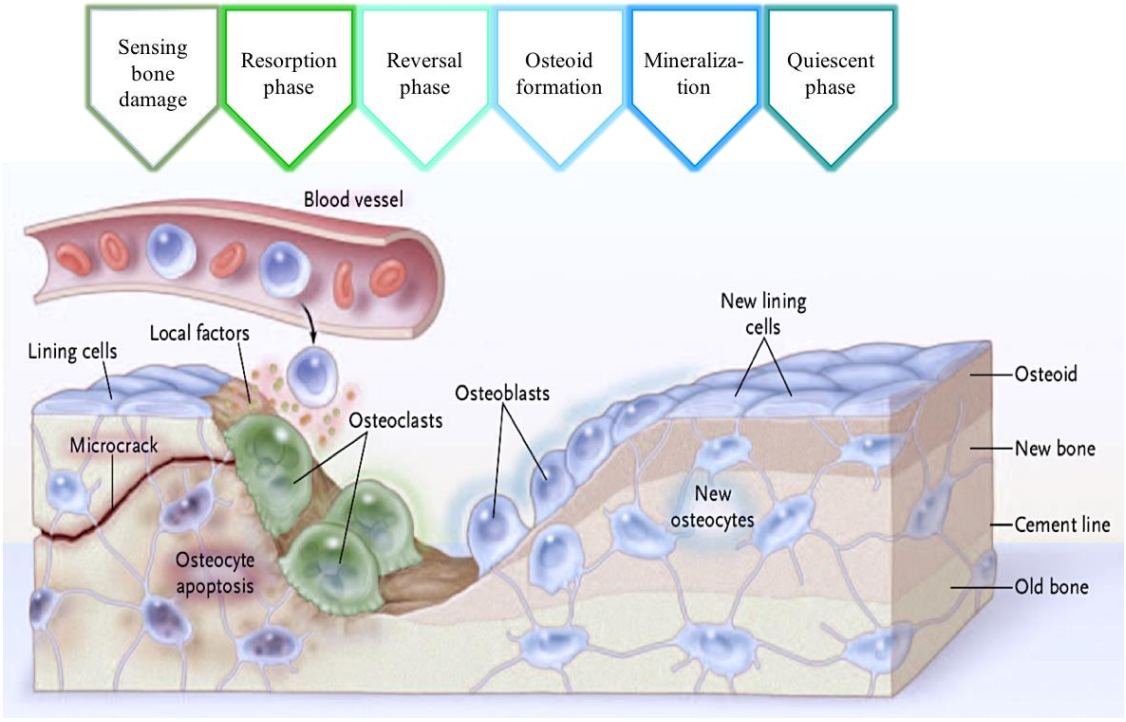


Figure 1.5 - Bone remodeling process. Pre-osteoclasts differentiate into osteoclasts under the presence of specific cytokines and growth factors, that are emitted by apoptotic osteocytes. Mature osteoclasts start bone resorption and degrade old or damaged bone. At the end of resorption, osteoclasts enter in a reversal phase and osteoblast start bone formation, followed by mineralization. In the end of the process, osteoblast differentiate into bone lining cells, and a resting period begins. [27]

In mature healthy bone, a balance between a bone resorption and bone formation is strictly maintained to ensure that there is no alteration in bone mass and homeostasis after each remodelling cycle. However, in some cases, an imbalance between these two processes may occur and lead to abnormal bone remodelling and the development of bone disorders, such as osteoporosis, osteopetrosis and Paget's disease of bone.

1.2. Paget Disease of Bone

1.2.1 Pathology

Paget's disease of bone (PDB) is a rare chronic and metabolic bone disorder characterized by focal areas of increased bone turnover. First, a lytic phase characterized by an abnormal excessive osteoclastic bone resorption activity is observed, and later followed by a mixed phase, where a rapid compensative activity of osteoblasts promotes bone reconstruction. In this second part, the activity of osteoblast is predominant, and bone is formed randomly and abnormally, resulting in a disorganized, enlarged and fragile structure. In the last phase, named sclerotic phase, the osteoblastic and osteoclastic activities diminish, and the pagetic remodelling process, that has accumulated several

abnormalities, may return to normal (Figure 1.6).[28-29] In pagetic bone lesions, osteoclasts are increased in number and size, presenting more nuclei than normal (up to 150), while osteoblasts remain structurally normal.[28,30-33]

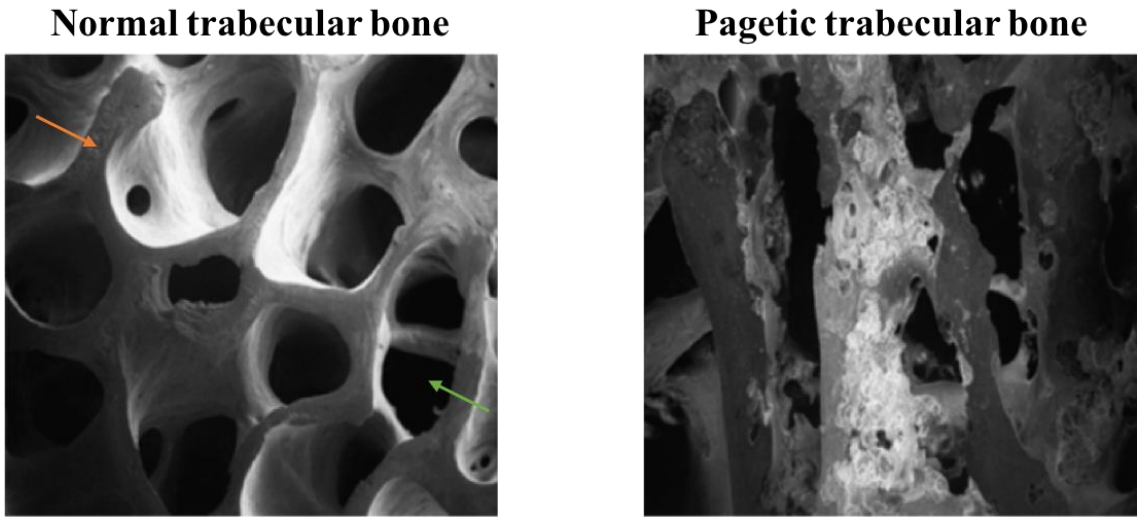


Figure 1.6 – Scanning electron micrographs of normal and pagetic trabecular bone of the iliac crest. Normal trabecular bone presents a normal microarchitecture composed with well-defined trabeculae (orange arrow) and pores (green arrow) that are filled with bone marrow. This spatial complexity provide strength to the bone. In contrast, the pagetic trabecular bone has lost its microarchitectural appearance. Alteration in bone structure, shape and density leads to changes in mechanical properties and increases bone susceptibility to fractures adapted from e. S. Siris et al. American society for bone and mineral research (2006).[34]

1.2.2 Clinical features

The disease may be monostotic or polyostotic, but in either case, the resulting bone structure presents a mixture of woven and lamellar bone called mosaic bone. The pelvis, femur and tibia are the most affected parts of the appendicular skeleton, whereas in the axial skeleton, the lumbar spine and skull are the most affected zones (Figures 1.7 and 1.8).

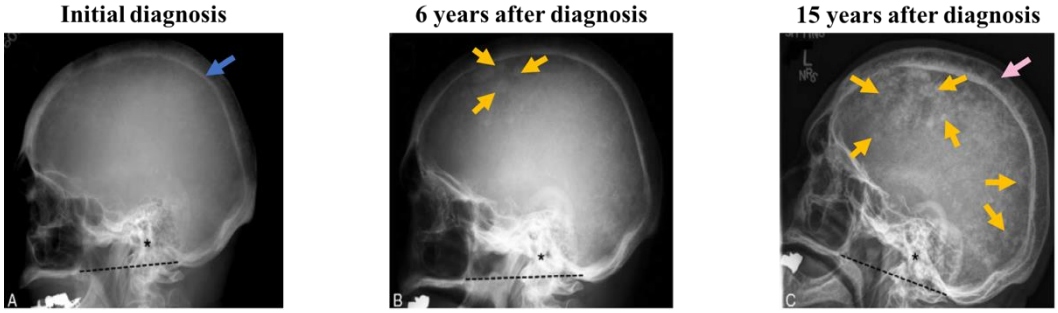


Figure 1.7 – lateral radiographs of a pdb patient skull showing evolution of the disease. (A) At initial diagnosis, the calvarium is thickened (blue arrow). (B) Six years after diagnosis, multiple patches of sclerotic bones appear (yellow arrows) (C) and this cotton wool appearance has expended throughout the skull 15 years after diagnosis. Expansion of the calvarium space can be observed (pink arrow) as well as cranio-cervical junction abnormality (asterisk). Dote line represent the Mc Gregor line. Adapted from Deep *et al.* American otological society, american neurotology society [and] european academy of otology and neurotology (2017).[35]

While some patients are asymptomatic, others experience a variety of symptoms

such as bone pain, bone deformation and fractures.[31] Complications of PDB are dependent on the affected site and common features include osteoarthritis, deafness, tinnitus and neurological defects.[31,35-36] In approximately 1% of cases, Pagetic tissue might also turn into osteosarcoma or, even more rarely, into giant cell tumour of bone (GCT).[37-39]

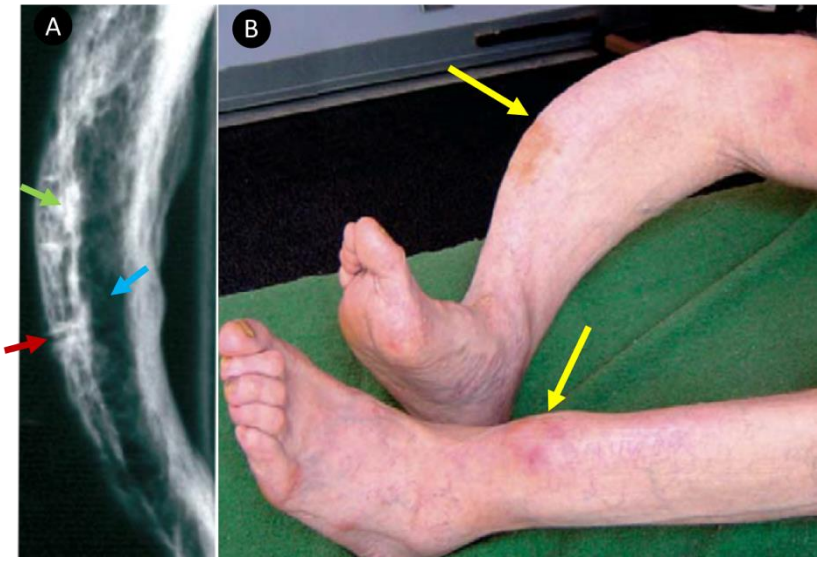


Figure 1.8 – Extension and deformation of both tibias of a PDB patient. (A) The radiograph of the right tibia shows expansion and deformity, resulting in a bowing bone. Alternating areas of osteolysis (blue arrow) and osteosclerosis (green arrow) can be observed, as well as a pseudo-fracture in the convex part of the bone (red arrow). (B) Picture showing the severe deformation of the right and left tibias (yellow arrows). Adapted from Lacet seminar 372 (2008).[31]

Asymptomatic PDB patients are often incidentally diagnosed through x-ray radiography when they are subjected to an investigation for another disease. The measurement of biochemical markers of bone turnover, such as alkaline phosphatase, which is the most widely used, is very important in the diagnosis of PDB.[28,30-33] Indeed, high level of ALP in the serum is present in 85% of patients and is correlated to the disease activity and proportion of affected bones.[28] The diagnose is confirmed by radiography, as well as nuclear scintigraphy of the affected site that proves continuing metabolic activity. Computerized tomography (CT) scans and magnetic resonance imaging (MRI) can provide additional information such as small fractures and malignant complications.

1.2.2 Epidemiology

PDB is a late-onset disease, affecting slightly more men than women, with a 3:2 (male:female) ratio, and its prevalence increases with age for both sexes. The disease is predominant in Caucasian population, mainly of European descent, and rare in Asian and

African populations.[30] This might suggest a genetic condition with the hypothesis of the existence of founder mutations, originating in northwestern Europe, that spread among different countries and continents through emigration.[39] Nonetheless, United Kingdom possesses the highest prevalence in the world.

1.2.3 Aetiology

The aetiology of PDB is not well understood, but genetic factors are strongly associated to PDB, and environmental factors have also been suggested as contributors.

The disease appears to be genetically heterogeneous and presents mutations in one or more genes. Different loci and genes located in chromosome 1, 5-10, 14, 18 and 20 have been recognized as possible candidates for PDB. Among them, mutations have been identified in *sequestosome 1 (SQSTM)*, *colony stimulating factor 1 (CSF1)*, *tumor necrosis factor receptor superfamily member 11a (TNFRSF11A or RANK), transmembrane 7 superfamily 4 (TM7SF4), optineurin (OPTN)*, *ras and rab intercator 3 (RIN3)*, *valosin containing protein (VCP), nucleoporin 205 (NUP205)* and more recently, the *zinc finger protein 687 (ZNF687)* and *heterogenous nuclear ribonucleoprotein A2/B1 (hnRNPA$_2$B$_1$)* genes, although most of them are not causative mutations.[38,40-43] Nonetheless, most of these genes are important in osteoclastogenesis, and consequently, the presence of mutations can affect expression and function of osteoclasts. Mutations in *SQSTM1* have been identified in up to 40% of familial PDB cases, and 10% in sporadic cases, and is necessary and sufficient to cause the disease. The mutation in c.1215C>T (p.Pro392Leu) is the most common mutation encountered in PDB patients.[41,44] The *SQSTM1* encode for a ubiquitin-binding protein, called sequestosome 1 (or p62), that is involved in IL-1, TNF and RANKL signaling pathways. Mutations in this gene are dominant and autosomal, presenting a high penetrance within families, and are generally correlated to a more severe form of the disease compared to unmutated PDB patients.[41,44] In the other hand and for the first time, a heterozygous missense mutation of *hnRNPA$_2$B$_1$* was reported in a Chinese family.[42] This heterozygous and dominant mutation occurs in the exon 10, c.929C>T leading to an amino acid substitution of a proline to a leucine. The proline residue is highly conserved in vertebrate species, and the substitution has been predicted *in silico* to be pathological and may alter the protein function. Clinical findings revealed that individuals affected with this mutation presented different clinical manifestation and incomplete penetrance. Nonetheless, the mutation was suggested to be

a candidate disease-causing.[42] Finally, four missense mutations in *ZNF687* have recently been identified and associated to PDB, one of which has been classified as necessary and sufficient for the development of PDB.[38,43]

Other factors have been related to PDB, such as mechanical charge, exposure to toxins such as lead and arsenic, and infection of osteoclast precursors by two paramyxovirus, measles virus and syncytial virus.[45-46] In the latter, nuclear inclusions found in pagetic osteoclasts, seem to closely resemble the nucleocapsids of paramyxovirus, but the role of viruses in the development of PDB is yet to be elucidate.[28,46]

1.3 Homo sapiens ZNF687

1.3.1 Role of *ZNF687*

ZNF687 is a zinc finger protein encoded by the *ZNF687* gene located in chromosome 1 (1q21.3). This nuclear protein contains various classical Cys2-His2 (C2-H2) zinc finger motifs, which are one of the most common DNA-binding motifs found in eukaryotic transcription factors.[47-49] Nonetheless, this zinc finger motif is also able to bind RNA. A single C2-H2 zinc finger motif is constituted by a short and compact alpha helix and beta hairpin fold, that is stabilized by the coordination of a central zinc atom by two cysteine (at one end of β-sheet) and two histidine (at the C-terminus of the α helix) (Figure 1.9). These four hydrophobic residues are highly conserved compared to the other residues in the motif. Since ZNF687 is a transcription factor, it can bind to the DNA of gene promoters, via C2-H2 zinc finger motifs, in a sequence specific manner.[47-49]

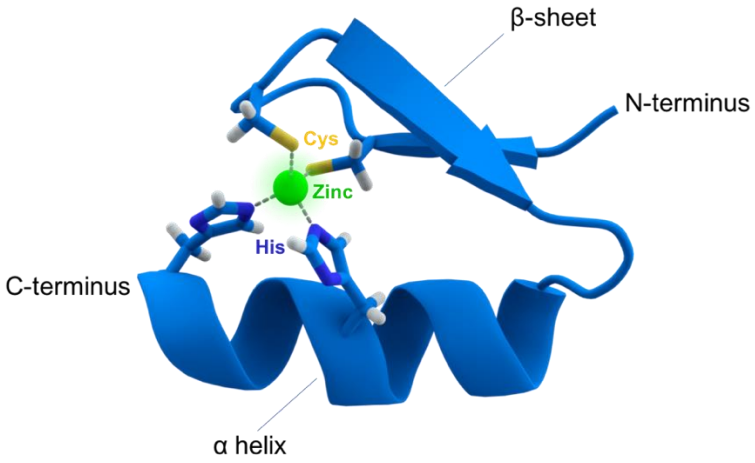


Figure 1.9 – C2H2 zinc finger motif. The zinc molecule (green) is bounded to two cysteine (yellow) and two histidine (dark blue) residues. His: histidine; Cys: cysteine adapted from Splettstoesser.[50]

Moreover, together with ZMYND8 and ZNF592, ZNF687 has also been described as being part of a transcriptional regulator complex, called Z3.[48] This complex interacts mainly with the integrator protein complex (involved in single nuclear RNA processing), but is also associated to H3K4 demethylation machinery, and, therefore, involved in the remodelling of the chromatin for transcription.[49]

ZNF687 is expressed in most normal tissues, such as bone, blood, brain, skeletal muscle, spleen, liver, kidney, lung, ovaries and testies.[38]Although very little is known about this gene, it has recently been demonstrated that *ZNF687* mRNA expression was significantly increased during osteoclastogenesis and osteoblastogenesis, in both human and zebrafish, suggesting a putative role in bone metabolism, in both species.[38] This seems to indicate that the function of this protein might be conserved by evolution, despite speciation. Furthermore, the *Homo sapiens ZNF687* has been related to several diseases such as acute myeloid leukaemia, hepatocellular carcinoma (HCC), and severe Paget's disease of bone associated with giant cell tumour of bone, advocating this time, a putative oncogenic role.[36,43,51-52] Indeed, a translocation of *ZNF687* with *runt-related transcription factor 1* (*RUNX1*) generated a protein with leukemogenicity potential, whereas overexpression of *ZNF687,* found in HCC tissue, enhanced tumour progression and metastasis, and was associated to poorer overall survival.[51-52] Several mutations in *ZNF687* were found in pagetic patients from Avellino geographic area in Italy, and will be discuss in the next paragraph.[38,43]

1.3.2 *ZNF687* Mutations and PDB

Recently, Divisato *et al.* studies have demonstrated that four missense mutations in *ZNF687* were correlated with PDB, one of which was determined as a casual mutation.[38,43] The mutation c.725G>T in exon 2, resulting in an amino acid change p.Ser242Ile, appears to be highly pathogenic, suggesting a causative role (Figure 1.10). Individuals with this mutation presented a polyostotic (2-4 lesion sites) phenotype of the disease, with an average onset of 50 years old. Two novel mutations, c.1994C>T (p.Pro665Leu) and c.2350C>G (p.Gln784Glu), located in exon 2 and 4 respectively, were found in two sporadic cases with a mild phenotype of the disease (polyostotic and late-onset) (Figure 1.10).[43] Both substituted residues are highly conserved and the mutation p.Gln784Glu is located in a zinc finger motif. Finally, the c.2810C>G mutation, located in exon 6, causes an amino acid change p.Pro937Arg, and was described as necessary and sufficient for PDB development (Figure 1.10).[38] The proline residue, highly conserved, especially among mammalian species, is located just before a nuclear localization signal. The mutation seems to increase the translocation of ZNF687 into the nucleus, where its accumulation leads to an overexpression of target genes. Furthermore, individuals affected with this mutation developed a more severe form of PDB, with an earlier onset and a greater number of skeletal lesions than other individual with or without *SQSTM1* mutations, and enhanced the prevalence of GCT up to 30%.[38] Whereas in most PDB, the overactivity of osteoblasts is regarded as a compensatory response of the organism to the overexpression of osteoclasts, here in PDB caused by *ZNF687* mutation c.2810C>G, osteoblasts are directly affected, as well as osteoclasts.

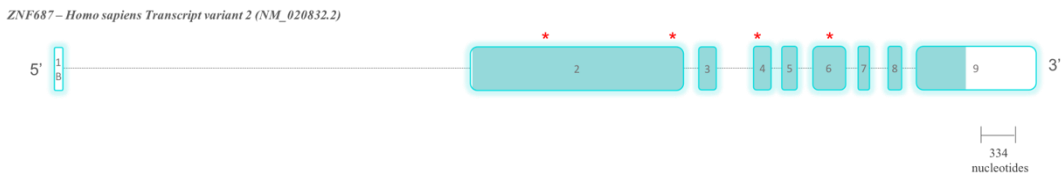


Figure 1.10 – Mutation encountered in *ZNF687*. The transcript variant represented here correspond to the variant 2 (NM_020832.2) identified in Pagetic patients by Divisato *et al.* (2016).[36] Boxes represent exons whereas dote line illustrates the introns. White and Caribbean blue filled boxes represent the non-coding and coding DNA sequence, respectively. Red asterisks depict the mutations identified by Divisato *et al.* (2016 and 2018).[38,43]

1.4 *Danio rerio znf687*

As mentioned before, Divisato *et al.* have demonstrated that *ZNF687* was upregulated during osteoclastogenesis and osteoblastogenesis in both human and zebrafish.[38] Since

there is no animal model, to our knowledge, for the study of *ZNF687*, it will be interesting to consider the zebrafish as a putative model.

The zebrafish, scientifically called *Danio rerio*, is found in tropical waters in Asia, and belongs to the infraclass of teleost fish. This vertebrate fish has been considered as a good biological model in a wide range of studies such as genetic, neurological, comparative, developmental and evolutionary biologies.[53-54] During evolution, the teleost lineage suffered a genome duplication which resulted in the appearance of several orthologs for mammals.[53] Despite this duplication, the zebrafish genome has been fully sequenced, and approximatively 70% of human genes possess at least one orthologue in zebrafish. Moreover, the high fecundity, fast life cycles and external fertilization of zebrafish, as well as transparency, external and fast development of embryos, provide easiness for genetic manipulations and analyses (Figure 1.11). These genetic manipulations include classical or insertional mutagenesis, and gene knock-in or knock-out mediated by CRISPR-Cas9, among others.[55-57] This genome editing is performed in early stages, allowing scientists to study and understand the effect or the importance of a gene involved, for example, in a human disease. Also, strong similarities such as biochemical and physiological processes, and anatomic and developmental aspects of the skeleton, between zebrafish and human, make this teleost fish an excellent model to study human pathologies, including those affecting bones.

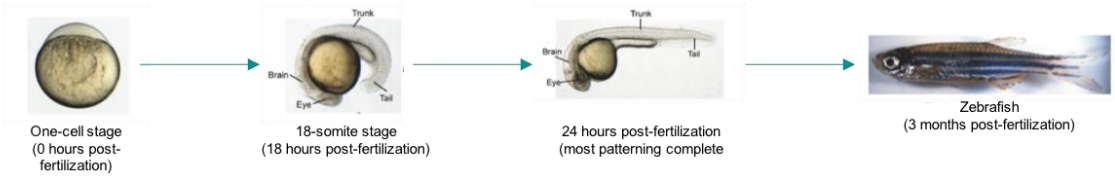


Figure 1.11 – Zebrafish development. The zebrafish development possesses several features that render it incredibly interesting for a wide range of studies. Adapted from Pyatia *et al.* (2007).[58]

The specific duplication of teleost fish genome resulted in the appearance of two *znf687* genes (*znf687a* and *znf687b*) in zebrafish, located in two distinct chromosomes. Indeed, *znf687a* is positioned in chromosome 16, whereas *znf687b* is located in chromosome 19 (Figure 1.12). Both genes produce two different, but quite similar, functional proteins, that appear to be upregulated during the regeneration of caudal fin, indicating a putative role in cell differentiation and proliferation.[36]

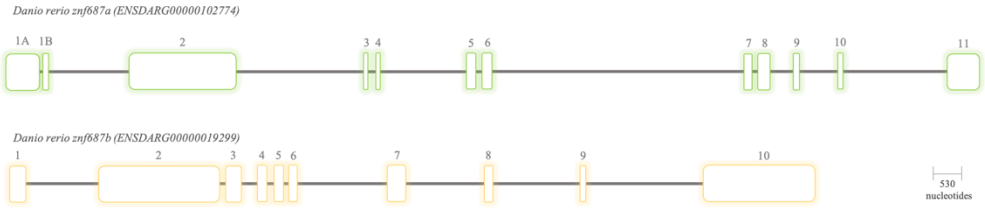


Figure 1.12 – Genomic structure of *znf687a* and *znf687b*. *znf687a* (green) possess 12 exons whereas its paralog, *znf687b* (yellow) possesses 10 exons. Nevertheless, these two genes present a similar genomic structure. Exons are represented by boxes and are at scale.

The *znf687* genes are highly expressed during early zebrafish developmental stages, but *znf687a* increases, in particular, during haematopoiesis and osteoclastogenesis (between 24h and 5 days post fertilization (dpf) and after 20 dfp). However, *znf687a* and *znf687b* are differently expressed in several organs. Indeed, *znf687a* was found to be more expressed in spleen, and kidney (organs responsible for the differentiation of hematopoietic cells, including osteoclast precursor), whereas *znf687b* was more expressed than its paralogue in the eye, gills and scales.[38] This different pattern of expression might be explained by the duplication of the genome, where one or both of the duplicated genes might have gained new functions or undergone sub-functionalization. Nonetheless their involvement in osteoclastogenesis and osteoblastogenesis is similar to human *ZNF687*.

OBJECTIVES

ZNF687, producing a zinc finger protein, has been associated to several diseases. Indeed, the four identified mutations on *ZNF687,* and in particular the c.2810C>G mutation, are correlated to a more severe form of Paget's disease of bone with a higher GCT occurrence, which arises within the Pagetic lesions. Still, very little is known about *ZNF687*, how is it regulated and which genes it regulates, but also about its function in the organism. Nonetheless, it has been demonstrated that *ZNF687* mRNA was upregulated during osteoclastogenesis and osteoblastogenesis, in both human and zebrafish, suggesting that this gene plays a role in bone metabolism in both species, but also, that the function of this protein might be conserved by evolution despite speciation.[38]

Therefore, in order to understand the role of *ZNF687,* and its involvement in osteoclastogenesis and osteoblastogenesis, but also in PDB, we performed different *in silico* and *in vitro* analyses:

- We have first investigated how the four mutations could affect, individually, the protein conformation and function.
- Since c.2810C>G mutation in *ZNF687* is associated with poorer prognostic of PDB, we induced an overexpression of both mutated and normal *ZNF687* on osteoblastic cells, in order to evaluate a differential expression of putative target genes.
- We have also performed a knock-out of the *ZNF687* gene on osteoblast (SaOS-2) and osteoclast (THP-1) cells, mediated by CRISPR-Cas9 genome editing system, and analyzed the resulting phenotype and the expression of putative target genes.
- Parallelly, we performed a comparative *in silico* analysis of ZNF687 protein among different species to contribute to validate zebrafish as a biological model for the study of *ZNF687*.
- Finally, we analyzed the genomic structure of *ZNF687* and identified, by *in silico* analysis, the putative transcription factors binding sites (TFBSs) in the promoter region. Then, we cloned the promoter region into a reporter vector (pGL3-Basic) and tested its effect on *in vitro* luciferase activity.

CHAPTER II – METHODS AND MATERIALS

2.1 *in silico* analysis

2.1.1 Genomic structure analyses of *ZNF687*

2.1.1.1 *ZNF687* transcript variants analysis

The genomic and transcript variant sequences of *ZNF687* from various species (*Homo sapiens, Pan troglodytes*, *Mus musculus*, *Heterocephalus glaber*, *Gallus gallus*, *Oryzias latipes*, *Xenopus tropicalis*, *Takifugu rubripes*, and *Danio rerio*) were retrieved from Ensembl (http://www.ensembl.org/index.html)[59] and Nacional Centre of Biotechnology Information (NCBI) databases (accession numbers are listed in Appendix V, Table A5.1). The structure variants were constructed through the alignment of each variant against its respective genomic sequence using the online program Splign (http://ncbi.nlm.nih.gov/sutils/splign/splign.cgi).[60]

2.1.1.2 Identification of putative binding sites for transcription factors in the promoter region of human *ZNF687*

Transcription factors were selected via UCSC Genome Browser on Human Assembly (https://genome.ucsc.edu/index),[61] using *ZNF687* gene as an input. Then, the putative binding sites of each selected transcription factor were assessed through several online software, such as TFBIND (http://tfbind.hgc.jp),[62] ConTra V3 (http://bioit2.irc.ugent.be/contra/v3/#/step/1),[63] and AliBaba2.1 (http://gene-regulation.com/pub/programs/alibaba2/index.html).[64]

2.1.1.3 Comparative analysis of putative transcription factor binding sites

Promoter region sequences, with 2kb of length, of *ZNF687* from several species were first treated with RepeatMasker web server (http://repeatmasker.org/cgi-bin/WEBRepeatMasker). Then, the putative binding sites of previously selected transcription factors (see section 2.1.1.2) were assessed into each promoter region of selected species, using different online databases such as TFBIND,[62] ConTra V3,[63] AliBaba2.1.[64]

2.1.2 Syntenic Analysis

The conservation of *ZNF687*'s neighbouring genes, among several species, was assessed using the genome browser Genomicus (http://www.genomicus.biologie.ens.fr/genomicus-91.01/cgi-bin/search.pl).[65]

2.1.3 Analysis of ZNF687 protein

2.1.3.1 Protein homology

Protein sequences of different species were retrieved from NCBI database (accession numbers are listed in Appendix V, Table A5.1) and aligned against the human ZNF687 using the multiple sequence alignment online tool, Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).[66] Then, aligned sequences were submitted to Sequence Identity and Similarity (SIAS, http://imed.med.ucm.es/Tools/sias.html) online tool in order to calculate pairwise sequence identity and similarity of ZNF687 protein among different species.

2.1.3.2 Prevision of the secondary structure of ZNF687

The secondary structure of the human ZNF687 protein was putatively assessed using several *in silico* databases, such as Predict Protein (http://predictprotein.org),[67] PSIPRED Protein Sequence Analysis Workbench (http://bioinf.cs.ucl.ac.uk/psipred),[68-69] RaptorX (http://raptorx.uchicago.edu),[70] Spider^2 (http://sparks-lab.org/yueyang/server/SPIDER2),[71-72] NetSurfP (http://cbs.dtu.dk/services/NetSurfP-1.0),[73] APSSP2 (http://crdd.osdd.net/raghava/apssp2),[74] and SABLE (http://sable.cchmc.org).[75]. On the other hand, the two protein sequences of zebrafish were submitted to RaptorX online program in order to generate a predicted secondary structure.[70] The putative structures obtained were then compared to the human wild type ZNF687.

2.1.3.3 Effect of *ZNF687* point mutations in the protein structure

We conducted different analyses in order to understand if the point mutations in *ZNF687,* encountered in Pagetic patients (c.725G>T, c.1994C>T, c.2350C>G, and c.2810C>G), generate a structural effect in the protein. We first used the HOPE web service (http://www.cmbi.ru.nl/hope/)[76] that collects and combines structural information from several sources and give a detailed analysis on the effect of a certain mutation on the protein structure. Then, the putative secondary structure of each mutated protein was

analysed using Predict Protein,[67] PSIPRED,[68-69] RaptorX,[70] Spider[2] (http://sparks-lab.org/yueyang/server/SPIDER2),[71-72] and SABLE (http://sable.cchmc.org).[75]

2.1.3.4 Protein domains analysis

The putative localization of zinc finger C2H2 domains in ZNF687 of different species, but also in human ZNF687 mutated sequences (p.Ser242Ile, p.Pro665Leu, p.Gln784Glu, and p.Pro937Arg), were identified using various web servers, including Pfam 31.0 (http://pfam.xfam.org)[77], Prosite (http://prosite.expasy.org/prosite_ref.html)[78], Uniprot (http://uniprot.org)[79], SMART (http://smart.embl-heidelberg.de)[80] and Conserved Domains NCBI (http://ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)[81].

The nuclear localization signals (NLS) in both wild type and mutated human ZNF687 protein sequences, but also in ZNF687 protein of different species, were putatively analysed by NucPred (http://nucpred.bioinfo.se/cgi-bin/single.cgi)[82], NLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi)[83] and NLStradamus (http://www.moseslab.csb.utoronto.ca/NLStradamus/).[84]

2.2. Molecular Biology

2.2.2 Genome editing

2.2.2.1 Site-directed mutagenesis

The QuickChange Lightning Site Directed Mutagenesis Kit (Agilent Technologies, #210518) was used to generate the point mutation in *ZNF687* (c.2810C>G) encountered in Pagetic patients.[38] In this procedure, we used a basic supercoiled double-stranded (dsDNA) vector, pCMVSPORT6 (Appendix 4, Figure A4.1), with *ZNF687* cDNA as its insert, that was kindly offered by Dr. F. Gianfrancesco (Institute of Genetics and Biophysics, Naples, Italy), and two mutagenic synthetic oligonucleotide primers. These two primers containing the desired mutation, were designed using the web-based QuickChange Primer Design Program available online (https://www.genomics.agilent.com/primerDesignProgram.jsp), and ordered to STAB VIDA (Caparica, Portugal).

2.2.2.1.1 Quick-Reference Protocol

Mutagenesis reaction was performed following supplier protocol found in the instruction manual. Briefly, we set up a sample reaction using 50 ng

pCMVsoprt6+*ZNF687* vector, 5 µl 10x reaction buffer, 125 ng of each mutagenic primer (Appendix 1, Table A1.1), 1 µl dNTP mix, 1.5 µl QuickSolution reagent, and ddH_2O to a final volume of 50 µl. Then 1 µl of QuickChange lightning enzyme was added to the sample and the reaction was cycled, in a thermocycler (2720 Thermal Cycler Applied Biosystems) following the parameters found in Table 1.

Table 1 – Cycling reaction conditions for the mutagenesis reaction.

Cycle number	Denaturation	Annealing	Extension
1	95ºC, 2 min		
2-20	95ºC, 20 s	60ºC, 10 s	68ºC, 4 min 30 s
21			68ºC, 5 min

After the generation of the mutated plasmid, a quick digestion was performed using *Dnp I* endonuclease to select for mutation-containing synthesized DNA, by digesting parental DNA templates. Therefore, at the end of the cycling reaction, *Dpn*I restriction enzyme was added to the mixture, and the digestion reaction was immediately incubated at 37ºC for 5 minutes. Then, *Dpn*I digested DNA was transformed into XL 10-Gold ultracompetent bacterial cells, previously supplemented with β-mercaptoethanol (14.3 M) in order to increase transfection efficiency, and reaction was incubated at 4ºC for 30 minutes. After, the sample suffered a heat-shock at 42ºC for 30 seconds and was immediately re-incubated at 4ºC for 2 minutes. Pre-heated (42ºC) room-temperature super optimal broth with catabolite repression (SOC, see composition in Appendix II, Table A2.1) was added and the mixture was incubated at 37ºC, for 1 hour under constant shaking at 250 rpm. Finally, the transformation reaction was plated on Luria-Bertani (LB) agar (Appendix II, Table A2.1) agar + ampicillin (50 µg/ml) and incubated overnight (>16 hours), at 37ºC.

2.2.2.1.2 Screening of mutated *ZNF687* (c.2810C>G)

Individual colonies were selected and inoculated into a culture LB-medium with ampicillin (50 mg/ml). Each sample was then incubated overnight at 37ºC, under constant agitation. After DNA extraction, following Sambrook *et al.* (1989) protocol[85] (Appendix III), samples were digested using 2 µl DNA extract, 2 µl buffer H 1x, 0.2 µl *EcoR*V (15 U/µl, Takara), 0.2 µl *Not*I (10 U/µl, Takara), 2 µl bovine serum albumin (0.1%, BSA) and ddH_2O for a total volume of 20 µl. The reaction was incubated at 37ºC for one hour.

Digested samples were then electrophoresed in a 1.2% agarose gel (Appendix II, Table A2.1), containing GreenSafe nucleic acid gel stain (NZYtech) After the addition of loading dye 1x (Appendix II, Table A2.1), samples were electrophoresed during approximatively 30 minutes at constant voltage (120V), together with a GeneRuler 1kb DNA ladder (Fermentas, Appendix III, Figure A3.2) marker, and further visualized with a trans-illuminator (Ultra-lum) under ultra-violet (UV) light. Identified positive fragments, e.g. with the expected size, were sequenced using ZNF687_Rev primer (Appendix I, Table A1.1). Positive samples were stored at -20ºC for downstream application (transfection assay 2.4.1).

2.2.2.2 CRISPR-Cas9/Cas9n system

The RNA-guided Cas9 nuclease, from the microbial clustered regularly interspaced short palindromic repeats (CRISPR), one of the recent genomes editing technologies used in eukaryotic cells, was selected to perform *ZNF687* knock-out. This system is composed by a nuclease Cas9 and a single-guided RNA (sgRNA) which comprised a ~20-nucleotide guide sequence and a scaffold (Figure 2.1). The CRISPR-Cas9 is directed to the genomic DNA target, by the sgRNA, where the guide sequence will bind, through base pairing, allowing the nuclease Cas9 to mediate a double-stranded break (DBS) approximatively 3 bp upstream a protospacer adjacent motif (PAM) (Figure 2.1). [86-90]

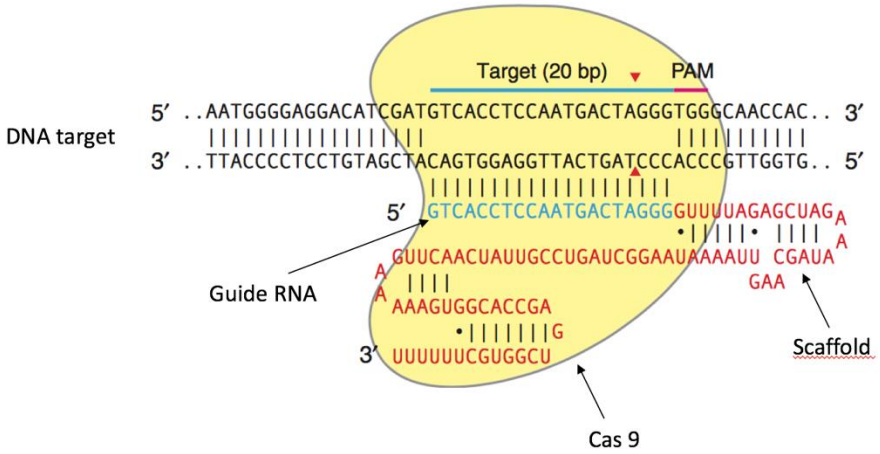


Figure 2.1 – CRISPR-Cas9 system. This system generates a DBS at a genomic DNA target, through base pairing, that will be repair by the cell either through non-homologous end joining repair or through homology direct repair. The blue and red sequences represent the guide RNA and the scaffold, respectively. The nuclease Cas9 is illustrated in yellow. Red triangles indicate the localization of the break in the DNA target. PAM: protospacer adjacent motif. Adapted from Ran et al. (2013).[86]

Additionally, by inflecting mutation in the Cas9 nuclease, the Cas9 mutant (Cas9n) rather nicks than cleaves the DNA target and yield a single-stranded break. Usually, two

sgRNA are designed to nick, at the same time, the DNA target strands. This method allows to decrease the numbers of off-targets and increase the specificity of target recognition. The generated DBS, or nick, is repaired by the cell, either through non-homologous end joining repair (NHJR), or through homology-directed repair (HDR) in the presence of a repair template. The NHJR, which is the preferential pathway, produces an indel mutation that leads to the appearance of a premature stop codon, and hence to a non-functional truncated protein.[86-90]

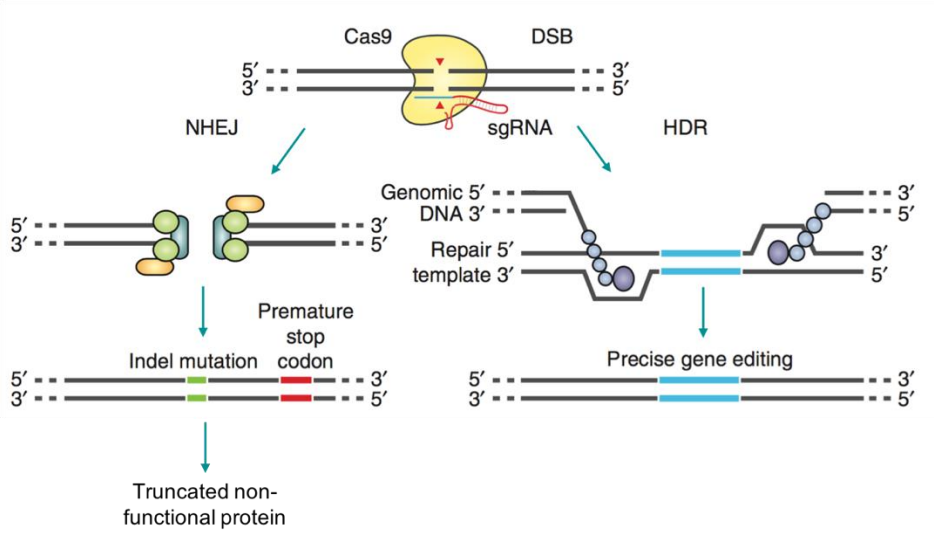


Figure 2.2 – Mechanisms of DNA repair after generation of DBS mediated by CRISPR-Cas9. The damaged target locus suffers a homology-directed repair, or, in the absence of a repair template, a non-homologous end joining repair. The latter will produce an indel mutation that will lead to a frameshift, and then to a premature stop codon. Finally, the synthesized protein is truncated and non-functional. Adapted from Ran et al. (2013).[86]

2.2.2.2.1 Oligos design

For the guide RNAs construction, oligos were retrieved from an online CRISPR design tool, the Oligo Optimized CRISPR Design (http://crispr.mit.edu). The input sequence was introduced (approximately 200 nucleotides downstream the ATG start codon of the CDS) and the design tool provided suitable target sites with their respective score and identified possible off-targets. The oligo presenting the best score is selected and its complementary reverse sequence is designed. Overhangs are added at the extremity of the gRNAs, with the top and bottom strands orientation matching those of the genomic strand as the top and bottom strands for each sgRNA design and were ordered from STAB VIDA. All oligos, one pair for Cas9 (oligo 1) and two pairs for Cas9n (oligo 2A & 2B), are listed in Table A1.1 (Appendix I).

2.2.2.2.2 Oligos annealing and phosphorylation

Each pair of oligo was annealed and phosphorylated according to the following reaction: 1 µl T4 ligation buffer 10x, 1 µl of each oligo (100 µM), and T4 Polynucleotide Kinase (PNK, 10U/µl, Fermentas) were mixed and samples were then subjected to 37ºC for 30 min, followed by 95ºC for 5 min in a thermocycler (C1000 Thermal Cycler, BIO-RAD) and temperature was ramped down overnight at room- temperature. This protocol was based in Ran *et al. (2013)* protocol.[86] Phosphorylated and annealed oligos were diluted 1:250 in nuclease free water.

2.2.2.2.3 Plasmid PX459 and PX462 digestion

PX459 (Addgene, #62988, Appendix III, Figure A.3) is a 9174 bp plasmid that contains the Cas9 gene from *S. pyogenes* together with a puromycin selectable marker (2A-puro) and a cloning backbone site for sgRNA (oligo1). On the other hand, PX462 (Addgene, #62987, Appendix III, Figure A.2) is a 9175 bp plasmid that possesses the Cas9n (D10A nickase mutant) from *S. pyogenes*, a puromycin cassette (2A-puro) for eukaryotic selection, and a cloning site for sgRNA (oligo2A & 2B). These two plasmids, kindly offered by Dr. W. Link (Center for Biomedical Research, University of Algarve) were digested using *Bpi*I restriction enzyme. The reaction was conducted for both plasmids as followed: 1 µg plasmid (PX459 or PX462), 1 µl FastDigest *Bpi*I (#FD1014, Thermo Scientific), 1 µl Calf Intestinal Alkaline Phosphatase (CIAP, 20U/µl, Invitrogen), 2 µl FastDigest buffer 10x (Thermo Scientific), and nuclease free water until final volume reaches 20 µl. Reactions were incubated at 37ºC for 30 minutes. Next, digested plasmids were screened by submitting the samples to an electrophoresis in a 1 % agarose gel, using non-digested plasmids (PX459 and PX462) as negative controls. Fragments corresponding to the digested plasmids were extracted and purified using a GeneJET Gel Extraction Kit (Thermo Scientific, #K0692)

2.2.2.2.4 Cloning reaction

The annealed oligo 1 was cloned into PX459 whereas oligo 2A and oligo 2B were separately cloned into PX462. The cloning reaction was set up as follow: 50 ng of *Bpi*I digested plasmid, 1 µl of diluted oligo duplex, 1 µl T4 DNA ligation buffer 10x, and 1 µl T4 ligase (1000 U/µl, Thermo Scientific), and nuclease-free water until final volume 10 µl. Samples were incubated at 22ºC, for 10 minutes. Final products were then transformed into *E.coli* Stbl4 competent bacteria cells, following Sambrook *et al.* (1989) protocol,[82] and plated into LB + ampicillin (50 mg/ml) medium.

2.2.2.2.5 Alternative digestion-ligation reactions

An alternative protocol was tested to increase cloning efficiency by obtaining higher percentage of correct colonies. The reaction was set up as follow: 100 ng PX459 (or PX462), 2 µl phosphorylated and annealed oligo duplex (1:250 dilution), 1 µl Dithiothreitol (DTT, 10 mM), 1 µl FastDigest *Bpi*I, 2 µl Bovine Serum Albumin (BSA, 0.1 mg/ml), 1.1 µl Salt solution (1.2 M NaCl), 2 µl T4 buffer 10x, 0.5 µl T4 DNA ligase, and nuclease free water until 20 µl final volume. The reaction was then incubated in a thermocycler fulfilling the conditions described in Table 2.

Table 2 – Digestion-ligation reaction conditions for CRISPR-Cas9/Cas9n constructs.

Cycle number	Step 1	Step 2	Step 3
1-6	37ºC, 5 min	23ºC, 5 min	
7			4ºC, until ready to proceed

Once the reaction was completed, final products were transformed into Stbl4 competent cells and plated on LB agar + ampicillin (50 mg/ml). The plates were incubated overnight at 37º'C.

2.2.2.2.6 Screening of CRISPR-Cas9/Cas9n constructs

Individual colonies were picked to check for the correct insertion of sgRNA and inoculated into a culture of LB-medium with ampicillin (50 mg/ml). Samples were then incubated at 37ºC overnight, under constant agitation. After DNA extraction, following previously mentioned protocol, a PCR was performed using 1 µl DNA extract, 2.5 µl 10x DreamTaq buffer, 0.5 µl $MgCl_2$ (50 mM), 0.5 µl dNTPs mix (10 mM), 0.5 µl primer U6 (10 mM, Appendix I, Table A1.1), 0.5 µl primer RevCas9/Cas9n (10 mM, Appendix I, Table A1.1), and 0.25 µl DreamTaq DNA polymerase (5 U/µl, Thermo Scientific). The PCR was executed in a thermocycler under the conditions described in Table 3.

Table 3 – PCR conditions for the amplification of sgRNAs

Cycle number	Denaturation	Annealing	Extension
1	94ºC, 3 min		
2-35	94ºC, 45 s	60ºC, 30 s	72ºC, 30 s
36			72ºC, 10 min

Final products were then separated by electrophoresis in a 1.4% agarose gel during approximatively 30 min at constant voltage (120V) and a GeneRuler 100 Plus DNA ladder (Fermentas) was used as marker. Samples were further visualized under UV light, and positive samples, i.e. with expected size, were then sent for sequencing using U6_Fwd primer (Appendix I, Table A1.1).

2.2.3 DNA extraction of SaOS-2 knock-out clones

Genomic DNA was isolated from ZNF687-KO SaOS-2 clones (see section 2.4.1), at full confluency in 100 mm dishes. First, cells were washed twice with Phosphate Buffered Saline (PBS) 1x (Appendix II, Table A2.1) to remove residual medium and then detached by trypsinization (Appendix II, Table A2.1). Then, cells were collected and transferred into a sterilized microcentrifuge tube and centrifuged at 250 x g, at 4ºC, for 5 minutes. After, pelleted cells were resuspended in a lysis buffer (Appendix II, Table A2.1) supplemented with proteinase K (10 µg/ml) and glycogen (20 µg/ml) and incubated in a humid incubator at 60ºC, for 3 hours. Next, precipitation buffer (Appendix II, table A2.1) was added to each lysate and the mixture was incubated for 30 min, at room-temperature. After the incubation time, the lysates were centrifuged at maximum speed in a refrigerated micro-centrifuge, for 15 minutes. The supernatant was discarded, and the precipitated DNA was washed with ice cold 70% ethanol, followed by a refrigerated centrifugation at maximum speed for 5 minutes. The supernatant was then removed, and pelleted DNA was allowed to dry at room-temperature. Once dried, DNA was resuspended in nuclease-free water, quantified in a photometer (Nanodrop Photometer 4.0), and finally stored at -20ºC, for further use (see section 2.2.9.1).

2.2.4 Protein extraction of SaOS-2 knock-out clones

Total protein extracts of SaOS-2 parental cells and ZNF687-KO SaOS-2-clones were obtained from confluent cell cultures, in 100 mm dishes. Cells were washed with cold PBS 1x to remove residual medium and then detached by trypsinization. Cells were then transferred into a sterilized microcentrifuge tube and centrifuged for 5 minutes at 250 xg, at 4ºC. Pelleted cells were resuspended in CST buffer (Appendix II, Table A2.1), supplemented with a proteinase inhibitor and Calyculin A (Santa Cruz Biotechnology), for cellular lysis. Lysate suspension were vortexed and incubated at 4ºC, for 20 minutes, under constant agitation. After a refrigerated centrifugation at 150 000 rpm, for 15 minutes, an equal volume of each supernatant was collected and transferred into a new

microcentrifuged tube. Protein extracts were quantified using Quick Start Bradford Protein Assay Kit (Biorad, cat# 500-0201, see protocol in Appendix III, section A3.2) that includes 1x dye reagent and Bovine Serum Albumin (BSA) standard at 2 mg/ml. After protein quantification, Laemmli buffer and CST buffer (supplemented with proteinase inhibitor and Calyculin A) were added to each sample in order to obtain a final protein concentration of 3 µg/µl.

2.2.5 Western Blot assay of knock-out clones

After denaturation at 100ºC for 5min, proteins (3 µg/µl) were separated in a 10% SDS-PAGE gel (Appendix II, Table A2.1) and then transferred into a nitrocellulose membrane (#10600001 GE life sciences). Proteins were blocked in a 5% non-fat milk (Blotting-Grade Blocker, BIO-RAD) in TBS-T 1x (Appendix II, Table A2.1), for 1 hour. Next, considering the predicted size of ZNF687 and β-actin proteins, the membrane was cut in two. Each separated membrane was incubated overnight at 4ºC with its respective antibody: anti-ZNF687 (1:500, #6861 ProSci) and anti-β-actin (1:500, #sc47778, Santa Cruz Biotechnology). Then, membranes were washed three times in TBS-T 1x, for 5 minutes and incubated for 1 h, at room-temperature, with anti-rabbit (1:10000, #NA934 Amersham) and anti-mouse (1:10000, #NA931 Amersham) antibodies for ZNF687 and β-actin protein detection, respectively. After incubation, membranes were washed three times in TBS-T 1x and dived in ECL solution (Appendix II, Table A2.1), for 5 minutes, and images were captured and visualized with ImageLab BIO-RAD software.

2.2.6 RNA extraction

For further gene expression assays by RT-qPCR (see section 2.2.8), the RNA from the KO clones, mineralized KO clones, ZNF687- and ZNF687m-overexpressed cells, and respective SaOS-2 cells controls, was isolated using a GeneJET RNA Purification Kit (Thermo Scientific, #K0732), and final purified RNA was quantified in a photometer.

2.2.7 cDNA synthesis by Reverse Transcriptase reaction

To enhance the purification of the RNA, 500 ng of RNA extract was first digested with RNA Qualified (RQ1) DNase (Promega), and the reaction was incubated for 30 minutes, at 37ºC. After the addition of RQ1 DNase stop solution (Promega), the reverse transcriptase reaction was performed using digested RNA extract, oligo dT (50µM), deoxyribonucleotide triphosphate (dNTPs, 10 mM), and the mixture was incubated for 5

minutes, at 65ºC, followed by 5 minutes, at 4ºC. Next, FS buffer 5x (Invitrogen), dithiothreitol (DTT, 0.1 M, Invitrogen), and Ribolock RI (Thermo Scientific) were added to the previous reaction, and the mixture was incubated at 37ºC, for 2 minutes. Finally, Moloney murine leukaemia virus reverse transcriptase (M-MLV RT, Invitrogen) was added and the reaction was incubated for 50 minutes, at 37ºC, and then for 15 minutes, at 70ºC.

2.2.8 Quantitative Real Time-PCR (qRT-PCR)

The expression of several genes was assessed by qRT-PCR using the isolated RNA described in section 2.2.4. The qRT-PCR was performed using 10 µl of Sensi Fast (BIO-RAD), 0.6 µl primer (10mM), 2 µl of cDNA (1:10), and nuclease free water until final volume of 20 µl. The samples were subjected, in a thermocycler (CFX, BIO-RAD), under the conditions presented in Table 4. Fluorescence was measured by the SYBR/FAM only channel. The levels of gene expression were normalized, using *GAPDH* and *β-actin* as reference genes, and calculated using the comparative method *ΔΔ*Ct. All the primers used in this reaction are listed in Table A1.1 (Appendix I).

Table 4 - qRT-PCR conditions for analysis of the *ZNF687* (over-expressed and mutated in SaOS-2 cell line) expression.

Cycle number	Denaturation	Amplification	Extention
1	95ºC, 20 s		
2-41	95ºC, 3 s	60ºC, 30 s	
42			65ºC, 5 s
43			95ºC, 50 s

2.2.9 Polymerase chain reaction (PCR)

The primers used in all PCRs were designed using the bioinformatic program perlprimer[84] or Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0/) and were ordered to STAB VIDA. All primers used in PCRs are listed in Table A1.1 (Appendix I).

2.2.9.1 PCR for indel mutation screening of KO SaOS-2 cells

For this reaction, we followed the *Taq* DNA Polymerase (Invitrogen) protocol,[85] in which 50 ng of DNA were added to 5 µl 10x PCR buffer minus Mg, 1 µl dNTPs mixture (10 mM), 1.5 µl $MgCl_2$ (50 mM), 2.5 µl primer mix (ZNF687mFwd1 and ZNF687mRev1,10 mM each. Table A1.1), 0.5 µl *Taq* DNA polymerase (5U/µl, Invitrogen), and nuclease

free water to a final volume of 50 µl. The PCRs were executed in a thermocycler under the conditions shown in Table 5.

Table 5 – PCR conditions for the amplification of a sequence harbouring the indel mutation in KO saos-2 cells.

Cycle number	Denaturation	Annealing	Extension
1	95ºC, 5 min		
2-35	98ºC, 20 s	58ºC, 15 s	72ºC, 2 min
36			72ºC, 7 min

2.2.9.2 Amplification of human and zebrafish promoter's region

Four fragments of the human promoter region were amplified using two different enzymes. Indeed, KAPA HiFi was used to amplify fragment 1 (F1, 961 bp) and fragment 2 (F2, 850 bp), whereas *Taq* DNA polymerase was employed to amplify fragment 3 (F3, 1795 bp), and fragment 4 (F4, 1416 bp) (Figure 2.3).

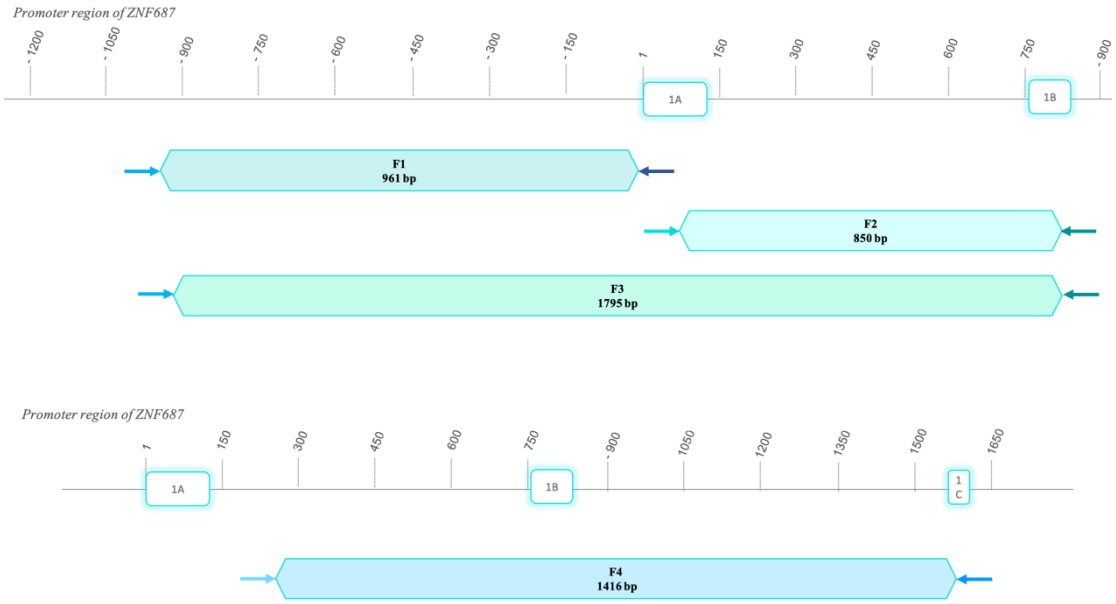


FIGURE 2.3 – Schematic representation of ZNF687 promoter fragments. Only partial sequences of *ZNF687* (from -1200 to 900 bp, and from 1 to 1650 bp) are represented here by a line. Square boxes represent exons, whereas hexagonal boxes represent the fragments that were amplified from the promoter region. Arrows represent primers used for the amplification of each fragment: HsZNF687_F1_Fwd and HsZNF687_F1_Rev for F1; HsZNF687_F21_Fwd and HsZNF687_F2_Rev for F2; HsZNF687_F1_Fwd and HsZNF687_F2_Rev for F3; HsZNF687_F4_Fwd_2 and HsZNF687_F4_Rev_2 for F4 (Appendix I, Table A1.1).

The amplification of F1 and F2 was performed following KAPA HiFi manufacturer protocol. Briefly, 50 ng DNA, 10 µl KAPA HiFi buffer 5x (Kapa Biosystems), 3 µl primer mix (HsZNF687_F1_Fwd and HsZNF687_F1_Rev for F1, and HsZNF687_F1_Fwd and HsZNF687_F2_Rev for F2, 10 mM each. Table A1.1), 1.5 µl

dNTPs Kapa mix (10mM, Kapa Biosystems), 1 µl KAPA HiFi HotStart polymerase (1U/µl, Kapa Biosystems), and nuclease free water (Sigma) to a final volume of 50 µl were used for each sample, and reaction was processed, in a thermocycler, under the conditions depicted in Table 6.

TABLE 6 - PCR conditions for the amplification of fragments F1 and F2 of the human *ZNF687* promoter region.

Cycle number	Denaturation	Annealing	Extension
Fragment 1 – F1			
1	95ºC, 5 min		
2-35	98ºC, 20 s	58ºC, 15 s	72ºC, 2 min
36			72ºC, 7 min
Fragment 2 – F2			
1	95ºC, 5 min		
2-35	98ºC, 20 s	58ºC, 15 s	72ºC, 2 min
36			72ºC, 7 min

The amplification of F3 and F4 were executed following *Taq* DNA Polymerase protocol,[85] in the same conditions depicted in section 2.2.9.1 and in Table 7. Primers HsZNF687_F1_Fwd and HsZNF687_F2_Rev were used for the generation of F3, and primers HsZNF687_F4_Fwd_2 and HsZNF687_F4_Rev_2 were used for F4 (F4 (Table A1.1).

Table 7 - PCR conditions for the amplification of F3 and F4 of the human *ZNF687* promoter region.

Cycle number	Denaturation	Annealing	Extension
Fragment 3 – F3			
1	94ºC, 3 min		
2-35	94ºC, 45 s	55ºC, 30 s	72ºC, 2 min
36			72ºC, 10 min
Fragment 4 – F4			
1	94ºC, 3 min		
2-35	94ºC, 45 s	55ºC, 30 s	72ºC, 1 min 30 s
36			72ºC, 10 min

On the other hand, one fragment corresponding to the promoter region of each *znf687* gene of the zebrafish was obtained (Figure 2.4). The fragment A (FA, 1231 bp), representing a portion of *znf687a* promoter region, and fragment B (FB, 1166 bp), representing a portion of *znf687b* promoter region, were amplified by the enzyme *Taq*

DNA polymerase, following the manufacturer protocol[85] as exemplified in section 2.2.9.1. The PCRs conditions are depicted in Table 8. To obtain the desired FB, a PCR was performed to amplify a fragment with 2134 bp followed by a nested PCR to generate a fragment with 1166 bp.

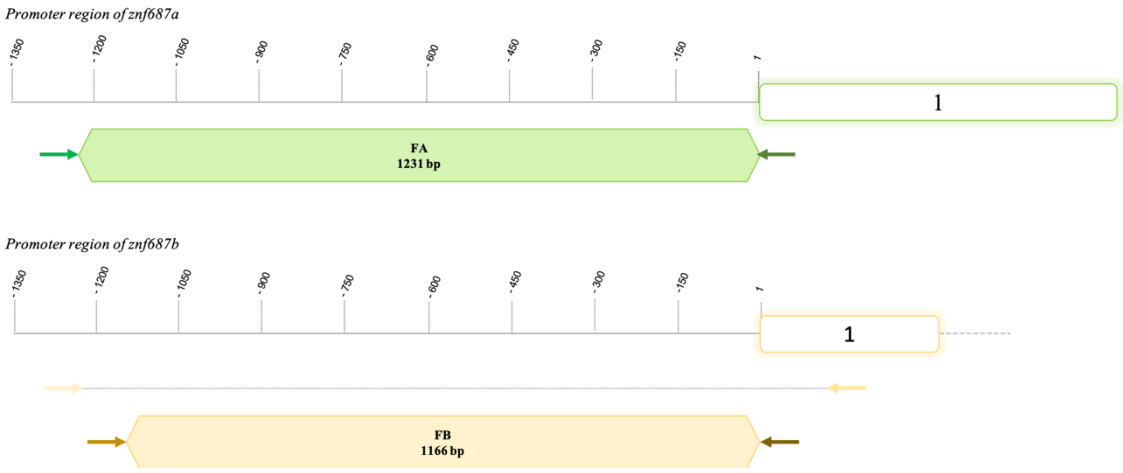


FIGURE 2.4 – Schematic representation of *znf687a* and *znf687b* promoter fragments. Only a partial sequence of *znf687a* (from -1350 to 900 bp) *and znf687b* (from -1350 to 900 bp) are represented here by a line. Square boxes represent exons, whereas hexagonal boxes represent the fragments that were amplified from the promoter region. Arrows represent primers used for the amplification of each fragment: HsZNF687_F1_Fwd and HsZNF687_F1_Rev for F1; HsZNF687_F21_Fwd and HsZNF687_F2_Rev for F2; HsZNF687_F1_Fwd and HsZNF687_F2_Rev for F3 (Table A1.1).

TABLE 8 – PCR conditions for the amplification of fragments A and B of the zebrafish *znf687a* and *znf687b* promoter regions, respectively.

Cycle number	Denaturation	Annealing	Extension
Fragment A – FA			
1	94ºC, 3 min		
2-35	94ºC, 45 s	53ºC, 30 s	72ºC, 1 min 30 s
36			72ºC, 10min
Fragment B – FB			
PCR to amplify the fragment with 2134 bp			
1	94ºC, 3 min		
2-35	94ºC, 45 s	50ºC, 30 s	72ºC, 2 min 30 s
36			72ºC, 10 min
Nested PCR to amplify the fragment with 1166 bp			
1	94ºC, 3 min		
2-35	94ºC, 45 s	50ºC, 30 s	72ºC, 1 min 30 s
36			72ºC, 10 min

2.2.9.3 Screening of amplified fragments by electrophoresed

Finally, samples were electrophoresed in an agarose gel containing GreenSafe nucleic acid gel stain (NZYtech). After the addition of loading dye 1x, samples were electrophoresed, during approximatively 30 minutes, at constant voltage (120V), together with a DNA ladder marker, and further visualized under UV light. Fragments with expected size were extracted using a GeneJET Gel Extraction Kit (Thermo Scientific, #K0692) and sequenced at the CCMAR sequencing facility, using respective primers.

2.2.10 Cloning reactions of human and zebrafish amplified promoter fragments

2.2.10.1 A-tailing reaction

Since KAPA HiFi polymerase has a proofreading function, an adenine was inserted to each 3'-end of purified DNA fragments, amplified with KAPA HiFi polymerase, in order to create sticky ends for further ligation into pCR®II-TOPO® vector (Invitrogen, Appendix III, Figure A4.5). Therefore, 4 µl *Taq* buffer, 0.8 µl dATP (10mM), 1.2 µl $MgCl_2$ (50 mM) and 1 µl *Taq* polymerase (5U/µl, Invitrogen), were added to the purified DNA. The reaction was finally incubated at 72ºC, for 5 minutes.

2.2.10.2 TOPO cloning reaction

The amplified fragment was cloned into a pCR®II-TOPO® vector. This ampicillin resistant plasmid possesses 3'-thymine overhangs for higher ligation efficiency, and a reporter gene Lac-Z which encodes the protein β-galactosidase (β-gal) for the blue-white screen. For the cloning reaction, 1 µl pCRII-TOPO vector, 1 µl Salt solution (1.2 M NaCl, 0.06 M $MgCl_2$, Invitrogen) and 4 µl purified PCR product were gently mixed and incubated at 4ºC, overnight.

2.2.10.3 Transformation of competent bacteria DH5α cells

The cloned fragment was inserted into competent *E.coli* DH5α cells (obtained as explained in Appendix III) following Sambrook *et al.* (1989)[85] protocol with modifications, as described below. Thus, 2 µl of cloned fragment were added to 50 µl competent bacteria and the reaction was incubated on ice, for 30 minutes. After, a heat shock at 42ºC was induced for 45 seconds to increase the membrane fluidity for DNA uptake. The reaction was then immediately incubated on ice for another 2 minutes. Under a Bunsen burner, 250 µl of room-temperature SOC were added, and the reaction was incubated at 37ºC under constant agitation (170 rpm), for 1 hour, to allow bacteria to

recover and express the antibiotic resistance marker encoded by the vector. Afterwards, 5 µl of isopropyl β-D-1-thiogalactopyranoside (IPTG; 200 mg/ml, Appendix II, Table A2.1) were added to the bacterial suspension that was then plated on a pre-warmed LB agar + ampicillin (50 µg/ml) medium, previously treated with 40 µl of 5-bromo-4-cloro-3-indoxil-β-D-galactopiranoside (X-Gal; 20 mg/ml, see composition in appendix II, Table A2.1). Bacteria were grown overnight at 37ºC.

2.2.10.4 Plasmid DNA extraction and purification

Individual white colonies were picked and inoculated into LB broth+ampicillin (50 µg/ml) medium at 37ºC overnight, under constant agitation (200 rpm). Subsequently, DNA of each bacterial suspension was extracted following manual mini-prep established procedure.[85] Briefly, the bacterial suspension was centrifuged at maximum speed and the pellet was then resuspended in solution P1 (Appendix II, Table A2.1) preventing the activation of DNases and allowing the degradation of RNA. After, solution P2 (Appendix II, Table A2.1) was added in order to lyse the cells wall and denature the cellular proteins and bacterial DNA. The reaction was incubated at room-temperature, for 5 minutes. Then, solution P3 (Appendix II, Table A2.1) was added to promote bacterial DNA and cellular protein precipitation, and the reaction was incubated on ice, for 10 minutes. After incubation, reaction was centrifuged for 8 minutes, at 14000 rpm. Plasmid DNA was then precipitated with ice cold 100% ethanol and pelleted DNA was washed with ice cold 70% ethanol. After centrifugation at 14000 rpm for 2 minutes, plasmid DNA was resuspended in 30 µl nuclease free water.

2.2.10.5 Screening purified plasmid DNA by restriction endonucleases digestion

To confirm the presence of the insert, purified DNA of each fragment was digested with restriction endonuclease *EcoR*I fast digestion (15U/ µl, Takara) in 10x buffer fast digestion for 20 minutes, at 37ºC in a thermoblock (Biometra Tri Heater lid). Then, final product was separated in a 1% agarose gel and visualized by GreenSafe staining under UV light. Identified clones with the correct size were sequenced at the sequencing facility of the CCMAR using the M13F primer (Appendix I, Table A1.1). Results were finally analysed *in silico* by Blastn (NCBI).

2.2.10.6 pGL3-Basic reporter plasmid cloning reaction

Each sequenced fragment was analysed in an online program, RestrictionMapper (http://restrictionmapper.org), in order to choose the desired pair of enzymes among pGL3-Basic (Invitrogen, Appendix IV, Figure A4.6) restriction enzymes. DNA fragments were then excised from TOPO vector using restriction enzymes:

- *Kpn* I (10 U/µl, Takara) and *Xho* I (10 U/µl, Takara) for F1, and
- *Hind* III (15 U/µl, Takara) and *Xho* I for F2 and FA

pGL3-Basic vector was also digested with the same restriction enzymes as its future insert (F1 or F2). Therefore, 2 µl 10x M buffer, 0.5 µl of each restriction enzyme, 1 µg cloned fragment DNA or pGL3-Basic vector, and ddH20 for a final volume of 20 µl, were used for the digestion reaction of TOPO vector and pGL3-Basic vector, respectively. The reactions were then incubated at 37ºC in a thermoblock, for 1 hour. Buffers used in these reactions are chosen in accordance with the double-digestion Takara chart available online at the following server: http://clontech.com/SV/Products/Molecular_Biology_Tools/Restriction_Enzymes/Double_Digestion_Buffers.

The cloning reaction in pGL3-Basic vector, was established as follow: 100 ng pGL3-Basic vector, *x* ng of insert DNA, 1 µl T4 ligase buffer 10x, and 0.5 µl T4 ligase (1000U/µl, Thermo Scientific), where:

$$x = \frac{(100ng\ pGL3\ vector\ \times\ nº\ bp\ of\ insert)}{nº\ bp\ pGL3\ vector} \times \frac{3}{1}$$

Each digestion reaction was incubated at 4ºC, overnight.

2.2.10.7 Screening positive cloned fragments in pGL3-Basic vector

After the cloning reaction, cloned fragments were inserted into competent *E.coli* DH5α cells, followed by DNA extraction, according to the same procedures as mention before. Then, a double digestion using respective restriction enzymes (*Kpn* I and *Xho* I for F1; *Hind* III and *Xho* I for F2 and FA) were conducted in the following conditions: purified DNA were added to a mix containing 10x buffer M, restriction enzymes (5U/µl each) and nuclease free water. Reactions were incubated at 37ºC, for 1 hour in a thermoblock. Afterwards, each final product was separated in a 1% agarose gel by

electrophoresis, and putative positive samples, *e.g.* with the correct DNA fragment size, were sent for sequencing using primer R240 (Appendix I, Table A1.1).

2.3 Cell culture

2.3.1 SaOS-2 cell line culture conditions

SaOS-2 are adherent human epithelial cells originated from osteosarcoma (ATCC®, HTB-85™) that were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma), 1% L-glutamine (L-Glu, 200mM, Gibco) and 1% (w/v) Penicillin/Streptomycin (Pen/Strep, Gibco). Cells were incubated in a humidified atmosphere, at 37ºC, enriched with 5% CO_2 (standard conditions) and subcultured every 3 days in a 100 mm culture dish.

2.3.2 Seeding and differentiation conditions for mineralization assays

SaOS-2 parental cells and SaOS-2-KO-clones were seeded at 1 x 10^5 cells per well in a 24-well plate (for calcium, phosphate and collagen detection assays) or at 5 x 10^5 cells per well in a 6-well plate (for alkaline phosphatase assay) and incubated in standard conditions for 48 hours. Then, differentiation reaction was performed by supplementing the culture medium with 50 µg/ml ascorbic acid and 10 mM β-glycerophosphate for 6 days in confluent cultures. Supplemented culture medium was renewed every two days. After 6 days of osteoblast differentiation, at least three independent experiments, of histological assays, were performed.

2.3.3 THP-1 cell line culture conditions

THP-1 are non-adherent human monocyte cells derived from acute monocytic leukemia (ATCC®, TIB-202™). Cells were maintained in T-25 flask with Roswell Park Memorial Institute (RPMI) medium 1640 (1X) + GlutaMAX (Gibco) complemented with 10% FBS and 1% (w/v) P/S. Cells were incubated in a humidified atmosphere, at 37 ºC, supplemented with 5% CO_2. Medium was changed every three days, and cells were sub-cultured after reaching density of approximatively 8 x 10^5 cells.

2.3.4 THP-1 differentiation assays

In order to turn non-adherent THP-1 cells into adherent cells, differentiation of THP-1 into macrophage-like cells was performed using Phorbol 12-myristate 13-acetate (PMA, Fisher Bioreagents). To determine the concentration of PMA required for stable differentiation of THP-1, cells were seeded at 1 x 10^6 in a 12-well plate, and incubated at different PMA concentrations, i.e. 2.5 ng/ml, 5 ng/ml, 10 ng/ml, 25 ng/ml, 50 ng/ml, and 100 ng/ml, for 48h. Cells were then observed in an inverted microscope (Axiovert 25, Zeiss) and optimal concentration was visually selected based on higher differentiation rate and lesser cellular toxicity.

2.3.5 HEK293 cell line culture conditions

HEK293 are adherent human embryonic kidney cells (ATCC® CRL-1573™), that were cultured in DMEM supplemented with 10% FBS, 1% L-Glu, and 1% (w/v) Pen/Strep. Cells were incubated at 37ºC with 5% CO_2 and subculture every 3-4 days in a 100 mm culture dish.

2.4 Transfection reactions

2.4.1 Transfection reaction of SaOS-2 cell line

SaOS-2 cells were seeded at 2.5 x 10^5 cells per well in a 6-well plate and incubated for 24 hours. Then, transfection reaction was performed using Lipofectamine LTX & PLUS Reagent (Invitrogen), following manufacturer instructions. Briefly, 2.5 µg of DNA were diluted into Opti-MEM (1X) Reduced Serum medium (Gibco), and PLUS Reagent. Then, Lipofectamine LTX reagent was added to the diluted DNA, and mixture was incubated at room-temperature for 5 minutes. After incubation, the DNA-lipid complex was incorporated into the cells, and finally, incubated in standard conditions. Transfection reactions for the overexpression assay were stopped at two-time points, i.e. 48 and 72 hours, whereas transfection reactions for knock-out assay were stopped after 48 hours.

2.4.2 Transfection reaction of THP-1 cell line

2.4.2.1 Chemical-based transfection

THP-1 cells were seeded at density of 5 x 10^5 cells per well in a 6-well plate and transfection was performed using Lipofectamine LTX & PLUS Reagent following same procedure as mentioned above.

2.4.2.2 Electroporation-based transfection

THP-1 cells were transfected using the Amaxa Cell Line Nucleofector Kit V (Lonza) for higher transfection efficiency, following the optimized protocol for THP-1 supplemented by the manufacturer. Briefly, 1 x 10^6 cells per samples were centrifuge at 90 x g, for 10 minutes, at room-temperature. Pelleted cells were then carefully resuspended in a room-temperature Nucleofector Solution, further combined with 0.5 µg pX459+oligo1 (or pX462+oligo2A and pX462+oligo2B, 1:1), and transferred into a certified cuvette. Electroporation was performed using the program V-001 (for high expression level) in the Nucleofector II device (Lonza). Then, pre-equilibrated THP-1 culture medium was added to the cuvette and the mixture was finally gently transferred into a 12-well plate (previously prepared with THP-1 culture medium). The pmaxGFP (0.5 µg/µl, Amaxa) positive control vector was transfected into THP-1 cells under the same conditions as described above and served as a positive control for transfection efficiency.

2.4.5. Selection of transfected cells and isolation

2.4.5.1 Selection of transfected cells for knock-out assay

In order to select transfected cells, SaOS-2 and THP-1 cell lines were treated with 2 µg/ml puromycin 48 hours after transfection reaction and were incubated for more 48 hours. Then, medium was discarded, cells were washed with PBS 1x and incubated in respective culture medium.

2.4.5.2 Isolation of clonal cells

After selection reaction, SaOS-2 were allowed to grow for approximatively three weeks in order to form clones. Culture medium was replaced every 2-4 days. Then, each clone population was isolated with plastic rings (edge of 20 µl tips), using purified and sterilized Vaseline as a glue, in order to fix the rings to the plate and enable trypsin from

licking. Cells were detached with trypsin and transferred into an individual well in a 24-well plate.

Since THP-1 cells are non-adherent cells, serial dilutions were performed in order to isolate cells. Therefore, cells were counted in a Nuebauer chamber and diluted in the appropriate culture medium at a density of 1 cell per 1.9 cm^2. Cells were plated in a 24-well plate and incubated for three weeks. Fresh culture medium was added once a week.

2.4.6 Transfection HEK293 cell line

2.4.6.1 Transient transfection assay

Transient transfection assays were performed using X-treme GENE HP transfection reagent (Roche). HEK293 were seeded on a 24-well plate at density of 5 x 10^4 cells per well and incubated for 24 hours. To measure the functional levels of promoter fragments, cells were primarily transfected with F1 or F2 constructs following manufacturing instructions. Briefly, 250 ng pGL3 construct is added to a mix containing 1 µl transfection reagent X-treme GENE HP, 5 ng pR-Null (*Renilla* plasmid) and 94.5 µl DMEM without supplement. As positive and negative controls, HEK293 were also co-transfected with pGL3-control and pGL3-Basic plasmids (Appendix IV, Figure A4.6 and A4.7), respectively.

For regulation analysis of each designed promoter, selected transcription factors were co-transfected with F1 or F2 (human) or FA (zebrafish)constructs into HEK293, following the same transfection assay described above. Empty expression vectors of each transcription factor were co-transfected together with F1 or F2 and used as control.

2.4.6.2 Measurement of luciferase activity

The Firefly & *Renilla* Luciferase Single Tube Assay kit (Biotium) was used to study the gene regulation of transfected cells by measuring the firefly luciferase activity. The *Renilla* luciferase activity was also measured as an internal control. Briefly, cells were washed with cold PBS and lysed in 1X Passive Lysis Buffer 2.0 (Biotium). Then, plates were rocked for 15 minutes, at room-temperature. Lysate were collected into a sterilized 1.5 ml microcentrifuge tube, cleared by centrifugation, and finally plated in a 96-well flat bottom white plate (Greiner). In this assay, firefly and *Renilla* luciferase activity were measured in the same sample: first the firefly luciferase activity was read,

and then *Renilla* luciferase was added to quench firefly luciferase activity and measured *Renilla* luciferase activity in a microplate reader (Bio-rad Benchmark).

2.5 Histology

2.5.1 XTT assay

The cytotoxic effect of puromycin on SaOS-2 and THP-1 cell lines was evaluated using a Cell Proliferation Assay XTT kit (AppliChem) following the manufacturer's protocol. This kit employs the 2,3-Bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide salt (XTT) that is reduced by mitochondria, forming an orange colored water-soluble dye. The concentration of the dye is directly proportional to the number of metabolically active cells, i.e, living cells. Briefly, SaOS-2 and THP-1 cells were plated at 1 x 10^4, and 2.5 x 10^5 cells per well, respectively, in a 96-well plate. Due to the difference in properties and sensibility of each cell line, cells were incubated with different concentrations of puromycin, *i.e*. 1 mg/ml, 1.5 mg/ml and 2 mg/ml for THP-1, and 1 mg/ml, 2 mg/ml, 2.5 mg/ml, 3 mg/ml, 3.5 mg/ml and 4 mg/ml for SaOS-2, for 48 hours, at 37ºC, in standard conditions (humid atmosphere enriched with 5% CO_2). After, 50 µl of XTT were added into each well and then incubated for more 2 hours, at 37ºC. Absorbance was read at 460 nm, and reference absorbance (to measure non-specific readings) was measured at a wavelength of 630 nm in a microplate reader.

2.5.2 Calcium mineral assay

Alizarin red S (AR-S) staining was used to define the extent of calcium deposit of differentiated osteoblast. Calcium binds to alizarin red staining by a chelation process, forming the calcium-AR-S complex which is a birefringent that appears red under light. Cells were fixed in 4% (v/v) formaldehyde (FA, Appendix II, Table A2.1) for 20 minutes at 4ºC and stained with 40 mM AR-S solution pH 4.2 (Appendix II, Table A2.1) for 15 minutes, at room-temperature, under gentle agitation. After removal of excessive staining, red nodules were observed under an inverted light microscope and pictures of each well were taken with a camera (Canon G7). Cells were distained in 10% cetylpyridinium chloride for 15 minutes under agitation and extract were collected in a 96-well plate to measure absorbance at 565 nm in a microplate reader for calcium deposits quantification. The same staining assay was performed in undifferentiated SaOS-2 parental cells and

SaOS-2-KO-cloned (seeded in the same conditions as differentiated cells) and served as a control for the establishment of basal calcium deposits.

2.5.3 Phosphate mineral assay

Von Kossa staining detects deposit of calcium. The method uses the ability of the staining to transform calcium salts into silver salts. The silver solution binds to the phosphate with silver ions, producing silver salts. The newly formed salts are reduced photochemically, and unreduced silver is removed by a sodium thiosulfate solution. Briefly cells were first fixed in 4% (v/v) FA at 4ºC for 20 minutes, and then stained with 5% (w/v) silver nitrate solution ($AgNO_3$, Appendix II, Table A2.1) for 30 minutes, under UV light. The staining was fixed with 2.5% (w/v) sodium thiosulfate solution ($Na_2S_2O_3$) at room-temperature for 5 minutes, and black nodules were observed under an inverted light microscope and capturied as photograph with a camera (Canon G7). Again, the same staining assay was performed in undifferentiated SaOS-2 parental cells and SaOS-2-KO-cloned (seeded in the same conditions as differentiated cells) in order to obtain basal level of phosphate deposits.

2.5.4 Collagen detection assay

Collagen was detected by using the Sirius red staining in both mineralized and unmineralized cells. The sirius red is an acidic hydrophilic staining that colors collagen fibers in red, by tight-binding of the stain sulfonic acid groups with the basic groups of collagen fibers. In resume, the cells were fixed in Bouin's fluid (Appendix II, Table A2.1) for 30 minutes at room-temperature and incubated with Sirius red staining solution (Appendix II, Table A2.1) for 1 hour. Unbound dye was removed by extensive wash with hydrochloric acid solution (HCl, 0.01 N) and cells were observed under an inverted light microscope. Then, bounded dye was re-suspended in sodium hydroxide solution (NaOH, 0.1 N) and re-suspended dye solution was transferred in a 96-well plate for optical density (OD) measurement at 565 nm in a microplate reader.

2.5.5 Alkaline phosphatase assay

Alkaline phosphatase activity was measured using the p-nitrophenyl phosphate (p-NPP) substrate, through a single-point spectrophotometric assay. The phosphatase catalyzes the hydrolysis of p-NPP to p-nitrophenol, a chromogenic product that gives a bright yellow color to the solution and absorbs maximally at 405 nm. Cells were first

fixed in 4% FA for 20 minutes at 4ºC, and then 5 mM p-NPP in reaction buffer was added to each well for 30 minutes at room-temperature. The reaction was stopped by adding 0.5 M NaOH solution and final solution was transferred into a 96-well plate to measure absorbance at 405 nm (pNP product peak of absorption) in a microplate reader. Undifferentiated cells underwent the same procedure and OD measurements were subtracted to the OD measured in differentiated cells.

2.6 – Statistical Analysis

Statistical analysis was performed with GraphPad Prism 7 (GraphPad, La Jolla, CA). One-way ANOVA followed by Tukey's post-hoc test was used for comparisons between several groups. A two-tailed Student's t-test was performed for comparison of two groups with less than three value. Differences were considered statistically significant when $p < 0.05$.

CHAPTER III – RESULTS

3.1 In silico analysis of ZNF687

3.1.1 Structure analysis of wild-type *ZNF687*

As mentioned earlier, very few is known about ZNF687 protein. Indeed, there is still no available structural information known for this protein, and its function is still poorly understood. Thus, the primary objective of this work was to try to understand how the mutations in *ZNF687*, identified by Divisato *et al.,* (2016) could affect the protein and lead to a metabolic bone disorder. Therefore, we first performed an *in silico* analysis in order to produce a putative secondary structure of ZNF687. The protein sequence of ZNF687 (NP_065883.1) was retrieved from NCBI and submitted to several online tools such as Predict protein, PSIPRED, Spider2, and Raptor X, among others. The data obtained from each program was compared and combined to further design a schematic representation of the putative secondary structure of the protein (Figure 3.1). We can observe that the protein presents several small α-helixes and β-sheets, which appear to be mostly coupled. These coupled motifs might represent the classical C2H2 motifs of the protein.

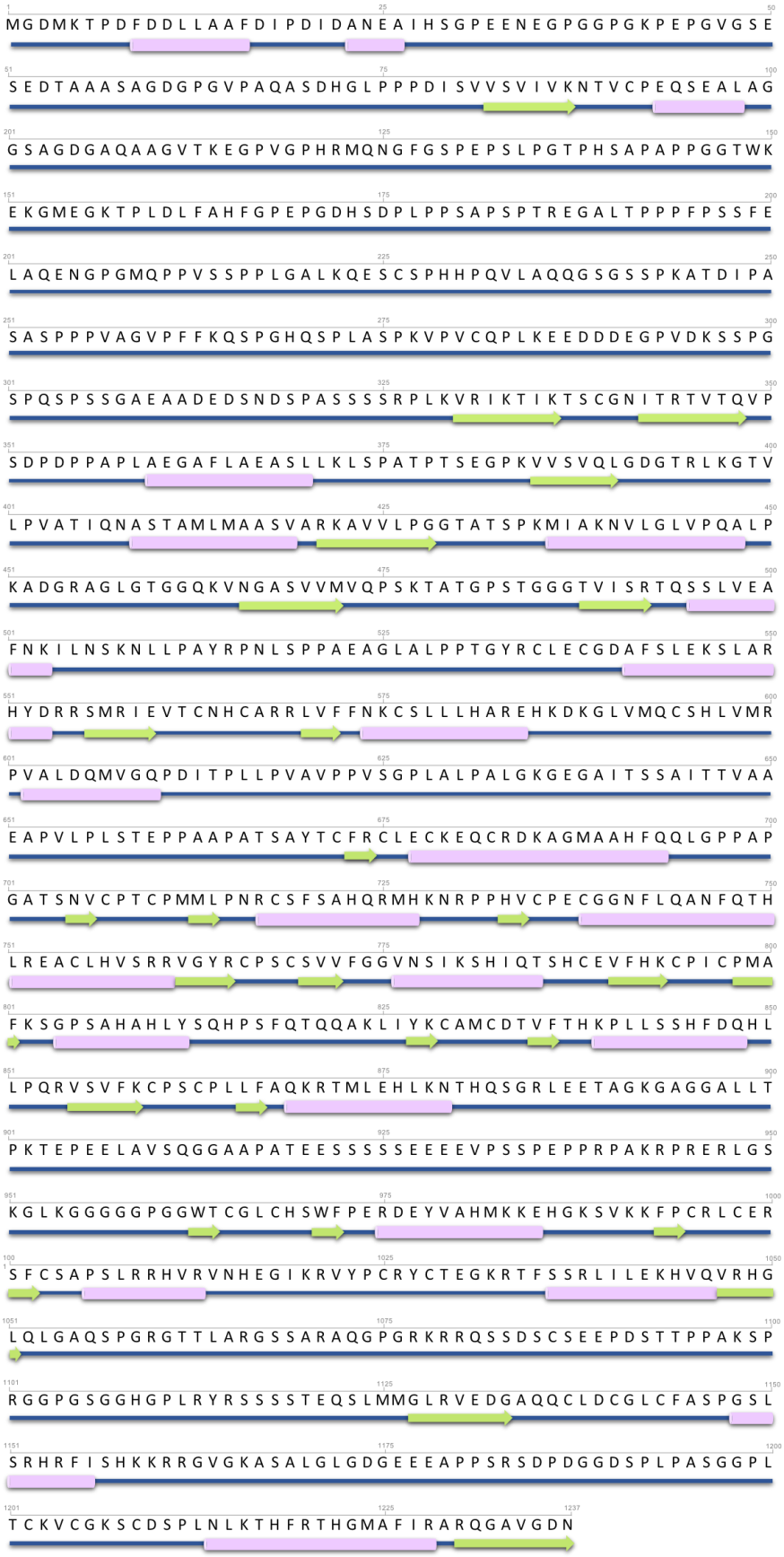


Figure 3.1 - Schematic representation of ZNF687 secondary structure. Each letter stands for an amino acid residue, and its respective position is indicated in grey. The dark blue lines represent a super coiled region, light purple cylinders

feature α-helixes and light green arrows illustrate β-sheets. Data obtained through Predict Protein, PSIPRED, Raptorx, Spider[2], NetSurfP, APSSP2 and SABLE.

To confirm such hypothesis, we performed another analysis that allows us to identify functional domains of ZNF687 by submitting the protein sequence to several online programs (Pfam, Prosite, Uniprot, SMART, and conserved domains NCBI). The only functional domain found in our protein of interest was C2H2 zinc finger domain. We compared and compiled all data obtained and have identified 14 zinc finger domains in the human ZNF687. All these domains are located in the second half of the protein sequence, and form at least three distinct zinc finger (ZF) clusters: one nearer the N-terminal (ZF1-2), another around the middle of the protein (ZF3-9) and another near the C-terminal (ZF10-12). The last two ZF motif (ZF13 and ZF14) are hard to say if they act as a cluster or individually. We also have submitted ZNF687 sequences of different mammals to the same conditions, to perform a comparative analysis. Results showed that all of the C2H2 domains, found in the human protein, are preserved between mammal species, and the same cluster pattern is maintained (Figure 3.2).

Figure 3.2 – C2H2 zinc finger domains in ZNF687 protein among different mammals. The zinc finger motifs are organized in three clustered that are conserved between mammal species. Protein sequences are linearly represented by a plain black line and zinc finger domains are illustrated in boxes. Zinc finger motifs were not identified with the same occurrence, therefore the more recurrent are represented in darker colour, while the less recurrent are illustrated in lighter colour. Protein sequences and zinc finger motifs are at scale. Data obtained through Pfam, Prosite, SMART, Uniprot and Conserved Domains NCBI.

Moreover, when we combined the results obtained from the secondary structure and the domain analyses, we can observe that almost all the β-sheets and α-helixes present from 540 to 1247 amino acids (aa) are associated to C2H2 zinc finger domains, which confirms our previous hypothesis (Figure 3.3).

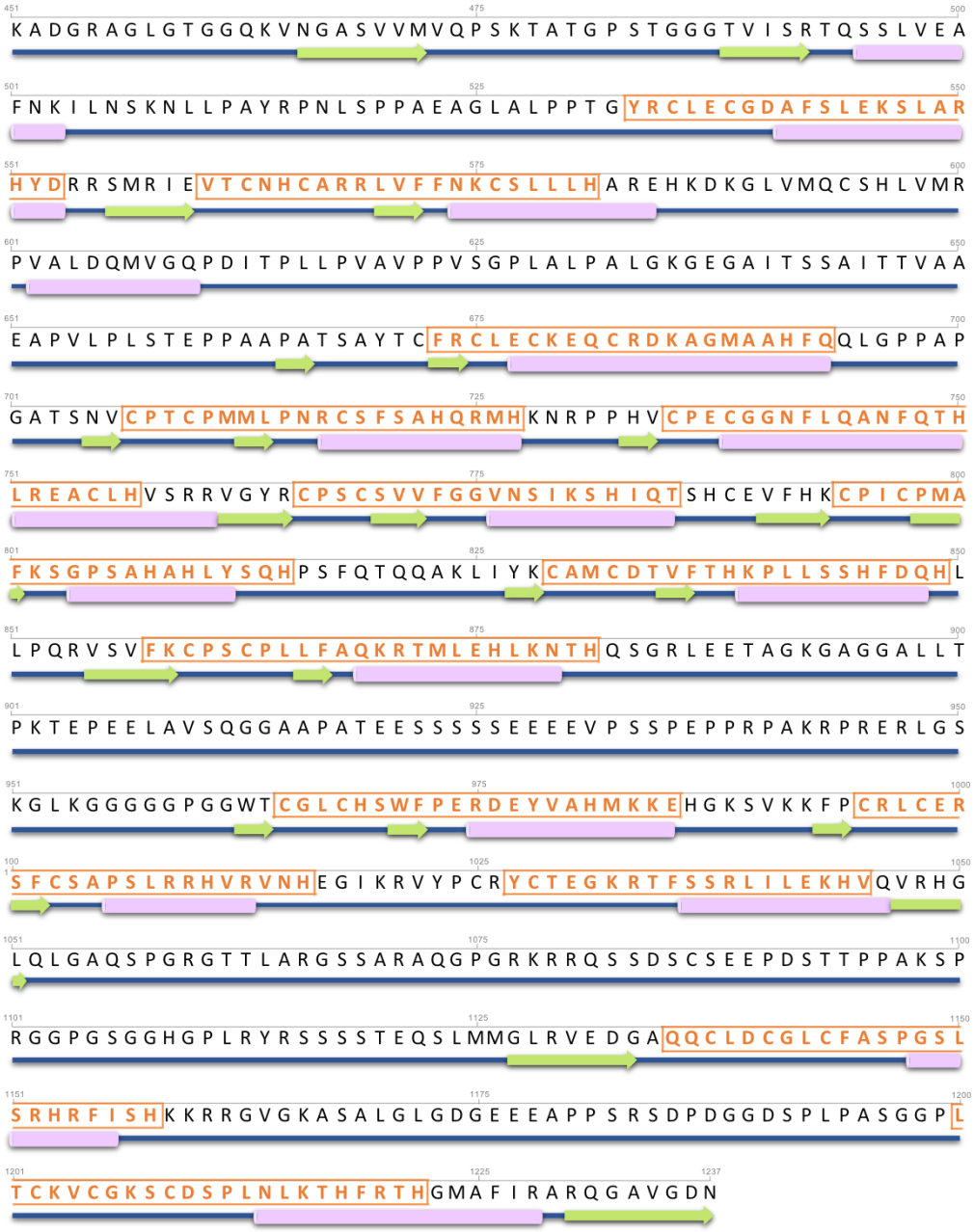


Figure 3.3 – Representation of half of ZNF687 protein sequence (501-1237 aa) depicting its secondary structure features and C2H2 zinc finger domains. The dark blue lines represent a super coiled region, light purple cylinders feature α-helixes and light green arrows illustrate β-sheets. Sequences C2H2 zinc finger domains are represented in orange. Scale is illustrated in grey.

Since ZNF687 is a transcription factor and is mostly localized in the nucleus, we search for putative nuclear signal localizations. There are two different kind of classical NLS: the monopartite, that is constituted of single cluster of basic amino acids, and the bipartite that possesses two clusters of basic residues separated by approximately 10 amino acids.[93] For its part, the monopartite comprises two classes: class 1 is represented by at least four consecutive basic residues, while class 2 only possesses three basic amino acids. These are based on putative consensus sequences such as PKKKRKV, and

K(K/R)X(K/R), for class 1 and 2, respectively. Nonetheless, not all experimentally identified NLS sequences comply to the consensus sequences. The NLS of our protein of interest were assessed by NucPred, NLStradamus and NLS Mapper and results showed that ZNF687 possesses three monopartite putative NLSs located in the end of the protein (Figure 3.4). The first monopartite NLS encountered in the ZNF687 protein, localized at position 938, correspond to the NLS identified by Divisato *et al.* (2016), possesses seven basic residues, and appears to be well preserved in mammal species (Figure 3.5).[38] The second monopartite NLS, placed at position 1071, belongs to the class 1, and present perfect identity with other mammal species. Finally, the last monopartite NLS, localized at position 1154, also belonging to class 2, is the weaker NLS, i.e. lower score, but is well conserved. Nonetheless, not all databases depict the same sequence, as some do not include the initial part, i.e RFISH, which, actually, has previously been identified as being part of a zinc finger domain (Figure 3.4).

Figure 3.4 – Representation of the three identified NLS in the human ZNF687. A third of the protein sequence is represented in this figure. NLS are characterized in blue and zinc finger motifs are boxed in orange. Scale is illustrated in grey. Data obtained through NLS mapper, NLStradamus and Nucpred.

Figure 3.5 – Evolutive conservation of the three putative NLS in ZNF687. Bold letters represent the NLS of the human ZNF687. NLS identified in other species are underlined. Blue highlight represent identity between human

residues (in the NLS region) and other species. Data obtained through Clustal Omega, NLS Mapper, NLStradamus and Nucpred.

3.1.2 Protein-protein interaction

We have also evaluated the interaction of ZNF687 with other proteins through STRING database. Results showed that ZNF687 interact with proteins that are involved in transcriptional regulation (GATAD2A, ZMYND8, INTS1, INTS3, INTS5, INTS6, and ZNF592), cell-cycle progression (TSPYL2, INTS3 and WDR26), gene regulation (ZNF687, ZMYND8, WDR26), and mono-ubiquitination (UBE2O) (Figure 3.6, and Appendix VI, Table A6.1).

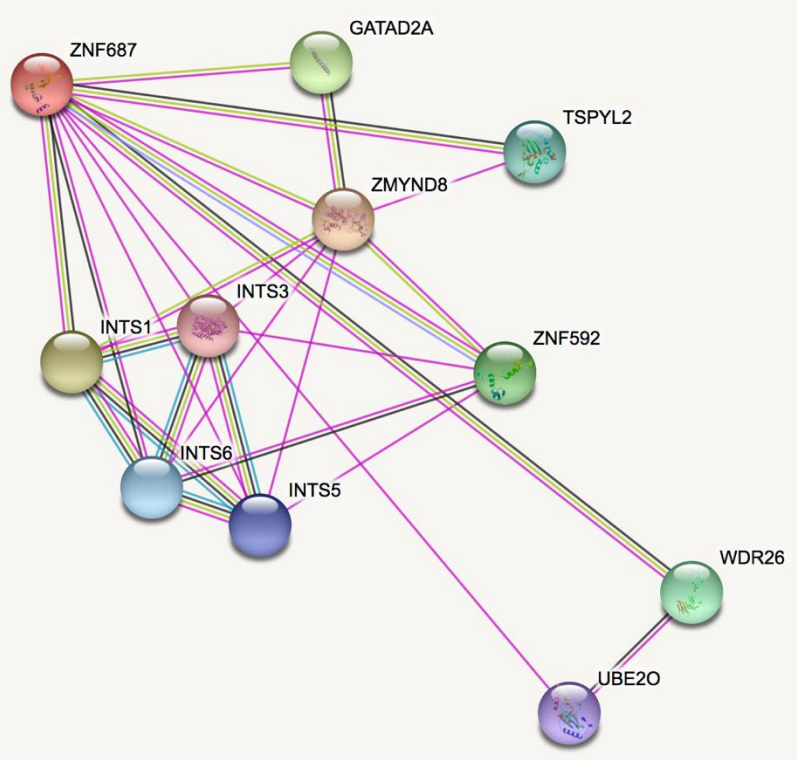


Figure 3.6 – Protein-protein interaction of ZNF687. Known interactions obtained experimentally or from curated databases are represented in pink and blue lines, respectively; interactions predicted by text-mining computational methodologies are represented in green lines, co-expression protein-protein interactions are featured in black lines, and homology-based prediction of interaction between proteins are illustrated in purple lines. Figure was obtained from String.

3.1.3 ZNF687 mutations analysis

To date, four mutations on ZNF687 have been reported and have been associated to PDB.[38,43] To visualize and understand if each point mutation could affect the protein structure and function, we analysed each mutant protein sequence using the mutant analysis server HOPE. Since there is no solved 3D-structure and modelling template,

reports obtained evaluate the modifications on the residues, contact made by the mutated residues, and structural domains in which the residue is located.

The mutations located in the residue 242 changes a serine into an isoleucine (Figure 3.7). The mutant residue is bigger and more hydrophobic than the wild type residue and may contribute to a slightly destabilized local conformation. Indeed, this increase in hydrophobicity might engender a loss of hydrogen bond which can disturb the correct folding of the protein. No functional domains are known at the location of this mutation.

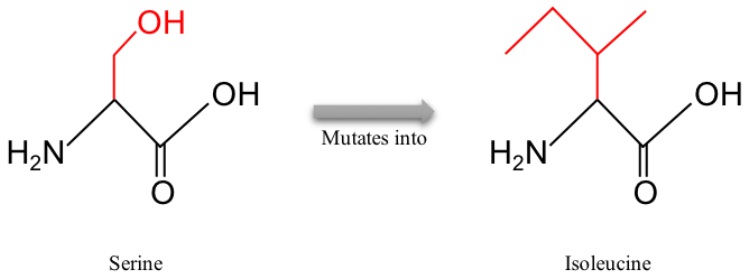


Figure 3.7 – Schematic structure of the wild-type (left) and mutant (right) residues in the p.Ser242Ile mutation. The side chain of each amino acid is represented in red whereas the backbone is illustrated in black. O: oxygen, H: hydrogen, N: nitrogen.

The mutation found in residue 665 results in a substitution of a proline into a leucine (Figure 3.8). The mutant residue is bigger than the wild-type, but both of them are neutral amino acids. Moreover, the proline residues are known to confer rigidity to the structure, sometimes forcing the backbone to fold into a specific conformation. Thus, the amino acid substitution might disturb the local structure. Again, no functional domains are known at the local of mutation.

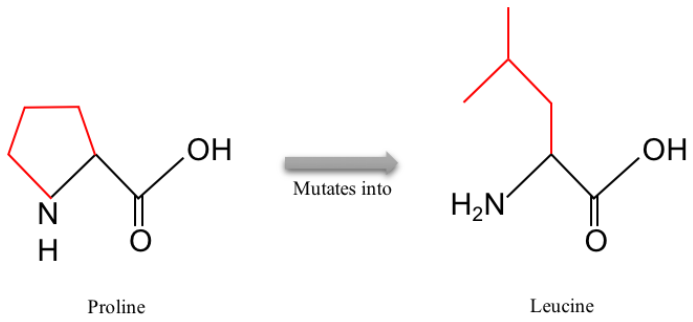


Figure 3.8 – Schematic structure of the wild-type (left) and mutant (right) residues in the p.Pro665Leu mutation. The side chain of each amino acid is represented in red whereas the backbone is illustrated in black. O: oxygen, H: hydrogen, N: nitrogen.

The mutation located at position 784 leads to an amino acid alteration of a glutamine into an acid glutamic (Figure 3.9). The mutant residue possesses a negative charge while the wild-type residue is neutral. The introduction of a negative charge might cause repulsion of ligands or other residues with the same charge. Besides, the mutation is located in a zinc finger domain and the differences in residues properties might disturb DNA binding.

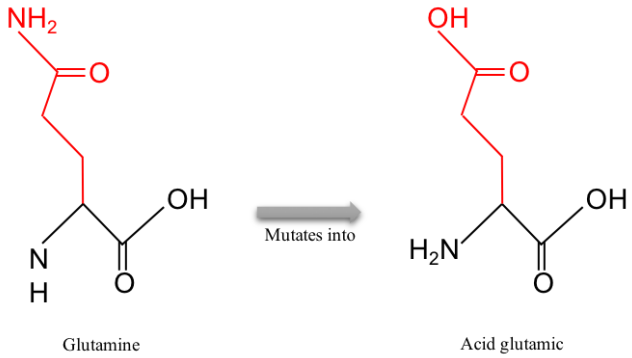


Figure 3.9 – Schematic structure of the wild-type (left) and mutant (right) residues in the p.Glu784Gln mutation. The side chain of each amino acid is represented in red whereas the backbone is illustrated in black. O: oxygen, H: hydrogen, N: nitrogen.

The mutation encountered in position 937 results in the substitution of a proline into an arginine (Figure 3.10). As mentioned earlier, the proline is a neutral residue and confers rigidity to the structure. This nonpolar amino acid tends to cluster its side chain inside the protein. On the other hand, the mutant residue is bigger in size and possesses a positive charge in its side chain that is stabilized by resonance. This hydrophilic amino acid is mostly found in the surface of the protein and can engaged ionic bonds, through electrostatic attraction. The differences in properties between these two residues might affect the interactions and the structure of the protein. Moreover, the mutation is found right before a putative identified NLS and might affect the translocation of the protein into the nucleus, as suggested by Divisato *et al.* (2016).[38]

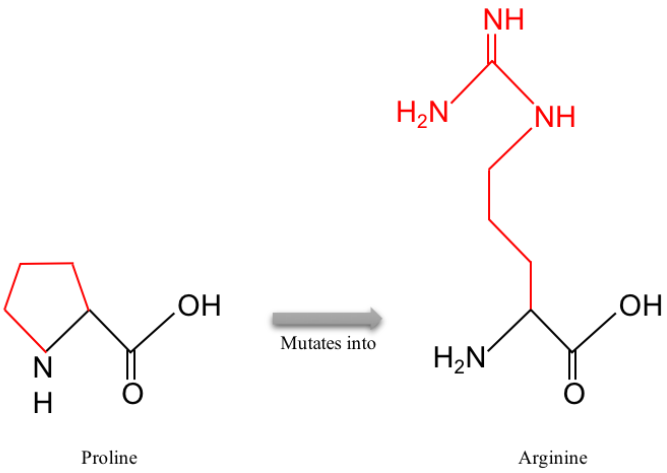


Figure 3.10 – Schematic structure of the wild-type (left) and mutant (right) residues in the p.Pro937Arg mutation. The side chain of each amino acid is represented in red whereas the backbone is illustrated in black. O: oxygen, H: hydrogen, N: nitrogen.

We have also searched for possible alterations, induced by each mutation, in the secondary structure. Therefore, we have submitted each mutated sequence to the same databases used in the secondary structure analysis of the wild-type ZNF687. After comparing results of each mutated sequence to the wild-type sequence of the protein, the putative secondary structure remains the unaltered, as no modification were observed (Figure 3.11).

In order to understand if the mutation p.Glu784Gln disturbs the zinc finger domain in which it occurs, we have submitted the mutated protein sequence to several protein domain databases and compared results with the wild-type protein. Interestingly, the C2H2 zinc finger domain, located at position 766-785 in the wild-type protein was not determined in the mutated protein by the database Prosite. Nonetheless, this region was defined as a C2H2 zinc finger domain by other databases (SMART and Conserved Domains NCBI) (Figure 3.11c).

Finally, since the mutation p.Pro937Arg befalls right before a putative NLS, we decided to assess if the mutant residue was able to disrupt this domain. Hence, the mutated sequence was submitted to NucPred, NLStradamus and NLS Mapper. Results revealed a stronger NLS, where the mutated residue was included in the NLS, in contrary to the wild-type (Figure 3.11d). The resulting sequence, **R**P**R**PA**KR**P**RR**E**R**LGS**K**GL**K**GG, possesses an additional positively-charged residue then the wild-type.

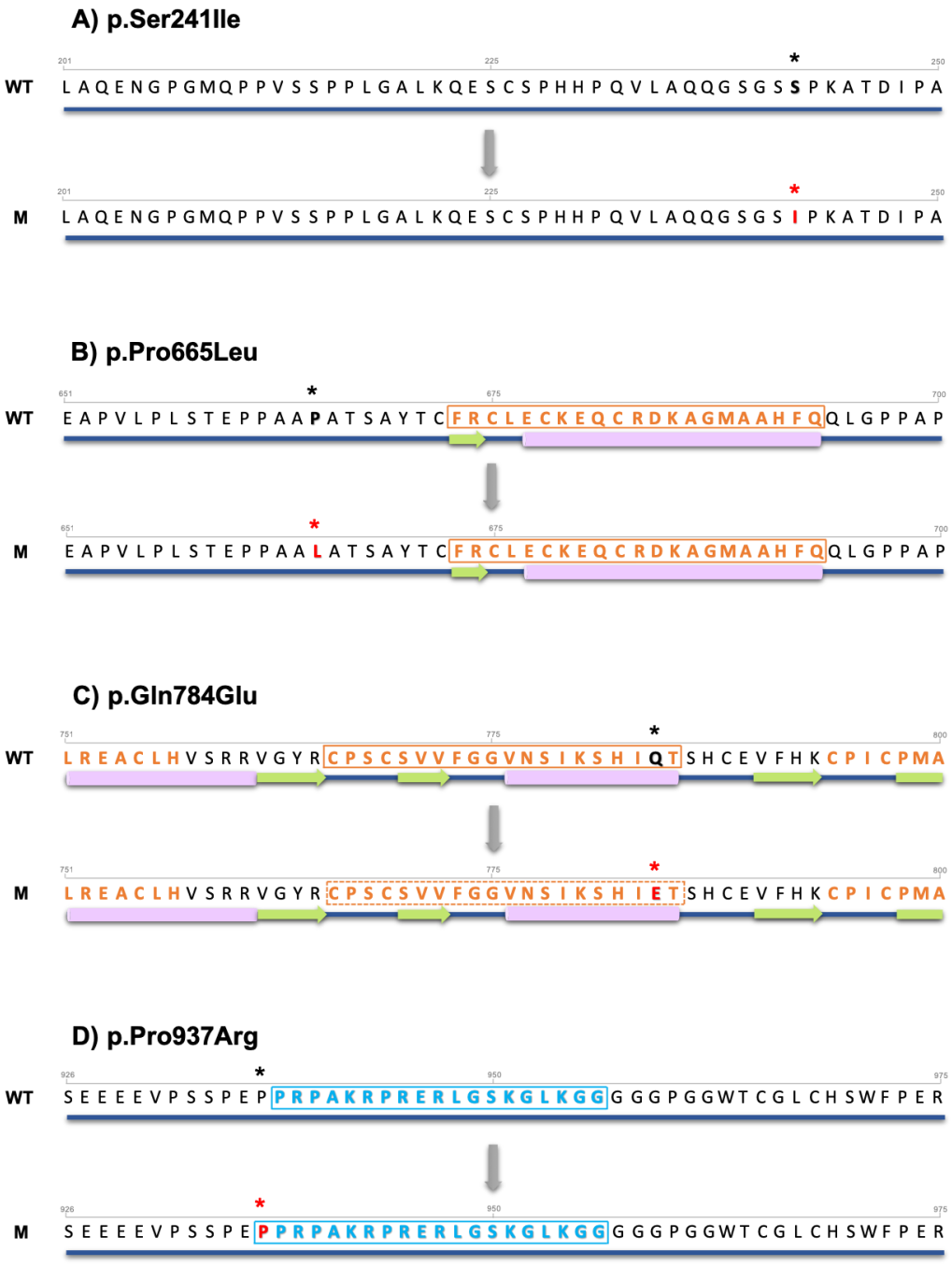


Figure 3.11 – Schematic representation of ZNF687 mutations in the protein structure. Only a portion (50 amino acids) of ZNF687 is represented for each case. Each letter stands for an amino acid residue. Super coiled region, α-helixes and β-sheets are represented in dark blue lines, light purple cylinders and light green arrows, respectively. Zinc finger domains and nuclear localization signal are characterized in orange and blue boxes, respectively. Wild-type residues that are to be mutated are featured in bold and signalled by a black asterisk. Mutated residues are illustrated in red and signalled by a red asterisk. Scales are represented in grey. WT: wild-type; M: mutant; Ser: serine; Ile: isoleucine; Pro: proline; Leu: leucine; Gln: glutamine; Glu: glutamic acid; Arg: arginine.

3.2 ZNF687 overexpression

The c.2810C>G mutation in *ZNF687* is the only mutation that has been determined as necessary and sufficient for PDB phenotype development, and patients presenting this mutation developed a more severe form of the disease. Since it has been suggested that ZNF687 might play an important role in osteoblastogenesis and osteoclastogenesis, we wanted to study the effect of an induced overexpression of ZNF687, in both osteoclastic and osteoblastic cells, and evaluate the differential expression of genes involved in bone metabolism and other putative target genes of ZNF687.

3.2.1 in vitro generation of the point mutation c.2810C>G in ZNF687

To induce the point mutation c.2810C>G in *ZNF687* we used the QuickChange Lightning Site Directed Mutagenesis Kit and the supercoiled dsDNA vector, pCMVSPORT6, containing the *ZNF687* cDNA (NM_020832.2) (kindly offered by Dr. F. Gianfrancesco; Institute of Genetics and Biophysics, Naples, Italy). Then, the mutation-containing synthesized DNA was selected by digestion with *Dpn I* and further transformed in XL10-Gold ultracompetent cells. After transformation, we performed a screening of the selected clones by digesting extracted DNA with *Not* I and *Eco*R V, to confirm the insert size. The screening products were analysed by electrophoresis on an agarose gel (Figure 3.12).

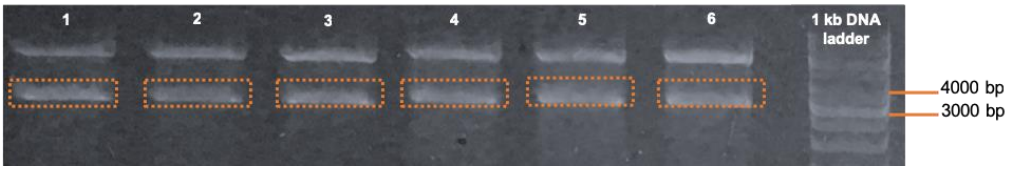


Figure 3.12 – Screening of mutated *ZNF687* (c.2810C>G). Screening of positive clones after site-directed mutagenesis. Samples were digested with *Not I* and *EcoR V*. 1-4: selected clones; 6: non-digested unmutated pcmvsport6; 7: 1kb dna ladder.

Clones from wells 1 to 4, depicted in Figure 3.12, presented the expected size of approximately 3750 bp, as *ZNF687* cDNA present in pCMVSport6. One of the clones was selected and sent for sequencing with primer ZNF687Rev (Table A.1) to confirm the presence of the mutation. Results confirmed the substitution of the nucleotide cytosine by the nucleotide guanidine (Figure 3.13).

Normal exon 6:

AACACCCATCAGTCTGGGCGCTTGGAGGAGACTGCTGGGAAAGGGGCCGGGGGTGCCC
TGCTGACCCCCAAGACTGAGCCTGAGGAGCTGGCTGTTTCTCAGGGAGGGGCAGCCCCT
GCTACTGAGGAGTCGTCTTCATCTTCAGAAGAGGAGGAAGTACCCAGCTCCCCTGAGCC
CCCCCGTCCAGCCAAACGGCCTCGGCGGGAACTAGGGAGCAAAGGCCTCAAGGGTGGG
GGTGGGGGGCCTGGAGGCTGGACCTGTGGCCTGTGTCACTCCTGGTTCCCTGAGCGTGA
TGAATACGTGGCCCACATGAAGAAGGAGCATGGCAAG

Sequencing results clone 3:

[...]AACACCCATCAGTCTGGGCGCTTGGAGGAGACTGCTGGGAAAGGGGCCGGGGGTGC
CCTGCTGACCCCCAAGACTGAGCCTGAGGAGCTGGCTGTTTCTCAGGGAGGGGCAGCCC
CTGCTACTGAGGAGTCGTCTTCATCTTCAGAAGAGGAGGAAGTACCCAGCTCCCCTGAG
CGCCCCCGTCCAGCCAAACGGCCTCGGCGGGAACTAGGGAGCAAAGGCCTCAAGGGTG
GGGGTGGGGGGCCTGGAGGCTGGACCTGTGGCCTGTGTCACTCCTGGTTCCCTGAGCGT
GATGAATACGTGGCCCACATGAAGAAGGAGCATGGCAAG[...]

Figure 3.13 – Sequencing results of the selected clone. Result of selected clone 3 sequenced with primer ZNF687Rev, after site-directed mutagenesis. The mutation c.2810C>G is located in exon 6 (NM_020832.2). The underline sequence corresponds to the designed mutagenic synthetic oligonucleotide primer. Dark bold letter corresponds to the wild-type nucleotide (c.2810 cytosine); red bold letter corresponds to the mutated nucleotide (c.2810 guanidine).

3.2.2 Analysis of relative gene expression involved in bone metabolism in overexpressed ZNF687 osteoblastic cells SaOS-2.

To understand how the mutation c.2810C>G in *ZNF687* (m*ZNF687*) affects the bone metabolism, we analysed the relative expression of several genes (Appendix VI, Table A6.2) involved in bone metabolism and in PDB, by real-time quantitative PCR, which were then normalized with two reference genes, *GAPDH* and *β-actin*.

Therefore, purified pCMVSPORT6+*ZNF687* and pCMVSPORT6+m*ZNF687* were transiently transfected in osteoblastic cell lines, SaOS-2. Transfection reactions were stopped at two time point, 48 and 72 hours. The same transfection reactions were attempted in osteoclastic cell lines THP-1, but unfortunately, these cells are quite sensitive and very difficult to transfect. As various protocols were tested, we decided to give up on transfected THP-1 and conducted the experiment only in SaOS-2. Transfection reactions were stopped at two time point, 48 and 72 hours. After RNA extraction and cDNA synthesis, RT-qPCRs were conducted for non-transfected SaOS-2, pCMVSPORT6+*ZNF687* and pCMVSPORT6+m*ZNF687*-transfected SaOS-2.

To verify the success of transfection, we first analysed the relative expression of *ZNF687*. Results demonstrated that 48 hours after transfection, a slight overexpression of *ZNF687* was visible. Nonetheless, 72 hours after transfection, *ZNF687* was expressed approximately 8-fold and 6-fold more, for pCMVSPORT6+*ZNF687* and

pCMVSPORT6+m*ZNF687*, respectively, than the wild-type SaOS-2 (i.e, non-transfected).

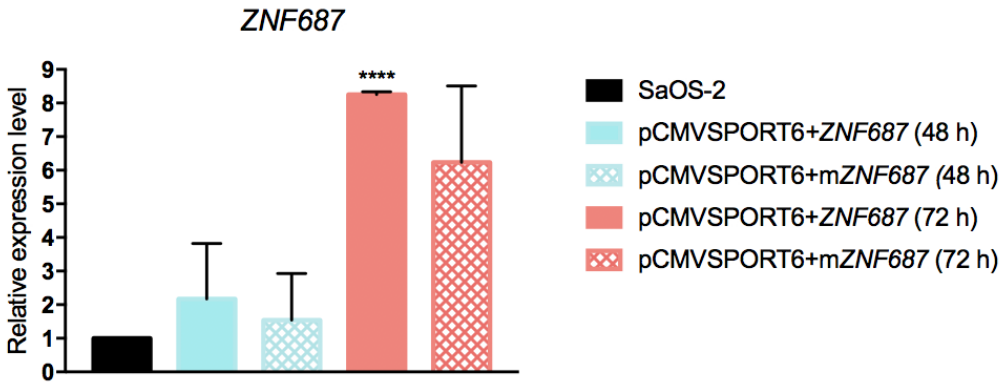


Figure 3.14 – Expression of *ZNF687* in wild-type SaOS-2 and SaOS-2-transfected cells. qPCR monitoring the expression of *ZNF687* in non-transfected (black) and transfected with pcmvsport6+*ZNF687* or with pcmvsport6+m*ZNF687* SaOS-2 cells (after 48 hours in Caribbean blue and 72 hours in coral). The gene expression levels were normalized with the mean expression of *GAPDH* and *β-actin*. The data are shown as the means ± SD from two independent experiments. Asterisks depict a response that is significantly different from the control SaOS-2 (****p< 0.0001).

Then, in order to understand the effect of ZNF687 overexpression in bone metabolism, we analysed the relative expression of several genes that are involved in bone metabolism and other putative target of ZNF687 (Appendix VI, Table A6.2). All genes selected presented an increased expression when *ZNF687* was overexpressed except for OPTN (Figure 3.15). Also, almost the same pattern persisted, in which cells transfected for 72h expressed higher gene expression levels than cells transfected for 48h. Interestingly, the expression of *RUNX2*, *RANK* and *FGF2* was extremely increased in cells expressing WT ZNF687, and a bit less expressed in mZNF687, but still with values over ZF687 basal level found in SaOS-2. Indeed, *RUNX2* appeared to be 30- to 20-fold more expressed than the control, *RANK* presented an expression 20- to 12-fold higher than the wild-type SaOS-2, and *FGF2* had its expression increased 18- to 8-fold than the control, for WT ZNF687 and mZNF687 cells. An increased in *OCN* expression has also been noticed, nevertheless, this increase was not statistically significant. We have obtained the same expression pattern in all genes, except for *OSX*, *CCDC3* and *OPTN*, where the expression was higher in transfected cells which transfection reactions had been stopped after 48 hours. *TWIST*, *OCT4* and *BMI 1* all seem to present a higher expression than the control, however neither one of them was statistically significant, as they all presented an important standard deviation.

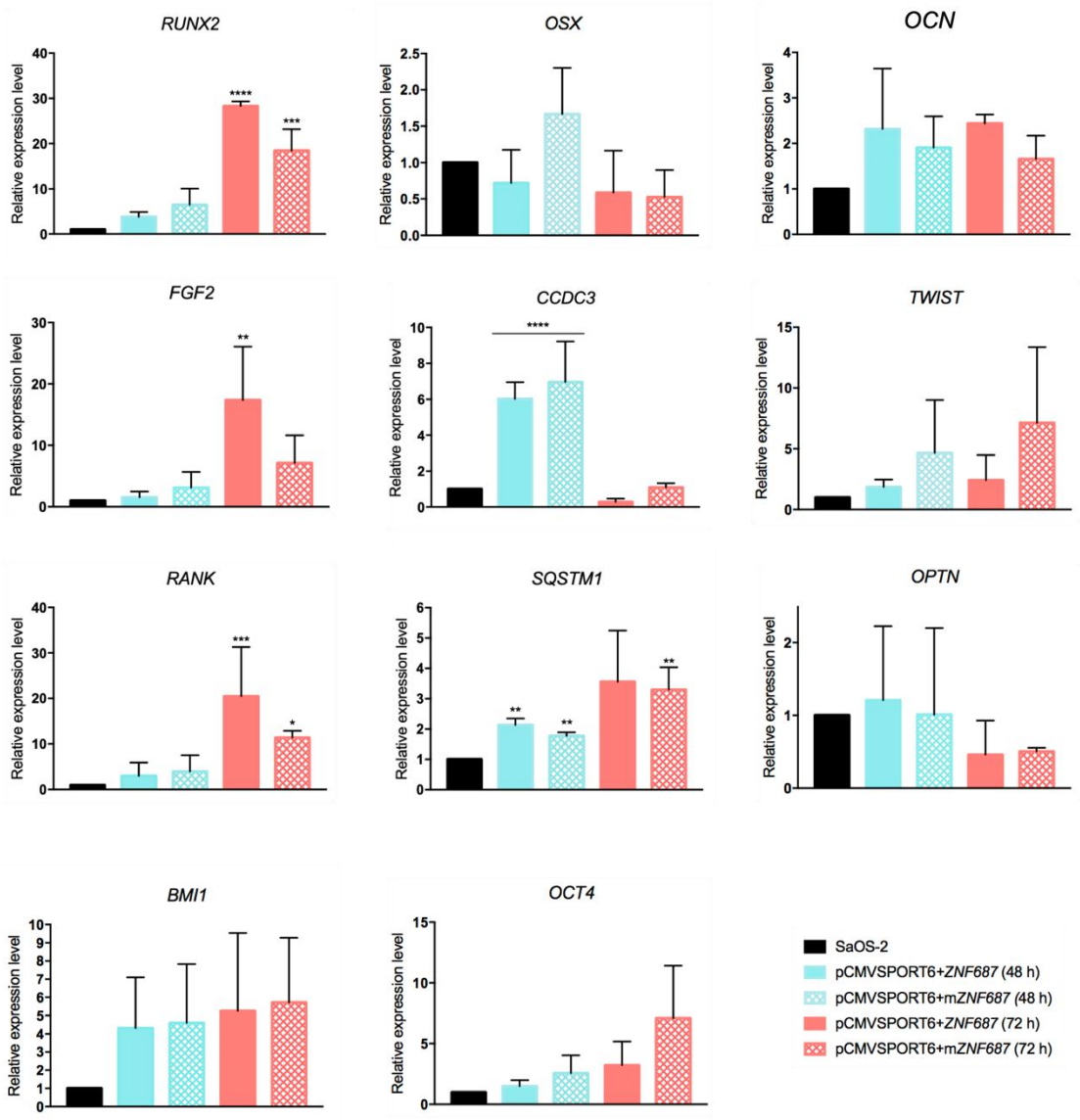


Figure 3.15 – Expression of several genes in wild-type SaOS-2 and SaOS-2-transfected cells. qPCR monitoring the expression of several genes in non-transfected (black) and transfected with pcmvsport6+*ZNF687* or with pcmvsport6+m*ZNF687* SaOS-2 cells (after 48 hours in Caribbean blue and 72 hours in coral). The genes expression levels were normalized with the mean expression of *GAPDH* and *β-actin*. The data are shown as the means ± SD from three independent experiments. Asterisks depict a response that is significantly different from the control SaOS-2 (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001).

3.3 *ZNF687* Knock-out mediated by CRISPR-Cas9

To study the effect of the *ZNF687* knock-out, we resorted to CRISPR-Cas9 genome editing system to produce an indel mutation, which leads to the appearance of a premature stop codon, and hence to a non-functional truncated protein.

3.3.1 CRISPR-Cas9/Cas9n systems Construction

To perform the indel mutation, we have constructed the CRISPR-Cas9 system following Ran *et al.* (2013) protocol. First, we have submitted a sequence, of approximatively 200 bp from the start codon (ATG) of *ZNF687* transcript, to the Oligo Optimized CRISPR Design on-line tool, in order to obtain the guide sequences. Results showed that the best oligo sequence for *ZNF687* presented a score of 86 with 84 off-target sites. The oligo sequence (oligo1), composed of 21 nucleotides, is located in the reverse strand, and its reverse complement is positioned 58 nucleotides downstream the ATG (Figure 3.16).

Figure 3.16 – Guide RNA sequence for CRISPR-Cas9 system. Result of the best oligo obtained via the oligo optimized CRISPR design on-line tool. Blue sequence corresponds to oligo1 and green letters illustrate the initiation codon.

Since the results obtained were not optimal, we decided to design a CRISPR-Cas9 nickase (CRISPR-Cas9n) system that reduces the number of off-targets. Thus, two other oligo sequences, oligo 2A and 2B, were also selected via Oligo Optimized CRISPR Design on-line tool (Figure 3.17). Independently, oligo 2A presented a score of 62 with 242 off-targets, and is located 22 nucleotides downstream the ATG, whereas oligo 2B had a score of 77 with 101 off-target sites and is located 67 nucleotides downstream the ATG. Both oligos are 20 nucleotides long. Oligo 2A and 2B presented a pair score of 47 and no off-targets. Since the score of the pair oligo is very low, the probability of success of this technique is also low.

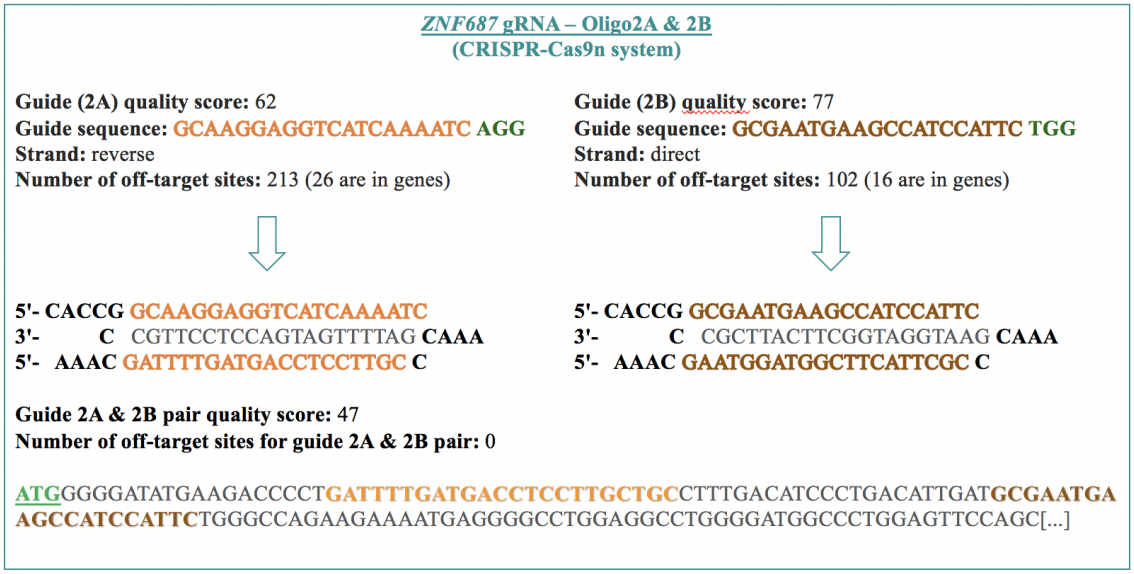


Figure 3.17 – Oligo2a and 2b sequences for CRISPR-Cas9n system. Results of the best pair oligo obtained via the oligo optimized CRISPR design on-line tool. Orange and brown letters correspond to oligo2a and 2b, respectively. Green letters correspond to the initiation codon.

Then, each pair of designed gRNAs were annealed and cloned into specific plasmids, namely pX459 and pX462, for CRISPR-Cas9 and CRISPR-Cas9n systems, respectively. To confirm the correct construction of each plasmids with its respective annealed oligo pair, pX459+oligo1, pX462+oligo2A, and pX462+oligo2B were subjected to a PCR, using primer U6 [designed in the U6 promoter of pX459 and pX463, located upstream the inserted gRNA (Figures A4.4 and A4.5)] and their specific reverse primers (reverse complementary designed gRNA for each construction) in order to amplify a DNA fragment of 285 bp. The screening products were analyzed by electrophoresis on an agarose gel (Figure 3.18).

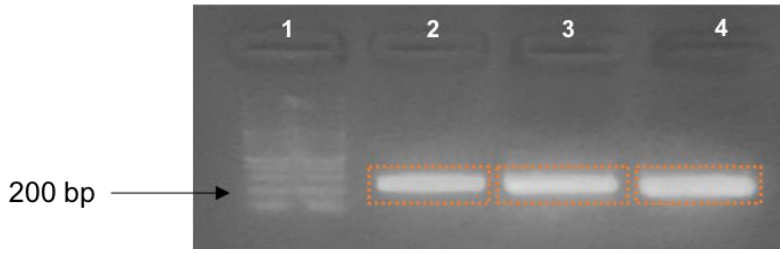


Figure 3.18 – Screening of the CRISPR amplified fragments. Amplified fragment of each CRISPR constructs were separated in an agarose gel. 1:100 bp DNA ladder; 2: amplified fragment from pX459+oligo1 construct; 3: amplified fragment from pX462+oligo2a construct; 4: amplified fragment from px462+oligo2b construct.

Results showed that all three fragments presented the expected size of 285 bp (Figure 3.17). Finally, after DNA extraction and purification, samples were sequenced, and results confirmed that each product corresponded to the correct gRNA sequence introduced in the construction (Figure 3.19).

Figure 3.19 – Sequencing results of each CRISPR constructs. Results showed that each designed gRNA was correctly inserted in its respective vector. Underlined bold letters correspond to the cleavage site of the BbsI restriction enzyme. Blue, orange and brown letters correspond to the oligo1, 2a and 2b, respectively.

3.3.2 Generation of an indel mutation in *ZNF687*

To generate a stable cell line, we have used puromycin antibiotic resistance as a positive selection marker. First, we used an XTT assay to determine the optimal concentration of puromycin to be used for the selection of transfected and un-transfected THP-1 and SaOS-2 cells. After being exposed to different concentration of puromycin for 48 hours, we observed a considerably decreased in the viability of both THP-1 and SaOS-2 cells (Figure 3.20). Indeed, at 1 $\mu g.ml^{-1}$ of puromycin, only ~30% of cells remained viable, in both cell lines. THP-1 and SaOS-2 cells presented lower viability at concentration of 2 $\mu g.ml^{-1}$, with values around 10 % and 19.5 %, respectively. Moreover, in SaOS-2 cells, concentrations above 2 $\mu g.ml^{-1}$ did not affect cell viability, probably due to the fact that cells had reached full confluency. Based on the results obtained, the concentration of 2 $\mu g.ml^{-1}$ was selected as the optimal concentration for the selection of transfected and non-transfected cells in both cell lines.

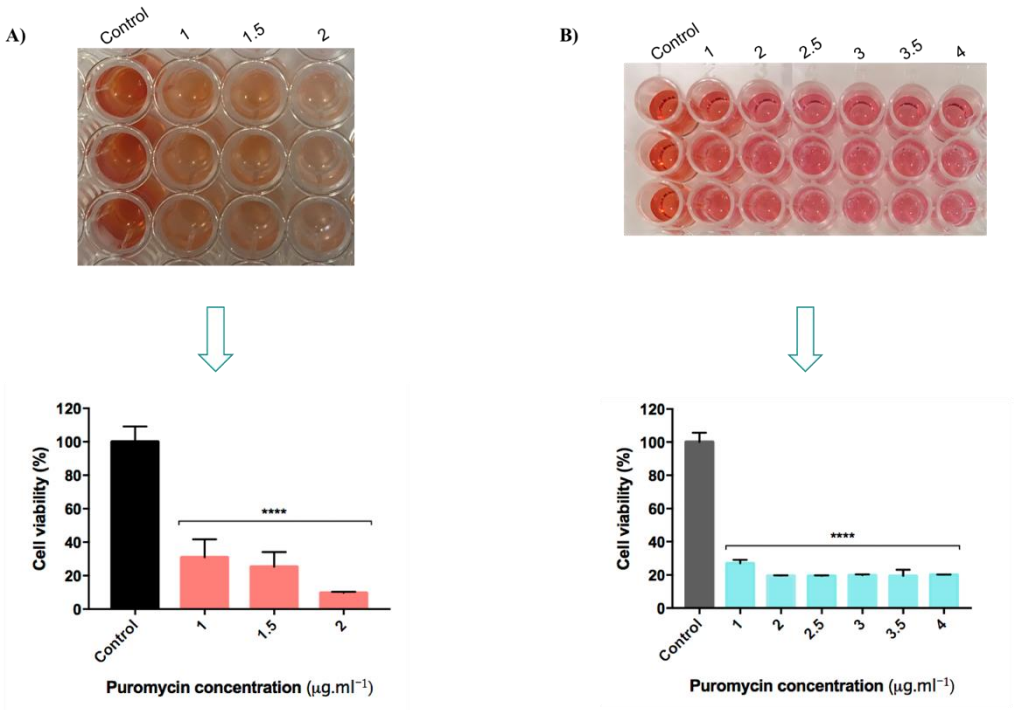


Figure 3.20 – Cytotoxic effect of puromycin on THP-1 and SaOS-2 cell lines. THP-1 (a) and SaOS-2 (b) cells were incubated with different concentration of puromycin for 48 hours. Since the cytotoxic effect of an antibiotic depends on the cell type, different concentrations of puromycin were used for each cell lines. Black and coral bars indicate puromycin-free and puromycin-treated THP-1 cells, respectively. Grey and Caribbean blue bars indicate puromycin-free and puromycin-treated saos-2 cells, respectively. Experiments were performed in triplicate and repeated two times with similar results. The data are shown as the mean ± SD. Asterisks depict a response that is significantly different from the control (****p = 0.0001).

The SaOS-2 and THP-1 cells were then transfected with pX459+oligo1 and co-transfected with pX462+oligo2A and pX462+oligo2B to generate an indel mutation in *ZNF687*. After selection with puromycin, THP-1 cells were submitted to serial dilutions and further plated in a 24-well plate. After one week, no cells could be observed in each well, and after three weeks, results remained the same. The experiment with THP-1 cells was then considered as unsuccessful and since THP-1 cells are described to be very difficult to transfect, we have then tried an electroporation-based transfection approach. To assess the THP-1 transfection efficiency we used the pmaxGFP expression vector, and after 24 hours the cells were observed under inverted fluorescence microscope (Figure 3.21). Electroporation-based transfection appeared to be a success since transfected cells could exhibit bright green fluorescence.

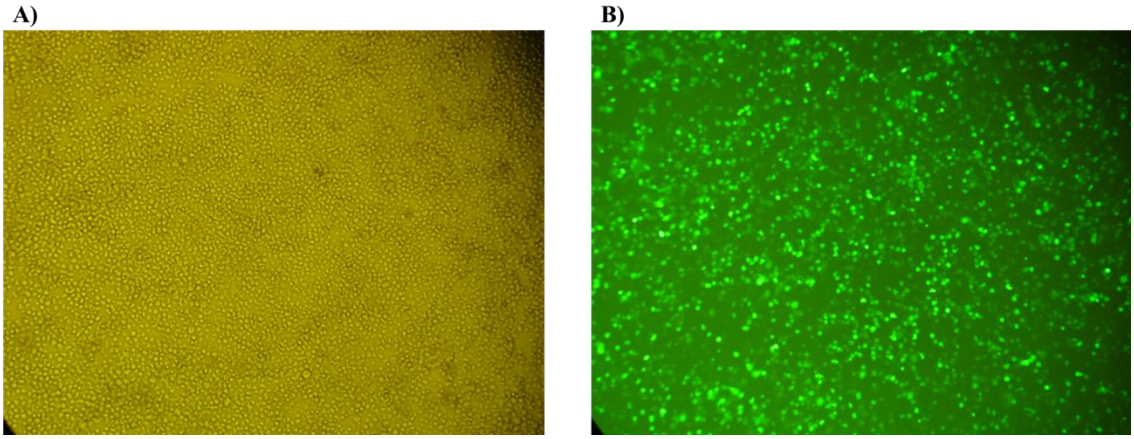


Figure 3.21 – THP-1 cells transfected by nucleofection with pmaxGFP expression vector. THP-1 cells were transfected with pmaxGFP vector and observed under an inverted microscope under normal light (a) and under ultraviolet light (b).

We then transfected the THP-1 cells with pX459+oligo1 alone or co-transfected with both pX462+oligo2A and pX462+oligo2B, using nucleofection. At the end of the transfection reaction, PMA was added to the culture medium in order to differentiate THP-1 cells into macrophage-like cells, which are adherent. This step was important to increase the efficiency of cell isolation by avoiding serial dilution. However, after puromycin treatment, most of transfected THP-1 cells appeared dead and undifferentiated (Figure 3.22).

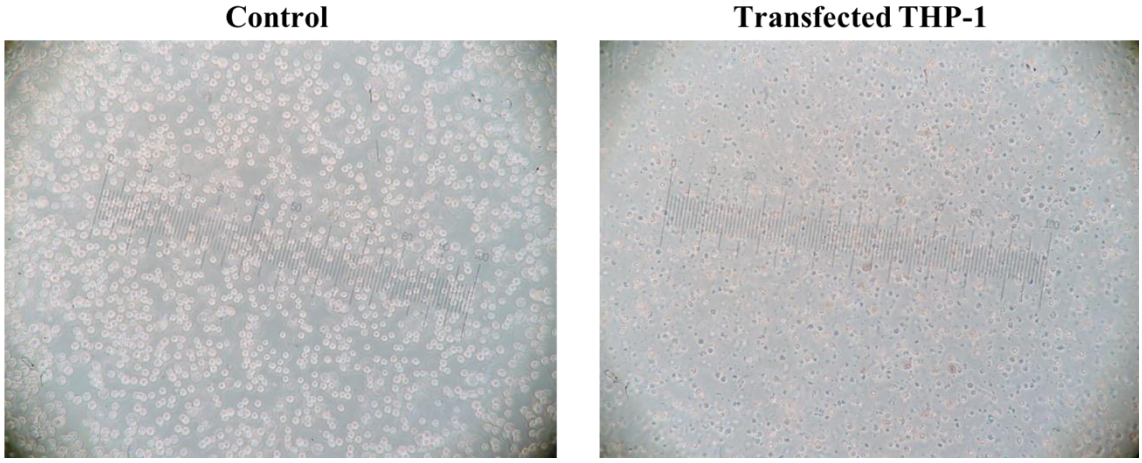


Figure 3.22 – Transfected THP-1 cells after puromycin treatment. The cells were transfected using nucleofection with pX459+oligo1 and treated with puromycin, for 48 hours. After puromycin treatment the non-transfected and untreated THP-1 cells (control) look perfectly viable whereas transfected THP-1 cells seem to be dead (transfected THP-1).

In order to optimize the transfection of THP-1 cells, we decided to first differentiate them with PMA, and, only after, proceed to the transfection reaction. Hence, cells were treated with different PMA concentrations, i.e. 2.5 $ng.ml^{-1}$, 5 $ng.ml^{-1}$, 10 $ng.ml^{-}$

[1], 25 ng.ml^{-1}, 50 ng.ml^{-1}, and 100 ng.ml^{-1}, for 48h, in order to visually determine the optimal concentration required for stable differentiation of THP-1 (Figure 3.23).

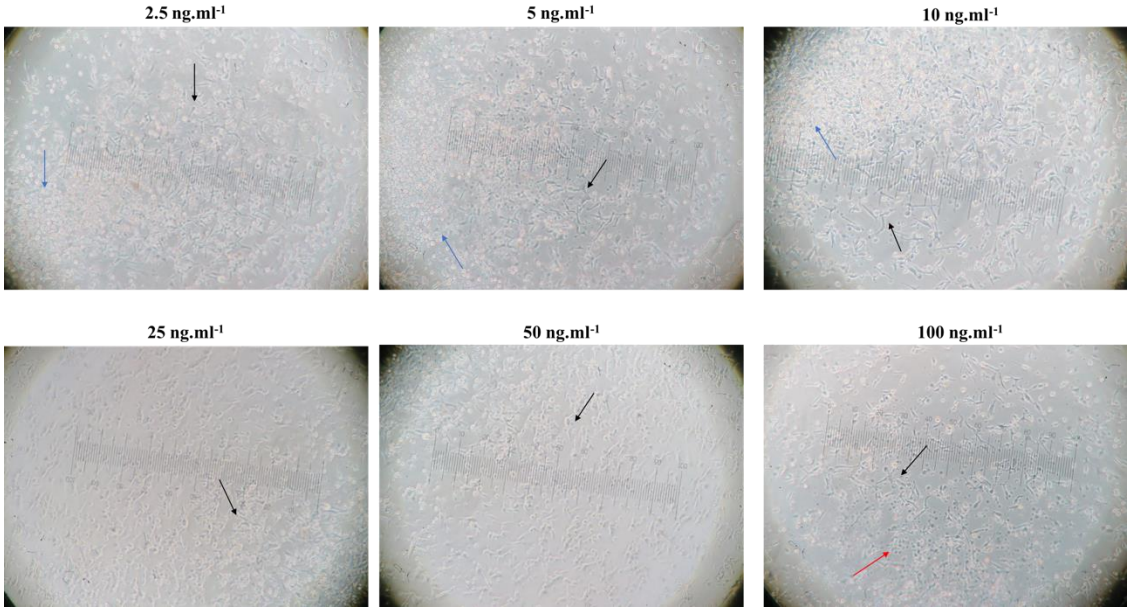


Figure 3.23 – Differentiation of THP-1 into macrophage-like cells under different concentration of PMA. THP-1 cells were incubated at different concentration of PMA, i.e. 2.5 ng.ml^{-1}, 5 ng.ml^{-1}, 10 ng.ml^{-1}, 25 ng.ml^{-1}, 50 ng.ml^{-1}, and 100 ng.ml^{-1}, for 48h. Blue, black and red arrows exemplify non-differentiated, differentiated and dead THP-1 cells, respectively.

The number of differentiated cells was proportional to the increase of PMA concentration. Indeed, at concentration of 2.5 ng.ml^{-1}, 5 ng.ml^{-1}, and 10 ng.ml^{-1}, only few cells had differentiated into macrophage-like. Nonetheless, at 100 ng.ml^{-1}, cells started to die due to the toxic effect of PMA. The differentiation rate of THP-1 at 25 ng.ml^{-1} and 50 ng.ml^{-1} seems quite similar. Therefore, to minimize the cytotoxic effect of PMA on cells, we selected the 25 ng.ml^{-1} as an optimal concentration for the differentiation of THP-1 cells.

One of the constringent about nucleofection is the density required for the reaction, which is actually hard to obtained by differentiated THP-1 into macrophage-like cells. In order to separate differentiated from undifferentiated cells, medium is discarded, and adherent cells are detached by trypsin. Hence, only a portion of the initial number of cells is harvest. This technique required a certain time to achieve, which is one of the limitation factors of this work. Therefore, the transfection of THP-1 with CRISPR-Cas9/Cas9n was abandoned, and we focused our work using the SaOS-2 cells.

Regarding SaOS-2, transfected cells were selected by puromycin, and only a very small number of viable cells remained (Figure 3.24). The result obtained after selection might reflect the transfection efficiency.

Figure 3.24 – Selection of CRISPR-Cas9 and CRISPR-Cas9n transfected SaOS-2 cells. Transfected cells were treated with puromycin for 48 hours. Compared to the control (non-transfected and untreated SaOS-2 cells), an incredible number of cells have died in transfected cells under puromycin treatment.

Since few cells remained after treatment with puromycin, instead of performing serial dilutions, cells were first transferred into a 10 mm round dish in order to scatter them and allow them to grow isolated (Figure 3.25). When clones reached a certain population, each one of them were individually transferred into another plate for harvesting for indel mutation analysis and downstream applications.

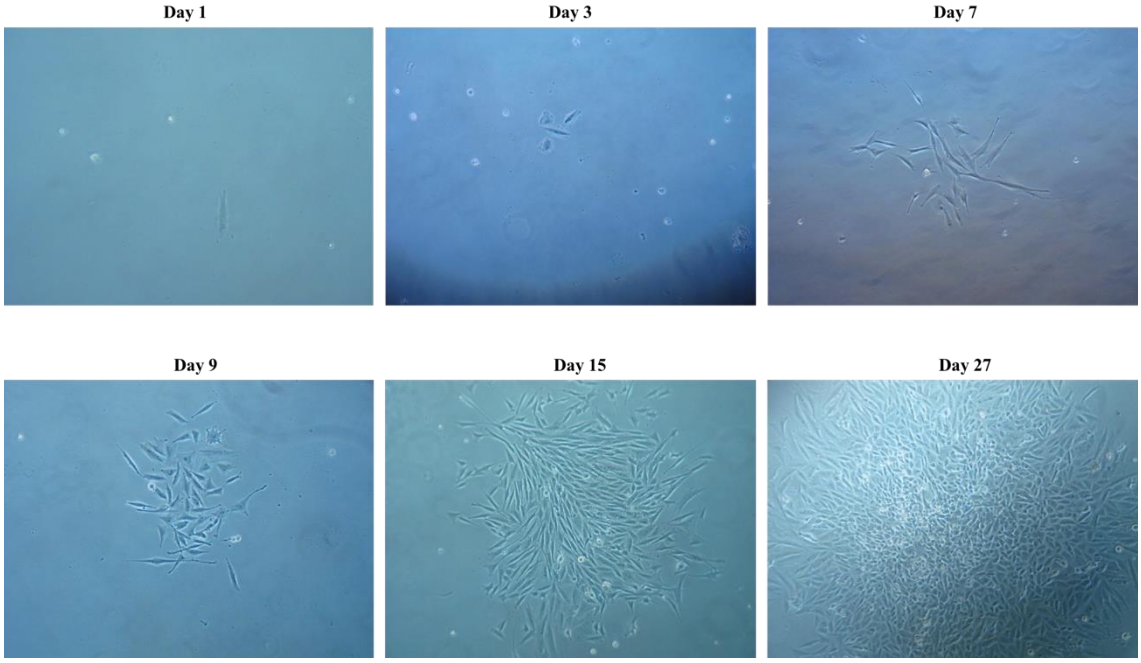


Figure 3.25 – Harvesting of SaOS-2 knock-out clones. Cells were harvest for weeks until they reached a certain confluency (day 27) that allow them to be individually transferred and grow into another plate for downstream applications.

3.3.3 Screening of the indel mutation in *ZNF687*

DNA of each isolated clone was collected and submitted to a PCR using a designed pair of primer (ZNF687mFwd and ZNF687mRev), in order to amplify a DNA fragment of 182 bp that flanks the local of the indel mutation (Figure 3.26).

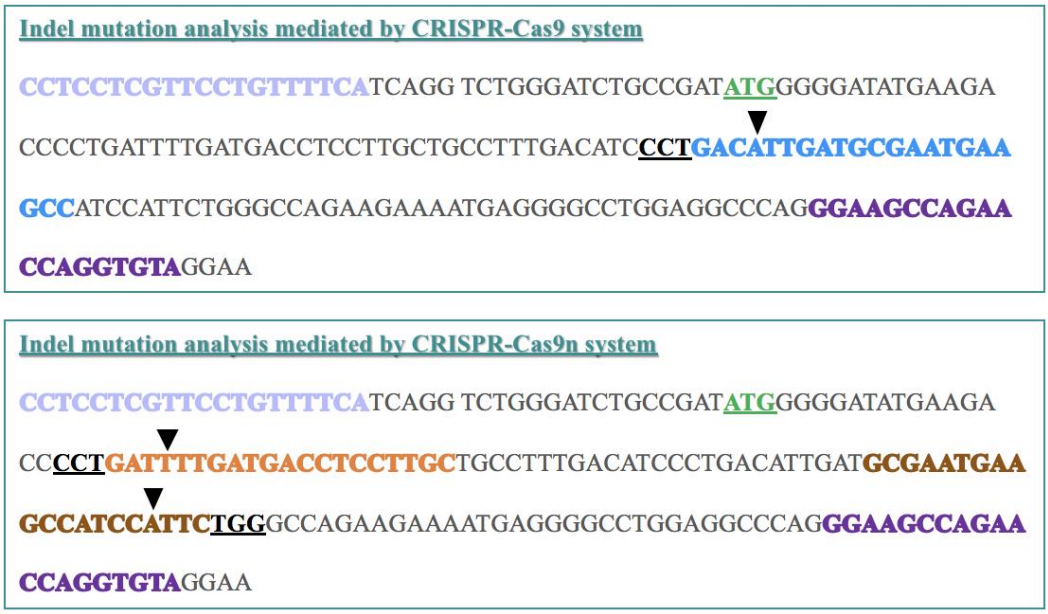


Figure 3.26 – Indel mutation analysis mediated by CRISPR-Cas9/Cas9n system. The designed pair of primer, ZNF687mFwd in light purple and ZNF687mRev in dark purple, flank the local of the indel mutation illustrated by a black triangle. Initial codon ATG is represented in green and the 5'-NGG-3' (or 5'-CCN-3') sequence is represented in bold black. Blue, orange and brown sequences depict oligo1, 2a and 2b respectively.

Each sample was further sequenced using either primer ZN687mFwd or ZNF687mRev. Unfortunately, among 24 CRISPR-Cas9n clones, no indel mutation was encountered in any of them. Regarding the CRISPR-Cas9 clones, six of the 22 clones presented an indel mutation in one or two chromosomes, resulting in a knock-down and a possible knock-out. Clones C1 and C8 harbour frameshift indels, a deletion of a thymine and an insertion of an adenine, five and four nucleotides downstream the 5'-CTT-3' sequence, respectively (Figure 3.27). By analysing the resultant sequences in both clones C1 and C8, we can predict that these frameshift indels will disrupt the reading frame by introducing premature stop codons, resulting in truncated proteins of only 46 and 20 amino acids, respectively.

Deletion of a thymine and insertion of a adenine in clones C1 and C8:

1) **ATG** GGG GAT ATG AAG ACC CCT GAT TTT GAT GAC CTC CTT GCT GCC TTT GAC ATC

CCT **GAC** **ATT** **GAT** **GCG** **AAT** **GAA** **GCC** **ATC** CAT TCT GGG CCA GAA GAA AAT GAG [...]

ATG ATG CGA ATG AAG CCA TCC ATT CTG GGC CAG AAG AAA ATG AGG

GGC CTG GAG GCC CAG GGA AGC CAG AAC CAG GTG **TAG** [...]

2) **ATG** GGG GAT ATG AAG ACC CCT GAT TTT GAT GAC CTC CTT GCT GCC TTT GAC ATC

CCT **GAC** **ATT** **GAT** **GCG** **AAT** **GAA** **GCC** **ATC** CAT TCT GGG CCA GAA GAA AAT GAG [...]

A

ATG [...] **AAT** **TGA** TGC GAA TGA AGC CAT [...]

Resulting protein sequences:

1) Met-Gly-Asp-Met-Lys-Thr-Pro-Asp-Phe-Asp-Asp-Leu-Leu-Ala-Ala-Phe-Asp-Ile-Pro-Asp-Met-Met-Arg-Met-Lys-Pro-Ser- Ile-Leu-Gly-Gln-Lys-Lys-Met-Arg-Gly-Leu-Glu-Ala-Gln-Gly-Ser-Gln-Asn-Gln-Val-**Stop**

➜**M**GDMKTPDFDDLLAAFDIPDMMRMKPSILGQKKMRGLEAQGSQNQV (46 aa)

2) **Met**-Gly-Asp-Met-Lys-Thr-Pro-Asp-Phe-Asp-Asp-Leu-Leu-Ala-Ala-Phe-Asp-Ile-Pro-Asp-Asn-**Stop**

➜**M**GDMKTPDFDDLLAAFDIP**D** (20 aa)

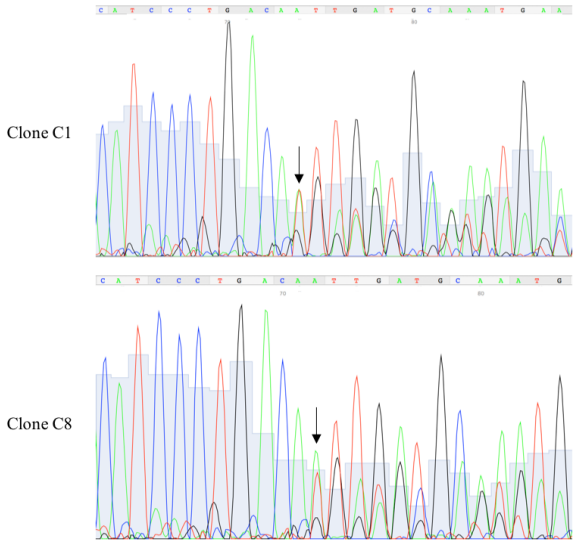


Figure 3.27 – Insertion of an adenine in two CRSIPR-Cas9 clones. The CRISPR-Cas9 system has mediated a DBS four nucleotides downstream the 5'-CCT-3'sequence in the clone C1, C8. To repair the DBS, an adenine was inserted in one allele and a thymine was deleted in another allele, both harbouring a shift in the reading frame and leading to the appearance of a premature stop codon. The newly synthesized proteins are 20 and 46 aa long. The initial and stop codons are represented in green and red colour, respectively. The gRNA1 (oligo1) is represented in blue, and the pam sequence, in this case 5'-CCT-3' (reverse complement), is illustrated in black. The deleted and inserted nucleotides, as well as the new amino acids are depicted in coral. The black arrow in both chromatograms depict the localization of the indel.

One the other hand, clone C21 has suffered an insertion of a thymine, four nucleotides downstream the 5'-CCT-3' sequence, where the CRISPR-Cas9 system has generated the DBS (Figure 3.28). We expect on the basis of the genomic alteration, the appearance of a premature stop codon, resulting in a truncated protein of only 20 amino acids.

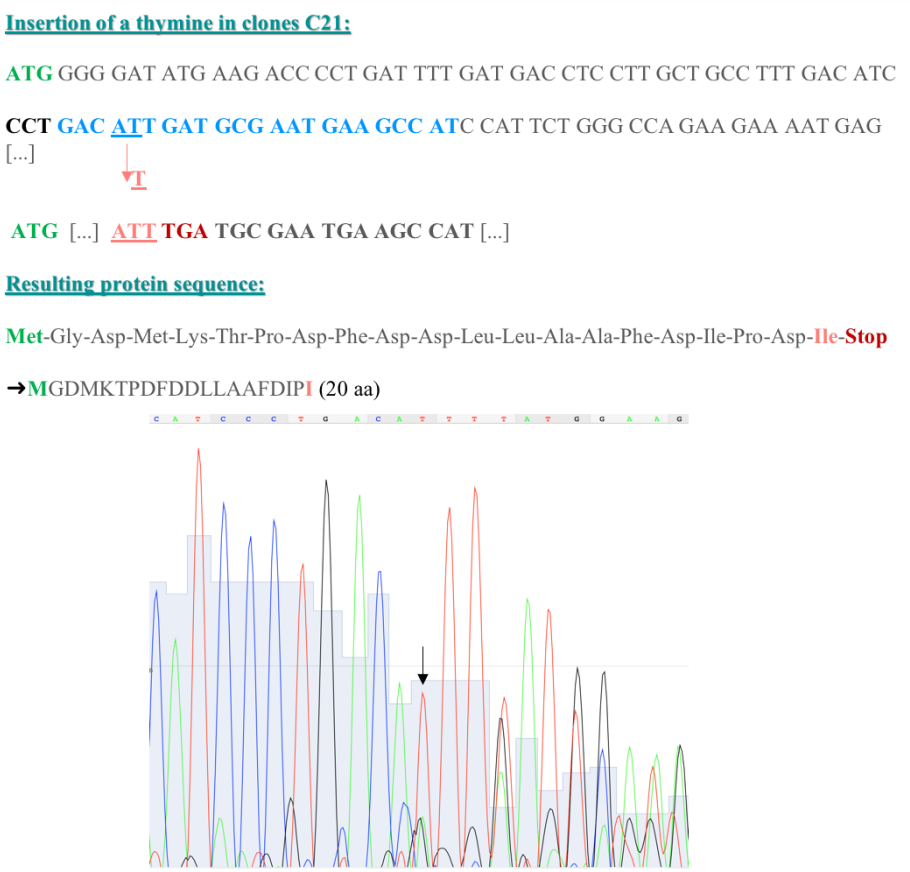


Figure 3.28 - Insertion of a thymine in one CRSIPR-Cas9 clone. The CRISPR-Cas9 system has mediated a DBS four nucleotides downstream the 5'-CTT-3'sequence in the clone c21. To repair the DBS, a thymine was inserted, changing the reading frame and leading to the appearance of a premature stop codon. The newly synthesized protein is only 20 aa long. The initial and stop codons are represented in green and red color, respectively. The gRNA1 (oligo1) is represented in blue, and the reverse complement PAM sequence, in this case 5'-CCT-3', is illustrated in black. The inserted nucleotide and the new amino acids are depicted in coral. The black arrow in the chromatogram depict the localization of the indel.

Clone 22 also suffered the same insertion as clone C1 and C8. Indeed, an adenine was inserted four nucleotides downstream the reverse complement of PAM (Figure 3.29).

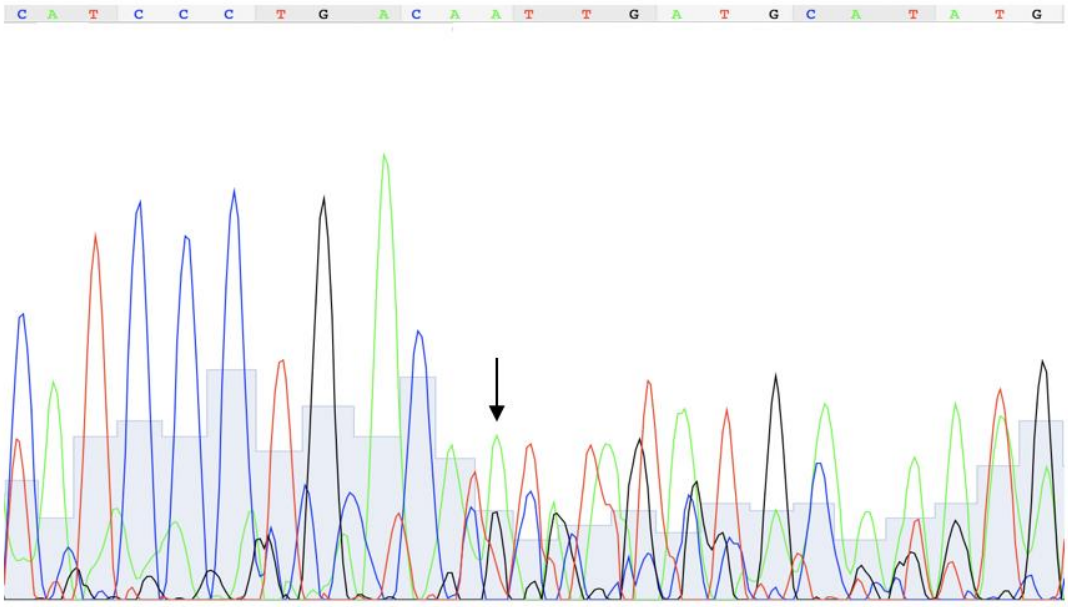


Figure 3.29 – Chromatogram of clone C22 depicting the insertion of an adenine. The black arrow in the chromatogram depicts the localization of the indel.

Finally, looking at the chromatogram of clones 15 and 17 we identified a deletion of an adenine downstream the fourth nucleotide of the 5'-CCT-3' sequence that will harbor a frameshift and lead to the appearance of a stop codon (Figure 3.30).

Deletion of an adenine in clones C15 and C17:

ATG GGG GAT ATG AAG ACC CCT GAT TTT GAT GAC CTC CTT GCT GCC TTT GAC ATC

CCT GAC ATT GAT GCG AAT GAA GCC ATC CAT TCT GGG CCA GAA GAA AAT GAG [...]

ATG [...] TTG ATG CGA ATG AAG CCA TCC ATT CTG GGC CAG AAG AAA ATG AGG

GGC CTG GAG GCC CAG GGA AGC CAG AAC CAG GTG TAG [...]

Resulting protein sequence:

Met-Gly-Asp-Met-Lys-Thr-Pro-Asp-Phe-Asp-Asp-Leu-Leu-Ala-Ala-Phe-Asp-Ile-Pro-Asp-Leu-Met-Arg-Met-Lys-Pro-Ser- Ile-Leu-Gly-Gln-Lys-Lys-Met-Arg-Gly-Leu-Glu-Ala-Gln-Gly-Ser-Gln-Asn-Gln-Val-Stop

→MGDMKTPDFDDLLAAFDIPDLMRMKPSILGQKKMRGLEAQGSQNQV (46 aa)

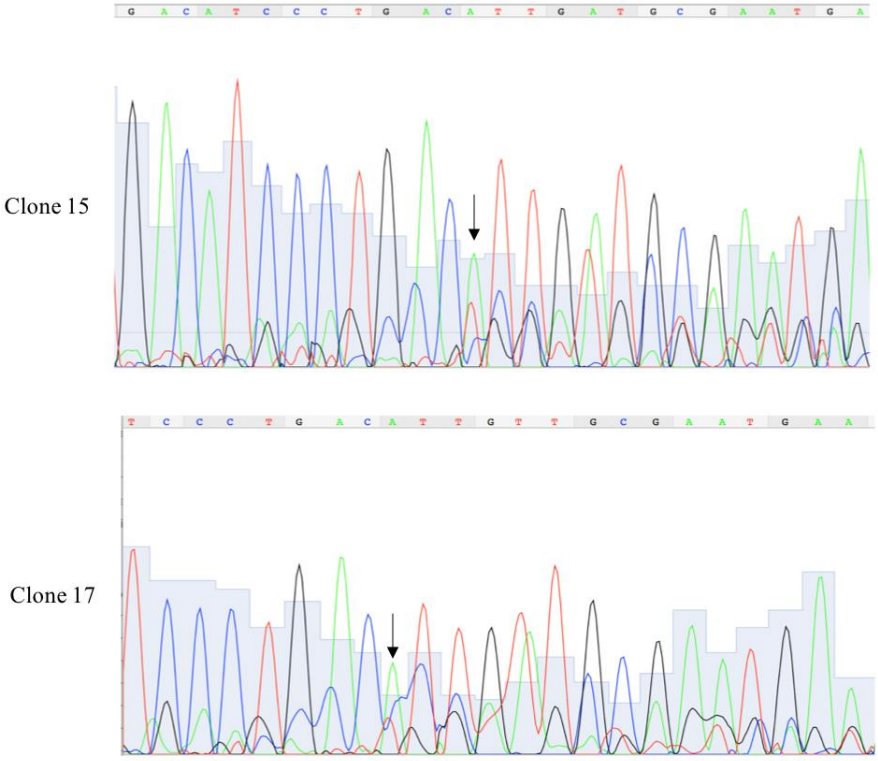


Figure 3.30 – Possible deletion of an adenine in CRISPR-Cas9 clones. The CRISPR-Cas9 system has mediated a DBS four nucleotides downstream the 5'-CTT-3'sequence in the clone C15 and C17. To repair the DBS, an adenine was deleted, shifting the reading frame and leading to the appearance of a premature stop codon. The newly synthesized protein is only 46 aa long. The initial and stop codons are represented in green and red colour, respectively. The gRNA1 (oligo1) is represented in blue, and the reverse complement pam sequence, in this case 5'-CCT-3', is illustrated in black. The inserted nucleotide and the new amino acids are depicted in coral. The black arrow in the chromatogram depict the local of the indel.

Then, in order to validate the results obtained from the SANGER sequencing, we have conducted a Western Blot assay and assessed the amount of protein in each of the six knock-down clones (Clone C1, C8, C15, C17, C21 and C22). Results showed that the band corresponding to ZNF687 appears to be the correct one, at the right size (129 kDa),

in all clones (Figure 3.31). No clone presented a knock-out, but clone C15 and C22 seem the most prominent ones for ZNF687 knock-down.

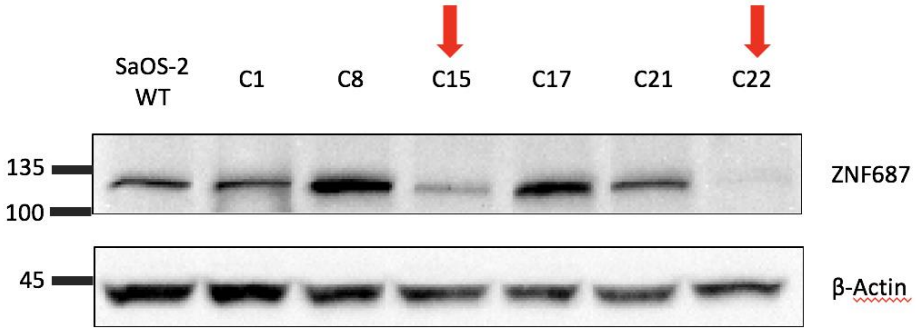


Figure 3.31 – Western Blot assay conducted on six CRISPR-Cas9 clones. A western blot assay was performed after extraction of total protein, from the six clones, previously identified with an indel mutation. β-Actin protein has been used as a loading control.

3.3.4 – Effect of ZNF687 knock-down in the expression of several genes

To understand the effect of the ZNF687 knock-down in bone metabolism, we analysed the relative expression of several genes (Appendix VI, Table A6.2) involved in bone metabolism and/or in PDB (e.g. *OPTN*, *SQSTM1*, *RANK*), by RT-qPCR, which were then normalized to two reference genes, *GAPDH* and *β-actin*. These selected genes are the same that we analysed in the overexpression assay. Total RNA extracted from normal SaOS-2 cells and the six CRISPR-cas9 clones was used for RT-qPCR analysis. Results showed that for all the six CRISPR-cas9 clones there were an overall decrease of expression in all the genes analysed, with the exception of *RANK* and *FGF* for certain clones (Figure 3.32). Statistical analysis demonstrated a significant decrease in *Osteocalcin*, *BMI1*, *SQSTM1* and *OPTN* expression levels, in all the clones analysed. The expression of *OSX/SP7* was lower in knock-down clones than in the control, but no statistical significance depicts the slight decrease in the gene expression for C22. All clones, except C22, showed a significant decrease in *CCDC3* expression. *RUNX2* expression appears diminished in C1, C8 and C21, with statistical significance.

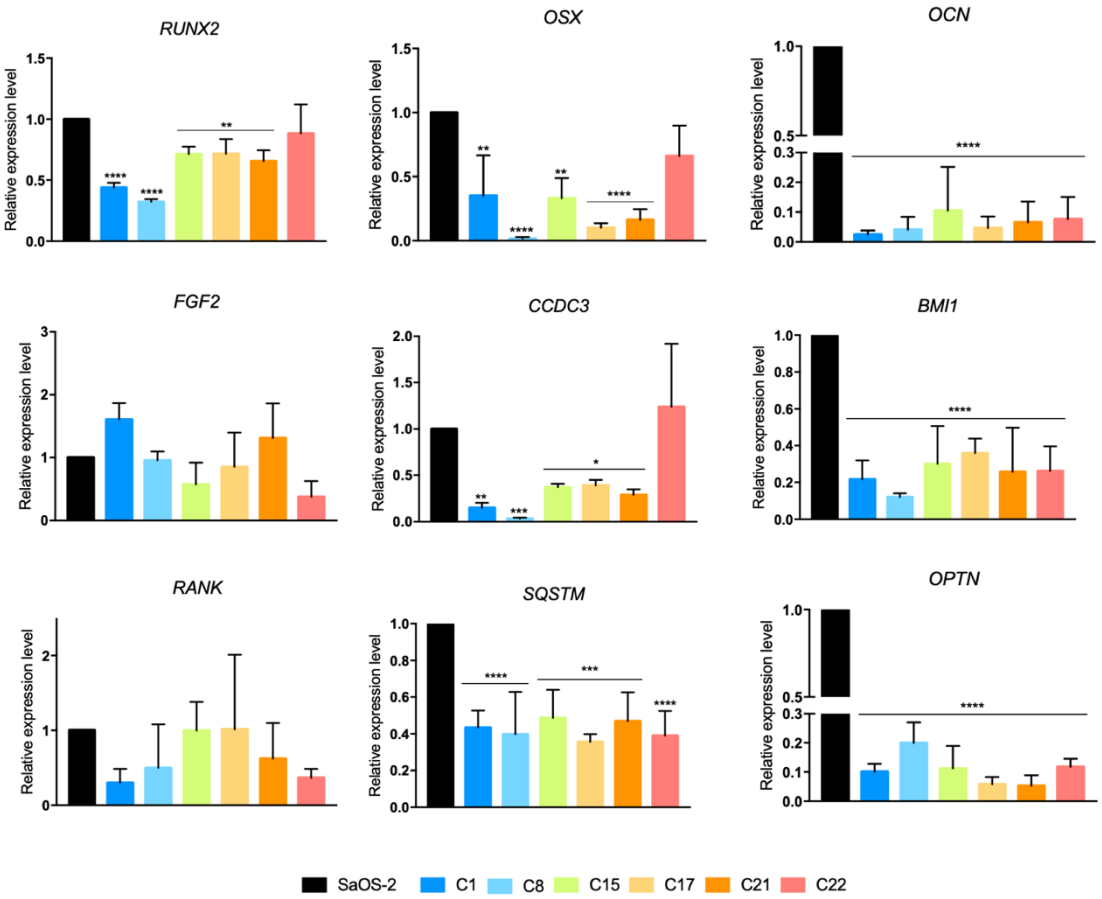


Figure 3.32 – Expression of several genes in knock-down clones. qPCR monitoring the expression of several genes in normal SaOS-2 (black) and clones C1, C8, C15, C17, C21, and C22. Gene levels were normalized to the mean expression of GAPDH and β-actin. The data are shown as the means ± SD from three independent experiments. Asterisk depict a response that is significantly different from the control SaOS-2 (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001).

3.3.5 Effect of ZNF687 knock-down in cell mineralization

3.3.5.1 Mineralization assays

To investigate the involvement of *ZNF687* gene expression in the mineralization process of SaOS-2 cells, Alizarin red (AR-S), von Kossa and Sirius red staining were performed. Alizarin red and von Kossa allows the quantification of the deposits of calcium, and Sirius red staining quantifies morphometric collagen. We also used p-NPP to measure the alkaline phosphatase activity. All these assays were performed in both unmineralized and mineralized cells.

Results from the AR-S clearly indicated an increase in calcium deposit in mineralized cells compared to control, i.e. unmineralized SaOS-2, but also to respective unmineralized cells, with statistical significance (Figure 3.33). Nonetheless, this increase was much less notable in clones C8M and C1M, and these two clones presented less

mineralization than mineralized SaOS-2. C22M, on the other hand, was the only clone who presented higher mineralization than the control SaOS-2M, with statistical significance.

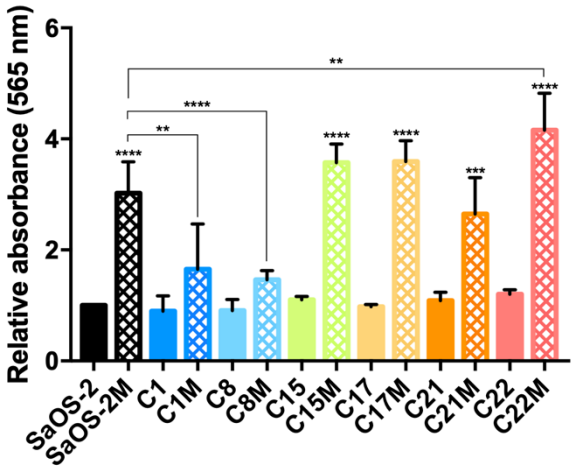


Figure 3.33 – Quantification of calcium salts by Alizarin red staining assay in unmineralized and mineralized cells. Data were all normalized to the control SaOS-2 (plain black) and are shown as the mean ± SD from four independent experiments. Plain colour bars indicate unmineralized cells, whereas the cross pattern in bars indicates mineralized cells. Black asterisks depict a response that is significantly different from the control SaOS-2 (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001). Grey asterisks depict a response that is significantly different from the mineralized SaOS-2 (SaOS-2M) (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001). Stripes in bars indicate mineralized cells, whereas plain color indicates unmineralized cells.

The von Kossa assay validated the results obtained in the previous experiment, since clones C1 and C8 precipitated fewer silver crystals than mineralized SaOS-2 and other clones (Figure 3.34). Moreover, clones C15M, C17M, C21M and C22M presented a similar deposition of silver crystal than SaOS-2 M.

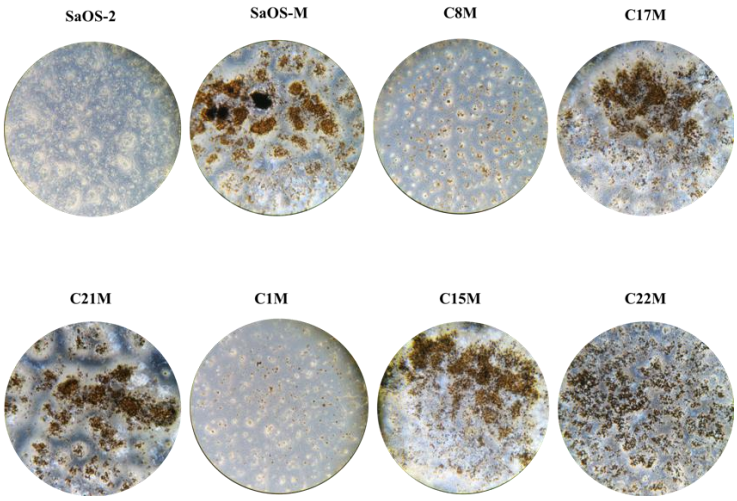


Figure 3.34 – Microscopic visualization of silver crystals in mineralized cells after von Kossa assay.

Results from the Sirius red staining showed slight differences between mineralized and unmineralized cells. Nonetheless, only clone C22M presented a statistically significant difference with SaOS-2M. (Figure 3.35).

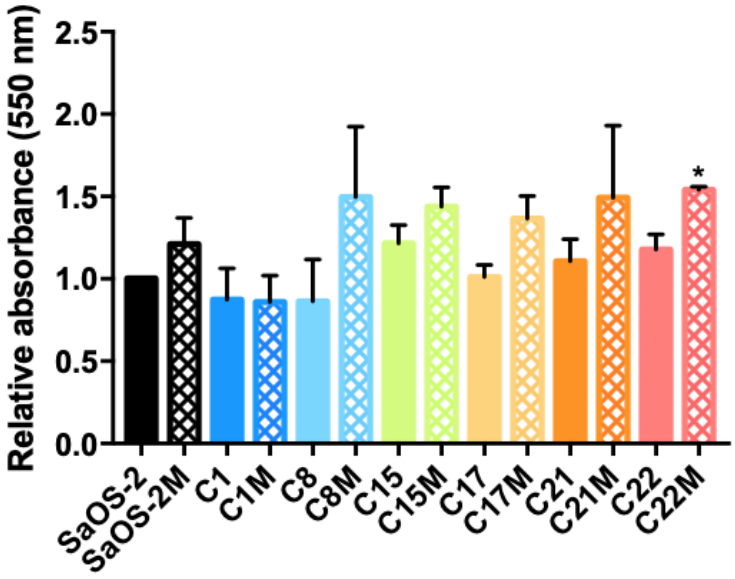


Figure 3.35 – Quantitative morphometric collagen measurements by Sirius red staining in unmineralized and mineralized cells. Data were all normalized to the control SaOS-2 (plain black) and are shown as the mean ± SD from five independent experiments. Plain colour bars indicate unmineralized cells, whereas the cross pattern in bars indicates mineralized cells (*p <0.05).

Finally, the alkaline phosphatase activity was rather similar between unmineralized and mineralized cells, except for clones C15 and C15M, where a difference can be observed between the two unmineralized and mineralized clones, with a statistical significance (Figure 3.36). The mineralized clones C1M, C8M, and C22M were the only clones who presented decreased levels of ALP compared to mineralized SaOS-2M, with statistical significance. The other clones, i.e. C15M, C17M and C21M presented levels of ALP similar to SaOS-2M. Interestingly, almost all unmineralized knock-down clones showed lower levels of ALP compared to the control SaOS-2.

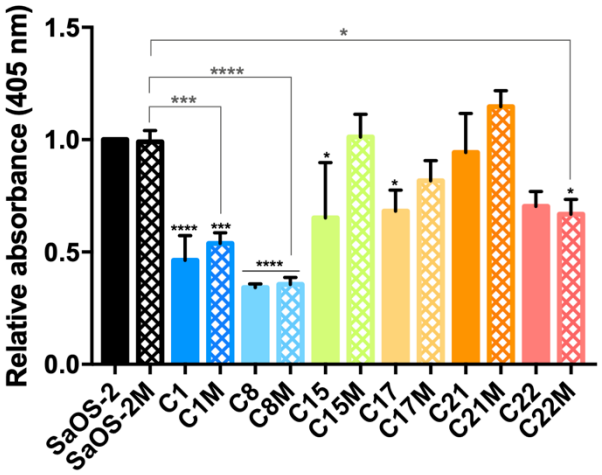


Figure 3.36 – Alkaline phosphatase in unmineralized and mineralized knock-down clones. Data were all normalized to the control SaOS-2 (plain black) and are shown as the mean ± SD from three independent experiments. Plain colour bars indicate unmineralized cells, whereas the cross pattern in bars indicates mineralized cells. Black asterisks depict a response that is significantly different from the control SaOS-2 (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001). Grey asterisks depict a response that is significantly different from the mineralized SaOS-2 (SaOS-2M) 2 (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001).

3.3.5.2 Effect of mineralization in gene expression

Mineralized clones were assessed for their gene expression. We performed a RT-qPCR in mineralized knock-down clones to evaluate any differences in the expression of certain genes previously studied. It should be noted that clone C8 was lost in the mineralization process, and thus was excluded from this part of the study. Therefore, the expression of *RUNX2*, *RANK*, *OSX*, *OSTEOCALCIN*, *CCDC3*, *FGF2*, *OPTN* and *SQSTM1* was appraised in mineralized clones C1M, C15M, C17M, C21M and C22M, using mineralized SaOS-2 as a control. The two reference genes *GAPDH* and *β-actin* were used for gene expression normalization.

Results showed that the expression of *RUNX2* and *RANK* in all clones were lower than in the control, with a statistical significance (Figure 3.37). The expression of the other genes did not appear significantly different than the control, except for clone C1M, which exhibited higher expression of *RANK*, *FGF2* and *SQSTM1*, and lower expression of *CCDC3* than SaOS-2M. Interestingly, clone C22M presented a lower expression in almost all gene tested, namely *SQSTM1* and *OPTN* than the control, with statistical significance, which is in accordance with the results obtained in the RT-qPCR in the knock-out assay.

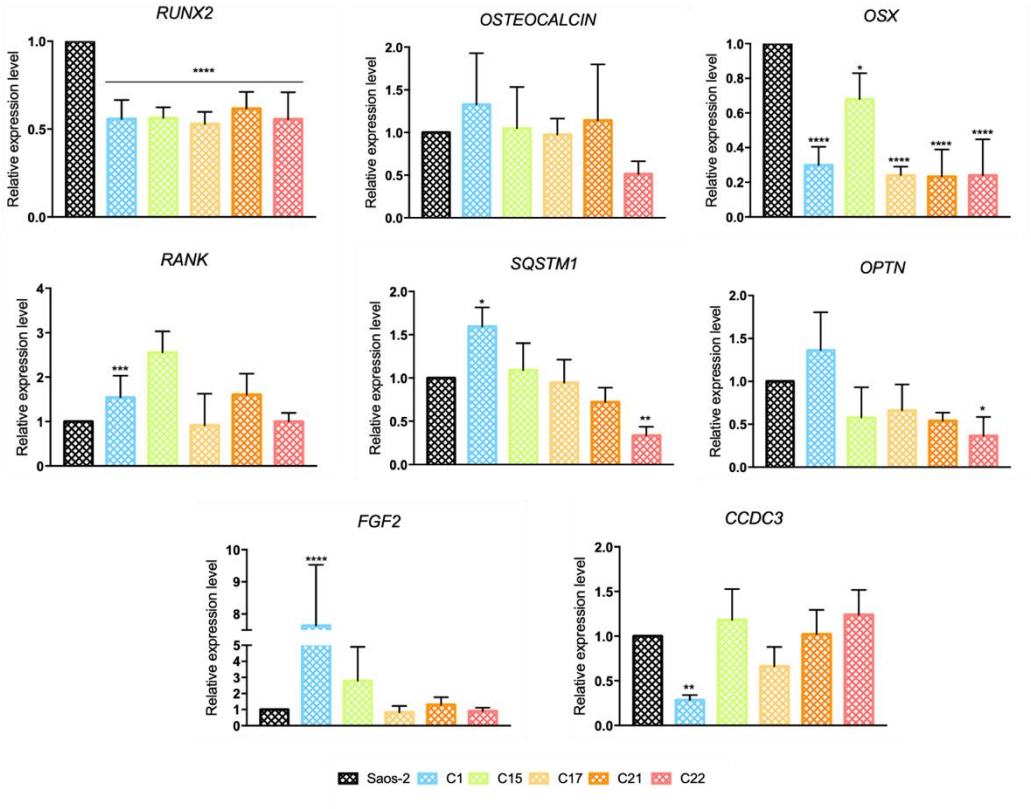


Figure 3.37 - Expression of several genes in mineralized knock-down clones. qPCR monitoring the expression of several genes in mineralized SaOS-2M (black) and mineralized clones C1M, C15M, C17M, C21M, and C22M. Gene expressions were normalized to the mean expression of GAPDH and β-actin. The data are shown as the means ± SD from three independent experiments. Asterisk depict a response that is significantly different from the control SaOS-2M (*p <0.05; **p <0.01; ***p <0.001; ****p< 0.0001).

3.4 Zebrafish as a good biological model for *ZNF687* study

Zebrafish appears to be a good animal model for a wide range of studies. Moreover, it has been demonstrated that ZNF687 was upregulated during osteoclastogenesis and osteoblastogenesis, in both human and zebrafish, suggesting that the function of this protein might be conserved by evolution despite speciation.[38] Since there is no animal model, to our knowledge, for the study of *ZNF687*, it will be interesting to consider the zebrafish as a putative model. Therefore, to evaluate this suggestion several *in silico* comparative analyses were performed between several species.

3.4.1 Genomic comparative analysis

During evolution, the teleost lineage suffered a genome duplication which resulted in the appearance of several orthologs for mammalians. Hence zebrafish possess two paralog *znf687* genes, one of them being the ortholog to human *ZNF687*.[53] Despite this genomic duplication, *ZNF687*genomic structure is similar to *znf687a* and *znf687b,* with higher resemblance with *znf687b* (Figure 3.38). The coding sequence also appeared to be

akin to human CDS. As expected, the genomic structure between the two mammalian species are very similar, presenting exons with the same number and size (exons 2 to 8). Also, all of the species presented here possess their *ZNF687* gene in different chromosomes (Figure 3.37).

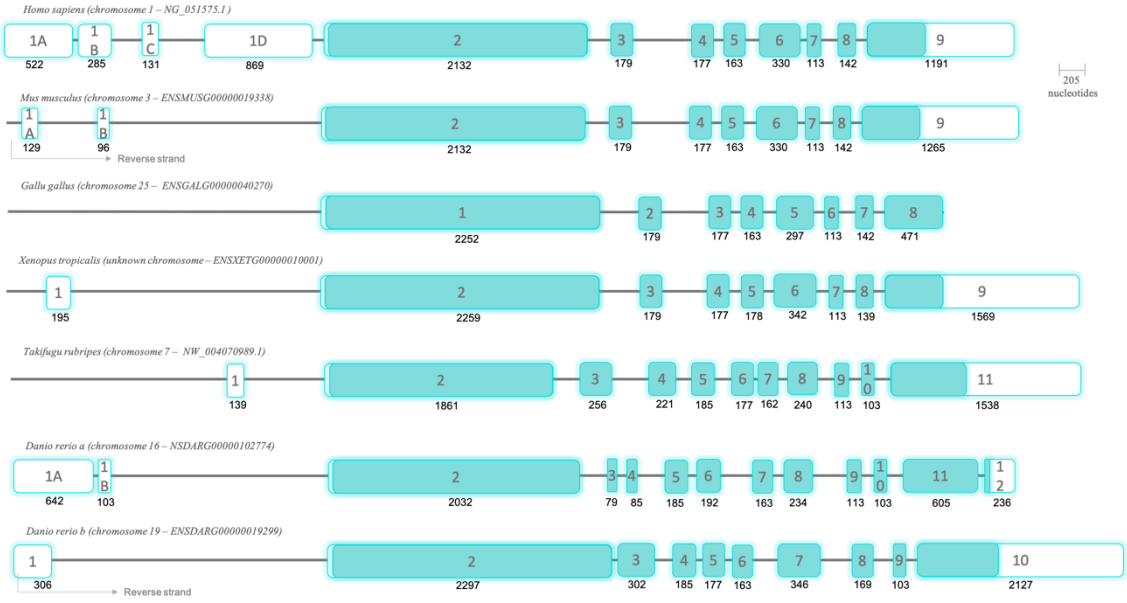


Figure 3.38 - Comparative analysis of *ZNF687* genomic structures between several species. All the *ZNF687* genes present similar structure and coding sequence. Coding exonic and non-coding exonic regions are represented by full and white boxes, respectively. Numbers below the boxes indicate exons length. Only exons are on scale. The genomic locations are according to Ensembl and NCBI.

In order to identified putative molecular players affecting zebrafish *znf687a* and *znf687b* gene transcription, we used a computational approach to search for cis-regulatory transcription factor binding sites (TFBSs) in the promoter regions of the *znf687a*, *znf687b* and human *ZNF687*. Therefore, 2 kb of promoter region between human and other species were analysed for the presence of several binding sites for transcription factors (TFs). The search of TFs was based on the information available in the ENCODE project (UCSC Genome Browser), for the human gene. Results demonstrated a few presences, within their promoter regions, of similar putative binding sites, for common transcriptional factors, such as SP1, C-EBPA, SRF, EGR-1 and NF-κB, among *znf687a* and *znf687b* zebrafish promoter region and human *ZNF687* genes, advocating a similar functional role (Figure 3.39).

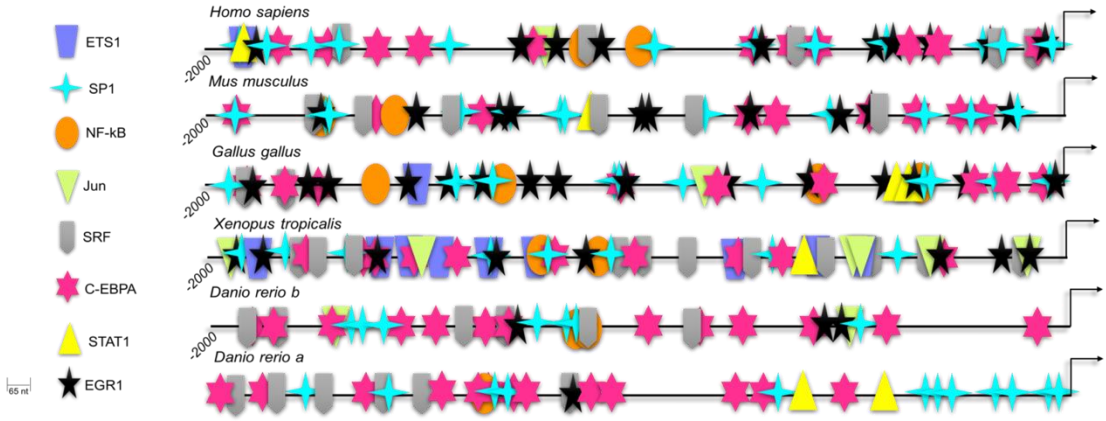


Figure 3.39 – Comparative analysis of 2 Kb promoter region between human and other species. Several transcription factors binding sites (TFBSs) are depicted in the promotor sequence of different species. The search of TFBSs was based on the information available in the ENCODE project (UCSC Genome Browser).

Finally, a comparative mapping of *ZNF687* genes and an analysis of syntenic regions and chromosomal locations that harbour *ZNF687* has been assessed between several species. Results demonstrated that the neighbouring genes of *ZNF687* orthologues are only conserved between mammalian and avian species, and only very few are preserved in amphibian and teleost species (Figure 3.40).

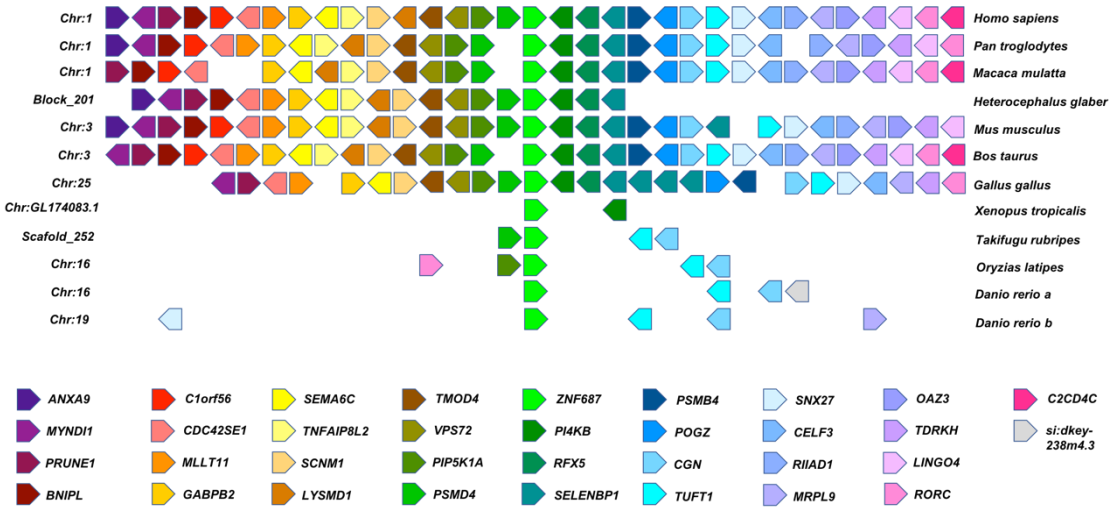


Figure 3.40 - Schematic representation of genetic synteny neighbouring *ZNF687* gene in the chromosome of various species. The cluster of surrounding genes of *ZNF687* is mainly conserved in all species except in amphibian and teleost species.

3.4.2 Comparative analysis of the ZNF687 protein

Despite the fact that zebrafish and human presented a similar genomic structure and CDS, we performed an alignment between the ZNF687 protein of several species (Appendix IV, Figure A6.8). Pairwise alignments and identity percentage revealed that the protein sequences were highly conserved among mammalian species, and less with other species, namely teleost (Figure 3.41). Indeed, human protein presented a percentage

of identity of 34.6 and 40.3 with Znf687a and Znf687b, respectively. Nonetheless, the protein homology between the two protein of zebrafish is also low, with 46.2% of identity.

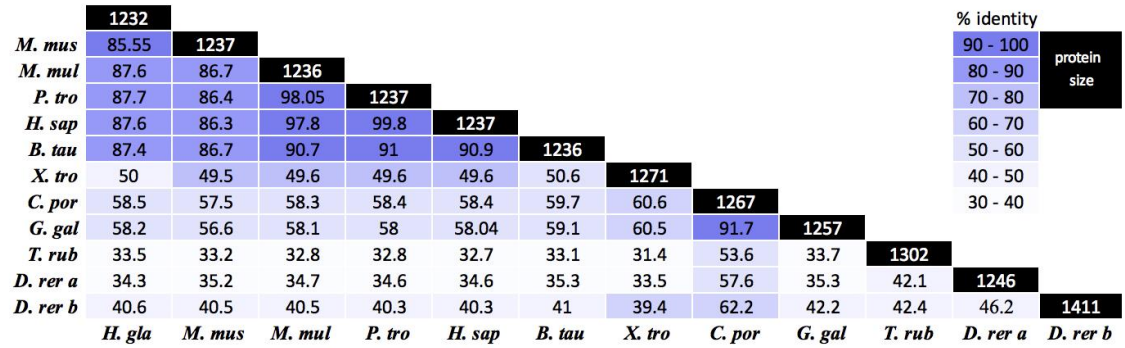


Figure 3.41 – Protein homology of ZNF687 between different species. Sequence identity of complete protein ZNF687 between species. Pairwise alignments were done using species ranging from fish to mammals and the identity percentage was acquired with Manipulation Suite facilities. Taxa are labeled: M. mus (Mus musculus), M. mul (Macaca mulata), P. tro (Pan troglodytus), H. sap (Homo sapiens), B. tau (Bos Taurus), X. tro (Xenopus tropicalis), C. por (Crocodylus porosus), G. gal (Gallus gallus), T. rub (Takifugu rubripes), and D. rer (Danio rerio).

Moreover, by analysing the pairwise alignment, we could observe that certain regions of the protein sequences are relatively well preserved between species. After identification of the human protein sequence in these regions, we noticed that they corresponded to the zinc finger motifs encountered in the human protein. Hence, these results suggest that the zinc finger motifs might be preserved throughout evolution. The identity percentage analysis of this protein region demonstrated a higher identity between human and other species. Indeed, the human protein presented 57.1% and 62% of identity with Znf687a and Znf687b, respectively (Figure 3.42).

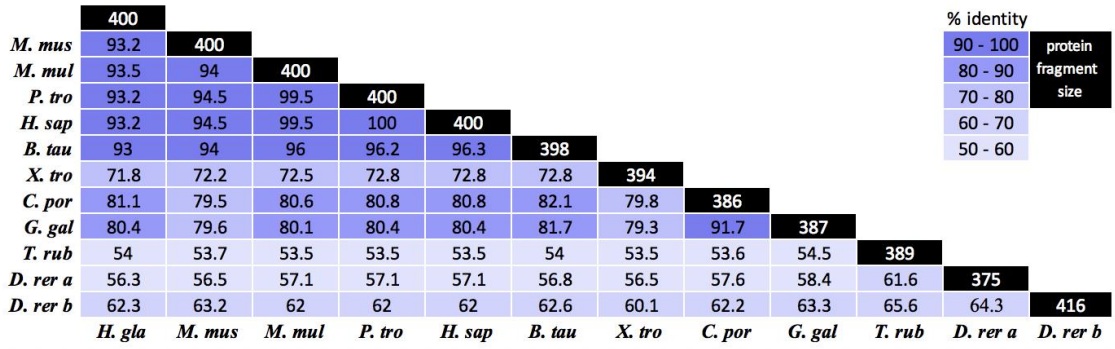


Figure 3.42 – Protein homology of the most conserved region. Sequence identity of partial ZNF687 corresponding to a cluster of C2H2 zinc finger in the human protein. Pairwise alignments were done using species ranging from fish to mammals and the identity percentage was acquired with Manipulation Suite facilities.

To determine the zinc finger motif in ZNF687 in several species, we performed a protein domain analysis. Results demonstrated that the 14 ZF motifs, that were identified

in the human protein, are conserved throughout evolution (Figure 3.43). An additional C2H2 motif was found in *Danio rerio* and other species, such as *Gallus gallus*, *Crocodylus rubripes*, and *Takifugu rubripes*. Even if znf687a lacks two of the 14 C2H2 ZF motif encountered in the human protein, zebrafish proteins presented a similar pattern of ZF domains that are organized in three clusters from the middle to the C-terminal end of the protein, when compared to the human ZNF687 (Figure 3.43).

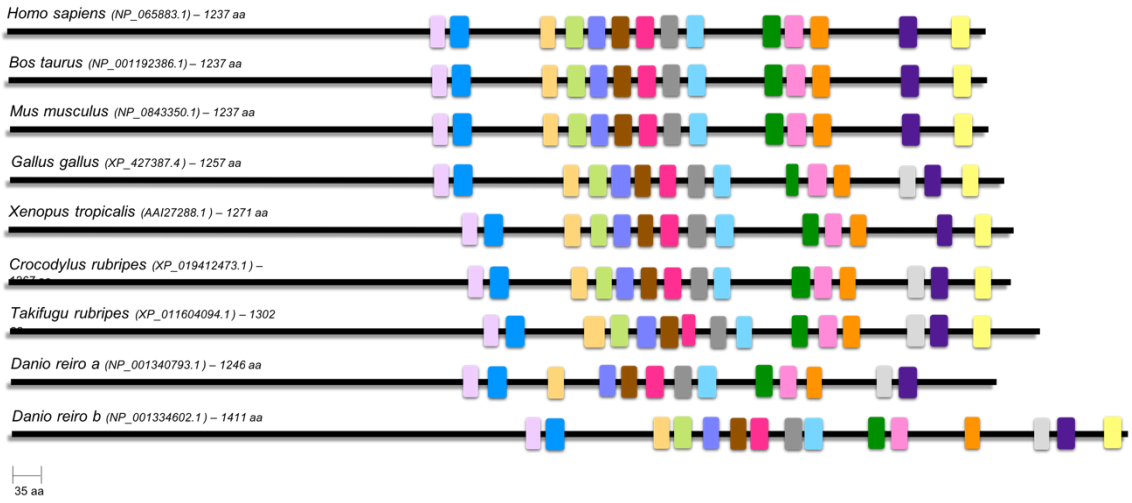


Figure 3.43 – Comparative analysis of zinc finger C_2H_2 motifs between various species. All the human zinc finger domains are conserved throughout evolution. Nonetheless, one domain (grey box) persisted in every species except in mammalians. Each zinc finger domain is represented by a box with its specific color.

Finally, through pairwise alignment, we searched if the amino acids that are found mutated in PDB patients, serine (p.Ser242Ile), proline (p.Pro665Leu) glutamine (p.Gln784Glu) and another proline (p.Pro937Arg) were conserved in zebrafish. Interestingly, only one of the four amino acids is preserved. Indeed, the glutamine is conserved in all species, and this amino acid is integrated in a zinc finger domain (Figure 3.44).

```
Danio rerio a          QNAVKTHMQTAHCEIFHKCPSCPMAFK
Takifugu rubripes      ---VKSHIQQAHCDVFHKCPSCPMAFK
Danio rerio b          ---IKSHIQTAHCEVFHKCPNCPMAFK
Mus musculus           ---IKSHIQASHCEVFHKCPICPMAFK
Pan troglodytes        ---IKSHIQTSHCEVFHKCPICPMAFK
Homo sapiens           ---IKSHIQTSHCEVFHKCPICPMAFK
Bos taurus             ---IKSHIQTSHCEVFHKCPICPMAFK
Xenopus tropicalis     ---IKSHIQTSHCEVFHKCPICPMAFK
Crocodylus porosus     ---IKSHIQTSHCEVFHKCPICPMAFK
Gallus gallus          ---IKSHIQTSHCEVFHKCPICPMAFK
```

Figure 3.44– Conservation of the glutamine residue involved in PDB. Pairwise alignment of a fragment of ZNF687 between species, depicting the conservation of the glutamine (highlighted in coral).

3.5 Functional analysis of the promoter region of human and zebrafish ZNF687

As no data are available on the regulation of *ZNF687* in any species, we performed a promoter analysis of *ZNF687*, *znf687a* and *znf687b* in order to provide additional evidence of the usefulness of nonmammalian model systems to elucidate the regulation of *ZNF687* gene transcription.

3.5.1 Analysis of ZNF687, znf687a and znf687b promoter activity in transfected cells.

ZNF687 possess eight transcript variants resulting from alternative splicing, five of them being classified as predicted or hypothetical (Figure 3.45). Nevertheless, the major differences found in these variants are localized upstream exon 2. Focusing our work on the three identified transcript variants (transcript variant 1, 2 and 3; i.e. NM_001304763.1, NM_001304764.1, and NM_001304764.1, respectively) we hypothesize the existence of two different promoter regions.

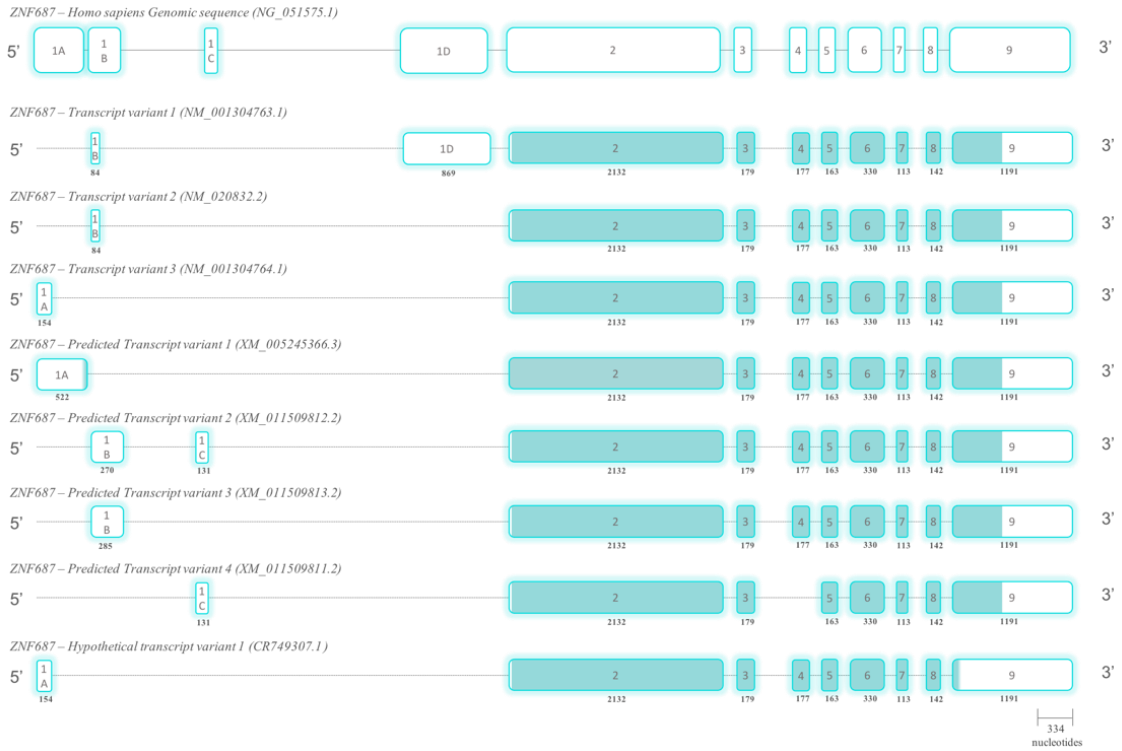


Figure 3.45 – Schematic representation of the genomic structure of *ZNF687* and its transcript variants. *ZNF687* possess eight transcript variants resulting from alternative splicing. Most of the variants present identical CDS and major differences are encountered upstream exon 2. Exons are represented in boxes, and introns are depicted in dashed lines. Filled boxes, in the transcript variants, represent the CDS. Both exon and introns are scaled.

Therefore, in order to test the ability of the human *ZNF687* promoters to direct transcription, we designed three fragments to be further cloned into the pGL3-Basic reporter vector: fragment 1 (F1) with 961 bp, located upstream exon 1A; fragment 2 (F2) with 850 bp, positioned between exon 1A and exon 1B; and fragment 3 (F3), with 1795 bp, representing F1 and F2 in one fragment (Figure 3.46). The insertion of the F3 into the pGL3-Basic is still ongoing.

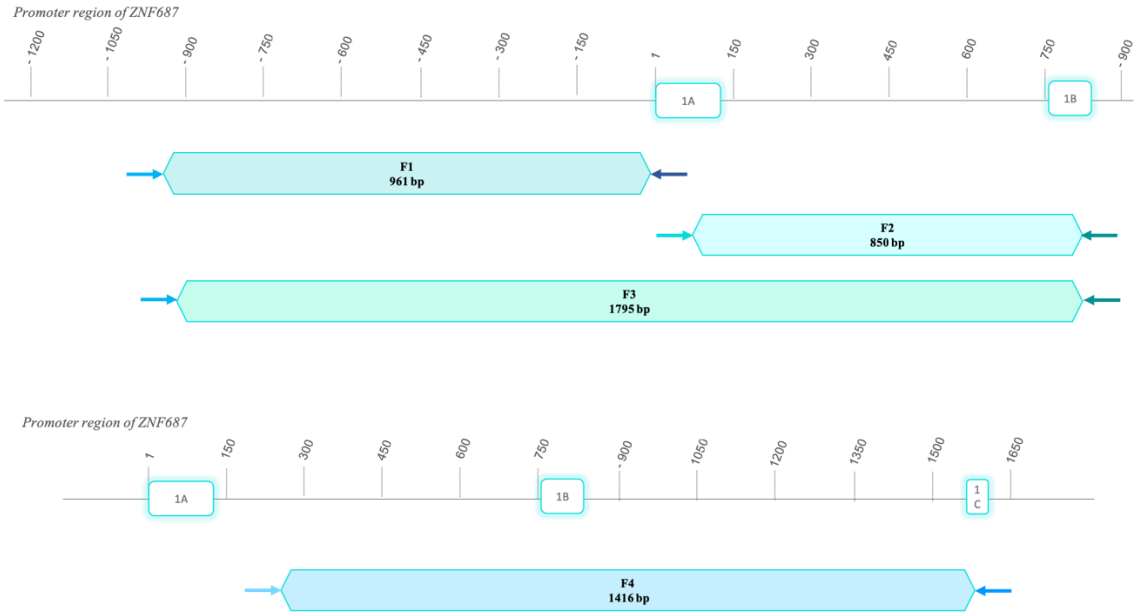


Figure 3.46 – The construction of the ZNF687 promoter-reporter plasmids. Each promoter fragment was amplified with designed primers to generate two constructions for ZNF687 promoter; F1 (with 961 bp), F2 (with 850 bp) and F3 (with 1795 bp). These fragments were linked to the Luc reporter gene.

Regarding zebrafish, both genes *znf687a* and *znf687b*, possess two transcript variants each (Figure 3.47). Transcript variants of each gene presented slight differences between them, but those differences do not affect the CDS. Based on these results we decided to analyse one promoter region for each zebrafish *znf687*. Fragments of promoter region of *znf687a* (FA) and *znf687b* (FB) were designed to be further cloned into the pGL3-Basic reporter vector containing the luciferase reporter gene (Figure 3.48). The FA with 1231 bp, located upstream exon1 of *znf687a*, and FB with 1166 bp, positioned upstream the exon1 of *znf687b*. After encountering several difficulties in the amplification of FB, the insertion of this fragment in pGL3 is still ongoing.

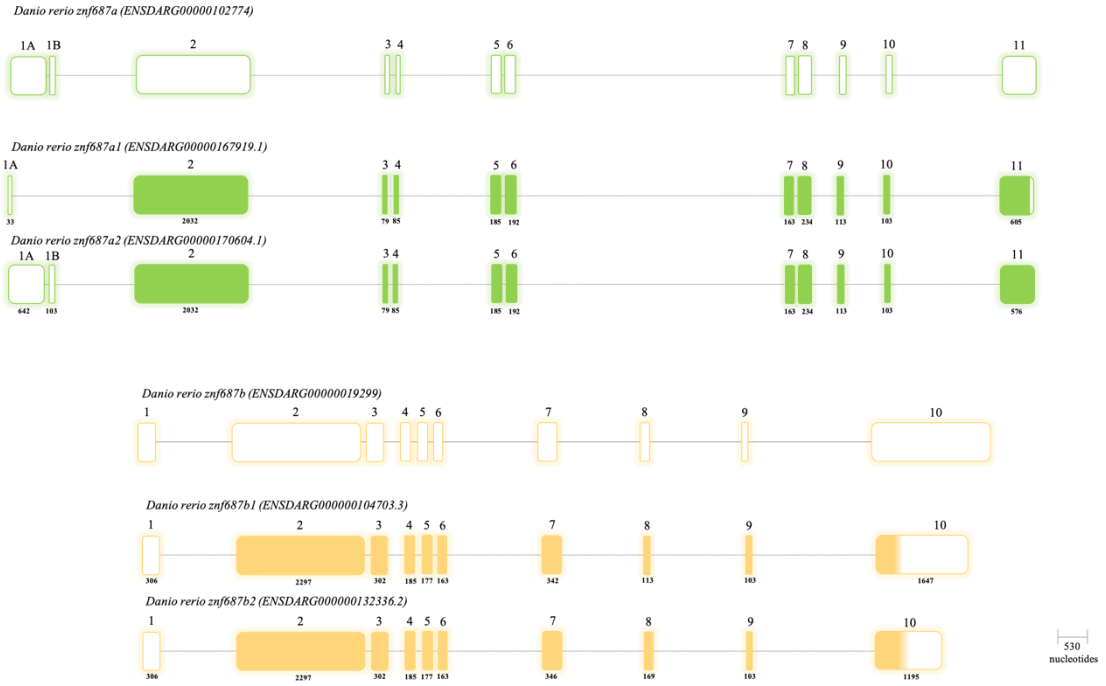


Figure 3.47 – Schematic representation of the genomic structures of *znf687a* and *znf687b* and their respective transcript variants. Genomic structure of *znf687a* and *znf687b* are illustrated in green and yellow, respectively. Exons are represented in boxes, and introns are depicted in dashed lines. Filled boxes, in the transcript variants, represent the CDS. Both exon and introns are scaled.

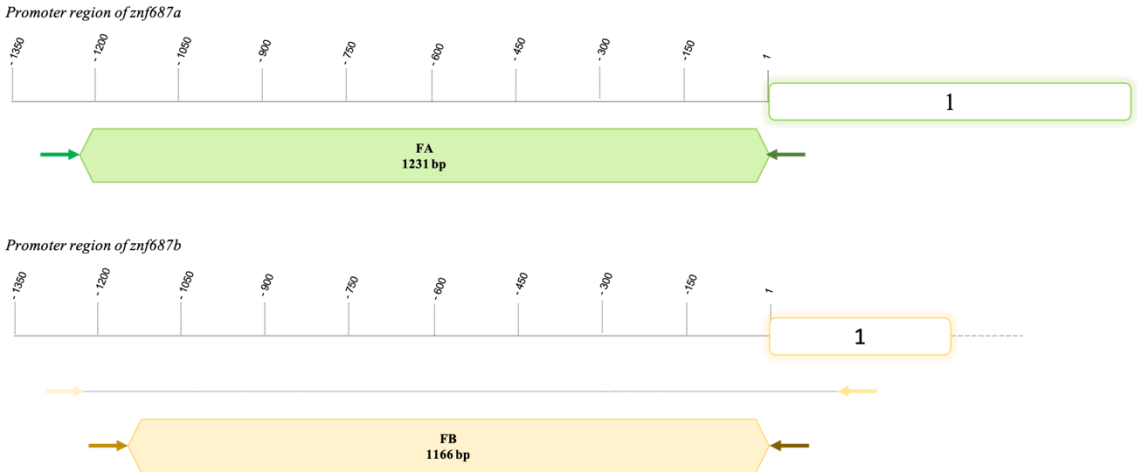


Figure 3.48 – The construction of *znf687a* and *znf687b* promoter-reporter plasmids. Each promoter was amplified with designed primers to generate two fragments: FA (green) with 1231 bp and FB (yellow) with 1166 bp. These fragments were linked to the Luc reporter gene.

To analyse the functionality of the cloned constructs we performed a luciferase assay in HEK293 cells. The promoterless luciferase reporter plasmid, pGL3-Basic, and the pGL3-control plasmids were also transfected into HEK293 and served as a negative and positive control respectively. Transfection reaction was stopped after 48 hours, and cells were harvest for firefly luciferase activity. After normalization against *renilla* luciferase activity, results showed that F1, F2 and FA were capable of promoting

transcriptional activity, compared to the negative control pGL3-Basic. Unfortunately, the promoter fragment FB did not prove to be functional (Figure 3.49).

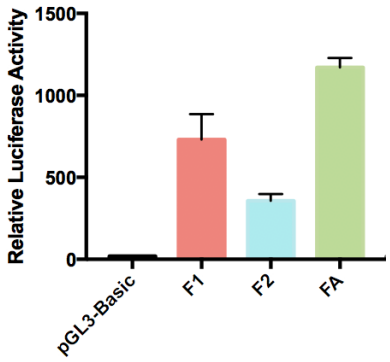


Figure 3.49 – Relative luciferase activity of *ZNF687*, *znf687a* and *znf687b* promoter fragments. Range of luciferase activity of F1, F2, FA, and FB, after transient transfection in HEK293 cells. With exception of FB, all promoter fragment constructs proved to be functional compared to the negative control pGL3-basic vector. The data are shown as the mean ± SD from two independent experiments.

3.5.2 Analysis of *ZNF687*, *znf687a* and *znf687b* promoter regulatory activity

We performed a co-transfection assay in HEK293 cells, in order to assess the regulatory activity of *ZNF687*, *znf687a* and *znf687b*. Hence, the human and zebrafish promoter-luciferase DNA constructs were transiently co-transfected with some of the transcription factors predicted in the *in silico* analysis (see Figure 3.40). The luciferase activity was then determined and normalized against *renilla* luciferase activity. The co-transfection of F1 construct together with SP1, EGR1, YY-1, IKK, FOXA1, E2F1, E2F4, STAT3, CTCF and AP1 showed no alteration in the transcriptional activity of this construct (Figure 3.50). However, since the low number of transfection assays performed, these results have to be carefully analysed and more assays have to be conducted to validate the results.

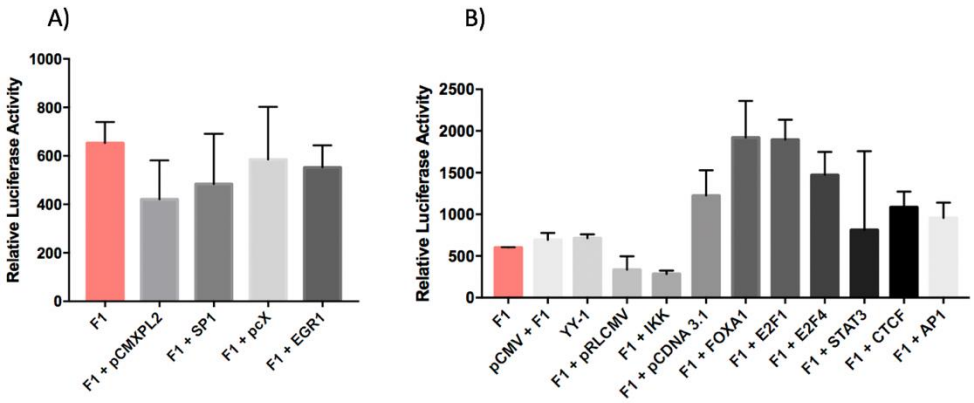


Figure 3.50 – Range of luciferase activity in co-transfected HEK293 with F1 construct. Cells were co-transfected with F1 construct and expression plasmid+TFs. Empty expression plasmids were also co-transfected with F1 and serve as control for putative regulatory activity of TFs. No representative alteration was noted in the transcriptional activity of F1**construct**. The data are shown as the mean ± SD in (A) two independent experiments (one duplicate per experiment) and in (B) one experiment (one duplicate per experiment).

The preliminary results involving F2 construct co-transfections with the TFs, showed that only AP1 seem to have an effect in F2 construct (Figure 3.51).

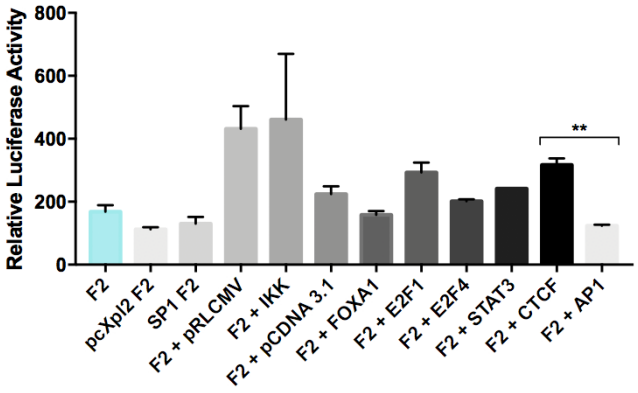


Figure 3.51 – Range of luciferase activity of co-transfected cells with F2. Cells were co-transfected with F2 construct and expression plasmid+TFs. Empty expression plasmids were also co-transfected with F2 and serve as control for putative regulatory activity of TFs. The data are shown as the mean ± SD in one experiment (with duplicate). An unpaired T test was performed between TFs and their respective empty plasmids. Black asterisks depict a response that is significantly different from the control, with **p <0.01.

Finally, results of FA construct co-transfections with STAT3, MEF2A and YY-1 TFs, showed two significant repressive effects, induced by STAT3 and YY-1 (Figure 3.53). Interestingly, in F1 construct regulatory analysis, STAT3 demonstrated similar behaviour than in the zebrafish *znf687a* promoter, but no statistical significance can be determined since results depict only one experiment.

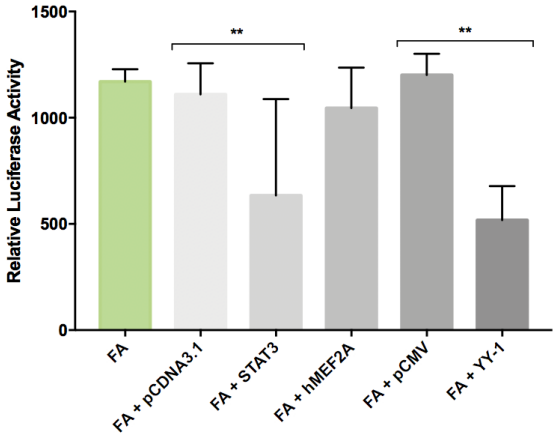


Figure 3.52 - Range of luciferase activity of co-transfected cells with FA construct. Cells were co-transfected with FA construct and expression plasmid+TFs. Empty expression plasmids were also co-transfected with FA construct and serve as control for putative regulatory activity of TFs. The data are shown as the mean ± SD from two independent experiments. The data are shown as the mean ± SD in three independent experiments (with duplicate). Black asterisks depict a response that is significantly different from the control, with **p <0.01.

CHAPTER IV – DISCUSSION

The aim of this work was to understand the role of *ZNF687,* and its involvement in osteoclastogenesis and osteoblastogenesis, but also in PDB. For this, we performed different *in silico* and *in vitro* analyses. After investigating how the four mutations, encountered in PDB patients, could affect the protein conformation and function, we have generated two types of *in vitro* analyses: (1) we performed an overexpression of WT *ZNF687* and of *mZNF687* (p.Pro937Arg, mediated by site-directed mutagenesis), independently, by transient transfection in SaOS-2 cell lines; and (2) we resorted to the CRISPR-Cas9 genome editing technique to cause a knock-down of the gene. These approaches allowed the assessment of the cell phenotype and gene expression, in proliferating and mineralization processes. Finally, comparative analyses were performed in order to define the usefulness of zebrafish as a biological model for the study of *ZNF687*. All these analyses provided information on ZNF687 that will be discussed in this section.

4.1 Mutations in ZNF687 might affect the protein-protein and protein-acid nucleic interactions

The secondary structure and domain analyses demonstrated that the transcription factor, ZNF687, possesses 14 putative zinc finger domains that are clustered in the second half of the protein, and appeared to be well conserved between mammalian species. These motifs are similar to the classical C2H2 motif, comprising approximatively 24 amino acids and a β-hairpin followed by an α-helix. They are organized in three distinct clusters that enable recognition of multi-nucleotide sequences. Moreover, our protein of interest seems to present three monopartite NLSs, located near the C-terminal part, and are well conserved in mammalian species.

The four mutations encountered in *ZNF687* by Divisato *et al.*, occur in different part of the protein (Figure 4.1).[38,43] While no alteration of the secondary structure of the protein was observed *in silico*, the difference in properties between the wild-type and the mutated residues might alter the local conformation or interaction. These alterations might affect the protein folding, protein-protein and protein-nucleic acid interactions. Indeed, the function and activity of certain proteins are modulated by the interaction with other proteins or nucleic acid. The p.Glu784Gln mutation occurs in the C-terminal end of

a zinc finger (ZF) domain. The residues located in this part of the protein are responsible for the folding of the α-helix. This conformation positions specific amino acids in a certain manner allowing their interaction with DNA.[94] Therefore, if a mutation occurs in this part of the ZF, it might disrupt DNA binding. Furthermore, we have observed that ZNF687 interacts with several proteins that are involved in gene regulation. It has been observed in Divisato *et al.* (2016) that mRNA of *ZNF687* was upregulated during osteoblastogenesis and osteoclastogenesis, suggesting a putative role of ZNF687 in the differentiation of these bone cells.[36] Meanwhile, its mechanism of regulation is still unknown, and even if our protein of interest is a transcription factor, we do not know if it regulates directly these genes on its own, or if it is involved in a complex, such as Z3.

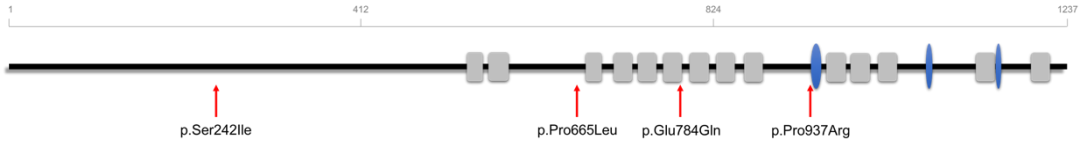


Figure 4.1 – Localization of the four mutations identified in the ZNF687. The four mutation, illustrated with red arrows, are localized in different part of the protein. The p.Pro665Leu occurs upstream a cluster of C2H2 motifs, while p.Glu784Gln is localized inside a zinc finger domain. The p.Pro937Arg arises just before a NLS. Zinc finger domains are represented by grey boxes and blue discs depict NLS.

Moreover, the p.Pro937Arg mutation generates an additional positive charge, and the mutated residue end up being part in a stronger NLS. The distal end of the arginine side chain is capped by a guanidinium group, which is often involved in ionic and hydrogen bonding that are essential to the structure, and stability of the protein and protein complexes.[95] Additionally, the equilibrium acid dissociation constant (pKa value) of the guanidinium group is very high (pKa=12,8), and hence, the side chain of arginine always appears protonated at physiological pH. Besides, the positive charge is very stable due to its delocalization and confer basicity to the side chain. Positively charged residues are important in NLS sequences for nuclear import. There are two types of protein translocation into the nucleus: the classical transport via the binding of a hetero dimeric import receptor complex, constituted of importin α and importin β, and a non-classical pathway where the NLS directly interact with the nucleoporins or with importin β homologues.[96] Moreover, Divisato *et al.* (2016) have suggested that the p.Pro937Arg mutation acts as a gain of function since PDB patients, with this mutation, presented a

higher nuclear fraction of ZNF687 then healthy individuals.[38] Therefore, the introduction of a positive charge by the mutation seems to enhance the nuclear translocation, by maybe increasing its interaction with importins or transportins. Altogether, the biochemical alterations produced by the mutations in ZNF687 might interfere with its protein-protein interaction, affecting the regulation of target genes promoting PDB.

4.2 ZNF687 might regulates genes responsible for osteoblastogenesis

It was suggested by Divisato *et al.* (2016) that *ZNF687* was involved in osteoblastogenesis, as its mRNA expression was increased during this differentiation process. Based on this idea, we have evaluated the expression of several genes that are involved in the osteoblastogenesis, in both ZNF687-overexpressed (normal or mutated) and -knock-down cells.

Our results showed that the expression of *RUNX2*, one of the master genes involved in osteoblastogenesis, was affected with the alteration of *ZNF687* expression. Indeed, overexpression of ZNF687 positively affected the expression of *RUNX2*, while in knock-down clones, its expression levels had decreased, compared to normal SaOS-2. The RUNX2 transcription factor is required during the early stages of osteoblast differentiation. Indeed, several studies have demonstrated its crucial role. For instance, mice presenting a homozygous mutation in *Runx2* showed total absence of bone formation and died right after birth, due to both intramembranous and endochondral ossification arrest.[97,98] Therefore, our results might suggest that the involvement of ZNF687 in osteoblastogenesis might be mediated by the regulation of *RUNX2*.

Also, knock-down clones demonstrated low expression levels of *OSX*, the second key gene for osteoblast differentiation, suggesting that it might be regulated directly or indirectly by ZNF687. Osterix/SP7 is a zinc finger transcription factor expressed in osteoblasts, in both endochondral and membranous bones. Nakashima *et al.* (2002) have demonstrated that $Osx^{-/-}$ mice did not present formation of endochondral, nor intramembranous bone, indicating the important role of *OSX* in bone formation.[99] The same study showed that $Osx^{-/-}$ mice expressed the *Runx2* gene, while $Runx2^{-/-}$ mice did not express *Osx*. Altogether, they have demonstrated that the arrest of osteoblast differentiation in $Osx^{-/-}$ mice occurred at a later step than *Runx2*, which indicates that *OSX*

acts downstream *RUNX2*.[99] Therefore, the diminishing of *OSX* levels found in our clones might be explained by the decrease of *RUNX2*, and that ZNF687 indirectly regulates *OSX*.

Finally, FGF2 acts as an external signal by stimulating the differentiation of osteoblasts, as well as bone formation, through the Wnt signalling pathway.[100] The increase in *FGF2* expression observed in our overexpressed cells might suggest that *ZNF687* regulates this gene, and that its regulation in bone metabolism might be through an indirect pathway.

Altogether, strong evidences showed that ZNF687 regulates directly or indirectly genes that are crucial to the osteoblast differentiation.

4.3 ZNF687 might also regulate osteoclastogenesis

The overexpression of ZNF687 has triggered the expression of RANK by osteoblastic cells SaOS-2, while the ZNF687-knock-down lead to the decrease of RANK, in certain clones (C1 and C22). This might suggest that ZNF687 regulates RANK, and hence osteoclast differentiation. Osteoblasts take part in the differentiation/activation of osteoclast by secreting RANKL that will bind to its cell-surface receptor RANK, expressed by osteoclasts.[11] RANKL-RANK binding activates a variety of downstream signalling pathways required for the differentiation and survival of osteoclasts. Li *et al.* (2000) have demonstrated that $RANK^{-/-}$ mice presented profound abnormalities in bone resorption and bone remodelling as they were lacking osteoclastic cells.[101]

Furthermore, our results showed that the decrease in *ZNF687* expression had also negatively affected the expression of two genes that have been associated to PDB and that are involved in osteoclastogenesis, namely *SQSTM1* and *OPTN*. Also, overexpression of *ZNF687* lead to increased levels of *SQSTM1* and *OPTN* expression. Despite the fact that our results were obtained in osteoblastic cells, there is a clear indication that ZNF687 regulates *SQSTM1* and *OPTN* expression. *SQSTM1* encode for a ubiquitin-binding protein, p62, that is involved in osteoclast signalling pathways where it activates the RANKL-induced NF-κB (Figure 4.2). Indeed, in response to activation of RANK by RANKL, a complex formed by the adaptor protein TRAF6, p62 and an atypical protein kinase c (aPKC), is associated to the cytoplasmic domain of the receptor. After auto-ubiquitination of TRAF6, that is regulated by p62, the TRAF6, p62 and aPKC complex activates and phosphorylates the IKKβ, IKKα and NEMO complex. Once

activated, IKb, which is associated to NF-κB, is phosphorylated by IKKβ, and suffers ubiquitination and is targeted for proteasomal degradation, liberating NF-κB that will translocate into the nucleus and activate target genes expression and promote osteoclastogenesis. On the other hand, OPTN, a ubiquitously expressed protein, is a negative regulator of both TNFα- and RANKL-induced NF-κB activation (Figure 4.2).[102,103] Indeed, by presenting a high domain homology with NF-κB essential modulator (NEMO), OPTN inhibits the TNFα-induced NF-κB signalling by competitively binding to polyubiquitinated RIP (receptor-interacting protein).[102] Also, Obaid *et al.* (2015) have demonstrated that mice, with a reduced expression or loss of function of Optn, presented an increase in osteoclastogenesis, indicating that Optn is a negative regulator of osteoclast differentiation triggered by RANKL-induced NF-κB activation.[103] Moreover, it has also been suggested that OPTN modulates the interferon β (IFN-β) signalling pathway, which has an inhibitory effect on osteoclastogenesis as it interferes with RANKL-induced expression of *c-Fos*.[103,104-107] Therefore, despite stimulating osteoclast differentiation through induction of RANK and SQSTM1, if ZNF687 positively regulates OPTN as our results suggest, then ZNF687 will be able to negatively auto-regulate itself (Figure 4.2).

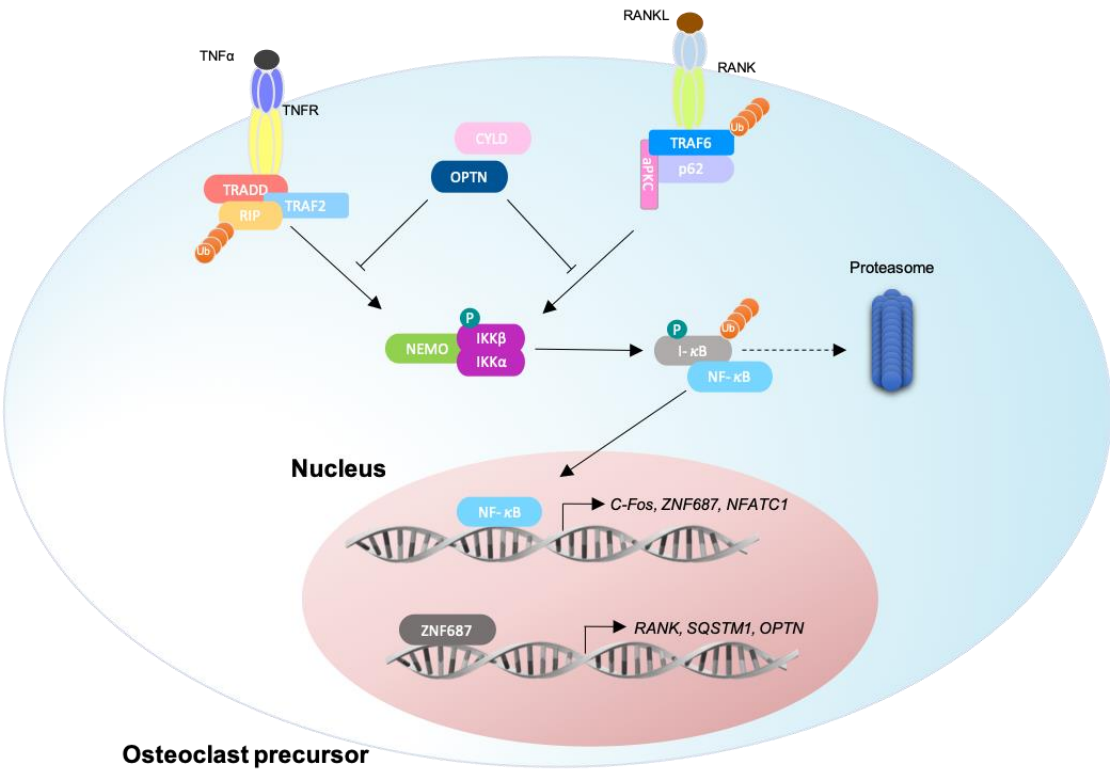


Figure 4.2 – SQSTM1/p62 and OPTN in osteoclast signalling pathways. p62 is involved in the activation of RANKL-induced NF-κB. The activation of the complex TRAF6/p62/aPKC activate in turn downstream signalling

pathway that culminate with the translocation of NF-κB into the nucleus to allow the regulation of genes involved in the osteoclastogenesis. OPTN recruits the deubiquitinating enzyme CYLD that will bind to its UBA domain in order to deubiquitinate TRAF6 disrupting the RANK-signalling. On the other hand, OPTN binds to RIP, and after recruiting CYLD deubiquitinate RIP and suppress the TNFR-signalling, preventing osteoclastogenesis. Adapted from Divisato *et al.* (2016) and Zhu *et al.* (2007).[38,102]

Yet, the regulation of *RANK*, *SQSTM1* and *OPTN* by ZNF687 observed in our results occurred in osteoblastic cells. In a future work, if the same result are obtained in osteoclastic cells, then the involvement of ZNF687 in osteoclastogenesis, as suggested by Divisato *et al.*(2016), could be determined.

4.4 The mineralization process might be indirectly modulated by ZNF687

It is still difficult to understand the role of ZNF687 in bone mineralization, since in our results some ZNF687-knock-down clones did not seem to present any alteration in the mineralization process compared to mineralized SaOS-2M, while other clones did. Indeed, in the quantification of calcium assays, only C22M presented higher level of calcium deposits, whereas C1M and C8M presented lower levels that might be due to the fact that mineralized SaOS-2 are easily detached during the removal of excessive dye, thus leading to a diminution of cell number. Nonetheless all mineralized cells presented higher levels of calcium deposit when compared to unmineralized SaOS-2. Moreover, all mineralized clones, except C1M presented a slight increase in collagen levels compared to mineralized SaOS-2. During differentiation of osteoblast, SaOS-2 cells exhibit higher expression of bone matrix protein such as *OCN*, *BSP I* and *II* and *type I collagen*, accompanied by high levels of ALP and increased levels of calcification.[108] Until now, C22M, the most prominent clone for ZNF687 knock-down is the only one who presented higher levels of calcium deposits and collagen than the unmineralized and mineralized clones. Nonetheless, clones C1M, C8 and C22M showed lower ALP levels compared to mineralized SaOS-2M, which was not expected for the latter.

The results obtained in the qPCR of mineralized clones showed that all clones depicted lower levels of *RUNX2* and *OSX* expression when compared to mineralized SaOS-2. Besides being responsible for early stage differentiation of osteoblasts, RUNX2 regulates the expression of most of the bone matrix proteins genes such as *OCN*, *BSP I* and *II* and *type I collagen* (*Col1a1*) but is not fundamental for their maintenance in mature osteoblasts.[109] Nonetheless, the expression levels of *OCN* of mineralized clones are similar to SaOS-2 level, whereas in unmineralized clones, the levels of OCN where quite

diminished compared to the control SaOS-2. Even if we can't really make a direct comparison between these results, it seems that under mineralization stimuli, the expression of OCN in knock-down cells increased. Moreover, the expression of OPTN and SQSTM1, which were lower than SaOS-2, are here also similar to the mineralized SaOS-2, except for clone C22M.

The differences found among clones in the mineralization assay and in the gene expression analysis are difficult to interpret and hypothesize a role o ZNF687 in the mineralization process of osteoblasts. Indeed, the diminution of ALP observed in certain mineralized clones might be explained by the low levels of *RUNX2*, since ALP is a direct target of RUNX2.[110] Thus, since ZNF687 seems to regulate *RUNX2*, the diminishing of ALP might indirectly be due to ZNF687. Yet, this cannot be concluded for the clones that presented ALP levels similar to mineralized SaOS-2.

4.5 The involvement of ZNF687 in PDB

The data that we have obtained in this study are still preliminary and more experiments must be conducted in order to confirm our suggestions and also to better understand the effect of overexpressed and down-regulated ZNF687 in the expression of the genes tested in this study, as well as in the mineralization process. Nonetheless, four mutation in ZNF687 (p.Ser242Ile , p.Pro665Leu, p.Gln784Glu, and p.Pro937Arg) have been identified by Divisato *et al.* (2016, 2017) but only one of them, p.Pro937Arg, has been identified as necessary and sufficient to cause PDB.[38,43] The exact mechanism by which, each one of them, predispose or cause the disease is still unknown. The p.Ser242Ile and the p.Pro665Leu occur in different parts of the protein, but no functional domain was found at their location. In addition, the residue substitution did not seem to affect the predicted secondary structure, rendering difficult to propose a putative cause for their involvement in PDB. The p.Gln784Glu mutation occurs in the C-terminal end of zinc finger motif, and might disrupt DNA binding ability. Therefore, this incapacity of binding DNA might be related to the predisposal or cause of PDB.

Nonetheless, it is very likely that ZNF687 regulates directly or indirectly *RUNX2*, *OSX*, *RANK, SQSTM1 and OPTN* expression. These genes are key genes in the differentiation process of osteoblast and osteoclasts, and any alteration of ZNF687 levels might lead to bone abnormalities, and most likely to disease. Indeed, the p.Pro937Arg mutation, causing-mutation encountered in PDB patients, seems to act as a gain of

function and enhance the translocation of ZNF687 into the nucleus, where its accumulation will increase the expression of downstream target genes. Therefore, the increase of gene expression involved in the differentiation of bone forming and resorbing cells might be the cause of the development of PDB in patients presenting this mutation. Most of mutations in *SQSTM1* that have been associated with PDB are located in the UBA domain, resulting in abnormalities in ubiquitin-mediated degradation of proteins that interacts with SQSTM1, thus causing an increased ubiquitination of TRAF6 and activation of NF- κB .[111-112] In addition, Divisato *et al.* (2016) have demonstrated that PDB patients with mutated *SQSTM1* presented a higher *ZNF687* expression than non-mutated PDB patients and healthy person, thus suggesting that ZNF687 might be a downstream target of NF-κB.[38] Also, an overexpression of SQSTM1/p62 was observed in PDB patients, with or without mutations, and might be involved in the pathophysiology of PDB.[43,112] In addition, in certain circumstances, OPTN has proved to potentiate NF-κB activation. Indeed, under viral infection of human T-lymphotropic virus type 1 (HTLV-1), OPTN cooperate with TAX1 binding protein to enhance the activation of NF-κB.[113] Also, Silva *et al.* (2018) have demonstrated that the T allele present in the rs1561570 variant, within *OPTN* and associated to PDB, lead to the loss of a methylation site resulting in an increase of *OPTN* expression.[114-115] This overexpression of *OPTN* was correlated to an increase of NF-κB translocation into the nucleus, resulting in higher expression of its target genes and hence an increase in osteoclastogenesis.

4.6 The gene *znf687b* is the ortholog for the human *ZNF687*

The genomic structure of *ZNF687* was rather similar among different species, with higher similarity among mammals as expected. Zebrafish *znf687b*, showed more similarity with the human genomic sequence than its paralog *znf687a*, in terms of number and size of the exons. Also, the coding DNA sequence was very much akin throughout evolution, with similar length, suggesting a certain conservation in the function of the protein. Nonetheless, the analysis of syntenic regions and chromosomal locations that harbour *ZNF687* have demonstrated that the neighbouring genes of *ZNF687* orthologues are only conserved between mammalian and avian species, and only few are preserved in amphibian and teleost species. Additionally, in our *in silico* attempt to identify conserved cis-regulatory TFBSs among *znf687a* and *znf687b* promoter sequences from zebrafish and human *ZNF687*, we verified that in some cases, the spatial disposition of the TFBSs

was not necessarily maintained but they were present in the sequence, enabling the prediction of conserved regulatory elements within the promoter sequences analysed.

It has been demonstrated that *ZNF687* mRNA was upregulated during osteoclastogenesis and osteoblastogenesis, in both human and zebrafish, suggesting that this gene plays a role in bone metabolism in both species, but also, that the function of this protein might be conserved by evolution despite speciation.[38] Therefore, the pairwise alignment of all sequences studied revealed an overall protein homology that is mainly conserved among mammalian species (86,3-99,8%) whereas lower sequence identity was observed between *Homo sapiens* and non-mammalian sequences, with the lowest percentages observed with teleost species (~36,5%). Once more, znf687b demonstrated higher identity with human protein (40.3%) than its paralog (34.6%). However, there is a region that appears to be more conserved between all selected species, with a mean percentage of identity reaching 62% between human and znf687b and 57.1% between human and znf687a. This conserved region corresponds to the zinc finger motifs encountered in the human protein, suggesting that the function of this transcription factor, endorsed by the ZF motifs, is well preserved despite speciation. The comparative domain analysis confirmed such hypothesis where all species maintain the same pattern of clustered ZF motif. znf687b accounts one additional ZF motif located in the C-terminal of the protein, whereas the Znf687a, which also possesses this extra ZF motif, lacks two of the 14 human ZF motifs.

Finally, among the four residues that have been found to be mutated and associated to PDB, only the glutamine (p.Gln784Glu) appears perfectly conserved throughout evolution. This result suggests that zebrafish could be used as a biological model to study this mutation and its involvement in PDB.

In the overall, znf687b presented more similarities with the human ZFN687 than its paralog znf687a, suggesting that it might be consider as the ortholog for human ZNF687.

4.7 zebrafish *znf687a* and human znf687 promoter are regulated by transcription factors involved in the regulation of bone development

In order to test the ability of the human and zebrafish *ZNF687* promoters to direct transcription, four fragments (F1, F2, F3 and F4) were design in the human promoter region whereas for zebrafish, only one fragment was draw for each of *znf687* gene (FA and FB). Unfortunately, after encountering difficulties in the amplification of F3, F4, and FB, the construction of reporter vectors with these fragments could not be achieved in time, thus preventing the measurement of their transcriptional activity. Nonetheless, our *in vitro* transient transfection-reporter analysis identified a region in the zebrafish *znf687a* (FA) and two regions in the human *ZNF687* (F1 and F2) as being responsible for its regulation. Moreover, upon co-transfection assay we observed a repression effect due to YY-1 in *znf687a*, and to AP-1 in *ZNF687*. The zinc finger transcription factorsYY-1 regulates positively and negatively the expression of several genes. It has been showed that YY-1 could repress the expression of bone morphogenic proteins that are responsible for bone formation.[116] The identification of a repressive activity mediated by YY-1 on the promoter region of *znf687a* sustained the role of this gene in bone formation, and in osteoblast differentiation, suggesting that *znf687a* depict a similar role as the human *ZNF687*. Moreover, the AP-1 components, such as c-Fos and Jun, play important function in osteoclast and osteoblast, supporting the involvement of ZNF687 in bone metabolism.[117] These results provide for the first time the identification and characterization of promoter region that regulate the transcription of *znf687a* and *ZNF687* genes and offer insights into the regulation of *znf687a* by YY-1 and *ZNF687* by AP-1.

Altogether, these studies support the usefulness of comparative genomic to uncover gene regulatory sequences based on evolutionary conservation and provide the basic information to explore and better understand the regulation and expression of *ZNF687*.

4.6 Limitation of this study

We have mediated a genome editing by CRISPR-Cas9 in order to obtain ZNF687 knock-out stable clones. The phenotype observed in six clones, presenting one or more indel mutations, and exhibiting a knock-down rather than a knock-out, might be explained by the homogeneity of a clone population, that can be called in question. Undeniably, there is a possibility that a clone population may have emerged from two or more distinct cells, with different genotypes. Moreover, the SaOS-2 cell lines that we used are cells derived from an osteosarcoma and do not present a normal diploidy. Indeed, the

hyperdiploidy of SaOS-2 might affect the generation of knock-out, as the cell may possess more than just one pair of chromosome 1.[118-119] Also, the CRISPR-Cas9 system present three possible out-comes: (1) a DBS is generated in the two alleles, resulting in a perfect knock-out, (2) only one of the two alleles suffered a DBS, causing a knock-down, or (3) the transfection failed and no DBS were generated due to the absence of CRISPR-Cas9 system in the cell. Looking at the phenotype of the cells, it is very likely that most of the clones depict the second out-come. All these possibilities could also explain the differences obtained between clones in the mineralization assays. Indeed, a loss of stable clones in heterogenic population, due to overgrowing of wild-type cells, with the number of passages, can influence the phenotype of the cell population analysed.

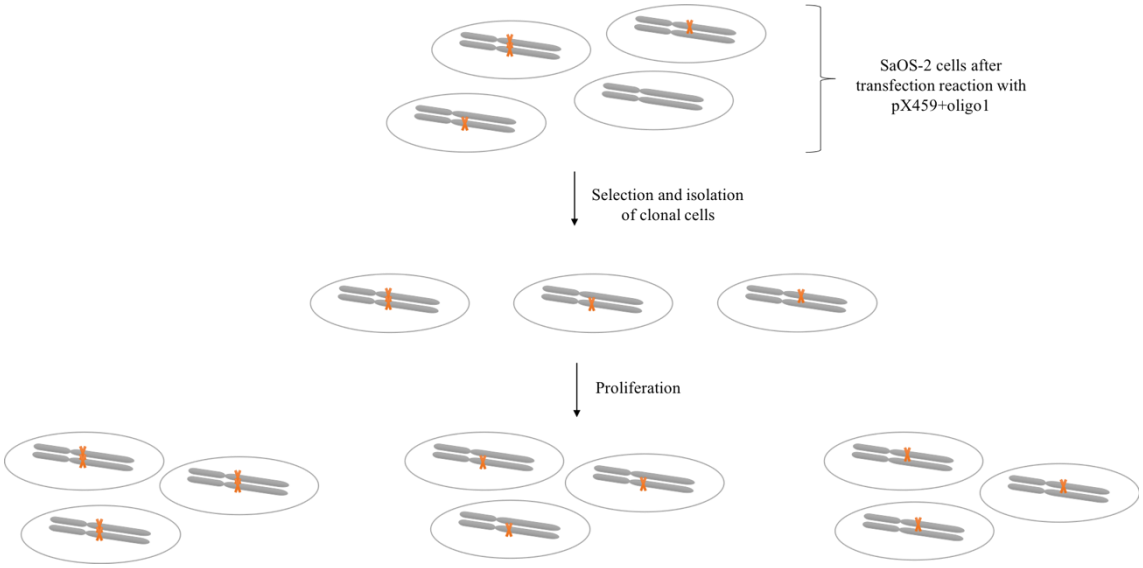


Figure 4.3 – Schematic representation of the CRISPR process.

In addition, the number of experiments in the gene expression analysis in both ZNF687 knock-down and overexpression assays need to be repeated in order to confirm the preliminary results obtained but also to try to understand the pattern of expression of certain genes that were difficult to interpret due to outlined values.

Another limitation of this work was the transfection of THP-1 cells. These monocyte cells are non-adherent, sensitive and very difficult to transfect. An optimized protocol for the transfection by nucleofection of these cells was proposed by Schnoor *et al* (2009) and should be experimented in future work in order to conduct a similar study as the one we performed in osteoblastic cells SaOS-2, but also to confirm the regulation of RANK, SQSTM1 and OPTN in osteoblastic cells by ZNF687.[120]

CHAPTER V – CONCLUSION AND FUTURE PERSPECTIVE

In conclusion, this work has provided new evidences about ZNF687 involvement in bone differentiation and development that might contribute to explain its role in PDB. Through transient transfections of overexpressed ZNF687, and genome editing mediated by CRISPR-Cas9, we have demonstrated that ZNF687 seems to regulate several genes that are key to the differentiation process of both osteoblastic and osteoclastic cells. Indeed, *RUNX2*, the master gene of early differentiation stages of osteoblasts, finds himself up- and down-regulated in ZNF687 overexpressed and knock-down cells, respectively. Moreover, *OSX/SP7*, another gene involved in osteoblastogenesis acting downstream RUNX2, was downregulated in knock-down clones, indicating that *RUNX2* and *OSX/SP7* are downstream targets of ZNF687. Additionally, *OPTN* and *SQSTM1* expression were downregulated, in the knock-down ZNF687 clones, while *RANK* expression was positively regulated by the overexpression of ZNF687. Since these three genes are involved in the NF-κB signalling pathway, responsible for the differentiation of osteoclasts, we can suggest that ZNF687 may, directly or indirectly, regulate *RANK*, *OPTN* and *SQSTM1*, and by doing so, play a role in osteoclastogenesis.

Moreover, our preliminary results obtained in the analysis of the promoter regions of *ZNF687* showed that AP-1, involved in the regulation of bone development, was able to negatively regulate the *ZNF687* promoter, sustaining the fact that ZNF687 is involved in osteoblastogenesis and osteoclastogenesis.

Unfortunately, the involvement of ZNF687 in the mineralization process still remains uncertain as clones did not present the same pattern in the mineralization assay and in the gene expression analysis. The expression of *RUNX2* and *OSX/SP7* remained under the basal level of mineralized SaOS-2, and might be due to the low levels of ALP encountered in certain clones. But our results in the mineralization assay are not very conclusive.

By regulating genes that are involved in osteoclastogenesis and osteoblastogenesis, mutations in ZNF687 might lead to abnormal bone remodelling and the development of bone disorders, such as Paget's disease of bone. The exact mechanism by which the four mutations identified by Divisato *et al.* (p.Ser242Ile, p.Pro665Leu,

p.Gln784Glu and p.Pro937Arg) predispose or cause the disease, is still unknown. Nonetheless, the p.Gln784Glu mutation that occurs in the C-terminal end of zinc finger motif might disrupt DNA binding ability, and therefore this incapacity of binding DNA might be related to the predisposal or cause of PDB. Finally, the expression of mutated ZNF687 (p.Pro937Arg) was rather similar to the overexpressed WT ZNF687 for all the genes studied. Therefore, the increase of gene expression involved in the differentiation of bone forming and resorbing cells might be the cause of the development of PDB in patients presenting the p.Pro937Arg mutation.

Despite similar genomic structure and coding sequence among human and zebrafish species, the comparative mapping of *ZNF687* genes and the analysis of syntenic regions and chromosomal locations that harbour *ZNF687* have demonstrated that the neighbouring genes of *ZNF687* orthologues are poorly preserved between human and teleost species. Additionally, the presence, within their promoter regions, of similar putative binding sites for common transcriptional factors, advocated a similar functional role of ZNF687, Znf687a and Znf687b. Moreover, even if the overall protein homology between *Homo sapiens* and zebrafish was poor, the region corresponding to zinc finger proteins presented higher identity, with a conserved pattern of ZF clusters. Finally, among the four residues that have been encountered mutated and associated to PDB, only the glutamine (p.Gln784Glu) appears perfectly conserved throughout evolution. This result demonstrates that zebrafish could be used as a biological model to study this mutation. In the overall, the *znf687b* presented more similarities with the human *ZFN687* than its paralog *znf687a*, suggesting that it might be considered as the ortholog for human *ZNF687*. Nonetheless, the identification of a repressive activity mediated by YY-1 on the promoter region of *znf687a* sustained the role of Znf687a in bone formation, and in osteoblast differentiation, suggesting that Znf687a depict a similar role as the human ZNF687. All these findings suggest that zebrafish could be a suitable biological model to further analyse *ZNF687* regulation and involvement in bone metabolism.

This study gave insight of ZNF687 role in bone metabolism and hence in PDB, but also gave a grasp of *ZNF687* regulation. Since the data that we have obtained are still regarded as preliminary results, due to the few *in vitro* experiments that have been conducted in the human and zebrafish promoter regulation, but also in the gene expression, more replicates have to be performed in order to validate the suggestions that have been made. Also, the transfection of THP-1 cells with CRSIPR-Cas9 constructs and

ovexpressed-ZNF687 plasmids have failed, preventing the realization of one part of this study. Therefore, an optimization protocol must be put in place in order to successfully transfect THP-1 cells, and after differentiation into osteoclasts, evaluate the gene expression as well as the mineralization process, as it was done in osteoblastic cells, SaOS-2. Once we have obtained conclusive findings regarding the target genes of ZNF687, in both osteoblastic and osteoclastic cells, we could perform a similar study using zebrafish as a model that would allow us to grip data on znf687a and znf687b target genes and their involvement in the bone metabolism. Therefore, we could generate *in vivo* zebrafish knock-out, mediated by CRISPR-Cas9. To accomplish this, a Cas9 protein and a sgRNA targeting *znf687a, znf687b* and *znf687a/b* would be introduced together into each target cell. The injected larvae would then be analysed to detect some phenotypic defects. The founders should be outcrossed in order to generate F1 embryos, that will then be genotyped and sequenced, by fin clipping. It would also be necessary to confirm if the mutants generated yield full loss-of-function phenotypes. This could be done by performing a Western blot to detect the presence or absence of protein.

Another study could be performed in order to evaluate if the other three mutations in *ZNF687* identified by Divisato *et al.* (p.Ser242Ile, p.Pro665Leu and p.Gln784Glu) could affect the gene expression of ZNF687 target genes. Indeed, each mutation could be performed by site-directed mutagenesis, and mutated DNA should be inserted in an expression vector possessing a eukaryotic selection cassette to generate stable clones. Then, osteoblastic and osteoclastic human cell lines would be transfected with the construction carrying one mutation and after selection of transfected cells, the mineralization process and the expression of target genes could be analysed. Results could provide some insight into the mechanism of action of each one of these mutations.

REFERENCES

[1] V. Tzelepi, A. C. Tsamandas, V. Zolota, and C. D. Scopa. Bone Anatomy, Physiology and Function. D. Kardamakis et al. (eds.) *Bone Metastases*, Cancer Metastasis – Biology and Treatment, vol. 12.

[2] J. E. Shea, and S. C. Miller, Skeletal function and structure: Implications for tissue-targeted therapeutics, Advanced Drug Delivery Reviews 57 (2005) 945–957

[3] Bart Clarke, Normal bone anatomy and physiology, Clinical Journal of the American Society of Nephrology, (2008) 3(Suppl 3): S131–S139.

[4] E. N. Marieb, P. B. Wilhem and J. B. Mallatt. Human Anatomy 7th edition. ISBN-13: 978-0321822413.

[5] M. Michael Cohen Jr., The new bone biology: Pathologic, olecular, and clinical correlates. American Journal of Medicl Genetics Part A 140A:2646–2706.

[6] Adele L. Boskey, Bone composition: relationship to bone fragility and antiosteoporotic drug effects, BoneKEy Reports 2, Article number: 447 (2013)

[7] Chiara Gentili and Ranieri Cancedda, Cartilage and Bone Extracellular Matrix, Current Pharmaceutical Design (2009), 15, 1334-1348.

[8] Andrea I. alford, Kenneth M. Kozloff, Kurt D. Hankenson, Extracellular matrix networks in bone remodeling, The International Journal of Biochemistry & Cell Biology 65 (2015) 20–31.

[9] Joana Caetano-Lopes, Helena Canhão and João Eurico Fonseca, Osteoblast and Bone Formation, Orgão Oficial da Sociedade Portuguesa de Reumatologia, Acta Reumatologica Portuguesa (2007), 32, 103-110.

[10] Arkady Rutkovskiy, Kåre-Olav Stensløkken, and Ingvar Jarle Vaage, Osteoblast Differentiation at a Glance, Medical Science Monitor Basic Research (2016), 22, 95-106.

[11] Akira Yamaguchi, Regulatory Mechanism of Osteoblast Differentiation and Bone Formation, Journal of Oral Pathology and Medicine (2000), 5, 1-14.

[12] L. Arboleya and S. Castañeda. Osteoimmunology: the study of the reliationship between the human system and bone tissue. Reumatologia Clinica (2014); 9(5):303-315.

[13] Kazuhisa Nakashima and Benoit de Crombrugghe, Transcriptional mechanisms in osteoblast differentiation and bone formation, TRENDS in Genetics (2003) Vol.19 No.8

[14] Christian Frantz, Kathleen M. Stewart, and Valerie M. Weaver, The extracellular matrix at a glance, Journal of Cell Science (2010); 123(24): 4195–4200.

[15] E.J. Mackie. Osteoblasts: novel roles in orchestration of skeletal architecture; The International Journal of Biochemistry & Cell Biology 35 (2003) 1301–1305

[16] M. Michael Cohen Jr. The new bone biology: pathologic, molecular, and clinical correlates; American Journal of Medical Genetics Part A 140A:2646-2706 (2006).

[17] T. Miyamoto and T. Suda. Differentiation and function of osteoclasts. The Keio Journal of Medicine (2003) 52, 1-7.

[18] B. F. Boyce, E. Rosenberg, A. E. de Papp and L. T. Duong. The osteoclast, bone remodeling and treatment of metabolic bone disease. European Journal of Clinical Investigation (2012).

[19] K. Inoue and Y. Imai. Identification of novel transcription factors in osteoclasts differentiation using genome-wide analysis of open chromatin determined by DNase-seq. American Society for Bone and Mineral Research (2014), vol.29, nº8, 1823-1832.

[20] B. Zhao and L.B. Ivashkiv. Negatie regulation of osteoclastogenesis and bone resorption by cytokines and transcriptional repressors. Arthritis Research Therapy (2011); 13:234.

[21] H. C. Blair. How the osteoclast degrades bone; BioEssays (1998), 20:837-846.

[22] M. Capulli, R. Paone and N. Rucci. Osteoblast and osteocyte: game without frontiers. Archive of Biochemistry and Biophysics (2015) 75:144-50.

[23] R. F. M. van Oers, H. Wang and R. Bacabac. Osteocyte shape and mechanical loading. Current Osteoporosis Reports (2015), 13:61-66.

[23] X. Feng. Vhemical and biochemical basis of cell-bone matrix interaction in health and disease; Current Chemical Biology (2009), 3(2):189-196.

[25] E. Seeman and P. D. Delmas; Bone quality – the material and structural basis of bone strength and fragility. The New England Journal of Medicine (2006), 354:2250-2261.

[26] C. Zuo, Y. Huang, R. Bajis, M. Sahih, Y-P Li, K. Dai and X. Zhang. Osteoblastogenesis signals in bone remodeling. Osteoporosis International (2012) 23:1653–1663

[27] E. Canalis, A. Giustina, and J. P. Bilezikian. Mechanism of anabolic therapies for osteoporosis. The New Engalnd Journal of Medicine (2007); 357(9):905-916.

[28] G. White and J. Rushbrook. Paget's Disease of Bone. Orthopaedics and trauma (2013) 27:42-54.

[29] K. Cortis, K. Micallef, and A. Mizzi. Imaging Paget's Disease of Bone – from head to toe. Clinical Radiology (2011), 66:662-672.

[30] S. H. Ralston. Paget's Disease of Bone. The New England Journal of Medicine (2013), 368:644-650

[31] S. H. Ralston, A. L. Langston, I. R. Reid. Pathogenesis and management of Paget's disease of bone. Lacet Seminar 372 (2008), 155-163.

[32] D. L. Galson and G. D. Roodman. Pathobiology of Paget's Disease of Bone. Journal of Bone Metabolism (2014), 21:85-98.

[33] G. D. Roodman and J. J. Windle. Paget's Disease of Bone. The Journal of Clinical Investigation (2005), 115:200-208.

[34] E. S. Siris, G. D. Roodman. Paget's disease of bone. Favus MJ, ed. Primer on the metabolic bone diseases and disorders of mineral metabolism, 6th ed. Washington D.C.: American Society for Bone and Mineral Research (2006), 320-9.

[35] N. Deep, C. L. W. Driscoll, J. I. Lane, M. Carlson and J. G. Besh-Stokes. Paget's Disease of the Temporal Bone: A Single-Institution Contemporary Review of 27 Patients. Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology (2017).

[36] M. H. Helfrich and L. J. Hocking. Genetics and aetiology of pagetic disorders of bone. Archives of Biochemistry and Biophysics 473 (2008), 172-182.

[37] F. Gianfrancesco, D. Rendina, D. Merlotti, T. Esposito, M. Amyere, D. Formicola, R. Muscariello, G. De Filippo, P. Strazzulo, R. Nuti, M. Vikkula and L. Gennari. Giant cell tumor occurring in familial Paget's disease of bone: report of clinical characteristics and linkage analysis of a large pedigree. American Society for Bone and Mineral Research (2013), 28(2):341-350.

[38] G. Divisato, D. Formicola, T. Esposito, D. Merlotti, L. Pazzaglia, A. Del Fattore, E. Siris, P. Orcel, J. P. Brown, R. Nuti, P. Strazzullo, M. S. Benassi, M. L. Cancela, L. Michou, D. Rendina, L. Gennari, and F. Gianfrancesco. ZNF687 mutations in severe Paget disease of bone associated with giant cell tumor. The American Journal of Human Genetics 98 (2016), 275–286.

[39] S. Mays. Archaeological skeletons support a northwest European origin for Paget's disease of bone. Journal of Bone and Mineral Research (2010), 25: 1839-1841.

[40] O. M. E. Albagha, S. H. Ralston. Chap. 3 - Genetics of Paget disease of bone. Advances on Pathobiology and Management of Paget's Disease of Bone (2016), 25-36.

[41] S. L. Rea, V. Majcher, M. Searle, and R. layfield. SQSTM1 mutations – bridging Paget disease of bone and ALS/FTLD. Experimental Cell Research (2014), 325:27-37.

[42] X. Qi, Q. Pang, J. Wang, Z. Zhao, O. Wang, L. Xu, J. Mao, Y. Jiang, M. Li1, X. Xing, W. Yu, Asan, and W. Xia. Familial Early-Onset Paget's Disease of Bone Associated with a Novel hnRNPA2B1 Mutation. Calcification Tissue International (2017), 101:159–169.

[43] G. Divisato, F. Scotto di Carlo, N. Petrillo, T. Esposito, and F. Gianfrancesco. ZNF687 mutations are frequently found in pagetic patients from South Italy: implication in the pathogenesis of Paget's disease of bone. Clinical Genetics (2018), 93:1240–1244.

[44] N. Laurin, J. P. Brown, J. Morissette, and V. Raymond. Recurrent Mutation of the Gene Encoding sequestosome 1 (SQSTM1/p62) in Paget Disease of Bone. American Journal of Human Genetics (2002), 70:1582–8.

[45] G. D. Roodman, J. J. Windle. Paget disease of bone. J. Clinicial Investigation 115 (2005), 200-208.

[46] F. R. Singer. Paget's of bone – genetic and environmental factors. Nature Reviews Endocrinology (2015), 138.

[47] M. Papworth, P. Kolasinska and M. Minczuk. Designer zinc-finger proteins and their applications. Gene (2006), 366:27-38.

[48] S. V. Razin, V. V. Borunova, O. G. Maksimenko, and O. L. Kantidze. Cys2His2 Zinc Finger Protein Family: Classification, Functions, and Major Members. Biochemistry (Moscow) (2012), 77(3): 277 288.

[49] A. Malovannaya, R. B. Lanz, S. Y. Jung, Y. Bulynko, N. T. Le, D. W. Chan, C. Ding, Y. Shi, N. Yucer, G. Krenciute, B. J. Kim, C. Li, R. Chen, W. Li, Y. Wang, B. W. O'Malley, and J. Qin. Analysis of the Human Endogenous Coregulator Complexome. Cell (2011), 145(5): 787–799.

[50] T. Splettstoesser (www.scistyle.com) - self-made, based on PDB structure 1A1L, the open source molecular visualization tool PyMol and Cinema 4D, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=3106866

[51] T. T. Nguyen, L. N. Ma, M. L. Slovak, C. D. Bangs, A. M. Cherry and D. A. Arber. Identifikation of novel *RUNX1* (*AML1*) translocation partner genes *SH3D19*, *YTHDF2*, and *ZNF687* in acute myeloid leukemia. Genes, Chromosomes, and Cancers (2006), 45:918-932.

[52] T. Zhang, Y. Huang, W. Liu, W. Meng, H. Zhao, Q. Yang, S-J. Gu, C-C. Xiao, C-C. Jia, B. Zhang, Y. Zou, H-P. Li and B-S. Fu. Overexpression of zinc finger protein 687 enhances tumorigenic capability and promotes recurrence of hepatocellular carcinoma. Oncogenesis (2017), 6.

[53] Y. M. Bradford, S. Toro, S. Ramachandran, L. Ruzicka, D. G. Howe, A. Eagle, P. Kalita, R. Martin, S. A. Taylor Moxon, K. Schaper, and M. Westerfield. Zebrafish Models of Human Disease: Gaining Insight into Human Disease at ZFIN. ILAR Journal (2017), 58(1): 4–16.

[54] R. L. Vaz, T. F. Outeiro, J. J. Ferreira. Zebrafish as an Animal Model for Drug Discovery in Parkinson's Disease and Other Movement Disorders: A Systematic Review. Front. Neurol. (2018), 9(347)

[55] T.O. Auer, K. Duroure, A. De Cian, J-P. Concordet, and F. Del Bene. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish byhomology-independent DNA repair. Genome Research (2018), 28.

[56] S. Ota, Y. Hisano, Y. Ikawa, and A. Kawahara. Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. Genes to Cells (2014), 19: 555–56

[57] S. Sivasubbu, D. Balciunas, A. Amsterdam, and S. C. Ekker. Insertional mutagenesis strategies in zebrafish. Genome Biol. (2007), 8:S9.

[58] U. J.Pyatia, A. T. Look,and M. Hammerschmidt. Zebrafish as a powerful vertebrate model system for in vivo studies of cell death. Seminar in Cancer Biology (2007); 17(2):154-165.

[59] D. R. Zerbino et al., Ensemble 2018, Nucleic Acids Research (2018) Vol.46, Issue D1, D754-D761.

[60] Y. Kapustin, A. Souvorov, T. Tatusova, & D. Lipman. Splign: algorithms for computing spliced alignments with identification of paralogs. Biology Direct 3, 20 (2008).

[61] UCSC Genome Browser: Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Research (2002);12(6):996-1006.

[62] T.Tsunoda, and T.Takagi. Estimating Transcription Factor Bindability on DNA. BIOINFORMATICS, Vol.15, No.7/8, pp.622-630, 1999.

[63] Ł. Kreft, A. Soete, P. Hulpiau, A. Botzki, Y.Saeys, P. De Bleser; ConTra v3: a tool to identify transcription factor binding sites across species, Nucleic Acids Res 2017

[64] N. Grabe, AliBaba2: Context Specific Identification of Transcription Factor Binding Sites, In Silico Biology 2 (2002) S1-S15

[65] N. T. T. Nguyen, P. Vincens, H. R. Crollius, A. Louis, Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing, Nucleic Acids Research (2018), Vol. 46, Issue D1, D816-D822.

[66] F. Sievers, A. Wilm, D.G. Dineen, T.J. Gibson, K. Karplus, W. L, R. Lopez, H. McWillia, M. Remmert, J. Söding, J.D. Thompson and D.G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology (2011); 7:539

[67] G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmi, A. Schafferhans, M. Roos, M. Bernhofer, and others. PredictProtein - an open resource for online prediction of protein structural and functional features. Nucleic acids research (2014).

[68] Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research. 41 (W1): W340-W348.

[69] D.T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology (1999) 292: 195-202.

[70] M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu & J. Xu. Template-based protein structure modeling using the RaptorX web server. Nature Protocols (2012), 7, 1511–1522.

[71] R. Heffernan, A. Dehzangi, J. Lyons, K. Paliwal, A. Sharma, J.a Wang, A. Satta, Y.Zhou and Y. Yang. Highly Accurate Sequence-based Prediction of Half-Sphere Exposures of Amino Acid Residues in Proteins. Bioinformatics (2016); 32 (6), 843-849

[72] R. Heffernan, K. Paliwal, J. Lyons, A.Dehzangi, A. Sharma, J. Wang, A.Sattar, Y. Yang and Y. Zhou. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Scientific Report 2015 (5),11476.

[73] B. Petersen, T. N. Petersen, P. Andersen, M. Nielsen and C. Lundegaard. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Structural Biology (2009); 9:51

[74] G. P. S. Raghava. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. CASP5. (2002) A-132.

[75] R. Adamczak, A. Porollo, J. Meller, Combining Prediction of Secondary Structure and Solvent Accessibility in Proteins, Proteins: Structure, Function and Bioinformatics, (2005), 59:467-75.

[76] H. Venselaar, T. A. H. te Beek, R. K.P. Kuipers, M. L. Hekkelman and G. Vriend. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics (2010) 11:548

[77] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman. The Pfam protein families database: towards a more sustainable future: Nucleic Acids Research (2016). BMC Structural Biology (2009), 9:51.

[78] CJA Sigrist, E. de Castro, L. Cerutti, B.A Cuche, N. Hulo, A. Bridge, L. Bougueleret, I. Xenarios. New and continuing developments at PROSITE Nucleic Acids Reserach. (2012).

[79] The UniProt Consortium, UniProt: the universal protein knowledgebase, Nucleic Acids Research (2017); 45: D158-D169.

[80] I. Letunic and P. Bork. 20 years of the SMART protein domain annotation resource Nucleic Acids Research (2017)

[81] A. Marchler-Bauer *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research (2017); 45(D)200-3.

[82] M. Brameier, A. Krings, R.M. Maccallum. NucPred - Predicting Nuclear Localization of Proteins.. Bioinformatics (2007); PubMed id: 17332022

[83] S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, and H. Yanagawa. Six classes of nuclear localization signals specific to different binding grooves of importin α. Journal of Biological Chemistry (2009) 284, 478-485.

[84] N. B. An, A. Pogoutse, N. Provart, A.M. Moses. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics. 2009 Jun 29;10(1):202.

[85] J. Sambrook, E. F. Fritshch, T. Maniatis. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press (1989).

[86] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang. Genome enineering using the CRISPR-Cas9 system. Nature Protocols, vol. 8, n. 11 (2013), 2281-2308.

[87] D. Rath, L. Amlinger, A. Rath and M. Lundgren. The CRISPR-Cas immune system: biology, mechanisms and applications. Biochimie 117 (2015), 119-128.

[88] Chap. 1 - Genome engineering overview. CRISPR-101: A Desktop Ressource, Created and Compiled by Addgene, 1st Edition (2016), 5-9.

[89] Chap. 2 - What is CRISPR. CRISPR-101: A Desktop Ressource, Created and Compiled by Addgene, 1st Edition (2016), 10-29.

[90] T. Sakuma and T. Yamamoto. Chap - 2 CRISPR/Cas9: The leading edge of genome editing technology. Targeted Genome Editing Using Site-Specific Nucleases (2015), 25-41.

[91] OJ. Marshall. PerlPrimer: cross-platform, graphical primer design for standart, bisulphite and real-time PCR. Bioinformatics 2004 20(15):2471-2472

[92] M.A. Innis, D. H. Gelfand, J.J. Sninsky, and T.J. White eds. (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego, CA.

[93] S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, and H. Yanagawa. Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin α. The Journal of Biological Chemistry (2009), 284(1):478-485.

[94] L. Stubbs, Y. Sun, and D. Caetano-Anolles. Function and Evolution of C2H2 Zinc Finger Arrays. T.R. Hughes (ed.), A Handbook of Transcription Factors, Subcellular Biochemistry 52.

[95] C.A. Fitch, G. Platzer, M. Okon, B. Garcia-Moreno, and L. P. McIntosh. Arginine: Its pKa value revisited. Protein Science (2015) 24:752-761.

[96] A. S. Philips, J. C. Kwok, and B. H. Chong. Analysis of the Signals and Mechanisms Mediating Nuclear Trafficking of GATA-4. The Journal of Biological Chemistry (2007), 282(34): 24915-24927.

[97] T. Komori, H. Yagi, S. Nomura, A. Yamaguch, K. Siasaki, K. Deguchi, Y. Shimizu, R. T. Bronson, Y. H. Gao, M. Inada, M. Sato, R. Okamoto, Y. Kitamura, S. Yoshiki, T. Kishimoto. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell (1997); 89: 755-764.

[98] F. Otto, A. P. Thornell, T. Crompton, A. Denzel, K. C Gilmour, I. R. Rosewell, G. W. Stamp, R, S. Beddington, S. Mundlos, B. R. Olsen, P. B Selby, M. J. Owen. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell (1997); 89: 765-771.

[99] K. Nakashima, X. Zhou, G. Kunkel, Z. Zhang, J. M. Deng, R. R. Behringer, and B. de Crombrugghe. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. Cell (2002); 108(1):17-29.

[100] Y. Fei, L. Xiao, T. Doetschman, D.J. Coffin, and M.M. Hurley. Fibroblast growth factor 2 stimulation of osteoblast differentiation and bone formation is mediated by modulation of the wnt signaling pathway. Journal of Biological Chemistry (2011); 286(47): 40575–40583.

[101] J. Li, I. Sarosi, X-Q. Yan, S. Morony, C. Capparelli, H-L. Tan, S. McCabe, R. Elliott, S. Scully, G. Van, S. Kaufman, S-C Juan, Y. Sun, J. Tarpley, L. Martin, K.

Christensen, J. McCabe, P. Kostenuik, H. Hsu, F. Fletcher, C. R. Dunstan, D.L. Lacey, and W. J. Boyle. RANK is the intrinsic hematopoietic cell surface receptor that controls osteoclastogenesis and regulation of bone mass and calcium metabolism. Proceeding of the National Academy of Sciences USA (2000); 97(4): 1566–1571.

[102] Zhu, G., Wu, C.J., Zhao, Y., and J.D. Ashwel. Optineurin negatively regulates TNFalpha- induced NF-kappaB activation by competing with NEMO for ubiquitinated RIP. Current Biology (2007), 17, 1438–1443.

[103] R. Obaid, S. E. Wani, A. Azfer, T. Hurd, R. Jones, P. Cohen, S. H. Ralston, and O. M.E. Albagha. Optineurin negatively regulates osteoclast differentiation by modulating NF-κB and interferon signaling: implications for Paget's disease. Cell Reports (2015); 13:1096–1102.

[104] H. Takayanagi, S. Kim, K. Matsuo, H. Suzuki, T. Suzuki, K.Sato, T. Yokochi, H. Oda, K. Nakamura, N. Ida, *et al*. RANKL maintains bone homeostasis through c-Fos-dependent induction of interferon-beta. Nature (2002); 416, 744–749.

[105] T. Hayashi, T. Kaneda, Y. Toyama, M. Kumegawa, and Y. Hakeda. Regulation of receptor activator of NF-kappa B ligand-induced osteoclastogenesis by endogenous interferon-beta (INF-beta) and suppressors of cytokine signaling (SOCS). The possible counteracting role of SOCSs- in IFN-beta-inhibited osteoclast formation. Journal of Biological Chemistry (2002); 277, 27880–27886.

[106] C.E. Gleason, A. Ordureau, R. Gourlay, J.S. Arthur, and P. Cohen. Polyubiquitin binding to optineurin is required for optimal activation of TANK-binding kinase 1 and production of interferon b. Journal of Biological Chemistry . (2011); 286, 35663–35674

[107] I. Munitic, M.L. Giardino Torchia, N.P. Meena, G. Zhu, C.C. Li, and J.D. Ashwell, Optineurin insufficiency impairs IRF3 but not NF-kB activation in immune cells. Journal of Immunology (2013); 191, 6231–6240.

[108] X. Yin, Z. Chen, Z. Liu, and C. Song. Tissue transglutaminase (TG2) activity regulates osteoblast differentiation and mineralization in the SAOS-2 cell line. Brazilian Journal of Medical and Biological Research (2012); 45: 693-700.

[109] T. Komori. Regulation of bone development and maintenance by Runx2. Frontiers Biosciences (2008); 1;13:898-903

[110] J.J. Weng and Y.Su. Nuclear matrix-targeting of the osteogenic factor Runx2 is essential for its recognition and activation of the alkaline phosphatase gene. Biochimica and Biophysica Acta (2013); 1830(3):2839-52.

[111] A. Daroszewska, and S. H. Ralston. Genetics of Paget's disease of bone. Clinical Science (2005); 109: 257–263.

[112] S. H. Ralston, & R. Layfield. Pathogenesis of Paget Disease of Bone. Calcified Tissue International (2012); 91(2), 97–113.

[113] C. Journo, J. Filipe, F. About, S.A. Chevalier, P.V. Afonso, J.N. Brady, D. Flynn, F. Tangy, A. Israël, P.O. Vidalain, R. Mahieux, and R. Wei. NRP/Optineurin Cooperates with TAX1BP1 to potentiate the activation of NF-kappaB by human T-lymphotropic virus type 1 tax protein. PLoS Pathogens (2009); 5.

[114] C. Collet, L. Michou, M. Audran, S. Chasseigneaux, P. Hilliquin, T. Bardin, J-L. Laplanche. Paget's Disease of Bone in the French Population: Novel SQSTM1 Mutations, Functional Analysis, and Genotype-Phenotype Correlations. Journal of Bone and Mineral Research (2006), 22(2), 310–317.

[115] I. A. L. Silva, N. Conceição, E. Gagnon, J. P. Brown, M. L. Cancela and L. Michou. Molecular effect of an OPTN common variant associated to Paget's disease of bone. PLOS One (2018), 13(5).

[116] K. Kurisaki, A. Kurisaki, U. Valcourt, A.A Terentiev, K. Pardali, P. Ten Dijke, C.H. Heldin, J. Ericsson, A. Moustakas. Nuclear factor YY1 inhibits transforming growth factor beta- and bone morphogenetic protein-induced cell differentiation. Mol Cell Biol. (2003);23(13):4494-510.

[117] E. F. Wagner. Functions of AP1 (Fos/Jun) in bone development. Ann Rheum Dis (2002);61(Suppl II):ii40–ii42

[118] C. Scheel1, K. Schaefer, A. Jauch, M. Keller, D. Wai1, C. Brinkschmidt, F. van Valen, W. Boecker, B. Dockhorn-Dworniczakand and C. Poremba. Alternative lengthening of telomeres is associated with chromosomal instability in osteosarcomas. Oncogene (2001); 20:3835- 3844.

[119] T. Ozakia, T. Neumann, D. Wai, K-L. Schäferd, F. Valen, N. Lindner, C. Scheel, W. Böcker, W. Winkelmann, B. Dockhorn-Dworniczak, J. Horst, C. Poremba. Chromosomal alterations in osteosarcoma cell lines revealed by comparative genomic hybridization and multicolor karyotyping. Cancer Genetics and Cytogenetics (2003) Vol. 140(2): 145-152.

[120] M. Schnoor, I. Buers, A. Sietmann, M.F. Brodde, O. Hofnagel, H. Robenek, and S. Lorkowski. Efficient non-viral transfectionf THP-1 cells.Journal of Immunological Methods (2009), 31(344):109-115.

A1. Appendix I – Primer´s List

TABLE A1.1 – LIST OF PRIMERS

Name	5'- Sequence –-3'	Size (bp)
Analysis of ZNF687 expression in SaOS-2 and THP-1 cell lines		
ZNF687_Fwd	GAGCATGGCAAGTCAGTGAA	20
ZNF687_Rev	GTTTCTCTAGGATCAGGCGG	20
Generation of site-directed mutagenesis		
HsZNF687mFwd	CTCCCCTGAGCGCCCCCGTCCAG	23
HsZNF687mRev	CTGGACGGGGGCGCTCAGGGGAG	23
gRNA 1		
ZNF_687_Cas9_F1	CACCGGCTTCATTCGCATCAATGTC	25
ZNF_687_Cas9_R1	AAACGACATTGATGCGAATGAAGCC	25
gRNA 2A		
ZNF_687_Cas9n_A_F1	CACCGGCAAGGAGGTCATCAAAATC	25
ZNF_687_Cas9n_A_R1	AAACGATTTTGATGACCTCCTTGCC	25
gRNA 2B		
ZNF_687_Cas9n_B_F1	CACCGGCGAATGAAGCCATCCATTC	25
ZNF_687_Cas9n_B_R1	AAACGAATGGATGGCTTCATTCGCC	25
Screening CRISPR-Cas9/Casn9 (pX459+gRNA1 and pX462+gRNA2A/2B) construction		
U6_Fwd	GAGGGCCTATTTCCCATGATTCC	23
Screening the indel mutation in KO clones		
ZNF687mFwd1	CCTCCTCGTTCCTGTTTTCA	20
ZNF687mRev1	GGAAGCCAGAACCAGGTGTA	20
RT-qPCR for KO, overexpression and mineralization assays		
Hsa_OSX CDS_01R	GGTAAAGAGAGTGATTGGCAAGCAGTGG	28
Hs_Osx_01F	GAAGCGACCACCTGAGCAAACACCAG	26
Hs_TWIST1_qPCRF1	CCTGCTAGTGGGACGCGGACAT	22
Hs_TWIST1_qPCRR2	CCTGCTAGTGGGACGCGGACAT	22
HsaFGF2FW1	CAAAAACGGGGGCTTCTTCCTG	22
HsaFGF2RV1	CCATCTTCCTTCATAGCCAGGTAACG	26
Hsa_RANK_qPCRF2	CCAGCCCCAGCCCCAACT	18
Hsa_RANK_qPCRR2	CTCCCATCAGCCCCATCCTC	20
Hsa_SQSTM1_qPCRF1	CTCTCCCCAACGTTCTTCAGGA	22
Hsa_SQSTM1_qPCRR1	CTCTCCCCAACGTTCTTCAGGA	22
Hsa_CCDC3_qPCRR1	ACAGCCAGTCTGGGGAGATGAC	22
Hsa_CCDC3_qPCRR1	GCTTCCATGGTTCTCCAAGCAC	22
HsRunx2Rev1	GCGGGACACCTACTCTCATACTGGG	25
HsRunx2Fw1	GGAGTGGACGAGGCAAGAGTTTCACC	26
OPTN_FW2	TATTCCGATTCATTCCTGCCCCAAG	25
OTNN_REV2	GTCTCAAACCCTGACCCCAAGTGA	24
Hs_OSTEOCAL_qPCRF1	CGGTGCAGAGTCCAGCAAAGGT	22
Hs_OSTEOCAL_qPCRR1	AGCCGATGTGGTCAGCCAACTC	22
Hs_BMI1_qPCRF1	TGCTGATGCTGCCAATGGCTCT	22
Hs_BMI1_qPCRR1	CAGTCATTGCTGCTGGGCATCG	22
Hs_OCT4_qPCRR1	CTCGTGCAGGCCCGAAAGAGAA	22

Hs_OCT4_qPCRR1	CTGGCGCCGGTTACAGAACCAC	24
hGAPDHForwRT1	TCAACGGATTGGTCGTATTGGGCG	24
hGAPDHRevRT1	CTCGCTCCTGGAAGATGGTGATGGG	25
HsaACTINβ_FW	CATGTACGTTGCTATCCAGGC	21
HsaACTINβ_Rev	CTCCTTAATGTCACGCACGAT	21
Amplification of the *Homo sapiens* promoter region fragment 1		
HsZNF687_F1_Fwd	AGGCAGGAGAATGGCGTGAA	20
HsZNF687_ F1_Rev	GGAGCATGGAAGGAATCGGG	20
Amplification of the *Homo sapiens* promoter region fragment 2		
HsZNF687_F2_Fwd	ATTCCTTCCATGCTCCAAATCC	22
HsZNF687_F2_Rev	CACGCTTACTTGTTCCGCTC	20
Amplification of the *Homo sapiens* promoter region fragment 3		
HsZNF687_F1_Fwd	AGGCAGGAGAATGGCGTGAA	20
HsZNF687_F2_Rev	CACGCTTACTTGTTCCGCTC	20
Amplification of the *Homo sapiens* promoter region fragment 4		
HsZNF687_F4_Fwd_2	GCGTCCCATTCTGTAACTTCTG	22
HsZNF687_F4_Rev_2	GAAACTACAAGGGAAACTGAGG	22
Amplification of the *Danio rerio* promoter region fragment A		
Zfznf687a For1	ATTAAAGCTCCTTTTTAAATGTA	24
Zfznf687a Rev1	GCCAAATGTTCTCGTCTACTCTC	23
Amplification of the *Danio rerio* promoter region fragment B		
Zfznf687bFW1	CAGTGTTGGGAATGAGATTGTATGG	25
Zfznf687bR1	GGTTGTTGTGTATTTATGTCCTTCC	25
Drznf687b_Fwd2	CTCAGCTACTCATGTTTATGTTCAG	25
Drznf687b_Rev2	GCATCTAAATCTCCTGTGGTATGTC	25
Screening cloning reaction in TOPO vector		
M13F	GTAAAACGACGGCCAG	14
Screening cloning reaction in pGL3 vector		
R240	CTTTATGTTTTTGGCGTCTTCCAT	24

A2. Appendix II – Solutions and reagents

TABLE A2.1 LIST OF SOLUTIONS AND REAGENTS

	Composition
SOC	0.02 M Glucose 0.02 M $MgCl_2$ 0.02 M MgSo4 in LB (0.015 $g.ml^{-1}$)
SOLUTION P1	50 mM glucose 25 mM Tris.Cl (pH 8.0) 10 mM EDTA (pH 8.0)
SOLUTION P2	0.2 N NaOH (freshly diluted from a 10 N stock) 1% SDS
SOLUTION P3	5 M potassium acetate glacial acetic acid ddH_2O
AGAROSE GEL	Agarose powder (Sigma) 1x TAE (0.04 M Tris-acetate and 0.001 M EDTA)
IPTG	1,2 g of IPTG powder (Sigma) ddH_2O to a final volume of 50 ml
LOADING DYE 6x	0.25% bromophenol blue 0.25% xylene cyanol FF 30% glycerol in water
LB bROTH	10 g/L tripton 5 g/L yeast extract 5 g/L Sodium chloride
X-Gal	40 mg X-Gal 2 ml N,N'-demiethylformamide
PBS 1x (pH 7.4)	137 mM NaCl 2.7 mM KCl 8.1 mM Na_2HPO_4 1.47 mM KH_2PO_4
TRYPSIN 0.2%	137 mM NaCl 2.7 mM KCl 8.1 mM Na_2HPO_4 1.47 mM KH_2PO_4 1.1 mM EDTA 0.2% (vol/vol) Trypsin (Gibco)
CST BUFFER	H_2O_2 1 M Tris pH 7.5 5 M NaCl 0.5 M EDTA 200 mM Pyro Pho 1 M b.g.p 100 mM ovo
TBT-T 1x	20 mM Tris pH 7.5

	50 mM NaCl 0.1% Tween-20
ECL	250 mM Luminol 90 mM p. Coumaric acid 1M Tris pH8.5 30 % H_2O_2 H_2Odd
SDS-Page gel	1.5 M TRIS pH 8.8 Bis-acrylamide 10 % SDS 10 % APS TEMED
4% PFA (V/V)	36.5 % formaldehyde solution PBS pH 7.4
5 % $AGNo_3$ (W/V)	1 g Silver nitrate (Sigma 209139) 20 ml milliQ water
2.5 % $NA_2S_2o_3$ (W/V)	0.5 g Sodium thiosulfate (Sigma S7143) 20 ml milliQ water
AR-s 40 MM (pH 4.2)	274 g alizarin red (Sigma A5533) 19 ml milliQ water
10 % CPC	10 g cetylpyridinium 100 ml sodium phosphate 10 mM pH 7.0
BOUIN'S FLUID	15 ml Saturated picric acid solution 5 ml Formaldheyde 35 % (w/v)
SIRIUS RED STAINING	10 mg Sirius dye 10 ml Saturated picric acid solution
GLYCINE BUFFER 0.1 M (pH 10.4)	0.1 M Glycine 1 mM MgCl2 1 mM ZnCl2 pH 10.4

A3. Appendix III – Additional protocols

A3.1. Competent *E.coli* DH5α & stbl4 Cells Protocol

5 µl of *E.coli* DH5α (or Stbl4) glycerol stock is inoculated into 5 ml of LB medium without antibiotic, and final solution was incubated at 37ºC overnight under constant agitation (200 rpm). The next day, 1 ml of the overnight grown bacterial suspension is inoculated into 100 ml of LB medium without antibiotic, and then incubated at 37ºC under constant shaking (200 rpm) for approximatively 2 hours and a half until optical density (OD) reach 0.4 at 600 nm. After, the bacterial suspension is immediately incubated on ice for 10 minutes and then centrifuged at 4000 ×g for 10 minutes at 4ºC. The cell pellet is gently resuspended with 50 ml of ice cold 100 mM calcium chloride solution and is incubated on ice for 20 minutes. After incubation, bacterial cells are centrifuged at 4000 x g for another 10 minutes at 4ºC. Then, the supernatant was discarded gently with a pipet without disturbing the cell pellet. The latter was carefully resuspended with 10 ml of ice cold 100 mM calcium chloride + 10 % glycerol solution. 100 µl of bacterial suspension were aliquot in sterile 1.5 ml Eppendorf tubes and stored at -80ºC.

Afterwards, we performed a sub-cloning efficiency assay to test the transformation efficiency. For this we add 1 ng of plasmid DNA into 100 µl of bacterial aliquot obtained before, and the mix was incubated on ice for 30 minutes. Then a heat-shock was performed for 1 minute at 42ºC, and cells were once again incubated on ice for 2 minutes. 250 µl of SOC were added and cell suspension was incubated at 37ºC for 1 hour under constant agitation (200 rpm). 10 µl and 100 µl of bacterial suspension was plated into LB plates with the appropriate antibiotic (ampicillin). Each plate was incubated at 37ºC overnight. The next day transformation efficiency (TE) was calculated for each plate following the next formula:

$$TE = \frac{number\ colonies / \mu g\ plasmid}{dilution\ factor}$$

To be considered competent cells, TE has to be superior or equal to 1×10^6.

A.3.2 – Quantification of protein

First, the standard samples were prepared from BSA protein (2mg.ml^{-1}), by preforming serial dilutions. In parallel, ZNF687 protein extracts were individually

diluted, at 1:5, with nuclease free water. Each of the six standard samples (2000 $\mu g.ml^{-1}$, 1000 $\mu g.ml^{-1}$, 800 $\mu g.ml^{-1}$, 600 $\mu g.ml^{-1}$, 300 $\mu g.ml^{-1}$, and 150 $\mu g.ml^{-1}$), together with the blank sample (0 $\mu g.ml^{-1}$) and ZNF687 protein extracts, were plated into separate wells in a 96-well microplate. ZNF687 protein extracts were plated in duplicate. Then, 250 µl of Bradford 1x dye reagent was added to each well. After agitation in the microplate reader, absorbance of each sample was measure at 595 nm.

A4. Appendix IV – Additional Figures

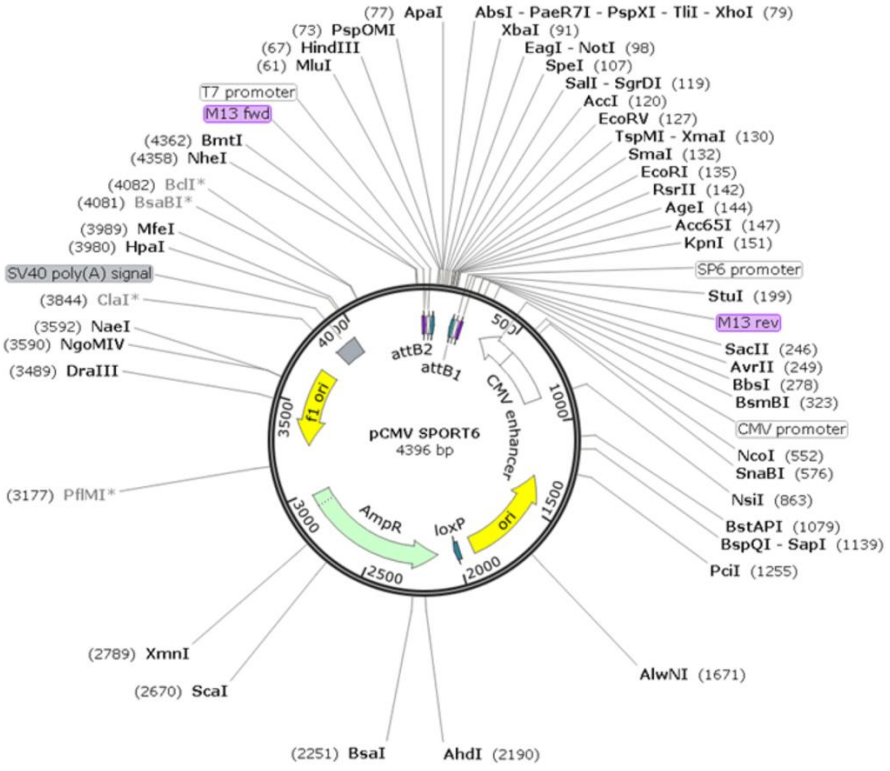


Figure A4.1 – pCMVSPORT6 vector map. Created with SnapGene (Source: Fetch my Gene®)

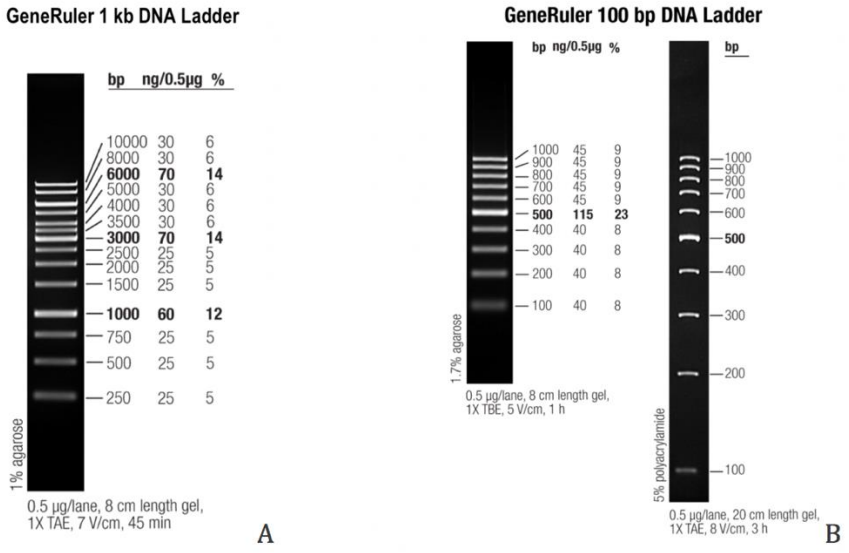


Figure A4.2 – DNA ladders used in electrophoresis. A - 1kb DNA ladder; B - 100 bp DNA ladder

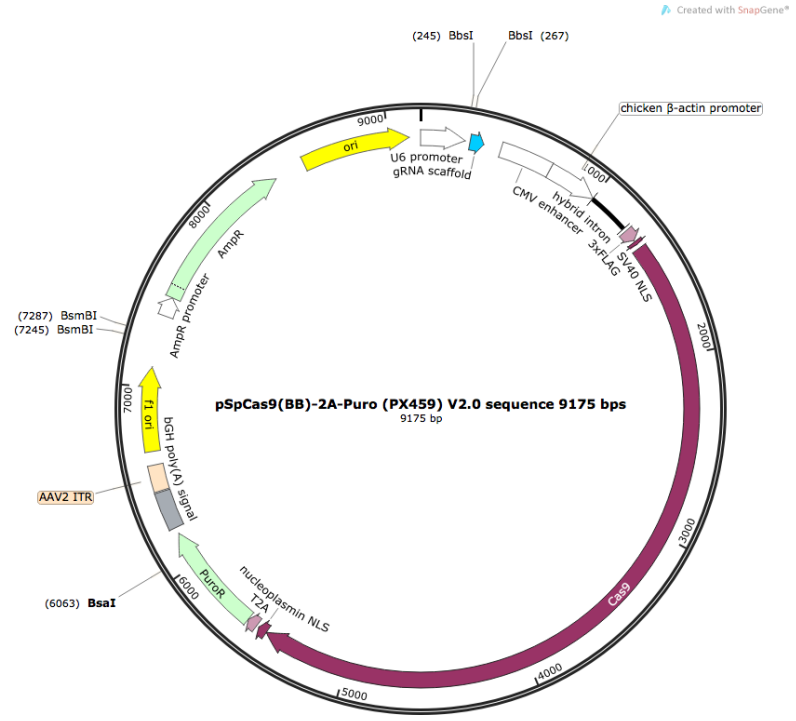


Figure A4.3 – pX459 vector map. (Addgene)

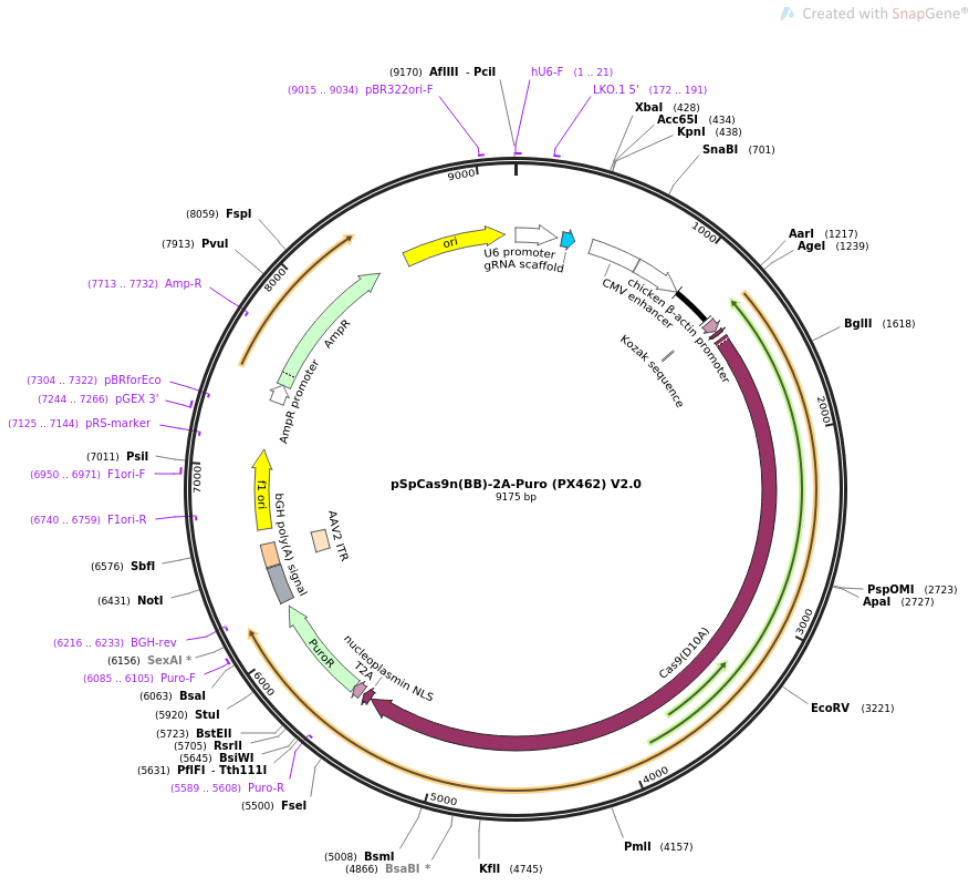


Figure A4.4 – pX462 vector map. (Addgene)

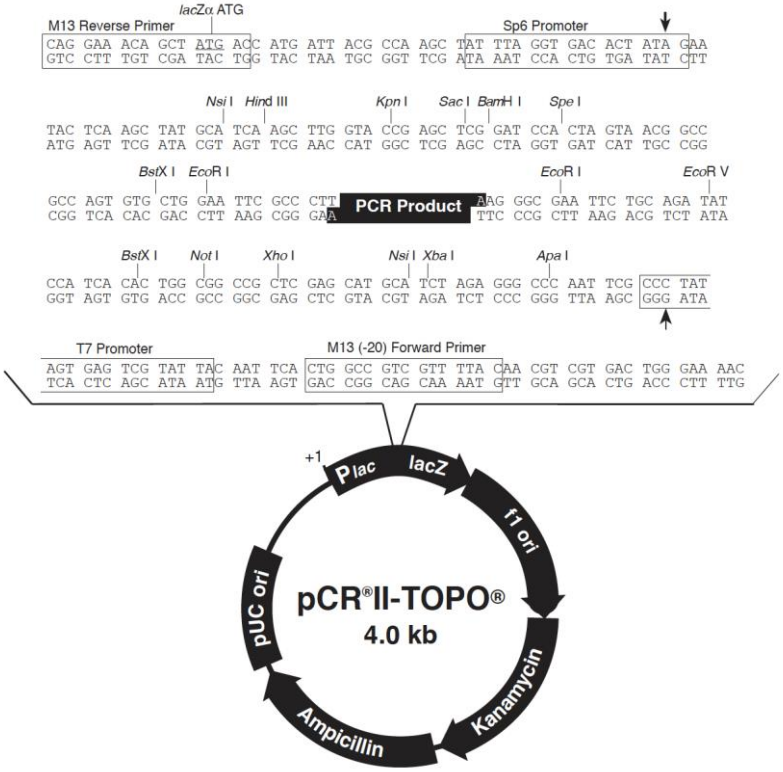


Figure A4.5 – Map of the pCRII-TOPO vector (Invitrogen)

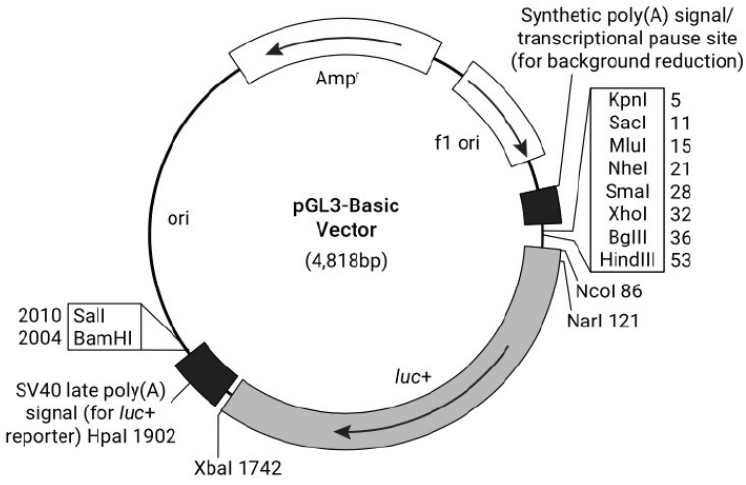


Figure A4.6 – pGL3-Basic vector (Invitrogen)

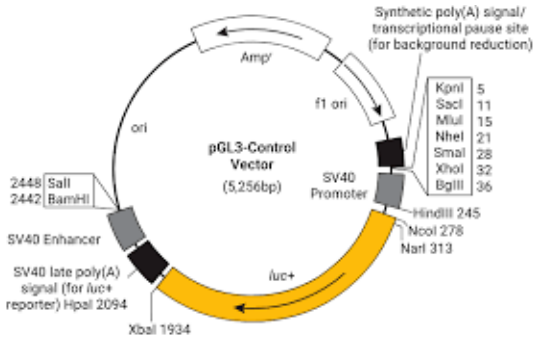
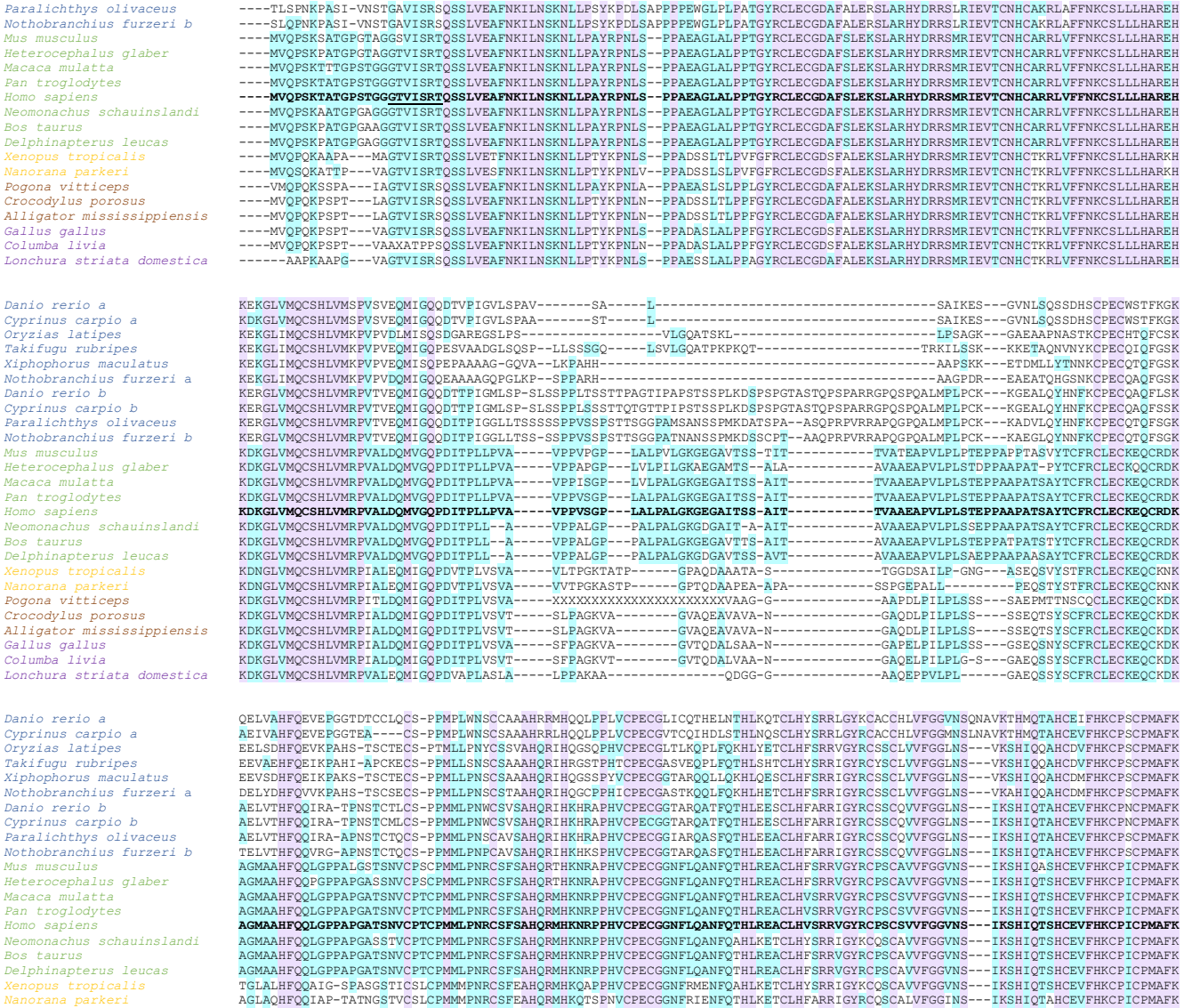


FIGURE A4.7 PGL3-CONTROL VECTOR (PROMEGA)


Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica


Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus


Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus_carpio_a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b


Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri


Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes


Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi
Bos taurus
Delphinapterus leucas
Xenopus tropicalis
Nanorana parkeri
Pogona vitticeps
Crocodylus porosus
Alligator mississippiensis
Gallus gallus
Columba livia
Lonchura striata domestica
Danio rerio a
Cyprinus carpio a
Oryzias latipes
Takifugu rubripes
Xiphophorus maculatus
Nothobranchius furzeri a
Danio rerio b
Cyprinus carpio b
Paralichthys olivaceus
Nothobranchius furzeri b
Mus musculus
Heterocephalus glaber
Macaca mulatta
Pan troglodytes
Homo sapiens
Neomonachus schauinslandi

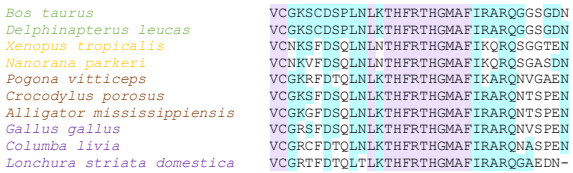


Figure A4.8 – Pairwise alignment of ZNF687 of several species. Blue clor represent amino acid that are identical to human. Pink color corresponds to perfect identity between all species. Green represent the NLS of human and yellow color illustrate the zebrafish NLS.

A5. Appendix V – Additional Tables

TABLE A5.1 – ACCESSION NUMBERS OF ZNF687 SEQUENCES (PROTEIN, GENOMIC AND TRANSCRIPT VARIANTS) OF SEVERAL SPECIES. * REFERS TO GENE *znf687a* AND ** REFERS TO *znf687b* FOR SPECIES WHICH GENOME SUFFERED DUPLICATION DURING EVOLUTION.

Type	Specie	Accession number
Protein	*Homo sapiens*	NP_065883.1
	Pan troglodytes	*JAA1593.1*
	Macaca mulatta	*NP_00124754.1*
	Bos taurus	*NP_001192386.1*
	Mus musculus	*NP_084350.1*
	Heterocephalus glaber	*JAO01528.1*
	Neomonachus schauislandi	*XP_021543693.1*
	Delphinapterus leucas	*XP_022419851.1*
	Xenopus tropicalis	*AAI27288.1*
	Nanorana parkeri	*XP_018421909.1*
	Crocodylus porosus	*XP_019412473.1*
	Pogona vitticeps	*XP_020653905.*
	Gallus gallus	*XP_427387.4*
	Columba livia	*XP_021136056.1*
	Lonchura striata domestica	*OWK51936.1*
	Danio rerio	*NP_001340793.1**
		*NP_001334602.1***
	Cyprinus carpio	*KTF83626.1**
		*KTG44759.1***
	Oryzas latipes	*XP_011483574.1*
	Takifugu rubripes	*XP_011604094.1*
	Xiphophorus maculatus	*XP_005809200.1*
	Nothobranchius furzeri	*SBP49295.1**
		*SBS53433.1***
	Paralychthys olivaceus	*XP_019957510.1*
Genomic	*Homo sapiens*	*NG_051575.1*
	Mus musculus	*ENSMUSG00000019338*

	Bos taurus	*NC_037330.1*
	Gallus gallus	*ENSGALG00000040270*
	Xenopus tropicalis	*ENSXETG00000010001*
	Oryzias latipes	*NC_019874.2*
	Takifugu rubripes	*NW_004070989.1*
	Danio rerio	*ENSDARG00000102774*
		ENSDARG00000019299
Transcript variants	*Homo sapiens*	*NM_001304763.1*
		NM_020832.2
		NM_001304764.1
		XM_005245366.3
		XM_0011509812.2
		XM_0011509813.2
		XM_0011509811.2
		CR749307.1
	Danio rerio	*ENSDARG00000167919.1**
		*ENSDARG00000170604.1**
		*ENSDARG000000104703.3***
		*ENSDARG000000132336.2***

Table A5.2 – List of proteins that interact with ZNF687 and their respective function. Ref

Name	Function
GATAD2A	Involved in DNA methylation and negative regulation of transcription.
TSPYL2	Chromatin remodeling and inhibitor of cell-cycle progression.
ZMYND8	Transcriptional regulator and chromatin remodeling.
INTS1	Component of the Integrator (INT) complex. Integrator cleaves small nuclear RNA (snRNAs) as part of their processing to their mature form in a mechanism that is intimately coupled with transcription termination.

INTS3	Component of the Integrator (INT) complex; component of the SOSS complex, a multiprotein complex that functions downstream of the MRN complex to promote DNA repair and G2/M checkpoint.
ZNF592	This gene is thought to play a role in a complex developmental pathway and the regulation of genes involved in cerebellar development. May be involved in transcriptional regulation.
INTS5	Component of the Integrator (INT) complex; may have a tumor suppressor role.
INTS6	Component of the Integrator (INT) complex
WDR26	Part of the WD repeat protein family that may facilitate formation of heterotrimeric or multiprotein complexes. Members of this family are involved in a variety of cellular processes, including cell cycle progression, signal transduction, apoptosis, and gene regulation.
UBE2O	Mediates monoubiquitination of target proteins; also able to monoubiquitinate the NLS of other chromatin-associated proteins, affecting their subcellular location; negatively regulates TRAF6-mediated NF-kappa-B activation.

TABLE A5.2 – LIST OF GENES USED IN THE GENE EXPRESSION ANALYSIS, WITH THEIR RESPECTIVE FUNCTIONS.

Name	Function
SQSTM1	Encodes a multifunctional protein that binds ubiquitin. Plays an important role in bone remodeling and regulates activation of the nuclear factor kappa-B (NF-kB) signaling pathway.
OSTEOCALCIN	Encodes a highly abundant bone protein secreted by osteoblasts. Involved in bone remodeling and energy metabolism.
BMI 1	Encodes a ring finger protein that is a major component of the polycomb group complex 1 (PRC1). Involved in chromatin

	remodeling, and functions as an epigenetic repressor of multiple regulatory genes involved in embryonic development and self-renewal in somatic stem cells. Also plays a central role in DNA damage repair.
RUNX2	Member of the RUNX family of transcription factors and encodes a nuclear protein. Involved in osteoblastic differentiation and skeletal morphogenesis. Also serves as a scaffold for nucleic acids and regulatory factors involved in skeletal gene expression.
CCDC3	Encodes a coiled-coil domain protein. Negatively regulates TNF-alpha-induced pro-inflammatory response in endothelial cells through inhibition of TNF-alpha-induced NF-kappaB activation. May be involved in bone metabolism.
RANK	Encodes a protein member of the TNF-receptor superfamily. Induces activation of NF-κB. Involved in osteoclast differentiation and activation.
FGF2	Encodes a protein member of the fibroblast growth factor (FGF). Plays an important role in the regulation of cell survival, cell division, cell differentiation and cell migration. Also implicated in tumor growth. Plays an important role in bone homeostasis.
OPTN	Encodes a multifunctional protein involved in several biological process, namely NF-κB regulation, autophagy, membrane trafficking, exocytosis, vesicle transport, transcriptional activation.
OCT4	Encodes a transcription factor containing a POU homeodomain that plays a key role in embryonic development and stem cell pluripotency.
OSX	Encodes a bone specific transcription factor required for osteoblast differentiation and bone formation.
TWIST1	Encodes a basic helix-loop-helix (bHLH) transcription factor that plays an important role in embryonic development. Also involved in osteoblast differentiation.

TABLE A6.3 – FUNCTION OF TRANSCRIPTION FACTORS

ETS-1	Transcription factor that can function either as transcriptional activator or repressor of numerous genes. Involved in stem cell development, cell senescence and death, and may control the differentiation, survival and proliferation of lymphoid cells.
SP1	A zinc finger transcription factor that binds to GC-rich motifs of many promoters. Involved in many cellular processes, such as cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling. Regulates the osteoblastic activity.
NFKB	NFKB is a transcription regulator that is activated by various intra- and extra-cellular stimuli. Once activated, it translocates into the nucleus and stimulates the expression of genes involved in a wide range of biological functions. Involved in osteoclastogenesis.
JUN (=AP1)	Transcription factor that interacts directly with specific target DNA sequences to regulate gene expression. Involved in bone metabolism and controls osteoblastic differentiation.
SRF	Ubiquitous nuclear protein that participates in cell cycle regulation, apoptosis, cell growth, and cell differentiation.
C-EBPA	Transcription factor that modulates proliferation arrest and the differentiation. Essential for the transition from common myeloid progenitors (CMP) to granulocyte/monocyte progenitors (GMP). Involved in osteoclastogenesis.
STAT-1	When phosphorylated, it translocates into the nucleus where it acts as a transcription factor. There, it coordinates the expression of a variety of genes, important for cell viability in response to external stimuli.
EGR1	C2H2 zinc finger protein that functions as a transcriptional regulator. Plays a role in the regulation of cell survival, proliferation and cell death.
STAT3	This protein mediates the expression of a variety of genes in response to cell stimuli, and thus plays a key role in many cellular processes. Osteoclastogenesis suppressor.

A6. Abstracts of panel communications

1. X Encontro Nacional de Biologia Evolutiva – Faro 2017

Title: Comparative analysis of the zinc finger protein ZNF687 throughout evolution.

Authors: Tatiana Authier, Natércia Conceição, e M Leonor Cancela.

Abstract:

ZNF687 is a zinc finger protein containing various Cys2-His2 (C2H2) zinc finger domains, which are one of the most common DNA-binding motifs found in eukaryotic transcription factors. The Homo sapiens ZNF687 has been related to several diseases such as acute myeloid leukemia, severe Paget's disease of bone associated with giant cell tumor of bone, and hepatocellular carcinoma, advocating a putative oncogenic role. Moreover, it has been demonstrated that *ZNF687* mRNA expression was significantly increased during osteoclastogenesis and osteoblastogenesis in both human and zebrafish, suggesting that this gene plays a role in bone metabolism in both species. This indicates that the function of this protein might be conserved by evolution despite speciation.

Therefore, ZNF687 sequences from various vertebrates were retrieved from databases and the protein homology was analysed through sequences alignment. These results showed that mammalian sequences presented high sequence identity (86,3-99,8%) and similarity (87,7-99,8%). Lower sequence identity was observed between Homo sapiens and non-mammalian sequences, with the majority of values around 50%, and the lowest percentages observed with teleost species (36,5%). Nonetheless, there is a region that appears to be more conserved between all selected species, with a percentage of identity reaching 59,0% between human and fish and 96,5% between human and other mammalian sequences. Among the 14 C2H2 zinc finger domains found in the human ZNF687, nine are localized in this conserved region. In parallel, since ZNF687 is a nuclear protein, we have identified putative nuclear localization signal (NLS) motifs and our results showed that among the three

predicted NLS encountered in the human ZNF687, only one seems to be partially conserved among the different species, with identities ranging from 45 to 100%.

We also performed comparative in silico analysis that showed not only the same *ZNF687* gene structure for fish and mammalian orthologs, but also the presence, within their promoter regions, of similar putative binding sites for common transcriptional factors known to affect bone metabolism. The synteny of *ZNF687* was also determined and results indicate that the cluster of surrounding genes was mostly conserved among mammals.

Comparative analysis of the zinc finger protein ZNF687 throughout evolution

T. Authier[1]; N. *Conceição*[2,3]; M.L. Cancela[2,3]

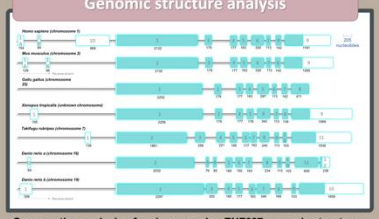
[1]*Master in Molecular and Microbial Biology, University of Algarve, Faro, Portugal*
[2]*Centre of Marine Sciences (CCMAR), University of Algarve, Faro, Portugal*
[3]*Dept of Biomedical Sciences and Medicine, University of Algarve, Faro, Portugal*

Overview

ZNF687 is a zinc finger protein containing various Cys2-His2 (C_2H_2) zinc finger domains, which are one of the most common DNA-binding motifs found in eukaryotic transcription factors. Although very few is known about ZNF687, together with ZMYND8 and ZNF592, ZNF687 has recently been described as part of the transcriptional regulator complex Z3.[1] The human *ZNF687* has been related to several diseases such as acute myeloid leukaemia, severe Paget's disease of bone associated with giant cell tumour of bone, and hepatocellular carcinoma, advocating a putative oncogenic role.[2-4] Moreover, it has been demonstrated that *ZNF687* mRNA expression was significantly increased during osteoclastogenesis and osteoblastogenesis in both human and zebrafish, suggesting that this gene plays a role in bone metabolism in both species.[3] This indicates that the function of this protein might be conserved by evolution despite speciation. In this present work we proceeded to a comparative *in silico* analysis of *ZNF687* gene and protein from different species in order to understand its trait of evolution.

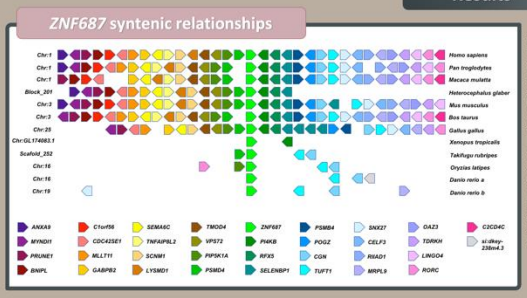
Results

Genomic structure analysis



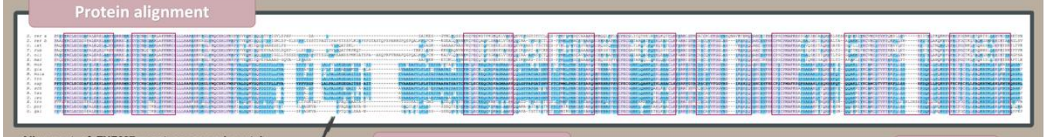
Comparative analysis of various species *ZNF687* genomic structures. All the *ZNF687* genes present similar coding sequence. Coding exons and non-coding exons are represented by full and white boxes, respectively. Numbers below the boxes indicate exons length. Only exons are on scale.

ZNF687 syntenic relationships



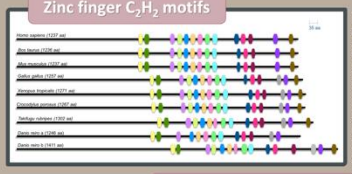
Schematic representation of genetic synteny neighbouring *ZNF687* gene in the chromosome of various species, as labeled. The cluster of surrounding genes of ZNF687 is mainly conserved in all species except in amphibian and teleost species.

Protein alignment



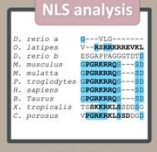
Alignment of ZNF687 most conserved protein region between several species. ZNF687 protein has an extensive conservation when comparing mammalian species. Residues identical in all species are filled in pink and mainly correspond to putative zinc finger domains (boxed). Blue highlight color represent residues that are identical to human ones (bold sequence).

Zinc finger C_2H_2 motifs



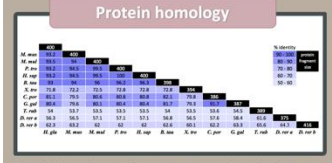
Comparative analysis of zinc finger C_2H_2 motifs among various species. All the human zinc finger domains are conserved throughout evolution. Nonetheless, one domain (grey box) persisted in every species except in mammalians. Each zinc finger domain is represented by a box with its specific color.

NLS analysis



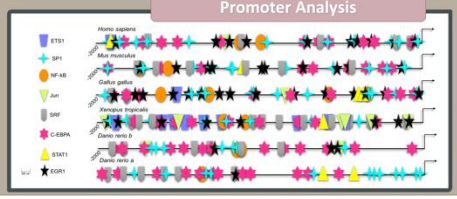
Nuclear localization signal (NLS) of ZNF687. Among the three putative NLS of human ZNF687 identified *in silico*, only one is conserved in all species but teleosts, except *O. latipes*. Bold sequences represent putative NLS.

Protein homology



Protein homology of the most conserved region. Pairwise alignments and identity percentage reveal that this ZNF687 protein region is relatively well conserved throughout evolution. Pairwise alignments were done using species ranging from fish to mammals and the identity percentage was acquired with Manipulation Suite facilities.

Promoter Analysis



Comparative analysis of 2 Kb promoter region between human and other species showing the presence of several binding sites for transcription factors (TFs). The search of TFs was based on the information available in the ENCODE project (UCSC Genome Browser).

Discussion

Despite similar genomic structure and coding sequence among various species, the comparative mapping of *ZNF687* genes and the analysis of syntetic regions and chromosomal locations that harbor *ZNF687* have demonstrated that the neighboring genes of *ZNF687* orthologues are only conserved between mammalian and avian species, and only few are preserved in amphibian and teleost species. Additionally, the presence, within their promoter regions, of similar putative binding sites for common transcriptional factors was assessed among different species, advocating a similar functional role of ZNF687. Overall protein homology (data not shown) is also mainly conserved among mammalian species (86,3-99,8%) and lower sequence identity was observed between *Homo sapiens* and non-mammalian sequences, with the lowest percentages observed with teleost species (36,5%). Nonetheless, there is a region that appears to be more conserved between all selected species, with a mean percentage of identity reaching 59,0% between human and fish and 96,5% between human when compared to other mammalians. Among the 14 C_2H_2 zinc finger domains found in the human ZNF687, nine are localized in this conserved region, suggesting that the zinc finger motifs are mainly conserved by evolution despite speciation. Indeed, the domain analysis performed confirmed that ZNF687 presented same pattern and almost same numbers of zinc finger domains. Finally, among the three predicted NLS encountered in the human ZNF687, only one seems to be conserved in all species but teleosts, except for one (*Oryzias latipes*).
Continuing to decipher the structure, regulation and biological role of *ZNF687* gene will increase our understanding of its functions and evolution.

References

(1) Malovannaya *et al*., Cell . (2011); 145(5): 787–799.
(2) Guyen et al., Genes, Chromosomes and Cancer (2006) 45:918–932.
(3) Gianfrancesco *et al*., The American Journal of Human Genetics 98 (2016):275–286 .
(4) Zhang et al., Oncogenesis (2017) 6, e363.

Acknowledgments

Centro de ciencias do mar (CCMAR); Susana Serrano, Biochemistry bachelor FCT, University of Algarve; Raquel Arroja, Biochemistry bachelor FCT, University of Algarve

2. Interdisciplinary Approaches in Fish Skeletal Biology - Tavira 2018

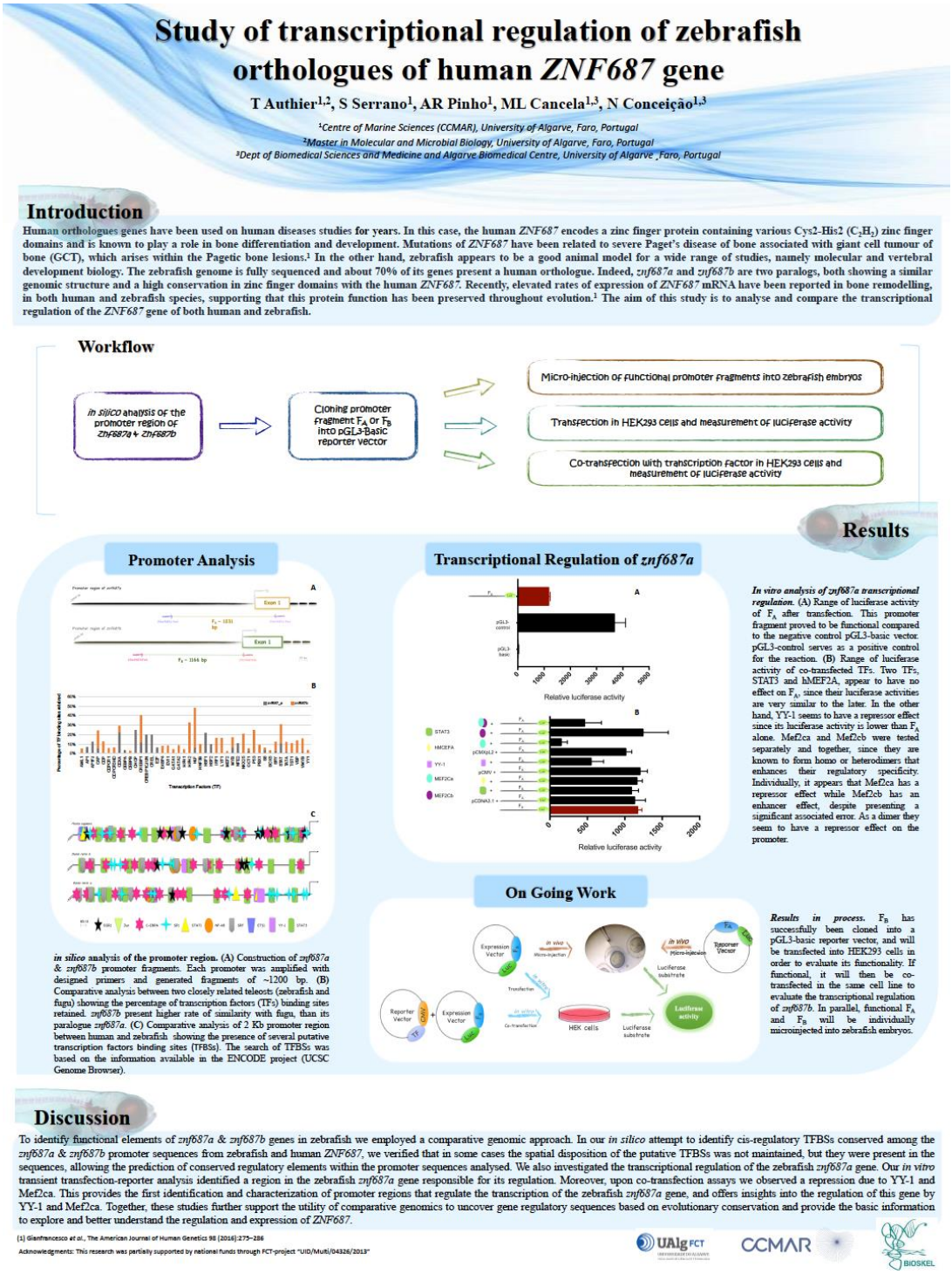
Title: Study of transcriptional regulation zinc finger protein ZNF687 gene in zebrafish

Authors: T. Authier, S. Serrano, A.R. Pinho, N. Conceição, e M.L. Cancela.

Abstract:

Zinc finger protein 687 (*ZNF687*) encodes for a zinc finger protein containing various Cys2-His2 (C_2H_2) zinc finger domains and known to play a role in bone differentiation and development. Mutations of *ZNF687* have been related to severe Paget's disease of bone associated with giant cell tumour of bone (GCT) which arises within the Paget bone lesions. Moreover, it has been demonstrated that *ZNF687* mRNA expression was significantly increased during osteoclastogenesis and osteoblastogenesis in both human and zebrafish, suggesting that the function of this protein might be conserved by evolution despite speciation. In this study we have identified putative molecular players affecting zebrafish *znf687a* and *znf687b* gene transcription by using a computational approach to search for cis-regulatory transcription factor binding sites (TFBSs) in the promoter regions of the *znf687a* and *znf687b* genes from zebrafish (*Danio rerio*) and fugu (*Takifugu rubripes*). We have used the DBA algorithm to obtain comparative alignment of promoter regions in order to detect meaningful conserved motifs, and then used MatInspector to determine putative TFBSs in those blocks. From the data obtained it was possible to identify the sites most likely involved in regulating expression of *znf687* in zebrafish. We also present the cloning of zebrafish *znf687a* and *znf687b* genes promoters, as well as the functional analysis of their promoters. Each promoter was capable of activating *znf687a* or *znf687b* transcription in transient transfections.

This work provides additional evidence of the usefulness of nonmammalian model systems to elucidate the regulation of *ZNF687* gene transcription.



3. XXXVIIIJornadas Portuguesas de Genética – 2018

Title : Analysis of ZNF687 gene expression in different tumors using the cancer genome atlas database

Authors: Helena Caiado, Ricardo Reis, Tatiana Authier, M. Leonor Cancela, Natércia Conceição

Abstract:

Zinc finger protein 687 (ZNF687), a transcription factor containing various Cys2-His2 (C2H2) zinc finger domains, has been found to be mutated and upregulated in different diseases such as Paget's disease of bone associated with giant cell tumor of bone and acute myeloid leukemia. Moreover, it has been recently reported that *ZNF687* was markedly upregulated in hepatocellular carcinoma (HCC) cell lines and tissues and was significantly correlated with relapse-free survival in those patients. Altogether, *ZNF687* seems to play an oncogenic role when upregulated, but the clinical significance and precise role of this gene in cancer progression remains unknown.

In this work we evaluated the expression levels of *ZNF687* in different specific cancers using available information in the cancer genome atlas (TCGA) data repository. Results demonstrated that *ZNF687* gene expression was significantly upregulated in primary tumor tissues from all the analyzed tumors comparing with the expression of the normal tissue ($P \leq 0.001$). Analysis of the methylation sites was also performed on *ZNF687* and data showed 17 CpG's methylation sites located along the gene, six in the promoter region, three in the 5'UTR region, one in the first non-coding exon, five in the first intron, one in the third exon and one in the eighth exon. Regarding the levels of methylation, CpG's cg17470103 (located in the distal promoter region) and cg02388783 (located in the proximal promoter region) presented an overall hypermethylation status in the normal tissue, while CpG's cg01435515, cg02290168 (located in the first intron) and cg23274217 (located in the third exon) presented an overall hypermethylation status in the primary tumor tissue.

Results from the bioinformatics analysis suggest that *ZNF687* might have an oncogenic role in cancer and thus may be a possible candidate as therapeutic target. However, further studies are required to better understand the inherent molecular mechanisms underlying *ZNF687* deregulation in cancer.

Analysis of *ZNF687* gene expression in different tumors using the cancer genome atlas database

Helena Caiado[1,2,4], Ricardo Reis[4], Tatiana Authier[1,3], M. Leonor Cancela[1,4,5], Natércia Conceição[1,4,5]

[1]Center of Marine Sciences (CCMAR). University of Algarve. Faro. Portugal
[2]ProRegeM PhD Programme in Regenerative Medicine. University of Algarve, Faro, Portugal
[3]Master in Molecular and Microbial Biology, University of Algarve, Faro, Portugal
[4]Department of Biomedical Sciences and Medicine University of Algarve, Faro, Portugal
[5]Algarve Biomedical Centre, University of Algarve, Faro, Portugal

Introduction

Zinc finger protein 687 (ZNF687), a transcription factor containing various Cys2-His2 (C2H2) zinc finger domains, has been found to be mutated and upregulated in different diseases such as Paget's disease of bone associated with giant cell tumor of bone[1,2] and acute myeloid leukemia[3]. Moreover, it has been recently reported that *ZNF687* was markedly upregulated in hepatocellular carcinoma cell lines and tissues and was significantly correlated with relapse-free survival in those patients[4]. Altogether, *ZNF687* seems to play an oncogenic role when upregulated, but the clinical significance and precise role of this gene in cancer progression remains unknown.
In this work we evaluated the expression levels of *ZNF687* in different specific cancers using available information in the cancer genome atlas (TCGA) data repository.

Workflow

Hypothesis

Is ZNF687 gene expression altered in different types of tumors by methylation sites?

Cancers selection from The Cancer Genome Atlas (TCGA) data repository

1. *ZNF687 gene expression analysis*
2. *Putative transcription factors gene expression analysis*
3. *Correlation of gene expression with methylation sites*

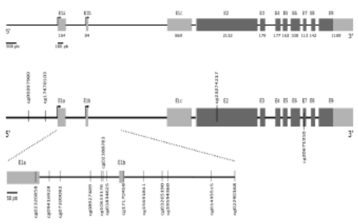
Results

ZNF687 gene expression analysis (1)

A) Schematic representation of *ZNF687* (ENSG00000143373) gene structure. *Both exons (boxes, 100bp=1cm) and introns (500bp=1cm) are at scale, and arrow indicates putative transcription starting sites. Dark grey area represents the coding region while the light grey area characterizes the non coding region.*

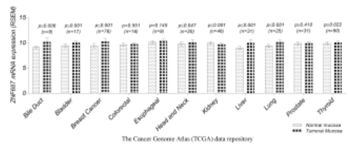
B) Location of the methylation sites of *ZNF687* gene, identified through the cancer genome atlas (TCGA) database.

Analysis of the methylation sites performed in ZNF687 showed 17 CpG's methylation sites located along the gene.



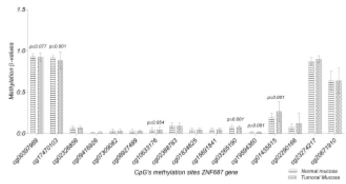
C) Analysis of *ZNF687* gene expression in normal and tumoral samples from TCGA database. *Gene expression analysis of ZNF687 in normal and tumoral mucosa, in 11 selected tumors. Mean±SD, Mann-Whitney statistic test, p value was considered statistically different when p≤0.05.*

Results demonstrated that ZNF687 gene expression was significantly upregulated in primary tumor tissues from all the analyzed tumors comparing with the expression of the normal tissue, except for colorectal, esophageal and prostate cancers.

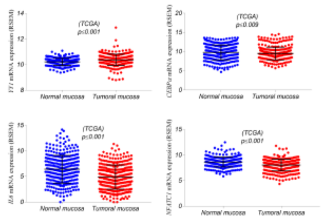


D) Analysis of *ZNF687* methylation levels in normal and tumoral samples from TCGA database. *Mean±SD, Mann-Whitney statistic test (n=331), p value was considered statistically different when p≤0.05.*

CpG's cg17470103 and cg02388783 presented an overall hypomethylation status in tumoral tissue, while CpG's cg01435515, cg02290168 and cg23274217 presented an overall hypermethylation status in the primary tumor tissue.



Putative transcription factors gene expression analysis (2)



Analysis of putative transcription factors (TFs) gene expression in normal and tumoral samples from TCGA. *Gene expression analysis of four TFs that might have a role in ZNF687 gene regulation on the previously selected 11 tumors from TCGA in normal and tumoral mucosa. Mean±SD, Mann-Whitney statistic test, p value was considered statistically different when (p≤0.05).*

Results demonstrated an up regulation of the transcription factors YY1 and CEBPα in the tumoral mucosa, while transcription factors IL6 and NFATC1 presented a down regulation in tumoral tissue, suggesting a possible regulation of ZNF687 gene expression through these TFs.

Correlation of gene expression with methylation sites (3)

To address the question if there is a correlation between ZNF687 gene expression and the identified transcription factors, we performed a Spearman's rank correlation coefficient.

ρ of Spearman		NFATC1	IL6	CEBPA	YY1
ZNF687	Correlation Coeficient	-0,267	-0,267	-0,312	0,194**
	Significance (bilateral)	,000	,000	,000	,000
	n	331	330	331	331

***The correlation is significant at level 0,01 (bilateral)*

Results demonstrated a positive correlation between ZNF687 gene expression and transcription factor YY1 as well as a negative correlation with NFATC1, IL6 and CEBPα. These correlations suggest that these TFs might be positive or negative regulators of ZNF687 gene expression depending on the methylation sites status.

Discussion

These analysis supported the usefulness to identify genes that might be deregulated in cancer and how this deregulation could be related with epigenetic events. However, further bioinformatics analyses of these data are required to apprehend if this upregulation could be related with an unfavorable prognosis of cancer.
Also, it is necessary to confirm by in vitro assays if the putative transcription factors have in fact a regulatory role in some molecular mechanisms underlying ZNF687 deregulation in cancer.

Acknowledgement

This research was partially supported by National Funds through Foundation for Science and Technology (FCT) under the project "CCMAR/Multi/04326/2013". HC is supported by a doctoral fellowship (PD/BD/128341/2017) from FCT.

References

[1]Divisato, G., Formicola, D., Esposito, T., Merlotti, D., Pazzaglia, L., Del Fattore, A., Siris, E., Orcel, P., Brown, J. P., Nuti, R., Strazzullo, P., Benassi, M. S., Cancela, M. L., Michou, L., Rendina, D., Gennari, L. and Gianfrancesco, F. (2016) 'ZNF687 Mutations in Severe Paget Disease of Bone Associated with Giant Cell Tumor', American Journal of Human Genetics. The Authors, 98(2), pp. 275–286. doi: 10.1016/j.ajhg.2015.12.016.

[2]Gianfrancesco, F., Rendina, D., Merlotti, D., Esposito, T., Amyere, M., Formicola, D., Muscariello, R., De Filippo, G., Strazzullo, P., Nuti, R., Vikkula, M. and Gennari, L. (2013) 'Giant cell tumor occurring in familial Paget's disease of bone: Report of clinical characteristics and linkage analysis of a large pedigree', Journal of Bone and Mineral Research, 28(2), pp. 341–350. doi: 10.1002/jbmr.1750.

[3]TuDung T. Nguyen, Lisa N. Ma, Marilyn L. Slovak, Charles D. Bangs, Athena M. Cherry, and D. A. A. (2006) 'Identification of Novel RUNX1 (AML1) Translocation Partner Genes SH3D19, YTHDF2, and ZNF687 in Acute Myeloid Leukemia', Genes, chromosomes & cancer, 45(July), pp. 918–932. doi: 10.1002/gcc.20355.

[4]Zhang, T., Huang, Y., Liu, W., Meng, W., Zhao, H., Yang, Q., Gu, S.-J., Xiao, C.-C., Jia, C.-C., Zhang, B., Zou, Y., Li, H.-P. and Fu, B.-S. (2017) 'Overexpression of zinc finger protein 687 enhances tumorigenic capability and promotes recurrence of hepatocellular carcinoma', Oncogenesis, 6(7), p. e363. doi: 10.1038/oncsis.2017.63.