MICHELE LOIODICE

TRANSCRIPTIONAL REGULATION OF THE OVERLAPPING HUMAN GENES *OPTN* AND *CCDC3*

UAlg
UNIVERSIDADE DO ALGARVE

Faculty of Science and Technology

2019

MICHELE LOIODICE

TRANSCRIPTIONAL REGULATION OF THE OVERLAPPING HUMAN GENES *OPTN* AND *CCDC3*

Master of Science in Molecular and Microbial Biology
Scientific supervisor Prof. Natércia Conceição

UAlg
UNIVERSIDADE DO ALGARVE
Faculty of Science and Technology

2019

Statement of work authorship

I declare to be the author of this work, which is original and unpublished. Authors and works consulted are properly cited in the text and are included in the reference list attached.

Michele Loiodice

Acknowledgment

I would like to sincerely express my gratitude to every single person who helped me in realizing and making possible this Master's thesis work and internship experience.

In first place, a heartfelt thank you to Professor Leonor Cancela for giving me the unique opportunity to join the BIOSKEL lab, and to significantly improve my work experience.

To professor Natércia, being my supervisor, for her endless patience and priceless support through the hurdles of science, including interrupted lunch breaks and weekend work; your guidance helped me grow a lot.

To every single member of the BIOSKEL lab, for making me feel welcome and part of the group from the very beginning.

To the sisters Débora and Tatiana, who were always available to guide me and solve my practical doubts; to Helena, my bench neighbour, for her precious help to overcome my problems.

To Marco and Alessio, the Italian team of the lab, and to Sunil, for the funny moments of light heartedness.

E infine, un enorme grazie alla mia famiglia, in particolare ai miei genitori. Non avete mai smesso di credere in me, e senza il vostro supporto niente di tutto questo sarebbe stato possibile.

Abstract

Human genes *OPTN* and *CCDC3*, encoding respectively for optineurin and coiled-coil domain-containing protein 3, are part of the *PDB6 locus*, a genetic hotspot strongly associated with Paget's disease of bone (PDB), the second most prevalent metabolic bone disease after osteoporosis. *OPTN* and *CCDC3* genes share a head-to-head configuration and partially overlapping sequences, located on opposite strands of this *locus*.

We first defined the molecular structure of the two genes based on the *in silico* identified mRNAs, which included several alternatively spliced transcripts. The task was performed with the aid of bioinformatics tools and online databases, such as expressed sequence tags (ESTs), AceView gene browser and Splign software; as a result, we obtained a comprehensive map of *OPTN* and *CCDC3*, emphasizing the size and position of introns and exons of each transcript.

Next, we assessed the activity of *CCDC3* and *OPTN* promoter regions; due to their head-to-head disposition, the two genes share a common regulatory sequence. A putative *CCDC3* alternative promoter, located downstream and exclusive for certain *CCDC3* transcripts, was identified by analysing the gene structure obtained *in silico*. The activity of the promoter regions was validated by transient transfecting pGL3 reporter constructs, containing the promoter sequences under analysis, into HEK 293 cells, followed by luciferase assays.

Trans-acting regulatory proteins, e.g. transcription factors (TFs), putatively involved in the regulation of the two genes, were identified *in silico* by analyzing the promoter sequences through bioinformatics software. The analysis revealed several putative TF binding sites, including for NF-κB, a TF known to play a role in the pathogenesis of PDB. Transient co-transfection of HEK 293 cells with pGL3 reporter constructs and transcription factor NF-κB expression vectors, followed by luciferase assays, have been performed in order to confirm their role as trans-regulators of the target promoters, and to unveil the presence of a possible co-regulation.

This study was funded by FCT through the project UID/Multi/04326/2019 (CCMAR).

Keywords: gene, expression, regulation, splicing, promoter, transcription factor, luciferase.

Abstract (in Portuguese language)

Os genes humanos *OPTN* e *CCDC3*, que codificam respectivamente para optineurina e coiled-coil domain-containing protein 3, fazem parte do locus *PDB6*, um hotspot genético fortemente associado à doença óssea de Paget (PDB), a segunda doença óssea metabólica mais prevalente após a osteoporose. Os genes *OPTN* e *CCDC3* compartilham uma configuração frente a frente e sequências parcialmente sobrepostas, localizadas em cadeias opostas desse locus.

Primeiro, definimos a estrutura molecular dos dois genes com os mRNAs identificados in silico, que incluíam vários transcritos alternadamente unidos. A tarefa foi realizada com o auxílio de ferramentas de bioinformática e bancos de dados on-line, como tags de sequência expressa (ESTs), navegador de genes AceView e software Splign; como resultado, obtivemos um mapa abrangente de *OPTN* e *CCDC3*, enfatizando o tamanho e a posição dos íntrons e exons de cada transcrição.

Em seguida, avaliamos a atividade das regiões promotoras de *CCDC3* e *OPTN*; devido à sua disposição frente a frente, os dois genes compartilham uma sequência reguladora comum. Um promotor alternativo de *CCDC3*, localizado a jusante e exclusivo para certos transcritos de *CCDC3*, foi identificado através da análise da estrutura genética obtida em silico. A atividade das regiões promotoras foi validada por construções repórteres de transfecção pGL3 transitórias, contendo as seqüências promotoras em análise, em células HEK 293, seguidas por ensaios de luciferase.

Proteínas reguladoras de ação trans, p. fatores de transcrição (TFs), potencialmente envolvidos na regulação dos dois genes, foram identificados in silico através da análise das seqüências promotoras através do software de bioinformática. A análise revelou vários locais de ligação a TF, incluindo NF-κB, um TF conhecido por desempenhar um papel na patogênese do PDB. A co-transfecção transitória de células HEK 293 com construções repórter pGL3 e vetores de expressão contendo o TF NF-κB, seguidos de ensaios de luciferase, foram realizados para confirmar seu papel como reguladores trans dos promotores alvo e para revelar a presença de um possível co-regulação.

Este estudo foi financiado pela FCT através do projeto UID/Multi/04326/2019 (CCMAR).

Palavras-chave: gene, expressão, regulação, união, promotor, TF, luciferase.

Table of Contents

List of figures and tables

I. Introduction

1.1 – The *PDB6 locus* and Paget's disease of bone

The genetic *locus* localized on the small arm of chromosome 10, corresponding to the cytoband number 13 (10p13), as shown in Figure 1.1, has been named as PDB6 due to its association with the Paget's disease of bone (PDB), though the single causal gene of the disease has not been identified yet [1]. The PDB6 *locus* contains over 70 genes (Figure 1.2), two of which have been selected to be analyzed: *OPTN*, coding for optineurin, and *CCDC3*, coding for coiled-coil domain containing 3 protein.

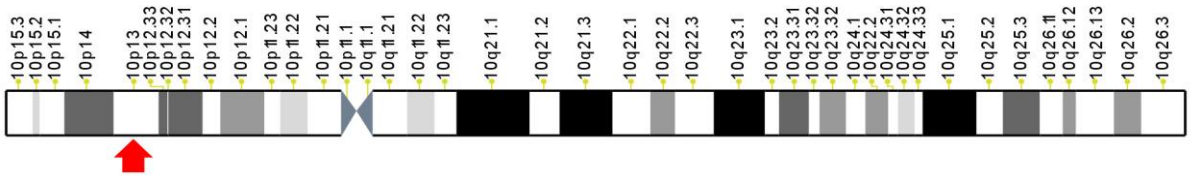

Fig. 1.1: Human chromosome 10 giemsa banding ideogram. Red arrow shows *PDB6 locus* genomic location (10p13). Adapted from Ensembl GRCh38.p10 ideogram, public domain.

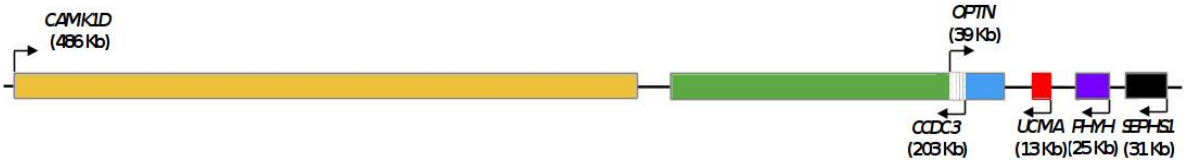

Fig. 1.2: Schematic representation of *PDB6 locus* genes. Arrows indicate the orientation of the genes. Size of each gene is indicated in brackets. Note the shared sequence between *OPTN* and *CCDC3* (in white). Adapted from Silva (2015).

The *PDB6 locus* is just one of the several chromosomal regions that was proven to be connected with PDB [1] [2]. Paget's disease of bone is the second most prevalent metabolic bone illness after osteoporosis [3]; the disease is characterised by an unbalanced and excessive bone turnover, during which the osteoblasts mineralization of the bone matrix and the osteoclasts bone erosion, a finely balanced modelling/remodeling activity in unaffected bone tissue, result instead increased and disregulated [3]. The process leads to the creation of abnormal bone tissue, with affected architectural properties and weakened structural strength of the organ. These elements result macroscopically in severe symptoms, which include bone deformities, increased tendency to bone fractures, chronic pain and impaired organs due to nerve compression [4].

Both environmental and genetic causes are able to influence the development of PDB [6] [7], which appears to be more common in people of Anglo Saxon origin [5]; overall, PDB has a prevalence between 3 and 3.7%, and its frequency intensifies with age, involving up to three percent of adults over 55 years of age [5].

Albagha et al. (2010) reported a strong genetic association between the 10p13 (*PDB6*) *locus* and PDB, specifically with three single nucleotide polymorphisms (SNPs) located in the *OPTN* gene [1]; Michou et al. (2012) also found genetic association of two different SNPs in *OPTN* with PDB, and identified a functional SNP located in the promoter of *UCMA*/*GRP* gene, which provided a weak genetic connection with the disease [9].

1.2 – The *OPTN* gene

The *OPTN* gene (size: 39 Kb), short for optineurin, takes its name from "optic neuropathy-inducing gene", as mutations of this gene were firstly found in patients with primary open-angle glaucoma, a cause of blindness [10]. The *OPTN* gene is almost ubiquitously expressed in human tissues [11].

OPTN gene structure consists in a variable number of non-coding exons at its 5'-UTR region, according to the transcript analysed, and 13 exons coding for a 66 kDa protein. Alternative splicing of the *OPTN* gene generates a large number of different transcripts (see Figure 1.3), which are translated into protein isoforms with the same open reading frame (ORF), giving rise to several protein variants [11].

By analysing the gene transcripts in Figure 1.3, it came to our attention that *OPTN* generates several mRNAs featuring an alternative first exon; these two first elements were named in the scheme as exon 1a (as alternative) and exon 1; many eukaryotic genes possess multiple transcriptional promoters with alternative first exons, located in the 5'-UTR region and far upstream of the coding exons [13]; even thought the protein primary structure might not be directly affected by these structural changes, 5'-UTR sequence can contain elements acting as molecular switch that regulate the translation of the mRNAs [14]; it would be interesting to understand to what extent these 5′ transcriptional decisions influence downstream alternative splicing events or mRNA translational regulation.

Fig. 1.3: *OPTN* alternative transcripts scheme. Comparison of alternatively spliced transcripts available in databases. Same sequences are in same colors. Red arrows indicate differences in exons sequences. Adapted from Michou et al. [9].

OPTN is a multifunctional cytoplasmic protein, characterized by a multi-domain structure. It contains coiled-coil motifs, a basic leucine-zipper motif (bZIP), a microtubule-associated protein 1 light chain LC3-interacting region (LIR), a ubiquitin-binding domain (UBAN), and a C-terminal zinc-finger domain. Such an articulated, multi-domain configuration gives the OPTN protein a complex role, being capable of interacting with many different proteins and resulting involved in several signaling pathways. An example as such is the role the OPTN protein plays in NF-κB signaling, or in several other complex cellular processes, such as selective autophagy or membrane vesicle trafficking [11] [17].

Due to the large number of crucial cellular activity associated to OPTN protein, mutations in *OPTN* gene are strongly connected with human degenerative diseases; besides the previously mentioned PDB and primary open-angle glaucoma, *OPTN* mutations were found in patients suffering amyotrophic lateral sclerosis (ALS), frontotemporal lobar dementia (FTD), and Crohn's inflammatory bowel disease (CD) [11] [12].

1.3 – The *CCDC3* gene

CCDC3 (size: 203 Kb), the second gene chosen for our transcriptional analysis, codes for the coiled-coil domain containing 3 protein.

As today, unlike *OPTN*, we do not possess much information on the functions of this gene. Azad et al. (2014) found a role for CCDC3 in tumor necrosis factor(TNF)-α-induced inflammatory response in endothelial cells, showing that knocking down the gene increases TNF-α-induced expression of VCAM-1 protein (vascular cell adhesion molecule-1), and that an induced overexpression of CCDC3 protein decreases the TNF-α-induced nuclear translocation of p50 and p65 and the transcriptional activity of NF-κB [15]. Moreover, Kobayashi et al. (2010) found that the homologous of *CCDC3* in mice positively regulates lipid accumulation in adipose cells [16].

Our choice of analysis fell on this gene due to its shared sequence with *OPTN (*see section 1.5), but also because of the low amount of structural information available: at the beginning of the project, only two different transcript variants of *CCDC3* were publicly present on the Ensembl database.

The mentioned transcripts represented a scarce and insufficient number of elements when compared with *OPTN* (see section 1.3) and the other adjacent genes, lying in the same *PDB6* locus (Figure 1.2); given the big variety of alternative mRNAs available for *OPTN* and the lack of *CCDC3* structural information, we decided to include this gene in our analysis.

CCDC3 gene does not seem to be directly connected with the Paget's disease of bones, but its peculiar shared gene structure with *OPTN* (see section 1.5) and its presence in the *PDB6* hotspot, make the *CCDC3* gene a good candidate for further investigation on its transcriptional variants and regulation.

1.4 - *OPTN* and *CCDC3* shared structure

The two genes *OPTN* and *CCDC3* have a head-to-head, divergent orientation and lie on different strands, designated as positive strand [+] for *OPTN* and negative [-] for *CCDC3*. Fukuda et al. defined two overlapping genes as two adjacent genes, lying on same or different strand, whose expressible nucleotide sequences are partially shared [18], which is the case in analysis.

Figure 1.4 shows the structure of the overlapping elements of the two genes. *OPTN* and *CCDC3* partially share the first exons of, respectively, 510 bp and 256 bp, with *CCDC3* first exon resulting embedded in *OPTN*; for both genes, the shared sequence falls in their 5'-UTR region, not involving their coding sequence.

OPTN and *CCDC3* feature different transcription initiation sites, as their their transcripts can differ substantially in their 5'-UTR regions. Such a variety is possibly originated by the existence of alternative promoters in the same gene and by events of exon skipping.

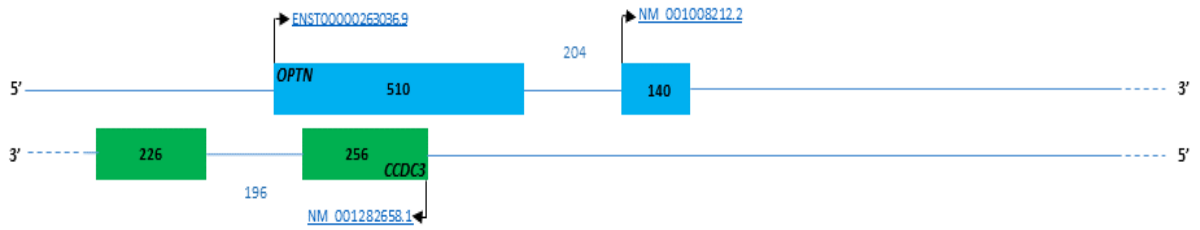

Fig. 1.4: *OPTN* and *CCDC3* overlapping elements. *OPTN* (in blue) and *CCDC3* (in green) lie on opposite strands. First *CCDC3* exon (256 bp) is embedded into first *OPTN* exon (510 bp). The first two exons of both genes are shown. Exon size depicted in black, intron size in blue. Arrows show transcription initiation sites, based on GeneBank and Ensemble accession numbers NM_001008212.2 and ENST00000263036.9 for *OPTN*, NM_001282658.1 for *CCDC3*.

The phenomenon of gene overlapping is a very common event in viral DNA, due to the strict restrictions imposed by their tightly packed structure; in non-viral organisms, though, the potential advantages coming from this particular structure are less clear. Such configuration, however, may suggest co-regulation: a head-to head overlapping structure could allow two genes to share the same promoter element and the same CpG island, resulting in a possible mutual regulation for coordinated expression [19].

In humans, the number of observed overlapping genes coding for a protein is increasing constatly, raising the idea it is a much more frequent phenomenon than previosly estimated [20].

II. Objectives

This work proposes a deeper study of the fine and complex transcriptional regulation of the *OPTN* and *CCDC3* human genes. More specifically, given the high variety of transcripts available in the databases for *OPTN*, we wanted to clarify the structure of both genes by identifying and analyzing new mRNA transcript variants, generated by the cells through alternative splicing or by the activity of putative alternative promoters and/or the effect of other cis-regulatory elements. The mentioned factors were investigated *in silico* with the aid of bioinformatics tools, online dedicated databases and programs for sequence analysis.

Following the *in silico* procedures, we aimed at creating gene reporter constructs containing the promoter sequences under study. In order to assess their activity as transcriptional regulators, their functionality was validated through the luciferase assay technique (section 4.2), after transfecting the gene reporter constructs into cultured HEK 293 cells.

Co-transfections of selected transcription factor expression vectors, together with the gene reporter constructs containing our promoter sequences were performed in the same cultured cells; the purpose of this procedure was to define and measure the effect of a specific transcription factor on the promoter sequence of our interest.

III. Material and Methods

3.1 *In silico* analysis

The very first phases of this thesis work consisted in acquiring more information related to the overall structure of the genes under analysis, performed *in silico* with the support of bioinformatics tools and online databases. Genes DNA is transcribed into mRNA, which is then translated into protein; this process characterizes the expression of a gene, and it is finely regulated at multiple levels.

Every gene in the human genome is organized in specialized sequence elements, involved in the process of gene expression. The control of gene expression is carried on by specific nucleotide sequences in the promoter region, known as *cis*-acting factors, and by proteins, the *trans*-acting factors, that interact with promoter DNA or RNA polymerase II enzyme to let the transcription begin.

At the beginning of this work, it was necessary to elucidate such key elements; in particular, the portions of the gene destined to be withheld in the mature mRNA, the exons, and the intervening portions of DNA between exons, the introns.

Successively, by analyzing the intron/exon structure of the gene, we proceeded to identify the promoter elements, part of the gene that controls the transcription and therefore its expression.

After having identified the sequences of the *OPTN*/*CCDC3* core promoter, and of a *CCDC3* alternative promoter, we performed an investigation on the putative transcription factors binding sites (TFBSs), short nucleotide sequences present in promoter sequences, to which trans-acting protein (transcription factors) can potentially bind and upregulate or downregulate the gene expression.

Completing our *in silico* investigation was the analysis of the different tissular expression of both *OPTN* and *CCDC3*, in order to gather more specific information regarding the prevalence of expression of each gene in different human tissues.

3.1.1 *CCDC3* gene structure analysis - EST and mRNA comparison

The two known *CCDC3* mRNA variants available at the beginning of this work on the Ensembl database (www.ensembl.org) were referring to the NCBI (www.ncbi.nlm.nih.gov) reference sequences (refseq) accession numbers NM_001282658.1 and NM_031455.4. The NM_001282658.1 variant provides a longer mRNA sequence: compared to variant NM_031455.4, it contains five additional alternative exons in the 5'-UTR, it lacks a portion of the 5' coding region, and it initiates the translation at a different downstream start codon. Interestingly, the longer mRNA variant (2300 bp) encodes for a smaller protein of 145 amino acid residues, while the shorter mRNA variant (700 bp) gives origin to a bigger protein of 270 residues.

Given the transcriptional complexity of the *OPTN* gene and its much higher number of transcript variants available, we wanted to understand if *CCDC3* would originate the same variety of transcribed alternative mRNAs. The first task consisted in using the bioinformatics tool Nucleotide BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi), to perform a search in the expressed sequence tags (ESTs) database for *Homo sapiens*. Specifically, we BLAST searched the nucleotide sequence of the exon 13 of the *CCDC3* gene, the first shared 175 bp exon (base pair 12998512 to 12998338 in NC_000010.11) which belonged to both the already known mRNA transcripts (see section 4.1.1).

The retrieved EST cDNA sequences containing the 175 bp common exon were successively analyzed one by one using the bioinformatics tool Splign (ncbi.nlm.nih.gov/sutils/splign/splign.cgi). Splign performs an alignment of the putative mature (spliced) mRNA sequence, here in the form of cDNA, with the whole genomic DNA sequence [21].

By directly comparing the size and position of the exons alignment supplied by Splign with the structure of the existing mRNA variants, we could determine whether each cDNA in analysis belonged to one of the two already known mRNA variants or it would be part of a newly found alternatively spliced mRNA.

3.1.2 *CCDC3* gene structure analysis - AceView database

The genomic database AceView browser (ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html) has been used as a convenient source to retrieve a comprehensive list of all the publicly available mRNA sequences associated to the gene under study; AceView collects and combines the data from GenBank, RefSeq, dbEST and Trace.

The sequences retrieved from AceView were aligned to the rest of the genome by using the Splign software tool, and successively compared to the previously known mRNA splicing patterns.

3.1.3 Analysis of putative TFBSs in *OPTN* and *CCDC3* promoters

Being *OPTN* and *CCDC3* two overlapping gene, the sequence shared and adjacent to the beginning of the two genes was considered as a common regulatory element, working bi-directionally; an analysis of the putative transcription factors binding sites (TFBSs) was performed on this 2120 bp long sequence (base pair 13099043 to 13101162 in NC_000010.11).

The sequence located 5' upstream of the exon 12 (501 bp) of *CCDC3*, corresponding to the position 13002689 to 13001633 in NC_000010.11, with 1057 bp, was identified as a potential alternative promoter and analysed for the presence of putative TFBSs that could potentially bind a corresponding TF and regulate the transcription of the *CCDC3* gene, allowing the expression of specific mRNAs (section 4.1.1).

For such purpose, three different online tools, AliBaba 2.1 (Transfac 4.0) (gene-regulation.com/pub/programs/alibaba2/index.html), PROMO (Transfac 8.3)

(alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) [22], and TFsitescan (www.ifti.org/cgi-bin/ifti/Tfsitescan.pl) were employed, and a list of predicted putative TFBSs was created for both the promoter regions.

3.1.4 *OPTN* and *CCDC3* tissular expression

Respectively 42 and 27 human tissues were examined for *OPTN* and *CCDC3* mRNA expression by mining experimentally verified human expression sequence tags (ESTs) deposited in Unigene database (www.ncbi.nlm.nih.gov/unigene).

The ESTs database is created via cDNA cloning from different cDNA libraries, followed by DNA sequencing. Gene mRNA levels are presented in the Unigene database as gene transcript units per million transcripts (TPM). The relative mRNA expression (REU) of *OPTN* (Hs.332706) and *CCDC3* (Hs.498720) were generated by normalizing the value of TPM of each gene with the one of *β-actin* (Hs.520640), used as reference gene in every tissue considered.

The intervals of the expression variation of the reference genes were generated by calculating the mean plus two times the standard deviation of the REU of three randomly selected reference genes, namely *ARHGDIA* (Rho GDP dissociation inhibitor alpha, number Hs.159161), *GAPDH* (glyceraldehyde-3-phosphatedehydrogenase, number Hs.544577) and *RPS27A* (ribosomal protein S27a, number Hs.311640); the mentioned reference genes were also normalized by *β-actin* values in the considered tissues.

3.2 Cloning of *OPTN* and *CCDC3* promoters sequences into a cloning vector

3.2.1 Amplification of promoter sequences

In order to clone the *OPTN-CCDC3* shared promoter and the *CCDC3* alternative promoter sequences into a cloning vector, the first step consisted in the amplification of the sequences through the polymerase chain reaction (PCR). The PCR technique uses repeated thermal cycles of heating and cooling to allow DNA melting and DNA replication through enzymatic reaction. Short oligonucleotides (primers) are deployed to begin the amplification process; a thermal-resistant DNA polymerase enzyme is applied to elongate the primer sequences by adding nucleotides using the complementary DNA strand as template. The use of

forward (FW) and reverse (REV) primers is necessary in order to set the beginning and the end of the fragment to amplify, and to overcome the functional limitation of the DNA polymerase, which can only catalyse the SN2 nucleophilic attack of the 3'-OH end to the α-phosphate of the incoming complementary nucleotide, being therefore able to extend an existing DNA strand and incapable of synthesizing a DNA strand *de novo* [23].

OPTN-CCDC3 shared promoter and *CCDC3* alternative promoter sequences were amplified by using genomic DNA (gDNA) extracted from T-47D human cell line as template. The PCR reaction mix was prepared as it follows: 1 μL of gDNA (100 ng/μL), 1 μL of Kapa® Hifi DNA polymerase enzyme (1 U/μL; Kapa Biosystems®), 1.5 μL dNTPs (10 μM; Kapa Biosystems®), 10 μL of 5X Hifi buffer, 1.5 μL of each FW and REV primer (10 μM), and 33.5 μL of ultra-pure water (Milli-Q®), to reach a final volume of 50 μL.

The primers were designed with the aid of PerlPrimer software, according to the functional elements present in the sequences. Table 3.1 shows in detail the primers used for the amplification.

Table 3.1: Primers names and sequences used to amplify the DNA fragments under analysis.

Construct name	FW and REV primer name	FW and REV primer sequence	Construct length
OPTN-CCDC3_F9R12	HsaOPTN_F9 HsaOPTN_R12	5'-TTAAATTCTCTATTTCTCCCCACTCC-3' 5'-TGACCCTGAGCGAAGCCAAGCCG-3'	1252 bp
OPTN-CCDC3_F9R2	HsaOPTN_F9 HsaOPTN_R2	5'-TTAAATTCTCTATTTCTCCCCACTCC-3' 5'-TTCTCTCCCTCTCTCCCTCC-3'	326 bp
OPTN-CCDC3_F18R9	HsaOPTN_F18 HsaOPTN_R9	5'-TGAGTGTATTTTAAAGCAAAAACGA-3' 5'-CCACTACGGGATCTGCGGGAAGA-3'	187 bp
OPTN-CCDC3_F18R20	HsaOPTN_F18 HsaOPTN_R20	5'-TGAGTGTATTTTAAAGCAAAAACGA-3' 5'-GAGAAGTCCCAGGGCAGAC-3'	1520 bp
CCDC3_AP1	HsaCCDC3_apF1 HsaCCDC3_apR2	5'-TAAATATTCAGGGTGGATGTGGG-3' 5'-GAGCAGCCGAGCGCCCAGGGCTGCCCTT-3'	1057 bp
CCDC3_AP2	HsaCCDC3_apF2 HsaCCDC3_apR2	5'-CTGAATGTATTTCTCAGGTGTACAG-3' 5'-GAGCAGCCGAGCGCCCAGGGCTGCCCTT-3'	656 bp
CCDC3_AP3	HsaCCDC3_apF3 HsaCCDC3_apR2	5'- TCTTTAATGAATGCCTTGCG -3' 5'-GAGCAGCCGAGCGCCCAGGGCTGCCCTT-3'	373 bp

The reaction was performed in a thermal cycler machine (model 2720; Applied Biosystems®). The initial denaturation step was set at 95 ºC for 5 minutes; a number of 30 cycles were programmed with the following parameters: denaturation at 98 ºC for 20 seconds, annealing at 55 ºC for 15 seconds, and extension at 72 ºC for 2 minutes; the final extension step ran at 72ºC for 5 minutes.

3.2.2 PCR products separation in gel electrophoresis and DNA extraction

The amplified sequences obtained by PCR were separated by size through agarose gel electrophoresis; the 1% agarose gel was obtained by dissolving 1 g of agarose powder (Sigma-Aldrich®) in 100 mL of 1X TAE buffer (tris base 40 mM, acetic acid 20 mM, EDTA sodium salt dihydrate 1 mM); 2 µl of intercalating agent (Green Safe®; Nzytech®) were added to the gel solution. The gel underwent 120V of electric field for 25 minutes.

An ultra-violet transilluminator was used to visualize and identify the fragments of the desired size, by running in the same gel 7 µL of GeneRuler 1kb DNA ladder marker (ThermoFisher Scientific®) and used as size reference.

The DNA fragments were extracted from the gel and purified with the aid of the GeneJET Gel extraction Kit® (ThermoFisher scientific®); the procedure of DNA extraction consists in melting the previously excised fragment of gel at 55 ºC in a provided binding buffer (1 µL of buffer for every mg of agarose), and then centrifuged through a purification column.

3.2.3 DNA ligation into pCR 2.1-TOPO® cloning vector

The following step consisted in ligating the purified PCR products in a cloning vector. The chosen vector was the pCR 2.1-TOPO® (Thermo Fisher Scientific®) (see the Appendix section). The ligation occurred between the 3'-adenine overhangs of the DNA fragments, added by the DNA polymerase during the PCR reaction, and the free thymine situated at the extremities of the TOPO vector.

The ligation reaction was performed with 4 µL of purified PCR products, 1 µL of salt solution and 1 µL of pCR 2.1-TOPO® cloning vector, and incubated at room temperature for 30 minutes.

3.2.4 Vector transformation in *E. coli* DH5α competent bacteria

In order to create multiple copies of the cloning vectors containing the promoter constructs of our interest, we performed bacterial transformation of the recombinant pCR 2.1-TOPO® vectors in *E. coli* strain DH5α competent bacteria. The DH5α cells were made competent, i.e. capable of receiving exogenous DNA, by treating them with calcium chloride ($CaCl_2$) during their exponential phase of growth; the treatment allows the temporary destabilization of the bacterial cell wall and increases its permeability to DNA [24].

To one microcentrifuge tube containing 100 µL of DH5α competent bacteria, 3 µL of the ligation product were added; the bacteria were at first incubated in ice for 30 minutes, and then underwent a heat shock at 42 ºC during 45 seconds, followed by 2 minutes back to ice; this heat shock procedure increased the capability of the bacterial cells to incorporate the exogenous DNA, in the form of cloning vectors containing the DNA fragments of interest [25].

Then, 300 µL of SOC nutritional media (20 mM of glucose, 20 mM of $MgCl_2$, and 2 mM $MgSO_4$) were added to the bacteria, and, successively, they were left to grow in suspension in a heated agitator (37 ºC at 200 RPM) for 1 hour.

Afterwards, 150 µL of grown bacteria were plated in LB agar culture plates containing 100 µg/mL of ampicillin antibiotic; 5 µL of IPTG (isopropyl β-D-1-thiogalactopyranoside, 10 mM), a lactose-metabolite-analog to allolactose that enhances the transcription of the *lac* operon by binding to the *lac* repressor, were added to the plate, in combination with 40 µL of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, 20 ng/µL), a chromogenic substrate of β-galactosidase analog of lactose; the two reagents were deployed in order to perform a blue-white screening: cells with white phenotype will have a disrupted *lacZ* gene and absence of β-galactosidase production, due to the insertion of the DNA fragments; the plates were then incubated overnight at 37 ºC.

As following, bacteria colonies showing a white phenotype were selected, extracted from the plate, and each colony grown separately in bacterial tubes with 2 mL of LB liquid medium, containing 2 µL of ampicillin (50 ng/µL); the colonies were left growing overnight in a heated agitator (37 ºC at 200 RPM).

3.2.5 Extraction of plasmidic DNA and confirmation of DNA incorporation

The extraction of plasmidic DNA from bacterial cells was performed by following a procedure of alkaline cell lysis, as described by Sambrook et al. [24]. The bacteria cultures containing the plasmids of interest were centrifuged at 14000 RPM for 2 minutes; the precipitated pellet was re-suspended in 100 µL of a first solution (P1) of Tris-HCl (50 mM, pH 8), EDTA (10 mM) and RNase A (100 µg/mL), followed by adding 100 µL of a second solution (P2) of 1% SDS (sodium dodecyl sulfate) and NaOH (200 mM), and incubated for 5 minutes at room temperature. Then, 100 µL of a third solution (P3) of KOAc (3 M, pH 5.5) and acetic acid (CH_3COOH) were added; the solution was incubated on ice for 10 minutes and later centrifuged at 14000 RPM for 8 minutes. The supernatant was collected, 600 µL of 100% ethanol were added and later centrifuged; the pellet was washed with 600 µL of 70% ethanol,

and finally re-suspended in 30 µL of pure water.

After extracting and isolating the DNA from the bacterial cells, *EcoR*I restriction enzyme digestions were performed on 3 µL of the extracted DNA samples, in order to visually screen the positive samples through a gel run and confirm the effective incorporation of the DNA fragments of interest into the plasmidic vectors. The restriction endonuclease *EcoR*I cuts the pCR 2.1-TOPO® vector in two specific regions, flanking at both 5' and 3' the incorporated DNA fragments. The restriction enzyme digestion was conducted by mixing 0.2 µL of *EcoR*I (15 U/µL; Takara®), 1.5 µL of buffer FD (Takara®), 3 µL of extracted DNA and 10.3 µL of sterile water for a final volume of 15 µL, and incubated for 1 hour at 37 ºC in a block heater (Stuart®).

The process originated two DNA fragments: the first with the size of the amplified fragments, and the second with the size of the plasmid. This was visible by separating the digested products by size through an agarose gel electrophoresis, and visualizing the gel on a UV transilluminator. We were able to distinguish the positive samples and select the ones which had incorporated our fragments, featuring two DNA bands in the gel.

The positive samples showing the expected size were sequenced to confirm the nucleotide sequence and the orientation of the inserts into the pCR 2.1-TOPO® plasmid vector. The Sanger sequencing reaction was performed in the CCMAR Molecular Biology Platform using primer SP6 (5'-ATTTAGGTGACACTATAG-3') as forward primer, and primer T7 (5'-TAATACGACTCACTATAGGG-3') as reverse primer.

3.3 Sub-cloning of promoter constructs to pGL3-Basic® reporter vectors

Once the nucleotide sequences and the orientation of the promoter fragments inside the plasmid had been confirmed, we proceeded with the sub-cloning of the inserts under analysis into the pGL3-Basic® vector (Promega®).

This plasmid is a luciferase reporter vector used for the quantitative analysis of cis-acting (promoters, enhancing sequence elements) and trans-acting factors (DNA-binding proteins, transcription factors) that regulate the eukaryotic gene expression (see Appendix section for the vectors structure).

The vector is engineered to provide a quantitative feedback on the regulatory activity of the sequences of interest; the elements under investigation, when acting as a promoter, will directly control the transcription of the luciferase gene and consequentially the quantity of the produced luciferase protein. To obtain this, the DNA fragments were inserted in the polylinker

site of the vector, upstream of the luciferase gene *luc+* contained in the pGL3 vector, and then the vector transfected into eukaryotic cells; therefore, the quantification of the luminescence produced by the transfected cells is proportional to the quantity of luciferase protein produced, which is directly regulated by the promoter activity of the sub-cloned sequences.

In order to perform the sub-cloning, both constructs, specifically 1 µg of the pCR 2.1-TOPO® containing the DNA fragments of interest and 1 µg of the empty pGL3-Basic®, were digested with the restriction endonucleases (Takara®) shown in table 2.2.

Tab. 3.2: List of restriction endonucleases used to digest the fragments contained in pCR 2.1-TOPO® vector and the empty pGL3-Basic® vector.

pCR 2.1-TOPO® source vector	**Restriction endonucleases (buffer)**	**pGL3® constructs**	**Constructs description**
OPTN-CCDC3_F9R12	*Kpn*I + *Xho*I (M buffer)	F9R12-*Luc*	Shared promoter, *OPTN* direction
OPTN-CCDC3_F9R12	*Xho*I + *Hind*III (M buffer)	*Luc*-F9R12	Shared promoter, *CCDC3* direction
OPTN-CCDC3_F9R2	*Kpn*I + *Xho*I (M buffer)	F9R2-*Luc*	Shared promoter, *OPTN* direction
OPTN-CCDC3_F18R9	*Xho*I + *Hind*III (M buffer)	*Luc*-F18R9	Shared promoter, *CCDC3* direction
OPTN-CCDC3_F18R20	*Xho*I + *Hind*III (M buffer)	*Luc*-F18R20	Shared promoter, *CCDC3* direction
CCDC3_AP1	*Kpn*I + *Xho*I (M buffer)	AP1-*Luc*	*CCDC3* alternative promoter
CCDC3_AP2	*Kpn*I + *Xho*I (M buffer)	AP2-*Luc*	*CCDC3* alternative promoter
CCDC3_AP3	*Xho*I + *Hind*III (M buffer)	AP3-*Luc*	*CCDC3* alternative promoter

To the plasmidic DNA we added 2 µL of the respective restriction buffers (10X, Takara®) according to the combination of restriction enzyme deployed, plus 0.5 µL (10 U/µL) of each endonuclease, and sterile water, up to a final volume of 20 µL. The mix was then incubated at 37 ºC for 2 hours. The products of the reaction were separated by electrophoretic run on agarose gel (1%), using the non-digested pGL3-Basic® as band reference; the products were extracted from gel and purified with the GeneJET Gel Extraction Kit® (ThermoFisher scientific®). The DNA concentrations of the extracted samples were later quantified by measuring the absorbance of ultraviolet light at a wavelength of 260 nm, using a NanoDrop One® microvolume UV-Vis spectrophotometer machine (ThermoFisher scientific®).

Next step consisted in ligating the DNA fragments into the pGL3-Basic® plasmids, digested with the same restriction endonucleases. For the purpose, a T4 DNA ligase enzyme was used; T4 ligase catalyzes the formation of a phosphodiester bond, which connects two adjacent nucleotides, between the contiguous 5'-phosphate and 3'-hydroxyl chain terminations in duplex DNA configuration, repairing the nicks and ligating the sequences. A ratio of 3:1 of insert:vector was used; the ligation reaction mix was prepared as follows: 100 ng of digested pGL3 basic vector, 300 ng of digested TOPO inserts, 1 µL of 10x T4 buffer (Promega®), 1 U of T4 DNA ligase (Promega®), and distilled water, up to a final volume of 10 µL; the mix is

then incubated overnight at 4 ºC.

The obtained recombinant pGL3® plasmids, containing the promoter fragments, were then replicated after bacterial transformation in competent DH5α strains of *E. coli* bacteria, followed by miniprep extraction and purification, applying the same procedures described in section 3.2.2.

The pGL3® constructs were screened by restriction endonuclease digestion, in order to select the positive samples: 3 µL of plasmidic miniprep DNA with 0.5 µL of *Bgl*II (10 U/µL), 0.5 µL of *Hind*III (15 U/µL), 1.5 µL of 10X K buffer, and sterile water, for a total volume of 15 µL, were incubated at 37 ºC for 1 hour; the reaction products were ran in electrophoresis agarose gel with 1X TAE buffer and 2 µL of Green Safe® intercalating agent, and then visualized by using a UV transilluminator. The positive samples, the ones in the gel showing two fragment bands (the pGL3-Basic plasmid and the promoter fragment inserted), were again sequenced by Sanger reaction, performed with R240 primer (5'-ATGGAAGACGCCAAAAACATAAAG-3') at the CCMAR Molecular Biology platform.

3.4 Transient transfection

3.4.1 HEK 293 cell culture maintenance

HEK 293 (human embryonic kidney) eukaryotic cell line was used for the transient transfection procedure. Cells were kept in DMEM (Dulbecco's modified eagle medium) culture medium (Gibco®) with added supplements (1% L-glutamine, 1% penicillin and streptomycin, and 10% fetal bovine serum), in a 5% CO_2-enriched incubator, at the temperature of 37 ºC.

Cell cultures were passed to new cell culture dishes and provided with fresh culture media every three and four days, to prevent lack of nutrients for the correct development of the cells, unwanted pH alterations of the media, and growth inhibition due to cell-cell interaction.

Each re-plating was performed following these steps: washing the cells with 10 mL of PBS (phosphate buffered saline), detaching cells with 1 mL of T solution (NaCl 137 mM + KCl 2.7 mN + Na_2HPO_4 8.1 mM + KH_2PO_4 1.47 mM + 0.2% volume/volume trypsin, at pH 7.4) applied at 37 ºC for a few seconds, re-suspending cells in 10 mL of supplemented DMEM, and seeding cells in a new 100 mm diameter cell culture plate (Sarstedt®), maintaining a density of $87x10^5$ or $54x10^5$ cells per plate. The desired densities were calculated by counting cells in a Neubauer® chamber with an inverted microscope (Zeiss® Axiovert 25), and then by diluting cells with supplemented DMEM according to the desired concentration.

3.4.2 Transient transfection of pGL3 recombinant vectors into HEK 293 cells

The process of transient transfection consists in the introduction of an exogenous DNA portion into eukaryotic cells. In this work, the exogenous DNA was the recombinant pGL3 vectors containing the promoter regions under analysis.

HEK 293 (human embryonic kidney) cells were plated in a 24-wells plate, at a density of $5x10^4$ per well, approximately 18 hours before the transfection occurred, in order to reach the desired value of confluency of 50 to 70%.

To begin the transfection, a mix of 1 µL of XtremeGene HP® DNA transfection reagent (Roche®) with an amount of 250 ng of each reporter construct, plus 5 ng of pRL-TK® *Renilla* luciferase expressing vector (Promega®) (see Appendix section for vector details) used to normalize the luciferase expression, plus DMEM without supplements, reaching a final volume of 100 µL, was incubated at room temperature for 15 minutes and then provided to the cells drop by drop, divided in two wells.

Co-transfection consisted in adding in a single well 125 ng of each reporter construct, 2.4 ng of Renilla luciferase normalizing vector, and 25 ng of each expression vectors containing p50 and p65 NF-κB isoforms (pCMV4_p50 and pCMV4_p65; Addgene plasmid 21965 and 21966, respectively), one of the transcription factors that were identified to putatively regulate the promoter activity. In an independent well, 25 ng of the empty expression vector (pCMV4; GeneBank AF239248) were added. The plasmids pGL3-Basic® and pGL3-Control® were transfected in separated wells and used respectively as negative and positive control. The plate was then incubated in a 5% CO_2-enriched atmosphere at 37 ºC.

3.5 Luciferase activity assay

The process of transfection and co-transfection is halted after 48 hours of incubation, by removing the plate from the incubator, aspiring out the growth media, and washing the cells with cold PBS. Cells were then lysated by adding 100 µL of 1X passive lysis buffer (Promega®), and placed for 20 minutes in a shaker (VWR®). With the aid of a scrapper, cells lysates were removed from the 24 wells plate and collected in microcentrifuge tubes. The lysate underwent a 14000 RPM centrifugation for 30 seconds, after which the supernatant was transferred to a new microcentrifuge tube, and the lysate pellet discarded.

Firefly luciferase is widely used as a reporter gene for studying gene regulation; the produced enzyme catalyzes the oxidation of luciferin to oxyluciferin (see Figure 3.3), by

converting the chemical energy of the substrate D-Luciferine into photon emission (hv).

The luminescence emitted by the cells, directly proportional to the amount of firefly luciferase protein synthesized, is connected to the activity of the promoters under study, thus allowing a quantification of the basal activity of each construct; the Renilla luciferase protein (RLuc), a second bioluminescent enzyme whose structure, function and substrate/product differ from the firefly luciferase (Figure 3.3), is used as a reference to normalize the luminescence values obtained.

The dual Firefly & Renilla luciferase single tube assay kit (Biotium®) was used for the luminescence reaction; specifically, 100 µL of firefly luciferase buffer, and 50 µL of Renilla luciferase buffer were transferred, together with 10 µL of each cell lysate, to a Bio-One 96 well microplate (Greiner®), and the luminescence activity measured in a microplate reader machine (BioTek® Synergy 4).

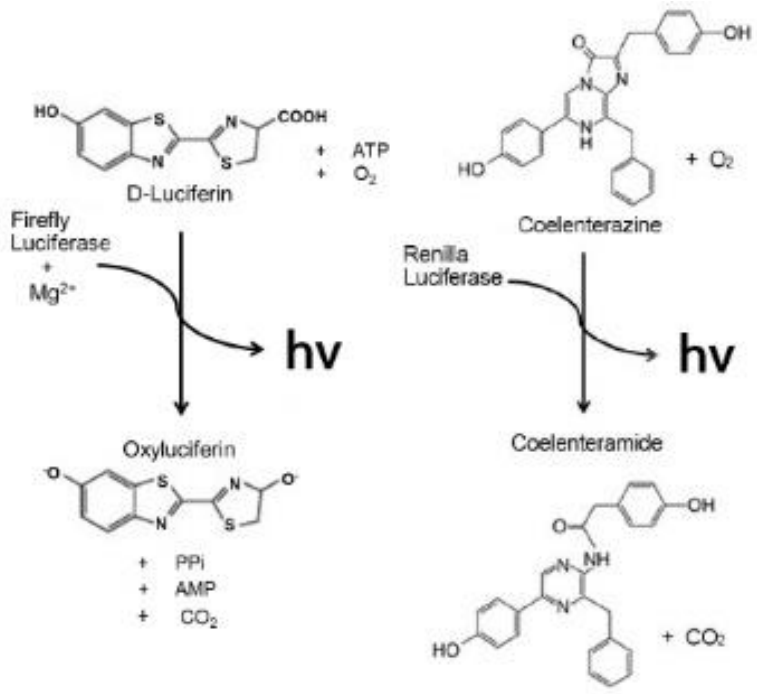

Fig. 3.3: Luminescence reaction of Firefly and Renilla luciferase. Note the differences in substrates, co-factors and products of the two bioluminescent enzymes. hv represents light as photon emission. Adapted from bpsbioscience.com.

Firefly/Renilla luciferase activity values were calculated over a minimum of two different readings, and completed with the calculation of the value of standard deviation.

Statistical data analysis of the reading results was elaborated using the software Prism 8 (Graphpad®) with one-way ANOVA and Tukey's multiple comparisons test; results were considered statistically significant for values of $P < 0.05$.

IV. Results and Discussion

4.1 *In silico* analysis results

4.1.1 *CCDC3* gene structure

The *in silico* search of new EST sequences in databases, performed to obtain more structural information about *CCDC3*, gave us partial sequences that were referring to the structures of the two previously known alternative mRNAs. The only exception to this trend was the EST sequence HY075572.1 (GenBank), retrieved from a human cDNA library, in which the *CCDC3* gene was expressed in thymus tissue (Figure 4.1).

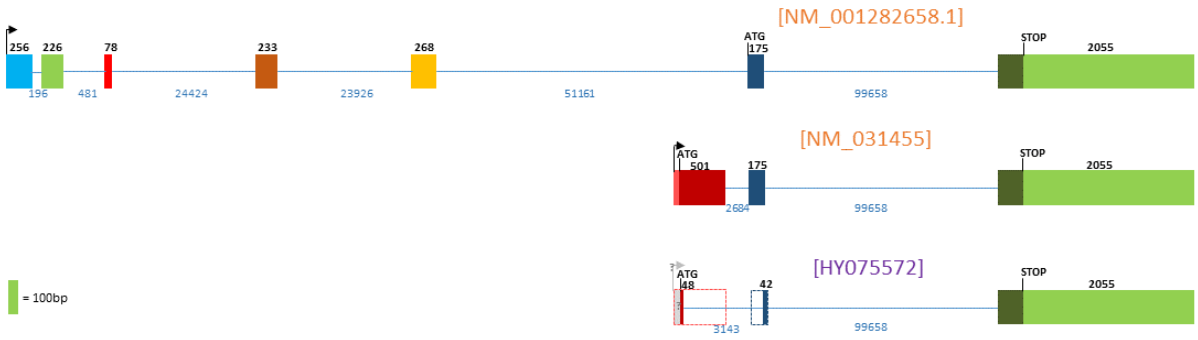

Fig. 4.1: *CCDC3* transcripts structural comparison. Previously known mRNA sequences (accession numbers in orange) and newly found HY075572.1 alternatively spliced mRNA variant (in purple) intron/exon structure are compared. In-frame ORF, same start and stop codon and alternative internal splicing sites are emphasized. Exon size depicted in black, intron size in blue. Due to space limitation, introns are not in scale.

The newly identified HY075572.1 alternatively spliced mRNA variant shares the same in-frame start and stop codons as the previously known NM_031455. Interestingly, the internal splicing sites of this mRNA variant do not follow the common GT-AG scheme, also known as the Chambon's rule, according to which the first two and the last two nucleotides of introns are GT and AG, respectively [26]; the nucleotide pattern we found, instead, is an unusual GC-TG. The possibility of more cases of alternative splice sites violating this "splicing rule" has been deeply analysed by Szafranski et al. in 2007 [27].

We created a final schematic representation of *CCDC3* alternative transcripts (Figure 4.2). The scheme was built by embracing the cDNA sequences of *CCDC3* that were retrieved from Ensembl, ESTs BLAST search, and AceView Browser combined, by assembling a total of 13 differentially spliced mRNAs. After analyzing the introns/exons structure of the

alternative *CCDC3* transcripts, we suggested the existence of an alternative promoter for the *CCDC3* gene, upstream exon number 12 (501 bp, base pair 13001697 to 13001197 in NC_000010.11), as shown in Figure 4.3. In transcripts NM_031455 and HY075572, exon 12 is the first exon identified.

Given the far distance of the core promoter from exon 12, located over 98 kb upstream, we suggested an alternative *CCDC3* promoter that can directly control the transcription of the shorter mRNA variant NM_031455 and the newly identified HY075572.1 transcript variant. The transcriptional activity of the longer mRNA variants, such as NM_001282658.1, would be, instead, under the control of the core promoter sequence found in the overlapping structure shared with *OPTN* (see section 1.5). The existence of this *CCDC3* alternative promoter can open a new scenario of differential gene expression, in which different *CCDC3* transcripts are under the independent control of a different promoter, possibly regulated by a variable set of transcription factors which could be solely expressed in specific tissues.

The mRNAs shown in Figure 4.2 differ at both 5' and 3' ends, including several cases of exon skipping, alternative 5' donor sites and 3' acceptor sites. Genes presenting differences in the 5'-UTR region of their transcripts, as for *CCDC3*, are relatively common: between 10 and 18% of human genes express mRNAs featuring alternative 5'-UTR by using alternative promoters [28] [29], as alternative splicing affects 13% of genes in the mammalian transcriptome [30]; these sequences variations in the 5′-UTR could function as important switch elements able to regulate gene expression, possibly allowing certain transcripts to be expressed in determined tissues instead of a systemic expression [14] [70].

In transcripts number DA200147, BY798997 and DA399174 (Figure 4.2), the lack of exon number 12 (501 bp) and 13 (175 bp) in their structure, might let the translation machinery choose alternative in-frame start and stop codons. Some of the incomplete mRNAs sequences could also be non-functional and targeted by the process of nonsense-mediated mRNA decay (NMD), a translation-coupled mechanism that eliminates mRNAs containing premature translation-termination codons or other nonsense aberrant mRNAs [31], in order to reduce errors in gene expression [32].

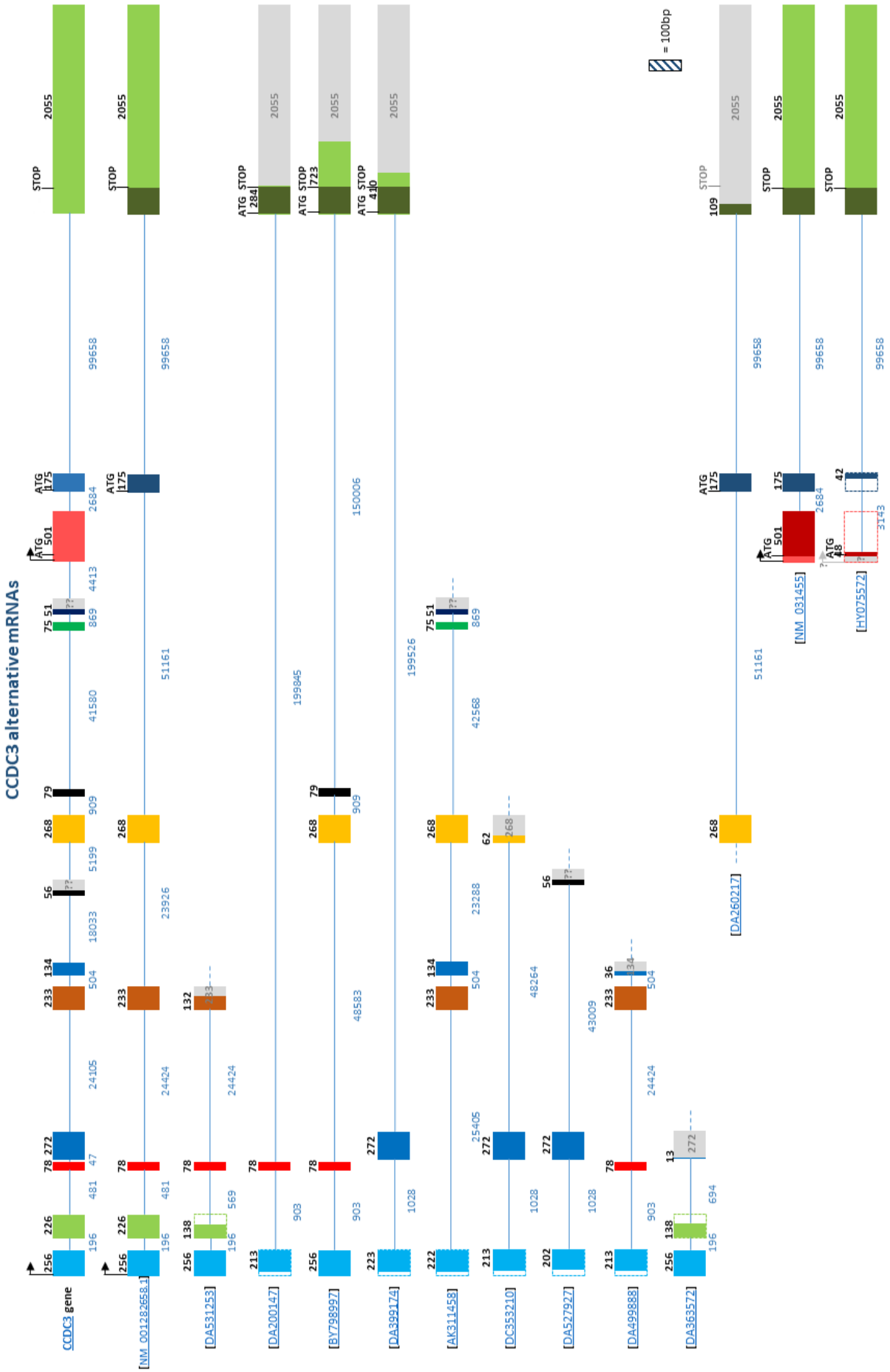

Fig. 4.2: *CCDC3* alternative mRNAs map. Representation of differential *CCDC3* gene expression. Same sequences are in same colors. Exons size is in black, introns size in blue (bp). Arrows show transcription initiation. Introns are not in scale.

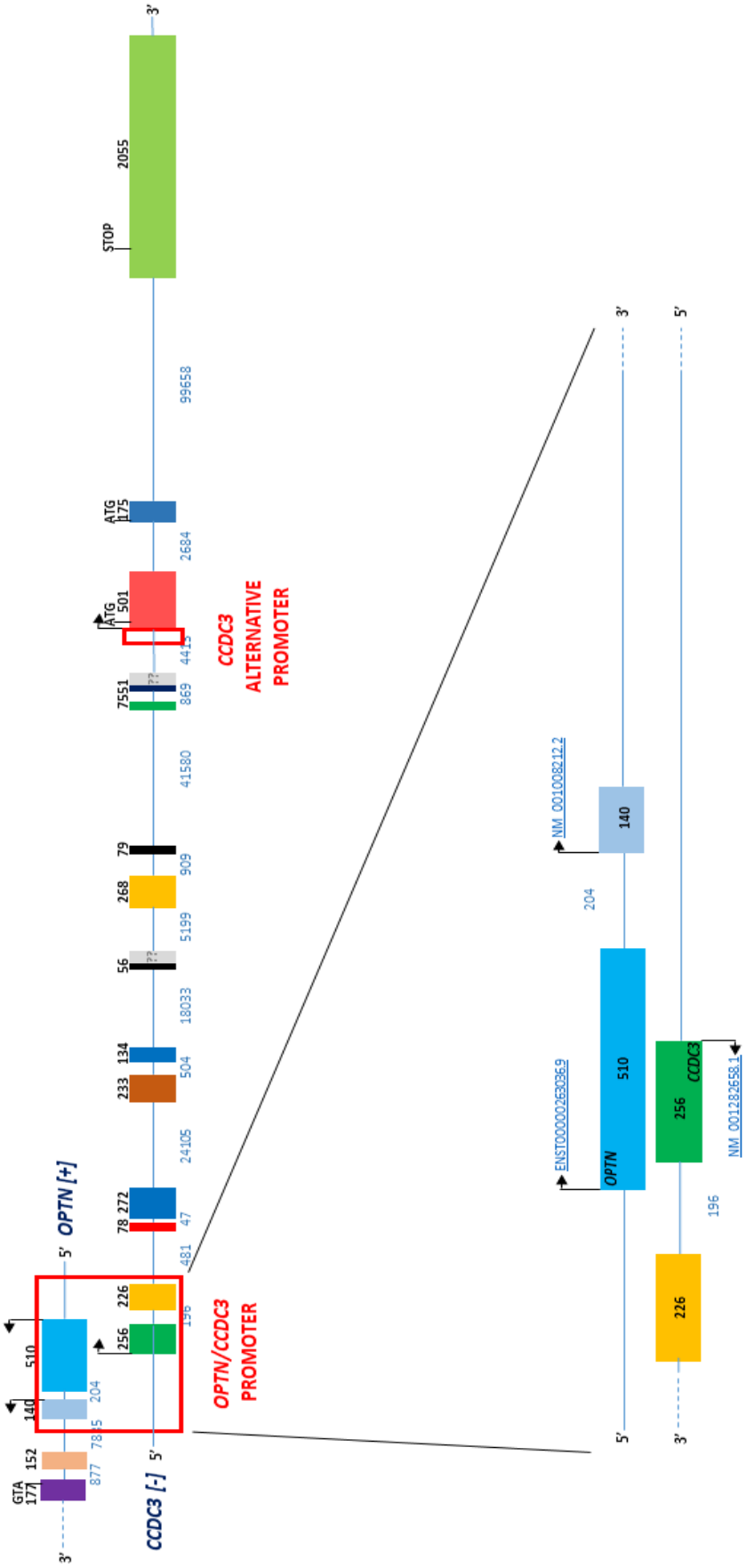

Fig. 4.3: *OPTN* and *CCDC3* promoter regions under analysis. Gene sequences used to build the reporter gene constructs. Exons size depicted in black, introns in blue (bp). Arrows show the transcription initiation site.

4.1.2 *OPTN* and *CCDC3* tissular expression

OPTN and *CCDC3* are widely expressed in several tissues; according to our results, *OPTN* gene is highly transcribed in muscle, pituitary gland and parathyroid, while *CCDC3* gene in pituitary gland, nerve and adipose tissue (Figure 4.4). Kobayashi et al. (2010), after performing northern blot analysis, found that *CCDC3* is highly expressed in the aorta (endothelial tissue) and in adipose tissue [16].

Based on our data, both *OPTN* and *CCDC3* genes appear expressed in the same tissues but in different amounts, such as in pituitary gland, adipose tissue, nerve and heart (Figure 4.4). Apparently, the expression of one gene does not seem to exclude the transcription of the other. Besides this, from the results obtained we cannot tell which promoter has been used by the transcription machinery when the gene is expressed in a specific tissue. Probably, when *OPTN* is transcribed, the transcription machinery precludes the expression of *CCDC3* gene from the same shared promoter, leaving the alternative promoter free to be accessed; in this scenario, the RNA polymerase may transcribe an alternative mRNA with a different 5'-UTR region, undergoing a diverse regulation [14].

Unfortunately, at this level of analysis we can only speculate upon a possible scenario, as unigene and other similar databases have a limitation: they collects tissue information from the ESTs in the databases, but they do not provide information about which transcripts, originated from the same gene, is being expressed in that given tissue; EST profiles, in fact, only show approximate gene expression patterns as retrieved from EST counts and cDNA library sources, as reported by sequence submitters. For the aforementioned reasons, the obtained data results incomplete and not totally accurate, but it still delivered a generic picture on the tissular distribution of the expressed genes under analysis.

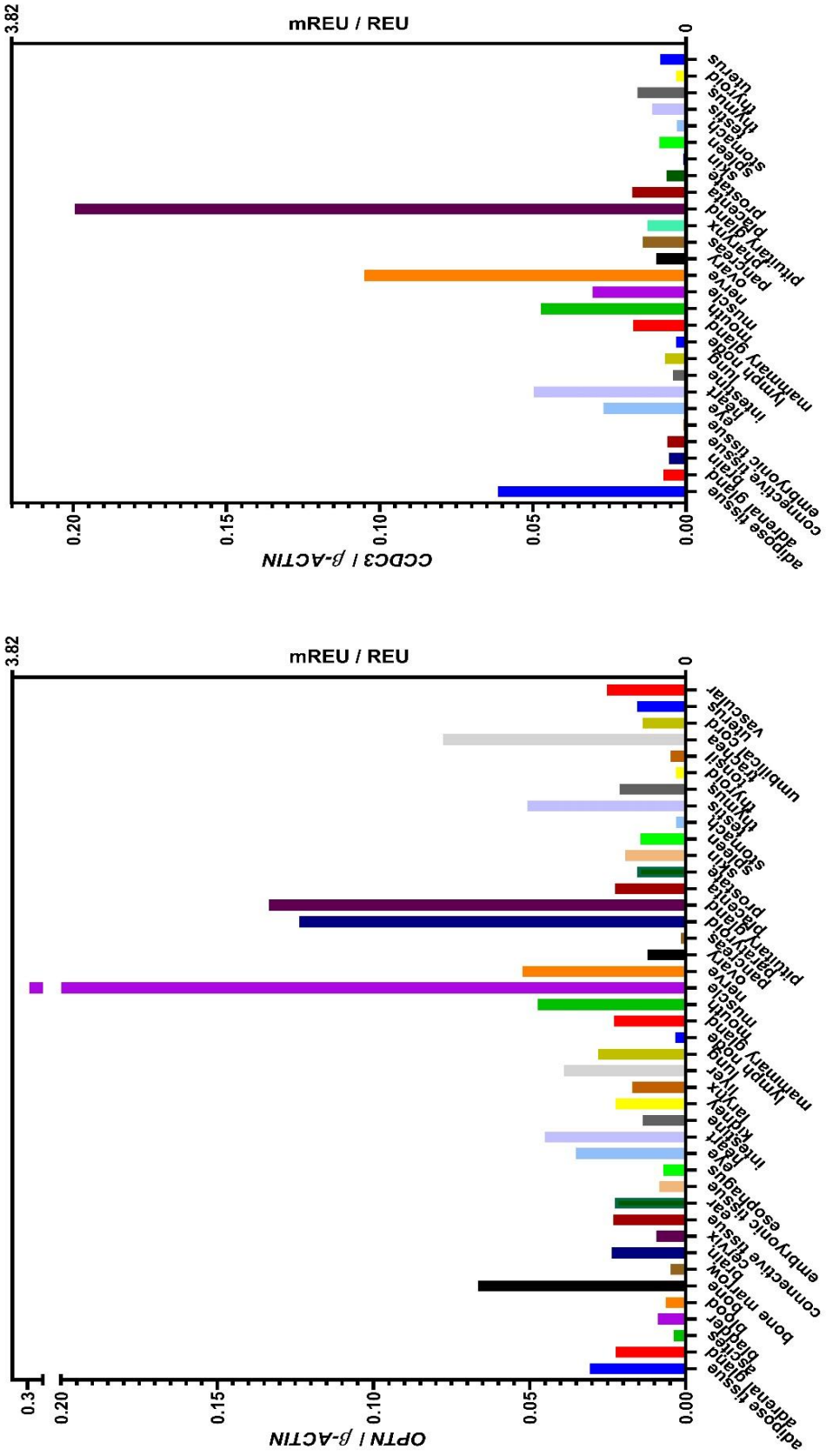

Fig. 4.4: Tissular expression of the genes *OPTN* and *CCDC3*. Graphics obtained based on Unigene database. Values indicated as gene transcript units per million transcripts (TPM).

4.1.3 TFBSs prediction

A list of the predicted transcription factor binding sites (TFBSs) was obtained by analysing the nucleotide sequences of the promoters with bioinformatics software (Table 4.5a and 4.5b). Three different online tools have been used (Alibaba, PROMO and Tfsitescan), and we selected and listed the TFs that were predicted by all of the three programs.

Table 4.5a: Putative *in silico* predicted TFBSs in *OPTN*/*CCDC3* promoter. Numbers represent the predicted occurrence of the binding site for each TF.

TF	Alibaba	PROMO	Tfsitescan
AP-1	1	9	14
AP-2αA	14	6	17
ATF	1	12	1
C/EBPα	10	7	2
C/EBPβ	1	4	2
CREB	1	9	1
Elk-1	1	9	2
GATA-1	4	6	1
HNF-3	1	8	6
NF-κB	**4**	12	19
PEA3	1	9	11
PR	1	7	1
PU.1	1	13	1
Sp1	82	10	160
T3R-β1	2	9	1
YY1	2	4	5

Table 4.5b: Putative *in silico* predicted TFBSs in *CCDC3* alternative promoter. Numbers represent the predicted occurrence of the binding site for each TF.

TF	Alibaba	PROMO	Tfsitescan
AP-1	3	9	1
AP-2αA	7	6	10
c-Jun	1	8	1
ER	2	5	1
GATA-1	4	6	5
HNF-3	3	8	8
NF-κB	**2**	**11**	**1**
T3R-β1	1	9	1
YY1	1	4	3

A map of the putative TFBSs, showing the specific sequences position in *OPTN*/*CCDC3* promoter and in *CCDC3* alternative promoter sequence, was built based on the results obtained with the software Alibaba, in order to allow a spacial recognition of the TFs (Figure 4.6).

We focused on analysing the function of this selected set of 16 TFs identified in the *OPTN*/*CCDC3* promoter region and 9 TFs in the *CCDC3* alternative promoter region.

Activator protein 1 (AP-1) is a family of dimeric TFs composed of Jun, Fos or ATF (activating transcription factor) subunits, that bind to a common AP-1-binding site [33]; it regulates gene expression in response to cytokines, growth factors, stress, and bacterial or viral infections [34], and controls cellular processes including differentiation, proliferation, and apoptosis [35]; in bone development, AP-1 Fos/Jun subunits regulate osteoblasts and osteoclast maturation [71].

Activating enhancer binding protein 2 alpha A (AP-2αA) plays a key role in gene expression regulation in early development, apoptosis and cell-cycle, and tumorigenesis [36]; AP-2 also downregulates the expression of Frizzled-1, responsible for osteoblasts and primary bone marrow stromal cells differentiation [72].

CCAAT-enhancer-binding proteins (C/EBPs), C/EBPα and C/EBPβ, are a ubiquitous TF family that promote gene expression by binding to the CCAAT (cytosine-cytosine-adenosine-adenosine-thymidine) motif of the promoter [37]; moreover, C/EBPβ has a role in osteoporosis: the upregulation of a long isoform of C/EBPβ decreases the number of osteoclasts and slows down the osteoporotic process, while the upregulation of a short isoform of C/EBPβ increases the loss of bone mass [38].

cAMP response element-binding protein (CREB) binds to DNA motifs called cAMP response elements (CRE), and modulate the transcription of the gene [39]; CREB plays a role in neuronal plasticity, in long-term and spacial memory in brain [40]; overexpression of CREB is considered as a possible therapy for Alzheimer's disease [41].

ETS Like-1 (ELK-1) belongs to the E26 transformation-specific (ETS) family of TF, and is involved in long-term memory formation, Alzheimer's disease, drug addiction, depression, Down syndrome and breast cancer [42] [43].

GATA-1 belongs to the GATA family, a group of TF able to bind to the DNA sequence guanosine-adenosine-thymidine-adenosine; the gene coding for GATA-1 is located in the X chromosome in humans and mice [44]; GATA-1 modulates the expression of genes involved in the maturation of blood cells, specifically erythroblasts into erythrocytes and thrombocytes [45].

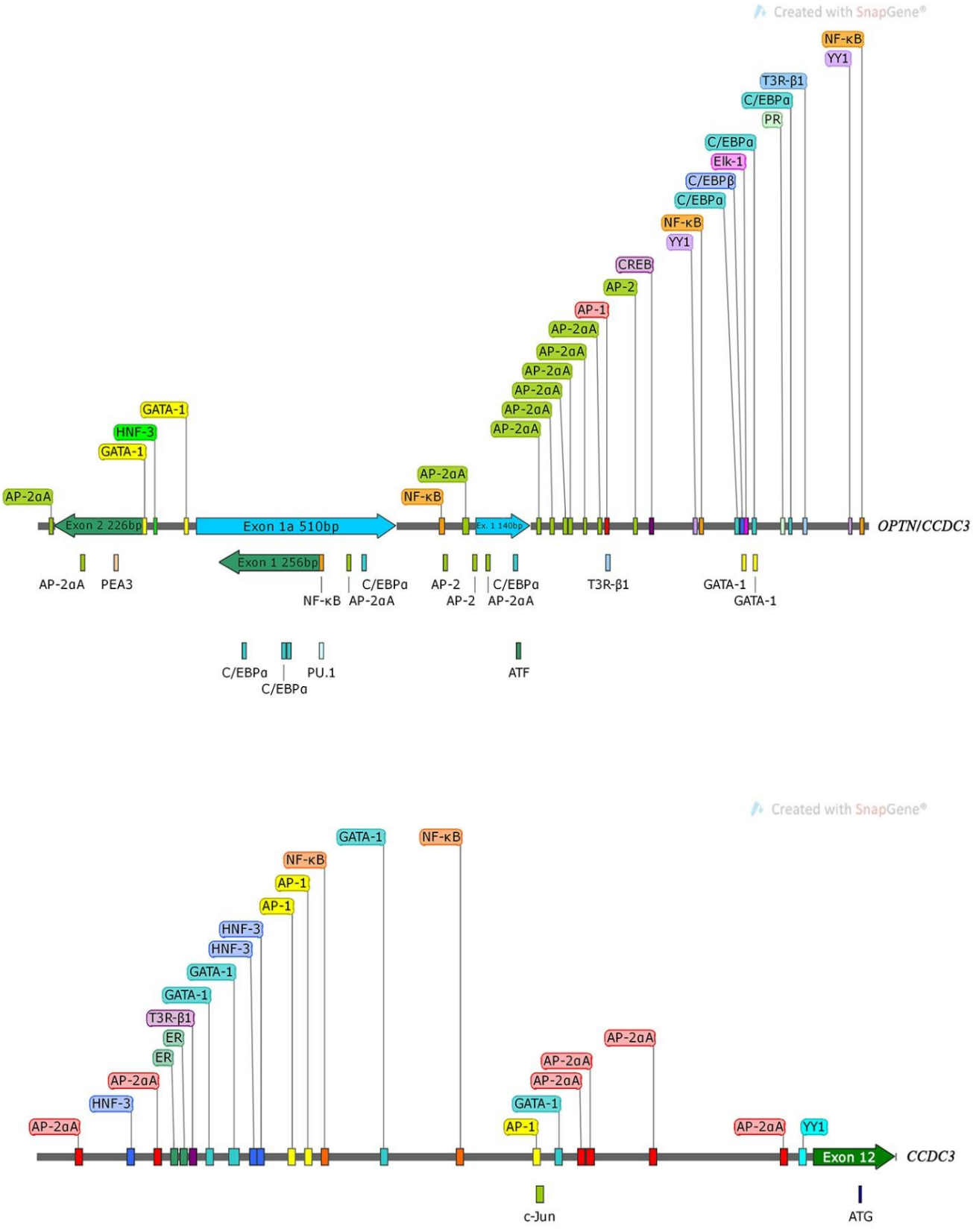

Fig. 4.6: Map of putative TFBSs in *OPTN*/*CCDC3* promoter (above) and in *CCDC*3 alternative promoter (below). *OPTN* and *CCDC3* exons are depicted in blue and in green, respectively. Map built based on the results obtained with AliBaba software.

Hepatocyte nuclear factors 3 (HNF-3) is a subfamily of HNF, mainly expressed in the liver and playing a role in the development of hepatic cells [46].

Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) is a protein that controls DNA transcription and the production of cytokine; NF-κB is almost ubiquitous, it can be found in most of the tissues and it plays a role in cellular responses to stress stimuli, pathogens or adverse environmental conditions, such as presence of free radicals, heavy metals, ultraviolet irradiation, and oxidized low-density lipoproteins [47]; given its role into inflammatory response, NF-κB has been targeted for therapies against cancer and inflammatory diseases [48] [49]; furthermore, the deletion of intermediates of NF-κB pathway resulted with aberrant skeletal development, as NF-κB signaling affects RANK ligand-induced osteoclastogenesis [50], and mediates osteoblast differentiation and bone formation [51].

Polyomavirus enhancer activator 3 (PEA3) is another subfamily of the previously described ETS TFs, involved in organogenesis in mammals [52].

Progesterone receptors (PRs) are a family of ligand-activated TFs, part of the SR (steroid hormone receptor) subfamily of nuclear receptors; two isoforms (PR-A and PR-B) are obtained from the same gene; they can regulate the same or different (isoform-specific) target genes and show both ligand-dependent and independent activities; PRs are involved in the development of breast cancer [53].

The protein PU.1 is expressed in humans by the gene SPI1, and belongs to the ETS family of TFs; PU.1 binds to a purine-rich sequence of a promoter (PU-box), and regulates the alternative splicing of the target genes in synergy with other TFs [54].

Specific protein 1 (Sp1) is part of the Sp/KLF transcription factor family, and it is involved in several cellular functions by interacting with other TFs [55] [56] [57] [58].

Thyroid hormone receptor β1 (T3R-β1) is an isoform of T3R, a family of ligand-inducible, hormone-regulated transcription factors; β-1 isoform, together with α-1 and β-0, repress target gene expression in absence of thyroid hormone [59].

Yin yang 1 (YY1) belongs to the GLI-Kruppel family of zinc-finger TFs; it regulates several genes during cell growth, it is required for the correct development of mammalian embryos, and it is regulated by acetylation (by p300 and P300/CBP-associated factor) and deacetylation (by histone deacetylases) [60]; it has been proved to have oncogenic potential [61], and plays a role in intellectual disability syndrome [62].

The transcription factor c-Jun, together with c-Fos, dimerizes constituting the TF AP-1 (activator protein 1); c-Jun was the first oncogenic TF discovered [63]; it appears overexpressed in cancer, suggesting it can be targeted during cancer therapy [64] [65].

Estrogen receptors (ERs), as PRs, are hormone-ligating receptors, that once activated are able to enter the nucleus and bind gene promoters to work as TF [66]; ER class is formed by ERα, found in in endometrium, breast cancer cells, ovarian stromal cells, efferent ducts and hypothalamus [67], and ERβ, in ovarian granulosa cells, kidney, brain, prostate, endothelial cells, heart [68], and bone [69].

By analysing our results (Table 4.4 and 4.5), and comparing the putative TFBSs available in the two promoters, we deducted that AP-1, AP-2αA, GATA-1, HNF-3, NF-κB, T3R-β1 and YY1 TFs have putative binding sites in both *CCDC3*/*OPTN* core promoter and *CCDC3* alternative promoter; on the other hand, ATF, C/EBPα, C/EBPβ, CREB, Elk-1, PEA3, PR, PU.1 and SP1 are putatively binding only in the *OPTN*/*CCDC3* promoter region, while just c-Jun and ER appear to have a putative TFBSs exclusively in the *CCDC*3 alternative promoter region.

Furthermore, AP-1, AP-2, C/EBPβ, ERβ and NF-κB are playing a role in skeletal development and in diseases related to the bone tissue [71] [72] [38] [69] [50] [51].

4.2 Luciferase Assay results

4.2.1 *OPTN*/*CCDC3* promoter constructs show different luciferase activity

We transfected the gene reporter pGL3 constructs containing the DNA fragments of our interest (Figure 4.7) into HEK 293 cells, and measured their luciferase activity, proportional to the amount of luciferase protein produced. We designed our constructs trying to cover different parts of the promoter region, in order to understand which portion of the vast sequence can have a stronger promoter activity and therefore a higher impact in the genes expression.

Constructs F9R12-Luc and Luc-F9R12 are characterized by the same nucleotide sequence but inserted in the opposite orientation, one for *OPTN* and the other for *CCDC3*; they both are 1252 bp long, and embrace a bigger potential promoting sequence.

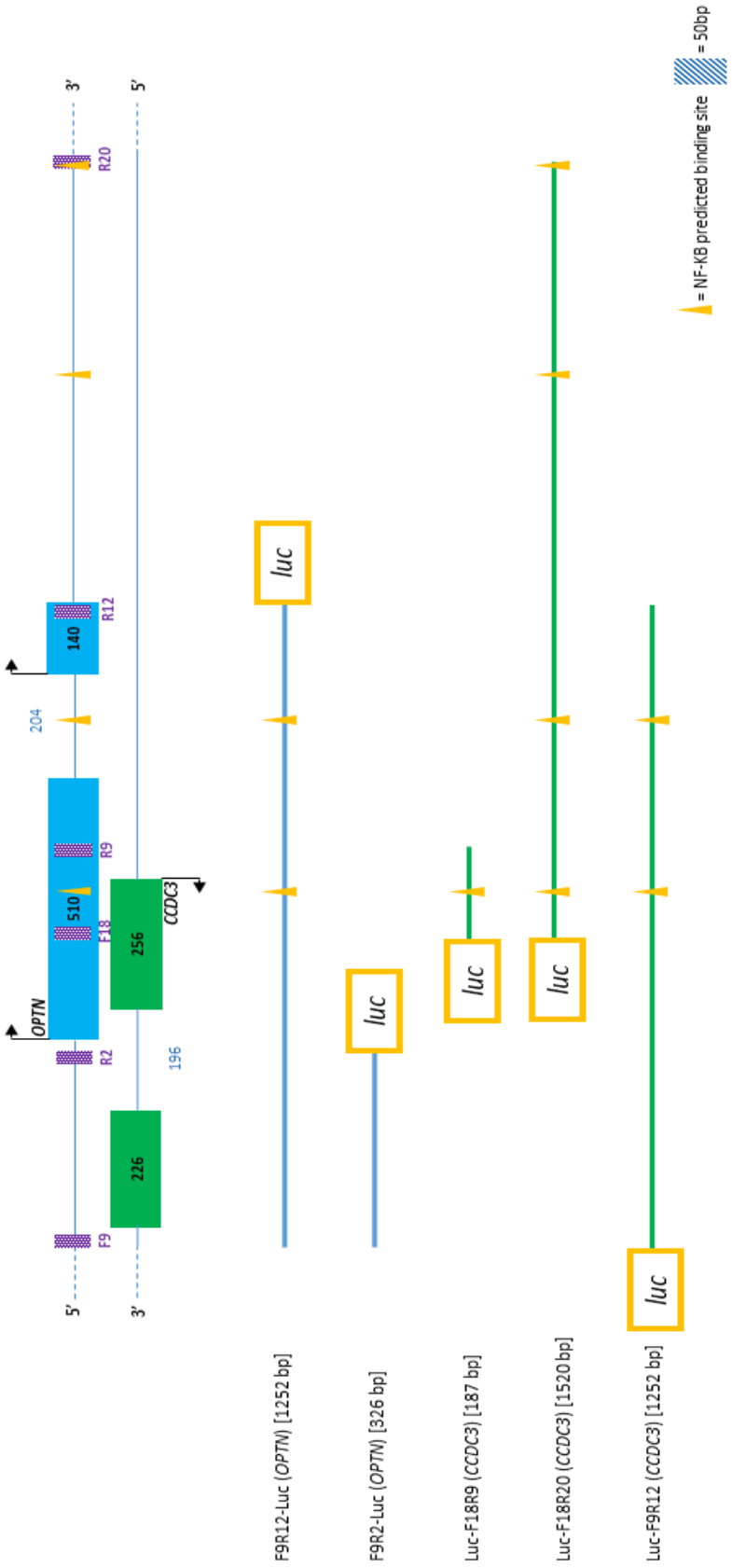

Fig. 4.7: *OPTN*/*CCDC3* promoter constructs. Graphical representation of the pGL3 constructs created from the *OPTN*/*CCDC3* promoter. Emphasis put on construct size, location, and *in silico* predicted NF-κB putative binding sites.

The luciferase activity of all of the *OPTN*/*CCDC3* promoter constructs show statistical significance when compared to negative control, which demonstrates that all of the constructs created contain functional promoter sequences (Figure 4.8).

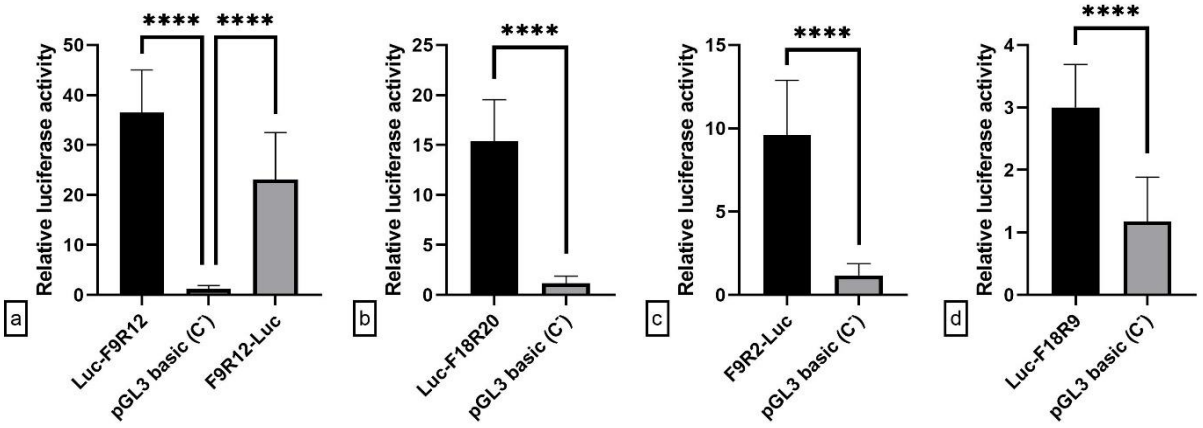

Fig. 4.8: Luciferase activity of *OPTN*/*CCDC3* promoter pGL3 constructs show statistical significance when compared to negative control. a. Comparison of luciferase activity between Luc-F9R12, F9R12-Luc and negative control. b. Comparison of luciferase activity between Luc-F18R20 and negative control. c. Comparison of luciferase activity between F9R2-Luc and negative control. d. Comparison of luciferase activity between Luc-F18R9 and negative control. **** indicates a P value ≤ 0.0001 (one-way ANOVA with Turkey's multiple comparisons test).

Constructs F9R12-Luc and Luc-F9R12, with 1252 bp, respectively in *OPTN* and *CCDC3* direction, showed comparable luciferase activity results, suggesting that the same region is functional in both directions (Figure 4.9). The results exhibited that both F9R12-Luc and Luc-F9R12 constructs were sufficient for the expression of the firefly luciferase gene regardless of their orientation. On the whole, the results demonstrated that the regulation of the *OPTN* and *CCDC3* gene expression could be coordinated through a bidirectional promoter (Figure 4.9).

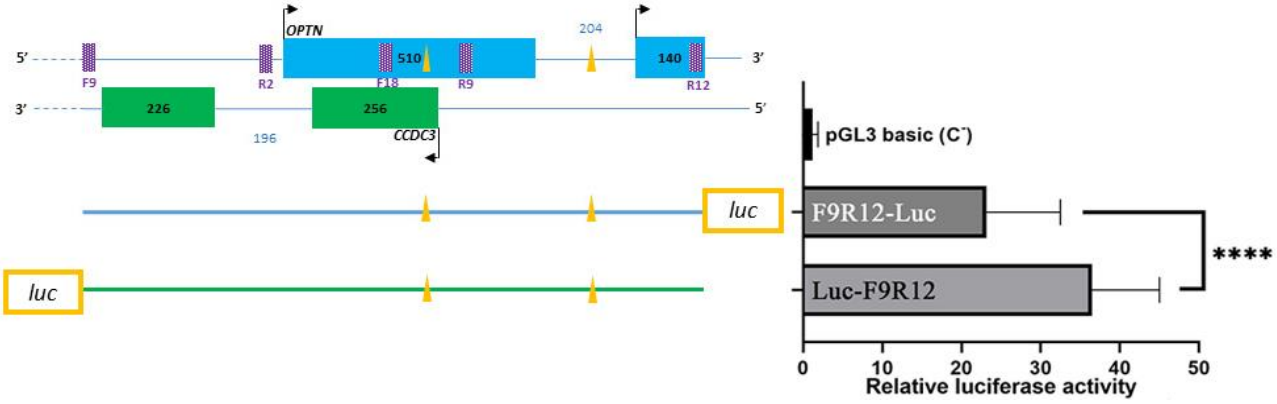

Fig. 4.9: Luciferase activity comparison of F9R12-Luc and Luc-F9R12 constructs. Both the constructs were sufficient for the expression of the firefly luciferase gene regardless of their orientation. **** indicates a P value ≤ 0.0001 (one-way ANOVA with Turkey's multiple comparisons test).

In order to assess the presence of positive or negative regulatory elements within each considered sequence, luciferase activity was performed in stepwise deletion mutants.

By comparing the constructs in the *OPTN* direction, we noticed that F9R12-Luc (size: 1252 bp) exhibited a much higher luciferase activity when compared to the activity registered by F9R2-Luc (326 bp) (see Figure 4.10). These results can be explained by the existence of elements within F9R12-Luc sequence that might strengthen the functionality of the promoter region, suggesting that positive regulators may bind in the deleted region (sequence between R2 and R12 in Figure 10).

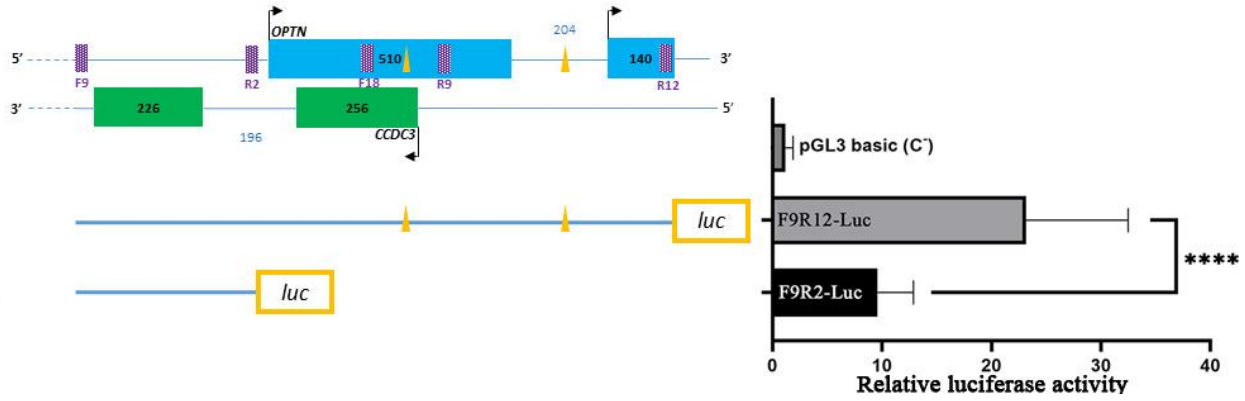

Fig. 4.10: Luciferase activity comparison of F9R12-Luc and F9R2-Luc constructs. For the constructs in the direction of *OPTN* gene, F9R12-Luc shows a much higher luciferase activity when compared to the shorter construct F9R2-Luc, possibly because positive regulators bind in the sequence between R2 and R12. **** indicates a P value ≤ 0.0001 (one-way ANOVA with Turkey's multiple comparisons test).

When analysing the constructs in the *CCDC3* direction, Luc-F9R12 (1252 bp) showed the highest luciferase activity, significantly higher when compared to the luciferase activity registered by the longer construct Luc-F18R20 (1520 bp) (see Figure 4.11). These results can be explained either by the fact that Luc-F9R12 sequence contains binding sites for positive regulators (sequence between F9 and F18) able to upregulate the promoter activity, lacking in Luc-F18R20, or that part of the sequence of Luc-F18R20 (between R12 and R20) features binding sites for negative regulators that downregulate the promoter activity, missing in Luc-F9R12.

In order to clarify the situation, a new fragment covering the region between F18 and R12 (Luc-F18R12), should be created and its luciferase activity assessed to be compared with the activity of the studied fragments.

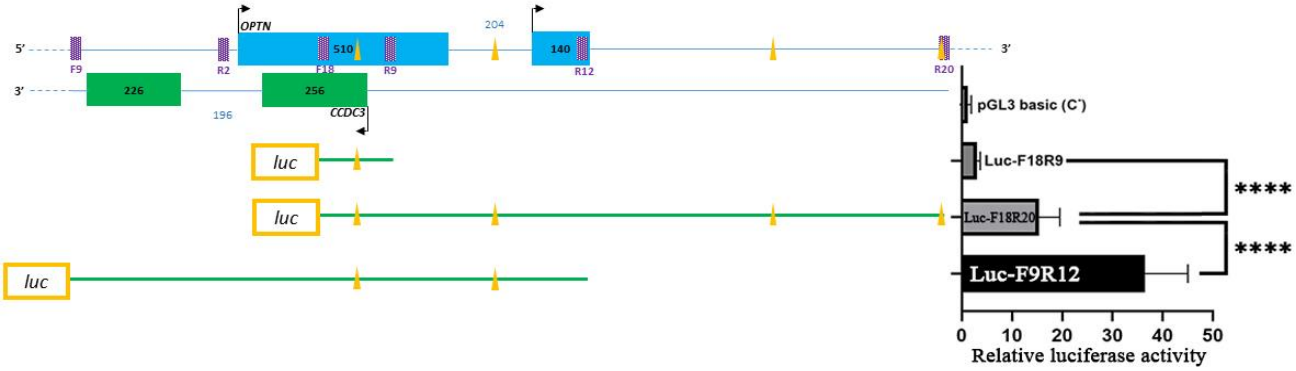

Fig. 4.11: Luciferase activity comparison of Luc-F18R9, Luc-F18R20 and Luc-F9R12 constructs. Luc-F9R12 shows the highest luciferase activity for the constructs in *CCDC3* direction. **** indicates a P value ≤ 0.0001 (one-way ANOVA with Turkey's multiple comparisons test).

By comparing the constructs Luc-F18R9 and Luc-F18R20, we noticed that the luciferase activity of the first sequence is much inferior than the second. Apparently, the deleted sequence belonging to Luc-F18R20 (between R9 and R20) contains binding sites for positive regulators, as stated previously. Moreover, the sequence Luc-F18R9, despite being the shortest with 187 bp, still features a significantly increased luciferase activity when compared to the negative control (pGL3 empty vector) (Figure 4.8d), proving that it is part of a fully functional promoter region.

4.2.2 NF-κB positively regulates *OPTN*/*CCDC3* promoter activity

We co-transfected HEK 293 cells with pGL3 vectors containing *OPTN*/*CCDC3* promoter sequences and NF-κB isoforms p50 and p65 expression vectors, in order to check in which way the transcription factor NF-κB could affect the luciferase activity of the promoter constructs.

Construct F9R12-Luc features two predicted binding sites for NF-κB, and its luciferase activity is clearly affected in the presence of the protein expression vectors (Figure 4.12). NF-κB has been previously proved to effectively bind in this region and act as a regulator of the *OPTN* gene expression, by Sudhakar et al. (2009) [17].

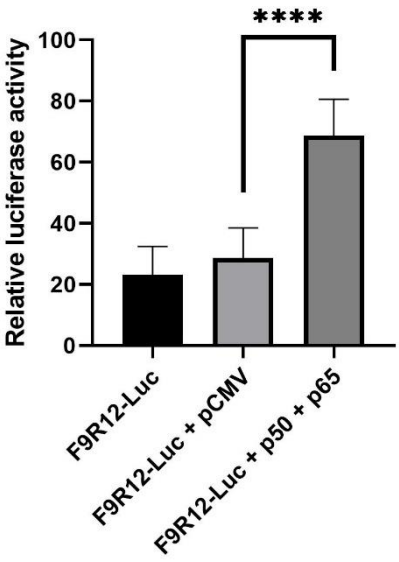

Fig. 4.12: Luciferase activity of construct F9R12-Luc in the presence of NF-κB. Increased luciferase activity registered when co-transfected with p50 and p65 NF-κB isoforms. **** indicates a P value ≤ 0.0001 (one-way ANOVA with Turkey's multiple comparisons test).

Construct F9R2-Luc, on the other hand, when co-transfected with NF-κB, did not show any significant variation in the luciferase activity compared with the empty expression vector; therefore, NF-κB does not seem to be involved in the regulation of this region of the promoter; the results can be explained with the lack of a predicted NF-κB binding site inside its relatively shorter sequence (Figure 4.13).

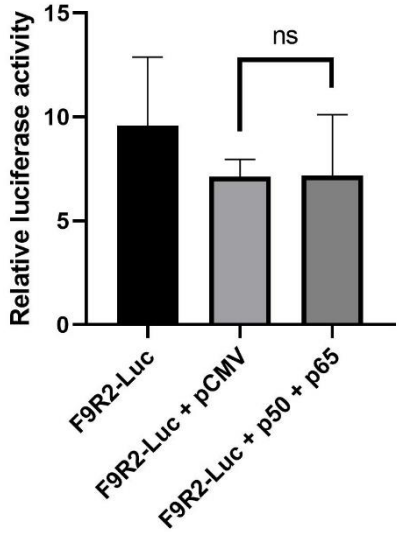

Fig. 4.13: Luciferase activity of construct F9R2-Luc in the presence of NF-κB. No luciferase activity variation registered when co-transfected with p50 and p65 NF-κB isoforms. ns indicates a non-significative P value > 0.05 (one-way ANOVA with Turkey's multiple comparisons test).

When comparing the luciferase activity results of the other co-transfected vectors (Figure 4.14), we observed that constructs Luc-F9R12, Luc-F18R9 and Luc-F18R20 showed significantly increased luciferase activity when co-transfected with NF-κB expression vectors, suggesting that NF-κB plays a role as a transcriptional regulator in these regions by potentially binding to one (or more) of the TFBSs predicted for NF-κB in these regions (see Figure 4.7).

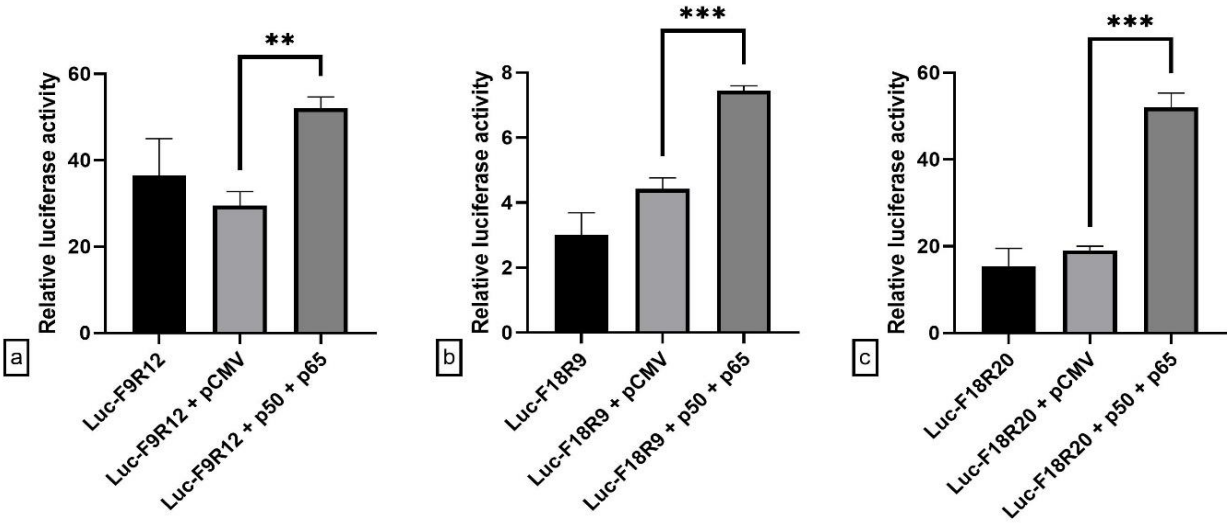

Fig. 4.14: Luciferase activity of constructs Luc-F9R12, Luc-F18R9 and Luc-F18R20 in presence of NF-κB. All of the constructs (a, b and c) showed an increased luciferase activity when co-transfected with p50 and p65 NF-κB isoforms. ** indicates a $P \leq 0.01$, and *** a $P \leq 0.001$ (one-way ANOVA with Turkey's multiple comparisons test).

4.2.3 *CCDC3* alternative promoter shows a high luciferase activity

CCDC3 AP1-Luc sequence (Figure 4.15) functions as a very strong promoter element due to its very high level of luciferase activity shown during the luminescence assay (Figure 4.16).

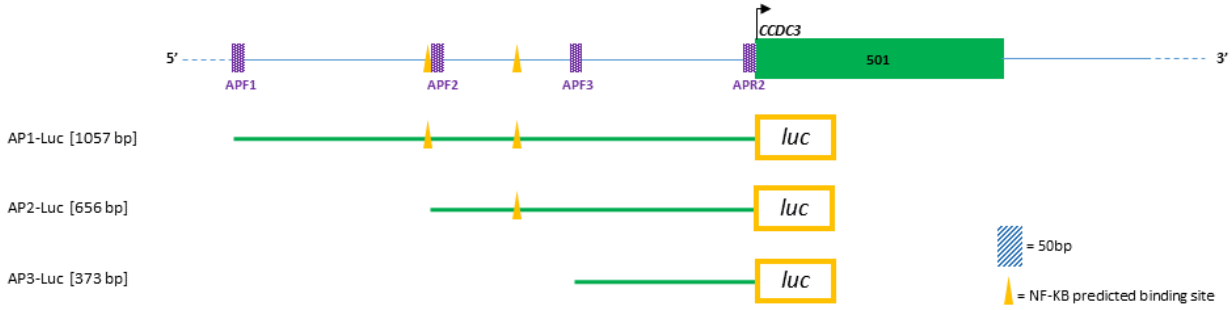

Fig. 4.15: Map of *CCDC3* alternative promoter pGL3 constructs. Graphical representation of the constructs created from *CCDC3* alternative promoter. Emphasis put on *in silico* predicted NF-κB putative binding sites. Primers used to build the constructs are depicted in purple.

The experimental results validated our initial hypothesis on the existence of an alternative promoter for the gene *CCDC3* (Figure 4.16).

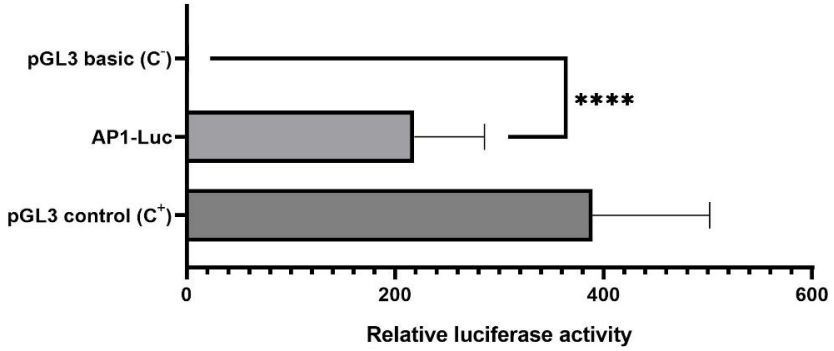

Fig. 4.16: Luciferase activity of AP1-Luc compared with pGL3 basic (C^-) and pGL3 control (C^+) . AP1-Luc showed a very high luciferase activity. **** indicates a P value ≤ 0.0001 (one-way ANOVA with Turkey's multiple comparisons test).

Shorter pGL3 constructs of *CCDC3* alternative promoter, named AP2 (656bp) and AP3 (373bp), are in process to be created by using a different set of primers (Figure 4.15), in order to assess which part of the *CCDC3* alternative promoter is in fact essential for its functionality.

4.2.4 NF-κB does not affect *CCDC3* alternative promoter activity

Despite acting as a very strong promoter element, *CCDC3* alternative promoter (AP1) sequence did not significantly vary the luciferase activity when co-transfected with NF-κB (p50 and p65) expression vectors, indicating that the predicted binding sites for NF-κB in this region are not functional, or that a possible molecular mechanisms exists that precludes the interaction TF-DNA (Figure 4.17).

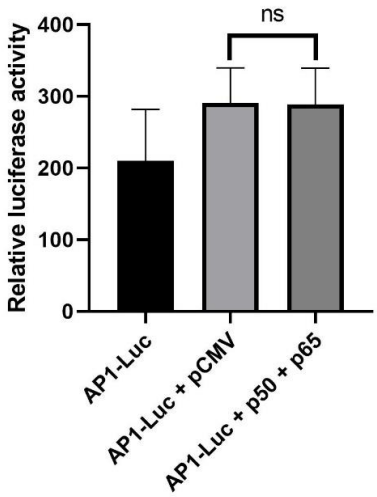

Fig. 4.17: Luciferase activity of construct AP1-Luc in presence of NF-κB. No luciferase activity variation registered when co-transfected with p50 and p65 NF-κB isoforms. ns indicates a non-significative P value > 0.05 (one-way ANOVA with Turkey's multiple comparisons test).

Note how there is no significant difference when comparing the luciferase activity of AP1-Luc construct co-transfected with pCMV (empty expression vector) and with p50 plus p65 (NF-κB expression vectors).

V. Conclusions and future perspectives

The results of the analysis on the transcriptional activity of *OPTN* and *CCDC3* human genes gave us to the opportunity to have a wider picture on the regulation of the human genes *OPTN* and *CCDC3* at a transcriptional level. The luciferase gene-reporter assay gave us an answer on the role of the analyzed sequences identified as promoter elements.

Essential to this work was the initial contribution of the *in silico* analysis, performed through bioinformatics tools; it gave us important early results from which we could draw an overall picture of the general structure of the genes; this supplied us of a solid starting point to propose the existence, later confirmed by gene reporter luciferase assay, of a *CCDC3* alternative promoter, and to develop further theories on the transcriptional regulation of the gene; *in silico* early results allowed us to choose the sequences that would be later cloned into the reporter vector, in a much more elaborated and time-costing process, finalized with the luciferase assay. Thus, in our opinion, bioinformatics and, in general, early *in silico* analysis represent a very powerful tool to be employed during the initial phases of genes analysis, especially when the available information in databases or online gene browsers on a specific gene are scarce, as it was for *CCDC3*.

The analysis work is still on-going, and not fully completed. More promoter-containing constructs are in process to be created and analysed, and consequentially more sequences to be investigated; as an example, shorter versions of the analysed *CCDC3* alternative promoter (AP) construct, previously mentioned as AP2 and AP3 in this work, are still in progress to be built, specifically to be sub-cloned in empty pGL3 vectors, in order to be later transfected into cultured cells and to evaluate their promoter activity through the gene-reporter luciferase assay.

The quantification of luminescence given by such shorter AP constructs, respectively 656 bp (AP2) and 373 bp long (AP3) (see Figure 4.12), will help us to define which portion of the original 1 kb-long *CCDC3* alternative promoter sequence is actually responsible for the strong promoter activity registered.

Furthermore, the action of different TFs, or their combined activity with one or more TF expressing vectors in co-transfected cells, are yet to be tested with the already existing constructs, aiming to unveiling further existing mechanisms involved in the complex protein-DNA regulatory interaction; it would be interesting to test more combinations of TFs and promoter constructs to see how the TF-DNA interaction can vary in presence of different binding sequence motifs, and how realistic *in silico* putatively predicted TFBS are in foreseeing the activity of a TF in regulating the expression of the gene.

Bibliography

1. Albagha, O. M. E., Visconti, M. R., Alonso, N., Langston, A. L., Cundy, T., Dargie, R., Ralston, S. H. (2010). Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. Nature Genetics, 42(6), 520–524. https://doi.org/10.1038/ng.562

2. Lucas, G. J., Riches, P. L., Hocking, L. J., Cundy, T., Nicholson, G. C., Walsh, J. P., & Ralston, S. H. (2007). Identification of a Major Locus for Paget's Disease on Chromosome 10p13 in Families of British Descent. Journal of Bone and Mineral Research, 23(1), 58–63. https://doi.org/10.1359/jbmr.071004

3. Hocking, L. J., Herbert, C. A., Nicholls, R. K., Williams, F., Bennett, S. T., Cundy, T., Ralston, S. H. (2001). Genome wide Search in Familial Paget Disease of Bone Shows Evidence of Genetic Heterogeneity with Candidate Loci on Chromosomes 2q36, 10p13, and 5q35. The American Journal of Human Genetics, 69(5), 1055–1061. https://doi.org/10.1086/323798

4. Ralston, S. H., Langston, A. L., & Reid, I. R. (2008). Pathogenesis and management of Paget's disease of bone. The Lancet, 372(9633), 155-163.

5. Dell'Atti, C., Cassar-Pullicino, V. N., Lalam, R. K., Tins, B. J., & Tyrrell, P. N. M. (2007). The spine in Paget's disease. Skeletal Radiology, 36(7), 609–626. https://doi.org/10.1007/s00256-006-0270-6

6. Van Staa T. P., Selby P. , Leufkens H. G. , Lyles K. , Sprafka J. M., Cooper C. (2002). Incidence and natural history of Paget's disease of bone in England and Wales. J Bone Miner Res 2002; 17:465–71.

7. Siris, E. S. (1994). Epidemiological aspects of Paget's disease: Family history and relationship to other medical conditions. Seminars in Arthritis and Rheumatism, 23(4), 222–225. https://doi.org/10.1016/0049-0172(94)90037-x

8. Morales-Piga, A. A., Rey-Rey, J. S., Corres-González, J., García-Sagredo, J. M., & López-Abente, G. (2009). Frequency and characteristics of familial aggregation of paget's disease of bone. Journal of Bone and Mineral Research, 10(4), 663–670. https://doi.org/10.1002/jbmr.5650100421

9. Michou, L., Conceição, N., Morissette, J., Gagnon, E., Miltenberger-Miltenyi, G., Siris, E. S., Cancela, M. L. (2012). Genetic association study of UCMA/GRP and OPTN genes (*PDB6* locus) with Paget's disease of bone. Bone, 51(4), 720–728. https://doi.org/10.1016/j.bone.2012.06.028

10. Rezaie, T. (2002). Adult-Onset Primary Open-Angle Glaucoma Caused by Mutations in Optineurin. Science, 295(5557), 1077–1079. https://doi.org/10.1126/science.1066901

11. Slowicka, K., Vereecke, L., & van Loo, G. (2016). Cellular Functions of Optineurin in Health and Disease. Trends in Immunology, 37(9), 621–633. https://doi.org/10.1016/j.it.2016.07.002

12. Osawa, T., Mizuno, Y., Fujita, Y., Takatama, M., Nakazato, Y., & Okamoto, K. (2011). Optineurin in neurodegenerative diseases. Neuropathology, 31(6), 569–574. https://doi.org/10.1111/j.1440-1789.2011.01199.x

13. Parra, M. K., Tan, J. S., Mohandas, N., & Conboy, J. G. (2007). Intrasplicing coordinates alternative first exons with alternative splicing in the protein 4.1R gene. The EMBO Journal, 27(1), 122–131. https://doi.org/10.1038/sj.emboj.7601957

14. Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science, 352(6292), 1413–1416. https://doi.org/10.1126/science.aad9868

15. Azad, A. K., Chakrabarti, S., Xu, Z., Davidge, S. T., & Fu, Y. (2014). Coiled-coil domain containing 3 (CCDC3) represses tumor necrosis factor-α/nuclear factor κB-induced endothelial inflammation. Cellular Signalling, 26(12), 2793–2800. https://doi.org/10.1016/j.cellsig.2014.08.025

16. Kobayashi, S., Fukuhara, A., Taguchi, T., Matsuda, M., Tochino, Y., Otsuki, M., & Shimomura, I. (2010). Identification of a new secretory factor, CCDC3/Favine, in adipocytes and endothelial cells. Biochemical and Biophysical Research Communications, 392(1), 29–35. https://doi.org/10.1016/j.bbrc.2009.12.142

17. Sudhakar, C., Nagabhushana, A., Jain, N., & Swarup, G. (2009). NF-κB Mediates Tumor Necrosis Factor α-Induced Expression of Optineurin, a Negative Regulator of NF-κB. PLoS ONE, 4(4), e5114. https://doi.org/10.1371/journal.pone.0005114

18. Fukuda, Y., Washio, T., & Tomita, M. (1999). Comparative study of overlapping genes in the genomes of Mycoplasma genitalium and Mycoplasma pneumoniae. Nucleic acids research, 27(8), 1847-53.

19. Nakayama, T., Asai, S., Takahashi, Y., Maekawa, O., & Kasama, Y. (2007). Overlapping of genes in the human genome. International journal of biomedical science: IJBS, 3(1), 14-9.

20. Sanna, C. R., Li, W.-H., & Zhang, L. (2008). Overlapping genes in the human and mouse genomes. BMC Genomics, 9(1), 169. https://doi.org/10.1186/1471-2164-9-169

21. Kapustin, Y., Souvorov, A., Tatusova, T., & Lipman, D. (2008). Splign: algorithms for computing spliced alignments with identification of paralogs. Biology Direct, 3(1), 20. https://doi.org/10.1186/1745-6150-3-20

22. Farré, D., Roset, R., Huerta, M., Adsuara, J., Roselló, L., Albà, M.M., Messeguer, X. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res, 31, 13, 3651-3653.

23. Ma, T. S. (1995). Applications and Limitations of Polymerase Chain Reaction Amplification. Chest, 108(5), 1393–1404. https://doi.org/10.1378/chest.108.5.1393

24. Chan, W., Verma, C. S., Lane, D. P., & Gan, S. K. (2013). A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. Bioscience Reports, 33(6), 931–937. https://doi.org/10.1042/bsr20130098

25. Sambrook, J. et al. (2000). Molecular Cloning: A Laboratory Manual (third edition). Cold Spring Harbor Laboratory Press. http://www.molecularcloning.com/

26. GT – AG rule (Chambon's rule). Encyclopedia of Genetics, Genomics, Proteomics and Informatics (pp. 828–828). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6754-9_7198

27. Szafranski, K., Schindler, S., Taudien, S., Hiller, M., Huse, K., Jahn, N., Schreiber, S., Backofen, R., Platzer, M. (2007). Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. Genome Biology, 8:R154. https://doi.org/10.1186/gb-2007-8-8-r154

28. Trinklein, N. D. (2003). Identification and Functional Analysis of Human Transcriptional Promoters. Genome Research, 13(2), 308–312. https://doi.org/10.1101/gr.794803

29. Zhang, T. (2003). Multiple Variable First Exons: A Mechanism for Cell- and Tissue-Specific Gene Regulation. Genome Research, 14(1), 79–89. https://doi.org/10.1101/gr.1225204

30. Carnici, P., Kasukawa, T., Katayama, S., et al. (2005). The Transcriptional Landscape of the Mammalian Genome. Science, 309(5740), 1559–1563. https://doi.org/10.1126/science.1112014

31. Conti, E., & Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Current Opinion in Cell Biology, 17(3), 316–325. https://doi.org/10.1016/j.ceb.2005.04.005

32. Chang, Y.-F., Imam, J. S., & Wilkinson, M. F. (2007). The Nonsense-Mediated Decay RNA Surveillance Pathway. Annual Review of Biochemistry, 76(1), 51–74. https://doi.org/10.1146/annurev.biochem.76.050106.093909

33. Karin, M., Liu, Z., & Zandi, E. (1997). AP-1 function and regulation. Current Opinion in Cell Biology, 9(2), 240–246. https://doi.org/10.1016/s0955-0674(97)80068-3

34. Hess, J. (2004). AP-1 subunits: quarrel and harmony among siblings. Journal of Cell Science, 117(25), 5965–5973. https://doi.org/10.1242/jcs.01589

35. Ameyar, M., Wisniewska, M., & Weitzman, J. B. (2003). A role for AP-1 in apoptosis: the case for and against. Biochimie, 85(8), 747–752. https://doi.org/10.1016/j.biochi.2003.09.006

36. Huang, W., Chen, C., Liang, Z., Qiu, J., Li, X., Hu, X. Zhang, J., et al. (2016). AP-2α inhibits hepatocellular carcinoma cell growth and migration. International Journal of Oncology, 48(3), 1125–1134. https://doi.org/10.3892/ijo.2016.3318

37. Miller, M., Shuman, J. D., Sebastian, T., Dauter, Z., & Johnson, P. F. (2003). Structural Basis for DNA Recognition by the Basic Region Leucine Zipper Transcription Factor CCAAT/Enhancer-binding Protein α. Journal of Biological Chemistry, 278(17), 15178–15184. https://doi.org/10.1074/jbc.m300417200

38. Smink, J. J., & Leutz, A. (2009). Rapamycin and the transcription factor C/EBPβ as a switch in osteoclast differentiation: implications for lytic bone diseases. Journal of Molecular Medicine, 88(3), 227–233. https://doi.org/10.1007/s00109-009-0567-8

39. Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A.-S., McNamara, J., White, L. (2008). Neuroscience (4th ed.). Sinauer Associates. pp. 170–6. ISBN 978-0-87893-697-7.

40. Silva, A. J., Kogan, J. H., Frankland, P. W., & Kida, S. (1998). Creb And Memory. Annual Review of Neuroscience, 21(1), 127–148. https://doi.org/10.1146/annurev.neuro.21.1.127

41. Pugazhenthi, S., Wang, M., Pham, S., Sze, C.-I., & Eckman, C. B. (2011). Downregulation of CREB expression in Alzheimer's brain and in Aβ-treated rat hippocampal neurons. Molecular Neurodegeneration, 6(1), 60. https://doi.org/10.1186/1750-1326-6-60

42. Besnard, A., Galan-Rodriguez, B., Vanhoutte, P., & Caboche, J. (2011). Elk-1 a Transcription Factor with Multiple Facets in the Brain. Frontiers in Neuroscience, 5. https://doi.org/10.3389/fnins.2011.00035

43. Chai, Y., Chipitsyna, G., Cui, J., Liao, B., Liu, S., Aysola, K., Rao, V. N., et al. (2001). c-Fos oncogene regulator Elk-1 interacts with BRCA1 splice variants BRCA1a/1b and enhances BRCA1a/1b-mediated growth suppression in breast cancer cells. Oncogene, 20(11), 1357–1367. https://doi.org/10.1038/sj.onc.1204256

44. Katsumura, K. R., DeVilbiss, A. W., Pope, N. J., Johnson, K. D., & Bresnick, E. H. (2013). Transcriptional Mechanisms Underlying Hemoglobin Synthesis. Cold Spring Harbor Perspectives in Medicine, 3(9), a015412–a015412. https://doi.org/10.1101/cshperspect.a015412

45. Katsumura, K. R., & Bresnick, E. H. (2017). The GATA factor revolution in hematology. Blood, 129(15), 2092–2102. https://doi.org/10.1182/blood-2016-09-687871

46. Costa, R. H., Kalinichenko, V. V., Holterman, A.-X. L., & Wang, X. (2003). Transcription factors in liver development, differentiation, and regeneration. Hepatology, 38(6), 1331–1347. https://doi.org/10.1016/j.hep.2003.09.034

47. Gilmore, T. D. (2006). Introduction to NF-κB: players, pathways, perspectives. Oncogene, 25(51), 6680–6684. https://doi.org/10.1038/sj.onc.1209954

48. Garg, A., & Aggarwal, B. (2002). Nuclear transcription factor-κB as a target for cancer drug development. Leukemia, 16(6), 1053–1068. https://doi.org/10.1038/sj.leu.2402482

49. Sethi, G., Sung, B., & Aggarwal, B. B. (2008). Nuclear Factor-κB Activation: From Bench to Bedside. Experimental Biology and Medicine, 233(1), 21–31. https://doi.org/10.3181/0707-mr-196

50. Abu-Amer, Y. (2013). NF-κB signaling and bone resorption. Osteoporosis International, 24(9), 2377–2386. https://doi.org/10.1007/s00198-013-2313-x

51. Yao, Z., Li, Y., Yin, X., Dong, Y., Xing, L., & Boyce, B. F. (2014). NF-κB RelB Negatively Regulates Osteoblast Differentiation and Bone Formation. Journal of Bone and Mineral Research, 29(4), 866–877. https://doi.org/10.1002/jbmr.2108

52. Chen, Q., Huang, S., Zhao, Q., Chen, R., & Zhang, A. (2010). Expression and function of the Ets transcription factor pea3 during formation of zebrafish pronephros. Pediatric Nephrology, 26(3), 391–400. https://doi.org/10.1007/s00467-010-1713-9

53. Daniel, A. R., Hagan, C. R., & Lange, C. A. (2011). Progesterone receptor action: defining a role in breast cancer. Expert Review of Endocrinology & Metabolism, 6(3), 359–369. https://doi.org/10.1586/eem.11.25

54. Hallier, M., Tavitian, A., & Moreau-Gachelin, F. (1996). The Transcription Factor Spi-1/PU.1 Binds RNA and Interferes with the RNA-binding Protein p54. Journal of Biological Chemistry, 271(19), 11177–11181. https://doi.org/10.1074/jbc.271.19.11177

55. Lin, S. Y., Black, A. R., Kostic, D., Pajovic, S., Hoover, C. N., & Azizkhan, J. C. (1996). Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. Molecular and Cellular Biology, 16(4), 1668–1675. https://doi.org/10.1128/mcb.16.4.1668

56. Foti, D., Iuliano, R., Chiefari, E., & Brunetti, A. (2003). A Nucleoprotein Complex Containing Sp1, C/EBPβ, and HMGI-Y Controls Human Insulin Receptor Gene Transcription. Molecular and Cellular Biology, 23(8), 2720–2732. https://doi.org/10.1128/mcb.23.8.2720-2732.2003

57. Botella, L. M. (2002). Transcriptional activation of endoglin and transforming growth factor-beta signaling components by cooperative interaction between Sp1 and KLF6: their potential role in the response to vascular injury. Blood, 100(12), 4001–4010. https://doi.org/10.1182/blood.v100.12.4001

58. Park, S.-Y., Shin, H.-M., & Han, T.-H. (2002). Synergistic interaction of MEF2D and Sp1 in activation of the CD14 promoter. Molecular Immunology, 39(1–2), 25–30. https://doi.org/10.1016/s0161-5890(02)00055-x

59. Yang, Z., Hong, S.-H., & Privalsky, M. L. (1999). Transcriptional Anti-repression. Journal of Biological Chemistry, 274(52), 37131–37138. https://doi.org/10.1074/jbc.274.52.37131

60. Yao, Y.-L., Yang, W.-M., & Seto, E. (2001). Regulation of Transcription Factor YY1 by Acetylation and Deacetylation. Molecular and Cellular Biology, 21(17), 5979–5991. https://doi.org/10.1128/mcb.21.17.5979-5991.2001

61. Thiaville, M. M., & Kim, J. (2011). Oncogenic Potential of Yin Yang 1 Mediated Through Control of Imprinted Genes. Critical Reviews− in Oncogenesis, 16(3–4), 199–209. https://doi.org/10.1615/critrevoncog.v16.i3-4.40

62. Gabriele, M., Vulto-van Silfhout, A. T., Germain, P.-L., Vitriolo, A., Kumar, R., Douglas, E., de Vries, B. B. A., et al. (2017). YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. The American Journal of Human Genetics, 100(6), 907–925. https://doi.org/10.1016/j.ajhg.2017.05.006

63. Vogt, P. K. (2002). Fortuitous convergences: the beginnings of JUN. Nature Reviews Cancer, 2(6), 465–469. https://doi.org/10.1038/nrc818

64. Rotondo, J. C., Borghi, A., Selvatici, R., Mazzoni, E., Bononi, I., Corazza, M., Martini, F., et al. (2018). Association of Retinoic Acid Receptor β Gene With Onset and Progression of Lichen Sclerosus–Associated Vulvar Squamous Cell Carcinoma. JAMA Dermatology, 154(7), 819. https://doi.org/10.1001/jamadermatol.2018.1373

65. Smith, L. M., Wise, S. C., Hendricks, D. T., Sabichi, A. L., Bos, T., Reddy, P., Birrer, M. J., et al. (1999). cJun overexpression in MCF-7 breast cancer cells produces a tumorigenic, invasive and hormone resistant phenotype. Oncogene, 18(44), 6063–6070. https://doi.org/10.1038/sj.onc.1202989

66. Levin, E. R. (2005). Integration of the Extranuclear and Nuclear Actions of Estrogen. Molecular Endocrinology, 19(8), 1951–1959. https://doi.org/10.1210/me.2004-0390

67. Couse, J. F., Lindzey, J., Grandien, K., Gustafsson, J.-A., & Korach, K. S. (1997). Tissue Distribution and Quantitative Analysis of Estrogen Receptor-α (ERα) and Estrogen Receptor-β (ERβ) Messenger Ribonucleic Acid in the Wild-Type and ERα-Knockout Mouse. Endocrinology, 138(11), 4613–4621. https://doi.org/10.1210/endo.138.11.5496

68. Babiker, F. (2002). Estrogenic hormone action in the heart: regulatory network and function. Cardiovascular Research, 53(3), 709–719. https://doi.org/10.1016/s0008-6363(01)00526-0

69. Khalid, A. B., & Krum, S. A. (2016). Estrogen receptors alpha and beta in bone. Bone, 87, 130–135. https://doi.org/10.1016/j.bone.2016.03.016

70. Leppek, K., Das, R., & Barna, M. (2017). Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nature Reviews Molecular Cell Biology, 19(3), 158–174. https://doi.org/10.1038/nrm.2017.103

71. Wagner, E. F. (2002). Functions of AP1 (Fos/Jun) in bone development. Annals of the Rheumatic Diseases, 61(Supplement 2), 40ii–42. https://doi.org/10.1136/ard.61.suppl_2.ii40

72. Yu, S., Yerges-Armstrong, L. M., Chu, Y., Zmuda, J. M., & Zhang, Y. (2015). AP2 suppresses osteoblast differentiation and mineralization through down-regulation of Frizzled-1. Biochemical Journal, 465(3), 395–404. https://doi.org/10.1042/bj20140668

Appendix

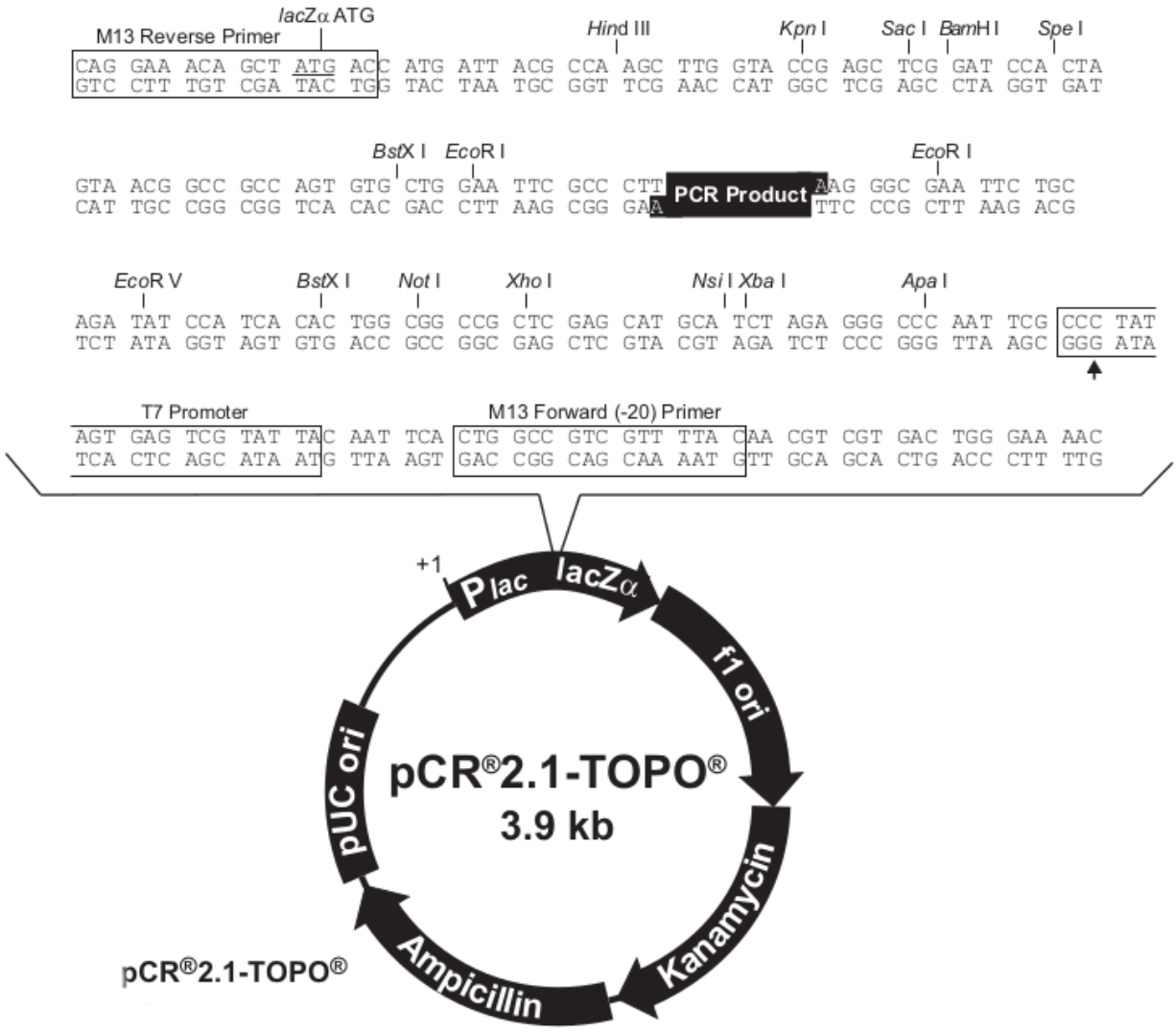

Fig. A1: pCR 2.1-TOPO® plasmid maps. Putting emphasis on the poly-linker site sequence with the restriction enzyme sites, the phagic origin of replication (f1 ori), the genes of antibiotic resistance (Kan^R and Amp^R), the *E. coli* origin of replication (pUC ori), the promoter for the lacZ gene (P lac), and the lacZ gene itself coding for the β-galactosidase enzyme.

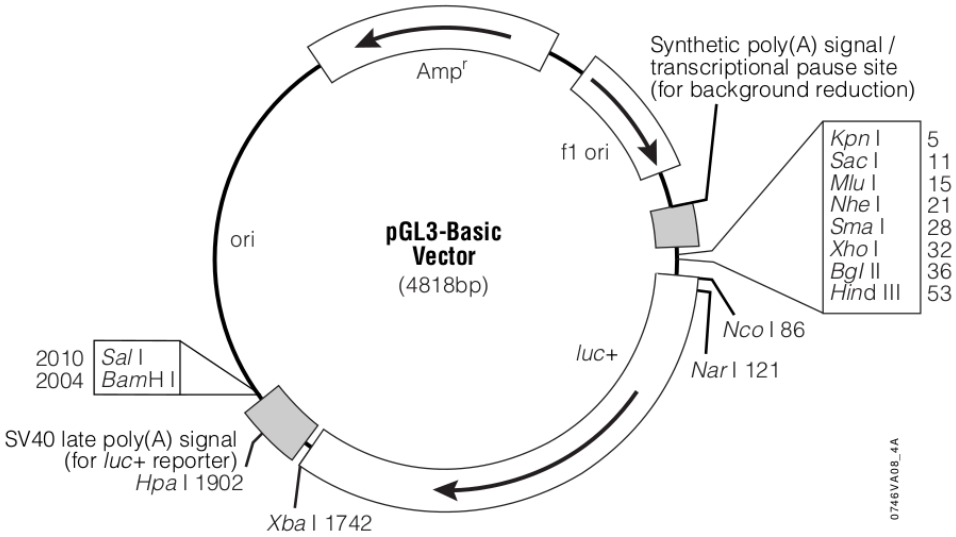

Fig. A2: pGL3-Basic® vector maps. The *luc+* gene encodes for the firefly luciferase protein. *Amp*^{R} provides ampicillin resistance; f1 ori is the phagic origin of replication; ori is the origin of replication in *E. coli*. Note the polylinker site located upstream the *luc+* gene.

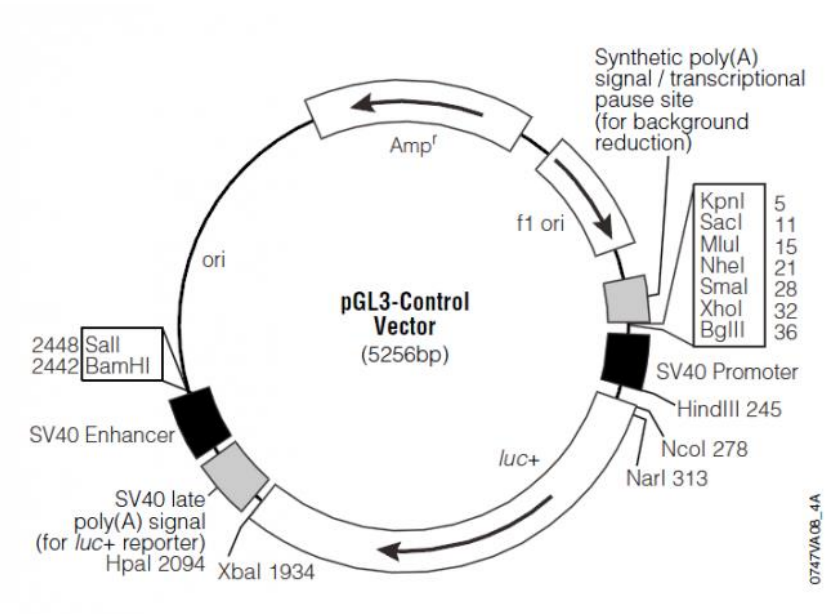

Fig. A3: pGL3-Control® vector maps. The *luc+* gene encodes for the firefly luciferase protein. *Amp*^{R} provides ampicillin resistance. SV40 is a strong viral promoter, conferring the vector the function of positive control.

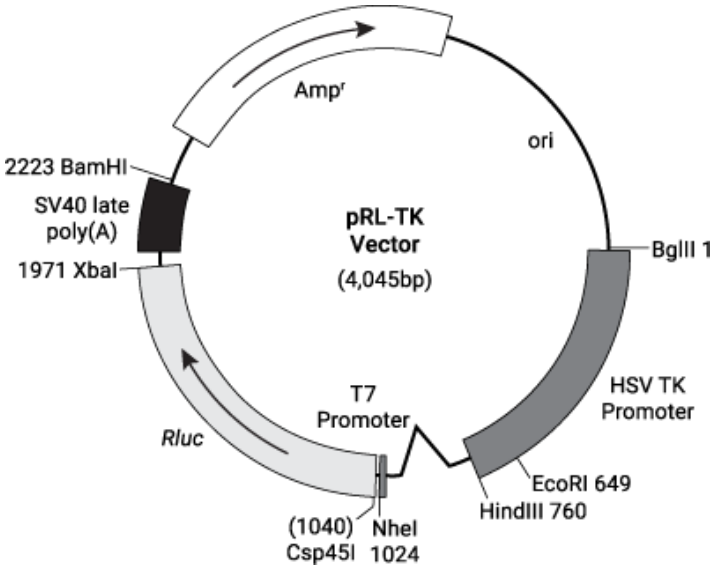

Fig. A4: pRL-TK ® vector maps. The *Rluc* gene encodes for the *Renilla* luciferase protein. *Amp^R* provides ampicillin resistance. HSV TK is a strong viral promoter, granting the expression of this vector.