UNIVERSIDADE DO ALGARVE

UNIVERSIDADE DO ALGARVE

Molecular structure and functional analysis of *runx3* in zebrafish

Brigite Sandra Nunes Simões Rodrigues

Tese para a obtenção do grau de Doutor em Ciências Biomédicas

Trabalho efectuado sob a orientação de:

Professora Doutora M. Leonor Cancela

Professor Doutor Robert Kelsh

Doutora Natércia Conceição

2015

UNIVERSIDADE DO ALGARVE

Molecular structure and functional analysis of *runx3* in zebrafish

Brigite Sandra Nunes Simões Rodrigues

PhD fellowship SFRH/BD/38083/2007

Tese para a obtenção do grau de Doutor em Ciências Biomédicas

Trabalho efectuado sob a orientação de:

Professora Doutora M. Leonor Cancela

Professor Doutor Robert Kelsh

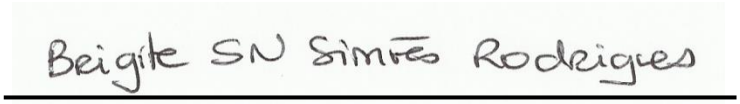
Doutora Natércia Conceição

2015

Molecular structure and functional analysis of *runx3* in zebrafish

Declaração de autoria de trabalho

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.



(Brigite Sandra Nunes Simões Rodrigues)

“Tudo o que um sonho precisa para ser realizado é alguém que acredite que ele possa ser realizado.”

Roberto Shinyashiki

Ao bem mais precioso, a minha família e amigos.

Agradecimentos/Acknowledgments

I would like to send my sincerest gratitude to following people for their contribution to this work in different ways.

First, I want to thank to Professor Leonor Cancela and Professor Robert Kelsh, for accepting me as their PhD student, because without that first "yes" this project could not have been done. Thank you all the support and encouragement and essentially for always believed that I could do it, even when sometimes I didn't believe it myself.

Second, I want to express my profound gratitude to Dr Natércia Conceição for accept to be my co-supervisor. More important than that, any words would never be enough to say THANK YOU for the long long long after dinner hours spending at my "side", giving me the precious push that enable me to finish this project. A big thank you to Sofia and Gonçalo for sharing with me, for so long, their mum's attention.

I would also like to thank all Lab members (present and past, in Portugal and in the UK) that have helped – directly and indirectly – to the execution of this project either with technical support or just with the right words in the right time, I am thankful to them all.

I want to say a special thanks to Cátia Marques, Cindy Fazenda, Íris Silva and Andreia Adrião for being the best bench colleagues ever!!! Also a special thanks to Iris, for her precious help with some figures presented here, to Cátia for her precious words and to Cindy that will always be my "little" girl.

Very special thanks to Marta Rafael, for always listening to my screams and to always been at a distance of a click.

Um muito obrigado à minha família pelo apoio incondicional, em especial ao meu marido e à minha filha que são maravilhosos e por sempre aceitarem da melhor forma possível todos os "nãos" que ouviram no decorrer deste trabalho.

Não menos importante, quero por fim agradecer à Fundação para a Ciência e Tecnologia, pelo financiamento, através da atribuição da bolsa de doutoramento com a referência SFRH/BD/38083/2007, que permitiu a realização deste trabalho.

Abstract

In mammals, runt-related family of transcription factors is encoded by the three distinct genes, RUNX1, RUNX2 and RUNX3 that share an evolutionarily conserved 128 amino acid Runt domain, which is responsible for the dual function of DNA binding and hetero-dimerization with the co-factor CBFβ. Gene ablation and gain of function experiments established all three proteins as key regulators of lineage-specific gene expression in major developmental pathways. Mutations in RUNX genes have been frequently associated with human diseases. Despite many studies to unveil mechanisms of RUNX3 action, available data is insufficient to fully understand its physiological role, particularly the importance of each isoform. We cloned for the first time a variety of transcript variants for both zebrafish *runx3* and *cbfβ* genes, identified their temporal expression by qPCR and localized *runx3* sites of expression by *in situ* hybridization in adult tissues and during early embryonic development. As *runx3*-P1 and *runx3*-P2 transcripts were found to be differentially expressed, we used a promoter *in silico* comparative approach for both P1 and P2 *runx3* gene promoter regions and *in vitro* and *in vivo* promoter analysis to identify regulatory regions and conserved transcription factor binding sites, allowing us to select, from an extensive list of putative transcription binding sites, the best candidates to regulate *runx3* promoter regions. Furthermore, through *in vitro* analysis, we have examined the possible cross- and auto-regulation of *runx3* promoters by Runx2 and Runx3 isoforms, respectively. In conclusion, we have used computational and molecular approaches to improve our understanding of the complexity of Runx and Cbfβ variants and their implication for function, using zebrafish as a model. Altogether, our structural and functional data provide further support to the assumption that the expression of *runx3* variants is tightly regulated, leading to a highly specific spatial-temporal expression pattern.

Keywords: Runx3, alternative promoters, splicing variants, Cbfβ, transcriptional promoter analysis, expression pattern, zebrafish.

Resumo

O termo CBF ("core binding factor") corresponde a um complexo heterodimérico de factores de transcrição, composto por duas subunidades, designadas de subunidade α e β. A subunidade α é caracterizada por possuir uma região de 128 aminoácidos no seu N-terminal, designada de domínio Runt, que se manteve bastante conservada ao longo da evolução. O domínio Runt possui uma dupla função, sendo essencial não só para a ligação da subunidade α ao ADN, que reconhece a sequência específica PyGPyGGT (sendo Py uma pirimidina), mas também para a heterodimerização com a subunidade β. A subunidade β não possui capacidade de ligação ao ADN, mas actua como um co-factor que se liga à subunidade α, sendo essencial para aumentar a sua afinidade na ligação às regiões promotoras dos genes alvo, e também para a regulação do seu "turnover", protegendo o complexo da degradação mediada por proteossomas via proteínas ubiquitinadas.

A subunidade α do complex CBF pode ser qualquer uma das três proteínas da família RUNX. Os genes *RUNX* surgiram muito cedo na evolução e mantiveram um elevado grau de semelhança nos vertebrados. Os metazoários primitivos, tais como o ouriço-do-mar e *C. elegans* parecem conter apenas um gene da família RUNX, enquanto que, até à data, foram já descritos três genes *RUNX* em mamíferos, quatro em insectos, no fugu e no peixe zebra, embora neste último caso seja por possuir dois genes *runx*2, *runx2a* e *runx*2b, para além do *runx1* e *runx3*. Os diversos genes *runx* sugerem a necessidade de uma rigorosa e elaborada regulação da actividade das proteínas RUNX ao especificar eventos complexos de desenvolvimento em organismos superiores. No entanto, a subunidade β, designada CBFβ, parece ser codificada apenas por um gene, excepto no caso dos insectos, em que até à data foram descritos dois genes. Os factores de transcrição RUNX têm sido comprovados por vários estudos como sendo determinantes na regulação de vários processos biológicos, de modo a orquestrar a correcta diferenciação celular durante o desenvolvimento embrionário, e são ainda responsáveis por uma diversidade de patologias.

Todos os membros da família RUNX possuem semelhanças estruturais, no entanto têm sido implicados em actividades biológicas distintas. A existência de uma variedade de transcritos para todos os membros da família RUNX foi anteriormente descrita. Esta diversidade de transcritos deve-se em parte ao facto dos genes *Runx* serem transcritos através de dois

promotores alternativos (P1 e P2), que se encontram separados por vários milhares de pares de bases (pb), mas sendo essa diversidade aumentada pelo grande número de "splicing" alternativo verificado nesses transcritos, assim como a presença de locais de poliadenilação alternativos. Cada gene *Runx* codifica para duas isoformas proteicas principais, que diferem na sua sequência N-terminal, dependendo se são codificadas a partir de transcritos derivados do promotor P1 ou do promotor P2. As proteínas que são geradas a partir de transcritos derivados do promotor P2 contêm uma sequência de cinco aminoácidos específicos no seu N-terminal – o pentapeptídeo M(R/H)IPV. Por sua vez, as proteínas que são geradas a partir de transcritos derivados do promotor P1 contêm uma sequência de 19 aminoácidos específicos no seu N-terminal, originados por "splicing" alternativo do exão 1 para um local the "splicing" conservado no exão 2, localizado 16 pb a montante do codão de iniciação (ATG) do transcrito P2. Estas duas isoformas são geralmente designadas como isoforma RUNX3-MA(S/D)NS e isoforma RUNX3-M(R/H)IPV, dependendo se são traduzidas de transcritos derivados do promotor P1 ou P2, respectivamente. Em termos de estrutura proteica, todos os membros da família RUNX possuem os mesmos domínios e motivos característicos: domínio Runt (RD), domínio de transactivação (TAD), domínio de inibição (ID), sinal de localização nuclear (NLS), motivo PY (ou PPxY) e o motivo VWRPY, que corresponde aos últimos cinco aminoácidos do C-terminal. A proteína RUNX2 apresenta ainda um domínio extra, apenas observado neste membro da família, o domínio QA, que é composto por um fragmento de resíduos Q e A.

Embora as proteínas RUNX sejam tradicionalmente descritas como sendo factores de transcrição, que podem actuar como homodímeros ou heterodímeros na ligação ao promotor dos genes alvo, foram reconhecidas mais recentemente como sendo moléculas multifacetadas que se podem associar com uma extensa variedade de proteínas, dependente do contexto celular. O resultado final da regulação da transcrição de determinado gene alvo parece ser afectado pela interação entre as proteínas RUNX e co-factores específicos de cada ambiente celular, dependendo da aproximação do local de ligação entre os factores ou da sua disponibilidade no núcleo. É evidente que cada uma das três proteínas RUNX tem funções diferentes, mas o facto de por vezes serem co-expressas no mesmo tecido indica que algumas dessas funções poderão actuar sinergisticamente ou corresponder a funções complementares que actuam em diferentes etapas temporais.

Enquanto o RUNX1 tem sido repetidamente comprovado como sendo um dos alvos mais frequentes de alterações genéticas associadas à leucemia, o RUNX2 é descrito como indispensável na formação normal do osso e na diferenciação de osteoblastos, sendo fundamental no desenvolvimento do esqueleto em mamíferos. O RUNX3, por sua vez, é essencial para a regulação de um tipo celular específico de neurónios proprioceptivos $TrkC^{+}$ nos gânglios das raízes nervosas dorsais ("dorsal root ganglia", DRGs) e para a maturação dos condrócitos durante a esqueletogénese. Está ainda envolvido na regulação da proliferação e da sobrevivência de vários tipos celulares, incluindo células epiteliais gástricas, desenvolvimento das células T e diferenciação de células imunes, incluindo as células "natural killer", células dendríticas e células B.

Dos três genes que pertencem à família dos factores de transcrição RUNX, o *RUNX3* é o mais pequeno em termos de sequência nucleotídica e o que apresenta menor número de exões. No entando, a proteína apresenta todos os domínios característicos dos membros desta família. Estudos filogenéticos usando sequências dos vários Runxs de uma diversidade de espécies indicam que a evolução destes genes foi possivelmente originada a partir de um gene *runx3* em invertebrados progredindo para a existência de múltiplos genes em espécies superiores. Embora o RUNX3 esteja predominantemente localizado no núcleo, onde exerce a sua função como factor de transcrição, a sua localização foi também observada no citoplasma em diferentes células cancerígenas, evidenciando a importância da localização celular das proteínas RUNX no exercício da sua função. O *RUNX3* é expresso numa variedade de tecidos, incluindo tecidos moles e cartilagíneos. A análise da expressão em termos de cada variante mostrou que diferentes transcritos do *RUNX3* são expressos de modo distinto nos vários tecidos; os mesmos autores mostraram ainda que os transcritos do *RUNX3* são co-expressos em alguns tecidos, juntamente com os transcritos dos outros membros da família RUNX. Apesar de nos últimos anos ter sido publicado um grande número de artigos focando o estudo do RUNX3 em diferentes aspectos da sua função, a informação recolhida continua a não ser suficiente para descrever exactamente quais os seus mecanismos de acção e qual a sua importância para o desenvolvimento, nomeadamente no que diz respeito à função de cada uma das múltiplas isoformas existentes para todas as proteínas desta família.

O peixe zebra foi escolhido como modelo de estudo para a realização deste projecto. Considera-se que os peixes teleósteos, grupo ao qual pertence o peixe zebra, tenham sido o primeiro grupo a desenvolver um esqueleto ósseo e, simultaneamente, toda a maquinaria necessária para a sua formação e manutenção. A combinação da genética e da embriologia estabeleceu o peixe zebra como sendo um organismo modelo importante para análises de desenvolvimento, fisiologia e comportamento em vertebrados. O desenvolvimento de técnicas especiais de clonagem, mutagénese e transgénese permitiu a identificação de um número importante de mutantes nesta espécie. A compreensão da inter-relação entre os genomas do peixe zebra e humano poderá ajudar na identificação da função de genes humanos a partir de mutações existentes nos genes correspondentes (ortólogos) do peixe zebra, tornando-o um bom modelo de estudo para certas doenças bem como para testes que resultem na identificação de novos agentes terapêuticos. Outras das vantagens da utilização deste modelo são, por exemplo, o seu desenvolvimento externo, a possibilidade de análise *in vivo* do desenvolvimento devido à transparência dos embriões, uma prole numerosa (200-300 ovos por fêmea) e um desenvolvimento rápido (em 48 a 72 horas evolui do estado de zigótico para larva e torna-se adulto aos 3 meses de vida) são atributos que favorecem a utilização deste modelo na investigação de inúmeras doenças humanas.

Para melhor conhecer e caracterizar as funções do gene *runx3*, a primeira etapa deste trabalho foi a clonagem, em peixe zebra, dos diferentes transcritos derivados de cada um dos promotores, P1 e P2, e a utilização dessas sequências para a realização de uma análise genómica comparativa. O resultado do alinhamento múltiplo entre as proteínas RUNX3 de várias espécies revelou que o domínio Runt, constituído por 128 aminoácidos, descrito como essencial para a dupla função de ligação ao ADN nas regiões reguladoras dos genes alvo e de heterodimerização com o co-factor CBFβ, evidencia um elevado grau de conservação ao longo da evolução das espécies. Verificou-se também que os restantes domínios da proteína são igualmente conservados entre as várias espécies analisadas.

Durante a clonagem dos diferentes transcritos, derivados do promotor P1 e do promotor P2 do gene *runx3*, foram obtidas dez novas variantes, aqui descritas pela primeira vez, estando de acordo com o previamente descrito para os outros genes desta família, para os quais

diversas variantes foram também identificadas, algumas com um papel fundamental na sua regulação.

O padrão de expressão do gene *runx3* foi determinado pela análise da expressão temporal específica dos transcritos derivados dos promotores P1 e P2, por hibridação *in situ* e por PCR quantitativo (qPCR), ao longo dos estadios de desenvolvimento embrionário e larvar e em tecidos do animal adulto. Os resultados da análise de qPCR demonstram que, embora os transcritos do *runx3* derivados dos promotores P1 e P2 sejam expressos numa variedade de tecidos, incluindo tecidos moles e tecidos mineralizados, e também nos diversos estadios embrionários e larvares testados, estes apresentam um padrão de expressão diferente. Enquanto os transcritos derivados do promotor P2 foram detectados em todos os tecidos estudados, no caso dos transcritos derivados do promotor P1 não se detectou expressão, ou a expressão relativa foi muito baixa em alguns tecidos não cartilagíneos. No entanto, para ambos os transcritos, os valores relativos de expressão foram mais acentuados nos tecídos cartilagíneos. É Interessante notar que os transcritos derivados do promotor P1 são expressos em estadios anteriores ao início da expressão zigótica, indicando a sua herança maternal, contrariamente aos derivados do promotor P2 que não foram detectados nestes estadios. Pela análise dos resultados da hibridação *in situ* usando uma sonda de ARN que reconhece uma região comum a ambos os transcritos, a expressão do *runx3* foi determinada nos gânglios trigerminais, neurónios sensoriais designados “Rohon-Beards” e em tecidos cartilagíneos craniofaciais. Usando uma sonda específica para os transcriptos do promotor P2, observou-se que a expressão coincide com a expressão observada quando uma sonda que detecta a região comum do *runx3* foi utilizada. No entanto, utilizando uma sonda específica para os transcritos derivados do promotor P1 não foi possível identificar qualquer sinal específico, provavelmente devido à sensibilidade do método utilizado.

As regiões reguladoras P1 e P2 do *runx3* permanecem até ao momento pouco estudadas, e portanto a regulação deste gene constitui ainda um tema de estudo em aberto e de importante significado. Fragmentos das regiões promotoras P1 e P2 do *runx3* de peixe zebra foram identificados e clonados, sendo depois sujeitos a uma análise comparativa *in silico* para identificação de possíveis motivos de ligação a factores de transcrição reguladores. Ambos os fragmentos dos promotores P1 e P2 analisados foram capazes de mediar a

transcrição da luciferase, usada como gene repórter, tanto *in vitro* em linhas celulares ósseas e não-ósseas, como *in vivo* em embriões de peixe zebra, verificando-se que ambos os promotores são activos nas condições testadas. Posteriormente, delecções de ambos os fragmentos foram testadas nas diferentes linhas celulares, permitindo a identificação de regiões reguladoras positivas e negativas em ambos os promotores do *runx3*, assim como a identificação de potenciais motivos de ligação a factores de transcrição reguladores nessas regiões, anteriormente determinados na nossa análise *in silico*.

Sabendo que o CBFβ é um co-factor que se liga ao domínio Runt de todos os membros da família RUNX, formando um heterodímero com maior afinidade de ligação ao ADN, decidimos amplificar o transcrito que codifica para esta proteína, para posteriormente introduzir num vector de expressão, de modo a testar a sua actividade em ensaios de transfecção. No decorrer dessa tarefa, 10 novas variantes, geradas por "splicing" alternativo, foram obtidas e analisadas em termos da sequência por comparação com variantes do CBFβ descritas noutras espécies. Após a análise de todas as variantes, quatro destas isoformas (isoformas 1-4) foram escolhidas como potencialmente interessantes para a análise da sua actividade *in vitro*. As isoformas 1 e 3 possuem uma região C-terminal distinta enquanto as isoformas 2 e 4 correspondem a variantes das isoformas 1 e 3, respectivamente. Para testar a capacidade de ligação e indução de cada uma dessas isoformas, foi usado um fragmento do ColXα1 previamente descrito como sendo regulado pela isoforma MASN-Runx2. Os resultados obtidos mostram que as isoformas 1 e 2 do Cbfβ possuem uma maior capacidade de indução do promotor do ColXα1 pela isoforma MASN-Runx2, comparado com o efeito observado pelas isoformas 3 e 4.

Foi ainda testada a hipótese do *runx3* ser regulado pelo Runx2 ou por si próprio (auto-regulação). Para isso, analisou-ses o efeito das isoformas do Runx2 e do Runx3, na presença ou ausência do CBFβ, na regulação dos promotores P1 e P2 do gene runx3. Os resultados preliminares obtidos indicam a existência de um efeito na regulação dos promotores do *runx3* pelas isoformas do Runx2 e do Runx3, na presença e na ausência da isoforma do Cbfβ testada, dependendo do promotor e das isoformas co-transfectadas. No entanto estes resultados precisam ainda de ser confirmados.

Por fim, foi testada a capacidade de cada um dos fragmentos dos promotores P1 e P2 do *runx3* reproduzir a expressão endógena em linhas trangénicas de peixe zebra. Para isso, foram construídos dois plasmídeos (Runx3-P1:GFP e Runx3-P2:GFP) contendo a sequência que codifica para a GFP ("green fluorescent protein") sob o controlo de cada um dos fragmentos do promotor do *runx3* a analisar. Estas construções foram injectadas em embriões de peixe zebra no estadio embrionário de 1 célula e a expressão da GFP analisada *in vivo* durante os estadios iniciais do desenvolvimento. Embora apenas se tenham analisado estes resultados transientemente, uma vez que no decorrer deste trabalho não se chegou a identificar nenhum portador do transgene, observou-se que nos peixes injectados com a contrução Runx3-P1:GFP a expressão da GFP foi muito reduzida, em contraste com a expressão observada nos embriões injectados com a contrução Runx3-P2:GFP, em que parte da expressão endógena do *runx3* obtida por análise de hibridação *in situ* foi reproduzida.

Com este trabalho pretendeu-se estudar a regulação do gene *runx3* de peixe zebra e a importância das suas diferentes isoformas para a respectiva função. Foram determinados padrões de expressão das isoformas do *runx3* tanto nos estadios iniciais de desenvolvimento, como em vários tecidos adultos, observando-se que as isoformas são diferencialmente expressas dependendo se derivam do promotor P1 ou promotor P2. Foi ainda realizada uma extensa análise das regiões reguladores dos promotores P1 e P2 do *runx3*, *in silico*, *in vitro* and *in vivo*, permitindo a identificação das regiões reponsáveis pela sua actividade basal, assim como a identificação de potenciais regiões reguladoras deste gene, tanto positivas como negativas, e a identificação de potenciais motivos de ligação de factores de transcrição que possam ser responsáveis pela regulação dessas regiões.

Palavras-chave: Runx3, promotores alternativos, análise da regulação de promotores, "splicing" alternativo, Cbfβ, padrões de expressão, peixe zebra.

Abreviations & Acronyms

aa	amino acid
AML	acute myeloid leukemia
bp	base pair
Co-IP	Co-immunoprecipitation
CBFα	core binding factor alpha
cDNA	complementary DNA
DBA	DNA block aligner
DMEM	Dulbecco's modified eagle's medium
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dpf	days post-fertilization
DRG	dorsal root ganglion
EDTA	ethylenediaminotetracetate
FBS	fetal bovine serum
hpf	hours post-fertilization
ISH	*in situ* hybridization
kb	kilobase pairs
kDa	kilodaltons
mRNA	messenger RNA
MW	molecular weight
ORF	open reading frame
OSF2	osteoblast-specific factor 2
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEBP	polyomavirus enhancer-binding protein
PWM	position weight matrices
qPCR	quantitative real-time PCR
RNA	ribonucleic acid
RUNX	runt homology domain transcription factor
Taq	*Thermus aquaticus* DNA polymerase
TFBS	transcription factor binding site
TFs	transcription factors
TSS	transcription start site
uATG	upstream ATG
uORF	upstream ORF
UTR	untranslated region

Index

Figure Index

Table Index

Chapter 1

General Introduction

1.1 Transcription in Eukaryotes - the first step in gene expression

The completion of "The Human Genome Project" in 2003 has revealed that the human genome encodes between 20.000 and 25.000 protein-coding genes, a number much lower than the initially predicted. The protein-coding sequences correspond to just 1.5% of the human genome, so all the remaining 98.5% correspond to non-coding DNA sequences.

One fundamental point in Biology is to understand how the different cell types are specified in a multicellular organism. In order to our bodies to be formed from a unique cell and develop into a broad range of very diverse cell types, sharing the same genetic information, there has to be a complex mechanism for controlling gene expression in a very precise way. In some cells, certain genes are "*turned off*" while in other cells they are "*turned on*" in order to be transcribed and translated into proteins. Eukaryotes have developed diverse chemical and physical mechanisms including chromatin condensation, DNA methylation, transcriptional initiation, alternative splicing of RNA and mRNA stability to control gene expression. Failure to do so can lead to dire consequences, so a tight controlled gene expression is essential to insure that the right gene is activated in the right cell at the right time during the development (Lodish et al, 2000).

One of the most important and highly regulated gene expression processes is the transcription of the DNA template to generate a RNA molecule. There are different elements in the transcription-control regions of the gene that regulate the transcription. In eukaryotes, transcription is carried out by three nuclear RNA polymerases (Patikoglou and Burley, 1997), RNA Polymerase I, II, and III, that transcribe different classes of genes. RNA polymerase I transcribes rRNA (ribosomal RNA), RNA polymerase II transcribes mRNA (messenger RNA), microRNAs and most snRNAs (small nuclear RNAs), and RNA polymerase III transcribes tRNA (transfer RNA), rRNA 5S and other small RNAs (Carter and Drouin, 2009). The vast majority of the genes in eukaryotic genomes encode functional proteins and contain a promoter region upstream from the gene, with certain specific motifs (nucleotide sequences) that recruit RNA Polymerase II (Zwan et al, 2003). The promoter region usually contains a highly conserved sequence called the TATA box (frequently at ≈25-35 base pairs upstream from the transcription start site) or an alternative promoter element called initiator that, unlike the TATA box, has an extremely degenerated consensus sequence

(Lodish et al, 2000). However, in some cases, genes do not contain a TATA box or an initiator but may contain a GC-rich sequence (20-50 nucleotides within ≈100 base pairs upstream of the transcription start site). The TATA-binding protein (TBP) initiates the formation of the basal transcription complex along with multiple TBP-associated proteins and multiple additional general transcription factors. Thus, TBP together with its TBP-associated proteins was originally identified as a fraction called TFIID and this group of proteins is required to build a basal transcription complex from TATA-containing or TATA-less promoters (Pugh and Tjian, 1991). The transcription by RNA polymerase II can be regulated by activators or repressors, called transcription factors (TFs) that bind to promoter elements and enhancers. These functional regulatory elements are generally found within several hundred bases of the transcription start site of the gene to which they are linked, but they can occasionally be located tens of kilobases upstream or downstream from a promoter, within an intron, or downstream from the final exon of a gene (Lodish et al, 2000; Fickett and Wasserman, 2000). The fundamental logic of transcriptional regulation in eukaryotes is that it is combinatorial: each gene has a particular combination of regulatory elements, where the nature, number, and spatial arrangement of which determine the gene unique pattern of expression. These promoter or enhancer elements control in which cell types the gene is expressed, the times during development in which it is expressed, and the level at which it is expressed in adults (Collingwood et al, 1999).

1.2 Transcription Factors

In eukaryotes, genes are usually in a default "*off*" state, so TFs serve mainly to turn gene expression "*on*". The TFs can bind alone or attract other TFs creating a complex that can either facilitate or repress the binding by RNA polymerase II to specific target genes, thus beginning or shutting the process of transcription (Roeder, 1996; Lee and Young, 2000; Levine and Tjian, 2003) (**Figure 1.1**). Because TFs are crucial to the regulation of gene expression, understanding their action of mechanisms is a major area of ongoing research in cell and molecular biology.

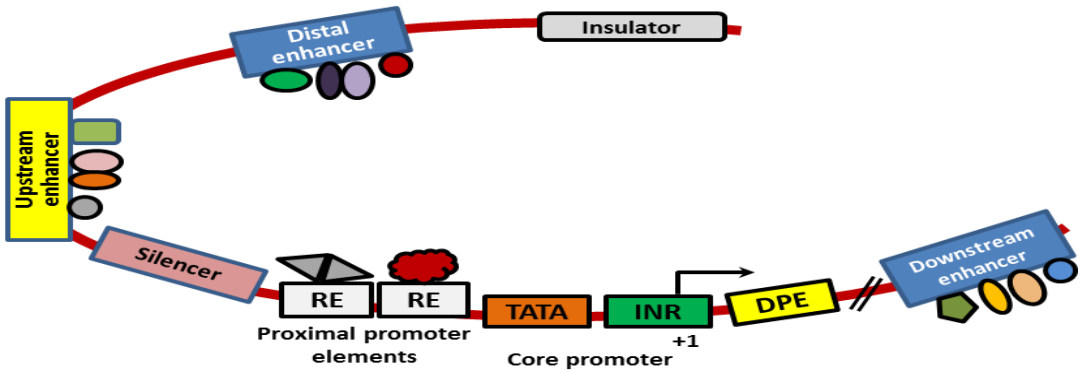


Figure 1.1 Complex metazoan transcriptional control modules. A complex arrangement of multiple clustered enhancer modules interspersed with silencer and insulator elements which can be located 10-50 kilobases (kb) either upstream or downstream of a composite core promoter containing TATA box (TATA), Initiator sequences (INR), and downstream promoter elements (DPE). Figure adapted from Levine and Tjian, 2003.

The control of gene expression in mammalian cells requires interactions among many combinations of transcription regulatory proteins (reviewed in Levine and Tjian, 2003). These interactions must be both versatile to enable individual TFs to interact with a variety of structurally unrelated proteins, as well as selective to enable different members of the same transcription factor family to have differential effects on gene expression (Hu et al, 2002).

1.3 The core binding factor (CBF) transcription complex

The core binding factor (CBF) is a heterodimeric transcription factor composed by two subunits, an α-subunit and a β-subunit. The α-subunit is characterized by containing a heterodimerization domain, responsible for the binding to the DNA as well as binding to the β-subunit (Kanno et al, 1998). The β-subunit does not bind the DNA itself, but acts as a co-factor, increasing the affinity of the α-subunit to the DNA of the target genes (Hart and Foroni, 2002).

1.4 The β-subunit of the CBF complex – CBFβ protein

CBFβ is the non-DNA binding subunit of the heterodimeric core binding factor (CBF). It is able to dimerize with its DNA-binding subunit, the product of each of the three identified *RUNX* genes, mediating a variety of biological responses according to the RUNX protein (1, 2 or 3) associated into the complex (Lund and van Lohuizen, 2002). One model for CBFβ function is that this protein helps to stabilize a conformation of the runt domain that has the highest affinity for DNA. CBFβ may have developmental and clinical significance beyond its association with the runt domain. Furthermore, the CBFβ protein is more widely distributed than either the mammalian runt domain proteins that have been identified (Ogawa et al, 1993; Wang et al, 1993). This suggests that CBFβ may participate in a wide variety of developmental pathways, possibly by interacting with other types of DNA-binding proteins (Kagoshima et al, 1993).

1.5 The α-subunit of the CBF complex - runx transcription factors

The α-subunit of the CBF complex corresponds to the runt-related transcription factors (RUNX) that comprise a small family of transcription factors identified by their highly conserved DNA binding domain, designated the runt domain. The first characterized member of this family was the Drosophila regulatory gene *runt,* discovered in a genetic screen (Nüsslein-Volhard and Wieschaus, 1980) and named for its mutant phenotype, which reflects its role as a “primary pair rule” gene in establishing the pattern of segments in the embryo (Gergen and Butler, 1988). Subsequently, the gene was found to have additional genetic functions in Drosophila, being implicated in sex determination and neurogenesis (Duffy and Gergen, 1991; Duffy et al, 1991). Due to its localization in the nucleus, it was predicted that the RUNX protein product could act as a transcription factor regulating the transcription of other genes (Kania et al, 1990). Some years later, a second Drosophila runt-related gene, *lozenge*, was identified, and shown to be required in cell patterning in the eye and developing antenna, as well as in cell fate specification during hematopoiesis (Gupta and Rodrigues, 1995; Daga et al, 1996; Flores et al, 1998; Canon and Banerjee, 2000; Bataillé et al, 2005). More recently, a Drosophila genome project revealed two additional runt-related

genes, CG42267 and CG34145 (Rennert et al, 2003; Bao and Friedrich, 2008), being described so far four *runx* genes in Drosophila.

Given the importance of *RUNX* genes, which play vital roles in the development of various species and human diseases, in the last years a large number of groups have focus their research in characterizing orthologues of *RUNX* genes from phylogenetically diverse species, trying to achieve a better understanding of their origin and functions. The evolution of *RUNX* genes begins with one more similar to *runx3* in invertebrates and then progress to multiple *RUNX* genes in higher species (Cohen, 2009). While primitive metazoans such as sea urchin and *C. elegans* appear to possess a single *RUNX* family member, so far, three have been described in mammals and four in insects, zebrafish and fugu. This can indicate (i) evolutionary conserved roles in metazoans and (ii) the necessity for a tight regulation of RUNX activity in specifying complex developmental events in higher organisms (Chuang et al, 2013).

RUNX proteins are crucial transcription factors that regulate a wide range of biological processes to orchestrate proper cell fate determination during the development of metazoans (Coffman, 2003). Traditionally described as a DNA-binding protein, RUNX is now viewed as a multifaceted protein that associates with diverse proteins to direct biological outcomes in a context-dependent manner (**Figure 1.2**) (Blyth et al, 2005; Chuang et al, 2013).

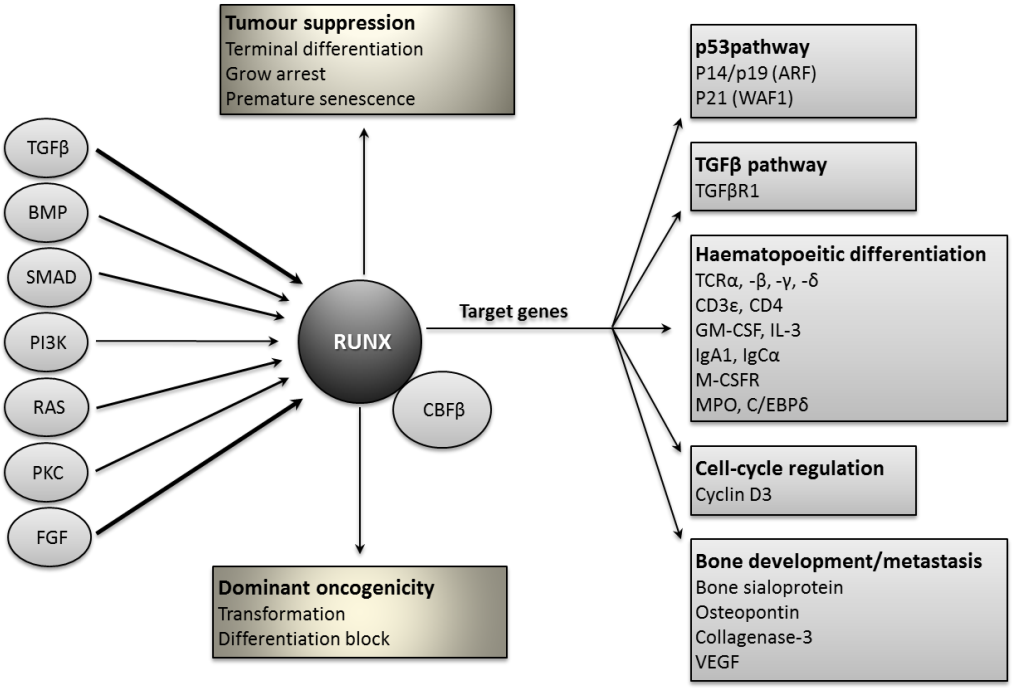


Figure 1.2 RUNX/CBFβ complexes in the determination of cell fate. This diagram illustrates the central role of RUNX/CBFβ complexes in the orchestration of cell fate in response to exogenous factors and environmental signals (circles in the left side). A subset of the known RUNX target genes is depicted (grey boxes), and these targets have been selected for their potential relevance to cancer. Figure adapted from Blyth et al, 2005.

1.6 Runx Aliases - Nomenclature for runt-related (RUNX) proteins

Since the first runt-related gene has been cloned and studied in Drosophila, there was an explosion in the number of independent laboratories focusing their studies in this family of transcription factors. So, in the literature, a diversity of family names can be found for these genes (**Table 1.1** and **1.2**) depending on the context of their study (van Wijnen et al, 2004).

Table 1.1 – Aliases for the runt-domain class of transcription factors.

Abbreviation	Gene Name	*Specie described*
Runt	Runt	*Drosophila*
Lz	Lozenge	*Drosophila*
AML	Acute myelogenous leukemia	Human
PEBP2alpha	Polyomavirus enhancer-binding protein 2 alpha subunit	Mouse
CBFalpha	Enhancer core-binding factor α of murine leukemia viruses (i.e SL3-3, AKV and Moloney MLV)	Mouse
PEA2	Polyoma enhancer A-binding factor 2	Mouse
SEF1	SL3-3 enhancer factor 1	Mouse
S/A-CBF	SL3-3 and AKV core-binding factor	Mouse
NF-deltaE3A	Nuclear factor delta E3A	Human
MyNF1	Myeloid nuclear factor 1	Mouse
NMP2	Nuclear matrix protein 2	Rat
OBSC	Osteoblast-specific complex	Rat
OSF2	Osteoblast-specific factor 2	Mouse
til-1	T-cell tumor integration locus 1 protein	Mouse
run	Runt domain-encoding gene	*Caenorhabditis elegans*

Table adapted from (van Wijnen et al, 2004).

Table 1.2 - Common synonyms for each Runx protein.

Runt-Related Transcription Factor 1	RUNX1; AML1; CBFα2; CBFA2; PEBP2αB; PEBP2A2; PEA2αB; SL3-3 enhancer factor 1αB subunit; Oncogene AML1; SL3/AKV core binding factor αB subunit; AML1-EVI-1; AMLCR1; EVI-1
Runt-Related Transcription Factor 2	RUNX2; AML3; CBFα1; CBFA1; PEBP2αA; PEBP2A1; PEA2αA; SL3-3 Enhancer Factor 1αA Subunit; Oncogene AML3; SL3/AKV core-binding factor αA Subunit; Osteoblast-specific transcription factor 2 (OSF-2); CCD
Runt-Related Transcription Factor 3	RUNX3; AML2; CBFα3; CBFA3; PEBP2αC; PEBP2A3; PEA2αC; SL3-3 Enhancer Factor 1αC Subunit; Oncogene AML2; SL3/AKV Core-Binding Factor αC Subunit

Table adapted from (http://www.genecards.org/)

Due to the existence of a wild diversity of names for the runt-related family of transcription factors, it was difficult to find in the literature all the references existing for this subject. So, in 1999, the Nomenclature Committee of the Human Genome Organization (HUGO) adopted

the use of the term 'RUNX' to refer to genes encoding the runt-related proteins. It was agreed by the investigators of this field, that any designations can be used to refer to the runt-related proteins, but they should be referred as RUNX proteins at least once in the Abstract and/or Title of new manuscripts submitted for publication (van Wijnen et al, 2004).

1.7 *RUNX* gene and protein structures

There are three *RUNX* genes described in mammals and all RUNX gene products share many structural similarities, but have distinct biological activities. A variety of transcripts have been described for all RUNX members, originated from the regulation of transcription from two alternative and distantly located promoters (P1 and P2), due to a high number of alternative splicing events and the presence of alternative polyadenylation (Miyoshi et al, 1995; Ahn et al, 1996; Levanon et al, 1996; Geoffroy et al, 1998). The *RUNX* gene structure is well conserved among all vertebrate species analysed (**Figure 1.3**) (Levanon and Groner, 2004).

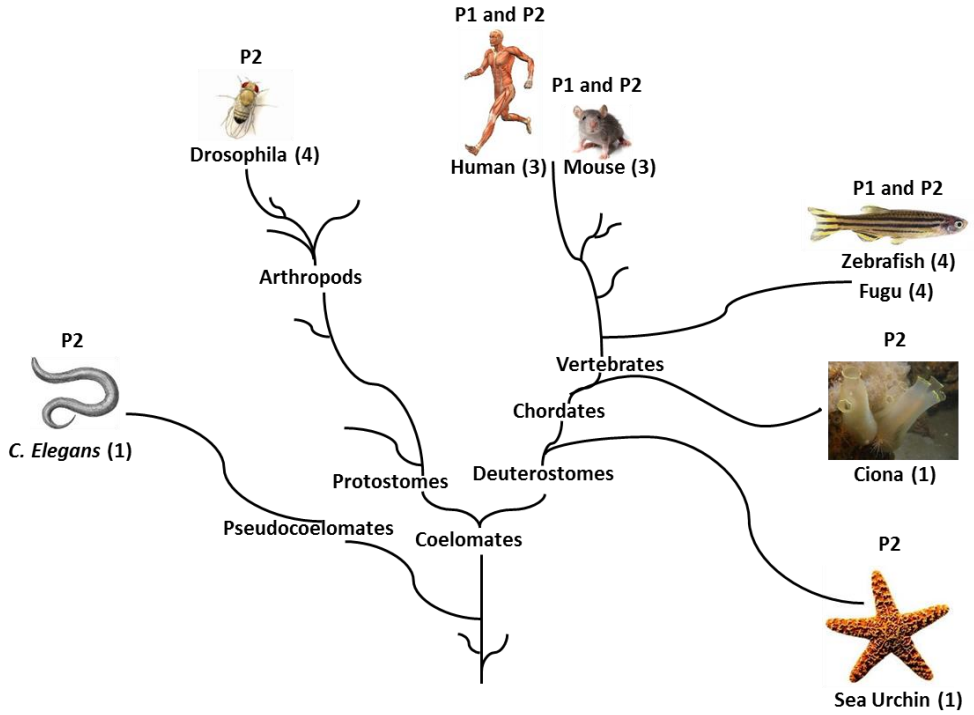


Figure 1.3 Phylogenetic illustration of *RUNX* genes showing the gene number and promoter usage in different species. The three lower non-designated branches in the phylogenetic tree represent the animal groups (from bottom up) sponges, cnidarians (e.g. jellyfish) and acoelomates (e.g. flatworms). The more primitive animals contain one gene regulated by the P2 promoter. Figure adapted from Levanon and Groner, 2004.

There are two major protein isoforms for each *RUNX* gene, with a different N-terminal sequence depending if it is encoded from the P1-derived or from the P2-derived transcript. The gene structure of the three human *RUNX* genes is represented in **Figure 1.4**.

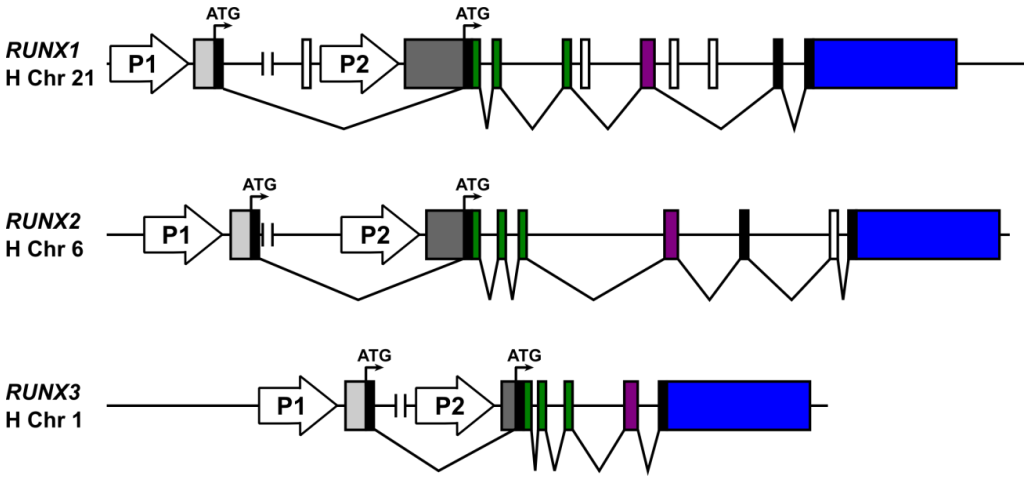


Figure 1.4 Genomic organization of the human *RUNX* genes. The two promoters P1 and P2 and initiator codons (ATG) are indicated. The P1-derived and P2-derived 5' UTRs are indicated in light and dark grey, respectively, common coding exons are shown in similar color and 3' UTRs are represented in blue. Figure adapted from Levanon and Groner, 2004.

The RUNX proteins that originate from the P2-derived transcripts contain a specific N-terminal region of five amino acids – the penta-peptide MRIPV. On the other hand, the RUNX isoforms that originate from the P1-derived transcripts contain a specific N-terminal region of 19 amino acids, originated by an alternative splicing from exon 1 onto the coding region of P2-derived transcript through a conserved in-exon splicing site located 16 bp downstream of the P2-ATG codon. This 19 amino acids long specific N-terminal is conserved in all P1-derived RUNX proteins, with the highest sequence similarity found between RUNX1 and RUNX3 (Bangsow et al, 2001). The isoforms are usually designated as isoform RUNX3-MA(S/D)NS and RUNX3-M(R/H)IPV, depending if they are encoded from the P1-derived or P2-derived transcripts, respectively.

1.8 Functional domains of RUNX proteins

Ito (2004) and, more recently, another report from the same group (Chuang et al, 2013), have described the common structure of the RUNX proteins by comparison of the amino acid sequences from the three RUNX proteins. They confirmed that all RUNX proteins share the same characteristic domains and motifs: runt domain (RD), transactivation domain (TAD), inhibitory domain (ID), nuclear localization signal (NLS), PY (or PPxY) motif and the VWRPY motif that corresponds to the last five amino acids of the RUNX proteins (**Figure 1.5**). RUNX2 possesses a unique QA domain composed of a stretch of Q and A residues (Ito and Miyazono, 2003). The N-terminal part of the molecules comprises the RD, a 128 amino acids long domain that is highly conserved between species and has an S-type immunoglobulin fold (Warren et al, 2000). The C-terminal part of the RUNX molecule contains the other domains (TAD and ID) that play a role in transcription regulation (reviewed in Downing, 1999; Ito, 1999), and a conserved PY motif, consisting of is a proline-rich peptide, that interacts with proteins harbouring the WW domain (Chuang et al, 2013) and a VWRPY motif at the end of the C-terminus of RUNX proteins, which mediate the interaction with the co-repressor Groucho/TLE (Liu et al, 2006; reviewed in Cohen, 2009) (**Figure 1.5**).

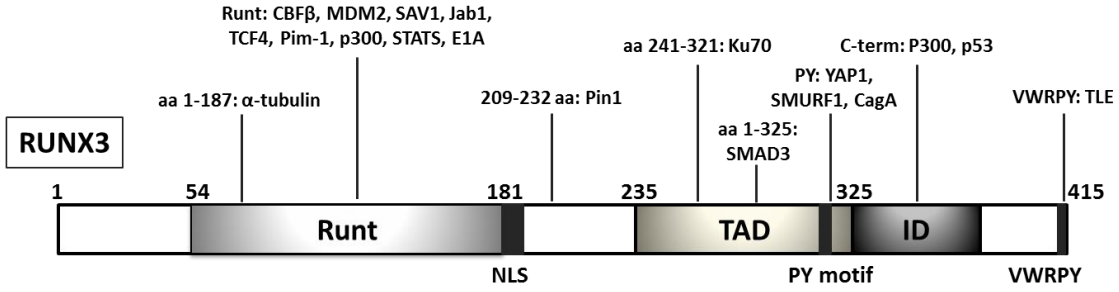


Figure 1.5 Schematic representation of functional domains of RUNX3 protein and examples of its interacting proteins. The domains represented are conserved among all RUNX proteins. The numbers indicate amino acid positions. TAD, ID, C-term and NLS refer to transactivation domain, inhibitory domain, C-terminus and nuclear localization signal, respectively. Figure adapted from Chuang et al, 2013.

Traditionally described as a DNA-binding protein that can act as a homodimer or heterodimer (with its partner CBFβ) to bind to the promoter of the target genes, RUNX is now viewed as a multifaceted protein that associates with a wide variety of proteins to

direct biological outcomes in a context-dependent manner. Those proteins can either be co-activators or repressors. Examples of some co-factors previously described to interact with RUNX proteins (Blyth et al, 2005; Chuang et al, 2013) are represented in **Figures 1.5** and **1.6**.

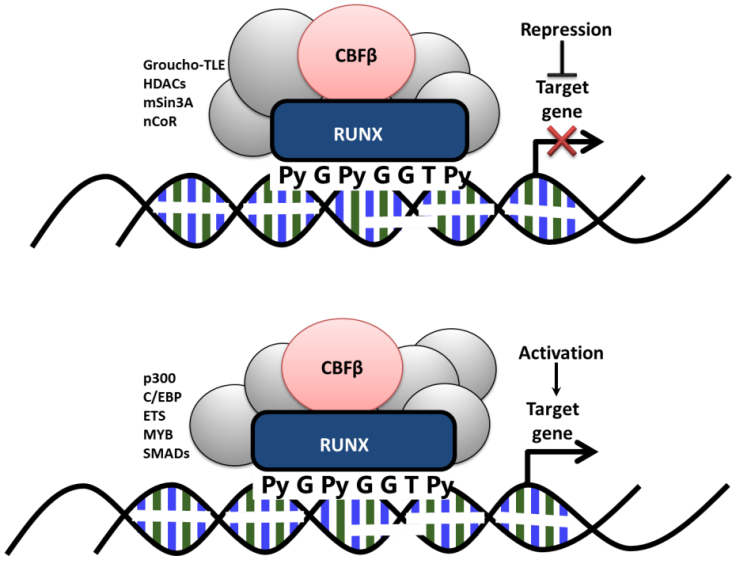


Figure 1.6 Schematic representation of the RUNX factors as organizers of transcription. RUNX factors bind to their consensus sequence (PyGPyGGTPy) in the promoter region of the target genes as a heterodimer with their binding partner CBFβ. A complex including other co-factors is generally formed, determining the outcome of the RUNX regulation on transcription of the target genes. CBFβ, core-binding factor-β; C/EBP, CCAAT/enhancer-binding protein; HDACs, histone deacytylases; nCoR, nuclear receptor corepressor; TLE, transducin-like enhancer of split. Figure adapted from Blyth et al, 2005.

The factors affecting the outcome of this interaction on transcription of the target gene seem to involve promoter-specific features such as the proximity of binding sites for co-activators or repressors as well as the availability of such cofactors in the nucleus.

1.9 Runt domain characteristics

All RUNX proteins have the characteristic runt domain that is known to have the dual function to bind to the promoter of target genes and also to bind to CBFβ protein forming the CBFα/β complex. This complex increases allosterically the DNA-binding affinity (Ogawa et al, 1993; Miller et al, 2002) and regulates the RUNX proteins turnover by protecting them

from ubiquitin-proteasome-mediated degradation (Huang et al, 2001). All RUNX proteins heterodimerize with CBFβ and all bind to the same consensus recognition sequence (5'-PyGPyGGT-3') (Py is any pyrimidine) found in a number of enhancers and promoters. The biological activity of the CBF complex is specified by the α-subunit according to the RUNX protein (1, 2 or 3) associated into the complex (Tahirov et al, 2001; Bäckström et al, 2002; Lund and van Lohuizen, 2002; Horsfield et al, 2007). The degree of homology between invertebrate and vertebrate runt domains is quite high, and this may reflect the conservation of dual functions within that domain, as described above.

There are two types of heterodimeric factors either (i) both subunits bind to specific DNA sequences (e.g. Jun/Fos and Myc/Max) or (ii) one subunit binds to DNA and one does not (e.g. RUNX/CBFβ). The molecular mechanisms of the DNA binding for the first type of heterodimeric factors was already described (Glover and Harrison, 1995), but for the second type it is still not well understood how non-DNA binding subunit contribute to DNA binding and whether the mechanism by which a non DNA binding subunit stimulates DNA binding is shared with different transcription factors (Tahirov et al, 2001).

Due to the important function of the heterodimeric RUNX/CBFβ complex in different developmental processes, regulating different promoters and enhancers of a variety of genes in a cell specific manner, different groups have used X-ray crystallography and nuclear magnetic resonance (NMR) to determine the tertiary structures of the runt domain and the runt/CBFβ heterodimeric complex, and the molecular mechanisms of their interaction (Nagata et al, 1999; Tahirov et al, 2001; Bravo et al, 2001; Bäckström et al, 2002). These studies revealed that the runt domain adopts the immunoglobulin (Ig)-like fold β-sandwich, which is homologous to other DNA-binding domains such as those found in p53, nuclear factor-kappa B (NF-kB), nuclear factor of activated T cells (NFAT) and other proteins (Nomura et al, 2013).

Determination of the tertiary structures together with mutational analysis was crucial to the identification of regions that are essential to the binding affinity between the runt domain and the DNA consensus recognition sequence. One example was the recent identification of the three guanines in the DNA consensus recognition sequence (5'-PyGPyGGT-3'), which are directly recognized by three arginine residues of the runt domain, as being important for the

efficient binding between the runt domain and the recognition element in the DNA (Nomura et al, 2013).

1.10 Biological roles of RUNX proteins

The proteins from the RUNX family show versatile functions being involved in several diseases and acting on target genes in a variety of tissues, and exhibiting both transcriptional activation and repression activities (Otto et al, 2003). These transcription factors are tissue-restricted and cancer-related, regulating cell proliferation and growth, as well as differentiation (Coffman, 2003; Miyazono et al, 2004; Stein et al, 2004; Young et al, 2007a, 2007b; Kagoshima et al, 2007). It is clear that each of the three RUNX proteins has different roles (Rennert et al, 2003), but since they were found to be co-expressed in some tissues, some of their functions may overlap, either synergistically or at different time points (Cohen, 2009).

1.10.1 The runt-related transcription factor 1 (RUNX1)

The human *RUNX1* was originally cloned from the breakpoint of chromosome 21 in t(8;21)(q22;q22) by Miyoshi and collaborators (Miyoshi et al, 1991). This translocation is frequently found in patients with acute myeloid leukemia (AML) with maturation (M2 subtype) (Miyoshi et al, 1991) and has subsequently been shown to be one of the most frequent targets of leukemic-associated gene aberrations (Reviewed in Okuda et al, 2001). RUNX1 has been proved to regulate, acting as a context-dependent transcriptional activator or repressor, a number of target genes, e.g. granulocyte macrophage colony-stimulating factor (GM-CSF) (Oakford et al, 2010), receptor for macrophage colony-stimulating factor (CSF-1R) (Sauter et al, 2013), tissue inhibitor of metalloproteinase 1 (TIMP1) (Bertrand-Philippe et al, 2004), early B-cell factor 1 (Ebf1) (Seo et al, 2012), cyclin D3 (Bernardin-Fried et al, 2004), CD11 integrin (Puig-Kröger et al, 2000), complement receptor type 1 (CR1) (Kim et al, 1999), insulin-like growth factor-binding protein 3 (IGFBP-3) (Iwatsuki et al, 2005), type B leukemogenic virus (TBLV) enhancer (Mertz et al, 2001), among others.

RUNX1 is expressed in a number of tissues at specific time windows during embryogenesis. In the mouse embryo, expression of *Runx1* is first detected in definitive hematopoietic stem cells (HSC) and in endothelial cells at HSC emergence sites (North et al, 1999; Cai et al, 2000; North et al, 2002) and later, *Runx1* is highly expressed in several hematopoietic lineages including myeloid, B- and T-lymphoid cells (Lorsbach et al, 2004). *Runx1* is also highly expressed in cranial and dorsal root ganglia (DRG), in the small-diameter nociceptive TrkA neurons (Simeone et al, 1995; Levanon et al, 2001, 2002), in the thymus (Satake et al, 1995; Komori et al, 1997; Levanon et al, 2001) and in cells of the chondrocyte lineage of the developing mouse skeleton (Sato el at, 2008). Tissue sections of the fetal liver or on smear preparations of peripheral blood of mutated mice embryos showed an absence of hematopoietic elements of definitive origin and analysing the yolk sac the erythropoiesis was minimally affected. So, this indicates that the gene targets of RUNX1 are essential for the development of all cell lineages of definitive hematopoiesis, but not required for erythropoiesis (Okuda et al, 2001). In zebrafish and Xenopus, the expression pattern of *runx1* is well conserved compared to the mammalian pattern, and was shown to be expressed in blood progenitors (Tracey et al, 1998; Kalev-Zylinska et al, 2002; Burns et al, 2005). In these species, *runx1* is also expressed in a few scattered cells within craniofacial cartilages (Kalev-Zylinska et al, 2003; Flores et al, 2006; Park and Saint-Jeannet, 2010) and in Rohon-Beard sensory neurons, that are a population of specialized neurons with mechanoreceptive properties only found in fish and amphibians and contribute importantly to the embryonic nervous system of these species (Rossi et al, 2009).

Knockout studies, using different Runx1 deficient mice lines, show that the lack of functional Runx1 protein production causes necrosis and haemorrhaging in the central nervous system and blocks definitive hematopoiesis resulting in death of the mice during the mid-gestational period (Okuda et al, 1996; Wang et al, 1996a). The importance of the heterodimerization of CBFβ with RUNX1 is supported by analysis of the disruption of CBFβ that causes a similar phenotype as the one obtained for the *Runx1* knockout (Wang et al, 1996b). Consistent with the *Runx1* knockout mice, the zebrafish mutant embryos show normal primitive hematopoiesis, but a blockage of the definitive hematopoiesis (Jin et al, 2009). However, the RUNX1 role in the generation and maintenance of HSCs during adult hematopoiesis remains poorly understood. Recently, Sood and colleagues (2010) reported that a zebrafish *runx1*-

deficient mutant line, carrying a truncated version of runx1, is capable of development of multilineage adult hematopoiesis (Sood et al, 2010). Apart the numerous reports describing the involvement of RUNX1 in hematologic malignancies (Taketani et al, 2003; Podgornik et al, 2007; Dicker et al, 2007), there are studies reporting the importance of RUNX1 in a wide range of diseases, for example, RUNX1 is involved among others in autoimmune diseases (Helms et al, 2003; Shen and Tsao, 2004; Chae et al, 2006), and in cancer (Yang et al, 2005; Sakakura et al, 2005; Moosavi et al, 2009).

1.10.2 The runt-related transcription factor 2 (RUNX2)

RUNX2 has been shown to exert a critical role in normal bone formation and osteoblast differentiation. It has been extensively studied and shown to be a key regulator in skeletal development in mammals, regulating the switch from cartilage to bone (Otto et al, 1997; Inada et al, 1999; Takeda et al, 2001; Ueta et al, 2001). RUNX2 is involved in a variety of functions such as in osteoblast differentiation (Komori et al, 1997), chondrocyte maturation (Enomoto et al, 2000; Takeda et al, 2001), bone remodelling (Gao et al, 1998; Liu et al, 2001) and in the transcriptional regulation of a variety of genes crucial for bone and cartilage development including osteocalcin (Ducy et al, 1997; Pinto et al, 2005), matrix Gla protein (Fazenda et al, 2010), collagen type X (Zheng et al, 2003; Simões et al, 2006), Indian hedgehog (Yoshida et al, 2004; Hecht et al, 2008), sclerosteosis (Sevetson et al, 2004), and in its own promoter (Drissi et al, 2000; Yoshida et al, 2002).

This complex genetic regulation seems to be due to a difference in distribution between MASN-Runx2 and MRIPV-Runx2 isoforms, not only resulting from a spatial-temporal expression but also due to species specificity (Harada et al, 1999; Li and Xiao, 2007) and to the existence of co-regulators that interact with Runx2 enhancing or inhibiting its function (Reviewed in Komori, 2005). Studies at cellular level in mice revealed that both Runx2 isoforms are expressed in osteoblasts and terminal hypertrophic chondrocytes, however, MRIPV-Runx2 isoform is also expressed in undifferentiated mesenchymal stem cells, preosteoblasts and chondrocyte precursors (Li and Xiao, 2007). Observations at the tissue level found that although both isoforms are expressed in bone and lung tissue, MRIPV-Runx2

is expressed more widely and can be detected in heart, brain, spleen and skeletal muscles (Xiao et al, 2003; Li and Xiao, 2007).

Komori and colleagues (1997) showed that the ossification is completely blocked in Runx2 mutated mice. Although the development of cartilage was nearly normal, the mutant mice died just after birth and showed a complete lack of bone formation (Komori et al, 1997). Diseases associated with RUNX2 include the bone development disorder cleidocranial dysplasia (CCD) (Golan et al, 2000; Yamachika et al, 2001), osteosarcoma (San Martin et al, 2009; Sadikovic et al, 2010), osteoporosis (Khalid et al, 2008) and cancer (Reviewed in Blyth et al, 2005).

1.10.3 The runt-related transcription factor 3 (RUNX3)

The mammalian *RUNX3* gene was cloned and localized to human and mouse chromosomes 1p36.1 and 4, respectively (Levanon et al, 1994; Avraham et al, 1995; Bae et al, 1995). From the three mammalian genes belonging to the RUNX family of transcription factors, *RUNX3* gene is the smallest and the one with fewest number of exons (6 exons), compared to 11 exons and 8 exons in *RUNX1* and *RUNX2*, respectively (Levanon and Groner, 2004). Phylogenetic studies of *RUNX* orthologs from diverse organisms point to *RUNX3* as the evolutionary founder with one *runx3* in invertebrates that progressed to multiple *RUNX* genes in higher species (Bangsow et al, 2001; Reyes et al, 2004, Cohen, 2009).

RUNX3 has been shown to be a transcriptional regulator, either activating or repressing the target genes, by binding to their promoters as a homodimer or heterodimer (with its co-factor partner CBFβ). Some examples of described RUNX3 target genes include survivin (Liu et al, 2014), thrombospondin-1 (Shi et al, 2013), aggrecan (Wigner et al, 2013), Foxp3 (Bruno et al, 2009), Runx1 (Brady and Farrell, 2009), osterix (Zheng et al, 2007), Bim (Yamamura et al, 2006), CD11a and CD49d integrins (Domínguez-Soto et al, 2005), among others.

Runx3 is the main focus of this report and so it will be described in more detail in the following sections.

1.10.3.1 Subcellular distribution of RUNX3 protein

The activity of RUNX is regulated by tissue-specific expression and post-translational modifications, and is also modulated by subcellular localization (e.g. nuclear localization signal and nuclear matrix targeting signal domains) (Pockwinse et al, 2006). As observed for Runx1 and Runx2, Runx3 is predominantly a nuclear protein that is punctately organized (Pande et al, 2009), however it was also localised in the cytoplasm of gastric cancer cells showing the importance of the spatial distribution for proper RUNX function (Ito et al, 2005). Pande and co-workers proposed that disruption of Runx3 subnuclear targeting may abrogate target gene activation and have important pathological implications for the development of gastric tumours (Pande et al, 2009). However, this dynamic shuttle between the nucleus and the cytoplasm is still poorly understood.

1.10.3.2 *Runx3* expression pattern

Various groups have published *in situ* hybridization expression patterns of *Runx3* in different species, describing detection of *Runx3* transcripts in a broad range of tissues, either alone or co-expressed with other *Runx* family members. Levanon and colleagues (2001) showed that, in mice, *Runx3* is the only Runx family member expressed in a small number of large-diameter neurons and in cranial and dorsal root ganglia (Levanon et al, 2001). These authors also showed *Runx3* to be expressed in hematopoietic precursors in the liver, thymus, eyelid mesenchyme, superficial cutaneous mesenchyme and mesenchymal element of *filiform papillae* of the tongue, where it is co-expressed with *Runx1*, and in cartilage, where it was expressed in pre-hypertrophic and hypertrophic chondrocytes, co-localized with *Runx2* (Levanon et al, 2001). *Runx3* was also shown to be expressed in the ovaries and uteri of mice (Tsuchiya et al, 2012), in Meckel's cartilage and other craniofacial cartilages, and in the osteogenic regions and ossified bone (Yamashiro et al, 2002). In the developing tooth, *Runx3* expression persisted in odontoblasts and was particularly intense in the dental mesenchyme in the cervical loop of continuously growing incisor (Yamashiro et al, 2002). Beside mammals, *Runx3* spatial-temporal expression pattern has also been analysed by mRNA *in situ* hybridization in other species such as Xenopus (Park and Saint-Jeannet, 2010) and

zebrafish (Kataoka et al, 2000; Kalev-Zylinska et al, 2003, Flores et al, 2006) and by RT-PCR in a number of other species, for example, in fugu (Ng et al, 2007), elephant shark (Nah et al, 2014a), Japanese lamprey (Nah et al, 2014b), among others. All together, *in situ* hybridisation results obtained for both zebrafish and Xenopus showed that *runx3* is expressed in the same cells: hematopoietic lineage, trigeminal ganglia, Rohon-Beard neurons and cartilaginous tissues in the head (Kataoka et al, 2000; Kalev-Zylinska et al, 2003; Flores et al, 2006; Park and Saint-Jeannet, 2010). *runx3* expression has also been detected in zebrafish intestinal bulb (Kalev-Zylinska et al, 2003). Although in mouse *Runx3* expression has been detected in a subset of the DRG sensory neurons, so far in zebrafish it is still not clear if *runx3* is expressed in these subset of sensory neurons. However, in zebrafish and Xenopus the *runx3* expression is strongly detected in Rohon-Beard spinal sensory neurons (Kataoka et al, 2000; Kalev-Zylinska et al, 2003; Park and Saint-Jeannet, 2010), that as described before, are a unique set of primary sensory neurons located in the dorsal spinal cord only present in lower vertebrates (Reyes et al, 2004). Rohon-Beard cells are known to derive from the same neural plate domain that generates neural crest cells (Cornel and Eisen, 2000). During development, these neurons undergo apoptosis and their death coincide with the development of the neural crest-derived dorsal root ganglia neurons (Reyes et al, 2004; Park and Saint-Jeannet, 2010). Taken together, the *runx3* expression pattern in fish and in Xenopus seem to be, at least in neural tissues, different from that of the mammalian *Runx3* expression pattern.

1.10.3.2.1 Tissue distribution of *Runx3* specific isoforms

Rini and Calabi (2001) performed a systematic survey (summarized in **Table 1.3**) for all three *Runx* loci, and showed that the mouse *Runx3* is expressed in a wide range of tissues, including soft and calcified tissues. In that study, they also showed that each *Runx3* promoter-specific transcript is expressed in the different tissues in a different way, and that the *Runx3* transcripts can be co-expressed with other *Runx* members in some tissues (**Table 1.3**) (Rini and Calabi, 2001). The expression pattern of the *Runx3* P1 and P2 derived transcripts suggests that RUNX3 has isoform specific functions since its transcript variants

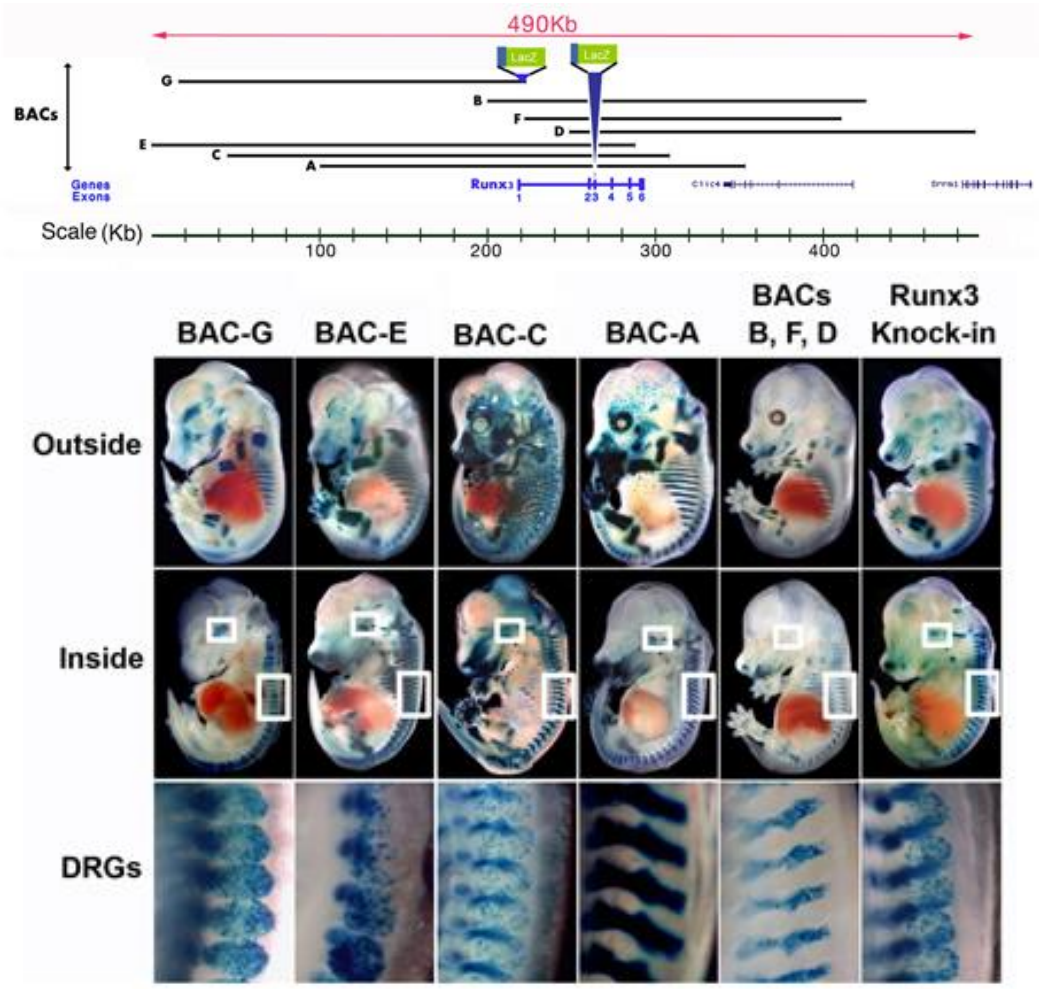
are differentially expressed in the different tissues, with *Runx3*-P2 been expressed in a broad range of tissues and the *Runx3*-P1 showing a more restricted expression pattern (**Table 1.3**).

Table 1.3 – Summary of the expression data for *Runx* transcripts obtained by Rini and Calabi (2001).

	E	BM	Sp	Th	Br	Ma	Gt	Ht	Kd	Lv	Lg	Sm	Ov	Ts
Runx1 P1	-	++	++	++	-	+	+	+	-	-	±	-	±	-
Runx1 P2	+	+	±	+	-	+	+	+	-	-	++	-	+	-
Runx2 P1	+	+	-	-	-	-	-	-	-	-	-	-	++	+
Runx2 P2	+	+	-	-	-	-	-	-	-	-	+	-	+	+
Runx3 P1	-	-	-	+	?	-	-	-	-	?	-	-	+	-
Runx3 P2	+	+	+	+	?	+	+	+	+	?	+	+	+	+

E, E12.5 embryo; BM, bone marrow; Sp, spleen; Th, thymus; Br, brain; Ma, breast; Gt, gut; Ht, heart; Kd, kidney; Lv, liver; Lg, lung; Sm, skeletal muscle; Ov, ovary; Ts, testis.

Groner´s group used a series of overlapping BACs spanning large regions of *Runx3* loci to identify and characterize the long-range regulatory elements mediating tissue and stage specific expression of *Runx3* during development and to address the question of how does transcription regulation of *Runx3* mediate cell-specific expression. Using transgenesis and deletions within *Runx3* reporter BACs, these authors discovered that *Runx3* expression in developing sensory neurons is regulated through concerted action of several long-range cis-regulatory elements located upstream of the gene (**Figure 1.7**) (http://www.weizmann.ac.il/molgen/Groner/research). However, so far, the expression pattern of *Runx3* isoform-specific transcripts and the function of each RUNX3 isoform are still poorly studied.



Tissues	G	E, C, A	B, F, D
Hair follicles	-	++	+/-
Vertebrae	-	+++	+
Palm	-	+++	+
Limbs, ribs	++	+++	+
Thymus	+/-	(E, C) +/- (A) +	-
Liver	+/-	(E, C) +/- (A) +	-
DRG	++	(E) ++ (A) +/-	-
Trigerminal ganglion	++	(E) +++ (C) ++	-

Figure 1.7 DRG TrkC neuron-specific enhancers are located hundred kb upstream of the P1 promoter/transcriptional start site. Series of transgenic mouse lines expressing the LacZ reporter gene under the regulation of the endogenous *Runx3* cis-regulatory elements and the different BAC constructs covering ≈0.5 megabase (Mb) around the *Runx3* gene. Top: Schematic representation of the mouse *Runx3* gene and the BAC constructs used to generate the transgenic lines and mouse transgenic E14.5 embryos, stained for lacZ, showing the spatial-temporal expression pattern of *Runx3*. Bottom: Table of *Runx3* expression pattern obtained by LacZ analysis in the transgenic embryos. (Image from http://www.weizmann.ac.il/molgen/Groner/research).

1.10.3.3 Functional studies to access RUNX3 biological functions

During mouse embryogenesis *Runx3* is expressed in hematopoietic organs, epidermal appendages, developing bones, and sensory ganglia (Levanon et al, 2001; Woolf et al, 2003). Over the last 15 years several authors have analysed the phenotypic effects of Runx3 silencing in the developing mouse embryo by analysis of the two currently available Runx3-deficient mice strains, obtained by Ito's and Groner's groups, and by knockdown studies using a morpholino targeting Runx3 (Littman's group). Although some discrepancies in the analysis of the Runx3 KO phenotype associated with gastric neoplasia have been observed, probably due to differences in the constructions of the KO mice, all strains show similar phenotypic features resulting from the Runx3 loss, with severe congenital limb ataxia as the more predominant phenotype, and high mortality soon after birth (Levanon et al, 2002; Taniuchi et al, 2002; Ito el al, 2003; Levanon and Groner, 2009, among others). *Runx3* was disrupted in these KO mice strains by inserting a LacZ-neomycin cassette into a region that affects both P1- and P2-derived RUNX3 isoforms, so the contribution to the phenotype of each isoform is still not understood. Based in gain and loss of function studies, Runx3 has been shown to be crucial for the regulation of the cell type specification of $TrkC^{+}$ proprioceptive dorsal root ganglion (DRG) neurons (Inoue et al, 2002; 2003; Levanon et al, 2002; Nakamura et al, 2008), to regulate both chondrogenesis and chondrocyte maturation (Yoshida et al, 2004; Soung et al, 2007) and to be involved in the regulation of the proliferation and survival of various cells including gastric epithelial cells (Li et al, 2002; Ito el al, 2008), in T-cell development (Taniuchi et al, 2002; Woolf et al, 2003; Wong et al, 2011) and in differentiation of immune cells including natural killer cells (Ohno et al, 2008), dendritic cells (Fainaru et al, 2004) and B cells (Watanabe et al, 2010).

As described above the mice $Runx3^{-/-}$ strains available exhibited high mortality in the first few days after birth, making it difficult to completely understand the effect of Runx3 depletion later on mice development. Zebrafish has been shown to be a powerful model to study human diseases and despite several hundred million years of evolutionary separation, the genes responsible for some of the teleost and mammalian developmental systems are highly conserved (Crosier et al, 2002). So, to overcome some of the difficulties using mouse and due to several advantages using zebrafish as a model of study, some researchers used

the latter model to analyse the phenotype of *Runx3* depletion (Crosier et al, 2002; Kalev-Zylinska et al, 2003; Flores et al, 2006). Morpholino knockdown studies showed that Runx3 is important for hematopoietic development in zebrafish, as abrogation of Runx3 function caused abnormalities in hematopoiesis (Croiser et al, 2002; Kalev-Zylinska et al, 2003). It was observed that morphants preserved development of primitive hematopoietic progenitors, however later in development a reduction of circulating blood cells was observed, suggesting that Runx3 is required for the maintenance of early blood cell numbers and for establishment of definitive hematopoiesis (Crosier et al, 2002; Kalev-Zylinska et al, 2003). The same group also studied the effect of Runx3 depletion in the onset of chondrogenesis in craniofacial endochondral bones, and observed a severe compromised craniofacial cartilage formation (Flores et al, 2006). The results obtained for Runx3 knockdown in zebrafish seem to be in agreement with the results observed for mouse knockout studies. However, so far no functional studies reported the effect of zebrafish Runx3 depletion in developing DRG sensory neurons, maintaining the role of Runx3 in these cells unclear. It is important to note that the importance of the different isoforms in zebrafish is still unknown, since the studies using morpholinos to target Runx3 only affected the isoforms derived from *runx3*-P2 and not those derived from the *runx3*-P1.

1.10.3.4 RUNX3 involvement in disease

RUNX3, as well as RUNX1, have been shown to be expressed in hematopoietic tissues and are frequently associated with leukemias, although the basis for the involvement of RUNX3 in hematopoiesis and leukemogenesis is still not fully understood (Crosier et al, 2002; Wang et al, 2013; 2014). A study published more than a decade ago, where the expression of RUNX3 was analysed in a diversity of mammalian cell lines, showed that RUNX3 is expressed predominantly in cells of hematopoietic origin, and non-hematopoietic cell lines that either do not express or express very low levels of RUNX3 protein (Le et al, 1999). In that report, the authors also showed that RUNX3 is induced in a human myeloid leukemia cell line through the retinoic acid signalling pathway, suggesting that RUNX3 may play a role in hematopoietic cell differentiation (Le et al, 1999). Recently, a Runx3 knockout study performed in mouse reveal a mild expansion of myeloid cells and hematopoietic stem cells

when aged (Wang et al, 2013) and a study using Runx1/Runx3 double-knockout mice showed that disruption of both genes leads to bone marrow failure and leukemia predisposition due to transcriptional and DNA repair defects (Wang et al, 2014).

In thymopoiesis, RUNX3 is known as an important regulator of T-cell differentiation and has also been shown to be highly expressed in dendritic cells, where it functions as a component of the transforming growth factor β (TGF-β) signalling cascade (Fainaru et al, 2004). Functional studies of Runx3 depletion revealed that the dendritic cells of the Runx3 KO mice do not respond to TGF-β and that loss of leukocytic cell-autonomous function of Runx3 results in spontaneous development of inflammatory bowel disease (IBD) and gastric lesion (Brenner et al, 2004). Moreover, RUNX3 and TGF-β are down-regulated in peripheral blood cells of Crohn's disease patients, which might suggest involvement of this pathway in the human pathogenesis of IBD (Weersma et al, 2008). More recently, RUNX3 has been shown to be implicated in the pathogenesis of ankylosing spondylitis (Zeng et al, 2013) and to be an additional susceptibility locus for psoriatic arthritis (Apel et al, 2013).

Similar to the other RUNX family members, RUNX3 has also been implicated in a variety of cancers where it can either function as a tumour suppressor or an oncogene, depending on the type of cancer (Li et al, 2002; Lund and van Lohuizen, 2002; Reviewed in Subramaniam et al, 2009). Although for more than a decade different laboratories have been focusing their research in evidences suggesting that lack of *RUNX3* expression is related to the development of a diversity of cancers, a recent paper that analysed published results based on whole-genome and whole-exome DNA sequencing studies on thousands of samples from different cancers, did not find *RUNX3* to be a gene significantly mutated in cancer (reviewed in Lotem et al, 2015). In fact, contradictions about *RUNX3* being expressed in gastrointestinal tract epithelium and functioning as a tumour suppressor have been disputed over a decade between different groups (Normile, 2011; Ito, 2012; Levanon et al, 2012). At the moment it is not completely defined how RUNX3 is related to cancer development.

1.11 Using zebrafish as a model organism to study gene regulation and function

Due to an increasing concern about animal welfare, when using animals in research, it is important to minimise animal suffering by using the least sentient organism possible to answer the question at hand and applying the 3R's rule (replacement, refinement and reduction; https://www.nc3rs.org.uk/). Zebrafish (*Danio rerio*) is a small (4-5 cm in length) tropical fish native from the rivers of northern and eastern India (Engeszer et al, 2007), that has long been a common feature in home aquariums, and shown in the last years to be an attractive research tool. Zebrafish has revealed several advantages over other models thus rapidly becoming an excellent model for biomedical studies, being widely used in research as an alternative to mammalian species.

A fundamental advantage in the use of zebrafish is that they are vertebrates and therefore share a considerable amount of genetic identity with mammals, including humans (Postlethwait et al, 2000). It was also observed that several zebrafish organ systems are remarkably similar to those in humans (Seth et al, 2013) and orthologues of most human genes and proteins can be found in zebrafish and show similar patterns of expression and a high conservation of amino acid residue sequence (Howe et al, 2013), while maintaining their cell biological and developmental processes conserved. Therefore, studies in fish can give great insight into human disease processes. The ability to accelerate genetic studies by gene knockdown or overexpression have led to the widespread use of zebrafish in the detailed investigation of vertebrate gene function and increasingly, the study of human genetic diseases (Grunwald and Eisen, 2002). Recently, the development of new mutagenesis techniques has increased exponentially the number of available zebrafish models of human genetic diseases. That, together with the recent availability of the complete zebrafish genome sequence, increased the ability to screen for modifier mutations (that suppress or enhance the disease phenotype) and allows dissection of the biological processes underlying the disease.

Other advantages of zebrafish as a model in research, compared to other vertebrate models, include their easy and relatively low cost maintenance in laboratory, easy breed and large number of offspring that can be obtained from a single spawning and their develop is *ex utero*, allowing easy access to the embryos. Additional advantages are the short generation

time and the fact that the young fish are small and relatively transparent which enables researchers to make microscopic observations in the living animal (Westerfield, 2000). The genetic tractability, the new transgenic and mutagenic tools available and the optical transparency of zebrafish makes it the vertebrate model system par excellence for *in vivo* analysis.

1.12 Main objectives

Although in recent years the number of published papers focusing on the study of RUNX3 in different aspects of its function has increased substantially, the exact mechanisms of action of RUNX3 proteins during development is still not completely understood. Even less clear is the mechanism of how *RUNX3* transcription is regulated and the implication for function of the complex variety of isoforms that are produced by this gene.

To characterise the variety of transcript variants and the encoded protein isoforms of the *runx3* gene and to investigate possible roles in zebrafish development, we propose to: **(i)** Identify and clone *runx3*-P1 and *runx3*-P2 transcript variants; **(ii)** Determine the expression pattern of *runx3* variants both during the embryonic development and in adult tissues; **(iii)** Analyse both *runx3* promoter regions using an *in silico*, *in vitro* and *in vivo* approach; **(iv)** Analyse the possibility of cross-regulation of the *runx3* promoter regions by Runx2 isoforms, as well as the auto-regulation by its own isoforms.

1.13 References

Ahn MY, Bae SC, Maruyama M, Ito Y, 1996. Comparison of the human genomic structure of the runt domain-encoding PEBP2/CBFα gene family. Gene. 168:279-280.

Apel M, Uebe S, Bowes J, Giardina E, Korendowych E, Juneblad K, Pasutto F, Ekici AB, McManus R, Ho P, Bruce IN, Ryan AW, Behrens F, Böhm B, Traupe H, Lohmann J, Gieger C, Wichmann HE, Padyukov L, Fitzgerald O, Alenius GM, McHugh NJ, Novelli G, Burkhardt H, Barton A, Reis A, Hüffmeier U, 2013. Variants in RUNX3 contribute to susceptibility to psoriatic arthritis, exhibiting further common ground with ankylosing spondylitis. Arthritis Rheum. 65:1224-1231.

Avraham KB, Levanon D, Negreanu V, Bernstein Y, Groner Y, Copeland NG, Jenkins NA, 1995. Mapping of the runt domain gene, Aml2, to the distal region of mouse chromosome 4. Genomics. 25:603-605.

Bäckström S, Wolf-Watz M, Grundström C, Härd T, Grundström T, Sauer UH, 2002. The RUNX1 Runt domain at 1.25A resolution: a structural switch and specifically bound chloride ions modulate DNA binding. J Mol Biol. 322:259-272.

Bae S-C, Takahashi E-i, Zhang YW, Ogawa E, Shigesada K, Namba Y, Satake M, Ito Y, 1995. Cloning, mapping and expression of PEBP2aC, a third gene encoding the mammalian runt domain. Gene. 159:245-248.

Bangsow C, Rubins N, Glusman G, Bernstein Y, Negreanu V, Goldenberg D, Lotem J, Ben-Asher E, Lancet D, Levanon D, Groner Y, 2001. The RUNX3 gene-sequence, structure and regulated expression. Gene. 279:221-232.

Bao R, Friedrich M, 2008. Conserved cluster organization of insect Runx genes. Dev Genes Evol. 218:567-574.

Bataillé L, Augé B, Ferjoux G, Haenlin M, Waltzer L, 2005. Resolving embryonic blood cell fate choice in Drosophila: interplay of GCM and RUNX factors. Development. 132:4635-4644.

Bernardin-Fried F, Kummalue T, Leijen S, Collector MI, Ravid K, Friedman AD, 2004. AML1/RUNX1 increases during G1 to S cell cycle progression independent of cytokine-dependent phosphorylation and induces cyclin D3 gene expression. J Biol Chem. 279:15678-15687.

Bertrand-Philippe M, Ruddell RG, Arthur MJ, Thomas J, Mungalsingh N, Mann DA, 2004. Regulation of tissue inhibitor of metalloproteinase 1 gene transcription by RUNX1 and RUNX2. J Biol Chem. 279:24530-24539.

Blyth K, Cameron ER, Neil JC, 2005. The runx genes: gain or loss of function in cancer. Nat Rev Cancer. 5:376-387.

Brady G, Farrell PJ, 2009. RUNX3-Mediated Repression of RUNX1 in B Cells. J Cell Physiol. 221:283-287.

Bravo J, Li Z, Speck NA, Warren AJ, 2001. The leukemia-associated AML1 (Runx1)-CBF beta complex functions as a DNA-induced molecular clamp. Nat Struct Biol. 8:371-378

Brenner O, Levanon D, Negreanu V, Golubkov O, Fainaru O, Woolf E, Groner Y, 2004. Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. Proc Natl Acad Sci U S A. 101:16016-16021.

Bruno L, Mazzarella L, Hoogenkamp M, Hertweck A, Cobb BS, Sauer S, Hadjur S, Leleu M, Naoe Y, Telfer JC, Bonifer C, Taniuchi I, Fisher AG, Merkenschlager M, 2009. Runx proteins regulate Foxp3 expression. J Exp Med. 206:2329-2337.

Burns CE, Traver D, Mayhall E, Shepard JL, Zon L, 2005. Hematopoietic stem cell fate is established by the Notch-Runx pathway. Genes Dev. 19:2331-2342.

Cai Z, de Bruijn M, Ma X, Dortland B, Luteijn T, Downing RJ, Dzierzak E, 2000. Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. Immunity. 13:423-431.

Canon J, Banerjee U, 2000. Runt and Lozenge function in Drosophila development. Semin Cell Dev Biol. 11:327-336.

Carter R, Drouin G, 2009. Structural differentiation of the three eukaryotic RNA polymerases. Genomics. 94:388-396.

Chae SC, Park BL, Park CS, Ryu HJ, Yang YS, Lee SO, Choi YH, Kim EM, Uh ST, Kim YH, Kim KK, Oh B, Chung HT, Kimm K, Shin HD, 2006. Putative association of RUNX1 polymorphisms with IgE levels in a Korean population. Exp Mol Med. 38:583-588.

Chuang LSH, Ito K, Ito Y, 2013. RUNX family: Regulation and diversification of roles through interacting proteins. Int J Cancer. 132:1260-1271.

Coffman JA, 2003. Runx transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol Int. 27:315-324.

Cohen MM Jr, 2009. Perspectives on RUNX genes: An update. Am J Med Genet Part A 149A:2629-2646.

Collingwood TN, Urnov FD, Wolffe AP, 1999. Nuclear receptors: coactivators, corepressors and chromatin remodeling in the control of transcription. J Mol Endocrinol. 23:255-275.

Cornell RA, Eisen JS, 2000. Delta signaling mediates segregation of neural crest and spinal sensory neurons from zebrafish lateral neural plate. Development. 127:2873-2882.

Crosier PS, Kalev-Zylinska ML, Hall CJ, Flores MV, Horsfield JA, Crosier KE, 2002. Pathways in blood and vessel development revealed through zebrafish genetics. Int J Dev Biol. 46:493-502.

Daga A, Karlovich CA, Dumstrei K, Banerjee U, 1996. Patterning of cells in the Drosophila eye by Lozenge, which shares homologous domains with AML1. Genes Dev. 10:1194-1205.

Dicker F, Haferlach C, Kern W, Haferlach T, Schnittger S, 2007. Trisomy 13 is strongly associated with AML1/RUNX1 mutations and increased FLT3 expression in acute myeloid leukemia. Blood. 110:1308-1316.

Domínguez-Soto A, Relloso M, Vega MA, Corbí AL, Puig-Kröger A, 2005. RUNX3 regulates the activity of the CD11a and CD49d integrin gene promoters. Immunobiology. 210:133-139.

Downing JR, 1999. The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance. Br J Haematol. 106(2):296-308.

Drissi H, Luc Q, Shakoori R, Chuva De Sousa Lopes S, Choi JY, Terry A, Hu M, Jones S, Neil JC, Lian JB, Stein JL, Van Wijnen AJ, Stein GS, 2000. Transcriptional autoregulation of the bone related CBFA1/RUNX2 gene. J Cell Physiol. 184:341-350.

Ducy P, Zhang R, Geoffroy V, Ridall A, Karsenty G, 1997. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. Cell. 89:747-754.

Duffy JB, Gergen JP, 1991. The Drosophila segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene Sex-lethal. Genes Dev. 5:2176-2187.

Duffy, JB, Kania MA, Gergen JP, 1991. Expression and function of the Drosophila gene runt in early stages of neural development. Development. 1230:1223-1230.

Engeszer RE, Patterson LB, Rao AA, Parichy DM, 2007. Zebrafish in the wild: a review of natural history and new notes from the field. Zebrafish. 4:21-40.

Enomoto H, Enomoto-Iwamoto M, Iwamoto M, Nomura S, Himeno M, Kitamura Y, Kishimoto T, Komori T, 2000. Cbfa1 is a positive regulatory factor in chondrocyte maturation. J Biol Chem. 275:8695-8702.

Fainaru O, Woolf E, Lotem J, Yarmus M, Brenner O, Goldenberg D, Negreanu V, Bernstein Y, Levanon D, Jung S, Groner Y, 2004. Runx3 regulates mouse TGF-beta-mediated dendritic cell function and its absence results in airway inflammation. EMBO J. 23:969-979.

Fazenda C, Simões B, Kelsh RN, Cancela ML, Conceição N, 2010. Dual transcriptional regulation by runx2 of matrix Gla protein in *Xenopus laevis*. Gene. 450:94-102.

Flores GV, Daga A, Kalhor HR, Banerjee U, 1998. Lozenge is expressed in pluripotent precursor cells and patterns multiple cell types in the Drosophila eye through the control of cell-specific transcription factors. Development. 125:3681-3687.

Flores MV, Lam EY, Crosier P, Crosier K, 2006. A hierarchy of Runx transcription factors modulate the onset of chondrogenesis in craniofacial endochondral bones in zebrafish. Dev Dyn. 235:3166-3176.

Fickett JW, Wasserman WW, 2000. Discovery and modeling of transcriptional regulatory regions. Curr Opin Biotechnol. 11:19-24.

Gao YH, Shinki T, Yuasa T, Kataoka-Enomoto H, Komori T, Suda T, Yamaguchi A, 1998. Potential role of cbfa1, an essential transcriptional factor for osteoblast differentiation, in osteoclastogenesis: regulation of mRNA expression of osteoclast differentiation factor (ODF). Biochem Biophys Res Commun. 252:697-702.

Geoffroy V, Corral DA, Zhou L, Lee B, Karsenty G, 1998. Genomic organization, expression of the human CBFA1 gene, and evidence for an alternative splicing event affecting protein function. Mamm Genome. 9:54-57.

Gergen JP, Butler BA, 1988. Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. Genes Dev. 2:1179-1193.

Glover JNM, Harrison SC, 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature. 373:257-261.

Golan I, Preising M, Wagener H, Baumert U, Niederdellmann H, Lorenz B, Müssig D, 2000. A novel missense mutation of the CBFA1 gene in a family with cleidocranial dysplasia (CCD) and variable expressivity. J Craniofac Genet Dev Biol. 20:113-120.

Grunwald DJ, Eisen JS, 2002. Headwaters of the zebrafish - emergence of a new model vertebrate. Nat Rev Genet. 3:717-724.

Gupta BP, Rodrigues V, 1995. Distinct mechanisms of action of the Lozenge locus in Drosophila eye and antennal development are suggested by the analysis of dominant enhancers. J Neurogenet. 10:137-151.

Harada H, Tagashira S, Fujiwara M, Ogawa S, Katsumata T, Yamaguchi A, Komori T, Nakatsuka M, 1999. Cbfa1 isoforms exert functional differences in osteoblast differentiation. J Biol Chem. 274:6972-6978.

Hart SM, Foroni L, 2002. Core binding factor genes and human leukemia. Haematologica. 87:1307-1323.

Hecht J, Stricker S, Wiecha U, Stiege A, Panopoulou G, Podsiadlowski L, Poustka AJ, Dieterich C, Ehrich S, Suvorova J, Mundlos S, Seitz V, 2008. Evolution of a core gene network for skeletogenesis in Chordates. PLoS Genet. 4:e1000025.

Helms C, Cao L, Krueger JG, Wijsman EM, Chamian F, Gordon D, Heffernan M, Daw JA, Robarge J, Ott J, Kwok PY, Menter A, Bowcock AM, 2003. A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nat Genet. 35:349-56.

Horsfield JA, Anagnostou SH, Hu JK, Cho KH, Geisler R, Lieschke G, Crosier KE, Crosier PS, 2007. Cohesin-dependent regulation of Runx genes. Development. 134:2639-2649.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, et al., 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 496:498-503.

Hu CD, Chinenov Y, Kerppola TK, 2002. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell. 9:789-798.

Huang G, Shigesada K, Ito K, Wee HJ, Yokomizo T, Ito Y, 2001. Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. EMBO J. 20:723-733.

Inada M, Yasui T, Nomura S, Miyake S, Deguchi K, Himeno M, Sato M, Yamagiwa H, Kimura T, Yasui N, Ochi T, Endo N, Kitamura Y, Kishimoto T, Komori T, 1999. Maturational disturbance of chondrocytes in Cbfa1-deficient mice. Dev Dyn. 214:279-290.

Inoue K, Ozaki S, Ito K, Iseda T, Kawaguchi S, Ogawa M, Bae SC, Yamashita N, Itohara S, Kudo N, Ito Y. 2003. Runx3 is essential for the target-specific axon pathfinding of trkc-expressing dorsal root ganglion neurons. Blood Cells Mol Dis. 30:157-160.

Inoue K, Ozaki S, Shiga T, Ito K, Masuda T, Okado N, Iseda T, Kawaguchi S, Ogawa M, Bae SC, Yamashita N, Itohara S, Kudo N, Ito Y, 2002. Runx3 controls the axonal projection of proprioceptive dorsal root ganglion neurons. Nat Neurosci. 5:946-954.

Ito K, Lim AC, Salto-Tellez M, Motoda L, Osato M, Chuang LS, Lee CW, Voon DC, Koo JK, Wang H, Fukamachi H, Ito Y, 2008. RUNX3 attenuates b-catenin/T cell factors in intestinal tumorigenesis. Cancer Cell. 14:226-237.

Ito K, Liu Q, Salto-Tellez M, Yano T, Tada K, Ida H, Huang C, Shah N, Inoue M, Rajnakova A, Hiong KC, Peh BK, Han HC, Ito T, Teh M, Yeoh KG, Ito Y, 2005. RUNX3, a novel tumor suppressor, is frequently inactivated in gastric cancer by protein mislocalization. Cancer Res. 65:7743-7750.

Ito Y, 1999. Molecular basis of tissue-specific gene expression mediated by the runt domain transcription factor PEBP2/CBF. Genes Cells. 4:685-696.

Ito Y, 2004. Oncogenic potential of the RUNX gene family. Oncogene. 23:4198-4208.

Ito Y, 2012. RUNX3 is expressed in the epithelium of the gastrointestinal tract. EMBO Mol Med. 4:541-542.

Ito Y, Miyazono K, 2003. RUNX transcription factors as key targets of TGF-superfamily signaling. Curr Opin Genet Dev. 13:43-47.

Iwatsuki K, Tanaka K, Kaneko T, Kazama R, Okamoto S, Nakayama Y, Ito Y, Satake M, Takahashi S, Miyajima A, Watanabe T, Hara T, 2005. Runx1 promotes angiogenesis by downregulation of insulin-like growth factor-binding protein-3. Oncogene. 24:1129-1137.

Jin H, Sood R, Xu J, Zhen F, English MA, Liu PP, Wen Z, 2009. Definitive hematopoietic stem/progenitor cells manifest distinct differentiation output in the zebrafish VDA and PBI. Development. 136:647-54.

Kagoshima H, Shigesada K, Kohara Y. 2007. RUNX regulates stem cell proliferation and differentiation: insights from studies of C. elegans. J Cell Biochem. 100:1119-1130.

Kagoshima H, Shigesada K, Satake M, Ito Y, Miyoshi H, Ohki M, Pepling M, Gergen P, 1993. The Runt domain identifies a new family of heteromeric transcriptional regulators. Trends Genet. 9:338-341.

Kalev-Zylinska ML, Horsfield JA, Flores MV, Postlethwait JH, Chau JY, Cattin PM, Vitas MR, Crosier PS, Crosier KE, 2003. Runx3 is required for hematopoietic development in zebrafish. Dev Dyn. 228:323-336.

Kalev-Zylinska ML, Horsfield JA, Flores MV, Postlethwait JH, Vitas MR, Baas AM, Crosier PS, Crosier KE, 2002. Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. Development. 129:2015-2030.

Kania MA, Bonner AS, Duffy JB, Gergen JP, 1990. The Drosophila segmentation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Genes Dev. 4:1701-1713.

Kanno Y, Kanno T, Sakakura C, Bae SC, Ito Y, 1998.Cytoplasmic sequestration of the polyomavirus enhancer binding protein 2 (PEBP2)/core binding factor alpha (CBFalpha) subunit by the leukemia-related PEBP2/CBFbeta-SMMHC fusion protein inhibits PEBP2/CBF-mediated transactivation. Mol Cell Biol. 18:4252-4261.

Kataoka H, Ochi M, Enomoto K, Yamaguchi A, 2000. Cloning and embryonic expression patterns of the zebrafish Runt domain genes, runxa and runxb. Mech Dev. 98:139-143.

Khalid O, Baniwal SK, Purcell DJ, Leclerc N, Gabet Y, Stallcup MR, Coetzee GA, Frenkel B, 2008. Modulation of Runx2 activity by estrogen receptor-alpha: implications for osteoporosis and breast cancer. Endocrinology. 149:5984-5995.

Kim JH, Lee S, Rho JK, Choe SY, 1999. AML1, the target of chromosomal rearrangements in human leukemia, regulates the expression of human complement receptor type 1 (CR1) gene. Int J Biochem Cell Biol. 31:933-940.

Komori T, 2005. Regulation of Skeletal Development by the Runx Family of Transcription Factors. J Cell Biochem. 95:445-453.

Komori TH, Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, Shimizu Y, Bronson RT, Gao YH, Inada M, Sato M, Okamoto R, Kitamura Y, Yoshiki S, Kishimoto T, 1997. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell. 89:755-764.

Le XF, Groner Y, Kornblau SM, Gu Y, Hittelman WN, Levanon D, Mehta K, Arlinghaus RB, Chang KS, 1999. Regulation of AML2/CBFA3 in hematopoietic cells through the retinoic acid receptor alpha-dependent signaling pathway. J Biol Chem. 274:21651-21658.

Lee TI, Young RA, 2000. Transcription of eukaryotic protein-coding genes. Annu Rev Genet. 34:77-137.

Levanon D, Bernstein Y, Negreanu V, Ghozi MC, Baram I, Aloya R, Goldenberg D, Lotem J, Groner Y, 1996. A large variety of alternatively spliced and differentially expressed messenger RNAs are encoded by the human acute myeloid leukemia gene AML1. DNA Cell Biol. 15:175-185.

Levanon D, Bettoun D, Harris-Cerruti C, Woolf E, Negreanu V, Eilam R, Bernstein Y, Goldenberg D, Xiao C, Fliegauf M, Kremer E, Otto F, Brenner O, Lev-Tov A, Groner Y, 2002. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. EMBO J. 21:3454-3463.

Levanon D, Brenner O, Negreanu V, Bettoun D, Woolf E, Eilam R, Lotem J, Gat U, Otto F, Speck N, Groner Y, 2001. Spatial and temporal expression pattern of Runx3 (Aml2) and Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis. Mech Dev. 109:413-417.

Levanon D, Groner Y, 2004. Structure and regulated expression of mammalian RUNX genes. Oncogene. 23:4211-4219.

Levanon D, Groner Y, 2009. Runx3-deficient mouse strains circa 2008: resemblance and dissimilarity. Blood Cells Mol Dis. 43:1-5.

Levanon D, Negreanu V, Bernstein Y, Bar-Am I, Avivi L, Groner Y, 1994. AML1, AML2, and AML3, the human members of the runt domain gene-family: cDNA structure, expression, and chromosomal localization. Genomics. 23:425-432.

Levanon D, Negreanu V, Lotem J and Groner Y, 2012. Author reply to: RUNX3 is expressed in the epithelium of the gastrointestinal tract. EMBO Mol Med. 4:543-544.

Levine M, Tjian R, 2003. Transcription regulation and animal diversity. Nature. 424:147-151.

Li QL, Ito K, Sakakura C, Fukamachi H, Inoue Ki, Chi XZ, Lee KY, Nomura S, Lee CW, Han SB, Kim HM, Kim WJ, Yamamoto H, Yamashita N, Yano T, Ikeda T, Itohara S, Inazawa J, Abe T, Hagiwara A, Yamagishi H, Ooe A, Kaneda A, Sugimura T, Ushijima T, Bae SC, Ito Y, 2002. Causal relationship between the loss of RUNX3 expression and gastric cancer. Cell. 109:113-124.

Li Y-L, Xiao Z-S, 2007. Advances in Runx2 regulation and its isoforms. Medical Hypotheses. 68:169-175.

Liu H, Carlsson L, Grundstrom T, 2006. Identification of an N-terminal transactivation domain that separates molecular function from global differentiation function. J Biol Chem. 281:25659-25669.

Liu W, Toyosawa S, Furuichi T, Kanatani N, Yoshida C, Liu Y, Himeno M, Narai S, Yamaguchi A, Komori T, 2001. Overexpression of Cbfa1 in osteoblasts inhibits osteoblast maturation and causes osteopenia with multiple fractures. J Cell Biol. 155:157-166.

Liu Z, Zhang X, Xu X, Chen L, Li W, Yu H, Sun Y, Zeng J, Jia J, 2014. RUNX3 inhibits survivin expression and induces cell apoptosis in gastric cancer. Eur J Cell Biol. 93:118-126.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J, 2000. Molecular Cell Biology. 4th edition. New York: W. H. Freeman.

Lorsbach RB, Moore J, Ang SO, Sun W, Lenny N, Downing JR, 2004. Role of RUNX1 in adult hematopoiesis: analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression. Blood. 103:2522-2529.

Lotem J, Levanon D, Negreanu V, Bauer O, Hantisteanu S, Dicken J, Groner Y, 2015. Runx3 at the interface of immunity, inflammation and cancer. Biochim Biophys Acta. 1855:131-143.

Lund AH, van Lohuizen M, 2002. RUNX: a trilogy of cancer genes. Cancer cell. 1:213-215.

Mertz JA, Mustafa F, Meyers S, Dudley JP, 2001. Type B leukemogenic virus has a T-cell-specific enhancer that binds AML-1. J Virol. 75:2174-2184.

Miller J, Horner A, Stacy T, Lowrey C, Lian JB, Stein G, Nuckolls GH, Speck NA, 2002. The core-binding factor β subunit is required for bone formation and hematopoietic maturation. Nat Genet. 32:645-649.

Miyazono K, Maeda S, Imamura T, 2004. Coordinate regulation of cell growth and differentiation by TGF-beta superfamily and Runx proteins. Oncogene. 23:4232-4237.

Miyoshi H, Ohira M, Shimizu K, Mitani K, Hirai H, Imai T, Yokoyama K, Soeda E, Ohki M, 1995. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. Nucleic Acids Res. 23:2762-2769.

Miyoshi H, Shimizu K, Kozu T, Maseki N, Kaneko Y, Ohki M, 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc Natl Acad Sci USA. 88:10431-10434.

Moosavi SA, Sanchez J, Adeyinka A, 2009. Marker chromosomes are a significant mechanism of high-level RUNX1 gene amplification in hematologic malignancies. Cancer Genet Cytogenet. 189:24-28.

Nagata T, Gupta V, Sorce D, Kim WY, Sali A, Chait BT, Shigesada K, Ito Y, Werner MH, 1999. Immunoglobulin motif DNA recognition and heterodimerization of the PEBP2/CBF Runt domain. Nat Struct Biol. 6:615-619.

Nah GS, Lim ZW, Tay BH, Osato M, Venkatesh B, 2014a. Runx family genes in a cartilaginous fish, the elephant shark (Callorhinchus milii). PLoS One. 9:e93816.

Nah GS, Tay BH, Brenner S, Osato M, Venkatesh B, 2014b. Characterization of the Runx gene family in a jawless vertebrate, the Japanese lamprey (Lethenteron japonicum). PLoS One. 9:e113445.

Nakamura S, Senzaki K, Yoshikawa M, Nishimura M, Inoue K, Ito Y, Ozaki S, Shiga T, 2008. Dynamic regulation of the expression of neruotrophin receptors by Runx3. Development. 135:1703-1711.

Ng CE, Osato M, Tay BH, Venkatesh B, Ito Y, 2007. cDNA cloning of Runx family genes from the pufferfish (Fugu rubripes). Gene. 399:162-173.

Nomura Y, Tanaka Y, Fukunaga J-i, Fujiwara K, Chiba M, Iibuchi H, Tanaka T, Nakamura Y, Kawai G, Kozu T and Sakamoto T, 2013. Solution structure of a DNA mimicking motif of an RNA aptamer against transcription factor AML1 Runt domain. J Biochem. 154:513-519.

Normile D, 2011. Cancer research. Dispute over tumor suppressor gene Runx3 boils over. Science. 334:442-443.

North T, Gu TL, Stacy T, Wang Q, Howard L, Binder M, Marin-Padilla M, Speck NA, 1999. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. Development. 126:2563-2575.

North TE, de Bruijn MF, Stacy T, Talebian L, Lind E, Robin C, Binder M, Dzierzak E, Speck NA, 2002. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. Immunity. 16:661-672.

Nüsslein-Volhard C, Wieschaus E, 1980. Mutations affecting segment number and polarity in Drosophila. Nature. 287:795-801.

Oakford PC, James SR, Qadi A, West AC, Ray SN, Bert AG, Cockerill PN, Holloway AF, 2010. Transcriptional and epigenetic regulation of the GM-CSF promoter by RUNX1. Leuk Res. 34:1203-1213.

Ogawa E, Inuzuka M, Maruyama M, Satake M, Naito-Fujimoto M, Ito Y, Shigesada K, 1993. Molecular cloning and characterization of PEBP2β, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2α. Virology. 194:314-331.

Ohno S, Sato T, Kohu K, Takeda K, Okumura K, Satake M, Habu S, 2008. Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation. Int Immunol. 20:71-79.

Okuda T, Nishimura M, Nakao M, Fujitaa Y, 2001. RUNX1/AML1: A Central Player in Hematopoiesis. Int J Hematol. 74:252-257.

Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR, 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell. 84:321-330.

Otto F, Libbert M, Stock M, 2003. Upstream and Downstream Targets of RUNX Proteins. J Cell Biochem. 18:9-18.

Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GW, Beddington RS, Mundlos S, Olsen BR, Selby PB, Owen MJ, 1997. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell. 89:765-771.

Pande S, Ali SA, Dowdy C, Zaidi SK, Ito K, Ito Y, Montecino MA, Lian JB, Stein JL, van Wijnen AJ, Stein GS, 2009. Subnuclear targeting of the Runx3 tumor suppressor and its epigenetic association with mitotic chromosomes. J Cell Physiol. 218:473-479.

Park B-Y, Saint-Jeannet J-P, 2010. Expression analysis of Runx3 and other Runx family members during Xenopus development. Gene Expr Patterns. 10: 159-166.

Patikoglou G, Burley SK, 1997. Eukaryotic transcription factor-DNA complexes. Annu Rev Biophys Biomol Struct. 26:289-325.

Pinto JP, Conceição NM, Viegas CS, Leite RB, Hurst LD, Kelsh RN, Cancela ML, 2005. Identification of a new pebp2alphaA2 isoform from zebrafish runx2 capable of inducing osteocalcin gene expression in vitro. J Bone Miner Res. 20:1440-1453.

Pockwinse SM, Rajgopal A, Young DW, Mujeeb KA, Nickerson J, Javed A, Redick S, Lian JB, van Wijnen AJ, Stein JL, Stein GS, Doxsey SJ, 2006. Microtubule-Dependent Nuclear-Cytoplasmic Shuttling of Runx2. J cellular Physiol. 206:354-362.

Podgornik H, Debeljak M, Zontar D, Cernelc P, Prestor VV, Jazbec J, 2007. RUNX1 amplification in lineage conversion of childhood B-cell acute lymphoblastic leukemia to acute myelogenous leukemia. Cancer Genet Cytogenet. 178:77-81.

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS, 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res. 10:1890-1902.

Pugh BF, Tjian R, 1991. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. 5:1935-1945.

Puig-Kröger A, López-Rodriguez C, Relloso M, Sánchez-Elsner T, Nueda A, Muñoz E, Bernabéu C, Corbi AL, 2000. Polyomavirus enhancer-binding protein 2/core binding factor/acute myeloid leukemia factors contribute to the cell type-specific activity of the CD11a integrin gene promoter. J Biol Chem. 275:28507-28512.

Rennert J, Coffman JA, Mushegian AR, Robertson AJ, 2003. The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms. BMC Evol Biol. 3:4.

Reyes R, Haendel M, Grant D, Melancon E, Eisen JS, 2004. Slow degeneration of zebrafish Rohon-Beard neurons during programmed cell death. Dev Dyn. 229:30-41.

Rini D, Calabi F, 2001. Identification and comparative analysis of a second runx3 promoter. Gene. 273:13-22.

Roeder RG, 1996. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci. 21:327-335.

Rossi CC, Kaji T, Artinger KB, 2009. Transcriptional control of Rohon-Beard sensory neuron development at the neural plate border. Dev Dyn. 238: 931-943.

Sadikovic B, Thorner P, Chilton-Macneill S, Martin JW, Cervigne NK, Squire J, Zielenska M, 2010. Expression analysis of genes associated with human osteosarcoma tumors shows correlation of RUNX2 overexpression with poor response to chemotherapy. BMC Cancer. 10:202.

Sakakura C, Hagiwara A, Miyagawa K, Nakashima S, Yoshikawa T, Kin S, Nakase Y, Ito K, Yamagishi H, Yazumi S, Chiba T, Ito Y, 2005. Frequent downregulation of the runt domain transcription factors RUNX1, RUNX3 and their cofactor CBFB in gastric cancer. Int J Cancer. 113:221-228.

San Martin IA, Varela N, Gaete M, Villegas K, Osorio M, Tapia JC, Antonelli M, Mancilla EE, Pereira BP, Nathan SS, Lian JB, Stein JL, Stein GS, van Wijnen AJ, Galindo M, 2009. Impaired cell cycle regulation of the osteoblast-related heterodimeric transcription factor Runx2-Cbfbeta in osteosarcoma cells. J Cell Physiol. 221:560-571.

Satake M, Nomura S, Yamaguchi-Iwai Y, Takahama Y, Hashimoto Y, Niki M, Kitamura Y, Ito Y, 1995. Expression of the Runt domain-encoding PEBP2 alpha genes in T cells during thymic development. Mol Cell Biol. 15:1662-1670.

Sato S, Kimura A, Ozdemir J, Asou Y, Miyazaki M, Jinno T, Ae K, Liu X, Osaki M, Takeuchi Y, Fukumoto S, Kawaguchi H, Haro H, Shinomiya K, Karsenty G, Takeda S, 2008. The Distinct Role of the Runx Proteins in Chondrocyte Differentiation and Intervertebral Disc Degeneration. Arthritis Rheum. 58:2764-2775.

Sauter KA, Bouhlel MA, O'Neal J, Sester DP, Tagoh H, Ingram RM, Pridans C, Bonifer C, Hume DA, 2013. The function of the conserved regulatory element within the second intron of the mammalian Csf1r locus. PLoS One. 8(1):e54935.

Seo W, Ikawa T, Kawamoto H, Taniuchi I, 2012. Runx1-Cbfβ facilitates early B lymphocyte development by regulating expression of Ebf1. J Exp Med. 209:1255-1262.

Seth A, Stemple DL, Barroso I. 2013. The emerging use of zebrafish to model metabolic disease. Dis Model Mech. 6:1080-1088.

Sevetson B, Taylor S, Pan Y, 2004. Cbfa1/RUNX2 directs specific expression of the sclerosteosis gene (SOST). J Biol Chem. 279:13849-13858.

Shen N, Tsao BP, 2004. Current advances in the human lupus genetics. Curr Rheumatol Rep. 6:391-398.

Shi X, Deepak V, Wang L, Ba X, Komori T, Zeng X, Liu W, 2013. Thrombospondin-1 is a putative target gene of Runx2 and Runx3. Int J Mol Sci. 14:14321-14332.

Simeone A, Daga A, Calabi F, 1995. Expression of runt in the mouse embryo. Dev Dyn. 203:61-70.

Simões B, Conceição N, Viegas CS, Pinto JP, Gavaia PJ, Hurst LD, Kelsh RN, Cancela ML, 2006. Identification of a promoter element within the zebrafish colXalpha1 gene responsive to runx2 isoforms Osf2/Cbfa1 and til-1 but not to pebp2alphaA2. Calcif Tissue Int. 79:230-244.

Sood R, English MA, Belele CL, Jin H, Bishop K, Haskins R, McKinney MC, Chahal J, Weinstein BM, Wen Z, Liu PP, 2010. Development of multilineage adult hematopoiesis in the zebrafish with a runx1 truncation mutation. Blood. 115:2806-2809

Soung do Y, Dong Y, Wang Y, Zuscik MJ, Schwarz EM, O'Keefe RJ, Drissi H, 2007. Runx3/AML2/Cbfa3 regulates early and late chondrocyte differentiation. J Bone Miner Res. 22:1260-1270.

Stein GS, Lian JB, van Wijnen AJ, Stein JL, Montecino M, Javed A, Zaidi SK, Young DW, Choi JY, Pockwinse SM, 2004. Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. Oncogene. 23:4315-4329.

Subramaniam MM, Chan JY, Yeoh KG, Quek T, Ito K, Salto-Tellez M, 2009. Molecular pathology of RUNX3 in human carcinogenesis. Biochim Biophys Acta. 1796:315-331.

Tahirov TH, Inoue-Bungo T, Morii H, Fujikawa A, Sasaki M, Kimura K, Shiina M, Sato K, Kumasaka T, Yamamoto M, Ishii S, Ogata K, 2001. Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. Cell. 104:755-767.

Takeda S, Bonnamy JP, Owen MJ, Ducy P, Karsenty G, 2001. Continuous expression of Cbfa1 in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues Cbfa1-deficient mice. Genes Dev. 15:467-481.

Taketani T, Taki T, Takita J, Tsuchida M, Hanada R, Hongo T, Kaneko T, Manabe A, Ida K, Hayashi Y, 2003. AML1/RUNX1 mutations are infrequent, but related to AML-M0, acquired trisomy 21, and leukemic transformation in pediatric hematologic malignancies. Genes Chromosomes Cancer. 38:1-7.

Taniuchi I, Osato M, Egawa T, Sunshine MJ, Bae SC, Komori T, Ito Y, Littman DR, 2002. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. Cell. 111:621-633.

Tracey WD, Pepling ME, Marko EH, Thomsen GH, Gerben JP, 1998. A Xenopus homologue of aml-1 reveals unexpected patterning mechanisms leading to the formation of embryonic blood. Development. 125:1371-1380.

Tsuchiya Y, Saito Y, Taniuchi S, Sakuma A, Maekawa T, Fukamachi H, Takeuchi S, Takahashi S, 2012. Runx3 expression and its roles in mouse endometrial cells. J Reprod Dev. 58:592-598.

Ueta C, Iwamoto M, Kanatani N, Yoshida C, Liu Y, Enomoto-Iwamoto M, Ohmori T, Enomoto H, Nakata K, Takada K, Kurisu K, Komori T, 2001. Skeletal malformations caused by overexpression of Cbfa1 or its dominant negative form in chondrocytes. J. Cell Biol. 153:87-99.

van Wijnen AJ, Stein GS, Gergen JP, Groner Y, Hiebert SW, Ito Y, Liu P, Neil JC, Ohki M, Speck N, 2004. Nomenclature for Runt-related (RUNX) proteins. Oncogene. 23:4209-4210.

Wang CQ, Krishnan V, Tay LS, Chin DW, Koh CP, Chooi JY, Nah GS, Du L, Jacob B, Yamashita N, Lai SK, Tan TZ, Mori S, Tanuichi I, Tergaonkar V, Ito Y, Osato M, 2014. Disruption of Runx1 and Runx3 leads to bone marrow failure and leukemia predisposition due to transcriptional and DNA repair defects. Cell Rep. 8:767-782.

Wang CQ, Motoda L, Satake M, Ito Y, Taniuchi I, Tergaonkar V, Osato M. 2013. Runx3 deficiency results in myeloproliferative disorder in aged mice. Blood. 122:562-566.

Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, Speck NA, 1996a. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci USA. 93:3444-3449.

Wang Q, Stacy T, Miller JD Lewis AF, Gu TL, Huang X, Bushweller JH, Bories JC, Alt FW, Ryan G, Liu PP, Wynshaw-Boris A, Binder M, Marín-Padilla M, Sharpe AH, Speck NA, 1996b. The CBFbeta subunit is essential for CBFalpha2 (AML1) function *in vivo*. Cell. 87:697-708.

Wang S, Wang Q, Crute BE, Melnikova IN, Keller SR, Speck NA, 1993. Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. Mol Cell Biol. 13:3324-3339.

Warren AJ, Bravo J, Williams RL, Rabbitts TH, 2000. Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBFβ. EMBO J. 19:3004-3015.

Watanabe K, Sugai M, Nambu Y, Osato M, Hayashi T, Kawaguchi M, Komori T, Ito Y, Shimizu A, 2010. Requirement for Runx proteins in IgA class switching acting downstream of TGF-beta 1 and retinoic acid signaling. J immunol. 184:2785-2792.

Weersma RK, Zhou L, Nolte IM, van der Steege G, van Dullemen HM, Oosterom E, Bok L, Peppelenbosch MP, Faber KN, Kleibeuker JH, Dijkstra G, 2008. Runt-related transcription factor 3 is associated with ulcerative colitis and shows epistasis with solute carrier family 22, members 4 and 5. Inflamm Bowel Dis. 14:1615-1622.

Westerfield M, 2000. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene.

Wigner NA, Soung do Y, Einhorn TA, Drissi H, Gerstenfeld LC, 2013. Functional role of Runx3 in the regulation of aggrecan expression during cartilage development. J Cell Physiol. 228:2232-2242.

Wong WF, Kohu K, Chiba T, Sato T, Satake M, 2011. Interplay of transcription factors in T-cell differentiation and function: the role of Runx. Immunology. 132:157-164.

Woolf E, Xiao C, Fainaru O, Lotem J, Rosen D, Negreanu V, Bernstein Y, Goldenberg D, Brenner O, Berke G, Levanon D, Groner Y, 2003. Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. Proc Natl Acad Sci USA. 100:7731-7736.

Xiao ZS, Simpson LG, Quarles LD, 2003. IRES-dependent translational control of Cbfa1/Runx2 expression. J Cell Biochem. 88:493-505.

Yamachika E, Tsujigiwa H, Ishiwari Y, Mizukawa N, Nagai N, Sugahara T, 2001. Identification of a stop codon mutation in the CBFA1 runt domain from a patient with cleidocranial dysplasia and cleft lip. J Oral Pathol Med. 30:381-383.

Yamamura Y, Lee WL, Inoue K-I, Ida H, Ito Y, 2006. RUNX3 Cooperates with FoxO3a to Induce Apoptosis in Gastric Cancer Cells. J Biol Chem. 281:5267-5276.

Yamashiro T, Aberg T, Levanon D, Groner Y, Thesleff I, 2002. Expression of Runx1, -2 and -3 during tooth, palate and craniofacial bone development. Mech Dev. 119:S107-110.

Yang G, Khalaf W, van de Locht L, Jansen JH, Gao M, Thompson MA, van der Reijden BA, Gutmann DH, Delwel R, Clapp DW, Hiebert SW, 2005. Transcriptional repression of the Neurofibromatosis-1 tumor suppressor by the t(8;21) fusion protein. Mol Cell Biol. 25:5869-5879.

Yoshida CA, Furuichi T, Fujita T, Fukuyama R, Kanatani N, Kobayashi S, Satake M, Takada K, Komori T, 2002. Core-binding factor interacts with Runx2 and is required for skeletal development. Nat Genet. 32:633-638.

Yoshida CA, Yamamoto H, Fujita T, Furuichi T, Ito K, Inoue K. Yamana K, Zanma A, Takada K, Ito Y, Komori T, 2004. Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. Genes Dev. 18:952-963.

Young DW, Hassan MQ, Pratap J, Galindo M, Zaidi SK, Lee SH, Yang X, Xie R, Javed A, Underwood JM, Furcinitti P, Imbalzano AN, Penman S, Nickerson JA, Montecino MA,

Young DW, Hassan MQ, Yang X-Q, Galindo M, Javed A, Zaidi SK, Furcinitti P, Lapointe D, Montecino M, Lian JB, Stein JL, van Wijnen AJ, Stein GS, 2007b. Mitotic retention of gene expression patterns by the cell fate determining transcription factor Runx2. Proc Natl Acad Sci USA. 104:3189-3194.

Zeng Z, Duan Z, Xu S, Pan F, 2013. Is RUNX3 a new player in the pathogenesis of ankylosing spondylitis? Rheumatol Int. 33:2449-2450.

Zheng L, Iohara K, Ishikawa M, Into T, Takano-Yamamoto T, Matsushita K, Nakashima M, 2007. Runx3 negatively regulates Osterix expression in dental pulp cells. Biochem J. 405:69-75.

Zheng Q, Zhou G, Morello R, Chen Y, Garcia-Rojas X, Lee B, 2003. Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression in vivo. J Cell Biol. 162:833-842.

Chapter 2

Cloning of Runt-related transcription factor 3 (*runx3*) transcript variants in zebrafish and their expression patterns during zebrafish development.

This chapter is in preparation for publication as a research paper:

Brigite Simões, Natércia Conceição, Robert Kelsh and M Leonor Cancela

Author's contribution:

All experimental work and writing of the paper was performed by B Simoes, with the exception of the pairwise identity analysis and some qPCR experiments performed by N Conceição. N Conceição, ML Cancela and RN Kelsh were responsible for the research concept and design, critical revision and final approval of the manuscript.

2.1 Abstract

The Runt domain transcription factor is known as polyomavirus enhancer-binding protein 2 (PEBP2)/core-binding factor (CBF) and is a heterodimer of two different subunits, α and β. In humans, three distinct genes, *RUNX1*, *RUNX2* and *RUNX3*, encode the DNA-binding α subunits while the single *CBFB* gene encodes the shared non-DNA binding β subunit, PEBP2β/CBFβ. The RUNX proteins share an evolutionarily conserved 128 amino acid region called the Runt domain which is responsible for DNA binding and hetero-dimerization with the co-factor CBFβ. This subunit enhances the DNA-binding activity of the RUNX transcription factors and regulates their turnover by protecting them from ubiquitin proteasome-mediated degradation. All *RUNX* genes contain two alternative promoters, P1 (distal) and P2 (proximal). RUNX transcription factors are involved in major developmental processes such as hematopoiesis, osteogenesis and neuropoiesis and mutations in these genes have been frequently associated with human hereditary diseases and development of cancer.

To improve our understanding of the structure and function of the Runx3 gene, we adopted a comparative genomic approach to analyse its domains and degree of conservation. A multiple alignment of the RUNX3 proteins from different species revealed high homology of the conserved protein domains. We have cloned and characterized several zebrafish *runx3* transcript variants and the corresponding protein isoforms that are derived from alternative promoter usage (P1 and P2 promoters) and alternative splicing events. We analyzed the temporal expression pattern of zebrafish P1 and P2 derived transcripts by qPCR during zebrafish embryonic development and in a variety of adult tissues. Our results show that the *runx3*-P1 and *runx3*-P2 transcripts are differentially expressed, with the latter showing a ubiquitous expression in all tissues analysed and expressed from 24 hpf onwards, in contrast to *runx3*-P1 transcripts that seem to be expressed in a tissue-specific manner and are maternally inherited. In addition, mRNA *in situ* hybridization was performed in whole mount zebrafish embryos and *runx3* expression was detected in neural and cartilaginous tissues. Altogether our data show that zebrafish can be a valuable vertebrate model to gain insight about the complex regulation of the *runx3* isoforms and to perform functional studies in order to clarify the importance of each isoform in development.

2.2 Introduction

The PEBP2/CBF is a small family of heterodimeric transcription factors composed of α and β subunits (Ito, 2004). The α-subunit is encoded by the runt domain genes (PEBP2α/AML/RUNX) that are characterized by containing a highly conserved 128 amino acid region (the runt domain). The *runx* genes were first described in *Drosophila melanogaster*, and shown to be essential for the formation of segmentation pattern in embryos (Gergen and Butler, 1988).

The RUNX proteins bind as homodimers to the promoter region of target genes recognizing the consensus DNA motif TGPyGGTPy (where Py is a pyrimidine) through the runt domain, and either act as activators or repressors. They can also heterodimerize with the non-DNA binding β-subunit (PEBP2β/CBFβ), that enhances the DNA-binding activity of the α subunit and stabilizes their proteins regulating their turnover by protecting them from ubiquitin proteasome-mediated degradation (Kagoshima et al, 1996).

So far, one α subunit has been identified in *Caenorhabditis elegans* (*rnt-1*), three in mammals (*Runx1*/*AML1*/*Cbfa2*, *Runx2*/*AML3*/*Cbfa1* and *Runx3*/*AML2*/*Cbfa3*) (Ito, 2004; van Wijnen et al, 2004) and four in Drosophila (*Runt*, *Lozenge*, *CG34145* and *CG42267*) (Canon and Banerjee, 2000; Rennert et al, 2003), in zebrafish (*runx1*, *runx2a*, *runx2b* and *runx3*) and in Fugu (*frRunx1*, *frRunx2*, *frRunx3*, and *frRunt*). For the β subunit, one gene has been described in all species (*bro-1* for *C. elegans* and *CBFβ* for mammals and fish), except for Drosophila, in which two genes were found (*Brother* and *Big-brother*) (Golling et al, 1996).

All *RUNX* genes are regulated from two distinct promoters, P1 and P2, giving rise to two major protein isoforms differing in the first amino acids of the N-terminal (Tighe and Calabi, 1994; Miyoshi et al, 1995; Geoffroy et al, 1998; Xiao et al, 1998; Levanon et al, 2001; Makita et al, 2008). The alternative promoter usage, in combination with alternative splicing contribution from exon skipping and multiple poly(A) tail sites, produce a large variety of alternatively spliced isoforms, resulting in the complex biological functions of Runx proteins (Miyoshi et al, 1995; Terry et al, 2004; Ng et al, 2007).

RUNX genes have been functionally characterized in many organisms, and are known to play essential roles in metazoan development (Coffman, 2003). Even if expression of the *Runx*

genes can overlap in some tissues (Le et al, 1999), they have distinct roles in tissue morphogenesis and homeostasis, controlling critical cell fate decisions in a number of different cell lineages (reviewed in Chuang et al, 2013) and being involved in the transcriptional control of developmental processes (Wheeler et al, 2000; Taniuchi et al, 2002; Coffman, 2003). Although many studies have been published about *Runx* genes, their biological function and regulation is still not completely understood. It was previously shown that *RUNX* genes are involved in hematopoiesis, neurogenesis and osteogenesis and are also implicated in disease and cancer. Runx1 is involved in regulation of hematopoiesis (Ichikawa et al. 2004; Komori, 2005; Growney et al, 2005), Runx2 is essential for bone and tooth development (Otto et al, 1997; Komori et al, 1997; D'Souza et al, 1999; Maruyama et al, 2007) and Runx3 is critical for gastric epithelial differentiation, neurogenesis of dorsal root ganglia and T cell differentiation (Levanon et al, 2002; Brenner et al, 2004; Komori, 2005; Friedrich et al, 2006; Lee et al, 2011; Reis et al, 2013).

Functional studies in different model organisms showed that alteration in the Runx3 expression levels affect the hematopoietic lineage (Kalev-Zylinska et al, 2003; de Bruijn and Speck, 2004; Wang et al, 2013), chondrocyte maturation (Yoshida et al, 2004; Hecht et al, 2008; Zhang et al, 2009; reviewed in Komori, 2015), and dorsal root ganglia (DRG) development (Inoue et al, 2002, 2003; Nakamura et al, 2008). In zebrafish, a *runx3* orthologue was previously identified and three splicing transcripts, encoding two protein isoforms with a different N-terminal, Runx3-MASN and Runx3-MHIPV, were described (Kataoka et al, 2000; Kalev-Zylinska et al, 2002). These studies characterized the zebrafish *runx3* embryonic expression, confirming a broad expression pattern similar to what was observed in mammals, and including Rohon-Beard neurons, trigeminal ganglia, craniofacial cartilage and hematopoietic cells. Some knockdown studies demonstrated that a reduction in the *runx3* transcript levels affect hematopoietic cells, neurons and craniofacial cartilage. It is still not clear whether the *runx3* transcripts are differentially expressed and if the encoded protein isoforms have different roles in development.

Here we describe the cloning of 12 new zebrafish *runx3* splicing transcript variants, derived from a combination of different promoter usage and alternative splicing. We analysed the expression pattern of the *runx3*-P1 and *runx3*-P2 derived transcripts in both zebrafish adult

tissues and developmental stages, using quantitative real-time PCR (qPCR) targeting the transcript-specific 5' UTR sequence, and the results confirm that they are differentially expressed. The mRNA expression pattern was also analysed by *in situ* hybridization, using a riboprobe that recognized the common region of both transcript variants, or a riboprobe that recognized the 5' UTR region specific to each transcript. The expression of the *runx3*-P2 derived transcripts appeared to overlap with that of the riboprobe recognizing the common region, but no conclusion could be taken about the expression of the *runx3*-P1 derived transcripts.

2.3 Material and Methods

2.3.1. Ethics statement

All zebrafish (*Danio rerio*) studies were conducted using the wild type (AB) strain, both at the aquarium at the University of Algarve (Faro, Portugal) or at the University of Bath (Bath, UK). The experiments were conducted following the legislation for animal experimentation and welfare either in Portugal, in accordance with the Portuguese law (Portaria 1005/02 and Portaria 1131/97) which transcribes the European Guideline 86/609/EC, and in UK, according to the Animals (Scientific Procedures) Act 1986 - Scientific Procedures on Living Animals, under a personal Home Office licence.

2.3.2. Zebrafish maintenance

Embryos were obtained from natural spawning and maintained according to each fish facility standard procedures. The embryos were collected in embryo medium (Westerfield, 2000), raised up to 5 days post fertilization (dpf) in an incubator with a constant temperature of 28.5°C and staged according to Kimmel et al (1995) (occasionally, their development was slowed down or speeded up at 23°C or 33°C, respectively, if needed for a specific experiment). For some experiments, embryos were treated with 0.003% 1-phenyl-2-thiourea (PTU; Sigma) to inhibit pigment formation and thus facilitating imaging and staining for

whole-mount *in situ* hybridization (ISH). All specimens were deeply anesthetised in Tricane (3-amino benzoic acid ethyl ester also called ethyl 3-aminobenzoate) (Sigma-Aldrich) before manipulation as described (Westerfield, 2000).

2.3.3 Sequence alignment and analysis

BLAST facilities at NCBI (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov) and Ensembl (http://www.ensembl.org/) databases were used to search for sequences showing similarity with zebrafish *runx3* gene and protein.

Alignments for Runx3 sequences were created using ClustalW (Thompson et al, 1994) or T-Coffee multiple sequence alignment software (Notredame et al, 2000; http://www.ebi.ac.uk/Tools/msa/tcoffee/) with parameters set to the default. Sequence logos were then created by T-Coffee from multiple alignments using WebLogo facilities at weblogo.berkeley.edu (Schneider and Stephens, 1990).

Percentage protein identity was calculated using the Sequence Manipulation Suite (Stothard, 2000) available at http://www.bioinformatics.org.

2.3.4 RNA preparation

Total RNA was prepared from different adult tissues using the acid guanidium thiocyanate procedure (Chomczynski and Sacchi, 1987). Due to the low weight of the individual embryonic and larvae samples RNA was extracted from pools of zebrafish embryos at different stages of development using the TRIzol Reagent (Sigma-Aldrich) as recommended in the manufacturer's protocol. RNA integrity was assessed through 1% (w/v) agarose/MOPS-formaldehyde gel electrophoresis with ethidium bromide staining (Sambrook et al, 1989) and RNA quantity was determined through spectrophotometry (NanoDrop 1000; Thermo Scientific). Total RNA (1 μg) was then treated with RQ1 RNase-free DNase I (Promega) for 30 min at 37°C, and reverse-transcribed at 37°C for 1 h using the

Moloney-murine leukemia virus (MMLV) reverse transcriptase, RNaseOUT (both from Invitrogen) and oligo(dT)-adapter or a gene specific primer (DrRunx3_R2, **Table 2.S1**).

2.3.5 Construction of cDNA library

Total RNA from a pool of embryos at different developmental stages (2 dpf, 20 dpf, 25 dpf, 40 dpf and adult male) was extracted as described in section 2.3.4. High-quality Poly A^+ RNA was purified using the Oligotex mRNA Midi Kit (Qiagen) and used in the construction of a Marathon cDNA library (Clontech) according to manufacturer's instructions. Briefly, the cDNAs were transcribed from 1 µg of purified Poly A^+ RNA blunt ends were created and the Marathon adaptors were ligated to both ends of the double-stranded cDNA.

2.3.6. cDNA Cloning

The zebrafish *runx3* sequences for P1-derived transcripts (AB043789 and AB043790) and P2-derived transcript (AB043788) were used as templates to design zebrafish *runx3* gene specific primers. All primer sequences used for cloning are shown in **Table 2.S1**.

2.3.6.1. Rapid amplification of cDNA ends (RACE)

The 5' ends of zebrafish *runx3*-P1 and *runx3*-P2 transcripts were amplified by RACE-PCR using gene specific reverse primers Runx3_R2 and Runx3_R4 and Marathon AP1 and AP2 adaptor primers (**Table 2.S1**), respectively in a primary and nested reaction, and the zebrafish Marathon cDNA library described in section 2.3.5. Amplifications were performed using the Advantage cDNA polymerase mix (Clontech) as suggested by the supplier. The resulting cDNA fragments were size-separated by electrophoresis on an agarose gel, purified using GeneJET Gel Extraction Kit (Thermo Scientific), cloned into pCRII-TOPO (Life technologies) and sequenced on both strands to confirm their identity.

2.3.6.2. Amplification of full-length cDNA

Amplification of the different zebrafish *runx3* transcripts was performed by two steps PCR with gene specific primers (**Table 2.S1**) and either with the Taq DNA polymerase (Life technologies) or the KOD Hot Start DNA Polymerase (Novagen), in a GeneAmp 2400 thermocycler (Perkin-Elmer), under conditions suggested by the suppliers. The template cDNA used was obtained, using the specific reverse primer DrRunx3_R2, from RNA obtained either from 48 hours post fertilization (hpf) embryos or adult male zebrafish, as described in section 2.3.4.

runx3 P1-derived isoforms 1 to 5 were amplified using cDNA from 48 hpf zebrafish embryos, in a primary PCR with primers DrRunx3II_F2 and DrRunx3_R8, and then used as template in a semi-nested PCR with the same sense primer and with the antisense primer DrRunx3_R6, and cloned into pCR-BluntII-TOPO vector (Life technologies).

runx3 P2-derived isoform 6 and isoforms 9 to 12 were amplified using cDNA from 48 hpf zebrafish embryos, in a primary PCR with primers DrRunx3I_F5 and DrRunx3_R8, and then in a nested PCR with primers DrRunx3I_F4 and DrRunx3_R6, and cloned into pCR-BluntII-TOPO vector. *runx3* P2-derived isoforms 7 and 8 were amplified using cDNA from an adult male zebrafish, in a primary PCR with primers DrRunx3I_F2 and DrRunx3_R2, and then a nested PCR using primers DrRunx3_F5 and DrRunx3_R6, and cloned into pCRII-TOPO vector.

Cloned fragments were identified by restriction digestion and sequenced on both strands to confirm their identity.

2.3.7 Measurement of relative gene expression by quantitative real-time PCR

Gene expression was performed by qPCR for *runx3* isoform specific transcripts using primers designed from 5'-unique sequences. For P1-derived transcripts we used the primers DrRunx3II&III_F1/DrRunx3II&III_R2 and for P2-derived transcripts we used the primers DrRunx3I_F3/DrRunx3I_R2 (**Table 2.S1**). RNA was prepared as described in section 2.3.4. qPCR was performed in a StepOnePlus apparatus (Applied Biosystems, Invitrogen). PCR

reactions, set up in duplicates, were carried out using SsoFast EvaGreen Supermix (Bio-Rad) according to manufacturer's instructions. Melting profiles and gel electrophoresis of each reaction were performed to check for unspecific product amplification. Levels of gene expression were calculated using the comparative method (ΔΔCt) and normalized using gene expression levels of *18S ribosomal RNA* or e*longation factor 1-alpha* (*ef1α*) housekeeping genes.

2.3.8 Probe synthesis and whole mount ISH staining

A fragment of 995 bp of zebrafish *runx3* transcript was PCR amplified with primers DrRunx3_F3/DrRunx3_R2 (**Table 2.S1**), using 2x BioMix (Bioline) solution as recommended by the manufacturer, and cDNA synthesized from 33 hpf zebrafish (as described in section 2.3.4). Then it was inserted into pGEM-T-easy vector (Promega) by standard TA-cloning. *runx3*-P1 and *runx3*-P2 specific fragments were amplified (178 bp and 589 bp, respectively) with primers DrRunx3II_F1/DrRunx3II&III_R1 and DrRunx3I_F2/DrRunx3I_R1 (**Table 2.S1**) respectively, using the Taq DNA polymerase as recommended by the manufacturer, and cDNA synthesized from adult male zebrafish (as described in section 2.3.4), and cloned into pCRII-TOPO. All fragments were sequenced on both strands to confirm their orientation and identity. Those constructs were used as templates to each probe synthesis. The plasmids were linearized using the appropriated enzyme (**Table 2.1**). Then, digoxigenin (DIG)-labeled antisense RNA probes were synthesized using either T7 or SP6 RNA polymerases (**Table 2.1**) according to the manufacturer's specifications (Roche). Whole-mount RNA ISH was performed as previously described (Thisse et al, 1993). After the hybridization, some embryos were embedded in paraffin, cut into 5 µm-thick sections, and imaged in an Eclipse E800 (Nikon) microscope.

Table 2.1 Information for ISH probe synthesis.

probe	Plasmid	Enzyme	Polymerase	size	Notes
***runx3* (+runt)**	pCRII-TOPO	ApaI SacI	SP6 T7	anti-sense sense	1175 bp
***runx3* (-runt)**	pGEM-T-Easy	SacI ApaI	T7 SP6	anti-sense sense	995 bp
P1-*runx3*	pCRII-TOPO	SacI XhoI	T7 SP6	anti-sense sense	178 bp
P2-*runx3*	pCRII-TOPO	SacI XhoI	T7 Sp6	anti-sense sense	589 bp

2.4 Results

2.4.1 Molecular characterization of zebrafish *runx3* gene

Kataoka and colleagues (2000) have previously published the cloning of three zebrafish *runx3* transcripts (that they designated *runx*b) generated by alternative promoter usage and alternative splicing (Kataoka et al, 2000). In order to amplify the corresponding zebrafish open reading frames (ORFs) we used the *runx3* transcripts (Accession Numbers: AB043788, AB043789 and AB043790) previously described as templates to design our primers. First we performed BLAST analysis against the zebrafish genome (*Danio rerio* strain Tuebingen chromosome 13, GRCz10) in NCBI database to determine the gene structure and the transcript variations. The zebrafish *runx3* gene was found on chromosome 13 in the reverse orientation (position 44960717 - 45064151) with a length of approximately 103.4 kb and it is organized in 7 exons and 6 introns (**Figure 2.1**), showing a structure slightly different from the *runx3* genomic structures known for other species, where the gene is organized in 6 exons and 5 introns.

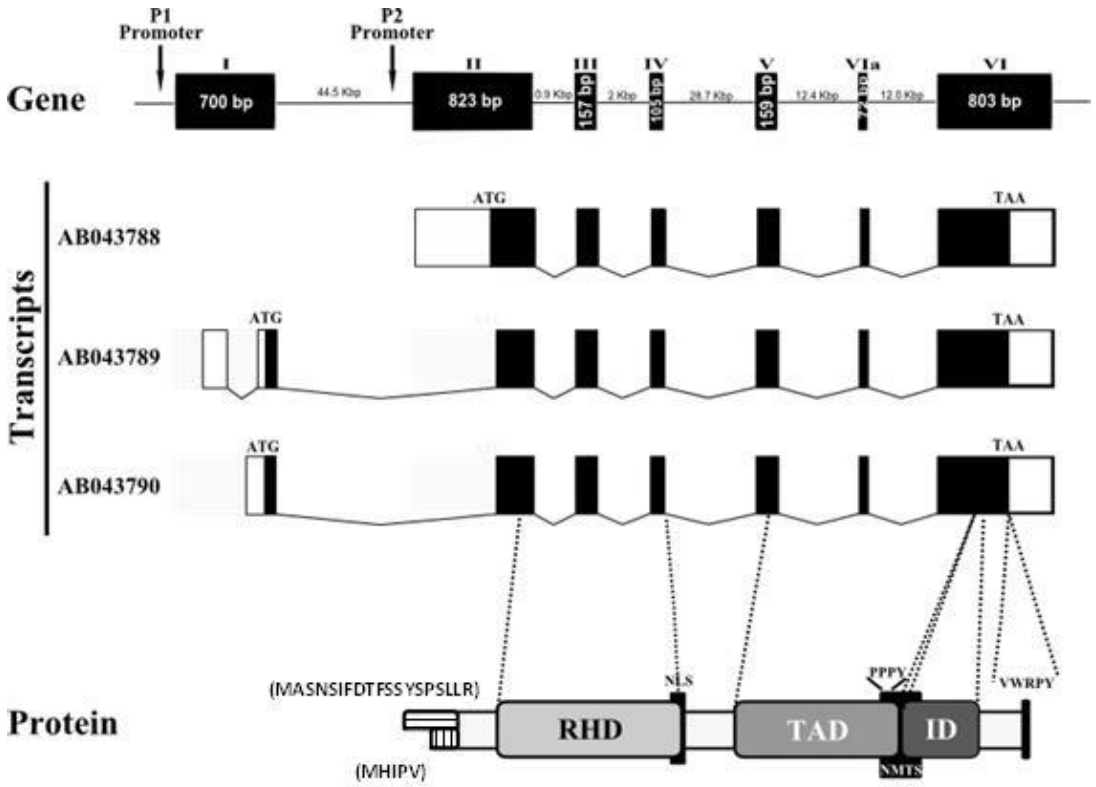


Figure 2.1 Schematic representations of the zebrafish *runx3* gene, the cDNA splicing variants and corresponding protein structure. In gene structure, exons are represented by black boxes and introns by solid black lines. Exon sizes are indicated inside the boxes and the intron sizes indicated above the lines. Exons are in scale but the introns are not. The two alternative promoter regions are also indicated (arrow). In cDNA structure, black represents the coding sequence and white the 5' and 3' untranslated regions. The ATG translation start and TAA stop codons are also indicated. In the protein structure, the Runx protein characteristic domains are represented as well as their corresponding region in the cDNA (dot lines). Box with horizontal and vertical lines indicates specific N-terminal from P1- and P2-derived isoforms, respectively. RHD – runt homology domain; NLS – nuclear localization signal; TAD – transactivation domain; ID – inhibition domain; PPPY - short proline-rich motif; NMTS - nuclear matrix targeting signal; VWRPY – Runx conserved penta-peptide motif.

The sites of exon-intron borders were deduced by comparison of the transcript sequences with the genomic sequence and it was observed that all exon-intron splicing junctions match consensus sequences for donor and acceptor sites (result not shown). The *runx3* transcripts are transcribed from two alternative promoters (P1/distal and P2/proximal) and encode two Runx3 protein isoforms that differ only in their N-terminal sequence (**Figure 2.1**). The P1-derived transcripts (AB043789 - type II and AB043790 - type III) generate a 438 amino acids protein containing a specific N-terminal of 19 amino acids (MASNSIFDTFSSYSPSLLR) present only in the P1-derived isoforms. The P2-derived transcript (AB043788 - Type I) does not

contain exon 1, been transcribed from an alternative promoter in intron 1, and so it is translated from a different ATG codon present in exon 2, generating a 424 amino acids protein containing a specific N-terminal of five amino acids (MHIPV) present only in the P2-derived isoforms (Kataoka et al, 2000). Due to their N-terminal sequences, in this work these isoforms are also referred as Runx3-MASN or Runx3-MHIPV isoform, depending if we are referring to a P1-derived or P2-derived isoform, respectively.

From the cDNA deduced primary structure and by comparative analysis between all RUNX proteins, we identified the typical domains present in all Runx proteins characteristic of this family of proteins: the runt domain (RHD), transcription activation domain (TAD), inhibition domain (ID), nuclear localization signal (NLS), nuclear matrix target signal (NMTS) and PPPY and VWRPY domains (**Figure 2.1**).

2.4.2 *in silico* analysis of Runx3 protein isoforms

Sequence databases at NCBI (www.ncbi.nlm.nih.gov) and Ensembl (www.ensembl.org) were searched for annotated RUNX3 sequences. A total of 23 RUNX3 sequences were collected (**Table 2.S2**). The full collection of sequences represents 18 species, from most classes of vertebrates (actinopterygii, chondrichthyes, amphibian and mammalia). We have calculated the pairwise percentage identities with Manipulation Suite facilities among RUNX3 isoform protein sequences of different species, observing that the amino acid sequence of zebrafish Runx3 has higher identity to those of grass carp (94%) and fugu (74%) and a sequence identity of 61-64% compared to mammals (including human) (**Figure 2.2**).

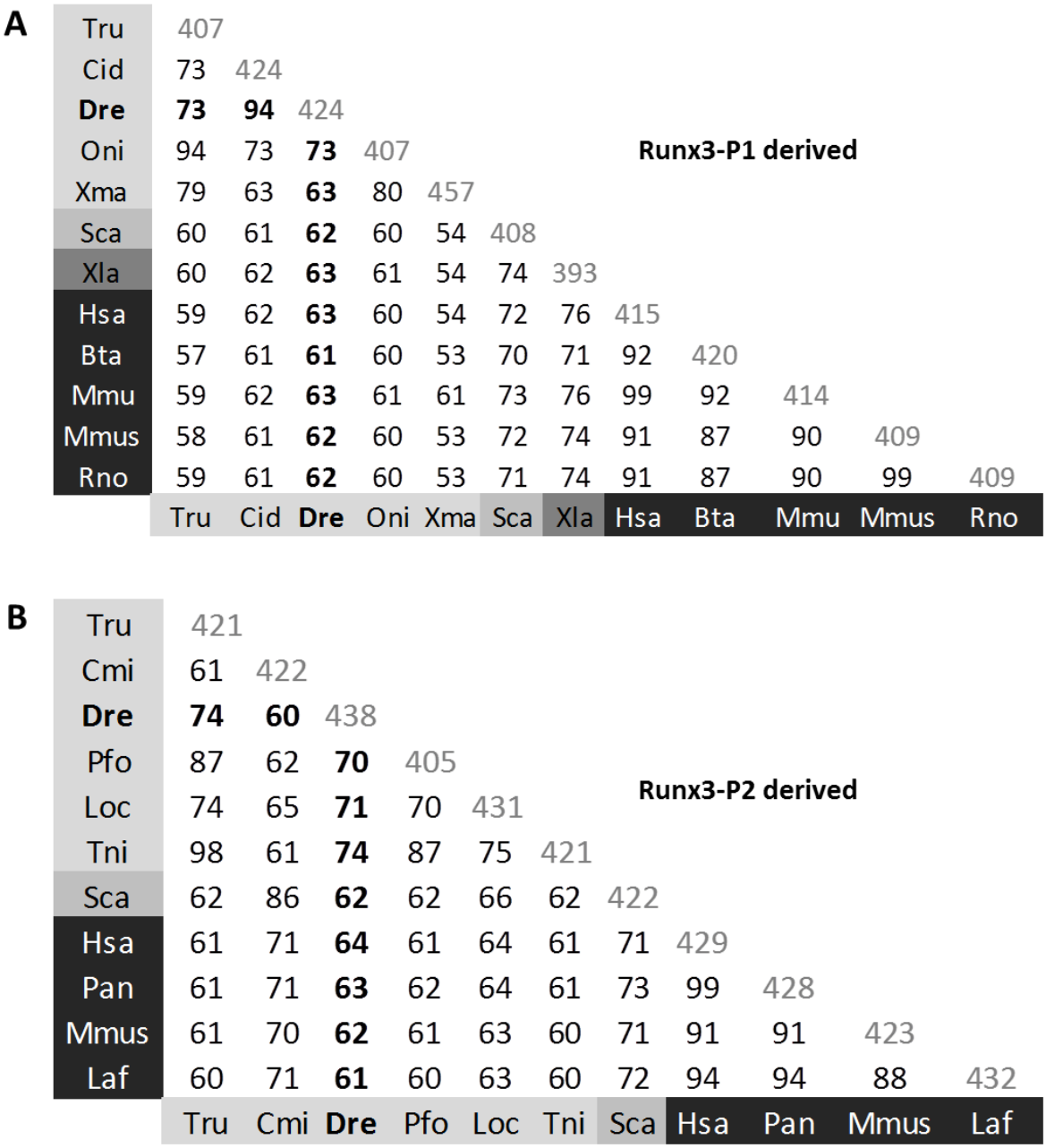


Figure 2.2 Pairwise percent identities among RUNX3 (A) P1-derived (MASN) and (B) P2-derived (M(H/R)IPV) isoforms. From light grey to black: actinopterygii, chondrichthyes, amphibia, mammalia. Hsa, *Homo sapiens* (human); Mmu, *Macaca mulata* (rhesus macaque); Bta, *Bos Taurus* (bovine); Mmus, *Mus musculus* (mouse); Rno, *Rattus norvegicus* (rat); Dre, *Danio rerio* (zebrafish); Cid, *Ctenopharyngodon idella* (grass carp); Oni, *Oreochromis niloticus* (tilapia); Loc, *Lepisosteus oculatus* (Spotted gar); Pfo *Poecilia formosa* (Amazon molly); Xma, *Xiphophorus maculatus* (Southern platyfish); Cmi, *Callorhinchus milii* (elephant shark); Xla, *Xenopus laevis* (African clawed frog); Tni, *Tetraodon nigroviridis* (spotted green pufferfish); Tru, *Tetraodon rubripes* (torafugu); Sca, *Scyliorhinus canicula* (lesser spotted catshark); Laf, *Loxodonta africana* (elephant); Pan, *Papio anubis* (Olive baboon). Diagonal grey values are sequence lengths.

As the overall Runx3 protein sequence identity between zebrafish and the mammals was not very high, we decided to analyse the degree of conservation between the domains of the RUNX3 proteins. The RUNX3 sequences (**Table 2.S2**) were aligned using T-Coffee and the result submitted to WebLogo software that generates sequence logos, a graphical representation of the patterns within a multiple sequence alignment (**Figure 2.3**). In the webLogo graph, the amino acid similarity is represented by the height of the each letter (logos that represent amino acid residues) that is directly proportional to the probability of the corresponding amino acid, so the conservation can be easily assessed (**Figure 2.3**).

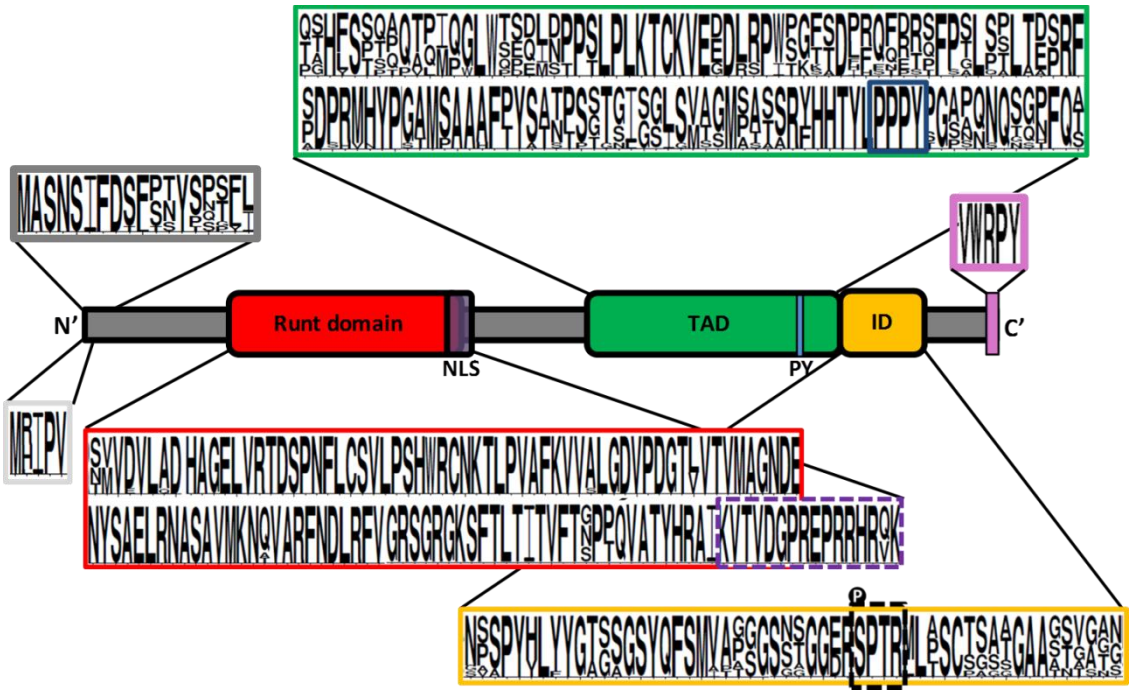


Figure 2.3 Schematic representation of the conservation in RUNX3 protein domains. The characteristic domains found in all RUNX proteins are indicated to show their relative positions along the protein. The multiple sequence alignment (using T-Coffee and WebLogo) using 23 RUNX3 proteins from 18 different species was used to determine the amino acid conservation in the different protein domains. NLS - nuclear localization signal, TAD - transactivation domain, ID - inhibitory domain, PY - short proline-rich motif, P -residue targeted for phosphorylation, VWRPY – Runx conserved penta-peptide motif.

As expected, we could observe a high degree of conservation of the runt domain as well as a 100% conservation of the NLS, PY and VWRPY motifs (**Figure 2.3**). The TAD and ID domains showed more variation, however some residues are still 100% conserved, indicating that they probably have a major contribution to the function of that domain (**Figure 2.3**).

2.4.3. Cloning of new zebrafish *runx3* transcript variants

In an attempt to amplify the zebrafish transcripts that derived from alternative promoter usage (P1/Distal or P2/Proximal) that generates different Runx3 isoforms, we hereby identified and cloned 12 different transcripts (**Figure 2.4**), 10 of each are new transcript variants that are described in this study for the first time, and found to have a structure different from that of any previously reported isoform of zebrafish *runx3* cDNAs. Two of the transcripts are similar to the P1-derived and P2-derived isoforms previously described (*runx3* isoform 1 and *runx3* isoform 6, respectively), four correspond to splicing variants of P1-derived transcripts (*runx3* isoform 2 to 5) and six correspond to splicing variants of P2-derived transcripts (*runx3* isoforms 7 to 12). These transcripts result from alternative exon skipping, generating different protein isoforms, depending on the splicing event. The identity of the *runx3* cDNA sequences obtained was confirmed using blast searches against GenBank (NCBI) database. The nucleotide sequences of these new spliced variants were deposited in GenBank as *runx3* isoforms 1 to 12 (GenBank ID: KP866665, KP866666, KP866667, KP866668, KP866669, KP866670, KP866671, KP866672, KP866673, KP866674, KP866675, KP866676, respectively).

The ORF sequence of the *runx3* isoform 1 and isoform 6 transcripts are similar to the ORF sequence of the *runx3* transcripts (AB043789 and AB043788, respectively) previously published in Kataoka et al (2000) (**Figure 2.4**). Comparison between both sequences shows that the *runx3* isoform 1 and the isoform 6 have nine nucleotides different from the published sequences, causing an ORF with two amino acid substitutions.

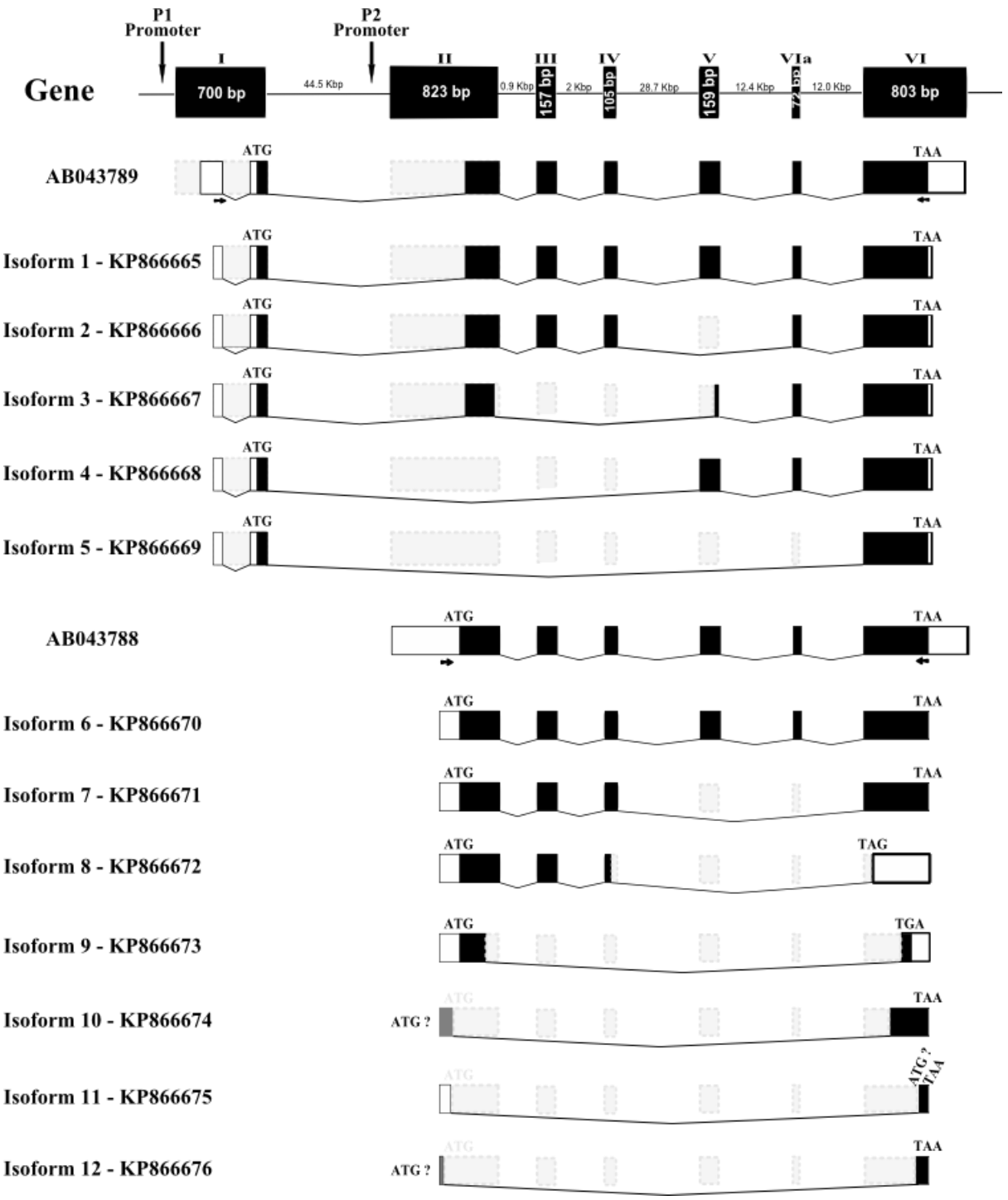


Figure 2.4 Schematic representation of zebrafish alternatively spliced *runx3* cDNA transcripts. (A) Schematic representation of the *runx3* gene structure. The exons are represented by black boxes, and the lines represent the introns and intergenic sequences. The exons are scale represented, but the introns are not. (B) Schematic representation of the alternative spliced *runx3* transcripts cloned. Black boxes represent coding exons, grey boxes represent spliced exons and white boxes represent UTRs. Start and stop codons are represented. Arrows represent the location of primers used for the amplification of the ORFs.

The transcripts of *runx3* isoforms 2 to 5 are splicing variants of isoform 1 that have a splicing event that eliminates one or more exons (**Figure 2.4**). Isoform 2 is a splicing variant that loses exon 5, and we predicted that it loses the NLS and part of the TAD domain, but keeps the coding region for the runt domain intact. On the other hand, transcripts of isoforms 3 to 5 have a splicing event that eliminates multiple exons: isoform 3 has a splicing from nucleotide 803 in exon 2 to nucleotide 122 in exon 6; isoform 4 has a splicing that eliminates exons 2 to 4; and isoform 5 has a splicing that eliminates exons 2 to 6a (**Figure 2.4**). The transcripts of *runx3* isoforms 7 to 12 are splicing variants of isoform 6, and all have multiple exons eliminated by alternative splicing events (**Figure 2.4**). The isoform 7 loses its exons 5 and 5a that corresponds to the coding of NLS and part of the TAD domains, but maintains its runt domain intact, suggesting that can be a functional isoform. The isoforms 8 and 9 have a splicing from nucleotide 52 of exon 4 to nucleotide 65 of exon 6 (isoform 8) and from nucleotide 725 of exon 2 to nucleotide 296 of exon 6 (isoform 9), causing a frame shift that produces a premature stop codon (**Figure 2.4**). On the other hand, the isoforms 10 to 12 have an extensive splicing event that eliminates the majority of the exons including the start codon ATG. As a consequence, these *runx3* transcripts are likely to result in loss-of-function mutants.

2.4.4 Molecular structure of zebrafish *runx3* 5' UTR regions

We employed RACE analyses to map the transcription-start sites (TSSs) downstream the two promoters, *runx3*-P1 and *runx3*-P2. A new putative TSS was mapped by 5'-RACE for the P1-transcript, and contains 642 bp of 5'-untranslated region from the translation initiation ATG codon (**Figure 2.5**). The sequence of this new P1-5' UTR splicing variant, named *runx3* isoform 14 (GenBank ID: KP866678) corresponds to the AB043789 transcript described by Kataoka et al (2000) carrying the 221 bp splicing event (nucleotides -277 to -57) but with more 182 bp of 5´ UTR region (**Figure 2.5**). Besides this, we also found a new P1-5' UTR regions splicing variant, named *runx3* isoform 13 (GenBank ID: KP866677), obtained from cDNA of an adult zebrafish speciemen (as described in section 2.3.6.2), that corresponds to the AB043789 transcript previously described by Kataoka et al (2000) but with more 313 bp of 5' UTR region (**Figure 2.5**). Regarding the P2-TSS, we also amplified and sequenced one

major DNA fragment that placed the major P2-TSS 562 bp upstream of the initiator ATG codon, matching the sequence available in the NCBI database (AB043788) (Results not shown).

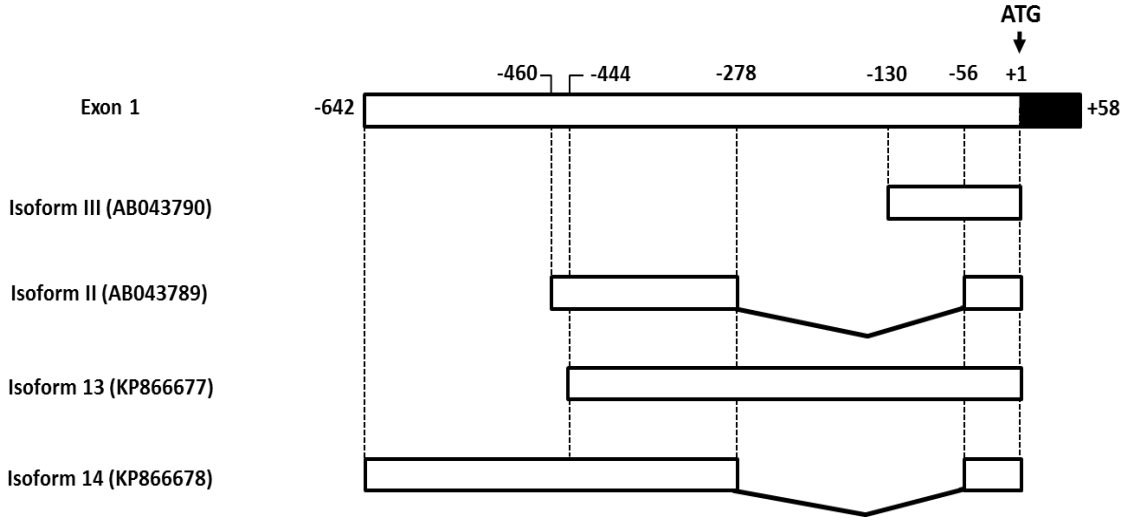


Figure 2.5 Schematic representations of the splicing events within the *runx3* exon 1 affecting the 5' UTR of the P1-derived transcripts. (Top) Representation of the *runx3* exon 1. White box represents 5' UTR region, black box represents coding region. ATG codon is also represented. The A from the ATG codon of the *runx3* P1-derived transcripts is represented as +1. (Bottom) Representation of the different 5' UTRs of the P1-derived transcripts described and the new transcripts obtained in this work, originated by alternative splicing. The rectangles represent transcript sequences and lines represent spliced sequences.

The analysis of both 5' UTR regions of *runx3* transcripts, revealed several ATG nucleotide triplets upstream of the ATG codon that is known to initiate Runx3 protein isoforms. Using a translation program (http://web.expasy.org/translate/), we could predict some putative upstream open reading frames (uORFs) in both 5' UTR regions, that can vary in size from very small (just one amino acid long) to longer peptides (70 amino acids long) (**Figure 2.S1**). The presence of one or more uORFs has been shown to repress the translation of the downstream ORF and this feature has been associated with tissue-specific gene regulation and developmental control (van der Velden and Thomas, 1999).

2.4.5 qPCR analysis of *runx3* P1 and P2 transcript variants

The gene expression of *runx3*-P1 and *runx3*-P2 derived transcripts has been evaluated in a variety of adult zebrafish tissues (including calcified and non-calcified soft tissues) by qPCR, using specific primers and normalized using the *ef1α* as housekeeping gene (**Figure 2.6**). The expression of P1-derived transcripts was detected mainly in the calcified tissues analysed, with the highest levels observed in the pectoral fins (≈6 fold) and in the branchial arches (≈15-fold), and with a similar relative expression in the brain, heart and skin and its expression in the liver and muscle is barely detectable (**Figure 2.6A**). P2-derived transcripts expression was detected in all tissues analysed, with the highest levels of expression in the calcified tissues (between ≈15-fold to ≈37-fold) compared to the expression in the soft tissues (between ≈1.7-fold to ≈12-fold) (**Figure 2.6B**).

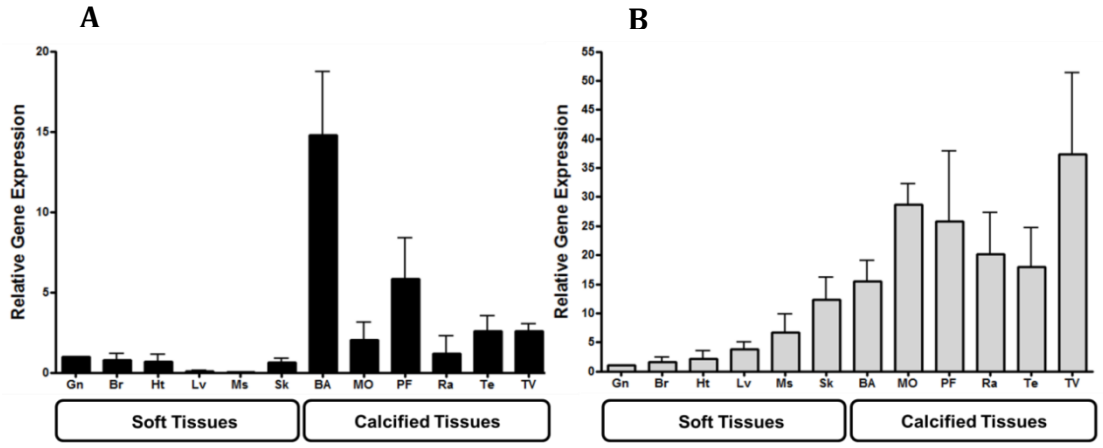


Figure 2.6 Levels of *runx3*-P1 and *runx3*-P2 derived transcripts in adult zebrafish tissues. Levels of gene expression were determined by qPCR in adult soft and calcified tissues and normalized using *ef1α* as housekeeping gene. The level in female gonads (Gn) was used as a reference and set to 1. Gn – female gonads; Br – brain; Ht – heart; Lv – liver; Ms – muscle; Sk – skin; BA – branchial arches; MO – mandibular operculum; PF – pectoral fins; Ra – rays; Te – teeth; TV – trunk vertebrae. Values are the mean of at least three replicates with error bars representing ± standard deviation (S.D.).

To analyse the temporal expression of *runx3*-P1 and *runx3*-P2 derived transcripts during zebrafish embryogenesis, mRNAs collected at seven different time points between 1-cell and 13 dpf were analysed by qPCR using specific primers and normalized using the *18S* as housekeeping gene. A different expression pattern was observed for *runx3*-P1 and *runx3*-P2 derived transcripts during the initial stages of zebrafish development (**Figure 2.7**).

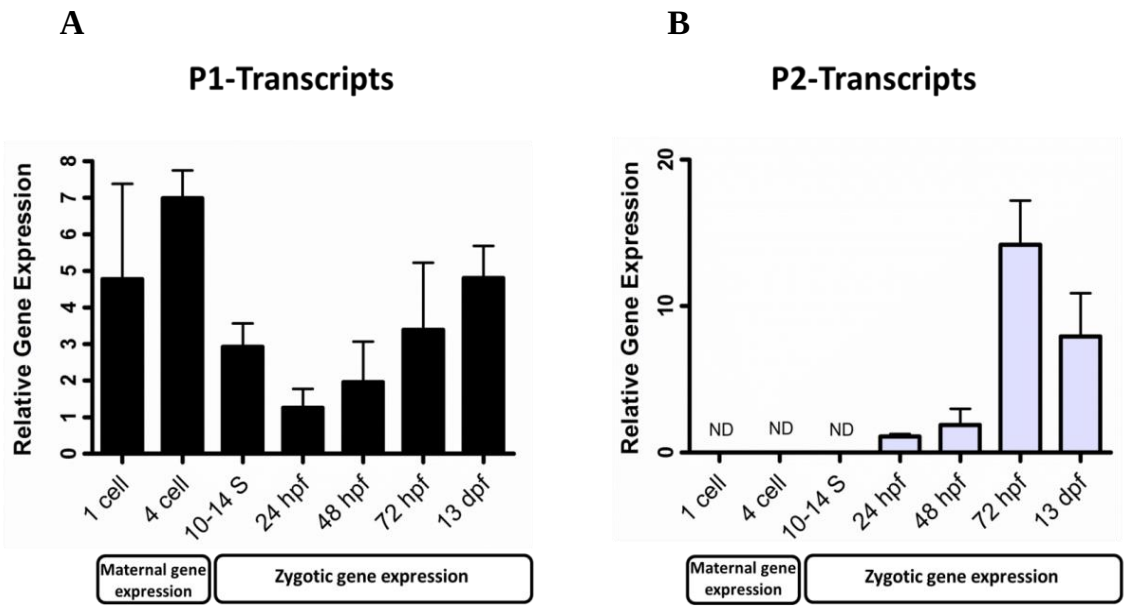


Figure 2.7 Levels of *runx3*-P1 and *runx3*-P2 derived transcripts in embryonic zebrafish development. Levels of gene expression were determined by qPCR in different embryonic developmental stages and normalized using *18S* as housekeeping gene. S means somites. An arbitrary unit was used as reference. Values are the mean of two replicates with error bars representing S.D.

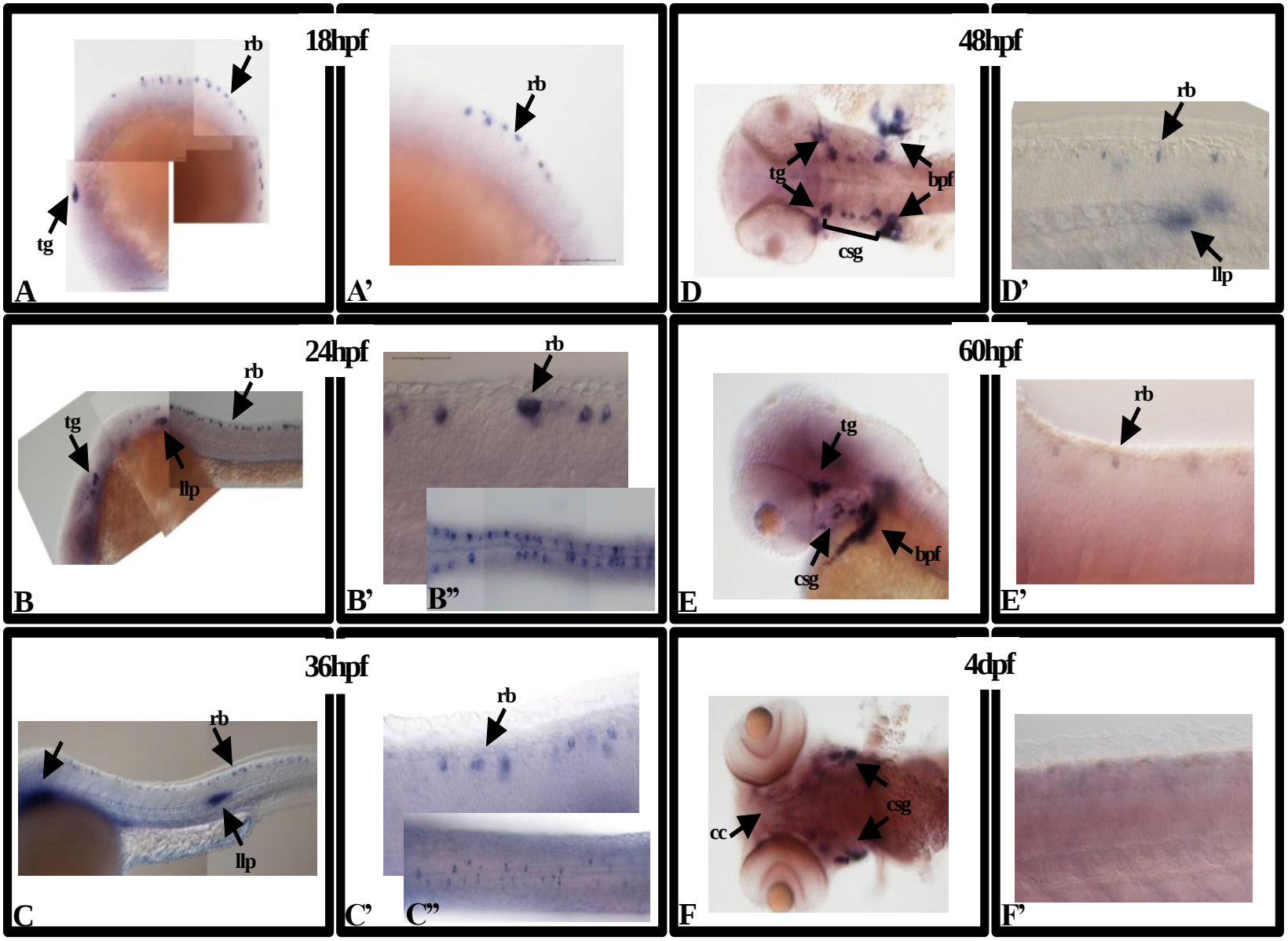
While *runx3*-P1 derived transcripts are detected already at 1 cell (**Figure 2.7A**), *runx3*-P2 transcripts can only be detected from 24 hpf onwards, showing its highest expression at 72 hpf (**Figure 2.7B**). Interestingly, the expression of *runx3*-P1 derived transcripts decreases until 24 hpf and then it increases again until 13 dpf (**Figure 2.7A**).

2.4.6 Expression patterns of zebrafish *runx3* gene expression

To determine the spatial and temporal expression pattern of *runx3*, whole-mount ISH time course was performed in wild-type zebrafish embryos. A *runx3* specific riboprobe (995 bp), which is complementary to the 3' of the *runx3* cDNA, including part of the sequence that codes for the runt domain (common part of all *runx3* transcripts) was used. Zebrafish wild type specimens, from 24 hpf up to 4 dpf were collected, fixed and submitted to whole mount ISH (**Figure 2.8, A-F**) and posterior cross-sections at 24 hpf and 36 hpf were also analysed (**Figure 2.8, G-H**).

The results of the ISH confirmed the *runx3* expression pattern previously published (Kataoka et al, 2000; Kalev-Zylinska et al, 2003). We observed *runx3* in the first stage analysed, at 18 hpf, in the trigeminal ganglions and in the Rohon-Beard neurons (rb), that are primary sensory neurons located dorsally in the embryos (**Figure 2.8A, A' to E, E'**). The expression in the rb neurons is still very strong at 24 hpf, but then gradually decreases being the number of rb expressing cells much lower at 60 hpf and almost non-existent at 4 dpf This is in agreement with previous studies showing that rb sensory neurons die over a protracted period of time (Williams et al, 2000; Cole and Ross, 2001; Reyes et al, 2004). Reyes and colleagues, showed that in zebrafish embryos rb neurons begin to show signs of programmed cell death as early as 1 dpf and that most rb cells died by 4 dpf (Reyes et al, 2004). At 24 hpf *runx3* expression is also detected in the lateral line primordia (llp), that migrates along the trunk (**Figure 2.8B, C, D'**). We also observed *runx3* expression at the cranial sensory ganglia from 48 hpf, as well as, in the base of the pectoral fins and in the cranial cartilages at 4 dpf (**Figure 2.8D, E, F**). The same expression pattern was observed using a probe spanning 1175 bp (**Table 2.1**), containing all sequence that codes for the runt domain (results not shown). However, we fail to detect *runx3* expression in the hematopoietic tissues, suggesting that maybe our protocol needs to be improved to increase sensitivity. Posterior cross-sections of the trunk at 24 hpf and 36 hpf confirmed the expression in the Rohon-Beard neurons (**Figure 2.8G, G', H**; arrow) and at 36 hpf we observed expressing cells that seems to lay outside of the neural tube and that may be part of the dorsal root ganglia (**Figure 2.8H, H'**; arrowhead).

To further assess the sites of expression for each *runx3*-P1 or *runx3*-P2 transcript variant, whole-mount ISH was performed using specific riboprobes for the 5' UTR of the P1-derived and P2-derived transcripts (**Table 2.1**; **Figure 2.8I, I', J, K, K'**). In the stages analysed we could not identify any structure expressing the P1-derived transcripts (**Figure 2.8I, I'**), but when a specific riboprobe for the P2-derived transcripts was used we could detect expression in trigeminal ganglions, Rohon-Beard neurons of neural tube, lateral line primordial, base of the developing pectoral fins, and cranial sensory ganglia (**Figure 2.8J, K, K'**), the same tissues that were detected when we used a riboprobe for the common part of *runx3* (**Figure 2.8A-F**).


18hpf
rb
tg
A
rb
A'
48hpf
tg
bpf
csg
D
rb
llp
D'
24hpf
rb
tg
llp
B
rb
B'
B"
60hpf
tg
bpf
csg
E
rb
E'
36hpf
rb
llp
C
rb
C'
C"
4dpf
cc
csg
F
F'

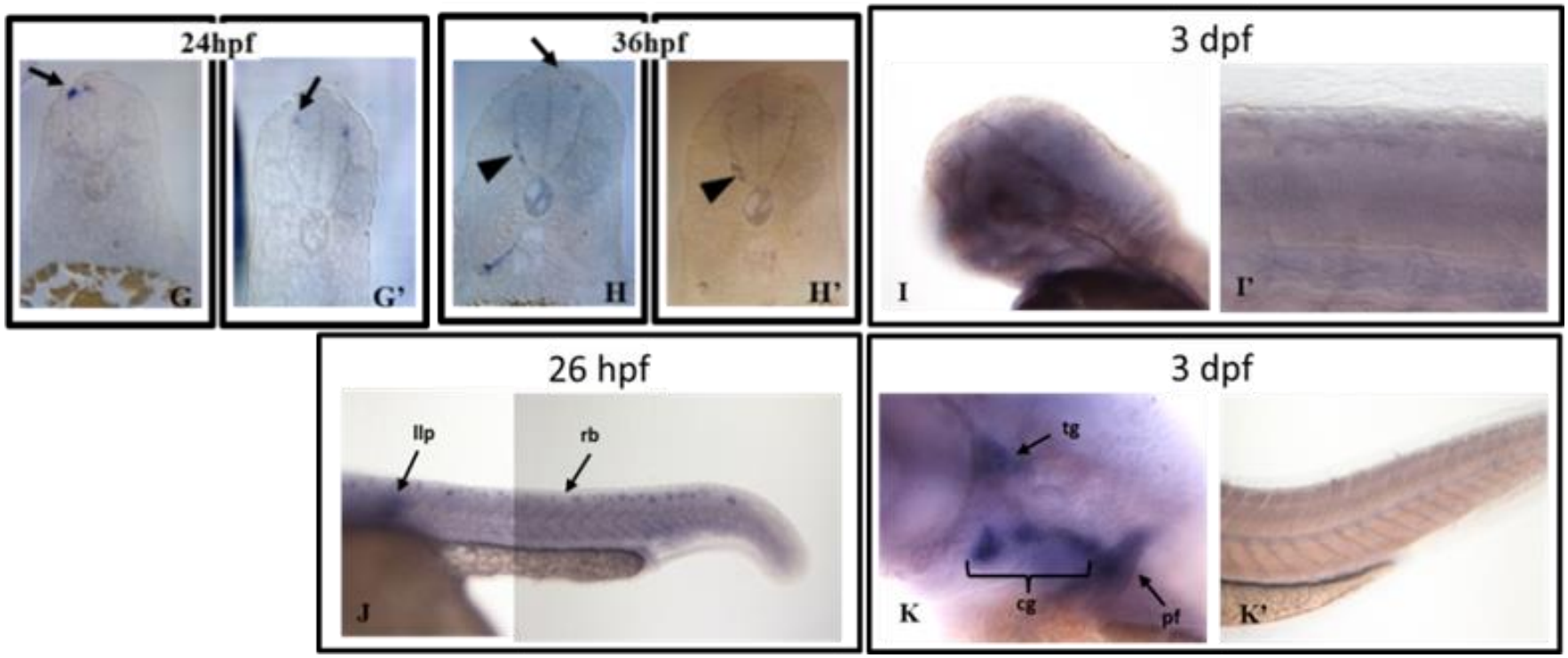


Figure 2.8 Analysis of *runx3* transcripts localization during early zebrafish developmental stages by ISH. Lateral (A, A', B, B', C, C', D', E, E', F', I, I', J, K and K') and dorsal (B'', C'', D and F) views of zebrafish embryos. In all pictures embryos are oriented anterior to the left and dorsal to the top. Transversal cross-sections of the trunk after whole mount ISH (G, G', H, H') showing *runx3* expression in Rohon-Beard cells (arrows in G' G', H) and expressing cells that lay outside the neural tube (arrowheads in H, H'). *runx3* probe complementary to the common region of the different *runx3* transcripts (A-H'), *runx3*-P1 transcripts specific probe (I, I') and *runx3*-P2 transcripts specific probe (J, K, K'). hpf – hours post-fertilization; dpf – days post-fertilization. tg - trigeminal ganglions; rb - Rohon-Beard neurons; llp - lateral line primordial; bpf - base of the developing pectoral fins; csg - cranial sensory ganglia; cc – cranial cartilages.

2.5 Discussion

Prior to this work, there were three major transcripts described for zebrafish *runx3* that are transcribed from two promoters and give rise to two proteins that differ in their N-terminal sequence. The *runx3* P2-derived transcript (named type I variant) is derived from the proximal (or P2) promoter and encodes the 424 amino acid P2-Runx3 protein isoform, also called the Runx3-shorter or Runx3-MHIPV isoform. On the other hand, due to 5' UTR splicing events two major transcripts are transcribed from the Distal (or P1) promoter (named *runx3* Type II and Type III variants) and encoding the 438 amino acid P1-Runx3 protein isoform, also called the Runx3-longer or Runx3-MASN isoform. The only difference between the two isoforms is the N-terminal – the first five amino acids of the Runx3-MHIPV isoform and the first 19 amino acids of the Runx3-MASN isoform are specific to each isoform. After that, both isoforms have exactly the same amino acid sequence and the alignment of the P1 and P2 full-length isoforms of RUNX3 proteins revealed several highly conserved protein domains. As expected, the 128 amino acid runt domain that is critical for DNA binding and heterodimerisation is highly conserved in all RUNX proteins. The NLS, VWRPY and PY motifs are also highly conserved. Our *in silico* results comparing the zebrafish Runx3 and vertebrate orthologs are in agreement with previous reports that have also studied the conservation of the RUNX3 proteins through evolution (Rini and Calabi, 2001; Ng et al, 2007; Nah et al, 2014).

The RUNX family of transcription factors have been extensively studied over the last years, due to their involvement in a variety of biological processes and disease, and they have been shown to be regulated by two alternative promoters that generated different transcript variants, and those are also subject to alternative splicing events generating a variety of RUNX protein isoforms (Coffman, 2003; Komori, 2005; Cohen, 2009; Wang et al, 2010; Chuang et al, 2013; among others). Although they are highly similar in respect to their genomic organization, DNA sequence and transcriptional regulation, functional studies have shown that the disruption of each *RUNX* gene causes different phenotypes, indicating tissue-specific roles Coffman, 2003; Chuang et al, 2013). The regulation of genes by alternative promoters and the existence of multiple alternative splicing variants have been associated with complex regulation of gene expression (Ayoubi and Van De Ven, 1996). RUNX3 is the

less studied of the RUNX family members. In zebrafish Runx3 has been shown to be required for hematopoietic development (Kalev-Zylinska et al, 2003) and to be involved in craniofacial development (Flores et al, 2006), but the importance of each of the P1-Runx3 and P2-Runx3 isoforms is not described.

In this study we investigated the expression pattern of each *runx3* isoform in different stages of zebrafish development and in different adult tissues, in order to identify if the isoforms are differentially regulated. In an attempt to clone the *runx3* ORFs, we identified and cloned multiple transcript variants. Besides the three different transcripts previously described for zebrafish *runx3* (Kataoka et al, 2000) we cloned 10 additional variants obtained by alternative splicing. Of the new variants cloned, four are P1-derived (isoforms 2 to 5) and six are P2-derived (isoforms 7 to 12). Interestingly, in contrast to the other RUNX family members, relatively few isoforms have been described for RUNX3. Moreover, the zebrafish *runx3* isoform 2 described in this study with a splice of exon 5 has already been described for human RUNX3 (Puig-Kroger et al, 2010). These alternatively spliced transcript variants that generate protein isoforms with one or more domains partially or totally absent, suggest that probably Runx3 splicing isoforms have different functions. In fact, Puig-Kroger and co-workers have shown that transcriptional activities of the human RUNX3/p33 isoform, generated by splicing a runt DNA-binding domain-encoding exon, differ from those of the non-spliced isoform (Puig-Kroger et al, 2010). Although some of these alternative spliced isoforms may have important modulatory functions in development or in critical cell fate decisions, others may not be translated due to the process of nonsense-mediated mRNA decay (NMD) that causes degradation of mRNAs containing premature translation termination codons. The NMD process was identified and studied also in zebrafish, and was shown to be essential for zebrafish embryonic development, preventing accumulation of potentially detrimental truncated proteins (Wittkopp et al, 2009). In this study, isoforms 8 to 12 showed splicing across multiple exons, generating either premature termination codons or splicing of the initiation codon, and thus may be potential targets for the NMD pathway, and not likely to be translated into proteins.

The *runx3* isoforms 13 and 14 correspond to transcripts with splicing events in the 5' UTR regions of the *runx3*-P1 transcript. The isoform 13 is similar to the AB043789 transcript

previously described (Kataoka et al, 2000) but with more 313 bp of 5´ UTR, that seems to correspond to an intron retention. Interestingly, one isoform similar to this was cloned for human RUNX1 (type I isoform; Zhang et al, 1997). A new TSS was identified from the P1-*runx3* transcript, isoform 14, that contains an extra 182 bp of the 5´ UTR region and carries the 221 bp splicing event observed in the previously described isoform AB043789 (Kataoka et al, 2000). Analysis of both *runx3* 5' UTR regions revealed the presence of several ATG nucleotide triplets upstream the ATG codon (uATG) that can generate several putative upstream open reading frames (uORFs) that can vary sizes. uATGs and uORFs are common features of mRNAs that encode regulatory proteins and have been shown to profoundly influence translation of the main ORF (Wang and Rothnagel, 2004). Although we did not test for the existence of these uORFs and their effect in the translation of the Runx3 proteins, it would be interesting to analyse if the regulation of the *runx3* isoforms is affected by their complex 5' UTR region.

We analysed the expression of *runx3*-P1 and *runx3*-P2 transcripts both in zebrafish adult tissues and during zebrafish embryonic development by qPCR. We observed that the relative expression of both variants is higher in the calcified tissues compared to the soft tissues. However, the variants showed a different expression pattern in the adult tissues analysed, with *runx3*-P1 showing a relatively similar expression in all tissues analysed, except in the liver and muscle where the expression was barely detectable, and in the pectoral fins and branchial arches where it showed the highest expression. In contrast, the *runx3*-P2 variant showed a more variable relative expression between the tissues analysed with the brain, heart and liver presenting the lowest expression values, and the pectoral fins, mandibular operculum and trunk vertebrae presenting the highest values. These observations are in agreement with a previously published report, where it was shown by S1 nuclease protection assay using RNA from mouse tissues that *Runx3*-P2 is nearly ubiquitously expressed, although *Runx3*-P1 is restricted to few lineages/physiological states (Rini and Calabi, 2001). These differences between the expression patterns of both *runx3*-P1 and *runx3*-P2 transcripts are also observed during the zebrafish embryonic development as the *runx3*-P1 variant showed to be expressed maternally as well as zygotically, in contrast to the *runx3*-P2 variant where the expression was first detected in the stage 24 hpf. In a previous report from Kataoka and colleagues, *runx3* transcript 1 and transcript 2 were detected

maternally, as well as zygotically and the expression of *runx3* transcript 3 was detected from the three somite stage (Kataoka et al, 2000). It is important to note that their nomenclature is somewhat confusing, since transcript 1 derives from the P2 promoter and transcripts 2 and 3 derive from the P1 promoter.

To identify the spatial-temporal expression of the *runx3* transcripts we performed ISH at different zebrafish developmental stages, using RNA specific probes to the common region of all variants, and to the 5' UTR specific of each variant. Using a probe that recognizes a common region of the *runx3* transcripts we confirmed the *runx3* expression previously reported (Kataoka et al, 2000; Burns et al, 2002; Kalev-Zylinkska et al, 2003; Flores et al, 2006) in neural and cartilaginous tissues in the developing embryo. Kalev-Zylinkska and co-workers (2003) have previously shown *runx3* to be expressed in hematopoietic tissues, however we and other authors (Kataoka et al, 2000; Burns et al, 2002) failed to detect this expression. Using the specific probes to each variant we always obtained a high background signal, making it very difficult to reach any conclusion. We could observe that the *runx3*-P2 specific probe reproduces at least partially the expression pattern obtained with the probe to the common region of the *runx3*; in contrast using a *runx3*-P1 specific probe it was not possible to identify any site of expression, since the background signal was always very intense. These high background levels of expression can be due to the small size of the probes (≈180 bp and ≈500 bp, respectively for P1- and P2-derived probes) or due to a low copy number of the transcripts, so optimization to find the better conditions need to be done or, alternatively new ISH techniques should be used. For example, to detect alternative spliced RNA variants that differ only in a short region, the locked nucleic acids (LNAs) probes that exhibits superior specificity, hybridization kinetics and stability (Darnell et al, 2010), are probably a better option. For the detection of low-copy number molecules the amplification of the signal can be essential for its detection, so the use of the recent technique, RNAscope, that uses a unique probe that enables simultaneous signal amplification and background suppression (Wang et al, 2012), may be an alternative. This technique has been recently used to successfully detect multiple RNA molecules in a high resolution and quantitatively manner in zebrafish whole-mount embryos (Gross-Thebing et al, 2014). Together, our qPCR and ISH results suggest that the *runx3*-P2 variants are probably expressed in higher levels and in a variety of tissues, but on the other hand, the *runx3*-P1 variants are likely expressed

in a tissue-specific way and/or in low levels, so that the sensitivity of this technique is not sufficient to detect that expression. In fact, knockdown analysis in zebrafish embryos using an antisense morpholino specific to the Runx3-P2 isoform has been reported to affect the hematopoietic and craniofacial cartilage development (Kalev-Zylinkska et al, 2003; Flores et al, 2006), but no studies have been reported where the Runx3-P1 isoform has been targeted.

For this reason, we attempted a knockdown analysis in zebrafish embryos, using antisense morpholinos specific to each Runx3 isoform to analyse if the resulting morphants have different phenotypes. Levanon and co-workers (2002) reported that Runx3 is expressed in the dorsal root ganglia (DRG) sensory neurons, specifically in a subset of neurons called the tyrosine kinase receptor C (TrkC) proprioceptive neurons, and these authors showed that Runx3-deficient mice develop severe limb ataxia due to disruption of monosynaptic connectivity between intra spinal afferents and motoneurons (Levanon et al, 2002). Using translation-blocking morpholinos targeting Runx3-P1 or Runx3-P2 protein isoforms we explored isoform-specific gene function in DRG development using the generic neuronal marker anti-Hu. Our preliminary results suggest that targeting the Runx3-P1 protein isoform resulted in a reduced number of DRG sensory neurons and in the presence of ectopic neurons in a position directly dorsal to the spinal cord. In contrast, the results targeting the Runx3-P2 protein isoform, showed no effect in DRGs (**Appendix 2.1**). However, we are aware of the increasing reports showing that morpholinos can cause artefacts due to the non-specific minority of morpholinos efficiently mimic mutant phenotypes (Eisen and Smith, 2008; Kok et al, 2015). So, at the moment our results must be treated as preliminary and more work has to be developed. Recently, new techniques to generate targeted mutations in genes of interest have been successfully used in zebrafish, namely transcription activator-like effector nucleases (TALENs; Cade et al, 2012; Xiao et al, 2013) and clustered regularly interspaced short palindromic repeats (CRISPR) RNA-guided Cas9 nuclease system (Hwang et al, 2013a; 2013b; Xiao et al, 2013; Jao et al, 2013; among others). As the latter system has been shown to be simple, robust and a very efficient reverse genetic tool, we intend to use it to confirm our preliminary results using morpholinos and to analyse in more detail the effect of depletion of the Runx3 isoforms in the zebrafish development.

In conclusion, we cloned for the first time new zebrafish *runx3* transcript variants that are generated by alternative splicing and derived from the alterative P1 and P2 promoters. We analysed the specific *runx3*-P1 and *runx3*-P2 spatial-temporal analysis either by qPCR or by ISH, and showed that the transcript variants are differentially expressed in zebrafish adult tissues and also during the development, where the *runx3*-P1 variant is maternally and zygotically expressed in contrast to the *runx3*-P2 variant that is only expressed after the zygotic transcription started. These results together with our preliminary results of Runx3 knockdown analysis strongly suggest that Runx3 protein isoforms are differently regulated and play different roles in the developing embryo.

Acknowledgments

This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Program and national funds through FCT – Foundation for Science and Technology, under the project "PEst-C/MAR/LA0015/2011. NC and BS are supported, respectively, by a post-doctoral and doctoral grant from FCT (SFRH/BPD/48206/2008 and SFRH/BD/38083/2007).

2.6 Supplementary Tables

Table 2.S1 PCR primers used in this chapter.

Sense primers	Primer sequence (5' to 3')[#]
DrRunx3_F3	GGCAGGAGCGGCAGAGGAAAGA
DrRunx3_F5	TGTGATGCATATTCCCGTAGACCCG
DrRunx3I_F1	CACAACGGACTTTCTGTAGCCT
DrRunx3I_F2	GCGGCGTTCGCGTAGTCG
DrRunx3I_F4	GTATTAACGCTTCTGCATCTGCATTG
DrRunx3I_F5	TGACAGGACGGAAAGTAACTGAATGA
DrRunx3II_F1	TGACCTCGGATACCTCAGACA
DrRunx3II_F2	GCAACCACTTCTGGAGAGAGAGAGAC
DrRunx3II&III_F1	AAGACGGTACGACGGCGGAGAGGC
DrRunx3I_F1_XhoI	CCGGAGCTCGAGCACAACGGACTTTCTGTAGCCT
DrRunx3II&III_F1_XhoI	CCGGAGCTCGAGGGCTGAAGATGGTACGACGG
Antisense primers	**Primer sequence (5' to 3')[#]**
DrRunx3_R2	TGCATCATGCGCAACTCTTCTGG
DrRunx3_R4	CGTCTGCTCGTGCTCGGGTCT
DrRunx3_R6	ATCTTAGTACGGCCTCCAGACAGACTC
DrRunx3_R8	ATGCAAGACGTTTCACCTCTTCTCTT
DrRunx3I_R1	CTCGTGCTCGGGTCTACGGGAA
DrRunx3II&III_R1	CGCAGCAAACTGGGCGAGTAG
DrRunx3I_R1_KpnI	CACGCGGTACCCTCGTGCTCGGGTCTACGGGAA
DrRunx3II&III_R1_KpnI	CACGCGGTACCCGCAGCAAACTGGGCGAGTAG
qPCR primers	**Primer sequence (5' to 3')**
DrRunx3II&III_F1	AAGACGGTACGACGGCGGAGAGGC
DrRunx3II&III_R2	CGCAGCAAACTGGGCGAGTAGCTAGAGA
DrRunx3I_F3	ACGCATTCATTTGGAAAGAACGCTTGAC
DrRunx3I_R2	CAGCAGGTGTGAGCTGCTCCTGAATATCAGT
DrEf1α_F1	ACGCCCTCCTGGCTTTCACCC
DrEf1α_R1	TGGGACGAAGGCAACACTGGC
Dr18s_F1	ACCACCCACAGAATCGAGAAA
Dr18s_R1	GCCTGCGGCTTAATTTGACT
Universal primers	**Primer sequence (5' to 3')**
Adaptor Primer 1 (AP1)	CCATCCTAATACGACTCACTATAGGGC
Adaptor Primer 2 (AP2)	ACTCACTATAGGGCTCGAGCGGC
oligo(dT)-adapter primer	ACGCGTCGACCTCGAGATCGATGTTTTTTTTTTTT

[#] indicates the recognition sequence for the restriction enzyme named on the primer.

Table 2.S2 Accession numbers of all Runx3 protein sequences used for the *in silico* analysis.

Species	Protein isoform	N-Terminal	Acession number
Amazon molly	P1-Runx3	MASN	ENSPFOT00000030105
Callorhinchus milii	P1-Runx3	MASN	AHW58145.1
Danio rerio	P1-Runx3	MASN	BAB17904.1
Homo sapiens	P1-Runx3	MASN	NP_001026850.1
Lepisosteus oculatus	P1-Runx3	MASN	ENSLOCT00000003239
Loxodonta africana	P1-Runx3	MASN	ENSLAFT00000008382
Mus musculus	P1-Runx3	MASN	NP_062706.2
Olive baboon	P1-Runx3	MASN	ENSPANT00000016693
Scyliorhinus canicula	P1-Runx3	MASN	ABL68117.1
Takifugu rubripes	P1-Runx3	MASN	AAU14193.1
Tetraodon nigroviridis	P1-Runx3	MASN	ENSTNIT00000001252
Bos taurus	P2-Runx3	MRIPV	NP_001180087.1
Ctenopharyngodon idella	P2-Runx3	MRIPV	AHJ59933.1
Danio rerio	P2-Runx3	MHIPV	BAB17903.1
Homo sapiens	P2-Runx3	MRIPV	NP_004341.1
Macaca mulatta	P2-Runx3	MRIPV	AFH34424.1
Mus musculus	P2-Runx3	MRIPV	EDL29993.1
Oreochromis niloticus	P2-Runx3	MHIPV	ENSONIT00000010344
Rattus norvegicus	P2-Runx3	MRIPV	NP_569109.1
Scyliorhinus canicula	P2-Runx3	MRIPV	ABL68116.1
Takifugu rubripes	P2-Runx3	MHIPV	NP_001092121.1
Xenopus laevis	P2-Runx3	MLIPV	NP_001182313.1
Xiphophorus maculatus	P2-Runx3	MHIPV	ENSXMAT00000018904

2.7 Supplementary Figures

> AB043788

GCGCGGCGTTCGCGTAGTCGTATGTCTGACTAACACAAACCACTTTTCTCGTTTTAACTCACTTTTACGCATTCATTTGGAAAGAACGCTTGACA
GCTGAAAGGGACTTTGCTGCAGAAACTCTTCTGAAACTGAGGGATTAATTTGTTTGGACTTTATTTTCCGAGACTCTGGATTGACAGGACGGAA
AGTAACTGAATGAAAGTAAAACTGATATTCAGGAGCAGCTCACACCTGCTGTACACGCTCTGGGATGAATTCAGTTTCATAAATAACTCTTAAA
TAGTAGTTGTATATGCTTTGAGAAGCATACAAGACTCAGTTTTTATGCCTTACTTTCATCTATCGTGCAGCACGAGATTGTATGTGCTTGCGTG
TGTGTGTGTGGCTAAAGTCAGTGGGGAACTCGCTCTGTATTAACGCTTCTGCATCTGCATTGCAACACAAAGTAAACGGTCACCGAGGAGTTTT
TTCCACTCTCCAACACAACGGACTTTCTGTAGCCTACTCAACCAACTGGGTTGCCCGAGCTTTTCACGAGGACTCATTTGAAATCTGTTGTGATG

5'3' Frame 1
ARRSRSR**M**SD**Stop**HKPLFSF**Stop**LTFTHSFGKNA**Stop**QLKGTLLQKLF**Stop**N**Stop**GINLFGLYFPRLWIDRTESN**Stop****M**KVKLIFRSSSHLLYTLWDEFSFI
NNS**Stop**IVVVYALRSIQDSVF**M**PYFHLSCSTRLYVLACVCVAKVSGELALY**Stop**RFCICIATQSKRSPRSFFHSPTQRTFCSLLNQLGCPSFSRGLI**Stop**NLL**Sto**
p

5'3' Frame 2
RGVRVVVCLTNTNHFSRFNSLLRIHLERTLDS**Stop**KGLCCRNSSETEGLICLDFIFRDSGLTGRKVTE**Stop**K**Stop**N**Stop**YSGAAHTCCTRSG**M**NSVS**Stop**ITL
K**Stop****Stop**LY**ML****Stop**EAYKTQFLCLTFIYRAARDC**M**CLRVCVWLKSVGNSLCINASASALQHKVNGHRGVFSTLQHNGLSVAYSTNWVARAFHEDSFEICCD

5'3' Frame 3
AAFA**Stop**SYV**Stop**LTQTTFLVLTHFYAFIWKERLTAERDFAAETLLKLRD**Stop**FVWTLFSETLD**Stop**QDGK**Stop**LNESKTDIQEQLTPAVHALG**Stop**IQFH
K**Stop**LLNSSCICFEKHTRLSFYALLSSIVQHEIVCACVCVCG**Stop**SQWGTRSVLTLLHLHCNTK**Stop**TVTEEFFPLSNTTDFL**Stop**PTQPTGLPELFTRTHLKS
VV**M**

> AB043789

GATACCGCGGGAGCTGTGACCTCGGATACCTCAGACATGCGACTGACAGGCCTTTACCACGGTGATCACCACTGTAAGAGCGTCAGACACCTGAC
ATTTCAACTCTCTGACTGAAGGCACGTAGACTCTACAGCAACCACTTCTGGAGAGAGAGAGACGCACAGAAGTGTCCATCAACAAAGGCTGAAG
ATGGTACGACGGCGGAGAGGCCAGCTCTTTTCCCTTCAGAGCGGGACACATG

5'3' Frame 1
DTAGAVTSDTSD**M**RLTGLYHGDHHCKSVRHLTFQLSD**Stop**RHVDSTATTSGERETHRSVHQQRLK**M**VRRRRGQLFSLQSGTH

5'3' Frame 2
IPREL**Stop**PRIPQTCD**Stop**QAFTTVITTVRASDT**Stop**HFNSLTEGT**Stop**TLQQPLLERERRTEVSINKG**Stop**RWYDGGEASSFPFRAGH**M**

5'3' Frame 3
YRGSCDLGYLRHATDRPLPR**Stop**SPL**Stop**ERQTPDISTL**Stop**LKARRLYSNHFWRERDAQKCPSTKAEDGTTAERPALFPSERDT

>AB043790

AAACCCCAAATTCTTGGTTGCAGGCTGTCTATGTAGTTTCTGTATTTCTCTTTCTCCGACTGTCATTCTCTCAGGCTGAAGATGGTACGACGGCG
GAGAGGCCAGCTCTTTTCCCTTCAGAGCGGGACACATG

5'3'Frame 1
KPQILGCRLS**M****Stop**FLYFSFSDCHSLRLK**M**VRRRRGQLFSLQSGTH

5'3'Frame 2
NPKFLVAGCLCSFCISLSPTVILSG**Stop**RWYDGGEASSFPFRAGH**M**

5'3'Frame 3
TPNSWLQAVYVVSVFLFLRLSFSQAEDGTTAERPALFPSERDT

>Runx3_iso13

TGACCTCGGATACCTCAGACATGCGACTGACAGGCCTTTACCACGGTGATCACCACTGTAAGAGCGTCAGACACCTGACATTTCAACTCTCTGAC
TGAAGGCACGTAGACTCTACAGCAACCACTTCTGGAGAGAGAGAGACGCACAGAAGTGTCCATCTACAAAGGTGAGTCCATTTCTATGAGACTT
AGTGCTGTCCTCACTTTGTCTTTCAGTCTTTCTCTCTTTCTCTCTTACTCTCGCTCGCTCTCACCACACGGCACATATGCTGTTCTGTTAGTCGGC
TTGTGGCGAAAGATTCCTGTGGCACTCTCAAACCCCAAATTCTTGGTTGCAGGCTGTCTATGTAGTTTCTGTATTTCTCTTTCTCCGACTGTCAT
TCTCTCAGGCTGAAGACGGTACGACGGCGGAGAGGCCAGCTCTTTTCCCTTCAGAGCGGGACACATG

5'3'Frame 1
StopPRIPQTCD**Stop**QAFTTVITTVRASDT**Stop**HFNSLTEGT**Stop**TLQQPLLERERRTEVSIYKGESIS**M**RLSAVLTLSFSLSLFLSYSRSLSPHGTYAVLLVGLW
RKIPVALSNPKFLVAGCLCSFCISLSPTVILSG**Stop**RRYDGGEASSFPFRAGH**M**

5'3'Frame 2
DLGYLRHATDRPLPR**Stop**SPL**Stop**ERQTPDISTL**Stop**LKARRLYSNHFWRERDAQKCPSTKVSPFL**Stop**DLVLSSLCLSVFLSFSLTLARSHHTAH**M**LFC**Sto**
pSACGERFLWHSQTPNSWLQAVYVVSVFLFLRLSFSQAEDGTTAERPALFPSERDT

5'3'Frame 3
TSDTSD**M**RLTGLYHGDHHCKSVRHLTFQLSD**Stop**RHVDSTATTSGERETHRSVHLQR**Stop**VHFYET**Stop**CCPHFVFQSFSLSLLLSLALTTRHICCSVSRLVA
KDSCGTLKPQILGCRLS**M****Stop**FLYFSFSDCHSLRLKTVRRRRGQLFSLQSGTH

>Runx3_iso14

CTCTTCAAGAAGGTCCCATCATCTTCTGGGAAACTCTTCAAACACCAGCCATGGGCCAATCAGATGAGTGAGAAAGGCACTGGGAAGAAAAGGA
GAACTTGAATGCATGCAAGCGAGGGAGAGAGAGAACGTCATGTGACTGGAGTTCCACCTATCAGATCATGGGATTGGCGAGCAGCAGTGATACC
GCGGGAGCTGTGACCTCGGATACCTCAGACATGCGACTGACAGGCCTTTACCACGGTGATCACCACTGTAAGAGCGTCAGACACCTGACATTTCA
ACTCTCTGACTGAAGGCACGTAGACTCTACAGCAACCACTTCTGGAGAGAGAGAGACGCACAGAAGTGTCCATCAACAAAGGCTGAAGATGGTA
CGACGGCGGAGAGGCCAGCTCTTTTCCCTTCAGAGCGGGACACATG

5'3'Frame 1
LFKKVPSSSGKLFKHQPWANQMSEKGTGKKRRTStopMHASEGERERHVTGVPPIRSWDWRAAVIPRELStopPRIPQTCDStopQAFTTVITTVRASDTStop
HFNSLTEGTStopTLQQPLLERERRTEVSINKGStopRWYDGGEASSFPFRAGHM

5'3'Frame 2
SSRRSHHLLGNSSNTSHGPIRStopVRKALGRKGELECMQARERENVMStopLEFHLSDHGIGEQQStopYRGSCDLGYLRHATDRPLPRStopSPLStopERQTP
DISTLStopLKARRLYSNHFWRERDAQKCPSTKAEDGTTAERPALFPSERDT

5'3'Frame 3
LQEGPIIFWETLQTPAMGQSDEStopERHWEEKENLNACKRGRERTSCDWSSTYQIMGLASSSDTAGAVTSDTSDMRLTGLYHGDHHCKSVRHLTFQLSDSto
pRHVDSTATTSGERETHRSVHQQRLKMVRRRRGQLFSLQSGTH

> AB043788

CGCGGCGTTCGCGTAGTCGTATGTCTGACTAACACAAACCACTTTTCTCGTTTTAACTCACTTTTACGCATTCATTTGGAAAGAACGCTTGACAG
CTGAAAGGGACTTTGCTGCAGAAACTCTTCTGAAACTGAGGGATTAATTTGTTTGGACTTTATTTTCCGAGACTCTGGATTGACAGGACGGAAA
GTAACTGAATGAAAGTAAAACTGATATTCAGGAGCAGCTCACACCTGCTGTACACGCTCTGGGATGAATTCAGTTTCATAAATAACTCTTAAAT
AGTAGTTGTATATGCTTTGAGAAGCATACAAGACTCAGTTTTTATGCCTTACTTTCATCTATCGTGCAGCACGAGATTGTATGTGCTTGCGTGT
GTGTGTGTGGCTAAAGTCAGTGGGGAACTCGCTCTGTATTAACGCTTCTGCATCTGCATTGCAACACAAAGTAAACGGTCACCGAGGAGTTTTT
TCCACTCTCCAACACAACGGACTTTCTGTAGCCTACTCAACCAACTGGGTTGCCCGAGCTTTTCACGAGGACTCATTTGAAATCTGTTGTGATG

5'3' Frame 1
R G V R V V V C L T N T N H F S R F N S L L R I H L E R T L D S **Stop** K G L C C R N S S E T E G L I C L D F I F R D S G L T G R K V T
E **Stop** K **Stop** N **Stop** Y S G A A H T C C T R S G **M** N S V S **Stop** I T L K **Stop Stop** L Y **M** L **Stop** E A Y K T Q F L C L T F I Y R A A R D C **M** C L R V C V
W L K S V G N S L C I N A S A S A L Q H K V N G H R G V F S T L Q H N G L S V A Y S T N W V A R A F H E D S F E I C C D

5'3' Frame 2
A A F A **Stop** S Y V **Stop** L T Q T T F L V L T H F Y A F I W K E R L T A E R D F A A E T L L K L R D **Stop** F V W T L F S E T L D **Stop** Q D G K **Stop** L N
E S K T D I Q E Q L T P A V H A L G **Stop** I Q F H K **Stop** L L N S S C I C F E K H T R L S F Y A L L S S I V Q H E I V C A C V C V C G **Stop** S Q W G T R S V
L T L L H L H C N T K **Stop** T V T E E F F P L S N T T D F L **Stop** P T Q P T G L P E L F T R T H L K S V V

5'3' Frame 3
R R S R S R **M** S D **Stop** H K P L F S F **Stop** L T F T H S F G K N A **Stop** Q L K G T L L Q K L F **Stop** N **Stop** G I N L F G L Y F P R L W I D R T E S
N **Stop** **M** K V K L I F R S S S H L L Y T L W D E F S F I N N S **Stop** I V V V Y A L R S I Q D S V F **M** P Y F H L S C S T R L Y V L A C V C V A K V S G E L A
L Y **Stop** R F C I C I A T Q S K R S P R S F F H S P T Q R T F C S L L N Q L G C P S F S R G L I **Stop** N L L **Stop**

Figure 2.S1 Both *runx3*-P1 and *runx3*-P2 5' UTR regions show the presence of multiple ATG nucleotide triplets upstream the ATG codon that initiates translation of the Runx3 proteins. The described ATG codon and respective Methionine is highlighted in black and the upstream ATG nucleotide triplets and respective putative uORFs are highlighted in grey. Putative uORFs were predicted using the Translate tool available in the ExPASy Portal (http://web.expasy.org/translate/).

2.8 References

Ayoubi TA, Van De Ven WJ, 1996. Regulation of gene expression by alternative promoters. FASEB J. 10:453-460.

Brenner O, Levanon D, Negreanu V, Golubkov O, Fainaru O, Woolf E, Groner Y, 2004. Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. Proc Natl Acad Sci USA. 101:16016-16021.

Burns CE, DeBlasio T, Zhou Y, Zhang J, Zon L, Nimer SD, 2002. Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. Exp Hematol. 30:1381-1389.

Canon J, Banerjee U, 2000. Runt and Lozenge function in Drosophila development. Semin Cell Dev Biol. 11:327-336.

Choi JY, Pratap J, Javed A, Zaidi SK, Xing L, Balint E, Dalamangas S, Boyce B, van Wijnen AJ, Lian JB, Stein JL, Jones SN, Stein GS, 2001. Subnuclear targeting of Runx/Cbfa/AML factors is essential for tissue-specific differentiation during embryonic development. Proc Natl Acad Sci USA. 98:8650-8655.

Chomczynski P, Sacchi N, 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 162:156-159.

Chuang LS, Ito K, Ito Y, 2013. RUNX family: Regulation and diversification of roles through interacting proteins. Int J Cancer. 132:1260-1271.

Coffman J, 2003. Runx transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol Int. 27:315-324.

Cohen MM Jr, 2009. Perspectives on RUNX genes: an update. Am J Med Genet A. 149A:2629-2646.

Cole LK, Ross LS, 2001. Apoptosis in the developing zebrafish embryo. Dev Biol. 240:123-142.

Darnell DK, Stanislaw S, Kaur S, Antin PB, 2010. Whole mount in situ hybridization detection of mRNAs using short LNA containing DNA oligonucleotide probes. RNA. 16:632-637.

D'Souza RN, Aberg T, Gaikwad J, Cavender A, Owen M, Karsenty G, Thesleff I, 1999. Cbfa1 is required for epithelial-mesenchymal interactions regulating tooth development in mice. Development. 126:2911-2920.

Eisen JS, Smith JC, 2008. Controlling morpholino experiments: don't stop making antisense. Development. 135:1735-1743.

Friedrich MJ, Rad R, Langer R, Voland P, Hoefler H, Schmid RM, Prinz C, Gerhard M, 2006. Lack of RUNX3 regulation in human gastric cancer. J Pathol. 210:141-146.

Geoffroy V, Corral DA, Zhou L, Lee B, Karsenty G, 1998. Genomic organization, expression of the human CBFA1 gene, and evidence for an alternative splicing event affecting protein function. Mamm Genome. 9:54-57.

Gergen JP, Butler BA, 1988. Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. Genes Dev. 2:1179-1193.

Golling G, Li L, Pepling M, Stebbins M, Gergen JP, 1996. Drosophila homologs of the proto-oncogene product PEBP2/CBF beta regulate the DNA-binding properties of Runt. Mol Cell Biol. 16:932-942.

Gross-Thebing T, Paksa A, Raz E, 2014. Simultaneous high-resolution detection of multiple transcripts combined with localization of proteins in whole-mount embryos. BMC Biol. 12: 55.

Growney JD, Shigematsu H, Li Z, Lee BH, Adelsperger J, Rowan R, Curley DP, Kutok JL, Akashi K, Williams IR, Speck NA, Gilliland DG, 2005. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood. 106:494-504.

Hecht J, Stricker S, Wiecha U, Stiege A, Panopoulou G, Podsiadlowski L, Poustka AJ, Dieterich C, Ehrich S, Suvorova J, Mundlos S, Seitz V, 2008. Evolution of a core gene network for skeletogenesis in chordates. PLoS Genet. 4(3):e1000025.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK, 2013a. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 31:227-229.

Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, Yeh JR, 2013b. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS One.8:e68708.

Ichikawa M, Asai T, Chiba S, Kurokawa M, Ogawa S, 2004. Runx1/AML-1 ranks as a master regulator of adult hematopoiesis. Cell Cycle. 3:722-724.

Inoue K, Ozaki S, Shiga T, Ito K, Masuda T, Okado N, Iseda T, Kawaguchi S, Ogawa M, Bae SC, Yamashita N, Itohara S, Kudo N, Ito Y, 2002. Runx3 controls the axonal projection of proprioceptive dorsal root ganglion neurons. Nat Neurosci. 5:946-954.

Inoue K, Ozaki S, Ito K, Iseda T, Kawaguchi S, Ogawa M, Bae SC, Yamashita N, Itohara S, Kudo N, Ito Y, 2003. Runx3 is essential for the target-specific axon pathfinding of trkc-expressing dorsal root ganglion neurons. Blood Cells Mol Dis. 30:157-160.

Ito Y, 2004. Oncogenic potential of the RUNX gene family: “overview”. Oncogene. 23:4198-4208.

Jao LE, Wente SR, Chen W, 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci USA. 110:13904-13909.

Kagoshima H, Akamatsu Y, Ito Y, Shigesada K, 1996. Functional dissection of the α and β subunits of transcription factor PEBP2 and the redox susceptibility of its DNA binding activity. J Biol Chem. 271:33074-33082.

Kagoshima H, Shigesada K, Satake M, Ito Y, Miyoshi H, Ohki M, Pepling M, Gergen P, 1993. The Runt domain identifies a new family of heteromeric transcriptional regulators. Trends Genet. 9:338-341.

Kalev-Zylinska ML, Horsfield JA, Flores MVC, Postlethwait JH, Vitas MR, Baas AM, Crosier PS, Crosier KE, 2002. Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. Development. 129:2015-2030.

Kalev-Zylinska ML, Horsfield JA, Flores MVC, Postlethwait JH, Chau JYM, Cattin PM, Vitas MR, Crosie PS, Crosier KE, 2003. Runx3 is required for hematopoietic development in zebrafish. Dev Dyn. 228:323-336.

Kataoka H, Ochi M, Enomoto K, Yamaguchi A, 2000. Cloning and embryonic expression patterns of the zebrafish Runt domain genes, runxa and runxb. Mech Dev. 98:139-143.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF, 1995. Stages of embryonic development of the zebrafish. Dev Dyn. 203:253-310.

Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, Kirchmaier BC, Peterson-Maduro J, Kourkoulis G, Male I, DeSantis DF, Sheppard-Tindell S, Ebarasi L, Betsholtz C, Schulte-Merker S, Wolfe SA, Lawson ND, 2015. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell. 32:97-108.

Komori T, 2005. Regulation of skeletal development by the Runx family of transcription factors. J Cell Biochem. 95:445-453.

Komori T, Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, Shimizu Y, Bronson RT, Gao YH, Inada M, Sato M, Okamoto R, Kitamura Y, Yoshiki S, Kishimoto T, 1997. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell. 89:755-764.

Komori T, 2015. The functions of Runx family transcription factors and Cbfb in skeletal development. Oral Science International. 12:1-4.

Le X-F, Groner Y, Kornblau SM, Gu Y, Hittelman WN, Levanon D, Mehta K, Arlinghaus RB, Chang K-S, 1999. Regulation of AML2/CBFA3 in Hematopoietic Cells through the Retinoic Acid Receptor α-Dependent Signaling Pathway. J Biol Chem. 274:21651-21658.

Lee J-M, Shin J-O, Cho K-W, Hosoya A, Cho S-W, Lee Y-S, Ryoo H-M, Bae S-C, Jung H-S, 2011. Runx3 is a crucial regulator of alveolar differentiation and lung tumorigenesis in mice. Differentiation. 81:261-268.

Levanon D, Bettoun D, Harris-Cerruti C, Woolf E, Negreanu V, Eilam R, Bernstein Y, Goldenberg D, Xiao C, Fliegauf M, Kremer E, Otto F, Brenner O, Lev-Tov A, Groner Y, 2002. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. EMBO J. 21:3454-3463.

Levanon, D, Glusman G, Bangsow T, Ben-Asher E, Male D, Avidan N, Bangsow C, Hattori M, Taylor TD, Taudien S, Blechschmidt K, Shimizu N, Rosenthal A, Sakaki Y, Lancet D, Groner Y, 2001. Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. Gene. 262:23-33.

Makita N, Suzuki M, Asami S, Takahata R, Kohzaki D, Kobayashi S, Hakamazuka T, Hozumi N, 2008. Two of four alternatively spliced isoforms of RUNX2 control osteocalcin gene expression in human osteoblast cells. Gene. 413:8-17.

de Bruijn MF, Speck NA, 2004. Core-binding factors in hematopoiesis and immune function. Oncogene. 23:4238-4248.

Maruyama Z, Yoshida C, Furuichi T, Amizuka N, Ito M, Fukuyama R, Miyazaki T, Kitaura H, Nakamura K, Fujita T, Kanatani N, Moriishi T, Yamana K, Liu W, Kawaguchi H, Nakamura K, Komori T, 2007. Runx2 determines bone maturity and turnover rate in postnatal bone development and is involved in bone loss in estrogen deficiency. Dev Dyn. 236:1876-1890.

Meyers S, Downing JR, Hiebert SW, 1993. Identification of AML-1 and the (8 ; 21) Translocation Protein (AML-1/ETO) as Sequence-Specific DNA-Binding Proteins: the runt Homology Domain Is Required for DNA Binding and Protein-Protein Interactions. Mol Cell Biol. 13:6336-6345.

Miyoshi H, Ohira M, Shimizu K, Mitani K, Hirai H, Imai T, Yokoyama K, Soeda E, Ohki M, 1995. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. Nucleic Acids Res. 23:2762-2769.

Nah GS, Lim ZW, Tay BH, Osato M, Venkatesh B, 2014. Runx family genes in a cartilaginous fish, the elephant shark (*Callorhinchus milii*). PLoS One. 3:9(4):e93816.

Nakamura S, Senzaki K, Yoshikawa M, Nishimura M, Inoue K, Ito Y, Ozaki S, Shiga T, 2008. Dynamic regulation of the expression of neurotrophin receptors by Runx3. Development. 135:1703-1711.

Ng CE, Osato M, Tay BH, Venkatesh B, Ito Y, 2007. cDNA cloning of Runx family genes from the pufferfish (Fugu rubripes). Gene. 399:162-173.

Notredame C, Higgins DG, Heringa J, 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302:205-217.

Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR, 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell. 84:321-330.

Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GWH, Beddington RS, Mundlos S, Olsen BR, Selby PB, Owen MJ, 1997. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell. 89:765-771.

Pepling ME, Gergen JP, 1995. Conservation and function of the transcriptional regulatory protein Runt. Proc Natl Acad Sci USA. 92:9087-9091.

Puig-Kröger A, Aguilera-Montilla N, Martínez-Nuñez R, Domínguez-Soto A, Sánchez-Cabo F, Martín-Gayo E, Zaballos A, Toribio ML, Groner Y, Ito Y, Dopazo A, Corcuera MT, Alonso Martín MJ, Vega MA, Corbí AL, 2010. The novel RUNX3/p33 isoform is induced upon monocyte-derived dendritic cell maturation and downregulates IL-8 expression. Immunobiology. 215:812-820.

Reis BS, Rogoz A, Costa-Pinto FA, Taniuchi I, Mucida D, 2013. Mutual expression of the transcription factors Runx3 and ThPOK regulates intestinal $CD4^+$ T cell immunity. Nat Immunol. 14:271-280.

Rennert J, Coffman JA, Mushegian AR, Robertson AJ, 2003. The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms. BMC Evol Biol. 3:4.

Reyes R, Haendel M, Grant D, Melancon E, Eisen JS, 2004. Slow degeneration of zebrafish Rohon-Beard neurons during programmed cell death. Dev Dyn. 229:30-41.

Rini D, Calabi F, 2001. Identification and comparative analysis of a second runx3 promoter. Gene. 273:13-22.

Sambrook J, Fritsch EF, Maniatis T, 1989. Molecular Cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor New York.

Schneider TD, Stephens RM, 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097-6100.

Stothard P, 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 28:1102-1104.

Taniuchi I, Osato M, Egawa T, Sunshine MJ, Bae SC, Komori T, Ito Y, Littman DR, 2002. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. Cell. 111:621-633.

Terry A, Kilbey A, Vaillant F, Stewart M, Jenkins A, Cameron E, Neil JC, 2004. Conservation and expression of an alternative 3’ exon of Runx2 encoding a novel proline-rich C-terminal domain. Gene. 336:115-125.

Tighe JE, Calabi F, 1994. Alternative, out-of-frame runt/MTG8 transcripts are encoded by the derivative (8) chromosome in the t(8;21) of acute myeloid leukemia M2. Blood. 84:2115-2121.

Thisse C, Thisse B, Schilling TF, Postlethwait JH, 1993. Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. Development. 119:1203-1215.

Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Ungos JM, Karlstrom RO, Raible DW, 2003. Hedgehog signaling is directly required for the development of zebrafish dorsal root ganglia neurons. Development. 130, 5351-5362.

van der Velden AW, Thomas AA, 1999. The role of the 5' untranslated region of an mRNA in translation regulation during development. Int J Biochem Cell Biol. 31:87-106.

Wang CQ, Jacob B, Nah GSS, Osato M, 2010. Runx family genes, niche, and stem cell quiescence. Blood Cells Mol Dis. 44:275-286.

Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ, Luo Y, 2012. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J Mol Diagn. 14:22-29.

Wang XQ, Rothnagel JA, 2004. 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Res. 32:1382-1391.

Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, Speck NA, 1996. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci USA. 93:3444-3449.

Wang CQ, Motoda L, Satake M, Ito Y, Taniuchi I, Tergaonkar V, Osato M, 2013. Runx3 deficiency results in myeloproliferative disorder in aged mice. Blood. 122:562-566.

Westerfield M, 2000. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene.

Wheeler J, Shigesada K, Petergergen J, Ito Y, 2000. Mechanisms of transcriptional regulation by Runt domain proteins. Semin Cell Dev Biol. 11:369-375.

Williams JA, Barrios A, Gatchalian C, Rubin L, Wilson SW, Holder N, 2000. Programmed cell death in zebrafish Rohon Beard neurons is influenced by TrkC1/NT3 signaling. Dev Biol. 226:220-230.

Wittkopp N, Huntzinger E, Weiler C, Saulière J, Schmidt S, Sonawane M, Izaurralde E, 2009. Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. Mol. Cell Biol. 29:3517-3528.

Van Wijnen AJ, Stein GS, Gergen JP, Groner Y, Hiebert SW, Ito Y, Liu P, Neil JC, Ohki M, Speck N, 2004. Nomenclature for Runt-related (RUNX) proteins. Oncogene. 23:4209-4210.

Xiao A, Wang Z, Hu Y, Wu Y, Luo Z, Yang Z, Zu Y, Li W, Huang P, Tong X, Zhu Z, Lin S, Zhang B, 2013. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic Acids Res.41:e141.

Xiao ZS, Thomas R, Hinson TK, Quarles LD, 1998. Genomic structure and isoform expression of the mouse, rat and human Cbfa1/Osf2 transcription factor. Gene. 214:187-197.

Yoshida CA, Yamamoto H, Fujita T, Furuichi T, Ito K, Inoue K, Yamana K, Zanma A, Takada K, Ito Y, Komori T, 2004. Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. Genes Dev. 18:952-963.

Zhang M, Xie R, Hou W, Wang B, Shen R, Wang X, Wang Q, Zhu T, Jonason JH, Chen D, 2009. PTHrP prevents chondrocyte premature hypertrophy by inducing cyclin-D1-dependent Runx2 and Runx3 phosphorylation, ubiquitylation and proteasomal degradation. J Cell Sci. 122:1382-1389.

Zhang YW, Bae SC, Huang G, Fu YX, Lu J, Ahn MY, Kanno Y, Kanno T, Ito Y, 1997. A novel transcript encoding an N-terminally truncated AML1/PEBP2 alphaB protein interferes with transactivation and blocks granulocytic differentiation of 32Dcl3 myeloid cells. Mol Cell Biol. 17:4133-4145.

Appendix 1 - Knockdown zebrafish Runx3 protein function to investigate effect on DRG sensory neurons

Methods

Zebrafish Runx3 protein function knockdown by morpholinos

Antisense morpholinos complementary to the 5' UTR near the ATG translation start codon of each Runx3 isoform and to the exon 3-intron 3 splice junction were designed and synthesised by Gene Tools LLC (www.gene-tools.com/). Runx3 morpholino sequences were as follows: Mo-L, 5'- AGATGCTGTTTGAAGCCATGTGTCC-3'; Mo-S, 5' CACAACAGATTTCAAATGAGTCCTC-3'; Mo-Sp, 5'- TGTTTAGATGTGGCACCTCTGCCGC-3'; Mo-M, 5'- AGATCCTCTTTCAAGCGATGTCTCC3' (mismatched bases are underlined).

EGFP fusion constructs

(Runx3-MASNS):EGFP and (Runx3-MHIPV):EGFP constructs were generated by PCR amplification using as template 24 hpf cDNA, prepared as described in section 2.3.4, in a primary reaction with sense primers (DrRunx3II&III_F1 and DrRunx3I_F1, respectively) and the same antisense primer (DrRunx3_R4). Then, a nested PCR was performed, using the amplified primary PCR as template, and the sense primers (Runx3II&III_F1_XhoI and Runx3I_F1_XhoI) and the antisense primers (RunxII&III_R1_KpnI and Runx3I_R1_KpnI) for isoforms Runx3-MASN and Runx3-MHIPV, respectively. All primer sequences are described in **Table 2.S1** DNA fragments were digested with *Xho*I and *Kpn*I restriction enzymes and cloned into pEGFP-N1 vector (Clontech) previously digested with the same enzymes. Constructs were verified by double stranded DNA sequence analysis and prepared for injection using the GFX^{TM} Micro Plasmid Prep kit (GE Healthcare).

Zebrafish microinjection

Injected DNA or morpholinos were diluted to the desired concentration (100 ng/µl and 5 µg/µl, respectively) in 1X Danieau solution (Westerfield, 2000) with 0.025% phenol red to help visualization. Morpholinos solution was heated at 65ºC prior to use to reduce

secondary structure and to dissolve any precipitation. Very thin needles were created by melting 31/2'' Drummond glass capillaries (Drummond Scientific Co., Broomall, PA) in a Narishige, PC-10 64 needle puller (heater 1 at 63, all weights on) and 1-cell stage zebrafish embryos were injected using a Drummond Nanoject II injector apparatus (Drummond Scientific Co.). The embryos and larvae were examined and imaged using an Eclipse E800 (Nikon) microscope.

Immunohistochemistry in whole mount zebrafish embryos

Whole-mount immunohistochemistry was performed as previously described (Ungos et al, 2003). The mouse monoclonal antibody [16A11] to Hu protein was used (1-700, Abcam, Cat#ab14370) and for fluorescent detection was used the AlexaFluor 488 goat anti-mouse IgG (1:700, Molecular Probes, Cat# A31619). Embryos were kept and imaged in 70% glycerol, in an Eclipse E800 (Nikon) microscope.

Results

Morpholino test using a tagged protein construct

Based in the analysis of all *runx3* transcripts, four morpholino were designed. Two translation morpholinos – Mo-L morpholino, that blocks the translation of Runx3-P1 derived isoform (Longer-isoform), Mo-S morpholino, specific to block the translation of Runx3-P2 derived isoform (Shorter-Isoform); one splicing morpholino – Mo-Sp morpholino that blocks the splicing of intron 3; and one control morpholino – Mo-M morpholino with five bases mismatch compared to Mo-L (**Figure 2.A1A**).

To test the efficiency of both translation morpholinos, Mo-L and Mo-S, two tagged protein constructs were made for both Runx3 isoforms, as GFP fusion proteins in pEGFP-N1 vector, and then microinjected into zebrafish 1-cell stage embryos in the presence or absence of the testing morpholinos (**Figure 2.A1B**).

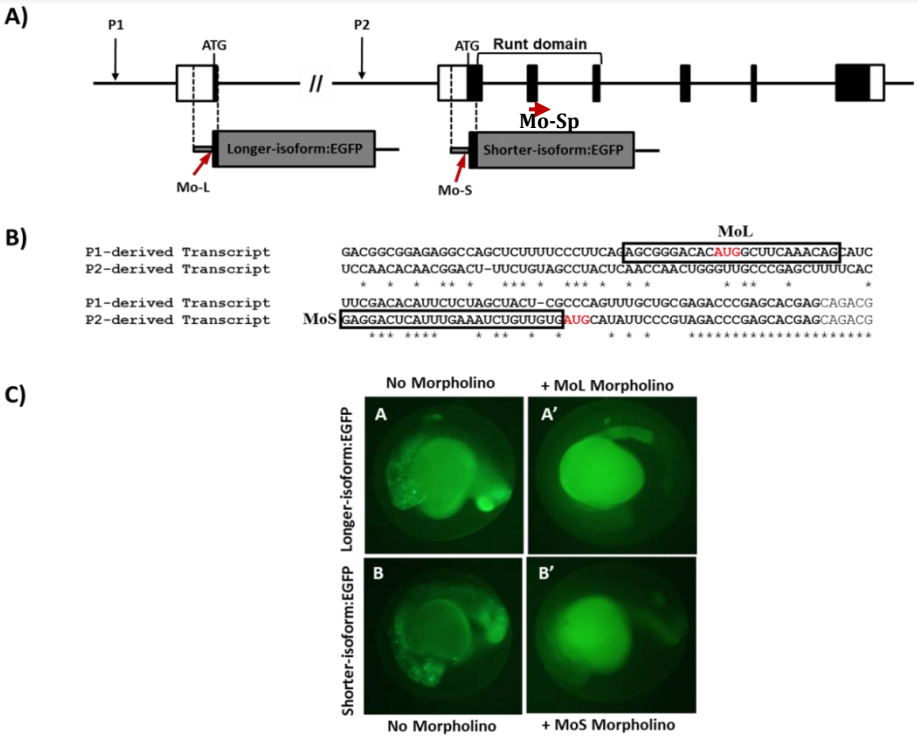


Figure 2.A1 Qualitative evaluation of the knockdown efficiency of translation Runx3 Morpholinos in zebrafish embryos using a Mo-target:GFP fusion assay. (**A**) Representation of the *runx3* gene structure with localization of the morpholinos target sequences and schematic representation of the runx3-GFP fusion constructs used in the assay. The *runx3* sequence placed upstream of the EGFP gene is indicated by the dotted lines. (**B**) Part of the nucleotide sequences of the zebrafish *runx3* region used in the fusion constructs. The Mo-L and Mo-S target sequences are indicated by the boxes and the start codons are also indicated in red. (**C**) Qualitative observation of the inhibition levels caused by the microinjection of Runx3-MOs using fluorescence microscopy.

We observed that the embryos injected with the tagged cDNA:GFP construct (100 pg/embryo) in the absence of morpholinos have high expression of GFP driven by the CMV promoter, in contract to the embryos co-injected with the tagged cDNA:GFP construct and the testing morpholinos (11.5 ng/embryo), that showed an almost complete blocking of the GFP expression (**Figure 2.A1C**). These results, confirm the inhibitory effect of the Runx3 morpholinos on the translation of the corresponding mRNA, whereas the control morpholinos with five nucleotides substitution (MoM) did not have a significant affect (data not shown).

Morpholino-mediated knockdown of specific zebrafish *runx3* isoforms to assess the function

We used three different *runx3* morpholinos to protein knockdown studies. Zebrafish injected with *runx3* testing morpholinos were compared with zebrafish at the same embryonic stage injected with the Danieau's solution and the negative control Mo-M, and the effects of both morpholinos on DRG development were observed at 5 dpf. As the morpholinos can have different efficiencies, a range of dilutions between ≈2 to 23 ng/embryo were tested to optimize the best concentration of each morpholino to use.

After injection, the embryos were raise as described previously and the embryos that were apparently normal were selected and fixed at 5 dpf for further analysis of the morpholinos effect in sensory neurons, analysed by immunohistochemistry using the neural marker Anti-Hu (**Figure 2.A2**).

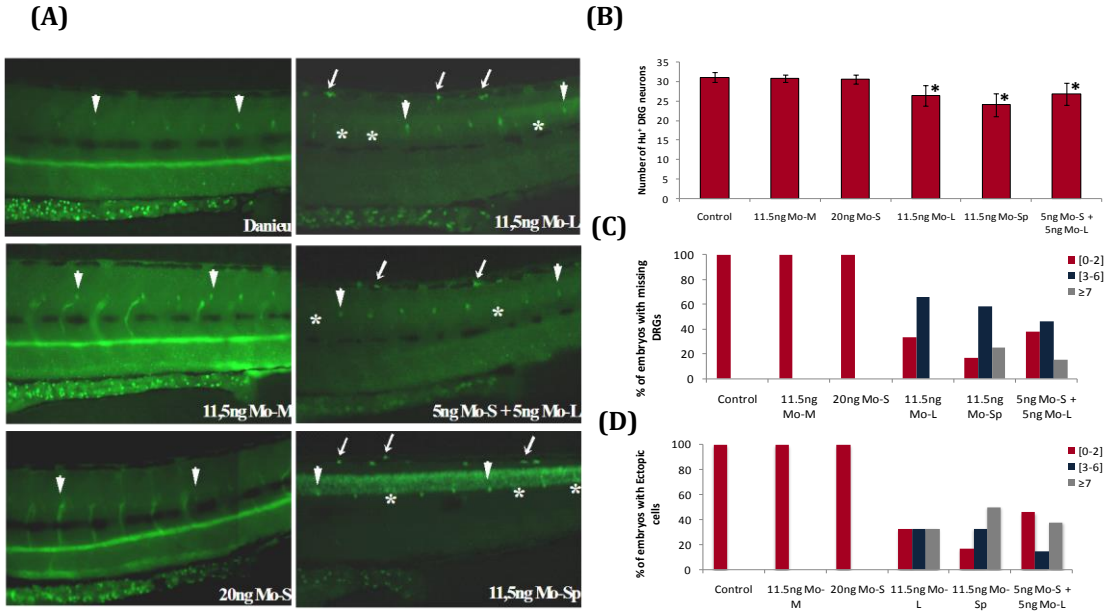


Figure 2.A2 Knockdown of *runx3* Longer-isoform causes a decrease in DGR number. (A) Lateral views of part of the trunk of 5 dpf embryos, showing staining for neurons with Hu specific antibody. (B, C, D) Quantitative measurement of Hu^+ DRG sensory neurons and ectopic Hu+ cells in control and morphant zebrafish embryos (* $P<0.05$, Student's t-test). Arrowheads represent DRG sensory neurons; asterisks represent missing DRG sensory neurons; arrows represent ectopic neurons. Mo-S – Translation morpholino specific to block Runx3 Shorter-isoform (Runx3-P2); Mo-L – Translation morpholino specific to block Longer-isoform (Runx3-P1); Mo-Sp – Splicing morpholino; Mo-M – Mismatch morpholino (5-base mismatch).

Our preliminary results show that knockdown of Runx3-P2 isoform, using a morpholino specific to this isoform (Mo-S), does not affect the total number of DRG sensory neurons (**Figure 2.A2A-C**), obtaining a total number of DRG sensory neurons similar to the number observed in control embryos (Danieu's and Mo-M; **Figure 2.A2A-C**). On the other hand, knockdown of Runx3-P1 isoform, using the specific morpholino Mo-L, resulted in a significant decrease in total number of DRG sensory neurons (**Figure 2.A2A-C**) and in the appearance of extra ectopic neuronal cells in a position directly dorsal to the spinal cord (**Figure 2.A2A, D**). The same phenotype was observed when a splicing morpholino (Mo-Sp) was used (**Figure 2.A2A-D**).

Chapter 3

Identification of *cis*-regulatory elements in the upstream regions of zebrafish runx3 gene through an *in silico* analysis: implications for function

This chapter is based on a published research paper:

Brigite Simões, Natércia Conceição, Robert Kelsh and M Leonor Cancela
Journal of Applied Ichthyology, 30 (2014), 661–670

Author's contribution:

Most of the experimental work and writing of the paper was performed by B Simoes, but N Conceição helped significantly with the analysis of the results and writing the paper. The research concept and design was performed by B Simões and N Conceição. N Conceição, ML Cancela and RN Kelsh were responsible for the critical revision and final approval of the manuscript.

3.1 Abstract

RUNX3 encodes a member of the runt domain family of transcription factors. In mammals this family includes three genes (*RUNX1-3*) and their protein products function as context-dependent transcription factors, either transcriptional activators or repressors, during developmental processes such as hematopoiesis, neurogenesis, and osteogenesis; all are proto-oncogenes or tumour suppressors. All three genes were shown to be transcribed from two promoters, giving rise to protein products bearing either the P1 or the P2 N-termini, translated respectively from transcripts originating from the distal (P1)- or the proximal (P2)-promoters. Understanding their differential regulation and interaction may help explain how RUNX factors contribute to such different and often opposing biological processes. In this study we have identified putative molecular players affecting zebrafish *runx3* gene transcription by using a computational approach to search for cis-regulatory transcription factor binding sites (TFBSs) in the promoter regions of the *runx3* gene from zebrafish (*Danio rerio*) and fugu (*Takifugu rubripes*). From the data obtained it was possible to identify the sites most likely involved in regulating expression of *runx3* in zebrafish. Our comparative approach reduced substantially the number of putative TFBSs in the *runx3* promoter regions; reassuringly, published TFs identified as transcriptional regulators of *Runx3* are confirmed by our *in silico* analysis. Our data now provides the basis for focused *in vitro* and/or *in vivo* experimental tests of the transcriptional regulatory activities of strong candidate regulators of zebrafish *runx3*.

3.2 Introduction

RUNX3 encodes a member of the runt domain family of transcription factors which also include *RUNX1* and *RUNX2*. RUNX proteins can bind DNA as a monomer to the core sequence 5'-PyGPyGGT-3' found in a number of enhancers and promoters, but their affinity for DNA is enhanced when the RUNX protein forms a heterodimer with its non-DNA binding partner CBFβ (Ogawa et al, 1993; Bae et al, 1994). The RUNX proteins also interact with other transcription factors, thus modulating their activity. Despite the recognized importance of this family in gene transcription, little is known about the factors regulating transcription of the *RUNX3* gene itself. Like the other two *RUNX* genes, *RUNX3* was shown to be transcribed from two promoters (Ghozi et al, 1996; Xiao et al, 1998; Rini and Calabi, 2001), giving rise to Runx3 protein products bearing either the P1 or the P2 N-termini, resulting from transcripts derived from the distal (P1) or the proximal (P2) promoters respectively (Bangsow et al, 2001; Rini and Calabi, 2001). The identification of several RUNX binding sites in the *RUNX* promoter regions (Ghozi et al, 1996; Levanon et al, 2001; Bangsow et al, 2001) led to the demonstration that auto- and cross-regulation of *RUNX* expression by RUNX proteins was likely to contribute to their regulation (Drissi et al, 2000; Spender et al, 2005). However, little work has addressed the regulatory processes that determine when RUNX proteins bind to the promoters of the genes of the other two family members to inhibit their expression in a kind of intrafamilial competition, nor when each RUNX protein acts mainly on its own promoter either promoting or inhibiting its own transcription, for example, to stabilise its levels of expression. Spender et al (2005) have shown that in human B lymphoid cell lines, RUNX3 represses *RUNX1* expression, thus contributing to their mutually exclusive expression in those cells. In this case, RUNX3 represses the *RUNX1* P1 promoter by binding specifically to the conserved RUNX sites located near the transcription start site of that promoter, thus confirming that cross-regulation between different RUNX family members is a means of controlling *RUNX* expression (Spender et al, 2005). The demonstration that RUNX transcription factors can be regulated by other members of the RUNX family may help explain their diverse functions and has important implications for the interpretation of pathologies associated with *RUNX* gene knockout or amplification. RUNX family proteins can function as context-dependent transcription factors during diverse developmental processes such as hematopoiesis (de Bruijn and Speck, 2004), neurogenesis

(Li et al, 2002; Fainaru et al, 2004), and osteogenesis (Karsenty, 2000; Komori, 2005). RUNX2 and RUNX3 have also been shown to regulate chondrocyte differentiation and maturation (Yoshida et al, 2004). In zebrafish, was shown that loss of function of *runx3* leads to severe reduction of head cartilage at 4 dpf (Flores et al, 2006; Dalcq et al, 2012). Furthermore, it was shown that a regulatory cascade formed by Runx3-Egr1-Sox9b controls late chondrogenesis by reducing expression of Follistatin A, a BMP inhibitor (Dalcq et al, 2012). This down-regulation allows the correct activation of BMP signalling required for expression of *runx2b* in developing chondrocytes (Dalcq et al, 2012). These observations were further investigated by Larbuisson et al (2013) using loss of function studies observed cartilage defects in Fgfr1a or Fgfr2 morphants that could be rescued by expression of exogenous Runx3 or Egr1. Recently, it was also shown using RNA-sequencing of Atlantic salmon notochord during segmentation that *runx3* was one of the genes expressed during and implicated in tissue mineralisation, alongside other genes such as the chondroblast-specific *sox6*, *sox5* and *sox9* (Wang et al, 2014).

Recently, we have cloned the full-length cDNA sequences of the zebrafish *runx3*, observed the tissue distribution pattern and analyzed their bioinformatic features. With the aim to characterize the genomic structure and to analyze the promoter activities, we have cloned the 5'-flanking regions of the *runx3* (P1 and P2) (Chapter 4). In this study we have analysed the promoter activities of the 5'-flanking regions of the zebrafish *runx3* (P1 and P2) and identified putative transcriptional regulators of zebrafish *runx3* by using a computational approach to search for *cis*-regulatory transcription factor binding sites (TFBSs) in the promoter regions (P1 and P2) of the *runx3* gene from zebrafish (*Danio rerio*) and fugu (*Takifugu rubripes*). Our *in silico* strategy provides a quick way to identify the most promising candidates among the large number of TFs that might potentially regulate zebrafish *runx3 in vivo*. Testing the functionality of these sites *in vitro* and *in vivo* will then be the priority for future studies.

3.3 Materials and Methods

3.3.1 Cell culture, transient transfection and luciferase assay

C6 cells (rat glioma cell line) were maintained in F-12K Nutrient Mixture Medium supplemented with 2.5% fetal bovine serum, 15% horse serum and 1% penicillin/streptomycin. U2OS cells (human osteosarcoma cell line) were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum, 2mM L-Glutamine and 1% penicillin/streptomycin. For both cell lines, incubation was carried out at 37 °C in a humidified atmosphere containing 5% CO2. The media, FBS, antibiotics and glutamine were obtained from Invitrogen. For all transfections, the C6 and the U2OS cells were seeded into 24-well plates at the density of 5 x 10^4 cells/well and 3 × 10^4 cells/well, respectively. Following 16 h of incubation, when the cells were about 70-80% confluent, transient transfection of the plasmids was carried out using Lipofectamine® LTX with Plus™ Reagent (Invitrogen) for C6 cells and X-tremeGENE HP DNA Transfection Reagent (Roche) for U2OS cells. To normalize the transfection efficiency, the pRL-null vector (Promega) encoding Renilla luciferase was co-transfected at the ratio of 1:10 relative to pGL3-basic vector. After incubation for 48 h, cells were harvested. The luciferase activities were measured by the Dual-luciferase Reporter Assay System (Promega). The data were normalized by calculating the ratio of the specific activity of firefly luciferase to that of Renilla luciferase.

3.3.2 Sequence collection

Sequence databases at GenBank (www.ncbi.nlm.nih.gov) and Ensembl (release v72; www.ensembl.org) were searched for annotated Runx3 sequences derived from zebrafish and fugu. The promoter sequences from these two genes were extracted for analysis by selecting 5000 base pairs (bp) upstream of the known translation start sites giving rise to both isoforms, P1 and P2. This length of sequence provided a reasonable assurance of containing the target gene's TFBSs. The promoter sequences were masked for repetitive elements by the program RepeatMasker (www.repeatmasker.org) with the default mode.

3.3.3 Comparative promoter TFBSs analysis

For each zebrafish-fugu orthologous promoter pair, the DNA Block Aligner (DBA) software (www.ebi.ac.uk/Tools/psa/promoterwise/) was used to extract blocks of aligned sequence using the default parameter settings based on the postulation that conserved regulatory blocks may be regions important for regulation of the gene. DBA alignments between orthologous promoters vary substantially, in many cases having no significant alignment, but others having several large sections of aligning sequences. This also clearly shows that many genes have multiple aligning blocks, sometimes spaced quite widely apart in the 5 kb region. With this software it is possible to obtain four different types of conserved blocks of a certain degree of similarity: type A, 60-70%; type B, 70-80%; type C, 80%-90%; and type D, 90-100%.

The promoter sequences (P1 and P2) of the zebrafish *runx3* genes were then assessed for TFBSs by running MatInspector (http://www.genomatix.de/) against TF binding site position weight matrices (PWM). For this study we used the default settings for the core similarity 0.75 and for matrix similarity 0.80.

For the multiple alignment plus prediction of TFBSs in the set of identified conserved blocks present within the promoters, we used the DiAlignTF software (http://www.genomatix.de/). We retained the same settings as used for the TFBSs prediction in the promoters (score similarity 0.75 and matrix similarity 0.8) and all the common TFBS matches located in aligned regions were determined. Then we used MatInspector for quantification of TFBSs common to all input sequences. The percentage of retention of putative TFBSs was calculated comparing the number of a given TFBS in the promoter to that from the conserved blocks in both zebrafish and fugu promoters.

3.4 Results

3.4.1 Promoter activity analysis and prediction of transcription factor binding sites in zebrafish *runx3*

To investigate the promoter activity of the promoter regions (P1 and P2) of the zebrafish *runx3* gene, we performed luciferase assays following the transient transfection of the human osteosarcoma U2OS cells and rat C6 glioma cells (**Figure 3.1**). The U2OS cell line was shown to express low levels of RUNX3 (Lai and Mager, 2012) and it was previously reported that RUNX3 expression is significantly decreased in human glioma (Mei et al., 2011). We have used P1 (from -5094 to -17 of the translation start site starting as MASN) and P2 (from -3930 to -474 of the translation start site starting as MHIPV) reporter constructs generated by inserting PCR fragments into the pGL3-basic vector (Chapter 5). Both the constructs exhibited higher luciferase activities than negative control of the pGL3-basic vector (**Figure 3.1**). The P2 promoter construct showed about 36-fold and 26-fold higher luciferase activities than pGL-3-basic in U2OS and C6 cells respectively, while the P1 promoter construct showed about 7-fold and 5-fold higher activity than the empty vector in U2OS and C6 cells respectively. Therefore, we expect that differences in the sequences of each promoter could affect their ability to function as a promoter.

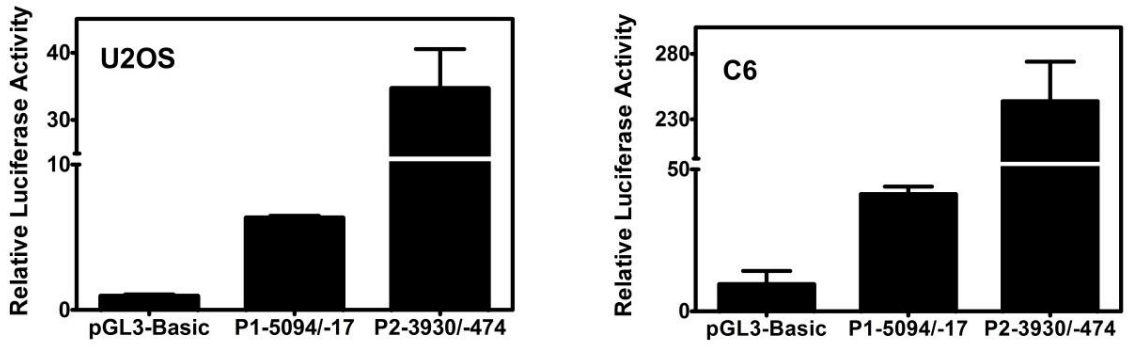


Figure 3.1 Relative transcriptional activity of zebrafish runx3 promoter constructs in U2OS and C6 cell lines. The results represent the ratio between firefly and Renilla luciferase determined. The mean and the SD for at least three independent transfections are shown.

In order to verify the relationship between sequence variation and promoter activity, we analyzed transcription factor binding sites (TFBSs) in the zebrafish *runx3* promoter regions by

using TRANSFAC® Public 6.0. We observed a huge number of putative TFBSs in both promoters, some of which were identified in both P1 and P2, while others were specifically found either in P1 or P2 promoter regions. To identify from this list of TFBSs which ones are more likely to be involved in the regulation of each promoter we decided to perform a comparative promoter analysis.

3.4.2 Zebrafish and fugu *runx3* promoter regions comparison

To identify the likely regulatory regions, we analyzed the conservation of the promoter regions of the zebrafish and fugu *runx3* genes using the DBA (DNA Block Aligner) web server. The output of DBA not only identifies discrete conserved blocks but also classifies them into four levels of conservation (A-D category, A showing lowest (60-70%) and D showing highest (90-100%) conservation). Among the 5 kb promoter sequences upstream of the translation start sites, an average of ten per cent of the total length of the P1 promoter region from zebrafish and fugu could be aligned by DBA; 0.7%, 3.2%, 1.9%, and 3.9% being of the A-D category, respectively. For the P2 promoter region only 3.3% was aligned by DBA; 0.5%, 0.9% and 1.9% being of the B-D categories, respectively. We plotted the sequence conservation as a function of the distance from the zebrafish *runx3* translation start sites (**Figure 3.2**). Sequence conservation in the first 1000 bp was distinctly higher (**Figure 3.2C**), a finding fully consistent with the typical pattern for protein coding genes (Conceição et al, 2009). Furthermore, in addition to being more abundant, conserved blocks in the first 1000 bp tended to be more conserved (i.e. Type D, with 90-100% sequence identity) in both promoters (**Figure 3.2**). Interestingly, we observed that the P1 promoter region shows more conserved blocks (23 blocks) than the P2 promoter (eight blocks) (**Table 3.1**), suggesting that the P1 promoter sequence is more conserved compared with P2 and that there are more conserved *cis*-acting regulatory elements in the P1 than in the P2 promoter region.

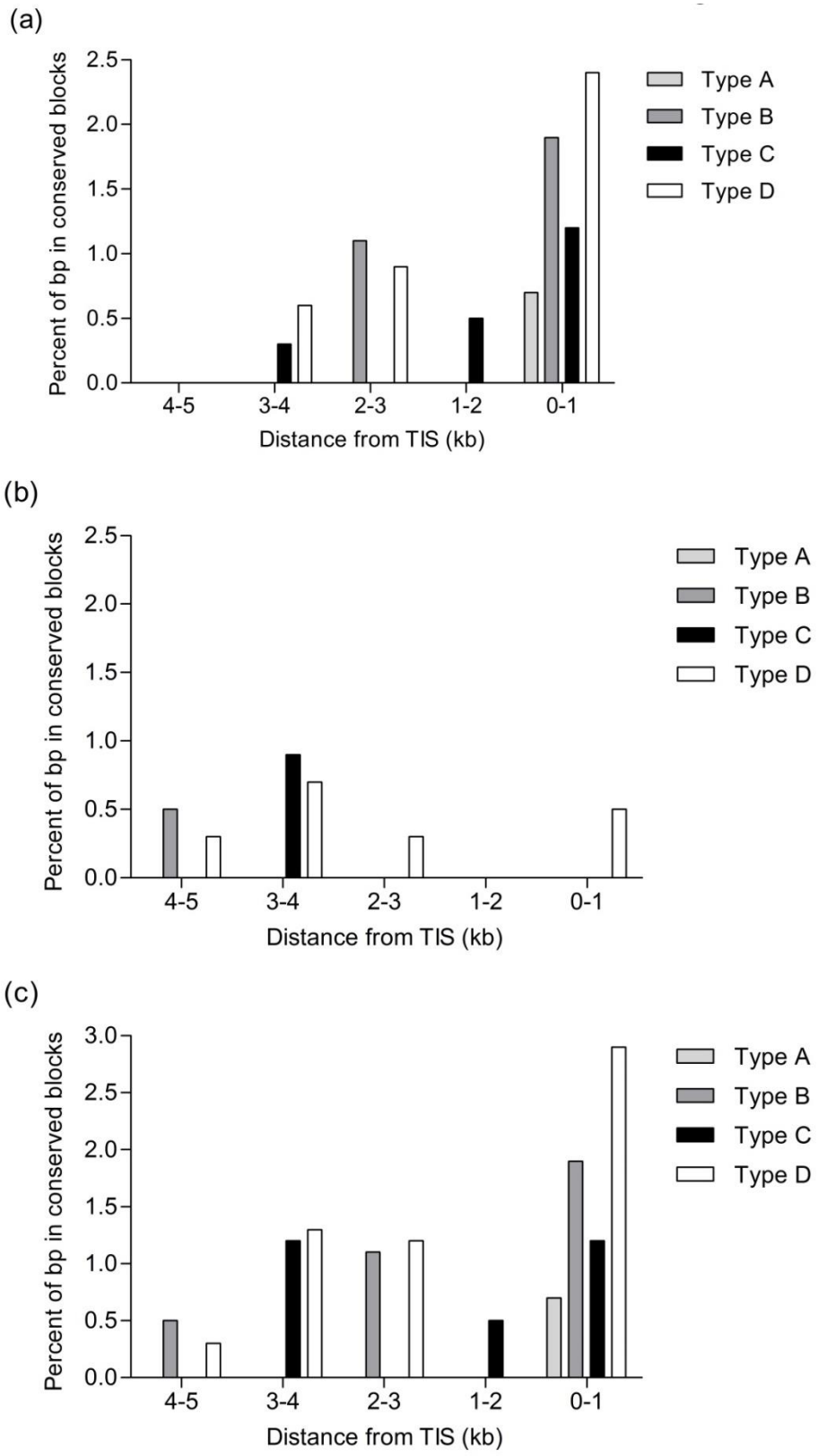


Figure 3.2 Distribution of percentage of base-pairs located in block A, B, C, or D located in zebrafish and fugu promoters, for each of five 1000 bp segments spanning up to -5000 bp upstream of translation start signal (TSS) in (a) P1, (b) P2 and (c) P1 and P2 runx3 promoter regions.

3.4.3 Analysis of regulatory elements using MatInspector

To analyze the conservation of the *cis* elements between the zebrafish and fugu 5 kb *runx3* promoter regions, the zebrafish promoter sequences and the set of identified conserved blocks presented within the zebrafish and fugu promoters were compared for TFBSs. Our comparison reveals the conservation of multiple potential *cis* elements between fugu and zebrafish promoters (**Figure 3.S1**). Although TFBSs are abundant in all sequences assessed, relatively few show conservation between zebrafish and fugu. Thus, for zebrafish, this criterion reduces the 3478 TFBSs in the P1 promoter sequence to only 142 in the conserved blocks, and the 8544 TFBSs in the P2 promoter sequence to only 104 in the conserved blocks. Thus, by identifying and analysing only conserved blocks, an average of 96% and 99% of all TFBSs in, respectively, the P1 and P2 promoter regions were eliminated (**Figure 3.3**). Likewise, this approach also substantially reduces the number of TF families implicated in *runx3* regulation, (from 170 to 66 (a reduction of 61 %) and from 175 to 49 (a reduction of 72 %), in P1 and P2 promoters, respectively).

To determine whether our comparative screening of TFBSs was likely to have identified TFs (**Table 3.1**) with roles in *runx3* regulation, we searched our lists for the presence of TFs previously characterized as regulating *RUNX3*. Key roles have been shown for signal transducer and activator of transcription STAT (Park et al, 2010), Sp1 transcription factor /TEA domain family member 2 (Sp1/ETF) (Bangsow et al, 2001), T helper transcription factor (Th-POK) (Egawa et al, 2009), CBF1/Suppressor of Hairless/Lag1 (CSL) (Fu et al, 2011), interferon regulatory factor 4 (IRF4) (Cao et al, 2010), E-twenty-six (Ets1) (Zamisch et al, 2009), cAMP-response element-binding protein (Creb) (Lim et al, 2011), microphthalmia-associated transcription factor (Mitf) (Hoek et al, 2008), Twist subfamily of class B bHLH transcription factors (Scl/Tal1) (Landry et al, 2008), Brn POU domain factor (Brn3a) (Dykes et al, 2010; 2011), recombination signal binding protein-J kappa (Rbpj) (Fu et al, 2011) and runt-related transcription factors (Runx) (Drissi et al, 2000) in regulation of mammalian *RUNX3.* Of these eight TFs, Ets1, Stat, Creb, Mitf, Scl, Brn3a, Rbpj and Runx show conserved binding sites in the conserved sequence blocks in the *runx3* promoter sequences analysed here, consistent with them having a functional role in the regulation of *runx3* in fish.

These TFs are then prime candidates for future functional studies assessing their ability to bind to and regulate activity of *runx3 in vivo*.

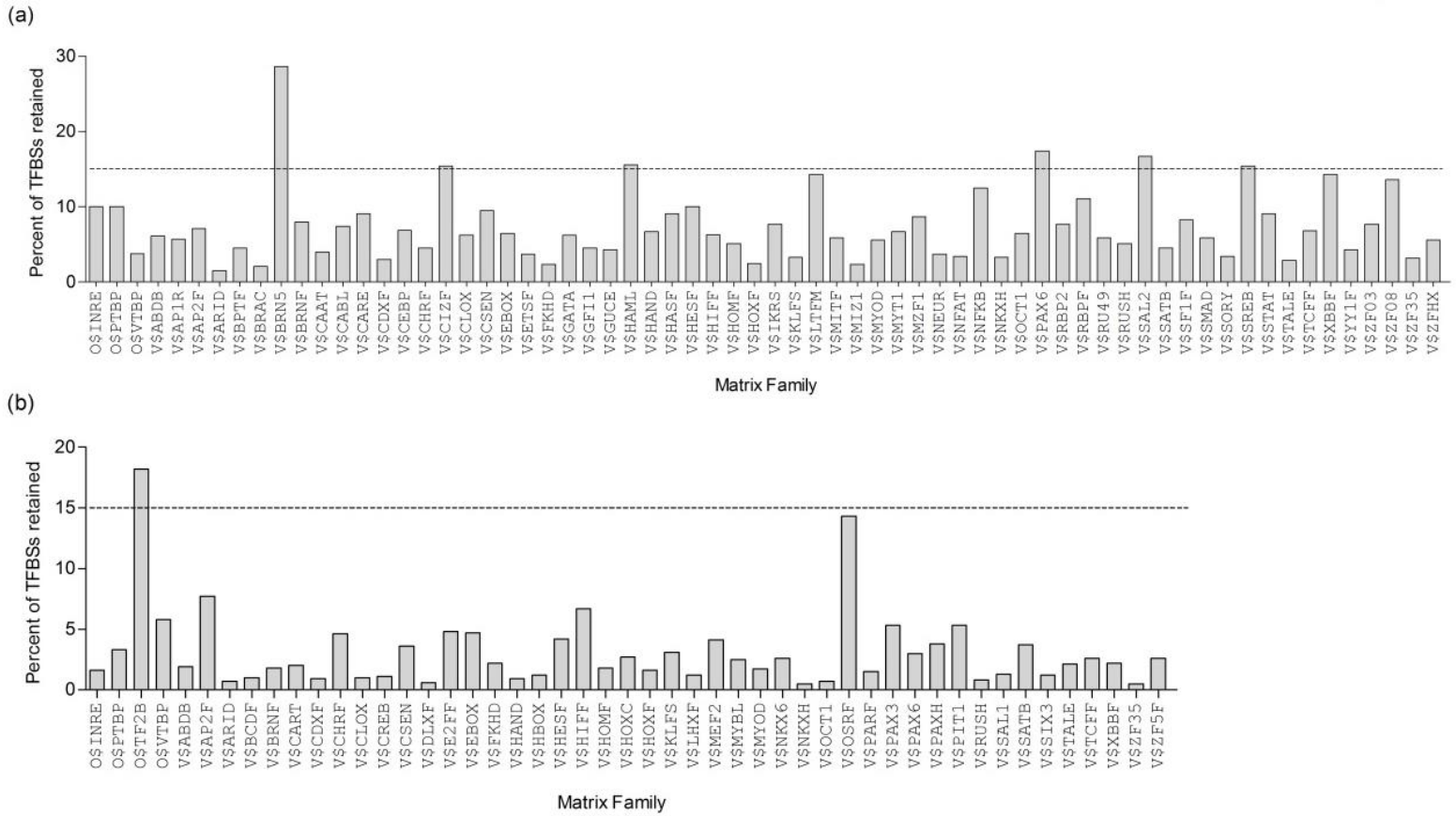


Figure 3.3 Retention of putative TFBSs after comparative analysis. For each TF listed along the x-axis, corresponding bars represent the percentage (y-axis) of putative TFBSs originally identified by MatInspector that also survived after DiAlignTF comparative analysis. The dashed line indicates the minimum percent chosen to consider TFBSs as most frequent.

Table 3.1 Transcription factor families conserved in each block obtained from the alignment of P1 and P2 *runx3* promoter regions between zebrafish and fugu.

Pr	Block Number	Block Position	Block Type				Conserved TF Families
			A	B	C	D	
P1	1	**Dr**: -3636 to -3620 **Fr**: -3044 to -3060 (-)			X		V$OCT1; O$PTBP
	2	**Dr**: -3557 to -3545 **Fr**: -1271 to -1259				X	V$ZF35; V$RUSH
	3	**Dr**: -3512 to -3494 **Fr**: -4874 to -4894 (-)				X	V$RU49; V$ZF03; V$HASF
	4	**Dr**: -3202 to -3194 **Fr**: -3149 to -3157 (-)				X	No common TF matches found
	5	**Dr**: -3084 to -3071 **Fr**: -3776 to -3789 (-)				X	V$RUSH; V$SMAD
	6	**Dr**: -3042 to -3032 **Fr**: -1084 to -1094 (-)				X	V$TCFF
	7	**Dr**: -2979 to -2967 **Fr**: -4396 to -4384				X	No common TF matches found
	8	**Dr**: -2957 to -2895 **Fr**: -1688 to -1626		X			V$CSEN; O$INRE; V$PAX6; V$AP1R; V$HAND; V$MYOD; V$SORY; V$NKXH; V$BRN5; VHOXF; V$YY1F; V$BRNF; V$ABDB; V$SF1F; V$LTFM; V$RBP2
	9	**Dr**: -1532 to -1503 **Fr**: -1274 to -1247			X		V$FKHD; V$NFAT; V$ABDB; V$BRNF; V$CDXF; V$ZF35; V$ARID; O$VTBP; V$SATB
	10	**Dr**: -1189 to -1171 **Fr**: -3275 to -3293 (-)			X		V$HAML; V$NFKB; V$ZFHX; V$ZF08; V$SAL2
	11	**Dr**: -1110 to -1100 **Fr**: -1494 to -1504 (-)				X	No common TF matches found
	12	**Dr**: -1073 to -1036 **Fr**: -4416 to -4379	X				V$CEBP; V$OCT1; V$PAX6; V$BRAC; V$HASF; V$RUSH
	13	**Dr**: -1021 to -1006 **Fr**: -3085 to -3100 (-)			X		V$CIZF; V$CHRF
	14	**Dr**: -852 to -803 **Fr**: -1037 to -1086 (-)		X			V$BRN5; V$SORY; V$BRNF; V$CLOX; V$GATA; V$HAND; V$HOMF; V$EBOX; V$HIFF; V$MYOD; V$HESF; V$MITF; V$NKXH; V$SREB; V$CAAT; V$NEUR; V$KLFS; O$INRE; V$CARE; V$SAL2; V$MIZ1
	15	**Dr**: -778 to -763 **Fr**: -499 to -516 (-)				X	V$HAML; V$ZF35; V$BPTF; V$ZFHX; V$CABL
	16	**Dr**: -701 to -684 **Fr**: -1249 to -1267 (-)				X	V$RUSH; V$CIZF; V$CABL
	17	**Dr**: -625 to -604 **Fr**: -3027 to -3006			X		V$STAT; V$XBBF; V$CEBP; V$RBPF; V$IKRS; V$MZF1; V$MYT1; V$ZF03; V$ZFHX
	18	**Dr**: -510 to -484 **Fr**: -544 to -518				X	V$TCFF
	19	**Dr**: -472 to -463 **Fr**: -3549 to -3540				X	No common TF matches found
	20	**Dr**: -454 to -443 **Fr**: -1597 to -1608 (-)				X	V$SMAD
	21	**Dr**: -202 to -188 **Fr**: -3392 to -3378				X	V$MZF1
	22	**Dr**: -162 to -98 **Fr**: -179 to -118		X			V$HASF; V$HAML; V$ZF08; V$GFI1; V$CLOX; V$ETSF; V$GATA; V$GUCE; V$EBOX; V$IKRS; V$KLFS; V$AP2F; V$HOMF; V$BRNF; V$SORY; V$STAT; V$PAX6; V$XBBF
	23	**Dr**: -80 to -61 **Fr**: -5053 to -5074 (-)				X	V$TALE; V$CSEN; V$TCFF

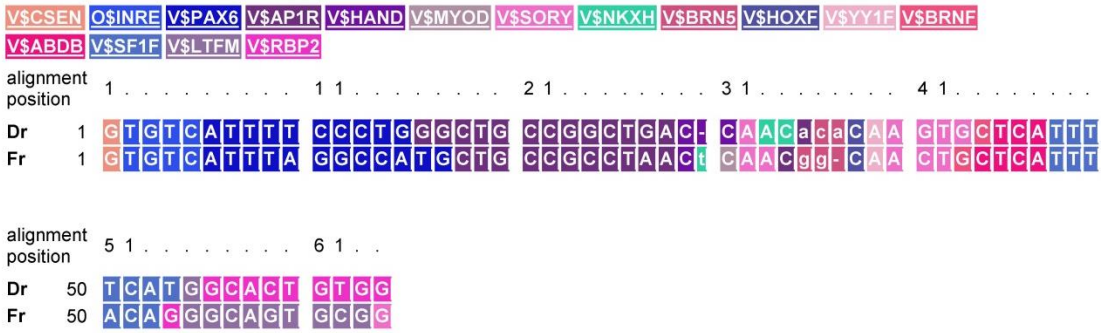
Pr	Block Number	Block Position	Block Type				Conserved TF Families
			A	B	C	D	
P2	1	**Dr**: -5505 to -5479 **Fr**: -294 to -320 (-)		X			O$VTBP; V$CDXF; V$PARF; V$FKHD; V$BRNF; V$HOMF; V$PAXH; V$RUSH
	2	**Dr**: -4930 to -4913 **Fr**: -2187 to -2205 (-)				X	V$RUSH; O$VTBP; V$PAXH; V$CHRF
	3	**Dr**: -4466 to -4446 **Fr**: -3630 to -3650 (-)			X		V$ARID
	4	**Dr**: -3926 to -3908 **Fr**: -2808 to -2790				X	V$MYBL; V$XBBF; V$OSRF
	5	**Dr**: -3900 to -3873 **Fr**: -2773 to -2745			X		V$CSEN; V$KLFS; V$TALE; V$AP2F; V$HAND; V$MYOD; V$EBOX; V$HESF; V$E2FF; V$ZF5F; O$TF2B
	6	**Dr**: -3862 to -3838 **Fr**: -2713 to -2689				X	V$CART; V$CREB; V$EBOX; V$HIFF; V$LHXF; V$ARID; V$NKXH; V$PAX3; V$SIX3; V$HOBX; V$HOMF; V$HOXF; V$PAX6; V$SATB; V$BCDF; V$BRNF; O$VTBP; V$DLXF; V$PIT1; V$NKX6; V$PAXH; V$HOXC; V$ABDB; V$OCT1; V$ZF35; O$INRE; V$TCFF
	7	**Dr**: -2982 to -2964 **Fr**: -2517 to -2535 (-)				X	V$BRNF; V$FKHD; O$VTBP; V$SATB; V$SAL1
	8	**Dr**: -974 to -950 **Fr**: -2322 to -2347 (-)				X	V$CART; V$HOMF; V$MEF2; V$ABDB; V$LHXF; O$PTBP; V$BRNF; V$CHRF; V$CLOX; V$FKHD; V$HOXC; O$VTBP; V$STAB

>P1_Block_8

(a)

```
Dr:     -2957    GTGTCATTTTCCCTGGGCTGCCGGCTGAC-CAACACACAAGTGCTCATT    -2910
           B     GTGTCATTT   C   GCTGCCG CT AC CAAC   CAA TGCTCATT
Fr:     -1688    GTGTCATTTAGGCCATGCTGCCGCCTAACTCAACGG-CAACTGCTCATT    -1641


Dr:     -2909    TTCATGGCACTGTGG    -2895
           B     T CA GGCA TG GG
Fr:     -1640    TACAGGGCAGTGCGG    -1626
```

(b)



>P2_Block_5

(a')

```
Dr:     -3900    GTCTGAATTTGACAGGCGCGCGGAG-TCA    -3873
           C     GTCT AAT TGACAGGCGC C GAG TCA
Fr:     -2773    GTCTAAATGTGACAGGCGCCCAGAGCTCA    -2745
```

(b')

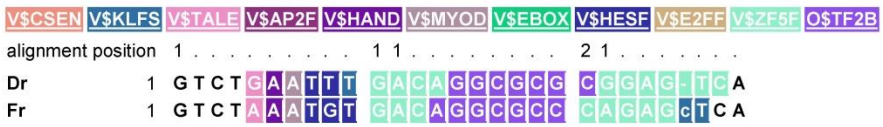


Figure 3.4 Representation of a DBA block obtained from the alignment of P1 (a) and P2 (a') runx3 promoter regions between zebrafish and fugu and overview of TFBS families detected by DiAlignTF on the conserved blocks analysed for P1 (b) and P2 (b') promoters. (a, a') Example of an alignment of one of the 23 blocks obtained for P1 and 8 blocks obtained for P2 using the DBA software. The block position in the respective promoter sequence is shown, considering the A of the translation initiation codon as +1. The block type is also represented as a bold letter next to the consensus sequence identified between the two blocks. (b, b') Overview of the TFBS conserved in the block showed as (a) or (a'), respectively for P1 or P2, detected by MatInspector using DiAlignTF program. Only upper-case letters are considered to be aligned. The colour code for each specific TFBS is shown above the alignment.

3.5 Discussion

In the present study we provide evidence for the transcriptional activity of zebrafish *runx3* promoters in two different cell lines, using in vitro transient transfection experiments. These findings support an earlier report from our laboratory. To gain insight into the regulatory mechanism of the *runx3* gene, the sequences of the genomic fragments (named P1 and P2) were analyzed *in silico* for potential recognition sites to transcription factors. Our analysis identified numerous putative cis regulatory elements that may serve as targets for sequence-specific enhancer/silencer transcription factors.

We have then used the DBA algorithm to obtain comparative alignment between zebrafish and fugu *runx3* promoter regions in order to detect conserved sequence blocks and then used MatInspector to determine putative TFBSs in those blocks, so as to enrich for likely functionally relevant TFBSs. Our *in silico* analysis of zebrafish *runx3* P1 and P2 promoter regions provides important clues as to factors likely to be involved in regulation of *runx3* expression. Although MatInspector can find most true positive TFBS matches in a promoter region (Cartharius et al, 2005), not all sites found are necessarily functional in a particular biological context. A first step in examining functionality is a comparative promoter analysis. The alignment obtained with DBA was then assessed for conserved TFBSs by DiAlignTF, a combination of MatInspector with the multiple alignment program DiAlign (Morgenstern et al, 1998). DiAlignTF displays TFBSs located at the same position within the alignment and then it can be used to reduce the list of potential TFBSs to the most likely functional matches. From the data obtained it was possible to identify the sites most likely involved in regulating expression of *runx3* in zebrafish. While a number of pathways regulating RUNX activity have been delineated, transcription factors binding to *RUNX* promoters are only beginning to be identified. From the list of 86 putative TFBSs families retained after the comparative analysis, CREB (family V$CREB) (Lim et al, 2011), Mitf (family V$MITF) (Hoek et al, 2008), Brn3a (family V$BRNF) (Dykes et al, 2010; 2011), Rbp-j (family V$RBPF) (Fu et al, 2011), Scl/Tal1 (family V$HAND) (Landry et al, 2008), Ets1 (family V$ETSF) (Zamisch et al, 2009), Stat (family V$STAT) (Park et al, 2010) and Runx (family V$HAML) (Drissi et al, 2000; Spender et al, 2005) are described in the literature as regulating the *RUNX3* gene. In addition to these data indicating likely conservation of a regulatory function for these TFs between

mammals and fish, our *in silico* analysis identified a number of novel potential regulatory TFs for the zebrafish *runx3* promoters (**Figure 3.3**). As described in Materials and Methods, several TFs were found to be retained within the conserved blocks (between zebrafish and fugu promoters) analysed (**Figure 3.4**). By focusing only on those, we selected a set having a retained score of 15 % or higher (**Figure 3.3**), which included Brn-5 POU domain factors (family V$BRN5), Cas interacting zinc finger (family V$CIZF), Runx or Human acute myelogenous leukemia factors (family V$HAML), PAX-4/PAX-6 paired domain binding sites (family V$PAX6), Spalt-like transcription factor 2 (family V$SAL2) and Sterol regulatory element binding proteins (family V$SREB) for P1 promoter and TFIIB or RNA polymerase II transcription factor II B (family O$TF2B) for P2 promoter. Interestingly, available data links some of these with either skeletal or neuronal development. In the context of the P1 promoter, five of these TFs have a function in neurogenesis and two in skeletogenesis. Brn-5 POU domain factors (Brn-5) is expressed in many central nervous system (CNS) neuron populations and may function as a transcriptional regulator involved in specifying the mature phenotype of CNS neurons (Cui and Bulleit, 1998). Spalt-like transcription factor 2 (Sall2) also plays a role in neuronal development (Pincheira et al, 2009) and is the only member of the family suggested to act as a tumor suppressor (Li et al, 2001; Ma et al, 2001). Sterol regulatory element binding protein 2 (Srebp2) was shown to interact with the *acetoacetyl-CoA synthetase* (*AACS*) promoter and knockdown experiments showed that SREBP-2 regulates *AACS* expression during neurite outgrowth in the neuroblastoma Neuro-2a cell line (Hasegawa et al, 2012). Paired box 6 (Pax6) also is expressed during neurogenesis (Gan et al, 2013), and it functions as a transcription factor with a major role in eye and brain development from Drosophila to humans (Callaerts et al, 1999; van Heyningen and Williamson, 2002). Lleras-Forero et al (2013) showed CNS *Pax6b* expression in zebrafish. The human acute myelogenous leukemia factors (Haml; also known as a runt-related Runx/AML protein) function as context-dependent transcription factors during developmental processes such as hematopoiesis, neurogenesis, and osteogenesis (Westendorf and Hiebert, 1999). Runx expression was shown in subtypes of dorsal root ganglion (DRG) neurons, suggesting their involvement in lamina-specific afferent differentiation and maturation (Inoue et al, 2003) and Runx2 and Runx3 have also been shown to regulate chondrocyte differentiation and maturation (Yoshida et al, 2004; Komori, 2005). In zebrafish, we have shown that Runx2 was able to transactivate the promoter of *osteocalcin*, an osteoblastic

marker gene (Pinto et al, 2005), as well as the promoter of *collagen Xα1*, a chondrocyte marker gene (Simões et al, 2006). Zebrafish *runx3* expression was observed in neuronal tissues including the trigeminal ganglia and Rohon-Beard neurons (Kalev-Zylinska et al, 2003) and also in the craniofacial region (Flores et al, 2006). Cas-interacting zinc finger protein (CIZ) is one of the suppressors of BMP signaling in osteoblastic differentiation (Shen et al, 2002). Besides these last two highly conserved TFBSs, our data shows the occurrence in the conserved sequences of binding sites for many more TFs that are described as having a role in skeletogenesis. These factors include NF-kB (Wu et al, 2011), NF-YB (Chen et al, 2009), NFATc1 (Lambertini et al, 2008), Ets-1 (Wenke et al, 2006), and Sox5 and Sox9 (Yang et al, 2011). Curiously, all these TFBSs are only detected in the P1 promoter and not in the P2 promoter.

Of the two putative TFBSs more conserved in the P2 promoter, that for the odd-skipped related (Osr) zinc finger transcription factor is notable since it has suggested to be involved in bone formation (Kawai et al, 2007). The other relates to TFIIB, a component of the basal transcription complex. Several other TFs identified in our analysis for the P2 promoter are known to play critical roles in zebrafish development, e.g. *dlx* (distal-less homeodomain; family V$DLXF) genes play a key role in the patterning of the forebrain, in peripheral structures of the head, and in the fins (Akimenko et al, 1994); *mef2* (myocyte enhancer factor 2; family V$MEF2) genes are essential for heart development (Hinits et al, 2012) and in cranial neural crest for proper head skeletal patterning (Miller et al, 2007); CREB (cAMP response element-binding protein; family V$CREB) have a role in neural development (Dworkin et al, 2007); Nkx6 (NK6 homeobox; family V$NKX6) proteins specify one zebrafish primary motoneuron subtype (Hutchinson et al, 2007); Pax3 (paired box 3; family V$PAX3) is induced early during neural development in progenitors of the dorsal spinal cord (Moore et al, 2013), and Six3 (sine oculis homeobox homolog 3; family V$SIX3) are involved in the left-right brain patterning (Inbal et al, 2007).

Since it is known that all Runx protein family members bind to the same DNA core sequence, their temporal and/or spatial expression has to be tightly regulated. Other investigators have reported that the two promoter regions, P1 and P2, regulate *Runx3* expression in a cell type-specific manner (Bangsow et al, 2001; Egawa, 2009). Soung et al (2007) showed that

both Runx3 isoforms in mice are expressed and regulated during chondrocyte differentiation, while Yoshida et al (2004) showed that *Runx3* mRNA was detected in both mouse CD8+ and CD4+ T cells, but only the CD8+ population expressed the P1 transcript isoform and detectable levels of RUNX3 protein.

Comparison of the zebrafish *runx2b* proximal promoter sequence that we identified and cloned previously (Pinto et al, 2005) with both those of zebrafish *runx3* showed some common consensus binding motifs, namely for NFAT, CREB, Runx, and CBF1. They are thus possible candidates for regulating expression of *runx3*. Of interest are the two putative RUNX-binding sites present at the beginning of the 5'UTR of the P1 promoter of all three RUNX genes perfectly conserved in mammals. These RUNX-binding sites were previously been shown by independent studies to have an effect on the transcriptional regulation of Runx genes, either positively or negatively, through the binding of RUNX proteins (Levanon et al, 1998; Ducy et al, 1999; Drissi et al, 2000; Bangsow et al, 2001; Levanon and Groner, 2004; Spender et al, 2005). In our present analysis we also found two Runx-binding sites in the *runx3* P1 promoter that are conserved between zebrafish and fugu. Taken together this may indicate important regulatory roles such as cross-regulation and/or auto-regulation.

In conclusion, our comparative *in silico* analysis of zebrafish *runx3* gene promoter regions, using the DBA and DiAlignTF softwares, predicts strong candidates TFBS likely to contribute to regulation of *runx3* transcription. These TFBSs include binding sites for TFs already known from work in mammals as transcriptional regulators of *Runx3*, but also include novel TFs. Thus, our data likely provide a powerful tool to guide future dissection of *runx3* transcriptional regulation *in vitro* and/or *in vivo*.

Acknowledgments

NC and BS are supported, respectively, by post-doctoral and PhD grants from FCT (SFRH/BPD/48206/2008 and SFRH/BD/38083/2007). This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Program and national funds through FCT – Foundation for Science and Technology, under the project "PEst-C/MAR/LA0015/2011.

3.6 Supplementary Figures

(a) Promoter P1

>P1_Block_1

```
Dr:      -3636    AATCTAATTTATATCTG    -3620
               C  AATCTAAT  ATAT TG
Fr:      -3044    AATCTAATAAATATGTG    -3060
```

V$OCT1 O$PTBP

```
alignment position  1 . . . . . .  1 1 . . . .
Dr            1  A A T C T A A T T T  A T A T C T G
Fr            1  A A T C T A A T A A  A T A T G T G
```

>P1_Block_2

```
Dr:      -3557    CCAGAAAATAAAA    -3545
               D  CCAGAAAATAAAA
Fr:      -1271    CCAGAAAATAAAA    -1259
```

V$ZF35 V$RUSH

```
alignment position  1 . . . . . .  1 1 .
Dr            1  C C A G A A A A T A  A A A
Fr            1  C C A G A A A A T A  A A A
```

>P1_Block_3

```
Dr:      -3512    CATTTCAGTACACA--TTTCA    -3494
               D  CATT CAGTACACA  TTTCA
Fr:      -4874    CATTGCAGTACACACATTTCA    -4894
```

V$RU49 V$ZF03 V$HASF

```
alignment position  1 . . . . . .  1 1 . . . . . .  2 1
Dr            1  C A T T T C A G T A  - - C A C A T T T C  A
Fr            1  C A T T G C A G T A  c a C A C A T T T C  A
```

>P1_Block_4

```
Dr:      -3202    TATCCTGTT    -3194
               D  TATCCTGTT
Fr:      -3149    TATCCTGTT    -3157
```

No common TF matches found.

```
alignment position  1 . . . . . .
Dr            1  T A T C C T G T T
Fr            1  T A T C C T G T T
```

>P1_Block_5

```
Dr:      -3084    TGTGTGTATGTGTG    -3071
               D  TGTGTGT TGTGTG
Fr:      -3776    TGTGTGTCTGTGTG    -3789
```

V$RUSH V$SMAD

```
alignment position  1 . . . . . .  11 . .
Dr          1  T G T G T G T A T G  T G T G
Fr          1  T G T G T G T C T G  T G T G
```

>P1_Block_6

```
Dr:      -3042    GAAGTCATTCT    -3032
               D  GAAGTCATTCT
Fr:      -1084    GAAGTCATTCT    -1094
```

V$TCFF

```
alignment position  1 . . . . . .  11
Dr          1  G A A G T C A T T C  T
Fr          1  G A A G T C A T T C  T
```

>P1_Block_7

```
Dr:      -2979    CACACACACATAG    2967
               D  CACACACACATAG
Fr:      -4396    CACACACACATAG    -4384
```

No TF matches found.

```
alignment position  1 . . . . . .  11 .
Dr          1  C A C A C A C A C A  T A G
Fr          1  C A C A C A C A C A  T A G
```

>P1_Block_8

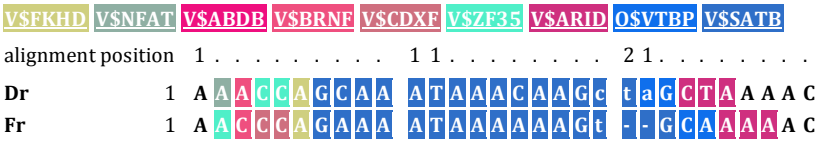
```
Dr:      -2957    GTGTCATTTTCCCTGGGCTGCCGGCTGAC-CAACACACAAGTGCTCATT    -2910
               B  GTGTCATTT   C   GCTGCCG CT AC CAAC   CAA TGCTCATT
Fr:      -1688    GTGTCATTTAGGCCATGCTGCCGCCTAACTCAACGG-CAACTGCTCATT    -1641

Dr:      -2909    TTCATGGCACTGTGG    -2895
               B  T CA GGCA TG GG
Fr:      -1640    TACAGGGCAGTGCGG    -1626
```

V$CSEN O$INRE V$PAX6 V$AP1R V$HAND V$MYOD V$SORY V$NKXH V$BRN5 V$HOXF V$YY1F V$BRNF V$ABDE V$SF1F V$LTFM V$RBP2

```
alignment position  1 . . . . . .  11 . . . . . .  21 . . . . . .  31 . . . . . .  41 . . . . . .
Dr          1  G T G T C A T T T T  C C C T G G G C T G  C C G G C T G A C -  C A A C a c a C A A  G T G C T C A T T T
Fr          1  G T G T C A T T T A  G G C C A T G C T G  C C G C C T A A C t  C A A C g g - C A A  C T G C T C A T T T
```

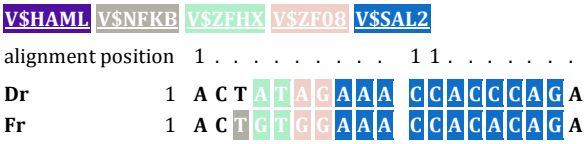
>P1_Block_9

```
Dr:      -1532    AAACCAGCAAATAAACAAGCTAGCTAAAAC    -1503
               C  AA CCAG AAATAAA AAG T GC AAAAC
Fr:      -1274    AACCCAGAAAATAAAAAAG-T-GCAAAAAC    -1247
```



>P1_Block_10

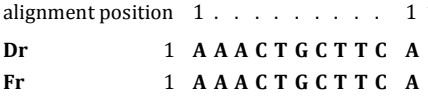
```
Dr:      -1189    ACTATAGAAACCACCCAGA    -1171
               C  ACT T GAAACCAC CAGA
Fr:      -3275    ACTGTGGAAACCACACAGA    -3293
```



>P1_Block_11

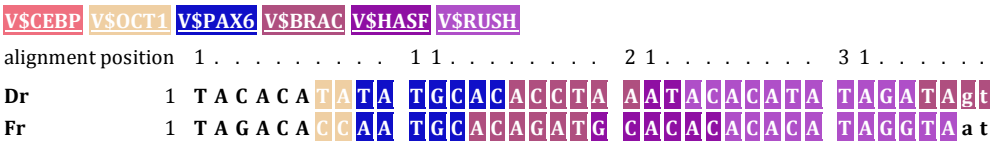
```
Dr:      -1110    AAACTGCTTCA    -1100
               D  AAACTGCTTCA
Fr:      -1494    AAACTGCTTCA    -1504
```

No common TF matches found.



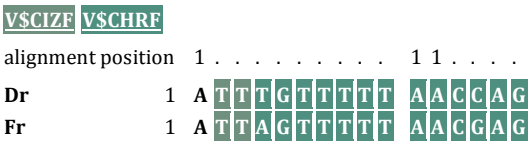
>P1_Block_12

```
Dr:      -1073    TACACATATATGCACACCTAAATACACATATAGATAGT    -1036
               A  TA ACA   ATGCACA  T  A ACACA ATAG TA T
Fr:      -4416    TAGACACCAATGCACAGATGCACACACACATAGGTAAT    -4379
```



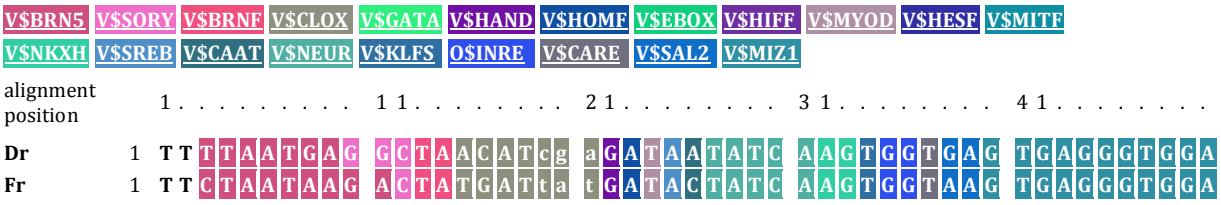
>P1_Block_13

```
Dr:      -1021    ATTTGTTTTTAACCAG    -1006
               C  ATT GTTTTTAAC AG
Fr:      -3085    ATTAGTTTTTAACGAG    -3100
```



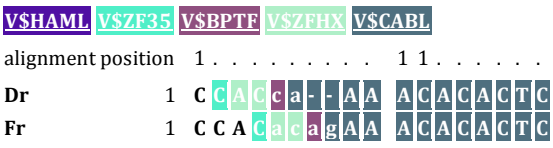
>P1_Block_14

```
Dr:      -852    TTTTAATGAGGCTAACATCGAGATAATATCAAGTGGTGAGTGAGGGTGGA    -803
            B    TT TAAT AG CTA  AT   GATA TATCAAGTGGT AGTGAGGGTGGA
Fr:      -1037   TTCTAATAAGACTATGATTATGATACTATCAAGTGGTAAGTGAGGGTGGA    -1086
```



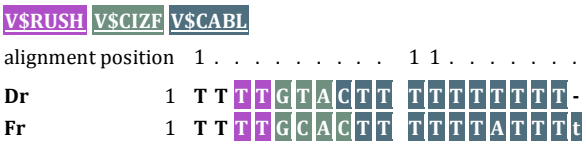
>P1_Block_15

```
Dr:      -778    CCAC-CA-AAACACACTC    -763
            D    CCAC CA AAACACACTC
Fr:      -499    CCACACAGAAACACACTC    -516
```



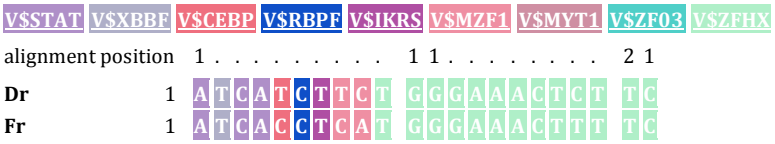
>P1_Block_16

```
Dr:      -701    TTTTGTACTTTTTT-TTTT    -684
            D    TTTTG ACTTTTTT TTTT
Fr:      -1249   TTTTGCACTTTTTTATTTT    -1267
```



>P1_Block_17

```
Dr:     -625     ATCATCTTCTGGGAAACTCTTC    -604
              C  ATCA CT  TGGGAAACT TTC
Fr:     -3027    ATCACCTCATGGGAAACTTTTC    -3006
```



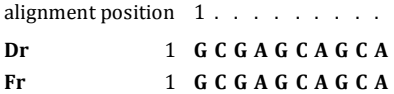
>P1_Block_18

```
Dr:     -510     CATGTGAC-TGGAGTTCCACCTA-TCAGA    -484
             D   CATGTGAC TGG GTTC ACC A TCAGA
Fr:     -544     CATGTGACGTGG-GTTCGACC-ACTCAGA    -518
```

>P1_Block_19

```
Dr:     -472     GCGAGCAGCA    -463
             D   GCGAGCAGCA
Fr:     -3549    GCGAGCAGCA    -3540
```

No common TF matches found.



>P1_Block_20

```
Dr:     -454     GCGGGAGCTGTG    -443
             D   GCGGGAGCTGTG
Fr:     -1597    GCGGGAGCTGTG    -1608
```

V$SMAD

```
alignment position  1 . . . . . .  1 1
Dr              1   G C G G G A G C T G  T G
Fr              1   G C G G G A G C T G  T G
```

>P1_Block_21

```
Dr:     -202     CTCGCTCTCACCACA    -188
           D     CTCGCTC CACCACA
Fr:     -3392    CTCGCTCCCACCACA    -3378
```

V$MZF1

```
alignment position  1 . . . . . .  1 1 . . .
Dr              1   C T C G C T C T C A  C C A C A
Fr              1   C T C G C T C C C A  C C A C A
```

>P1_Block_22

```
Dr:     -162     GGCTTGTGGCGAAAGATTCCTGTGGCACTCTCAAACCCCAAATTCTTGG -114
           B     GGCTTGTGGC    GATTCCTGTGG   TC CAAACC  A  TTCT GG
Fr:     -179     GGCTTGTGGCTCTGGATTCCTGTGGTGGTCACAAACCG-A--TTCTCGG -134

Dr:     -113     TTGCAGGCTGTCTATG    -98
           B     TTGC  GC  TCT TG
Fr:     -133     TTGCGCGCCATCTCTG    -118
```

V$HASF V$HAML V$ZF08 V$GFI1 V$CLOX V$ETSF V$GATA V$GUCE V$EBOX V$IKRS V$KLFS V$AP2F
V$HOMF V$BRNF V$SORY V$STAT V$PAX6 V$XBBF

```
alignment position  1 . . . . . .  1 1 . . . . . .  2 1 . . . . . .  3 1 . . . . . .  4 1 . . . . . .
Dr              1   G G C T T G T G G C  G A A A G A T T C C  T G T G G C A C T C  T C A A A C C c c a  a A T T C T T G G T
Fr              1   G G C T T G T G G C  T C T G G A T T C C  T G T G G T G G T C  A C A A A C C g - -  - A T T C T C G G T

alignment position  5 1 . . . . . .  6 1 . . .
Dr             51   T G C A G G C T G T  C T A T G
Fr             48   T G C G C G C C A T  C T C T G
```

>P1_Block_23

```
Dr:     -80      TTTC-TCC-GACTGTCATTCTC    -61
           D     TTTC TCC GA TGTCATTCTC
Fr:     -5053    TTTCCTCCAGAATGTCATTCTC    -5074
```

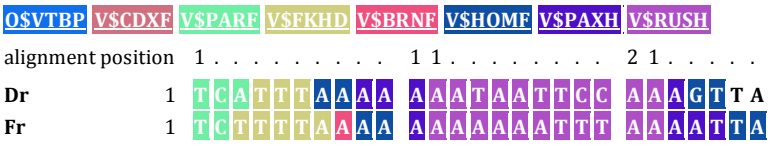
V$TALE V$CSEN V$TCFF

```
alignment position  1 . . . . . .  1 1 . . . . . .  2 1
Dr              1   - T T T C T C C - G  A C T G T C A T T C  T C
Fr              1   t T T C C T C C a G  A A T G T C A T T C  T C
```

(b) Promoter P2

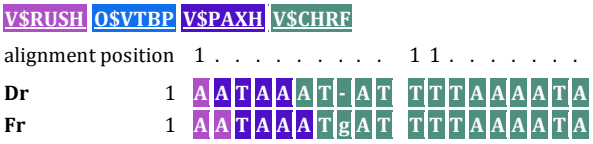
>P2_Block_1

```
Dr:      -5505    TCATTTAAAAAAATAATTCCAAAGTTA    -5479
               B  TC TTTAAAAAAA AA T  AAA TTA
Fr:      -294     TCTTTTAAAAAAAAAAATTTAAAATTA    -320
```



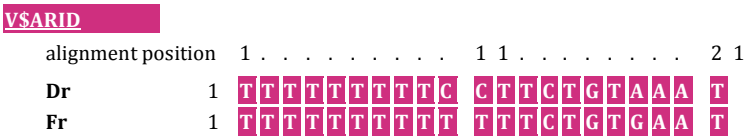
>P2_Block_2

```
Dr:      -4930    AATAAAT-ATTTTAAAATA    -4913
               D  AATAAAT ATTTTAAAATA
Fr:      -2187    AATAAATGATTTTAAAATA    -2205
```



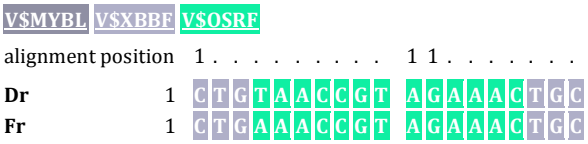
>P2_Block_3

```
Dr:      -4466    TTTTTTTTTCCTTCTGTAAAT    -4446
               C  TTTTTTTTT  TTCTGT AAT
Fr:      -3630    TTTTTTTTTTTTTCTGTGAAT    -3650
```



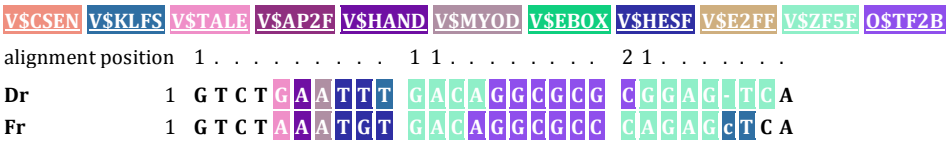
>P2_Block_4

```
Dr:      -3926    CTGTAACCGTAGAAACTGC    -3908
               D  CTG AACCGTAGAAACTGC
Fr:       2808    CTGAAACCGTAGAAACTGC    -2790
```



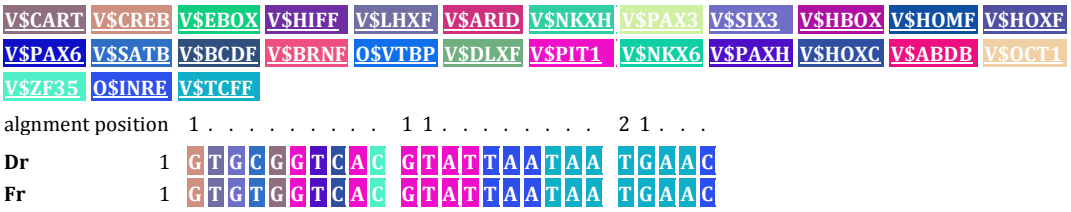
>P2_Block_5

```
Dr:      -3900    GTCTGAATTTGACAGGCGCGCGGAG-TCA    -3873
               C  GTCT AAT TGACAGGCGC C GAG TCA
Fr:      -2773    GTCTAAATGTGACAGGCGCCCAGAGCTCA    -2745
```



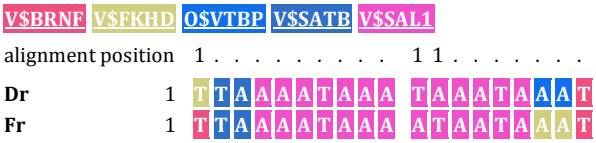
>P2_Block_6

```
Dr:      -3682    GTGCGGTCACGTATTAATAATGAAC    -3838
               D  GTG GGTCACGTATTAATAATGAAC
Fr:      -2713    GTGTGGTCACGTATTAATAATGAAC    -2689
```



>P2_Block_7

```
Dr:      -2982    TTAAAATAAA-TAAATAAAT    -2964
               D  TTAAAATAAA TAA TAAAT
Fr:      -2517    TTAAAATAAAATAA-TAAAT    -2535
```



>P2_Block_8

```
Dr:      -974     ATAACTAT-TTTAAATAGATTATTTA    -950
             D    ATAA TAT TTTAAATAGA TATTTA
Fr:      -2322    ATAAATATATTTAAATAGAATATTTA    -2347
```

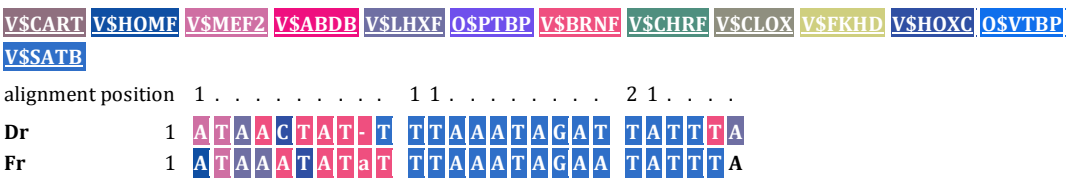


Figure 3.S1 TFBS families detected by DiAlignTF that are common in all conserved blocks obtained from the alignment of *runx3* P1 (a) and P2 (b) promoter regions between zebrafish and fugu. In each block is represented the alignment obtained by DBA software (upper alignment) and the DiAlignTF output (lower alignment) showing the TFBSs conserved in each block. The colour code for each specific TFBS is shown above the alignment.

3.7 References

Akimenko MA, Ekker M, Wegner J, Lin W, Westerfield M, 1994. Combinatorial expression of three zebrafish genes related to distal-less: part of a homeoboxgene code for the head. J Neurosci. 14:3475-3486.

Bae SC, Ogawa E, Maruyama M, Oka H, Satake M, Shigesada K, Jenkins NA, Gilbert DJ, Copeland NG, Ito Y, 1994. PEBP2 alpha B/mouse AML1 consists of multiple isoforms that possess differential transactivation potentials. Mol Cell Biol. 14:3242-3252.

Bangsow C, Rubins N, Glusman G, Bernstein Y, Negreanu V, Goldenberg D, Lotem J, Ben-Asher E, Lancet D, Levanon D, Groner Y, 2001. The RUNX3 gene - sequence, structure and regulated expression. Gene. 279:221-232.

Callaerts P, Munoz-Marmol AM, Glardon S, Castillo E, Sun H, Li WH, Gehring WJ, Salo E, 1999. Isolation and expression of a pax-6 gene in the regenerating and intact planarian Dugesia(G)tigrina. Proc Natl Acad Sci USA. 96:558-563.

Cao Y, Li H, Sun Y, Chen X, Liu H, Gao X, Liu X, 2010. Interferon regulatory factor 4 regulates thymocyte differentiation by repressing Runx3 expression. Eur J Immunol. 40:3198-3209.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T, 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 21:2933–2942.

Chen YH, Lin YT, Lee GH, 2009. Novel and unexpected functions of zebrafish CCAAT box binding transcription factor (NF-Y) B subunit during cartilages development. Bone. 44:77-784.

Conceição N, Cox CJ, Simões B, Viegas M, Cancela ML, 2009. Comparative promoter analysis and its application to the identification of candidate regulatory factors of cartilage-expressed genes. J Appl Ichthyol. 26:245-250.

Cui H, Bulleit RF, 1998. Expression of the POU transcription factor Brn-5 is an early event in the terminal differentiation of CNS neurons. J Neurosci Res. 52:625-632.

Dalcq J, Pasque V, Ghaye A, Larbuisson A, Motte P, Martial JA, Muller M, 2012. RUNX3, EGR1 and SOX9B form a regulatory cascade required to modulate BMP-signaling during cranial cartilage development in zebrafish. PLoS One. 7:e50140.

De Bruijn MF, Speck NA, 2004. Core-binding factors in hematopoiesis and immune function. Oncogene. 23:4238-4248.

Drissi H, Luc Q, Shakoori R, Lopes SCS, Choi JY, Terry A, Hu M, Jones S, Neil JC, Lian JB, Stein JL, Van Wijnen AJ, Stein GS, 2000. Transcriptional autoregulation of the bone related CBFA1/RUNX2 gene. J Cell Physiol. 184:341-350

Ducy P, Starbuck M, Priemel M, Shen J, Pinero G, Geoffroy V, Amling M, Karsenty G, 1999. A Cbfa1-dependent genetic pathway controls bone formation beyond embryonic development. Genes Dev. 13:1025-1036.

Dworkin S, Heath JK, deJong-Curtain TA, Hogan BM, Lieschke GJ, Malaterre J, Ramsay RG, Mantamadiotis T, 2007. CREB activity modulates neural cell proliferation, midbrain-hindbrain organization and patterning in zebrafish. Dev Biol. 307:127-141.

Dykes IM, Lanier J, Eng SR, Turner EE, 2010. Brn3a regulates neuronal subtype specification in the trigeminal ganglion by promoting Runx expression during sensory differentiation. Neural Dev. 5:1-18.

Dykes IM, Tempest L, Lee SI, Turner EE, 2011. Brn3a and Islet1 act epistatically to regulate the gene expression program of sensory differentiation. J Neurosci. 31:9789-9799.

Egawa T, 2009. Runx and ThPOK: a balancing act to regulate thymocyte lineage commitment. J Cell Biochem. 107:1037-1045.

Fainaru O, Woolf E, Lotem J, Yarmus M, Brenner O, Goldenberg D, Negreanu V, Bernstein Y, Levanon D, Jung S, Groner Y, 2004. Runx3 regulates mouse TGF-beta-mediated dendritic cell function and its absence results in airway inflammation. EMBO J. 23:969-979.

Flores MV, Lam EY, Crosier P, Crosier K, 2006. A hierarchy of Runx transcription factors modulate the onset of chondrogenesis in craniofacial endochondral bones in zebrafish. Dev Dyn. 235:3166-3176.

Fu Y, Chang AC, Fournier M, Chang L, Niessen K, Karsan A, 2011. RUNX3 maintains the mesenchymal phenotype after termination of the Notch signal. J Biol Chem. 286:11803-11813.

Gan Q, Lee A, Suzuki R, Yamagami T, Stokes A, Nguyen BC, Pleasure D, Wang J, Chen HW, Zhou CJ, 2013. Pax6 mediates ß-catenin signaling for self-renewal and neurogenesis by neocortical radial glial stem cells. Stem Cells. 32:45-58.

Ghozi MC, Bernstein Y, Negreanu V, Levanon D, Groner Y, 1996. Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. Proc Natl Acad Sci USA. 93:1935-1940.

Hasegawa S, Kume H, Iinuma S, Yamasaki M, Takahashi N, Fukui T, 2012. Acetoacetyl-CoA synthetase is essential for normal neuronal development. Biochem Biophys Res Commun. 427:398-403.

Hinits Y, Pan L, Walker C, Dowd J, Moens CB, Hughes SM, 2012. Zebrafish Mef2ca and Mef2cb are essential for both first and second heart field cardiomyocyte differentiation. Dev Biol. 369:199-210.

Hutchinson SA, Cheesman SE, Hale LA, Boone JQ, Eisen JS, 2007. Nkx6 proteins specify one zebrafish primary motoneuron subtype by regulating late islet1 expression. Development. 134:1671-1677.

Inbal A, Kim SH, Shin J, Solnica-Krezel L, 2007. Six3 represses nodal activity to establish early brain asymmetry in zebrafish. Neuron. 55:407-415.

Inoue K, Ozaki S, Ito K, Iseda T, Kawaguchi S, Ogawa M, Bae SC, Yamashita N, Itohara S, Kudo N, Ito Y, 2003. Runx3 is essential for the target-specific axon pathfinding of trkc-expressing dorsal root ganglion neurons. Blood Cells Mol Dis. 30:157-160.

Kalev-Zylinska ML, Horsfield JA, Flores MV, Postlethwait JH, Chau JY, Cattin PM, Vitas MR, Crosier PS, Crosier KE, 2003. Runx3 is required for hematopoietic development in zebrafish. Dev Dyn. 228:323-336.

Karsenty G, 2000. Role of Cbfa1 in osteoblast differentiation and function. Semin. Cell Dev Biol. 11:343-346.

Kawai S, Yamauchi M, Wakisaka S, Ooshima T, Amano A, 2007. Zinc-finger transcription factor odd-skipped related 2 is one of the regulators in osteoblast proliferation and bone formation. J Bone Miner Res. 22:1362-1372.

Komori T, 2005. Regulation of Skeletal Development by the Runx Family of Transcription Factors. J Cell Biochem. 95:445-453.

Lai CB, Mager DL, 2012. Role of runt-related transcription factor 3 (RUNX3) in transcription regulation of natural cytotoxicity receptor 1 (NCR1/NKp46), an activating natural killer (NK) cell receptor. J Biol Chem. 287:7324-7334.

Lambertini E, Penolazzi L, Tavanti E, Pocaterra B, Schincaglia GP, Torreggiani E, Franceschetti T, Vecchiattini R, Gambari R, Piva R, 2008. Modulation of expression of specific transcription factors involved in the bone microenvironment. Minerva Biotec. 20:69-77.

Landry JR, Kinston S, Knezevic K, de Bruijn MF, Wilson N, Nottingham WT, Peitz M, Edenhofer F, Pimanda JE, Ottersbach K, Göttgens B, 2008. Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood. 111:3005-3014.

Larbuisson A, Dalcq J, Martial JA, Muller M, 2013. Fgf receptors Fgfr1a and Fgfr2 control the function of pharyngeal endoderm in late cranial cartilage development. Differentiation. 86:192-206.

Levanon D, Goldstein RE, Bernstein Y, Tang H, Goldenberg D, Stifani S, Paroush Z, Groner Y, 1998. Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. Proc Natl Acad Sci USA. 95:11590-11595.

Levanon D, Brenner O, Negreanu V, Bettoun D, Woolf E, Eilam R, Lotem J, Gat U, Otto F, Speck N, Groner Y, 2001. Spatial and temporal expression pattern of Runx3 (Aml2) and

Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis. Mech Dev. 109:413-417.

Levanon D, Groner Y, 2004. Structure and regulated expression of mammalian RUNX genes. Oncogene. 23:4211-4219.

Li D, Dower K, Ma Y, Tian Y, Benjamin TL, 2001. A tumor host range selection procedure identifies p150 (sal2) as a target of polyoma virus large T antigen. Proc Natl Acad Sci USA. 98:14619-14624.

Li QL, Ito K, Sakakura C, Fukamachi H, Inoue K, Chi XZ, Lee KY, Nomura S, Lee CW, Han SB, Kim HM, Kim WJ, Yamamoto H, Yamashita N, Yano T, Ikeda T, Itohara S, Inazawa J, Abe T, Hagiwara A, Yamagishi H, Ooe A, Kaneda A, Sugimura T, Ushijima T, Bae SC, Ito Y, 2002. Causal relationship between the loss of RUNX3 expression and gastric cancer. Cell. 109:113-124.

Lim B, Ju H, Kim M, Kang C, 2011. Increased genetic susceptibility to intestinal-type gastric cancer is associated with increased activity of the RUNX3 distal promoter. Cancer. 117:5161-5171.

Ma Y, Li D, Chai L, Luciani AM, Ford D, Morgan J, Maizel AL, 2001. Cloning and characterization of two promoters for the human HSAL2 gene and their transcriptional repression by the Wilms tumor suppressor gene product. J Biol Chem. 276:48223-48230.

Mei PJ, Bai J, Liu H, Li C, Wu YP, Yu ZQ, Zheng JN, 2011. RUNX3 expression is lost in glioma and its restoration causes drastic suppression of tumor invasion and migration. J Cancer Res Clin Oncol. 137:1823-1830.

Miller CT, Swartz ME, Khuu PA, Walker MB, Eberhart JK, Kimmel CB, 2007. mef2ca is required in cranial neural crest to effect Endothelin1 signaling in zebrafish. Dev Biol. 308:144-157.

Moore S, Ribes, V, Terriente J, Wilkinson D, Relaix F, Briscoe J, 2013. Distinct regulatory mechanisms act to establish and maintain Pax3 expression in the developing neural tube. PLoS Genet. 9:e1003811.

Morgenstern B, Frech K, Dress A, Werner T, 1998. DIALIGN: finding local similarities by multiple sequence alignment. Bioinformatics. 14:290-14294.

Park JH, Adoro S, Guinter T, Erman B, Alag AS, Catalfamo M, Kimura MY, Cui Y, Lucas PJ, Gress RE, Kubo M, Hennighausen L, Feigenbaum L, Singer A, 2010. Signaling by intrathymic cytokines, not T cell antigen receptors, specifies CD8 lineage choice and promotes the differentiation of cytotoxic-lineage T cells. Nat Immunol. 11:257-264.

Pincheira R, Baerwald M, Dunbar JD, Donner DB, 2009. Sall2 is a novel p75NTR-interacting protein that links NGF signalling to cell cycle progression and neurite outgrowth. Embo J. 28:261-273.

Pinto JP, Conceição NM, Viegas CS, Leite RB, Hurst LD, Kelsh RN, Cancela ML, 2005. Identification of a new pebp2alphaA2 isoform from zebrafish runx2 capable of inducing osteocalcin gene expression in vitro. J Bone Miner Res. 20:1440-1453.

Rini D, Calabi F, 2001. Identification and comparative analysis of a second runx3 promoter. Gene. 273:13-22.

Shen ZJ, Nakamoto T, Tsuji K, Nifuji A, Miyazono K, Komori T, Hirai H, Noda M, 2002. Negative regulation of bone morphogenetic protein/Smad signaling by Cas-interacting zinc finger protein in osteoblasts. J Biol Chem. 277:29840-29846.

Simões B, Conceição N, Viegas CS, Pinto JP, Gavaia PJ, Hurst LD, Kelsh RN, Cancela ML, 2006. Identification of a promoter element within the zebrafish colXalpha1 gene responsive to runx2 isoforms Osf2/Cbfa1 and til-1 but not to pebp2alphaA2. Calcif Tissue Int. 79:230-244

Soung do Y, Dong Y, Wang Y, Zuscik MJ, Schwarz EM, O'Keefe RJ, Drissi H, 2007. Runx3/AML2/Cbfa3 regulates early and late chondrocyte differentiation. J Bone Miner Res. 22:1260-1270.

Spender LC, Whiteman HJ, Karstegl CE, Farrell PJ, 2005. Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells. Oncogene. 24:1873-1881.

van Heyningen V, Williamson KA, 2002. PAX6 in sensory development. Hum Mol Genet. 11:1161-1167.

Wang S, Wang Q, Crute BE, Melnikova IN, Keller SR, Speck NA, 1993. Cloning and Characterization of Subunits of the T-Cell Receptor and Murine Leukemia Virus Enhancer Core-Binding Factor. Mol Cell Biol. 13:3324-3339.

Wang S, Furmanek T, Kryvi H, Krossøy C, Totland GK, Grotmol S, Wargelius A, 2014. Transcriptome sequencing of Atlantic salmon (Salmo salar L.) notochord prior to development of the vertebrae provides clues to regulation of positional fate, chordoblast lineage and mineralisation. BMC Genomics. 15:141.

Wenke AK, Rothhammer T, Moser M, Bosserhoff AK, 2006. Regulation of integrin alpha10 expression in chondrocytes by the transcription factors AP-2epsilon and Ets-1. Biochem Biophys Res Commun. 345:495-501.

Westendorf JJ, Hiebert SW, 1999. Mammalian Runt-Domain Proteins and Their Roles in Hematopoiesis, Osteogenesis, and Leukemia. J Cell Biochem. Supplements 32/33: 51-58.

Wu S, Zang W, Li X, Sun H, 2011. Proepithelin stimulates growth plate chondrogenesis via nuclear factor-{kappa}B-p65-dependent mechanisms. J Biol Chem. 286:24057-24067.

Xiao ZS, Thomas R, Hinson TK, Quarles LD, 1998. Genomic structure and isoform expression of the mouse, rat and human Cbfa1/Osf2 transcription factor. Gene. 214:187-197.

Yang HN, Park JS, Woo DG, Jeon SY, Do HJ, Lim HY, Kim SW, Kim JH, Park KH, 2011. Chondrogenesis of mesenchymal stem cells and dedifferentiated chondrocytes by transfection with SOX Trio genes. Biomaterials. 32:7695-7704.

Yoshida CA, Yamamoto H, Fujita T, Furuichi T, Ito K, Inoue K, Yamana K, Zanma A, Takada K, Ito Y, Komori T, 2004. Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. Genes Dev. 18:952-963.

Zamisch M, Tian L, Grenningloh R, Xiong Y, Wildt KF, Ehlers M, Ho IC, Bosselut R, 2009. The transcription factor Ets1 is important for CD4 repression and Runx3 up-regulation during CD8 T cell differentiation in the thymus. J Exp Med. 206:2685-2699.

Chapter 4

Molecular characterization of CBFβ gene and identification of new transcription variants: Implications for function

This chapter is based in a published research paper:

B Simões, N Conceição, AC Matias, J Brangança, RN Kelsh, ML Cancela

Archives of Biochemistry and Biophysics (2015) 567:1–12

Author's contribution:

Most of the experimental work and writing of the paper was performed by B Simoes, except the gene synteny and co-transfection analysis performed by N Conceição and the Co-IP performed by AC Matias and J Bragança. The research concept and design was performed by B Simões, N Conceição and ML Cancela. N Conceição, ML Cancela and RN Kelsh were responsible for the critical revision and final approval of the manuscript.

4.1 Abstract

The *CBFβ* gene encodes a transcription factor that, in combination with CBFα (also called Runx, runt-related transcription factor) regulates expression of several target genes. CBFβ interacts with all Runx family members, such as RUNX2, a master-regulator of bone-related gene transcription that contains a conserved DNA-binding domain (Runt domain). CBFβ stimulates DNA binding of the Runt domain, and is essential for most of the known functions of RUNX2.

A comparative analysis of the zebrafish *cbfβ* gene and protein, and of its orthologous identified homologous proteins in different species indicates a highly conserved function. We cloned eleven transcripts of the zebrafish *cbfβ* gene, one resulting in the known Cbfβ protein (with 187 amino-acids (aa)), and three additional variants resulting from skipping exon 5a (resulting in a protein with 174 aa) or exon 5b resulting in a protein with 201 aa), both observed for the first time in zebrafish, and a completely novel isoform containing both exon 5a and 5b (and resulting in a protein with 188 aa). Functional analysis of these isoforms provides insight into their role in regulating gene transcription. From the other seven variants two correspond to premature early termination of Cbfβ forms, while the others show in-frame exon-skipping causing changes in the Cbfβ domain that may affect its function.

4.2 Introduction

Chondrocytes, osteoblasts, and osteoclasts are the major cell types that contribute to the development and maintenance of the skeleton (Erlebacher et al, 1995). Vertebrate skeletons are constructed by the formation of bone and cartilage structures that can occur via two distinct mechanisms: intramembranous and endochondral ossification. During intramembranous (or dermal) ossification, mesenchymal cells condense and differentiate into osteoblasts, the bone-forming cells. In contrast, during chondral ossification, mesenchymal cells condense and differentiate into chondrocytes to form a cartilage template. Subsequently, this template is either replaced by bone (endochondral ossification) or it becomes surrounded by bone (perichondral ossification) (Spoorendonk et al, 2010).

The importance of runt-related transcription factor 2 (RUNX2), in skeletal development was first suggested by studies of the autosomal dominant disease cleidocranial dysplasia (CCD) (Mundlos et al, 1995; reviewed in Martin et al, 2011). RUNX2 is a known master transcription factor for bone and hypertrophic cartilage formation expressed very early in bone development and continues to be present through the later phases of development (Ducy et al, 1997). It is essential for osteoblast differentiation as well as a critical regulator for chondrocyte maturation (Komori et al, 1997; Kishimoto et al, 1997; Otto et al, 1997; Kim et al, 1999; Inada et al, 1999; Takeda et al, 2001; Hinoi et al, 2006). *RUNX2* belongs to the Runt-related transcription factor (RUNX) family of genes which are also called core binding factor-α (*CBFα*). The other two members identified are RUNX1 (AML1/CBFα2/PEBP2αB) and RUNX3 (AML2/CBFα3/PEBP2αC). The RUNX proteins can bind DNA as a monomer *in vitro*, but their affinity for DNA is enhanced when binding to the DNA as a CBFα:β heterodimers (Ogawa et al, 1993; Wang et al, 1993). Unlike CBFα, the CBFβ subunit does not contact DNA directly, but rather stabilizes and enhances *in vitro* DNA binding of the runt domain of the α subunit (Ogawa et al, 1993; Wang et al, 1993), which is a DNA binding domain conserved amongst the Runx family (Ogawa et al, 1993). Earlier studies have indicated that CBFβ and RUNX2 can cooperatively activate transcription (Harada et al, 1999; Zhang et al, 2000). Kundu et al (2002) carried out a series of experiments to determine whether CBFβ and Runx2 could interact physically and function in a cooperative manner, and have shown that the addition of CBFβ strongly induced the DNA binding of Runx2 (Kundu et al, 2002).

Runx2 initiates and mediates the entire process of hypertrophic differentiation of chondrocytes (Stricker et al, 2002; Smith et al, 2005) by regulating the transcription of genes important for this process, e.g. collagen type X gene (*Col10α1*) (Enomoto et al, 2000; Zheng et al, 2003; Higashikawa et al, 2009). RUNX2 regulation of cell-specific *Col10α1* expression may impact the process of chondrocyte maturation and represent the major mechanistic basis of multiple skeletal pathologies, such as CCD, fracture healing, and osteoarthritis (Higashikawa et al, 2009; Zheng et al, 2005; Kamekura et al, 2006; Tu et al, 2007). Zheng et al (2005) have previously reported abnormal endochondral ossification in a fetal case of CCD, possibly due to altered RUNX2 regulation of chondrocyte hypertrophy and down-regulation of its target genes, including type X collagen. The above observations clearly demonstrate that both *Runx2* and *Col10α1* genes play important roles upon chondrocyte maturation

during endochondral bone formation. The interaction between Runx2 and the *Col10α1* proximal or core promoters in different species has previously been described extensively (Dourado and LuValle, 1998; Zheng et al, 2003; Simões et al, 2006; Higashikawa et al, 2009).

Most studies in the areas of osteogenesis and mineral research have been performed in mice and chicken, or using *in vitro* cell culture systems. Although it has been shown that there are some characteristics in teleost bones that differ from mammals (Witten and Huysseune, 2009), the origin of cells that contribute to the various bone elements and the key regulators of bone formation are highly conserved between mammals and teleosts. Furthermore, the corresponding orthologs share significant sequence similarities and an overlap in expression patterns (Flores et al, 2004; Yan et al, 2005; Li et al, 2009) when compared to mammals. As a result of this finding, in the last few years zebrafish was demonstrated to be a powerful model especially in forward genetics to identify novel gene functions and to study their role in numerous processes including osteogenesis. Accordingly, zebrafish can be used as a tool to complement genetic and embryological studies in mice and chicken in order to clarify the molecular mechanisms underlying bone development and disease. In addition, zebrafish and medaka are ideally suited and currently the only model systems available to allow visualization of chondrocytes and osteoblasts *in vivo* over time.

Thus far, different CBFβ isoforms have been described in mammals, but just one zebrafish Cbfβ protein has been reported. Here we report the cloning and characterization of ten novel zebrafish isoforms, which are generated by alternative splicing. A structural conservation during evolution from fish to mammals was confirmed, by a comparative analysis between zebrafish *cbfβ* gene and protein and its orthologs in different species. Previously, we have shown that zebrafish *col10α1* expression is up-regulated by Runx2 (Simões et al, 2006) through its binding to specific motifs within the *col10α1* promoter region. So, we tested the ability of some of these newly identified Cbfβ isoforms to enhance Runx2-dependent up-regulation of *col10α1* promoter. The transcriptional activity determined by luciferase reporter assays was enhanced by transfection of Runx2-MASN isoform and increased even more potently by the co-transfection of both Runx2-MASN and the co-activator Cbfβ (isoforms 1 and 2) as compared with the control. Furthermore, this indicates that Cbfβ exon 5a is not required for interaction with Runx2-MASN and

transcription activation. Moreover, we analysed the expression pattern of the Cbfβ isoforms 1-4 in various adult tissues and at different embryonic developmental stages.

4.3 Material and Methods

4.3.1 Zebrafish RNA extraction and RNA reverse transcription

Total RNA was extracted from ZFB1 cell line as described by Vijayakumar et al (2013) and from pools of zebrafish embryos at different stages of development and from a variety of adult zebrafish tissueswith TRIzol (Sigma Aldrich) as recommended in the manufacturer's protocol. RNA integrity was assessed through 1% (w/v) agarose/formaldehyde gel electrophoresis and RNA quantity was determined through spectrophotometry (NanoDrop 1000; Thermo Scientific). Total RNA (1 µg) was then treated with RQ1 RNase-free DNase I (Promega) for 30 min at 37ºC, and reverse-transcribed at 37°C for 1 h using the Moloney-murine leukemia virus (MMLV) reverse transcriptase, RNaseOUT (both from Invitrogen) and oligo(dT)-adapter primer (**Table 4.1**).

Table 4.1 Oligonucleotides used for PCR amplification.

Name	Sequence
ZfCBFbFw1	GAGCGTCTGTTGTCAGCAGTCGGA
ZfCBFbFw2	CGTTCAAGATGCCTCGGGTGGTCC
ZfCBFbFw3	GAGGACTCGTGATTTCGAGGACAG
ZfCBFbFw4	GGAGCAGATGCCGATGGCACAGCT
ZfCBFbRev1	CCCCAAAACTCCCCCAGCGGTGTG
ZfCBFbRev2	CTGCCCACTTTGGTGAATGCCGCT
ZfCBFbRev3	GTGATCATCAGTGTTGCCCATGTT
ZfCBFbRev5	GTCCTTGAAGGCCATCAGTCCCAGA
Gapdh_F	GTGGAGTCTACTGGTGTCTTC
Gapdh_R	GTGCAGGAGGCATTGCTTACA
Oligo(dt) primer	ACGCGTCGACCTCGAGATCGATG(T)13

4.3.2 Zebrafish *cbfβ* cDNA cloning using RT-PCR

The primer sequences used for cloning are shown in **Table 4.1** and were synthesized and purchased from Sigma-Aldrich. Specific primers (zfCBFbFw1, zfCBFbFw2, zfCBFbRev1 and zfCBFbRev2) were designed to amplify zebrafish *cbfβ* complete cDNA coding region, according to its cDNA sequence available in the NCBI database (GenBank NM_199209.1). Amplification was performed by two steps PCR with zfCBFbFw1 and zfCBFbRev1 primers (0.3 µM each), and either with the Taq DNA polymerase (Invitrogen) or the KOD Hot Start DNA Polymerase (Novagen), in a GeneAmp 2400 thermocycler (Perkin-Elmer), under conditions suggested by the suppliers and using as template cDNA from either ZFB1 cell line or 24 hpf zebrafish. The amplified product was used for the second step PCR. For this step, the reaction mix and PCR conditions were similar to the first step except in that the primer pairs zfCBFbFw2 and zfCBFbRev2 or the zfCBFbFw2 and zfCBFbRev1 (0.3 µM each) were used instead. Amplified fragments were cloned into pCRII-TOPO vector (Invitrogen) by standard TA-cloning or into pJet1.2 vector (Fermentas, Thermo Scientific) by standard blunt-cloning. Cloned fragments were identified by restriction digestion and by sequencing at CCMAR sequencing facilities (University of Algarve, Faro, Portugal). All sequence alignments were performed with ClustalW (Thompson et al, 1994) or using AlignX, from Vector NTI Advance® 11.5 (Invitrogen).

4.3.3 Sequence alignment and analysis

GenBank and Ensembl databases were searched for CBFβ sequences. Amino acid sequence alignments were created using AlignX, from Vector NTI Advance® 11.5 (Invitrogen) or Clustal Omega (http://www.ebi.ac.uk/). Final adjustments to the alignments were made manually to obtain highly accurate consensus sequences. Percentage protein identity was calculated using the Sequence Manipulation Suite (Stothard, 2000) available at http://www.bioinformatics.org. The alternative splicing events in both human and zebrafish, as also for the other species, whose genomic sequence was available, were revealed by aligning the cDNAs against the genomic sequences, using the mRNA alignment tool Spidey (ncbi.nlm.nih.gov/spidey/spideyweb.cgi).

4.3.4 Genomic structure of zebrafish and human *CBFβ* gene

Exon-intron architecture of zebrafish *cbfβ* gene was determined through mRNA-to-genomic alignment using Zv9 zebrafish genome assembly and transcript sequences determined within the scope of this work. Similarly, human gene structure was determined using GRCh37 genome assembly, mRNA and expressed sequence tag (EST) sequences retrieved from NCBI (on 2014-04-13).

4.3.5 Assessment of conserved synteny

To examine patterns of conserved synteny, chromosomal loci of *CBFβ* genes in human, and zebrafish were compared by identifying all neighbour genes of *CBFβ*. The position of each of these genes was searched in both species using the Ensembl database search function.

4.3.6 Isoform expression profile

To determine the presence of the *Cbfβ* alternative transcripts (isoforms 1-4) during various zebrafish developmental stages and in a broad number of adult tissues, primers were designed in order to amplify all the four splice variants. A first RT-PCR amplification was performed with the primers CBFβ_F3 and CBFβ_R3 (**Table 4.1**). Then, 1 µl of the first amplification product was used to perform a second amplification with the CBFβ isoforms 1 and 2 specific primers (CBFβ_F3 and CBFβ_R5; **Table 4.1**) and the CBFβ isoforms 1 and 3 specific primers (CBFβ_F4 and CBFβ_R3; **Table 4.1**). The zebrafish *gadph* was used as control (Gapdh_F and Gapdh_R; **Table 4.1**). The RT-PCR amplification was performed with the DreamTaq PCR Master Mix (Thermo Scientific) in a 25 µl reaction for 40 cycles.

4.3.7 Plasmid construction

The zebrafish collagen Xα1 luciferase reporter plasmid [4x(-822/-794)TATALuC] and zfrunx2 P1-MASN (*til-1*ORF-pCMX-PL1) were previously described (Simões et al, 2006). The expression vectors of the full length and splicing variants of zebrafish *cbfβ* (*cbfβ* isoform 1-4) were obtained by cloning all the corresponding open reading frames into the pCMX-PL2 expression vector (kindly provided by Dr. R. Schüle laboratory).

The zebrafish HA-tagged *cbfβ* and and Flag-tagged *runx2* expression constructs (pcDNA3.1-HA-*cbfβ* isoform 1-4 and pcDNA3.1-Flag-runx2 P1-MASN) were generated by subcloning PCR-amplified full-length *cbfβ* isoform 1 to 4 and *runx2* P1-MASN cDNAs into the *BamH*I and *Xba*I sites of a pcDNA3.1 expression vector containing at the N-terminal portion of the proteins.

All constructs were verified by DNA sequencing. Plasmids used for transfection studies were prepared using the plasmid GFX™ Micro Plasmid Prep kit (GE Healthcare).

4.3.8 Cell transfection and luciferase assays

Human embryonic kidney (HEK) 293 cell line (ATCC number CRL-1573) was cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 1% (v/v) penicillin/streptomycin, 2 mM L-glutamine and 10% (v/v) fetal bovine serum at 37°C in a 5% CO2 humidified atmosphere. Cells were seeded at approximately 40% of confluence in 24-well plates (5×10^4 cells/well) and transient transfection assays were carried out using the X-TREME reagent (Roche). Typically, 125 ng, 25 ng and 2.5 ng of (i) promoter-reporter construct, (ii) transcriptional regulator expression vector and (iii) pRL-null internal control vector (Promega) were used, per well. The amount of transfected DNA was kept constant in both positive and negative control cells, by transfecting them with the same amount of DNA: 125 ng of pGL3-control vector or pGL3-basic vector (both from Promega), respectively; 25 ng of pCMX-PL2 expression vector; and 2.5 ng of pRL-null internal control vector. Luciferase activity was assayed 48 h after transfection using the standard protocol provided with the Dual-luciferase reporter assay system (Promega) in a Synergy 4 microplate reader (Biotek).

Luciferase activity assays were performed in duplicate and are the mean of at least three separate experiments.

4.3.9 Co-immunoprecipitation (Co-IP) Assay

For co-immunoprecipitation assays, ≈0.1-0.2 mg of whole cell extracts from HEK293T cells transfected with expression vectors for the four HA-CBFβ isoforms alone or together with FLAG-Runx2, were prepared in buffer containing 50 mM Tris pH7.5, 150 mM NaCl, 1% Triton-X100, and Complete protease inhibitors (Roche), and incubated with M2 Flag-resin (Sigma) overnight at 4°C. The resin was washed five times with wash buffer (20mM Tris pH 7.5, 100mM NaCl, 0.1mM EDTA, 0.05% Tween-20) and the bound material was eluted with 200 µg/ml solution of flag peptide (Sigma), for 30 minutes at 4°C. Samples were subjected to western blot analysis.

4.3.10 Western Blot Assay

For western blot assays, protein extracts were subjected to 12% SDS-PAGE, and thereafter transferred onto a PVDF membrane (GE Healthcare) with a semi-dry blot system (BioRad). Mouse monoclonal 16B12 antibody against HA epitope (Covance) was used at 1:2000 dilution and anti-flagM2 (Sigma) antibody was used at 1:5000 dilution. Blotted proteins were visualized using horseradish peroxidase-conjugated goat anti-mouse (Southern Biotech), the chemiluminescence blotting substrate detection system from Roche and X-ray films.

4.3.11 Statistical analysis

The data was presented as average and standard deviation of measurements taken at least in three separate experiments. Statistical significance of data was determined wherever indicated by analysis of variance (ANOVA) followed by a Tukey test for multiple comparisons within a group. Differences were considered to be significant at $p<0.001$.

4.4 Results

4.4.1 Molecular cloning of novel spliced variants of zebrafish *cbfβ*

Using a combination of bioinformatics and RT-PCR approaches, we cloned a cDNA fragment encoding the zebrafish *cbfβ* open reading frame (ORF). Sequencing predicts an ORF of 564 bp encoding a 187 aa polypeptide. From the cDNA deduced primary structure, we identified the typical CBFβ domain characteristic of this family of proteins, and sequence comparison with the available zebrafish *cbfβ* cDNA (accession number: NM_199209.1) showed 100% identity. This isoform was named in this work as isoform 4. So far, just one transcript has been described for the zebrafish *cbfβ* gene, contrasting with mammals where two major transcripts have been described for this gene that generate two different protein isoforms with a different C-terminal sequence.

In the course of amplifying the cDNA for the entire ORF of zebrafish *cbfβ* using RT-PCR, we observed multiple amplified products. By cloning and sequencing each one of them we were able to identify ten novel transcript variants for zebrafish *cbfβ* that are described here for the first time. **Figure 4.1** shows both the simple and compound deletion events discovered in this study. These transcripts result from alternative exon skipping, generating different protein isoforms, depending on the splicing event. The identity of the *cbfβ* cDNA sequences obtained was confirmed using blast searches against GenBank (NCBI). The nucleotide sequences of these new spliced variants were deposited in GenBank as *cbfβ* isoforms 1 through 11 (GenBank ID: KF709194, KF709195, KF709196, KF709197, KF709198, KF709199, KF709200, KJ704807, KJ704808, KJ704809 and KJ704810, respectively).

The transcript *cbfβ* isoform 1 corresponds to the longest transcript we cloned and has an additional 90 nucleotides compared to the *cbfβ* transcript described previously (**Figure 4.1**). This extra nucleotide sequence in *cbfβ* isoform 1 within the 3' coding region generates a stop codon located in exon 5b, resulting in a protein isoform with a different C-terminal sequence from the form described previously (**Figure 4.2**).

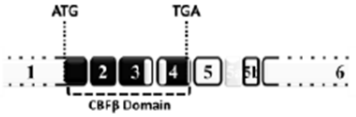
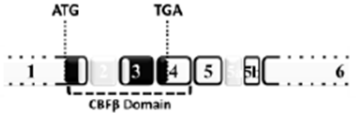
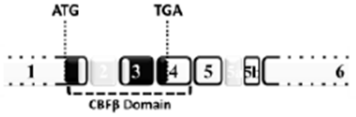
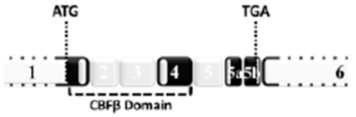
Isoforms	Description	Schematic diagram of cbfβ splice forms	mRNA* (bp)	Predicted amino acids	GenBank accession n°
Isoform 1	WT		684	188	KF709194
Isoform 2	Δ6		642	174	KF709195
Isoform 3	Δ7		636	201	KF709196
Isoform 4	Δ6,7		594	187	KF709197
Isoform 5	pΔ3,4; Δ6		544	94	KF709198
Isoform 6	pΔ1; Δ2; pΔ3; Δ6		431	26	KF709199
Isoform 6	pΔ1; Δ2; pΔ3; Δ6		431	26	KF709199
Isoform 7	pΔ1; Δ2,3; pΔ4; Δ5; pΔ6		300	60	KF709200
Isoform 8	pΔ8		781	188	KJ704807
Isoform 9	pΔ6		783	187	KJ704808
Isoform 10	Δ2; pΔ6,8		587	158	KJ704809
Isoform 11	pΔ3; Δ4,5,6,7; pΔ8		261	>84	KJ704810

Figure 4.1 Schematic representation of zebrafish *cbfβ* transcripts. The different transcripts originated by alternative splicing are indicated as isoforms 1-11 with the respective accession numbers. Black boxes indicate coding regions, white boxes represent non coding regions and grey boxes indicate DNA fragments removed following splicing. * Size of the cloned cDNA fragments obtained.

The transcripts of *cbfβ* isoform 2 and 3 are spliced variants generated by skipping exon 5a (i.e. a 42 bp fragment) or 5b (i.e. a 48 bp fragment), respectively. The resulting protein products of these alternative splicing variants are similar to *cbfβ* isoforms 1 and 4, respectively, in terms of the stop codon used (**Figure 4.1**). All the *cbfβ* isoforms 1-4 preserve the heterodimerization domain intact, suggesting that they produce functional proteins.

Interestingly, we found two predicted isoforms similar to the *cbfβ* isoform 1 and 2 in the NCBI database (XM_005159048.1 and XM_005159049.1), supporting together with our results the existence of these alternative splicing isoforms in zebrafish *cbfβ.*

The transcripts *cbfβ* corresponding to the isoform 5, 6 and 7, are generated by a complex splicing of multiple sequence fragments from different exons, resulting in truncated isoforms, lacking an extensive part of the characteristic CBFβ heterodimerization domain. As a consequence, all of these *cbfβ* transcripts are likely to result in loss-of-function mutants.

The transcript *cbfβ* isoform 8 results from an alternative splicing event that involves a partial deletion of the exon 6 (5 bp). This deletion is in the 3' UTR and does not affect the coding region, so the predicted protein isoform is exactly the same as the one produced by *cbfβ* isoform 1.

The transcript *cbfβ* isoform 9 is also similar to the *cbfβ* isoform 1, differing only in a deletion of three nucleotides in the beginning of exon 5a and expected to result in deletion of one amino acid (Q166) from the encoded protein. Interestingly, an isoform similar to the *cbfβ* isoform 9 was recently submitted to the NCBI database as a predicted isoform (XP_005159104.1), but presenting the splicing of the exon 5b in addition to the 3 nucleotides deletion we characterized.

The *cbfβ* isoform 10 harbours both the deletion of the 3 nucleotides present in *cbfβ* isoform 9 as well as the deletion of the 5 nucleotides observed in *cbfβ* isoform 8. This isoform presents also the complete deletion of exon 2 (**Figure 4.1**). Since exon 2 codes for part of the heterodimerization domain, we hypothesize that this change leads to loss of function of this isoform.

The new transcript *cbfβ* isoform 11 has a splice starting at the 3'-end of exon 3 (pΔ3 51 bp) and utilizes a splice acceptor site within exon 6 instead of the "native" acceptor site. Analysing the cDNA sequence obtained for this isoform, we observe that this deletion event causes an out-of-frame translation and furthermore, introduces a stop codon after the alternative splicing (in exon 6) that is located 67 nucleotides downstream the annotated stop codon for the *cbfβ*, thereby introducing a late termination of protein translation, resulting in a protein isoform with a different C-terminus sequence.

Summarizing, we found in this study simple exon-skipping events that can be categorized as follows: (a) simple deletions, that is skipping of complete single exons or consecutive exons (for example, Δ5a; Δ5b; Δ5a,5b; isoforms 2-4, respectively) and (b) partial exon deletions (for example, pΔ6 5 bp; pΔ5a 3 bp; isoforms 8 and 9). In addition to the five simple exon-skipping events shown in **Figure 4.1**, five more splicing events were identified that involved multiple exon-skipping events. In these compound exon-skipping transcripts, several splicing events were evident from the pre-mRNA processing (**Figure 4.1**). These compound splicing events include combinations of the whole exon-skipping and partial exon deletion (pΔ3 56 bp, pΔ4 42 bp plus Δ5a; pΔ1 39 bp, Δ2, pΔ3 85 bp, plus Δ5a; pΔ1 39 bp, Δ2,3, pΔ4 27 bp, Δ5, plus pΔ5a 18 bp; Δ2, plus pΔ5a 3 bp plus pΔ6 5 bp; pΔ3 51 bp, Δ4, 5,5a,5b plus pΔ6 69 bp; isoforms 5-7, 10 and 11). These multiple exon skipping events all involved the entire or partial deletion of exon 5a, which implies that the skipping of this exon is a common event in the pre-mRNA processing.

As mentioned previously, we identified in this study alternative splicing events that involve partial exon deletions instead of the skipping of complete exons. The splice sites used are shown in **Table 4.2**. Three of these variants (isoforms 9 and 10, pΔ5a 3bp and isoforms 8 and 10, pΔ6 5bp) still keep their original splicing donor site (GT) in the boundary of the exon and intron but utilize the next possible legitimate splice acceptor site (AG) within the exon in the immediate vicinity area, instead of the 'native' acceptor site. These two partial deletion events do not cause out-of-frame translation and, furthermore, they use the same stop codon as the full transcript (isoform 1). The splicing events of the other four variants (isoforms 5, 6, 7 and 11) are more complex. In the partial deletions pΔ3-4 (isoform 5) and pΔ1-4 (isoform 7) not only are none of the original splice donor/acceptor sites used, but

those that are used are very atypical (**Table 4.2**). For the remaining variants pΔ1-3 (isoform 6) and pΔ3-6 (isoform 11), do not use a pair of legitimate splice donor site within exon 1 (isoform 6) or exon 3 (isoform 11) but utilize the next possible legitimate splice acceptor site (AG) within the exon 3 (isoform 6) or exon 6 (isoform 11), instead of the 'native' acceptor site, causing a 211 bp (isoform 6) or a 423 bp (isoform 11) deletion.

Table 4.2 Splice code usages of the partial exon-skipping in *cbfβ* mRNA.

Transcripts	Partial exon-skipping	Splice donor sites	Splice acceptor sites	Deletion caused by alternative splicing
Isoform 9 and 10	pΔ5a	TTGGAGgtgagagct	catttagcagATGCCGATG	3 bp
Isoform 8 and 10	pΔ6	GGACCAGGgtctgtctcc	cccagcccagGCGACAGCAG	5 bp
Isoform 5	pΔ3,4	TTCATGGGGatcagcggc	agtatgtgtgATCTGGAGAG	98 bp
Isoform 6	pΔ1,3	GTTCGAGaacgagga	cgcccacccgagAATATGTGG	211 bp
Isoform 7	pΔ1,4	GTTCGAGAACGaggagttctt	atcctgaacgGAGTATGTGTG	270 bp
Isoform 7	pΔ5a	GCACAGgtacagcaaaat	gatggcacagCTCAATCAT	270 bp
Isoform 11	pΔ3,6	GATCAGCGgcaggcg	gcgctagcgGCATTCAC	126 bp

4.4.2 Translation potential of the *cbfβ* spliced variants

The predicted protein sequences translated from these alternative spliced transcripts are summarized in **Figure 4.2**. Two of these variants (isoforms 5 and 6) introduce early termination codons to the open reading frames after the alternative splicing event(s), which may produce premature proteins with only short peptides (94 aa and 26 aa, respectively) of Cbfβ.

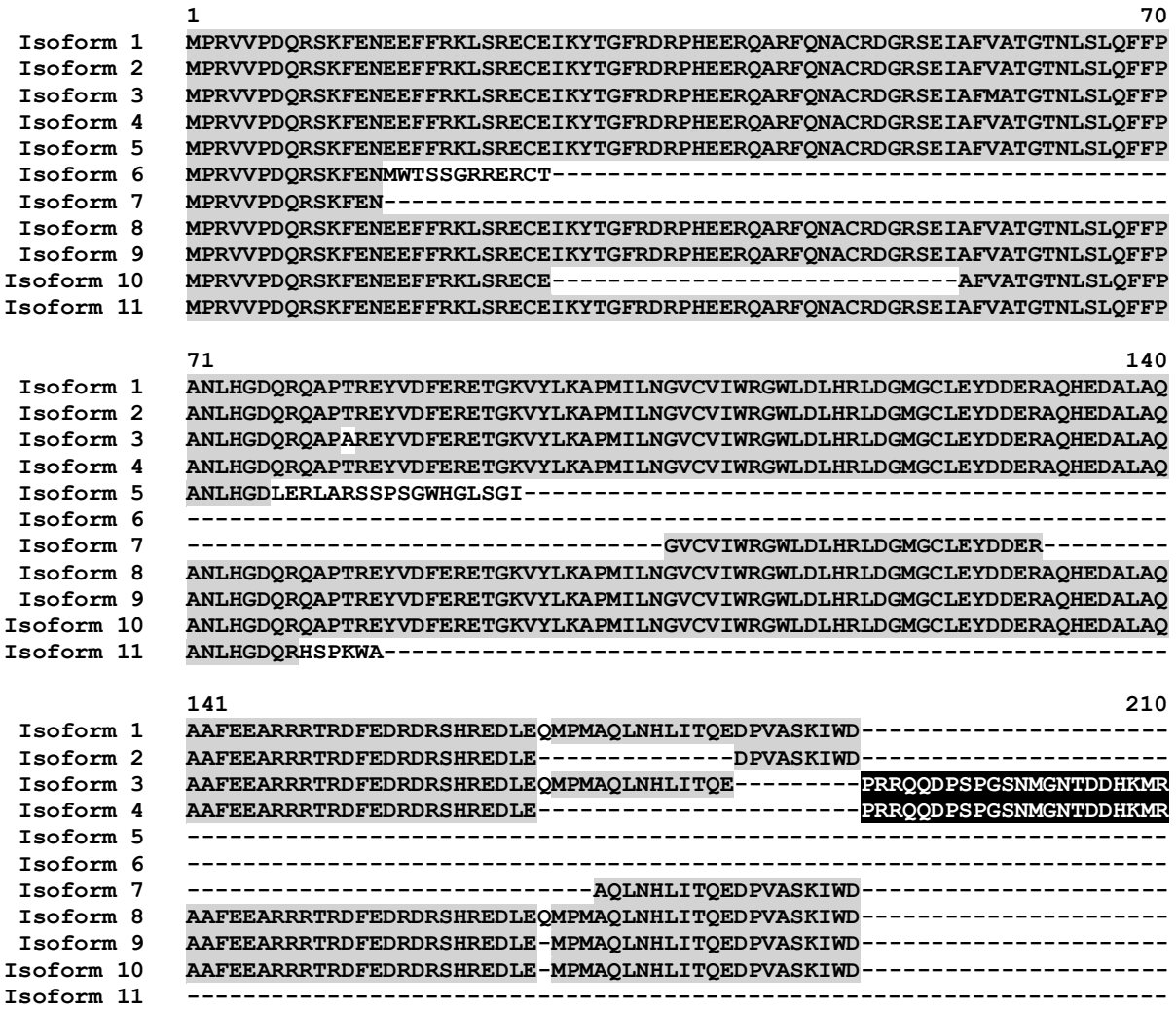


Figure 4.2 Alignment analysis of zebrafish Cbfβ protein isoform sequences. Cbfβ amino acid sequences were analysed using AlignX. Isoforms 5 and 6 show premature stop codons due to alternative splicing. Isoform 3 lacks exon 5b and isoform 4 lacks exons 5a and 5b, presenting a different C-terminal (white letters in black) with the occurrence of the stop codon in exon 6. Numbering is according to the first residue of the protein.

4.4.3 Chromosomal localization and structural organization of the zebrafish *cbfβ* gene and cDNA

Chromosomal assignment of the zebrafish *cbfβ* gene was performed by BLAST against NCBI database. The zebrafish major *cbfβ* transcript cloned (isoform 1) was aligned with the zebrafish genomic sequence, and sites of exon-intron borders were deduced by comparison. The zebrafish *cbfβ* gene was found on chromosome 18 (position 22774824-22852021) with a

length of approximately 77.200 kb, and based on the data in this study it is organized in 8 exons and 7 introns (**Figure 4.3**).

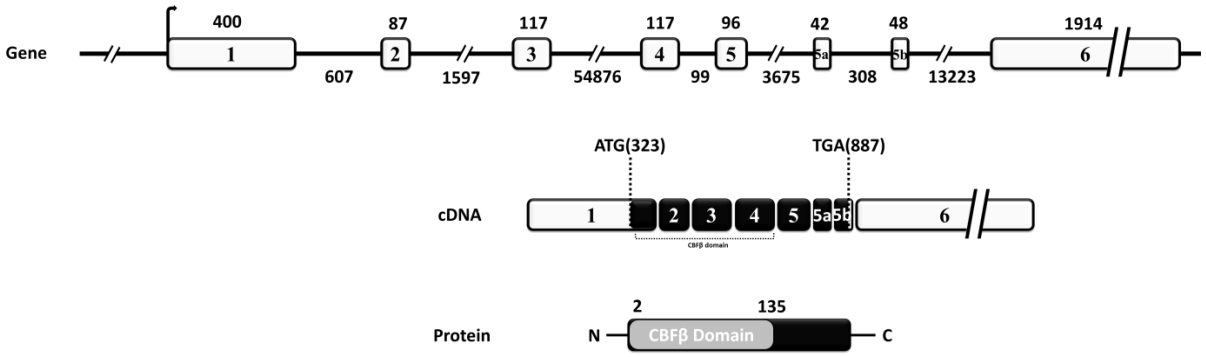


Figure 4.3 Schematic representation of zebrafish *cbfβ* gene, isoform 1 and protein structures. In gene structure: exons and introns are represented by boxes and lines, respectively. Numbers (in bp) above the boxes indicate size of the exons and numbers below the lines indicate size of introns. In transcript structure: black boxes represent the coding exons and white boxes the 5' and 3' untranslated regions; in protein structure: CBFβ heterodimerization domain is represented by a light grey box.

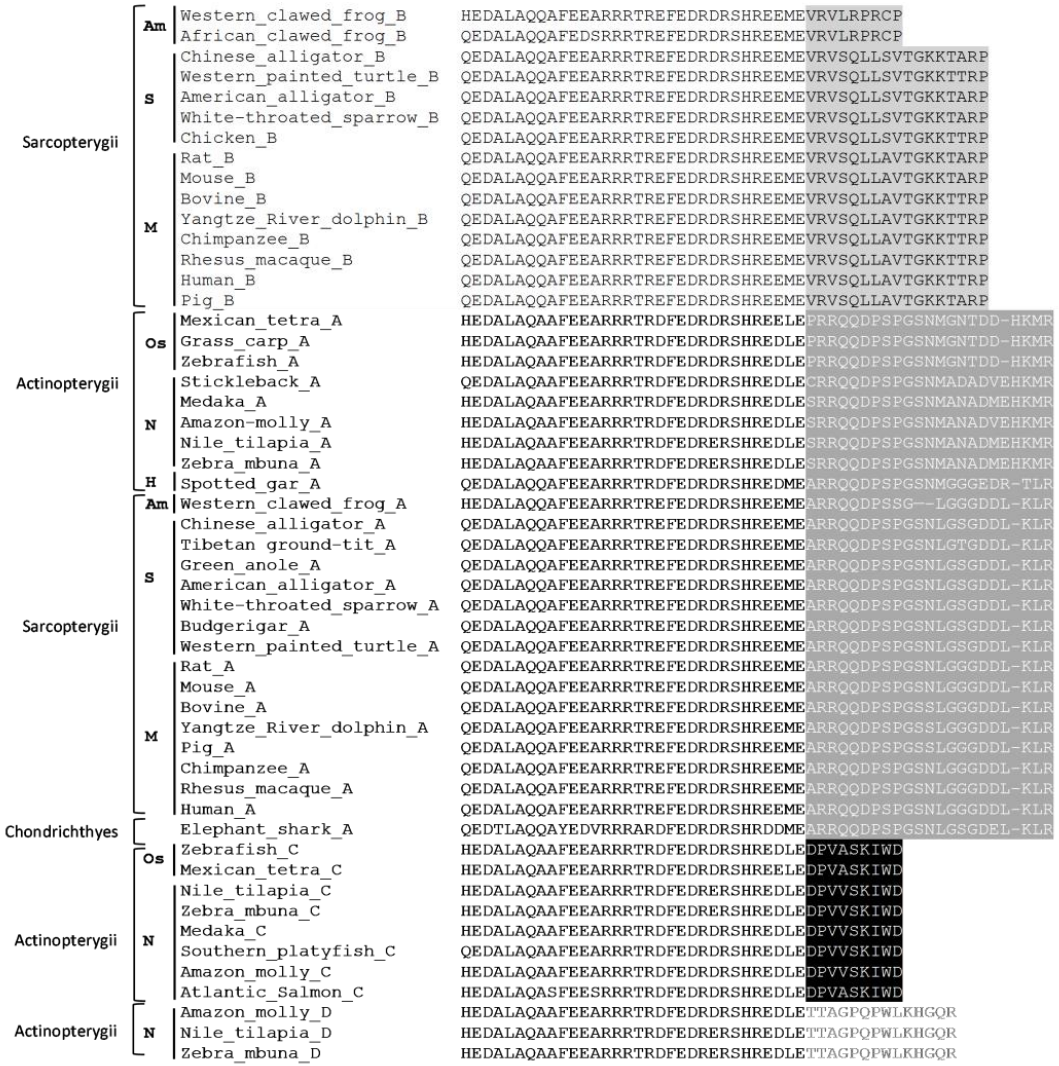
All splice junctions follow the GT/AG rule (Breathnach and Chambon, 1981). The zebrafish *cbfβ* isoform 1 contains all eight exons (exons 1-6, including exons 5a and 5b) with the start codon in exon 1 and the termination codon in exon 5b, and exon 6 contains the 3´ UTR. The protein deduced from this major Cbfβ isoform is 188 amino acids long. It contains a heterodimerization domain of 135 amino acids starting with the first methionine, and spanning sequences from exon 1 through exon 4.

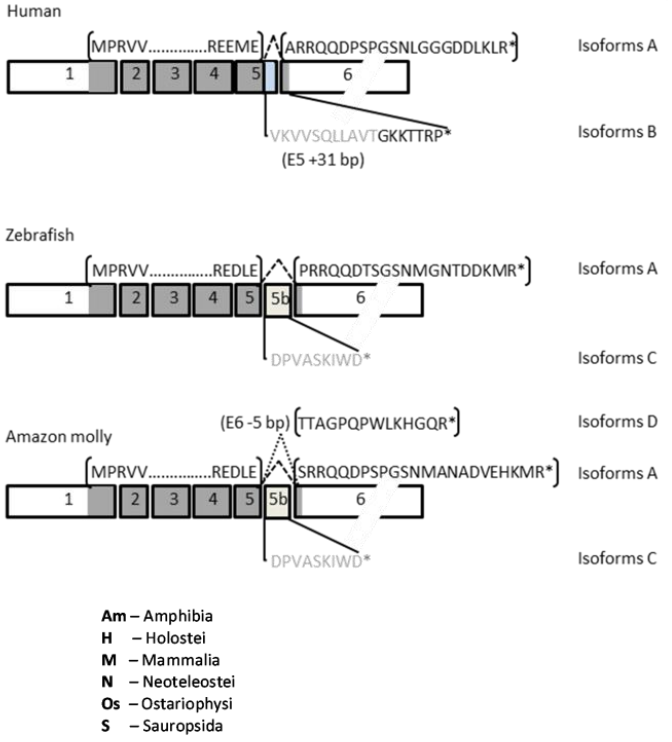
4.4.4 Protein sequence alignment between zebrafish and orthologs

Sequence databases at NCBI (www.ncbi.nlm.nih.gov) were searched for annotated CBFβ sequences. A total of 59 CBFβ sequences (containing the complete coding sequence) were collected. The full collection of sequences represents 29 species, including most classes of vertebrates (mammalia, sauropsida, amphibia, chondrochthyes and actinopterygii). Although this analysis was performed using sequences from a large set of organisms with diverse evolutionary pathways, CBFβ alignment revealed a remarkable conservation of protein primary structure **(Figure 4.4**), confirming the existence of domains in the protein

that are highly conserved. Interestingly, we found four different protein isoforms (labelled A-D) that differ only in the C-terminal region (**Figure 4.4**) that result from alternative splicing at the 3´-end. Zebrafish Cbfβ_A (isoform 4) is highly conserved between all vertebrate CBFβ_A (isoform 187) used in this alignment, containing exons 1-6 (excluding exons 5a and 5b).

Figure 4.4 Protein sequences comparison for CBFβ C-terminal. Sequences were aligned using Clustal Omega. The different C-terminal sequences are grouped and shown in different tons of grey to black. GenBank and Ensembl accession numbers for CBFβ: NP_074036.1 and NP_001746.1 (human A and B, respectively; *Homo sapiens*); JAA28496.1 and JAA42562.1 (chimpanzee A and B, respectively; *Pan troglodytes*); AFE80636.1 and AFH29554.1 (rhesus macaque A and B, respectively; *Macaca mulata*); DAA20211.1 and DAA20210.1 (bovine A and B, respectively; *Bos Taurus*); JAA74282.1 and JAA74187.1 (pig A and B, respectively; *Sus scrofa*); NP_071704.3 and NP_001154930.1 (mouse A and B, respectively, *Mus musculus*); AAH40752.2 and AAH81946.1 (rat A and B, respectively; *Rattus norvegicus*); XP_007457364.1 and XP_007457365.1 (Yangtze River dolphin A and B, respectively; *Lipotes vexillifer*); XP_002937211.2 and XP_004913586.1 (Western clawed frog A and B, respectively; *Xenopus tropicalis*); AFH75431.1 (grass carp; *Ctenopharyngodon idella*); AAI62159.1 and KF709194 (zebrafish A and C, respectively; *Danio rerio*); ABA42830.1 (Atlantic salmon; *Salmo salar*); NP_001087047.1 (African clawed frog; *Xenopus laevis*); NP_989901.2 (chicken; *Gallus gallus*); ENSAMXT00000021049 and XP_007256271.1 (Mexican tetra A and C, respectively; *Astyanax mexicanus*); ENSORLT00000017254 and ENSORLT00000017256 (medaka A and C, respectively; *Oryzias latipes*); ENSGACT00000018489 (Stickleback; *Gasterosteus aculeatus*); ENSONIT00000012829, XP_003447081.1 and XP_005471238.1 (Nile tilapia A, C and D, respectively; *Oreochromis niloticus*); XP_007553234.1, XP_007553236.1 and XP_007553235.1 (Amazon molly A, C and D, respectively; *Poecilia formosa*), XP_005795843.1 (Southern platyfish; *Xiphophorus maculatus*), XP_004569219.1, XP_004569222.1 and XP_004569220.1 (Zebra mbuna A, C and D, respectively; *Maylandia zebra*), XP_007902879.1 (Elephant shark; *Callorhinchus milii*), XP_006019105.1 and XP_006019106.1 (Chinese alligator A and B, respectively; *Alligator sinensis*), XP_005306333.1 and XP_005306334.1 (Western painted turtle A and B, respectively; *Chrysemys picta bellii*), XP_006268225.1 and XP_006268226.1 (American alligator A and B, respectively; *Alligator mississippiensis*), XP_005490832.1 and XP_005490833.1 (white-throated sparrow A and B, respectively; *Zonotrichia albicollis*), XP_005526382.1 (Tibetan ground-tit; *Pseudopodoces humilis*), XP_006641568.1 (spotted gar; *Lepisosteus oculatus*) and XP_005152308.1 (budgerigar; *Melopsittacus undulatus*).


Sarcopterygii
Actinopterygii
Chondrichthyes
Actinopterygii
Actinopterygii
Am
S
M
Os
N
H
Am
S
M
Os
N
N
Western_clawed_frog_B HEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVLRPRCP
African_clawed_frog_B QEDALAQQAFEDSRRRTREFEDRDRSHREEMEVRVLRPRCP
Chinese_alligator_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTARP
Western_painted_turtle_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTTRP
American_alligator_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTARP
White-throated_sparrow_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTARP
Chicken_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTTRP
Rat_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTARP
Mouse_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTARP
Bovine_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Yangtze_River_dolphin_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Chimpanzee_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Rhesus_macaque_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Human_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Pig_B QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTARP
Mexican_tetra_A HEDALAQAAFEEARRRTRDFEDRDRSHREELEPRRQQDPSPGSNMGNTDD-HKMR
Grass_carp_A HEDALAQAAFEEARRRTRDFEDRDRSHREDLEPRRQQDPSPGSNMGNTDD-HKMR
Zebrafish_A HEDALAQAAFEEARRRTRDFEDRDRSHREDLEPRRQQDPSPGSNMGNTDD-HKMR
Stickleback_A QEDALAQAAFEEARRRTRDFEDRDRSHREDLECRRQQDPSPGSNMADADVEHKMR
Medaka_A HEDALAQAAFEEARRRTRDFEDRDRSHREDLESRRQQDPSPGSNMANADMEHKMR
Amazon-molly_A HEDALAQAAFEEARRRTRDFEDRDRSHREDLESRRQQDPSPGSNMANADVEHKMR
Nile_tilapia_A HEDALAQAAFEEARRRTRDFEDRERSHREDLESRRQQDPSPGSNMANADMEHKMR
Zebra_mbuna_A HEDALAQAAFEEARRRTRDFEDRERSHREDLESRRQQDPSPGSNMANADMEHKMR
Spotted_gar_A QEDALAQQAFEEARRRTRDFEDRDRSHREDMEARRQQDPSPGSNMGGGEDR-TLR
Western_clawed_frog_A HEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSSG--LGGGDDL-KLR
Chinese_alligator_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Tibetan_ground-tit_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGTGDDL-KLR
Green_anole_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
American_alligator_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
White-throated_sparrow_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Budgerigar_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Western_painted_turtle_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Rat_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Mouse_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Bovine_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSSLGGGDDL-KLR
Yangtze_River_dolphin_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSSLGGGDDL-KLR
Pig_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSSLGGGDDL-KLR
Chimpanzee_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Rhesus_macaque_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Human_A QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Elephant_shark_A QEDTLAQQAYEDVRRRARDFEDRDRSHRDDMEARRQQDPSPGSNLGSGDEL-KLR
Zebrafish_C HEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVASKIWD
Mexican_tetra_C HEDALAQAAFEEARRRTRDFEDRDRSHREELEDPVASKIWD
Nile_tilapia_C HEDALAQAAFEEARRRTRDFEDRERSHREDLEDPVVSKIWD
Zebra_mbuna_C HEDALAQAAFEEARRRTRDFEDRERSHREDLEDPVVSKIWD
Medaka_C HEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVVSKIWD
Southern_platyfish_C QEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVVSKIWD
Amazon_molly_C HEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVVSKIWD
Atlantic_Salmon_C HEDALAQASFEESRRRTRDFEDRDRSHREDLEDPVASKIWD
Amazon_molly_D HEDALAQAAFEEARRRTRDFEDRDRSHREDLETTAGPQPWLKHGQR
Nile_tilapia_D HEDALAQAAFEEARRRTRDFEDRERSHREDLETTAGPQPWLKHGQR
Zebra_mbuna_D HEDALAQAAFEEARRRTRDFEDRERSHREDLETTAGPQPWLKHGQR


Human
MPRVV.....REEME
ARRQQDPSPGSNLGGGDDLKLR*
Isoforms A
1 2 3 4 5 6
VKVVSQLLAVTGKKTTRP*
(E5 +31 bp)
Isoforms B
Zebrafish
MPRVV.....REDLE
PRRQQDTSGSNMGNTDDKMR*
Isoforms A
1 2 3 4 5 5b 6
DPVASKIWD*
Isoforms C
Amazon molly
(E6 -5 bp)
TTAGPQPWLKHGQR*
Isoforms D
MPRVV.....REDLE
SRRQQDPSPGSNMANADVEHKMR*
Isoforms A
1 2 3 4 5 5b 6
DPVASKIWD*
Isoforms C
Am – Amphibia
H – Holostei
M – Mammalia
N – Neoteleostei
Os – Ostariophysi
S – Sauropsida

In contrast, the C-terminal of the zebrafish Cbfβ_C (isoform 2) shows high homology with the C-terminal of Cbfβ from other fish (all neoteleostei: Atlantic salmon, tilapia, Mexican tetra, zebra mbuna, Southern platyfish, Amazon molly, Atlantic salmon, medaka, Atlantic cod, turquoise killifish, Burton´s mouthbrooder, red Mwanza and lyretail cichlid (**Figure 4.4** and results not shown). The residues from 166 to 174 in Cbfβ_C were encoded by exon 5b, that contains the stop codon. We also observed the presence of a C-terminal that is different from the named Cbfβ_A or Cbfβ_C that we called Cbfβ_B and was only found in Sarcopterygii, which is obtained from a long exon 5 (more 31 bp in the 3´-end) that ends in the exon 6 coding for the seven last amino acids and the stop codon. A fourth variant named Cbfβ_D has been identified in fish (all neoteleostei: Amazon molly, Nile tilapia, zebra mbuna, Burton´s mouthbrooder and red Mwanza) (**Figure 4.4** and results not shown), that results from the occurrence of an alternative splicing from exon 5 to a cryptic site in exon 6. The transcription of exons 1, 2, 3 and 4 does not undergo any modifications and remains constant. We have calculated the pair-wise percentage identities among all CBFβ protein sequences used in this study, and we can observe a high identity between all the species (**Table 4.S1**), even if we take in account the C-terminal differences observed in the alignment.

4.4.5 Conserved gene synteny of zebrafish *cbfβ* gene

Synteny-based analysis of zebrafish *cbfβ* gene shows strong syntenic conservation between human chromosome 16 and zebrafish chromosome 18. In both cases, the genes *DNAJA2L*, *BBS2*, *GOT2*, *CCDC79*, *PDP2*, *CES2*, *CES3, B3GNT9*, *HSF4*, *PARD6A*, *TSNAXIP1*, *NUTF2*, *EDC4*, *PSKH1*, *NRN1LA*, *CIRH1A*, *AARS*, *TAT*, *BCAR1*, *NOC4* and *KIAA1049* were found in the region of *CBFβ* gene, but they appear in a different order (**Figure 4.5** and **Table 4.S2**). Interestingly, from this list of genes only *noc4* and *kiaa1049* are present downstream side of *cbfβ* gene in zebrafish chromosome 18 (**Figure 4.5**). This syntenic conservation supports the identification of *cbfβ* as ortholog to human *CBFβ*.

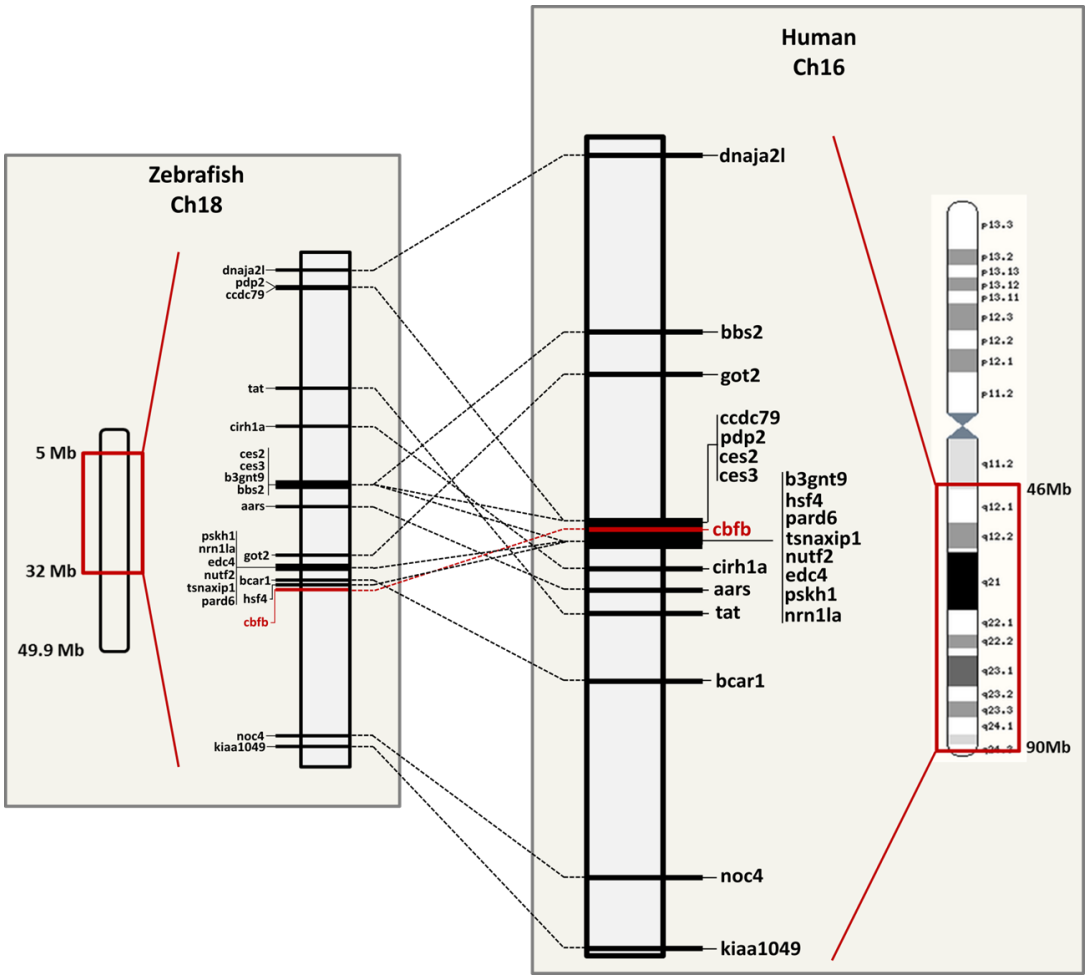


Figure 4.5 Comparison of genomic environment and gene positional order in zebrafish and human chromosomes containing CBFβ. Comparison of the chromosomal locations of 22 ortholog gene pairs between zebrafish chromosome 18 and human chromosome 16. Lines between the compared chromosomes connect positions of ortholog gene pairs in the two species. Distances between markers on a single chromosome are shown to scale, but compared chromosomes have been scaled to equivalent lengths. Map positions for the genes were obtained from http://www.ensembl.org/.

4.4.6 Expression profiles of zebrafish *cbfβ* variants

Expression patterns of *cbfβ* mRNA variants were analyzed using RT-PCR with gene-specific primers located on ORF exons and variant-specific primers located on respective leader/terminal exons (**Figure 4.6**). To assess the expression of the zebrafish *cbfβ* transcript variants, the coding region between exons 5 and 6, comprising the full and the alternatively splicing exons 5a and 5b were amplified (**Figure 4.6A**). The *cbfβ* mRNA was widely distributed. In a first PCR round the amplification of all four possible splicing variants

(isoforms 1-4) was tested using the primers CBFβ_F3 and CBFβ_R3 located in exon 5 and exon 6, respectively. Two amplicons corresponding to isoform 1 (199 bp) and isoforms 2 and/or 3 (157 bp and 151 bp, respectively) were observed in all samples tested (**Figure 4.6B**). A third amplicon corresponding to isoform 4 (109 bp) was observed in all tissue samples and developmental stages tested, except at 1 cell stage. In order to distinguish the expression of isoform 2 and 3, a second amplification was performed with isoform specific primers. The expression of isoform 2 mRNA was analyzed using a forward primer located on the exon 5 and a reverse primer on exon 5b (**Figure 4.6A**).

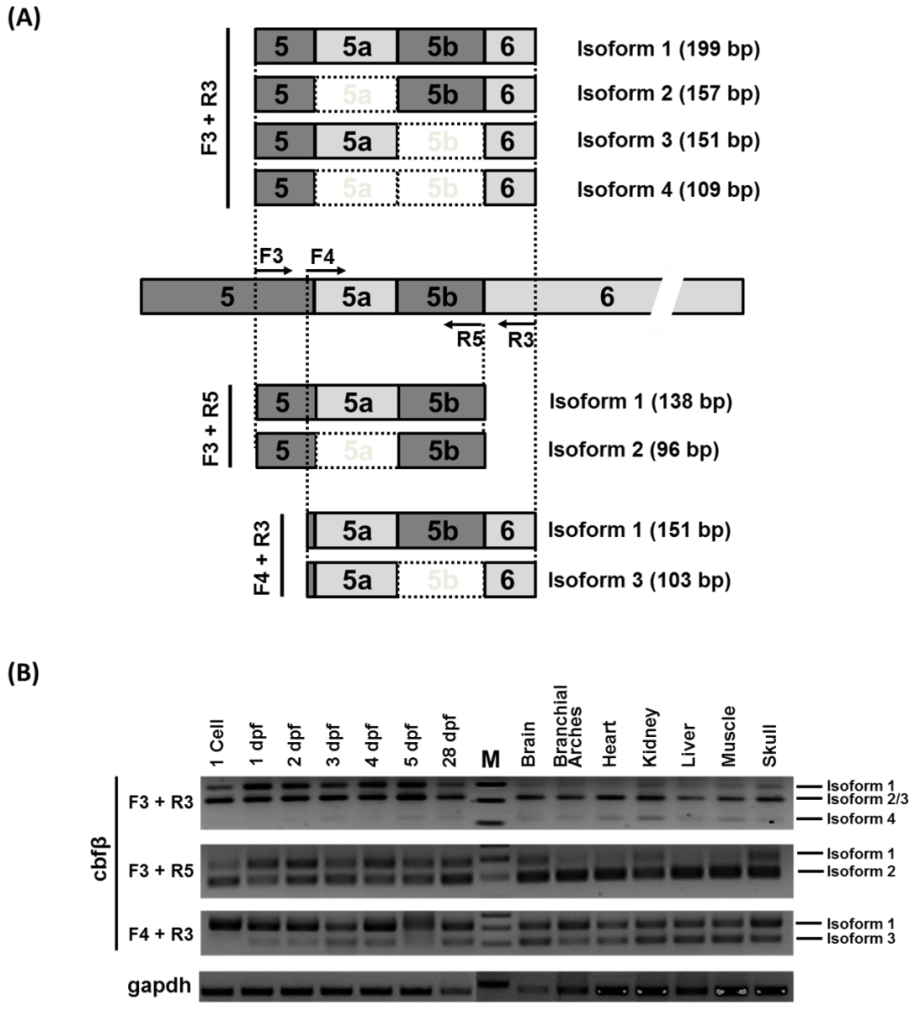


Figure 4.6 Identification of the expression profile of zebrafish cbfβ splicing variants (isoform 1-4). (**A**) Schematic representation of partial RNA structure and PCR products resulting from each amplification. Dotted boxes with white background correspond to spliced exons. The pair of primers used for amplification and sizing of the resulting products are represented (in the left and right side of scheme, respectively) (**B**) Qualitative expression profile of the cbfβ isoforms (1-4) investigated by RT-PCR in zebrafish developmental stages and adult tissues. Zebrafish *gapdh* was used as control for sample integrity. Sample designations are indicated above and primer pairs used are indicated in the left side. M corresponds to the marker (Thermo Scientific GeneRuler 50 bp DNA Ladder).

This amplification generates two amplicons corresponding to isoform 1 (138 bp), and isoform 2 (96 bp). It was observed that the isoform 2 is expressed in all the developmental stages and tissues analysed (**Figure 4.6B**). The expression of isoform 3 mRNA was analysed using a forward primer located on the frontier of exons 5/5a and a reverse primer on exon 6, resulting in two amplicons corresponding to isoform 1 (151 bp) and isoform 3 (103 bp). The expression of isoform 3 was observed in all the developmental stages and tissues analysed with the exception at 1 cell stage where the corresponding amplicon was not observed (**Figure 4.6B**).

4.4.7 Functional analysis of the different *cbfβ* splicing variants

Given that CBFβ is a transcription co-factor, and is able to bind mammalian CBFα proteins and enhance their DNA binding affinity (Wang et al, 1993), we wanted to test if the zebrafish Cbfβ protein isoforms cloned in this work had a similar function. The newly identified isoforms 1-3, and also the isoform 4 that corresponds to the one previously characterized (AF278758) were cloned in an expression vector and used in co-transfection assays with a fragment of the zebrafish *col10α1* promoter described previously (Simões et al, 2006). This promoter was previously reported to be regulated by the Runx2 transcription factor (zebrafish isoform MASN-Runx2) (Simões et al, 2006). To this end, HEK293 cells were transiently co-transfected with the pTATALuC-4×ColX(-822/-794) vector containing four repeated copies of the Runx2 binding site, in the presence of expression vectors containing MASN-Runx2 and the zebrafish *cbfβ* isoforms 1 to 4. Our previous studies showed that in the *Xenopus laevis* A6 cell line the transcriptional activity of the pTATALuCColX(-822/-794) construct is induced by MASN-Runx2 isoform, and a further increase was observed when four copies of this sequence element were present (Simões et al, 2006). In the present work we showed in HEK293 cells that the ability of MASN-Runx2 to transactivate the 4×ColX construct, although smaller than previously seen in A6 cells, was strongly stimulated when Cbfβ isoforms 1 or 2 were co-expressed (**Figure 4.7**). Furthermore, Cbfβ isoforms 3 and 4 seem to have lost the ability to regulate Runx2 (**Figure 4.7**). All together, these results clearly indicate that the presence of the different amino acids in the C-terminal of the Cbfβ that are generated by the presence of the exon 5b, are likely to be essential for protein binding to

Runx2-MASN isoform and so to enhance Runx2-induced transcription. We also show that Cbfβ alone has no effect on 4×ColX transcription (**Figure 4.7**).

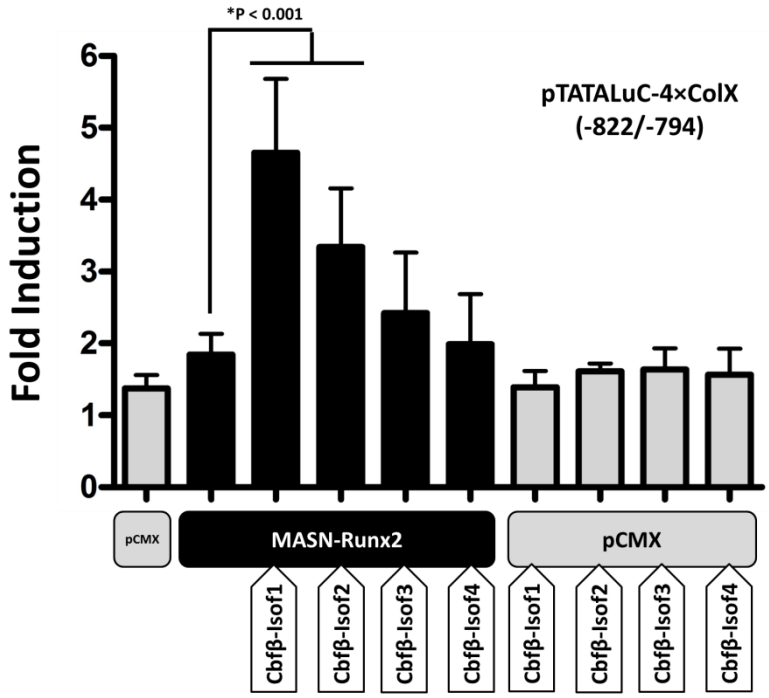


Figure 4.7 Transcriptional co-activation of collagen type X promoter by Runx2-MASN/Cbfβ. HEK 293 cells were transfected with zebrafish pTATALuC-4×ColX(-822/-794) promoter construct, a reporter plasmid derived from the *colXα1* promoter that contains four copies of putative Runx-binding site. Cells were cotransfected with the indicated Cbfβ (isoforms 1-4) expression plasmids in the presence of zebrafish Runx2-MASN isoform. The graph shows the fold induction expression of *colX* promoter construct, alone or co-transfected with Runx2 and/or Cbfβ. The data indicated is a representative plot that shows the average and standard deviation (error bars) from at least three independent experiments, each done in duplicate. Significance was determined by One Way Anova. Asterisk (*) indicates that the value is statistically different ($p<0.001$).

4.4.8 Co-immunoprecipitation of Cbfβ splicing variants and runx2

To assess the heteromeric assembly of zebrafish Cbfβ protein isoforms 1-4 and Runx2 by an independent biochemical approach, co-immunoprecipitation experiments were performed. Protein lysates prepared from HEK293 cells expressing HA-tagged Cbfβ isoforms 1-4 alone or together with Flag-tagged runx2 were immunoprecipitated with an anti-Flag monoclonal antibody. Immunoprecipitates were subjected to SDS-gel electrophoresis and probed with anti-HA and anti-FLAG antibodies to visualize HA-Cbfβ isoforms 1-4 and Flag-Runx2 (**Figure**

4.S1). Flag-Runx2 was specifically co-immunoprecipitated with HA-tagged Cbfβ isoforms 1, 2 and 4, but not HA-Cbfβ isoform 3 (**Figure 4.S2**). These experiments clearly indicated that Cbfβ isoforms 1, 2 and 4 are present in protein complexes with Runx2 in HEK293 cells, suggesting an interaction between these isoforms and Runx2, while the isoform 3 of Cbfβ failed to interact with Runx2 under these conditions.

4.5 Discussion

In this study we describe 11 different spliced variants of zebrafish *cbfβ* mRNA (including the one previously known (Blake et al, 2000) corresponding to our isoform 4, Δ5aΔ5b). These 10 novel spliced variants greatly expand our knowledge of the isoforms of *cbfβ* at the level of mRNA in zebrafish and provide evidence for a conserved structure and splicing events between zebrafish and human *CBFβ* genes. Alternative pre-mRNA splicing plays an important role in regulating gene expression by generating multiple transcripts from a single gene with specific spatial and temporal patterns, thus contributing to generate proteome diversity and increasing flexibility for gene expression and regulation (Graveley, 2001; Black, 2003). Nonetheless, much remains to be understood about the mechanisms and functional significance of this process. The *CBFβ* gene encodes a transcription factor (CBFβ) that plays important roles in hematopoiesis, osteogenesis and leukemia (Liu et al, 1995; Speck et al, 1999; Miller et al, 2002). The biological relevance of CBFβ has been demonstrated in a knock-out mouse model that exhibits embryonic lethality due to defective fetal liver hematopoiesis and central nervous system bleeding, recapitulating the *Runx1* null phenotype (Sasaki et al, 1996; Wang et al, 1996). Conversely, heterozygous *Cbfβ*$^{+/-}$ knock-in mice survive gestation but die soon after birth with bone developmental defects comparable to those observed in *Runx2*$^{-/-}$ mice although less severe (Kundu et al, 2002). In zebrafish, *cbfβ* is expressed during embryogenesis in early hematopoietic cells and in the lateral plate mesoderm at tail bud stage, as well as in Rohon-Beard cells, cranial nerve ganglia, hindbrain, retina, branchial arches, jaw, and fin buds (Blake et al, 2000). Recently it was shown that zebrafish *cbfβ* knockout mutants (*cbfβ*$^{-/-}$) retained primitive hematopoiesis and erythro-myeloid progenitors but completely lacked all definitive blood lineages (Bresciani et al, 2014), confirming the importance of Cbfβ in the onset of definitive hematopoiesis. Our RT-

PCR analysis in zebrafish developmental stages and adult tissues shows that *cbfβ* is widely expressed, been detected in all samples analyzed. This is in agreement with a previous study (Blake et al, 2000) where they show by Northern blot hybridization that *cbfβ* expression is first detected at 3-somite stage and then continued through to at least 48 hpf and also in an adult sample. Our gene expression profile data demonstrate that at 1 cell stage just the *cbfβ* isoforms 1 and 2 are detected, but not isoforms 3 and 4 (**Figure 4.6B**). The fact that isoforms 1 and 2 are detected at 1 cell stage indicates that they are maternally inherited, in contrast to isoforms 3 and 4 that are not expressed at this time, emphasizing that the biological function of Cbfβ splice variants should be further evaluated throughout development. Blake and co-workers (2000) also showed that *cbfβ* is expressed in the kidney as they used a kidney cDNA library to clone the transcript. Our data shows that *cbfβ* expression persists in adult, as we could detect all four transcript variants (isoforms 1-4) in all the tissues analyzed (**Figure 4.6B**).

Translated variants of such an important mRNA species may have important modulatory functions in development or in critical cell fate decisions, although some of these isoforms may not be translated due to the process of nonsense-mediated mRNA decay (NMD) that promotes degradation of mRNAs containing premature translation termination codons. This process was identified and studied also in zebrafish, and shown to be essential for zebrafish embryonic development, preventing accumulation of potentially detrimental truncated proteins (Wittkopp et al, 2009). Two of the transcript variants described in this report present premature termination codons (isoforms 5 and 6; **Figure 4.1**), and thus may be potential targets for the NMD pathway, and not likely to be translated into protein.

In human and mice, *CBFβ* resides on chromosomes 16 and 8, respectively, and both species show two major isoforms resulting from distinct alternative splicing events that produce, in each case, a frame-shift generating a termination codon so that the two proteins (of 187 and 182 amino acids, respectively) differ in several amino acids at the carboxy terminus (Adya et al, 2000; Ogawa et al, 1993) (**Figure 4.S3**). A search of the human dbEST and non-redundant data bases identified three more exons in the human *CBFβ* gene (**Figure 4.S4**), giving a gene structure of nine exons whose alternative splicing creates ten human *CBFβ* isoforms.

Multiple alignments between major CBFβ isoforms described in different vertebrates (**Figure 4.4 and Figure 4.S5**), show that zebrafish Cbfβ_A (isoform 4) is highly conserved in all species analysed (CBFβ_A isoform containing 187 aa). In contrast, the C-terminal of the zebrafish Cbfβ_C (isoform 2) shows high homology with the C-terminal of Cbfβ from other fish (e.g. Atlantic salmon, tilapia, and medaka isoform_C) but differs from the C-terminal of the other vertebrates CBFβ_B (isoform containing 182 aa) used in the alignment. This divergence in the C-terminal between the different species may indicate that this region has a functional relevance that could be species specific, possibly mediating interactions with different proteins from the CBF regulatory complex. Different groups (Wang et al, 1996; Kagoshima et al, 1996; Zhou et al, 2012; Du et al, 2013) have studied the CBFβ binding capacity to Runx co-factors throughout the heterodimerization domain (N-terminal region), but the exact function of the C-terminal region of the CBFβ isoforms is still unknown at this time. Interestingly, an association between breast cancer and mutations in the heterodimerization domain of *CBFβ* were previously reported (Banerji et al, 2012; Taniuchi et al, 2012; Ellis et al, 2012). Accordingly, all these *CBFβ* genetic changes are likely to result in loss-of-function mutants. Oncogenic rearrangements of *CBFβ* are common in acute myeloid leukemia where the CBFB–MYH11 translocation produces a protein product that fuses the first 165 aa of CBFβ to the MYH11 resulting in a hybrid molecule believed to have dominant negative function (Shigesada et al, 2004).

It was previously shown that Cbfβ interacts with Runx2 in bone and cartilage and enhances Runx2-mediated transcription (Kundu et al, 2002; Yoshida et al, 2002; Nakashima and Crombrugghe, 2003; Kanatani et al, 2006; Han et al, 2010). Higashikawa et al (2009) showed that human *COL10A1* promoter activity, which was enhanced by RUNX2, was further potentiated by RUNX2 in combination with the co-activator CBFβ. The same was observed with the osteocalcin promoter (Kanatani et al, 2006). According to previous studies, the C-terminal amino acids that are different between the two major CBFβ isoforms are in a region of the protein that is not required for the heterodimerization with the RUNX partner (Ogawa et al, 1993; Kagoshima et al, 1996) and so it was suggested that the amino acid differences in this region are not expected to affect the ability of the α/β subunits to heterodimerize (Blake et al, 2000). From the spliced variants cloned in this work, four of them seem to be potentially interesting from a functional point of view: isoforms 1-4 (the

complete, Δ5a, Δ5b and Δ5aΔ5b isoforms, respectively), and so their capacity for transcription transactivation was further analysed. Our co-transfection experiments demonstrate that zebrafish Cbfβ isoforms carrying the exon 5b (isoforms 1 and 2) have a higher capacity to enhance the induction of *ColX* promoter by Runx2-MASN isoform, compared to the isoforms lacking exon 5b (isoforms 3 and 4) (**Figure 4.7**). Immunoprecipitation data allowed us to explain the transactivation data by the direct interaction of Runx2 with Cbfβ isoforms 1 and 2 and not with Cbfβ isoform 3. However, an interaction was also observed between Runx2 and Cbfβ isoform 4, although this interaction does not result in a Runx2 stimulated transcription of *ColX* promoter in the conditions tested. The differences between these four Cbfβ isoforms reside in their C-terminal region (**Figure 4.2**). Isoforms 1 and 3 have distinct C-terminal sequences, while isoforms 2 and 4 represent spliced variants of isoforms 1 and 3, respectively. These results suggest that isoforms 1 and 2 have a functional motif that is lacking in isoforms 3 and 4, likely located in exon 5b. Alternatively, the distinct C-terminal domain found in isoforms 3 and 4, (**Figure 4.2**) may be interfering with its binding to the Runx2 protein by either affecting the stabilization of the heterodimer, enabling the binding of some other co-factor still not identified and important to the function of the CBF complex, or affecting its translocation to the nucleus, which is required for acting as a co-factor of Runx2.

Zebrafish Cbfβ (isoform 4; Δ5aΔ5b) has previously been shown to induce the human CBFα2 (RUNX1-MRIPV isoform) as efficiently as the human CBFβ protein (isoform 187) (Blake et al, 2002). In contrast with these findings, our results show no significant enhancement of runx2-MASN transcriptional activity in the *ColXα1* promoter fragment when co-transfected with the Cbfβ isoform 4 (**Figure 4.7**). This apparent discrepancy may indicate that the different CBFα subunits (runx1, 2 and 3) have distinct affinities for the different Cbfβ isoforms. In fact, it was shown that in mammals the CBFβ (isoform 187) and CBFβ (isoform 182) interact with RUNX1 similarly, although CBFβ (isoform 187) in conjunction with RUNX1 transactivates *SL3-3MLV* enhancer more strongly (Zaiman et al, 1995). In addition, and as suggested previously, CBFβ proteins apart from their well-known function as co-factors of RUNX associated DNA-binding affinity, may have additional functions such as, (i) when bound to the runt domain, CBFβ proteins may induce a conformational change allowing it to interact with other transcriptional activators or (ii) it can act as an interacting factor between

RUNX proteins and other protein cofactors (Adya et al, 2000; Li and Gergen, 1999). Altogether it seems that CBFβ isoforms function can be modulated by the RUNX isoform present and thus also depends on the cell type used in each study (Adya et al, 1998). Higashikawa et al (2009) showed that the effect of RUNX2 in human *COL10A1* promoter activity observed in human cells were not reproducible in the mouse chondrogenic ATDC5 cells, in which neither RUNX2 alone nor in combination with CBFβ affected *COL10A1* promoter activity. Indeed, more recently Du et al (2013) showed that when HEK293T cells were co-transfected with the C-terminal-truncated CBFβ constructs and the viral infectivity factor (Vif) of HIV-1 (Vif-expressing vector) following repression of endogenous expression of CBFβ by an shRNA approach, Vif expression appeared quite variable, depending on the co-transfected CBFβ variant. The authors concluded that different lengths of CBFβ are required for its role in Vif function and for its role in RUNX-mediated gene transcription and hypothesized that different CBFβ domains may be required for regulation of different target genes (Du et al, 2013). It is also possible that still another co-activator, as yet unidentified, may be involved in this process of transactivation but further studies are required to clarify the precise mechanism of this phenomenon.

Overall, in this work we have cloned and described for the first time a variety of zebrafish *cbfβ* alternative spliced variants. Using a bioinformatic approach we have determined the structures of both the zebrafish *cbfβ* gene and predicted protein products and shown a high degree of sequence identity between zebrafish Cbfβ and the mammalian CBFβ proteins, indicating conserved functions. Using luciferase assays, we showed that the Runx2-MASN mediated activation of the Col10α1 promoter is differentially co-activated by Cbfβ isoforms, although further work will be needed to clarify the significance of the biological function of these *cbfβ* variants.

Acknowledgments

This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Program and national funds through FCT – Foundation for Science and Technology, under the project “PEst-C/MAR/LA0015/2011. NC and BS were supported, respectively, by a post-doctoral and doctoral grant from FCT (SFRH/BPD/48206/2008 and SFRH/BD/38083/2007).

4.6 Supplementary Tables

Table 4.S1 Pairwise per cent identities among CBFβ sequences. A - From light grey to black: actinopterygii, chondrichthyes, sarcopterygii (amphibia, sauropsida, mammalia). Hsa, *Homo sapiens* (human); Ptr, *Pan troglodytes* (chimpanzee); Mmu, *Macaca mulata* (rhesus macaque); Ssc, *Sus scrofa* (pig); Bta, *Bos Taurus* (bovine); Mmus, *Mus musculus* (mouse); Rno, *Rattus norvegicus* (rat); Lve, *Lipotes vexillifer* (Yangtze River dolphin); Xtr, *Xenopus tropicalis* (Western clawed frog); Dre, *Danio rerio* (zebrafish); Ola, *Oryzias latipes* (medaka); Ame, *Astyanax mexicanus* (Mexican tetra); Phu *Pseudopodoces humilis* (Tibetan ground-tit); Mun, *Melopsittacus undulatus* (budgerigar); Gac, *Gasterosteus aculeatus* (three spined stickleback); Cid, *Ctenopharyngodon idella* (grass carp); Ssa, *Salmo salar* (Atlantic salmon); Oni, *Oreochromis niloticus* (Nile tilapia); Loc, *Lepisosteus oculatus* (Spotted gar); Mze *Maylandia zebra* (Zebra Mbuna); Pfo, *Poecilia formosa* (Amazon molly); Xma *Xiphophorus maculatus* (Southern platyfish); Aca, *Anolis carolinensis* (green anole); Cmi, *Callorhinchus milii* (elephant shark); Gga, *Gallus gallus* (chicken); Xla, *Xenopus laevis* (African clawed frog); Cpi, *Chrysemys picta belli* (Western painted turtle); Asi, *Alligator sinensis* (Chinese alligator); Zal, *Zonotrichia albicollis* (white-throated sparrow); Ami, *Alligator mississippiensis* (American alligator).

A

	Ame	Cid	Dre	Gac	Loc	Mze	Ola	Oni	Pfo	Cmi	Aca	Ami	Asi	Cpi	Mun	Phu	Zal	Xtr	Bta	Hsa	Lve	Mmu	Mmus	Ptr	Rno	Ssc
Ame																										
Cid	98																									
Dre	97	99																								
Gac	92	93	92																							
Loc	91	93	92	90																						
Mze	93	94	93	93	88																					
Ola	93	94	93	94	89	96																				
Oni	93	94	93	93	88	100	96																			
Pfo	94	95	94	95	90	97	97	97																		
Cmi	86	86	84	85	88	85	84	85	84																	
Aca	90	89	88	87	92	86	86	86	86	91																
Ami	90	89	88	87	92	86	86	86	86	91	100															
Asi	89	88	87	86	91	85	85	85	85	91	99	99														
Cpi	90	89	88	87	92	86	86	86	86	91	99	99	98													
Mun	90	89	88	87	92	86	86	86	86	91	100	100	99	99												
Phu	90	89	88	87	92	86	86	86	86	91	99	99	98	99	99											
Zal	90	89	88	87	92	86	86	86	86	91	100	100	99	99	100	99										
Xtr	87	86	85	81	87	82	81	82	82	84	91	91	90	91	91	91	91									
Bta	89	88	87	86	93	85	85	85	86	90	98	98	97	98	98	98	98	91								
Hsa	89	88	87	86	93	85	85	85	86	90	98	98	97	98	98	98	98	91	99							
Lve	89	88	87	86	92	85	84	85	86	89	97	97	96	98	97	97	97	91	99	99						
Mmu	89	88	87	86	93	85	85	85	86	90	98	98	97	98	98	98	98	91	99	100	99					
Mmus	88	87	87	85	91	84	84	84	85	89	97	97	96	97	97	97	97	90	98	99	98	99				
Ptr	89	88	87	86	93	85	85	85	86	90	98	98	97	98	98	98	98	91	99	100	99	100	99			
Rno	88	87	87	85	91	84	84	84	85	89	97	97	96	97	97	97	97	90	98	99	98	99	100	99		
Ssc	89	88	87	86	92	85	84	85	86	89	97	97	99	98	97	97	97	91	99	99	100	99	98	99	98	

B

	Ami	Asi	Cpi	Gga	Zal	Xla	Xtr	Bta	Hsa	Lve	Mmu	Mmus	Ptr	Rno	Ssc
Ami															
Asi	99														
Cpi	99	98													
Gga	99	98	99												
Zal	100	99	99	99											
Xla	86	85	86	85	86										
Xtr	86	85	86	86	86	94									
Bta	98	97	99	98	98	86	86								
Hsa	97	96	98	97	97	85	85	99							
Lve	97	96	98	97	97	85	85	99	100						
Mmu	97	96	98	96	97	85	85	99	100	100					
Mmus	97	96	97	97	97	84	84	98	98	98	98				
Ptr	97	96	98	97	97	85	85	99	100	100	100	98			
Rno	97	96	97	96	97	84	84	98	98	98	98	100	98		
Ssc	98	97	98	97	98	85	85	98	99	99	99	99	99	99	

C

	Ame	Dre	Mze	Ola	Oni	Pfo	Ssa	Xma
Ame								
Dre	97							
Mze	94	94						
Ola	95	95	97					
Oni	94	94	100	97				
Pfo	96	95	97	98	97			
Ssa	94	92	91	91	91	91		
Xma	95	95	97	97	97	99	91	

D

	Mze	Oni	Pfo
Mze			
Oni	100		
Pfo	97	97	

Table 4.S2 Zebrafish-human ortholog gene pairs.

Locus	LG	location	Human Gene ID	Human Chromosome	location	name
dnaja2l	18	5,943,008	ENSG00000069345	16	46,989,299	DnaJ (HSP40)homolog, subfamily A, member 2
pdp2	18	6,916,184	ENSG00000172840	16	66,912,492	Pyruvate dehydrogenase phosphatase isoenzyme 2
ccdc79	18	6,919,972	ENSG00000249961	16	66,788,879	Coiled-coil domain containing 79
tat	18	12,640,431	ENSG00000198650	16	71,599,563	Tyrosine aminotransferase
cirh1a	18	14,205,614	ENSG00000141076	16	69,165,194	Cirrhosis, autosomal recessive 1A (cirhin)
ces2	18	17,067,638	ENSG00000172831	16	66,968,347	Carboxylesterase 2
ces3	18	17,077,342	ENSG00000172828	16	66,995,140	Carboxylesterase 3
b3gnt9	18	17,171,163	ENSG00000237172	16	67,182,008	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 9
bbs2	18	17,175,737	ENSG00000125124	16	56,500,748	Barbet-Biedl syndrome 2
aars	18	18,416,984	ENSG00000090861	16	70,286,293	Alanyl-tRNA synthetase
got2	18	20,959,432	ENSG00000125166	16	58,741,035	Glutamic-oxaloacetic transaminase 2 mitochondrial (aspartate aminotransferase 2)
pskh1	18	21,508,905	ENSG00000159792	16	67,918,708	Protein serine kinase H1
nrn1la	18	21,564,830	ENSG00000188038	16	67,918,708	Neuritin 1-like a
edc4	18	21,641,228	ENSG00000038358	16	67,906,926	Enhancer of mRNA decapping 4
nutf2	18	21,694,396	ENSG00000102898	16	67,880,635	Nuclear transport factor 2
tsnaxip1	18	21,712,452	ENSG00000102904	16	67,840,668	Translin-associated factor x interacting protein 1
pard6a	18	21,809,385	ENSG00000102981	16	67,694,851	Par-6 partitioning defective 6 homolog alpha
bcar1	18	22,284,911	ENSG00000050820	16	75,262,928	Breast cancer anti-estrogen resistance 1
hsf4	18	22,523,979	ENSG00000102878	16	67,197,288	Heat shock transcription factor 4
cbfb	**18**	**22,774,824**	**ENSG00000067955**	**16**	**67,063,019**	**Core-binding factor, beta subunit**
noc4	18	30,444,963	ENSG00000131148	16	85,805,364	ER membrane protein complex subunit 8 (EMC8)
kiaa1049	18	31,005,946	ENSG00000141002	16	89,940,000	TCF25 transcription factor 25 (basic helix-loop-helix)

4.7 Supplementary Figures

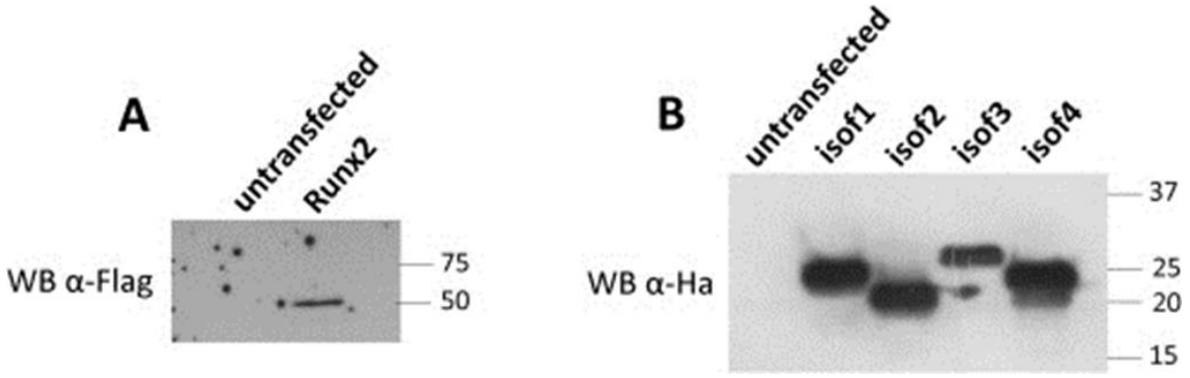


Figure 4.S1 Preparation of fusion proteins. Whole cell extracts from untransfected HEK293 cells or transiently expressing the indicated proteins were analysed by western blotting. Each lane was loaded with equivalent amounts of protein extracts (10 µg). (**A**) Expression of Runx2 detected with anti-flag antibody and (**B**) expression of the four isoforms of Cbfβ detected with anti-Ha antibody.

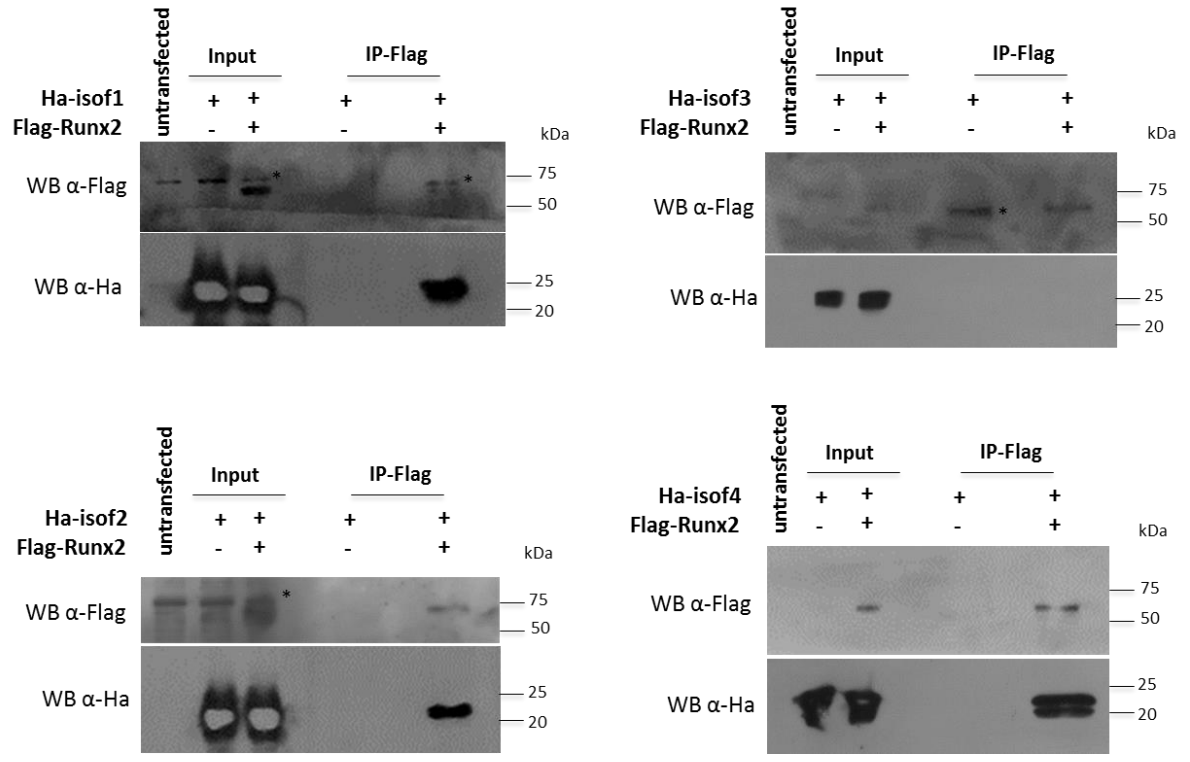


Figure 4.S2 Runx2 binds to isof1, isof2 and isof4 but not to isof3 of Cbfβ. Whole cell extracts from HEK293 cells transiently expressing Ha-Cbfβ isoforms 1-4 (isof1, isof2, isof3 and isof4) alone or in combination with Flag-Runx2 were immunoprecipitated with an anti-flag antibody and proteins were detected by western blotting with anti-Ha and anti-flag antibodies. IP indicates immunoprecipitation and WB indicates western blot. The position of non-specific proteins (*) is indicated.

```
                          1                                                                  70
   Hs_NM_001755.2     (1) MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
   Hs_NM_022845.2     (1) MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
Hs_XM_005256212.2     (1) MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
Hs_XM_006721321.1     (1) MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEI---------------

                          71                                                                 140
   Hs_NM_001755.2    (71) ASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQQEDALAQ
   Hs_NM_022845.2    (71) ASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQQEDALAQ
Hs_XM_005256212.2    (71) ASWQGEQRQTPSREYVDLEREAGK-----------------------------------QEDALAQ
Hs_XM_006721321.1    (71) ------------------------VYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQQEDALAQ

                         141                                                         187
   Hs_NM_001755.2   (141) QAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP----------------------
   Hs_NM_022845.2   (141) QAFEEARRRTREFEDRDRSHREEME-----------------ARRQQDPSPGSNLGGGDDLKLR
Hs_XM_005256212.2   (102) QAFEEARRRTREFEDRDRSHREEME-----------------ARRQQDPSPGSNLGGGDDLKLR
Hs_XM_006721321.1   (102) QAFEEARRRTREFEDRDRSHREEME-----------------ARRQQDPSPGSNLGGGDDLKLR
```

Figure 4.S3 Alignment analysis of human CBFβ protein isoform sequences. CBFβ amino acid sequences were analysed using AlignX. Numbering is according to the first residue of the protein.

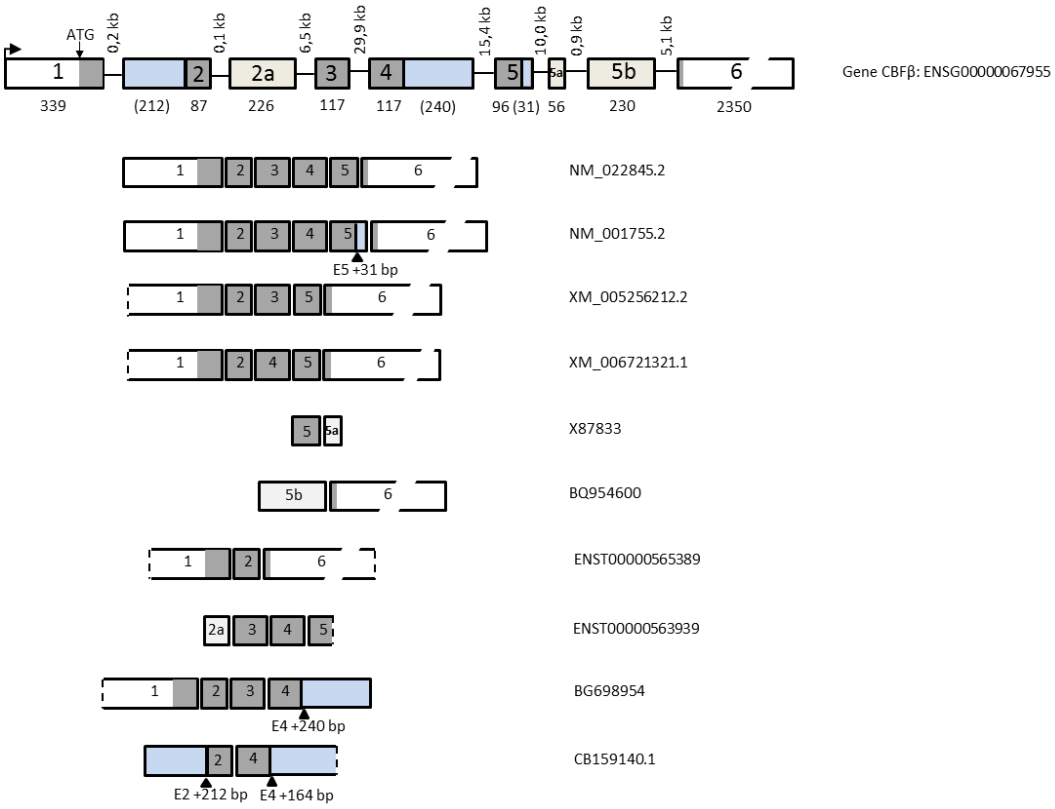


Figure 4.S4 Schematic representation of human *CBFβ* gene and corresponding transcripts. Curved arrow indicates site of transcription initiation, from exon 1. The gene structure of nine exons (boxes numbered 1-6) was obtained after assembly of all the identified transcripts. Numbers below the gene indicate the size of exons (in bp) and numbers in vertical on the top of the lines indicate the size of the introns (in kb). Ten different transcripts originated by alternative splicing are indicated below the gene. The corresponding GenBank or Ensembl accession numbers are indicated to the right of each transcript. Grey boxes represent coding regions; white boxes represent non coding regions.

```
Western_clawed_frog_B     MPRVVPDQRAKFENEEFFRKLNRESEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANWQGEPRQAPTREYVDFEREPGKVHLKAPMILNGVCVLWKGWIDLQRLDGMGCLEFDDDRAQ
African_clawed_frog_B     MPRVVPDQRAKFDNEEFFRKQSREGEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEPRQAPAREYVDFERESGKVYLKAPMILNGVCVLWKGWIDLQRLDGMGCLEFDDDRAQ
Chinese_alligator_B       MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEEPQGRFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Western_painted_turtle_B  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
American_alligator_B      MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
white-throated_sparrow_B  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Chicken_B                 MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRSHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Rat_B                     MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQTRFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLHRLDGMGCLEFDEERAQ
Mouse_B                   MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQTRFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLHRLDGMGCLEFDEERAQ
Bovine_B                  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDFEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Yangtze_River_dolphin_B   MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Chimpanzee_B              MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Rhesus_macaque_B          MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Human_B                   MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Pig_B                     MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Mexican_tetra_A           MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGEQRQMPTREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGCLEYDDERAQ
Grass_carp                MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGCLEYDDERAQ
Zebrafish_A               MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQ
Stickleback_A             MPRVVPEQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGDQRQLPAREYVDFERETGKVYLKAPMILNGVCVMWRGWIDLQRLDGMGYLEYDEERAQ
Medaka_A                  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERETAQVYLKAPMILNGVCVIWRGWIDLQRLDGMGYLEYDQERAQ
Amazon-molly_A            MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPSREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGYLEYDDERAQ
Nile_tilapia_A            MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERESGRVYLKAPMILNGVCVIWKGWIDLQRLDGMGYLEYDDERAQ
Zebra_mbuna_A             MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERESGRVYLKAPMILNGVCVIWKGWIDLQRLDGMGYLEYDDERAQ
Spotted_gar_A             MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGDQRQAPSREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGCLEFDEERAQ
Western_clawed_frog_A     MPRVVPDQRAKFENEEFFRKLNRESEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANWQGEPRQAPTREYVDFEREPGKVHLKAPMILNGVCVLWKGWIDLQRLDGMGCLEFDDDRAQ
Chinese_alligator_A       MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEEPQGRFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Tibetan_ground-tit_A      MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Green_anole_A             MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
American_alligator_A      MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
White-throated_sparrow_A  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Budgerigar_A              MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREGGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Western_painted_turtle_A  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPTREYVDFEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Rat_A                     MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQTRFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLHRLDGMGCLEFDEERAQ
Mouse_A                   MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQTRFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLHRLDGMGCLEFDEERAQ
Bovine_A                  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDFEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Yangtze_River_dolphin_A   MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Pig_A                     MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Chimpanzee_A              MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Rhesus_macaque_A          MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Human_A                   MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPASWQGEQRQTPSREYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ
Elephant_shark_A          MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQMRFQNASRDGRSEIAFVATGTNLSLQFFPANWHGEQRQTPTREYVDFDRESGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGFLEFDEERAQ
Zebrafish_C               MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQ
Mexican_tetra_C           MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFPANLHGEQRQMPTREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGCLEYDDERAQ
Nile_tilapia_C            MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERESGRVYLKAPMILNGVCVIWKGWIDLQRLDGMGYLEYDDERAQ
Zebra_mbuna_C             MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERESGRVYLKAPMILNGVCVIWKGWIDLQRLDGMGYLEYDDERAQ
Medaka_C                  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERETAQVYLKAPMILNGVCVIWRGWIDLQRLDGMGYLEYDQERAQ
Southern_platyfish_C      MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPSREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGYLEYDDERAQ
Amazon_molly_C            MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPSREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGYLEYDDERAQ
Atlantic_Salmon_C         MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNASRDGRSEIAFVATGTNLSLQFFPANLHGEQRQTPTRDYVDFDRETGKVYLKAPMILNGVCVIWKGCIDLQRLDGMGCLEFDEERAQ
Amazon_molly_D            MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPSREYVDFERETGKVYLKAPMILNGVCVIWRGWIDLQRLDGMGYLEYDDERAQ
Nile_tilapia_D            MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERESGRVYLKAPMILNGVCVIWKGWIDLQRLDGMGYLEYDDERAQ
Zebra_mbuna_D             MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEVAFVATGTNLSLQFFPANLHGDQRQVPTREYVDFERESGRVYLKAPMILNGVCVIWKGWIDLQRLDGMGYLEYDDERAQ
```

```
Western_clawed_frog_B     HEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVLRPRCP
African_clawed_frog_B     QEDALAQQAFEDSRRRTREFEDRDRSHREEMEVRVLRPRCP
Chinese_alligator_B       QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTARP
Western_painted_turtle_B  QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTTRP
American_alligator_B      QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTARP
White-throated_sparrow_B  QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTARP
Chicken_B                 QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLSVTGKKTTRP
Rat B                     QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTARP
Mouse B                   QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTARP
Bovine_B                  QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Yangtze_River_dolphin_B   QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Chimpanzee_B              QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Rhesus_macaque_B          QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Human_B                   QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTTRP
Pig_B                     QEDALAQQAFEEARRRTREFEDRDRSHREEMEVRVSQLLAVTGKKTARP
Mexican_tetra_A           HEDALAQAAFEEARRRTRDFEDRDRSHREELEPRRQQDPSPGSNMGNTDD-HKMR
Cave_fish_A               HEDALAQAAFEEARRRTRDFEDRDRSHREELEPRRQQDPSPGSNMGNTDD-HKMR
Grass_carp_A              HEDALAQAAFEEARRRTRDFEDRDRSHREDLEPRRQQDPSPGSNMGNTDD-HKMR
Zebrafish_A               HEDALAQAAFEEARRRTRDFEDRDRSHREDLEPRRQQDPSPGSNMGNTDD-HKMR
Stickleback_A             QEDALAQAAFEEARRRTRDFEDRDRSHREDLECRRQQDPSPGSNMADADVEHKMR
Medaka_A                  HEDALAQAAFEEARRRTRDFEDRDRSHREDLESRRQQDPSPGSNMANADMEHKMR
Amazon-molly_A            HEDALAQAAFEEARRRTRDFEDRDRSHREDLESRRQQDPSPGSNMANADVEHKMR
Nile_tilapia_A            HEDALAQAAFEEARRRTRDFEDRERSHREDLESRRQQDPSPGSNMANADMEHKMR
Zebra_mbuna_A             HEDALAQAAFEEARRRTRDFEDRERSHREDLESRRQQDPSPGSNMANADMEHKMR
Spotted_gar_A             QEDALAQQAFEEARRRTRDFEDRDRSHREDMEARRQQDPSPGSNMGGGEDR-TLR
Western_clawed_frog_A     HEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSSG--LGGGDDL-KLR
Chinese_alligator_A       QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Tibetan ground-tit_A      QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGTGDDL-KLR
Green_anole_A             QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
American_alligator_A      QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
White-throated_sparrow_A  QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Budgerigar_A              QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Western_painted_turtle_A  QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGSGDDL-KLR
Rat_A                     QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Mouse_A                   QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Bovine_A                  QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSSLGGGDDL-KLR
Yangtze_River_dolphin_A   QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSSLGGGDDL-KLR
Pig_A                     QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSSLGGGDDL-KLR
Chimpanzee_A              QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Rhesus_macaque_A          QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Human_A                   QEDALAQQAFEEARRRTREFEDRDRSHREEMEARRQQDPSPGSNLGGGDDL-KLR
Elephant_shark_A          QEDTLAQQAYEDVRRRARDFEDRDRSHRDDMEARRQQDPSPGSNLGSGDEL-KLR
Zebrafish_C               HEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVASKIWD
Mexican_tetra_C           HEDALAQAAFEEARRRTRDFEDRDRSHREELEDPVASKIWD
Nile_tilapia_C            HEDALAQAAFEEARRRTRDFEDRERSHREDLEDPVVSKIWD
Zebra_mbuna_C             HEDALAQAAFEEARRRTRDFEDRERSHREDLEDPVVSKIWD
Medaka_C                  HEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVVSKIWD
Southern_platyfish_C      QEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVVSKIWD
Amazon_molly_C            HEDALAQAAFEEARRRTRDFEDRDRSHREDLEDPVVSKIWD
Atlantic_Salmon_C         HEDALAQASFEESRRRTRDFEDRDRSHREDLEDPVASKIWD
Amazon_molly_D            HEDALAQAAFEEARRRTRDFEDRDRSHREDLETTAGPQPWLKHGQR
Nile_tilapia_D            HEDALAQAAFEEARRRTRDFEDRERSHREDLETTAGPQPWLKHGQR
Zebra_mbuna_D             HEDALAQAAFEEARRRTRDFEDRERSHREDLETTAGPQPWLKHGQR
```

Figure 4.S5 Protein sequence comparison of CBFβ from different species. (For description see legend of Figure 4.4).

4.8 References

Adya N, Stacy T, Speck NA, Liu PP, 1998. The leukemic protein core binding factor beta (CBFbeta)-smooth-muscle myosin heavy chain sequesters CBFalpha2 into cytoskeletal filaments and aggregates. Mol. Cell Biol. 18: 7432-7443.

Adya N, Castilla LH, Liu PP, 2000. Function of CBFβ/Bro proteins. Semin Cell Dev Biol. 11:361-368.

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, Cortes ML, Fernandez-Lopez JC, Peng S, Ardlie KG, Auclair D, Bautista-Piña V, Duke F, Francis J, Jung J, Maffuz-Aziz A, Onofrio RC, Parkin M, Pho NH, Quintanar-Jurado V, Ramos AH, Rebollar-Vega R, Rodriguez-Cuevas S, Romero-Cordoba SL, Schumacher SE, Stransky N, Thompson KM, Uribe-Figueroa L, et al., 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 486:405-409.

Black DL, 2003. Mechanisms of alternative pre-messenger RNA splicing Annu Rev Biochem. 72:291-336.

Blake T, Adya N, Kim C-H, Oates AC, Zon L, Chitnis A, Weinstein BM, Liu PP, 2000. Zebrafish homolog of the leukemia gene *CBFB*: its expression during embryogenesis and its relationship to *scl* and *gata-1* in hematopoiesis. Blood. 96:4178-4184.

Breathnach R, Chambon P, 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu Rev Biochem. 50:349-383.

Bresciani E, Carrington B, Wincovitch S, Jones M, Gore AV, Weinstein BM, Sood R, Liu PP, 2014. CBFβ and RUNX1 are required at two different steps during the development of hematopoietic stem cells in zebrafish. Blood. 124:70-78.

Chomczynski P, Sacchi N, 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 162:156-159.

Dourado G, LuValle P, 1998. Proximal DNA elements mediate repressor activity conferred by the distal portion of the chicken collagen X promoter. J Cell Biochem. 70:507-516.

Du J, Zhao K, Rui Y, Li P, Zhou X, Zhang W, Yu XF, 2013. Differential requirements for HIV-1 Vif-mediated APOBEC3G degradation and RUNX1-mediated transcription by core binding factor beta. J Virol. 87:1906-1911.

Ducy P, Zhang R, Geoffroy V, Ridall AL, Karsenty G, 1997. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. Cell. 89:747-754.

Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, Margenthaler J, Babiera GV, Marcom PK, Guenther JM, Leitch M, Hunt K, Olson J, Tao Y, Maher CA, Fulton LL, Fulton RS, Harrison

M, et al., 2012. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature. 486:353-360.

Enomoto H, Enomoto-Iwamoto M, Iwamoto M, Nomura S, Himeno M, Kitamura Y, Kishimoto T, Komori T, 2000. Cbfa1 is a positive regulatory factor in chondrocyte maturation. J Biol Chem. 275:8695-8702.

Erlebacher A, Filvaroff EH, Gitelman SE, Derynck R, 1995. Toward a molecular understanding of skeletal morphogenesis. Cell. 80:371-378.

Flores MV, Tsang VW, Hu W, Kalev-Zylinska M, Postlethwait J, Crosier P, Crosier K, Fisher S, 2004. Duplicate zebrafish runx2 orthologues are expressed in developing skeletal elements. Gene Expr Patterns. 4:573-581.

Graveley BR, 2001. Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 17:100-107.

Han MS, Kim HJ, Wee HJ, Lim KE, Park NR, Bae SC, van Wijnen AJ, Stein JL, Lian JB, Stein GS, Choi JY, 2010. The cleidocranial dysplasia-related R131G mutation in the Runt-related transcription factor RUNX2 disrupts binding to DNA but not CBF-beta. J Cell Biochem. 110:97-103.

Harada H, Tagashira S, Fujiwara M, Ogawa S, Katsumata T, Yamaguchi A, Komori T, Nakatsuka M, 1999. Cbfa1 isoforms exert functional differences in osteoblast differentiation. J Biol Chem. 274:6972-6978.

Higashikawa A, Saito T, Ikeda T, Kamekura S, Kawamura N, Kan A, Oshima Y, Ohba S, Ogata N, Takeshita K, Nakamura K, Chung UI, Kawaguchi H, 2009. Identification of the core element responsive to runt-related transcription factor 2 in the promoter of human type X collagen gene. Arthritis Rheum. 60:166-178.

Hinoi E, Bialek P, Chen YT, Rached MT, Groner Y, Behringer RR, Ornitz DM, Karsenty G, 2006. Runx2 inhibits chondrocyte proliferation and hypertrophy through its expression in the perichondrium. Genes Dev. 20:2937-2942.

Inada M, Yasui T, Nomura S, Miyake S, Deguchi K, Himeno M, Sato M, Yamagiwa H, Kimura T, Yasui N, Ochi T, Endo N, Kitamura Y, Kishimoto T, Komori T, 1999. Maturational disturbance of chondrocytes in Cbfa1-deficient mice. Dev Dyn. 214:279-290.

Kagoshima H, Akamatsu Y, Ito Y, Shigesada K, 1996. Functional dissection of the alpha and beta sub- units of transcription factor PEBP2 and the redox susceptibility of its DNA binding activity. J Biol Chem. 271:33074-33082.

Kamekura S, Kawasaki Y, Hoshi K, Shimoaka T, Chikuda H, Maruyama Z, Komori T, Sato S, Takeda S, Karsenty G, Nakamura K, Chung UI, Kawaguchi H, 2006. Contribution of runt-related transcription factor 2 to the pathogenesis of osteoarthritis in mice after induction of knee joint instability. Arthritis Rheum. 54:2462-2470.

Kanatani N, Fujita T, Fukuyama R, Liu W, Yoshida CA, Moriishi T, Yamana K, Miyazaki T, Toyosawa S, Komori T, 2006. Cbfbeta regulates Runx2 function isoform-dependently in postnatal bone development. Dev Biol. 296:48-61.

Kim IS, Otto F, Zabel B, Mundlos S, 1999. Regulation of chondrocyte differentiation by Cbfa1. Mech Dev. 80:159-170.

Komori T, Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, Shimizu Y, Bronson RT, Gao YH, Inada M, Sato M, Okamoto R, Kitamura Y, Yoshiki S, Kishimoto T, 1997. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell. 89:755-764.

Kundu M, Javed A, Jeon JP, Horner A, Shum L, Eckhaus M, Muenke M, Lian JB, Yang Y, Nuckolls GH, Stein GS, Liu PP, 2002. Cbfbeta interacts with Runx2 and has a critical role in bone development. Nat Genet. 32:639-644.

Li LH, Gergen JP, 1999. Differential interactions between Brother proteins and Runt domain proteins in the Drosophila embryo and eye. Development. 126:3313-3322.

Li N, Felber K, Elks P, Croucher P, Roehl HH, 2009. Tracking gene expression during zebrafish osteoblast differentiation. Dev Dyn. 238:459-466.

Li F, Lu Y, Ding M, Napierala D, Abbassi S, Chen Y, Duan X, Wang S, Lee B, Zheng Q, 2011. Runx2 contributes to murine Col10a1 gene regulation through direct interaction with its cis-enhancer. J Bone Miner Res. 26:2899-2910.

Liu PP, Hajra A, Wijmenga C, Collins FS, 1995. Molecular pathogenesis of the chromosome 16 inversion in the M4Eo subtype of acute myeloid leukemia. Blood. 85:2289-2302.

Martin JW, Zielenska M, Stein GS, van Wijnen AJ, Squire JA, 2011. The role of RUNX2 in osteosarcoma oncogenesis. Sarcoma. 2011:282745.

Miller J, Horner A, Stacy T, Lowrey C, Lian JB, Stein G, Nuckolls GH, Speck NA, 2002. The core-binding factor beta subunit is required for bone formation and hematopoietic maturation. Nat Genet. 32:645-649.

Mundlos S, Mulliken JB, Abramson DL, Warman ML, Knoll JHM, Olsen BR, 1995. Genetic mapping of cleidocranial dysplasia and evidence of a microdeletion in one family. Hum Mol Genet. 4:71-75.

Nakashima K, de Crombrugghe B, 2003. Transcriptional mechanisms in osteoblast differentiation and bone formation. Trends Genet. 19:458-466.

Ogasawara T, 2013. Involvements of runt-related transcription factor 2 with cell-cycle machinery in osteoblasts and skeletal development: a review of the literature. Hard Tissue. 2:29.

Ogawa E, Inuzuka M, Maruyama M, Satake M, Naito-Fujimoto M, Ito Y, Shigesada K, 1993. Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. Virology. 194:314-331.

Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GW, Beddington RS, Mundlos S, Olsen BR, Selby PB, Owen MJ, 1997. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell. 89:765-771.

Pinto J, Conceicao N, Viegas C, Leite R, Hurst L, Kelsh R, Cancela M, 2005. Identification of a new pebp2alphaA2 isoform from zebrafish runx2 capable of inducing osteocalcin gene expression in vitro. J Bone Miner Res. 20:1440-1453.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual, 2 ed. Cold Spring Harbor Laboratory Press, New York, 1989.

Sasaki K, Yagi H, Bronson RT, Tominaga K, Matsunashi T, Deguchi K, Tani Y, Kishimoto T, Komori T, 1996. Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. Proc Natl Acad Sci USA. 93:12359-12363.

Shigesada K, van de Sluis B, Liu PP, 2004. Mechanism of leukemogenesis by the inv(16) chimeric gene CBFB/PEBP2B-MHY11. Oncogene. 23:4297-4307.

Simões B, Conceição N, Viegas CSB, Pinto JP, Gavaia PJ, Hurst LD, Kelsh RN, Cancela ML, 2006. Identification of a promoter element within the zebrafish colXα1 gene responsive to Runx2 Isoforms Osf2/Cbfa1 and til-1 but not to pebp2aA2. Calcif Tissue Int. 79:230-244.

Smith N, Dong Y, Lian JB, Pratap J, Kingsley PD, van Wijnen AJ, Stein JL, Schwarz EM., O'Keefe RJ, Stein GS, Drissi MH, 2005. Overlapping expression of Runx1(Cbfa2) and Runx2(Cbfa1) transcription factors supports cooperative induction of skeletal development. J Cell Physiol. 203:133-143.

Speck NA, Stacy T, Wang Q, North T, Gu TL, Miller J, Binder M, Marín-Padilla M, 1999. Core-binding factor: a central player in hematopoiesis and leukemia. Cancer Res. 59:1789s-1793s.

Spoorendonk KM, Hammond CL, Huitema LFA, Vanoevelen J, Schulte-Merker S, 2010. Zebrafish as a unique model system in bone research: the power of genetics and in vivo imaging. J Appl Ichthyol. 26:219-224.

Stothard P, 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 28:1102-1104.

Stricker S, Fundele R, Vortkamp A, Mundlos S, 2002. Role of Runx genes in chondrocyte differentiation. Dev Biol. 245:95-108.

Takeda S, Bonnamy JP, Owen MJ, Ducy P, Karsenty G, 2001. Continuous expression of Cbfa1 in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues Cbfa1-deficient mice. Genes Dev. 15:467-481.

Taniuchi I, Osato M, Ito Y, 2012. Runx1: no longer just for leukemia. EMBO J. 31:4098-4099.

Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Tu Q, Zhang J, James L, Dickson J, Tang J, Yang P, Chen J, 2007. Cbfa1/Runx2-deficiency delays bone wound healing and locally delivered Cbfa1/Runx2 promotes bone repair in animal models. Wound Repair Regen. 15:404-412.

Wang S, Wang Q, Crute BE, Melnikova IN, Keller SR, Speck NA, 1993. Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. Mol Cell Biol. 13:3324-3339.

Wang Q, Stacy T, Miller JD, Lewis AF, Gu TL, Huang X, Bushweller JH, Bories JC, Alt FW, Ryan G, Liu PP, Wynshaw-Boris A, Binder M, Marin-Padilla M, Sharpe AH, Speck NA, 1996. The CBFβ subunit is essential for CBFα2 (AML1) function in vivo. Cell. 87:697-708.

Witten PE, Huysseune A, 2009. A comparative view on mechanisms and functions of skeletal remodelling in teleost fish, with special emphasis on osteoclasts and their function. Biol Rev Camb Philos Soc. 84:315-346.

Wittkopp N, Huntzinger E, Weiler C, Saulière J, Schmidt S, Sonawane M, Izaurralde E, 2009. Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. Mol Cell Biol. 29:3517-3528.

Yan YL; Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH, 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development. 132:1069-1083.

Zaiman AL, Lewis AF, Crute BE, Speck NA, Lenz J, 1995. Transcriptional activity of core binding factor-alpha (AML1) and beta subunits on murine leukemia virus enhancer cores. J Virol. 69:2898-2906.

Zhang YW, Yasui N, Ito K, Huang G, Fujii M, Hanai J, Nogami H, Ochi T, Miyazono K, Ito Y, 2000. A RUNX2/PEBP2alpha A/CBFA1 mutation displaying impaired transactivation and Smad interaction in cleidocranial dysplasia. Proc Natl Acad Sci USA. 97:10549-10554.

Zheng Q, Zhou G, Morello R, Chen Y, Garcia-Rojas X, Lee B, 2003. Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression in vivo. J Cell Biol. 162:833-842.

Zheng Q, Sebald E, Zhou G, Chen Y, Wilcox W, Lee B, Krakow D, 2005. Dysregulation of chondrogenesis in human cleidocranial dysplasia. Am J Hum Genet. 77:305-312.

Zhou X, Evans SL, Han X, Liu Y, Yu XF, 2012. Characterization of the interaction of full-length HIV-1 Vif protein with its key regulator CBFβ and CRL5 E3 ubiquitin ligase components. PLoS One. 7:e33495.

Chapter 5

Identification of regulatory regions in the two *zebrafish runx3* promoters by *in silico*, *in vitro* and *in vivo* functional analysis

This chapter is based in a paper in its final stage of preparation to be submitted:

B Simões, N Conceição, RN Kelsh, ML Cancela

BBA Gene Regulatory Mechanisms

Author's contribution:

Most of the experimental work and writing of the paper was performed by B Simoes, except some transfection analysis performed by N Conceição. The research concept and design was performed by B Simões, N Conceição, ML Cancela and RN Kelsh. N Conceição, ML Cancela and RN Kelsh were responsible for the critical revision and final approval of the manuscript.

5.1 Abstract

RUNX3 encodes a member of the runt domain-containing family of transcription factors. RUNX3 forms a heterodimeric complex with CBFbeta subunit and binds to the core DNA sequence 5'-PYGPYGGT-3' found in a number of enhancers and promoters, and can either activate or suppress transcription. *RUNX3* expression is regulated by two promoter regions, designated P1 and P2. Although the importance of this transcription factor in the regulation of many genes is proven, not much is known about the factors that regulate *RUNX3* transcription. We showed that both Runx2 and Runx3 interact with Cbfβ to regulate zebrafish *runx3* promoters. Using a variety of cell lines, a number of upstream regulatory regions in the zebrafish *runx3* gene were identified. Functional analysis of these regions allowed the identification of four regions in *runx3*-P1 promoter, including two positive regulatory regions from -5094 to -1766 and -948 to -786 and two negative regulatory regions from -1766 to -948 and -662 to -559, and three regions in *runx3*-P2 promoter, including two positive regulatory regions from -1232 to -699 and -554 to -474 and one negative regulatory region from -3930 to -1232. These results were complemented with *in silico* analysis of the putative transcription binding sites that are present in each identified region. In addition, *in vivo* analysis of zebrafish *runx3* gene expression was performed and the tissue-specific expression of each specific promoter fragment analysed. Taken together, the results of the functional differences between *runx3*-P1 and *runx3*-P2 promoters suggest differences in overall transcriptional activity. As these differences are probably dependent on regulation by different transcription factors, the first detailed map of the transcriptional regulatory elements of the zebrafish *runx3*-P1 and *runx3*-P2 promoter regions described in this study provides an important reference for further functional analysis to identify exactly how each factor regulates each promoter region.

5.2 Introduction

Runx (runt-related) family of transcription factors comprises Runx1/AML1/Cbfa2, Runx2/AML2/Cbfa1 and Runx3/AML3/Cbfa3 (van Wijnen et al, 2004). RUNX proteins are crucial transcription factors that regulate a wide range of biological processes to orchestrate proper cell fate determination during the development of metazoans (Coffman, 2003). Gene ablation and gain of function experiments established all three proteins as important regulators of cell fate decisions in early embryonic development and tissue differentiation - blood, neurons, and bone (Westendorf and Hiebert, 1999; Ito, 2008). Runx2 and Runx3 act early in development and play fundamental roles in skeletal development by regulating chondrocyte maturation and proliferation (Komori, 2005; 2015; Soung et al, 2007). Runx3 isoforms have also been shown to affect early neuron development (Chen et al, 2006; Kramer et al, 2006). Accordingly, deregulation of RUNX proteins is also involved in human disease (Blyth et al, 2005; Ito, 2008).

The RUNX genes arose early in evolution and maintained extensive structural similarities in vertebrates. All have two promoters, located 5' from each of the two ATG-containing exons and all heterodimerize with CBFβ that enhances their affinity to bind to the common DNA motif TGPyGGTPy (Py is a pyrimidine) (Ito, 2004; Westendorf and Hiebert, 1999). Multiple RUNX putative binding sites have been predicted in both promoters of all *RUNX* genes, and the cross-regulation and/or auto-regulation between the RUNX family members has been reported (Drissi et al, 2000; Otto et al, 2003; Spender et al, 2005; Brady et al, 2009). In zebrafish, a *runx3* orthologue was identified and its embryonic expression described (Burns et al, 2002; Kataoka et al, 2000; Simões et al, unpub.). In this study, we performed an *in silico* analysis of zebrafish *runx3* gene Distal (P1) and Proximal (P2) promoters with MATCH software that revealed the existence of several putative transcription factors binding sites (TFBSs) in both regions. We cloned the 5'-upstream regions of *runx3* gene promoters in a reporter vector (runx3-P1:LuC and runx3-P2:LuC constructs) and demonstrated that the putative promoter regions, P1 and P2, are capable of mediating gene transcription both in cultured cells and *in vivo*, in zebrafish embryos. We performed an extensive *in vitro* functional analysis in different cell lines using several deletion mutant constructs from both promoter regions allowing us to identify positive and negative regulatory regions in each promoter and by promoter comparative analysis using DiAlignTF software we were able to

identify several putative TFBSs conserved in that regions, potentially involved in the transcriptional regulation of the *runx3* gene. As our *in silico* analysis predicted various putative TFBSs for the RUNX transcription factors, the possible cross-regulation and/or auto-regulation of the *runx3* gene by Runx2 and by its own isoforms was also tested *in vitro* by co-transfection of both *runx3* promoter constructs into HEK293 cells, with or without the Runx co-factor Cbfβ, indicating that in fact, *runx3* promoters are regulated by Runx2/3 isoforms. Additionally, using Tol2-mediated technology, *runx3*-P1 and *runx3*-P2 promoters fused upstream of a mutant version of the green fluorescent protein (EGFP) reporter were analysed *in vivo* for their capacity to drive EGFP expression. We observed that *runx3*-P2:EGFP construct was capable to recapitulate part of the endogenous expression pattern of *runx3* mRNA observed by *in situ* hybridization.

5.3 Materials and Methods

5.3.1. Ethics statement

All zebrafish (Danio rerio) studies were conducted using the wild type (AB) strain available at the CCMAR zebrafish facility, University of Algarve (Faro, Portugal) and the experiments were conducted following the legislation for animal experimentation and welfare in Portugal, in accordance with the Portuguese law (Portaria 1005/02 and Portaria 1131/97) which transcribes the European Guideline 86/609/EC.

5.3.2. Zebrafish maintenance

Embryos were obtained from natural spawning, collected in embryo medium (Westerfield, 2000), raised at constant temperature of 28.5°C and staged according to Kimmel and co-workers (1995) (occasionally, when necessary their development was slowed down or speeded up by incubation at 23°C or 33°C, respectively). For some experiments, embryos were treated with 0.003% 1-phenyl-2-thiourea (PTU; Sigma) to inhibit pigment formation for further imaging improvement. All specimens were deeply anesthetised in Tricane (3-amino

benzoic acid ethyl ester also called ethyl 3-aminobenzoate, Sigma) before manipulation as described in the Zebrafish Book (Westerfield, 2000).

5.3.3. *In silico* sequence analysis

The *runx3* P1 and P2 upstream regions were identified using the BLAST tool available at NCBI database (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov). The GenBank database was searched for annotated zebrafish *runx3* sequences and the promoter sequences were extracted for analysis by selecting 5000 base pairs (bp) upstream of the translation start site of each isoform. Core promoter elements and binding sites for transcription factors were predicted by *in silico* analysis using the Match™ (public version 1.0) program that uses a library of positional weight matrices from TRANSFAC® Public 6.0, available at ww.gene-regulation.com/pub/programs. A Cut-off of 0.9 for core and matrix similarity was used.

5.3.4. Genomic DNA library preparation

The genomic DNA of an adult zebrafish specimen was prepared using the QIAGEN DNeasy Tissue kit and the DNA quality and quantity was evaluated by measurement of absorbance at 260 nm, using a NanoDrop 1000 spectrophotometer (Thermo-Fisher).

Two genomic DNA libraries were constructed using the previous genomic DNA, digested with *Stu*I and *Eco*RV, and prepared using standard protocols according to the manufacturer's instructions for the Universal GenomeWalker™ Kit (Clontech).

5.3.5 Cloning of the 5' upstream regions (P1 and P2) of z*frunx3*

The zebrafish *runx3* sequences for P1-derived transcripts (AB043789 and AB043790) and P2-derived transcript (AB043788) were aligned against the zebrafish genome (*Danio rerio*

strain Tuebingen chromosome 13, GRCz10; Sequence ID: ref|NC_007124.6|) and the 5 kb region upstream the ATG codon for each isoform was used as template to design the *runx3* specific primers, listed in **Table 5.S1**. The 5' upstream regions (P1 and P2) of z*frunx3* were obtained using as template either zebrafish genomic DNA or GenomeWalker *Eco*RV and *Stu*I libraries, in a nested PCR.

The amplifications using the GenomeWalker *Eco*RV and *Stu*I libraries as templates were performed using Advantage Polymerase Mix (Clontech), Adaptor Primer 1 (AP1) and gene-specific primers. A nested PCR was then performed using a dilution of the primary PCR product as template, the Adaptor Primer 2 (AP2) and respective gene-specific primers following the conditions suggested by the supplier. Three different size fragments of the *runx3* P2 (or PP) promoter region, named PP665, PP134 and PP104, were obtained using the *Stu*I genomic library and the adaptor primers (AP1 and AP2) and the gene-specific antisense primers (GWZfRunx3_R4 and GWZfRunx3_R3), respectively, in primary and nested PCR reaction. Similarly, one fragment (PD1200) of the *runx3* P1 (or PD) promoter region was obtained using the *Stu*I genomic library in a primary amplification reaction with AP1 adaptor and the antisense gene-specific primer (GWZfRunx3_R1) and then this product was used in a nested PCR with a pair of gene-specific primers (GWZfRunx3_F1 and GWZfRunx3_R5). PCR products were size-separated by electrophoresis on an agarose gel, purified using GeneJET Gel Extraction Kit (Thermo Scientific), cloned into the pCRII-TOPO vector using the TOPO TA cloning kit (Invitrogen) and sequenced on both strands to confirm their identity.

Zebrafish genomic DNA was used to amplify larger fragments of each P1 and P2 5' upstream region, using the KOD Hot Start DNA Polymerase (Novagene) and gene-specific primers. Amplification conditions were those suggested by the supplier. The fragments designated runx3-PD(5kb) and runx3-PD(4.4kb) from the *runx3*-P1 promoter region were amplified using the same sense primer (DrRunx3PD_F2) and the antisense primers (DrRunx3II&III_R2 and GWZfRunx3_R5, respectively) in a primary PCR and then used as template in a nested PCR with the sense primer (DrRunx3PD_KpnI_F5) and the antisense primers (DrRunx3II&III_R2 and DrRunx3PD_XhoI_R5, respectively). The fragments designated runx3-P2(3.4kb) and runx3-P2(3.3Kb) from the runx3-P2 promoter region were amplified using the same specific sense primer (DrRunx3PP_F4) and antisense (DrRunx3_R4 and

Runx3IntExR1_XhoI, respectively) in a primary PCR reaction and then used as template for a nested PCR using the sense primer (DrRunx3PPKpnI_F3) and the antisense primers (GWZfRunx3_R3 and Runx3IntExR1_XhoI, respectively). PCR products were size-separated by electrophoresis on an agarose gel, purified using GeneJET Gel Extraction Kit (Thermo Scientific), cloned into the pJet1.2 vector using the CloneJET PCR Cloning Kit (Thermo Scientific) and sequenced on both strands to confirm their identity. All the primers used are listed in **Table 5.S1**.

5.3.6 Plasmid construction for promoter functional analysis assays

The zebrafish *runx3* luciferase reporter plasmid PD(-1766/-559)LuC was obtained by direct extraction from the respective fragment (PD1200) cloned in pCRII-TOPO vector using *Kpn*I and *Xho*I restriction enzymes, whereas the constructs PD(-5094/-17)LuC and PP(-5094/-701)LuC were obtained by direct extraction from the respective fragments (PD5Kb and PD4.4Kb) cloned in pJet1.2 vector using *Kpn*I and *Bgl*II restriction enzymes. The PD(-1766/-662)LuC, PD(-948/-662)LuC, PD(-786/-662)LuC and PD(-478/-17)LuC constructs were generated by PCR amplification using as template the construct PD(-1766/-559)LuC with sense primers (pGL3_FW2; DrRunx3PD_KpnI_F1; DrRunx3PD_KpnI_F2 and DrRunx3PD_KpnI_F3, respectively) and a common antisense primer (DrRunx3PD_XhoI_R2). The PD(-1766/-17)LuC, PD(-948/-17)LuC and PD(-786/-17)LuC constructs were generated by PCR amplification using as template the construct PD1200 cloned in pCRII-TOPO vector with sense primers (Universal M13Rev; DrRunx3PD_KpnI_F1; and DrRunx3PD_KpnI_F2, respectively) and a common antisense primer (DrRunx3PD_XhoI_R1). DNA fragments were digested with *Kpn*I and *Xho*I restriction enzymes.

The zebrafish *runx3* luciferase reporter plasmids PP(-1232/-713)LuC, PP(-1232/-474)LuC and PP(-699/-474)LuC were obtained by direct extraction using *Xho*I and *Hind*III restriction enzymes, from the fragments (PP665, PP134 and PP104, respectively) cloned in pCRII-TOPO vector. The constructs PP(-3930/-474)LuC and PP(-3930/-554)LuC were obtained by direct extraction using *Kpn*I and *Bgl*II restriction enzymes, from the respective fragments (PP3.4Kb and PP3.3Kb) cloned in pJet1.2 vector. The PP(-855/-554)LuC construct was generated by

PCR amplification using as template the construct PP(-1232/-474)LuC with the primers Runx3_PPF1_KpnI and Runx3IntExR1_XhoI, while the PP(-855/-18)LuC construct was generated by PCR amplification using zebrafish genomic DNA as template with the primers Runx3_PPF1_KpnI and Runx3PP_R1_XhoI. DNA fragments were digested with *Kpn*I and *Xho*I restriction enzymes.

All the DNA fragments described above were inserted into pGL3-Basic vector previously digested with the respective restriction enzymes. Constructs were verified by double stranded DNA sequence analysis. Plasmids used for transfection studies were prepared using the GFX^{TM} Micro Plasmid Prep kit (GE Healthcare). All the primers used are listed in **Table 5.S1**.

5.3.7 Cell culture and growth conditions

The cell lines used on transient transfection assays were maintained in the appropriate cell culture medium according to their specific requirements. Human embryonic kidney 293 (HEK293, ATCC CRL-1573), human bone osteosarcoma U2OS (ATCC HTB-96) and mouse neuroblastoma Neuro2a (ATCC CCL-131) cell lines were cultured in Dulbecco's modified eagle medium (DMEM, Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS, Sigma) and 2mM L-Glutamine, whereas mouse chondrogenic ATDC5 (Sigma 99072806) and rat glial C6 (ATCC CCL-107) cell lines were maintained in DMEM:F12 (1:1) (Gibco) medium supplemented with 5% (v/v) FBS and F-12K (Gibco) medium supplemented with 2.5% (v/v) FBS plus 15% (v/v) horse serum (Gibco), respectively. 1% (v/v) penicillin/streptomycin (P/S) was added to cell culture media. Cells were allowed to grow in a humidified incubator at 37°C and an atmosphere of 5% CO_2 and passaged at regular intervals on reaching approximately 80% confluence.

5.3.8 Runx3 amplification from cell lines using RT-PCR

Total RNA was extracted from A6 (*Xenopus laevis*; ATCC#CCL102), HEKL293, Neuro2a and C6 cell lines with TRIzol (Sigma) as recommended by the manufacturer. RNA integrity was assessed through 1% (w/v) agarose/formaldehyde gel electrophoresis and RNA quantity was determined through spectrophotometry (NanoDrop). Total RNA (1 µg) was DNase I (Promega) treated before reverse-transcribed using the Moloney-murine leukemia virus (MMLV) reverse transcriptase, RNaseOUT (both from Invitrogen) and oligo(dT)-adapter primer (**Table 5.S1**). Runx3 and control (GAPDH or Actin) genes were amplified from each cell line cDNA sample by PCR using the SsoFast EvaGreen Supermix (Bio-Rad) and the respective pair of gene-specific primers (**Table 5.S2**). PCR results were analysed in a 1.8% agarose gel.

5.3.9 Transient transfection and functional promoter analysis assays

Each cell line was seeded and transfected according to its optimal number of cells per well and transfection reagent. Cells were seeded at approximately 40%-70% of confluence in 24-well plates (5×10^4 cells/well for HEK293 cells and 3×10^4 cells/well for U2OS) or in 12-well-plates (1×10^4 cells/well for ATDC5 cells) and transient transfection assays were carried out after approximately 16 hours using the X-TREME reagent (Roche). Neuro2a and C6 cells were seeded in 24-well plates (5×10^4 cells/well) and transfected with the Lipofectamine LTX with PLUS reagent (Invitrogen). The cells were transient transfected in the day after seeding with the promoter luciferase reporter constructs and the Tk Renilla vector (Promega), as an internal control. The cells were harvested 48 h post-transfection and the luciferase activity was obtained in a Synergy 4 microplate reader (BioTek) using the Dual-Luciferase reporter assay system (Promega) according to the manufacturer's instructions. Each transfection was normalized using the ratio between the firefly luciferase and Renilla luciferase activities.

5.3.10 Functional promoter analysis *in vivo*

5.3.10.1 Transient luciferase assays *in vivo*

Transient luciferase assays *in vivo* were performed by co-injection of *runx3*-P1 and *runx3*-P2 promoter luciferase constructs (≈50 pg/embryo) and the internal control Tk Renilla vector (≈5 pg/embryo), into 1-cell stage zebrafish embryos, using a Nanoject II (Drummond Scientific Co.) apparatus and incubated at 28.5°C. Embryos were collected for analysis using the Dual-Luciferase reporter assay system 24 hours post-injection in groups of five embryos per tube, anesthetized, lysated and luciferase activities (firefly and Renilla) were assessed as described previously (**Section 5.3.8**).

5.3.10.2 Generation of transgenic animals

The construct PD(-5094/-17):EGFP was obtained by direct extraction from the construct PD(-5094/-17)LuC (in pGL-3-Basic vector; Section 5.3.6) using *Kpn*I and *Hind*III restriction enzymes and the construct PP(-3930/-554):EGFP was obtained by direct extraction from the construct runx3-P2(3.3Kb) (in pJet1.2 vector; Section 5.3.5) using *Kpn*I and *Bgl*II restriction enzymes, and independently cloned into pminiTol plasmid (Balciunas et al, 2006), a modified Tol2 expression vector that contains the EGFP reporter transgene. The constructs for injection were verified by double stranded DNA sequence analysis and purified using the GFX^{TM} Micro Plasmid Prep kit. The Tol2 transposase mRNA was obtained from the template pCSTZ2.8 plasmid (Kawakami et al, 1998) by *in vitro* transcription using the SP6 mMessage mMachine kit (Ambion).

To generate the zebrafish transgenic lines, each expression construct was co-injected with transposase mRNA into the cytoplasm of single-cell stage wild-type AB strain zebrafish embryos (≈125 pg/embryo each). The injected embryos were screened for GFP expression and the transgenic founders (F0) were raised to sexual maturity and outcrossed to wild-type, and their progeny (F1) analysed for GFP expression. GFP positive F1 embryos, carriers of the transgene, were raised to adulthood to outcross and generate the transgenic stable line. The embryos were mounted in 1.2% agarose and imaged in a Zeiss LSM710 confocal microscope

or mounted in 3% methylcellulose (Sigma) and imaged using an Olympus IX81 fluorescence microscope coupled with an F-view II camera (Olympus) .

5.3.11 Statistical analysis

For the promoter-reporter assays all samples were tested in duplicates in at least three independent experiments and the data was presented as mean and standard deviation (mean ± SD) of that measurements. Statistical significance of data was determined by analysis of variance (ANOVA) with Tukey's post test using GraphPad Prism version 4.00 (GraphPad Software, www.graphpad.com). A value $P<0.05$ was assumed to denote a significant difference.

5.4. Results

5.4.1 *In silico* identification of transcriptional regulators in *runx3* P1 and P2 promoter regions

Similar to the other two members of the Runx family of transcription factors, the transcription of *runx3* is regulated by two distinct promoter regions (P1/distal and P2/proximal), located upstream of exons 1 and 2, respectively, generating two major protein isoforms that differ in their N-terminal sequence (Chapter II of this thesis). To predict putative core promoter elements and transcription factor binding sites (TFBSs) in each region, an *in silico* analysis on the regions located 1 kb upstream the translating starting codons of the zebrafish *runx3* gene was performed using the Match^{TM} (public version 1.0) program (**Figure 5.1**). The results revealed a consensus TATA box and two inverted CCAAT boxes located at positions -43, -72 and -103, respectively, upstream the *runx3*-P2 TSS (**Figure 5.1B**). No consensus TATA box was predicted in the *runx3*-P1 upstream region, but two CCAAT boxes were detected at positions -476 and -587 upstream the ATG codon in the 5'UTR region in the proximity of the *runx3*-P1 TSSs (**Figure 5.1A**). The *in silico* analysis failed

to predict an Sp1 binding site, which is usually associated with TATA-less promoters, in the *runx3*-P1 promoter region analysed, but one Sp1 binding site was predicted in *runx3*-P2 region located at position -88 upstream the P2 TSS in the proximity of the TATA and CAAT boxes (**Figure 5.1A** and **B**). Several other TFBSs were also predicted in the 1 kb upstream region of each zebrafish *runx3* promoters, such as Ap1, Ikaros, C-Ets, C/EBP, NF-Y, Runx, NFAT, Nkx, SRY, FoxD3, MyoD, among others (**Figure 5.1** and data not shown).

5.4.2 Characterization of the 5' regions of the zebrafish *runx3* gene

The *RUNX3* gene has been extensively studied in different organism models, but although several authors described the *in silico* analysis of *RUNX3* promoters, little information is available about the transcriptional regulation of this gene. To date, a few studies in mammals have shown the functionality of both *RUNX3* promoters *in vitro*, but no information is still available about the regulation of the zebrafish *runx3* promoter regions. To address this question we have analysed each promoter for the presence of positive and negative regulatory regions. Various deletion constructs between -5094 and -17 of the *runx3*-P1 region and between -3930 and -18 of the *runx3*-P2 region were sub-cloned into the pGL3-Basic vector, upstream the luciferase reporter gene. Basal promoter activity was evaluated for each fragment by transfecting the different promoter constructs into human embryonic kidney HEK293 cell line (**Figure 5.2A** and **5.2B**). The activity of each promoter region was also assessed *in vivo*, by injection of *runx3*-P1 and *runx3*-P2 constructs into 1 cell stage zebrafish embryos and determined 24 hours post-injection by luciferase assay (**Figure 5.S1**).

Figure 5.1 *In silico* analysis of 1 kb of *runx3* P1 and P2 upstream regions. Identification of core promoter elements and putative transcription factors binding sites by Match software. The putative TATA box is represented by red letters in a red box, the CAAT box highlighted in grey, the Sp1 highlighted in green, the Ap1 highlighted in blue, the Ets highlighted in pink, the C/ebp highlighted in yellow, the Ikaros in double underline letters and the Runx in blue circles. The translational start codon is represented highlighted in red.

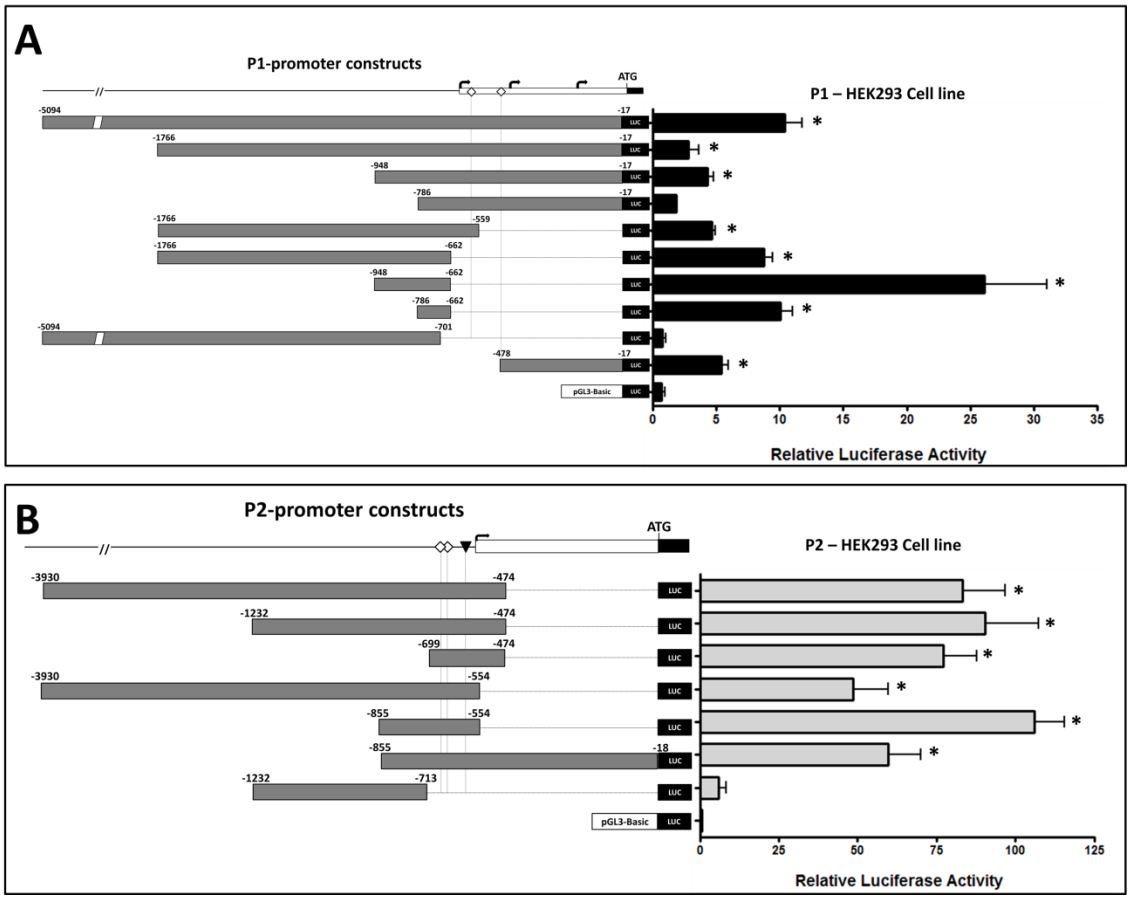


Figure 5.2 Relative transcriptional activity of zebrafish (A) *runx3*-P1 and (B) *runx3*-P2 promoter constructs. Schematic representation of the deletion constructs transfected into HEK293 cells and assessed by luciferase assay (left) and respective relative luciferase activity (right). The *runx3*-P1 and *runx3*-P2 regions are represented in the top of each figure A and B, respectively. Each sample was done in duplicates and at least three times. The promoter-less pGL3-Basic vector was used to determine the baseline. Significance (*$p<0.05$) compared to the baseline was determined by One Way ANOVA followed by Tukey's test. The TSSs are represented by a black arrow; the 5' UTR and part of the coding region are represented by white and black rectangles, respectively; the putative TATA and CAAT boxes are indicated by a filled triangle and white diamonds, respectively.

5.4.2.1 Identification of regulatory regions in *runx3* P1 promoter in HEK293 cell line

The results of the longest P1 construct, comprising 5 kb of 5' flanking region [PD(-5094/-17)LuC], showed a promoter activity of approximately 12 fold compared to the control pGL3 empty vector. However this promoter activity was significantly decreased when the region -5094 to -1766 was deleted in construct PD(-1766/-17)Luc. Further deletion of the region -1766 to -948 in PD(-948/-17)LuC slightly increased the activity of the promoter, although without significance, but an additional deletion in PD(-786/-17)LuC further

decreased significantly the activity of the promoter (**Figure 5.2A**). These results suggest that binding sites for positive regulators may be present in both regions -5094 to -1766 and -948 to -786. The *runx3*-P1 promoter region has been characterized as a TATA-less promoter and our *in silico* analysis predicted two putative CAAT boxes, located in the 5'UTR region, positions -476 and -587 upstream from the ATG codon, as well as various putative TFBSs in this region (**Figure 5.1**). This led us to test the PD(-478/-17)LuC construct, detecting luciferase activity approximately 5 fold compared to the promoter-less pGL3-Basic vector (**Figure 5.2A**). This result showed that the region -478 to -17, comprising the 5'UTR, was capable of driving basal promoter activity. We then tested PD(-5094/-701)LuC, in which the two putative CAAT boxes and the 5' UTR region were deleted. The promoter activity dropped to the levels of the promoter-less pGL3-Basic vector (**Figure 5.2A**), indicating that the region between -17 to -701 is essential to the basal activity of the *runx3*-P1 promoter. In order to identify a more restricted region essential to the basal activity of this promoter, two 3' deletion constructs were assessed for promoter activity: PD(-1766/-559)LuC, that lacks one putative CAAT box, and PD(-1766/-662)Luc that lacks both putative CAAT boxes. PD(-1766/-559)LuC showed an activity similar to that of PD(-1766/-17)LuC, but PD(-1766/-662)Luc significantly increased the promoter activity compared to PD(-1766/-17)LuC (**Figure 5.2A**), indicating that the two putative CAAT boxes are not necessary for the basal activity of this promoter. This result also suggests that the region -662 to -559 may have binding sites for negative regulators, as its deletion significantly increased the promoter activity. PD(-948/-662)LuC showed the highest promoter activity (**Figure 5.2A**), indicating that an extra region containing binding sites for negative regulators may be present between -1766 to -948 as the deletion of this region dramatically increased the promoter activity when compared to the activity of the PD(-1766/-662)LuC. Result obtained with the PD(-786/-662)LuC confirmed those obtained with the PD(-786/-17)Luc indicative of a positive regulatory region between -948 and -786 (**Figure 5.2A**).

Together, our results show that the region essential to drive P1 basal activity in HEK293 cells is comprised between -701 and -17, as the deletion of this region completely abolishes the promoter activity; suggesting the presence of binding sites for positive elements in the -5094

to -1766 and -948/-786 regions, while binding sites for repressive elements are observed in the regions -1766 to -948 and -662 to -559 of the zebrafish *runx3*-P1 promoter (**Figure 5.2A**).

5.4.2.2 Identification of regulatory regions in *runx3* P2 promoter in HEK293 cell line

A strong activity of the *runx3*-P2 promoter was observed in HEK293 cells (**Figure 5.2B**). All constructs were significantly highly expressed in these cells when compared to the activity of the promotor-less pGL3-Basic vector, with the exception of PP(-1232/-713)Luc, that lacks the predicted TATA and CAAT boxes and the 5'UTR region (**Figure 5.2B**). The constructs PP(-3930/-474)LuC, PP(-1232/-474)LuC and PP(-699/-474)LuC did not show significant differences in promoter activity among them, despite the fact that the region -3930 to -699 is deleted. However, when the region -3930 to -855 was deleted and the PP(-855/-554)LuC construct used, a significant increase in promoter activity was observed compared to the activity of the PP(-3930/-554)LuC (**Figure 5.2B**), suggesting the presence of negative regulator in the region between -3930 to -855. Moreover, when the region -554 to -474 was deleted and the PP(-3930/-554)LuC construct was used we observed a significant decrease in the promoter activity compared to the activity of the PP(-3930/-474)LuC construct (**Figure 5.2B**), indicating that the region -554 to -474 may have binding sites for positive regulators. Taking together, these results show that the region between -855 and -699 may also have binding sites for positive regulators. The capacity of the 5'UTR region to drive promoter activity was also assessed by the PP(-855/-18)LuC construct. The result showed that this construct is capable to drive higher basal promoter expression compared to the activity of the promoter-less pGL3-Basic vector. Nevertheless, the addition of the 5'UTR region significantly reduced the basal activity of the promoter compared to the activity of the PP(-855/-554)LuC construct (**Figure 5.2B**). This suggests that the region between -474 and -18 may have binding sites for negative regulators.

Together, our results for the activity of *runx3*-P2 promoter constructs, show that the region between -713 and -554 is essential to drive *runx3*-P2 basal activity in HEK293 cells, and suggest that (i) the regions -3930 to -855 and -474 to -18 might include binding sites for

repressive elements, while (ii) the regions -855 to -699 and -554 to -474 most possibly possesses binding sites for positive regulatory elements (**Figure 5.2B**).

5.4.3 Functional characterization of Runx2/3 and Cbfβ interaction to regulate *runx3* promoters

To address the possible cross-regulation between Runx2 and Runx3 and the putative auto-regulation of *runx3* promoter gene by its own isoforms, we have transfected HEK293 cells, that do not express Runx3 (**Figure 5.S2**) or any other *RUNX* gene (Zheng et al, 2007; Bone et al, 2010), with zebrafish *runx3*-P1 promoter construct PD(-5094/-17)LuC and *runx3*-P2 promoter construct PP(-3930/-554)LuC alone or in combination with zebrafish Runx2 and Runx3 isoforms (MASN or M(R/H)IPV) in the presence or absence of their transcriptional partner Cbfβ (**Figure 5.3**).

Our results for the co-transfection of *runx3*-P1 construct PD(-5094/-17)LuC showed a significant repression effect when this construct was co-transfected with the Runx3-MASN isoform in the presence of Cbfβ isoform 4, but no significant difference was observed when this construct was co-transfected with the other Runx2 and Runx3 isoforms in the present or absence of the Cbfβ, when the activity was compared to the activity of the control (promoter construct co-transfected with the empty pCMX vector) (**Figure 5.3A**). On the other hand, the results for the *runx3*-P2 construct PP(-3930/-554)LuC showed a significant decrease in the activity of the promoter when co-transfected with all Runx2 and Runx3 isoforms and compared to the expression of the control (promoter construct co-transfected with the empty pCMX vector) (**Figure 5.3B**). The results also showed that, in general, the presence of Cbfβ appeared to be associated to a more consistent repression of the activity of the runx isoforms, but this effect was not statistically significant.

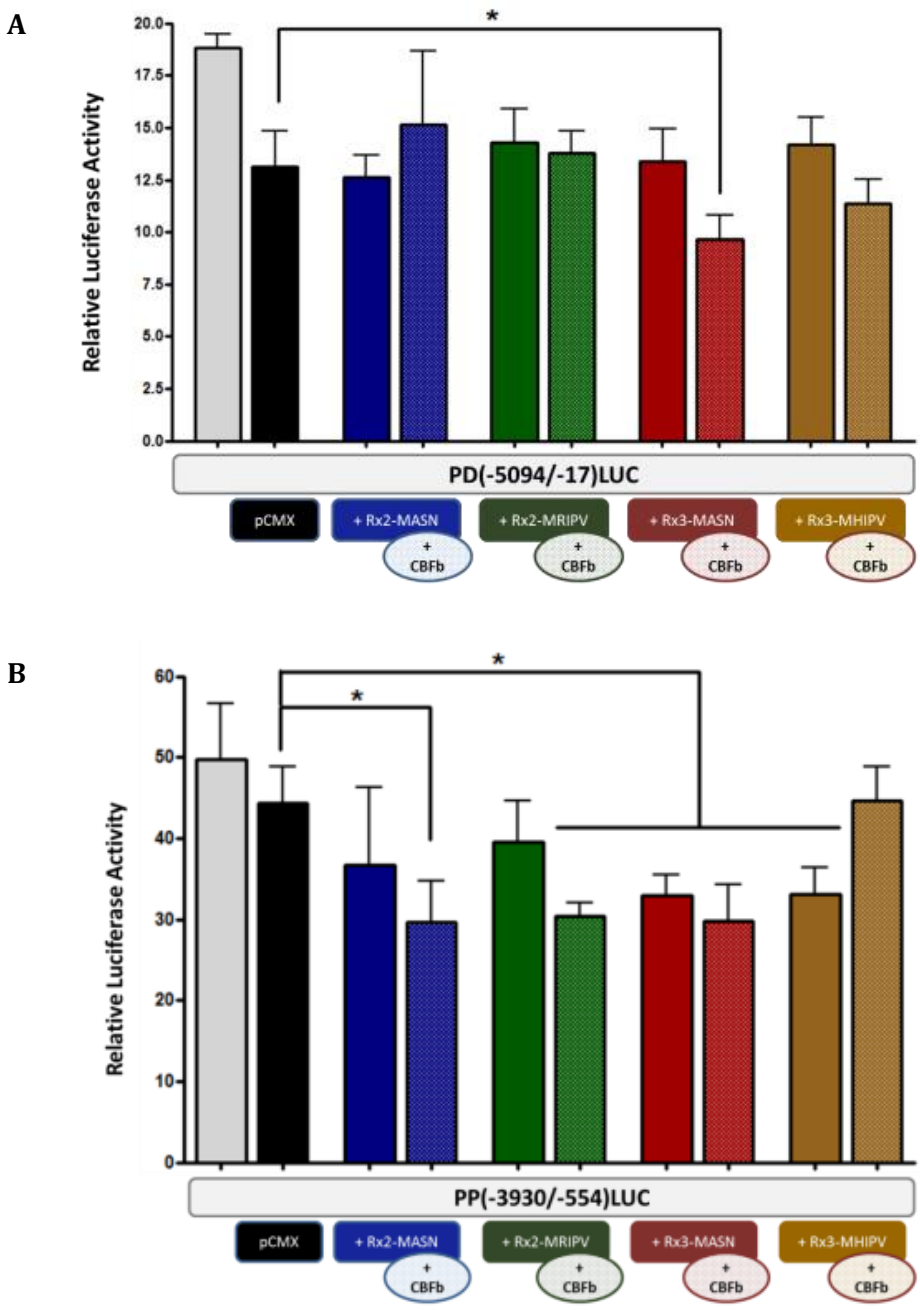


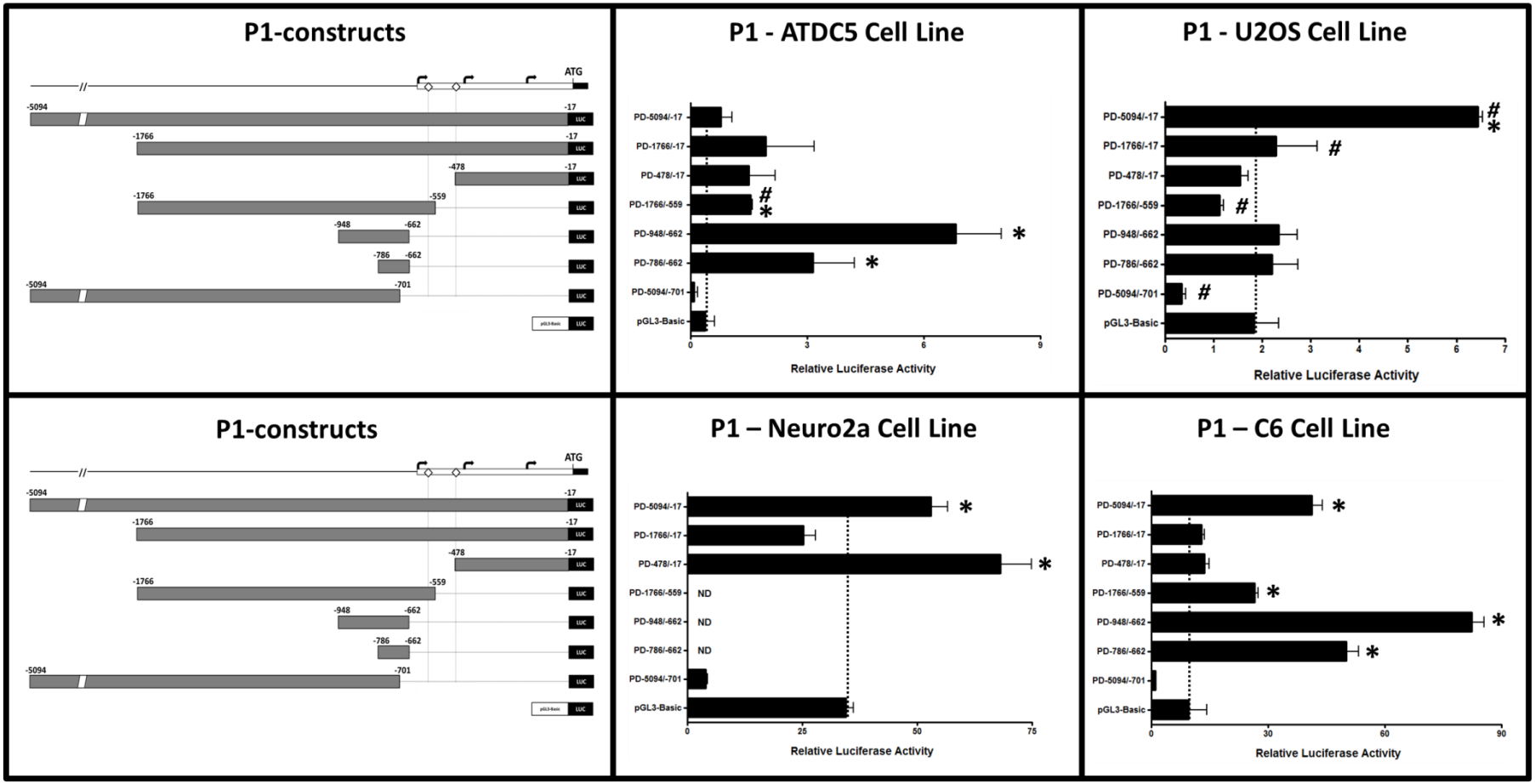
Figure 5.3 The transcriptional regulation of *runx3* promoter regions are affected by co-transfection with Runx2 and Runx3 isoforms. HEK293 cells were transfected with either zebrafish *runx3*-P1 [PD(-5094/-17)LuC] or *runx3*-P2 [PP(-3930/-554)LUC] constructs. Cells were co-transfected with zebrafish Runx2 or Runx3 isoforms (MASN or M(R/H)IPV) in the presence or absence of Cbfβ expression plasmid. The data indicated is a representative plot from two experiments that were done in duplicate. Significance was determined by One Way ANOVA followed by Tukey test. *p<0.05.

5.4.4 *in vitro* analysis of *runx3* promoters in different cell lines

Because *Runx3* is expressed in a variety of cell types, to analyse if *runx3* promoter regions respond differently to different cell micro-environments we characterized their expression in a number of cell lines (**Figure 5.4**). The mouse chondrogenic ATDC5, human bone U2OS, mouse neural Neuro2a and rat glial C6 were selected since they are derived from different cell types and are well documented for *in vitro* transfection assays. It was previously reported that ATDC5 (Wigner et al, 2013) and U2OS cells (Lai and Mager, 2012) express all Runx family members. As no information was available about *Runx3* expression in Neuro2a and C6 cell lines, a PCR approach was used to address this question, confirming that both cell types express *runx3* transcripts (**Figure 5.S2**). We first tested the functionality of zebrafish *runx3* promoter regions in the different cellular contexts, by transfecting the larger P1 and P2 constructs, PD(-5094/-17)LuC and PP(-3930/-474)LuC respectively. The results showed that the larger P1 construct [PD(-5094/-17)LuC] had variable activity in osseous vs non-osseous cell lines and that the larger P2 construct [PP(-3930/-474)LuC] showed high activity in all cell lines tested (**Figure 5.S3**). Next, *runx3*-P1 and *runx3*-P2 deletion mutants were tested to further investigate their activity in the cells and analyse how they respond to different cell environmental stimuli (**Figure 5.4**).

The results showed that in general all constructs from *runx3*-P2 promoter were active in all cells tested (**Figure 5.4B**). In contrast, a high variability in the expression levels of the *runx3*-P1 constructs across the different cell lines was observed (**Figure 5.4A**). Because the promoter-less pGL3-Basic vector, used to define the baseline, produced variable activity in the different cell lines, the absolute fold over background values cannot be compared across the different cell lines; rather, the relative changes in activities among different deletion constructs within a cell line were examined.

A


P1-constructs
ATG
-5094
-17
-1766
-478
-559
-948
-662
-786
-701
LUC
pGL3-Basic
P1 - ATDC5 Cell Line
PD-5094/-17
PD-1766/-17
PD-478/-17
PD-1766/-559
PD-948/-662
PD-786/-662
PD-5094/-701
pGL3-Basic
Relative Luciferase Activity
P1 - U2OS Cell Line
Relative Luciferase Activity
P1-constructs
P1 – Neuro2a Cell Line
ND
Relative Luciferase Activity
P1 – C6 Cell Line
Relative Luciferase Activity

B

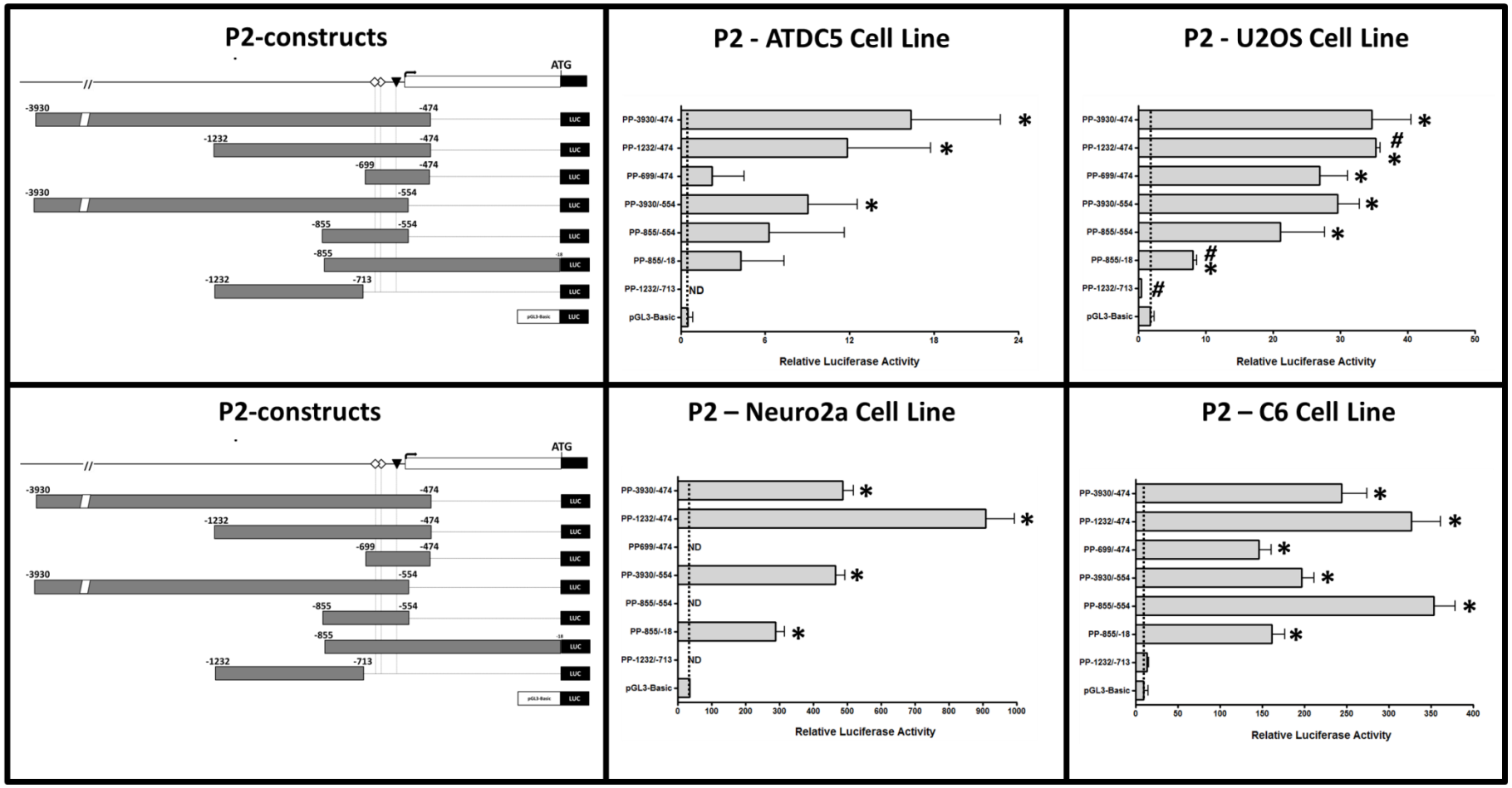

P2-constructs
ATG
-3930
-474
-1232
-699
-554
-855
-18
-713
LUC
pGL3-Basic
P2 - ATDC5 Cell Line
PP-3930/-474
PP-1232/-474
PP-699/-474
PP-3930/-554
PP-855/-554
PP-855/-18
PP-1232/-713
pGL3-Basic
ND
Relative Luciferase Activity
P2 - U2OS Cell Line
Relative Luciferase Activity
P2-constructs
P2 – Neuro2a Cell Line
PP699/-474
Relative Luciferase Activity
P2 – C6 Cell Line
Relative Luciferase Activity

Figure 5.4 Basal activity of zebrafish (**A**) *runx3*-P1 and (**B**) *runx3*-P2 promoter regions in various *runx3* expressing cells. Mouse chondrogenic ATDC5, human osteosarcoma U2OS, human neuroblastoma Neuro2a and rat glioma C6 cell lines were transiently transfected with *runx3*-P1 and *runx3*-P2 promoter constructs. The *runx3*-P1 and *runx3*-P2 promoter regions and the schematic representation of the deletion constructs are represented in the left. The TSSs are represented by a black arrow; the 5' UTR and part of the coding region are represented by white and black rectangles, respectively; the putative TATA and CAAT boxes are indicated by a filled triangle and white diamonds, respectively. The promoter-less pGL3-Basic vector was used to determine the baseline (dotted line) and samples with significant higher relative expression compared to the baseline are indicated (*). n≥3 (except n=2 for samples indicated with #). Significance (*p<0.05) was determined by One Way ANOVA followed by Tukey's test. ND - not determined.

5.4.4.1. Analysis of *runx3*-P1 constructs activity in the different cell lines

The results for the expression of *runx3*-P1 constructs showed that in ATDC5 cells only the PD(-948/-662)LuC and PD(-786/-662)LuC constructs had an expression significantly higher than the baseline (**Figure 5.4A**), while in U2OS cells, all constructs, except the PD(-5094/-17)LuC, have a relative expression activity similar to the baseline (**Figure 5.4A**). In Neuro2a cells only the constructs PD(-5097/-17)LuC and PD(-478/-17)LuC show a significant relative expression, however not all constructs were tested in this cell line and in C6 cells, all constructs, except the PD(-5094/-701)LuC have an expression higher than the baseline (**Figure 5.4A**), showing a relative expression pattern similar to that observed in HEK293 cells (**Figure 5.2A**). We compared the relative expression of the deletion constructs in each line and identified the regulatory regions of the *runx3*-P1 promoter in each cell context. These results are summarized in **Table 5.1A**. Overall, from the results obtained from all the cell lines tested, it was possible to identify two positive regulatory regions [A1 (-5094 to -1766) and A2 (-948 to -786)] and two negative regulatory regions [R1 (-1766 to -948) and R2 (-662 to -559)] in *runx3*-P1 promoter (**Figure 5.2A** and **5.4A** and **Table 5.1A**).

We then analysed which putative TFBSs are conserved in the identified regulatory regions by our *in silico* comparative analysis (Section III of this thesis) (**Table 5.2**). We found that those regions are highly conserved and contain multiple putative TFBSs. Interestingly, we observed that the PD(-5094/-17)LuC construct, that showed high relative expression in all cell lines tested, with the exception of ATDC5 cells, contains the regulatory region A1, with several

putative TFBSs for known positive regulators (**Table 5.2**). The same occurs regarding the PD(-948/-662)LuC construct that showed the higher relative expression in three out of the five cell lines tested, and includes the identified positive regulatory region A2, containing some conserved putative TFBSs known to be important for transcription regulation (**Table 5.2**). Moreover, the regions R1 and R2 identified in our results contain putative TFBSs for known repressors (**Table 5.2**). A full description of the putative TFBSs present in these identified regions is found in Section III of this thesis, as well as their known role in skeletogenesis.

5.4.4.2. Analysis of *runx3*-P2 constructs activity in the different cell lines

The results obtained upon the transfection of the *runx3*-P2 promoter constructs in different cell lines show that this promoter is highly active and the expression patterns of the deletion constructs are similar in all cells tested (**Figure 5.2B** and **5.4B**). We compared the relative expression of the deletion constructs in each line and identified the regulatory regions of the *runx3*-P2 promoter region in each cell context. These results are summarized in **Table 5.1B**. Overall, we identified two positive regulatory regions [A1 (-1232 to -699) and A2 (-554 to -474)] and one negative regulatory region [R1 (-3930 to -1232) in *runx3*-P2 promoter (**Figure 5.2B** and **5.4B** and **Table 5.2B**), A1 appearing to be a positive regulatory region in all cells tested (**Figure 5.2B** and **5.4B**). The second positive regulatory region, A2, was observed in all cell lines tested with the exception of Neuro2a cells, but the effect was only significant in HEK293 and C6 cells (**Figure 5.2B** and **5.4B** and **Table 5.1B**). On the other hand, the presence of a negative regulatory region R1 was only observed in the Neuro2a and C6 cells lines (**Figure 5.4B** and **Table 5.1B**). The highest activity of the *runx3*-P2 promoter was observed with the construct PP(-3930/-474)LuC in ATDC5 and U2OS cells, but no significant difference was observed compared to the activity of the PP(-1232/-474)LuC construct, and for the construct PP(-1232/-474)LuC in Neuro2a and C6 cells (**Figure 5.4B** and **Table 5.1B**). When the *runx3*-P2 promoter regulatory regions identified were analysed for the presence of conserved putative TFBSs, determined by promoter comparative analysis in Chapter III of this thesis, both A1 and R1 regions contained various conserved putative TFBSs. However, no putative TFBSs were detected in the region A2 (**Table 5.1B**).

Table 5.1A. Activating and repressing regions identified in *runx3*-P1 promoter in different cell lines.

Cell line	HEK293	ATDC5	U2OS	Neuro2a	C6
Activating regions					
(A1) -5094 to -1766	**Yes**	ns	**Yes**	**Yes**	**Yes**
(A2) -948 to -786	**Yes**	**Yes**	ns	nd	**Yes**
Repressing regions					
(R1) -1766 to -948	**Yes**	**Yes**	ns	-1766 to -478[a]	**Yes**
(R2) -662 to -559	**Yes**	**Yes**	ns		**Yes**
Construct showing higher activity	**PD(-948/-662)LuC**	**PD(-948/-662)LuC**	**PD(-5094/-17)LuC**	**PD(-478/-17)LuC**	**PD(-948/-662)LuC**

[a] Deletion constructs of the region -1766 to -786 were not determined, so a smaller region could not be determined. ns – not significant. nd – not determined.

Table 5.1B. Activating and repressing regions identified in *runx3*-P2 promoter in different cell lines.

Cell line	HEK293	ATDC5	U2OS	Neuro2a	C6
Activating regions					
(A1) -1232 to -699	ns	**Yes**	**Yes**	nd	**Yes**
(A2) -554 to -474	**Yes**	ns	ns	ns	**Yes**
Repressing region					
(R1) -3930 to -1232	ns[a]	ns	ns	**Yes**	**Yes**
Construct showing higher activity	**PP(-3930/-474)LuC** **PP(-1232/-474)LuC** **PP(-855/-554)LuC**[b]	**PP(-3930/-474)LuC** **PP(-1232/-474)LuC**[b]	**PP(-3930/-474)LuC** **PP(-1232/-474)LuC**[b]	**PP(-1232/-474)LuC**	**PP(-1232/-474)LuC** **PP(-855/-554)LuC**[b]

[a] not significant for this region, but significant for the region -3930 to -855. ns – not significant. nd – not determined. [b] The activity between the constructs is not statistically different.

Table 5.2. Identification of the putative TFBSs, obtained by promoter comparative analysis, in the predicted regulatory regions of *runx3*-P1 and *runx3*-P2 promoters.

Activating regulatory regions		**Repressing regulatory regions**	
***runx3*-P1 promoter region**			
-5094 to -1766 (Region A1)	**-948 to -786 (Region A2)**	**-1766 to -948 (Region R1)**	**-662 to -559 (Region R2)**
(**Blocks 1 to 8**)[a] V$OCT1; O$PTBP; V$ZF35; V$RUSH; V$RU49; V$ZF03; V$HASF; V$SMAD; V$TCFF; V$CSEN; O$INRE; V$PAX6; V$AP1R; V$HAND; V$MYOD; V$SORY; V$NKXH; V$BRN5; V$HOXF; V$YY1F; V$BRNF; V$ABDB; V$SF1F; V$LTFM; V$RBP2	(**Block 14**)[a] V$BRN5; V$SORY; V$BRNF; V$CLOX; V$GATA; V$HAND; V$HOMF; V$EBOX; V$HIFF; V$MYOD; V$HESF; V$MITF; V$NKXH; V$SREB; V$CAAT; V$NEUR; V$KLFS; O$INRE; V$CARE; V$SAL2; V$MIZ1	(**Blocks 9 to 13**)[a] V$FKHD; V$NFAT; V$ABDB; V$BRNF; V$CDXF; V$ZF35; V$ARID; O$VTBP; V$SATB; V$HAML; V$NFKB; V$ZFHX; V$ZF08; V$SAL2; V$CEBP; V$OCT1; V$PAX6; V$BRAC; V$HASF; V$RUSH; V$CIZF; V$CHRF	(**Blocks 17 and 18**)[a] V$STAT; V$XBBF; V$CEBP; V$RBPF; V$IKRS; V$MZF1; V$MYT1; V$ZF03; V$ZFHX; V$TCFF
***runx3*-P2 promoter region**			
-1232 to -699 (Region A1)	**-554 to -474 (Region A2)**	**-3930 to -1232 (Region R1)**	
(**Block 8**)[a] V$CART; V$HOMF; V$MEF2; V$ABDB; V$LHXF; O$PTBP; V$BRNF; V$CHRF; V$CLOX; V$FKHD; V$HOXC; O$VTBP; V$STAB	No conserved Blocks in this region[a]	(**Blocks 1 to 7**)[a] O$VTBP; V$CDXF; V$PARF; V$FKHD; V$BRNF; V$HOMF; V$PAXH; V$RUSH; V$CHRF; V$ARID; V$MYBL; V$XBBF; V$OSRF; V$CSEN; V$KLFS; V$TALE; V$AP2F; V$HAND; V$MYOD; V$EBOX; V$HESF; V$E2FF; V$ZF5F; O$TF2B; V$CART; V$CREB; V$HIFF; V$LHXF; V$NKXH; V$PAX3; V$SIX3; V$HOBX; V$HOXF; V$PAX6; V$SATB; V$BCDF; V$DLXF; V$PIT1; V$NKX6; V$PAXH; V$HOXC; V$ABDB; V$OCT1; V$ZF35; O$INRE; V$TCFF; V$SATB; V$SAL1	
	V$CAAT; V$TATA; V$SP1; V$AP1; V$CEBP[b]		

[a] Blocks and TFBSs conserved between zebrafish and fugu *runx3* promoters (Described in chapter III of this thesis). [b] TFBSs identified in zebrafish *runx3* promoter by MATCH software from TRANSFAC.

5.4.5 Tools for *in vivo* temporal and spatial analysis of zebrafish *runx3* gene expression

Although *in vitro* assays using cultured cell lines are powerful tools for promoter analysis and identification of regulatory regions, these systems have the disadvantage of poorly replicating the complex *in vivo* micro-environment. For this reason, we attempted to analyse the expression of *runx3*-P1 and -P2 promoters *in vivo*, by co-injecting into 1-cell stage zebrafish embryos the *runx3*-P1 [PD(-5094/-17):EGFP] and *runx3*-P2 [PP(-3930/-554):EGFP] constructs together with the transposase mRNA. The transposase enzyme is essential for successful delivery of the transposon transgene into the host genome after DNA-mRNA co-microinjection (Kawakami, 2007) and GFP expression reveals the expression pattern associated with each specific promoter fragment (**Figure 5.5**).

Although the P0 generation will be transgenic chimaeras, we expected to observe the tissue-specific expression of GFP. As shown in **Figure 5.5**, transient *runx3*-P2 transgenic embryos show GFP expression in a variety of tissues. At 1 dpf, we could detect GFP expression driven by *runx3*-P2 promoter in retinal neuroepithelial cells in the eye, otic epithelium (ot), cranial ganglia (cg) as well as migrating cranial neural crest (arrow in B), in cells within the heart (ht) (**Figure 5.5A** and **5.5B**) and in the trunk, with GFP expression being detected in migratory neural crest cells (mNCCs) (**Figure 5.5C**). At 2 dpf the embryos showed GFP expression in a small fraction of retinal neuroepithelial cells in the eye, cells within the heart (ht) and non-neural cells in the brain (**Figure 5.5D** and **5.5E**), in the Rohon-Beard cell (arrow), Schwann cells of the spinal nerves (arrowhead) and of the dorsal branch of the posterior lateral line nerve (plln), and expression in the muscle cells (**Figure 5.5F**). At 3 dpf GFP expression could still be detected in cells within the heart (**Figure 5.5G**), in the trigeminal ganglion (arrowhead) and in the optic vesicle (arrow) (**Figure 5.5H**), in growth cone of the nearby Rohon-Beard neuron (arrow) (**Figure 5.5I**), and in interneurons (**Figure 5.5J** and **5.5K**). The same pattern of expression was observed when another runx3-P2 construct [PP(-1232/-474):EGFP] was injected (data not shown).

Even after several attempts injecting different concentrations of the *runx3*-P1 construct and mRNA transposase, the transient transgenic embryos just occasionally showed some green positive cells (**Figure 5.5L-P**). Since this is a transient analysis we would not expect to see all *runx3*-P1 derived positive cells labelled, however the few green positive cells detected were

not consistent comparing the multiple transient embryos, so we asked if this construct was working. To verify its functionality, we transfected A6 cell line with the *runx3*-P1 construct, and the *runx3*-P2 construct and the pEGFP-N1 vector as controls, and observed GFP expression 24 hours after transfection. We could detect GFP expressing cells in all three samples tested, confirming that the constructs are driving GFP expression (**Figure 5.S4**).

The results of the transient transgenesis indicate that the tool works (at least for the *runx3*-P2 construct), but stable transgenic lines are needed to have unambiguous results and to be able to analyse in detail the GFP expression pattern under control of the *runx3* promoter during development. In order to make stable transgenic lines from each *runx3*-P1 and *runx3*-P2 promoter constructs, the injected embryos (F0) were screen for GFP positive expressing cells and raised to sexual maturity. As few GFP positive cells were observed in the embryos injected with *runx3*-P1 construct, we raised all embryos injected that at 5 dpf were phenotypically normal. Then they were outcrossed with WT fish and the progeny (F1) screened for GFP expression. Despite the fact that the Tol2 system has been extensively used to create transgenic zebrafish lines and is reported to result in an integration ratio of 50-70%, unfortunately we failed to find F1 transgenic fish expressing GFP driven by the *runx3* promoter fragments in a pattern similar to that of the endogenous *runx3* gene. In total, from 118 F0 embryos of *runx3*-P1 and 84 F0 embryos of *runx3*-P2 that were raised just 61 and 47, respectively, survived to adulthood. Up to the time of submission of this thesis, it was just possible to test 25 *runx3*-P1 and 18 *runx3*-P2 founders for the transgene integration. Analysing the progeny (F1) from the 25 *runx3*-P1, we found no embryos expressing GFP. On the other end, from the 18 *runx3*-P2 founders we found 3 that expressed GFP, but the expression pattern did not correspond to the expression of the endogenous *runx3*, as it was expressing GFP ubiquitously (data not shown).

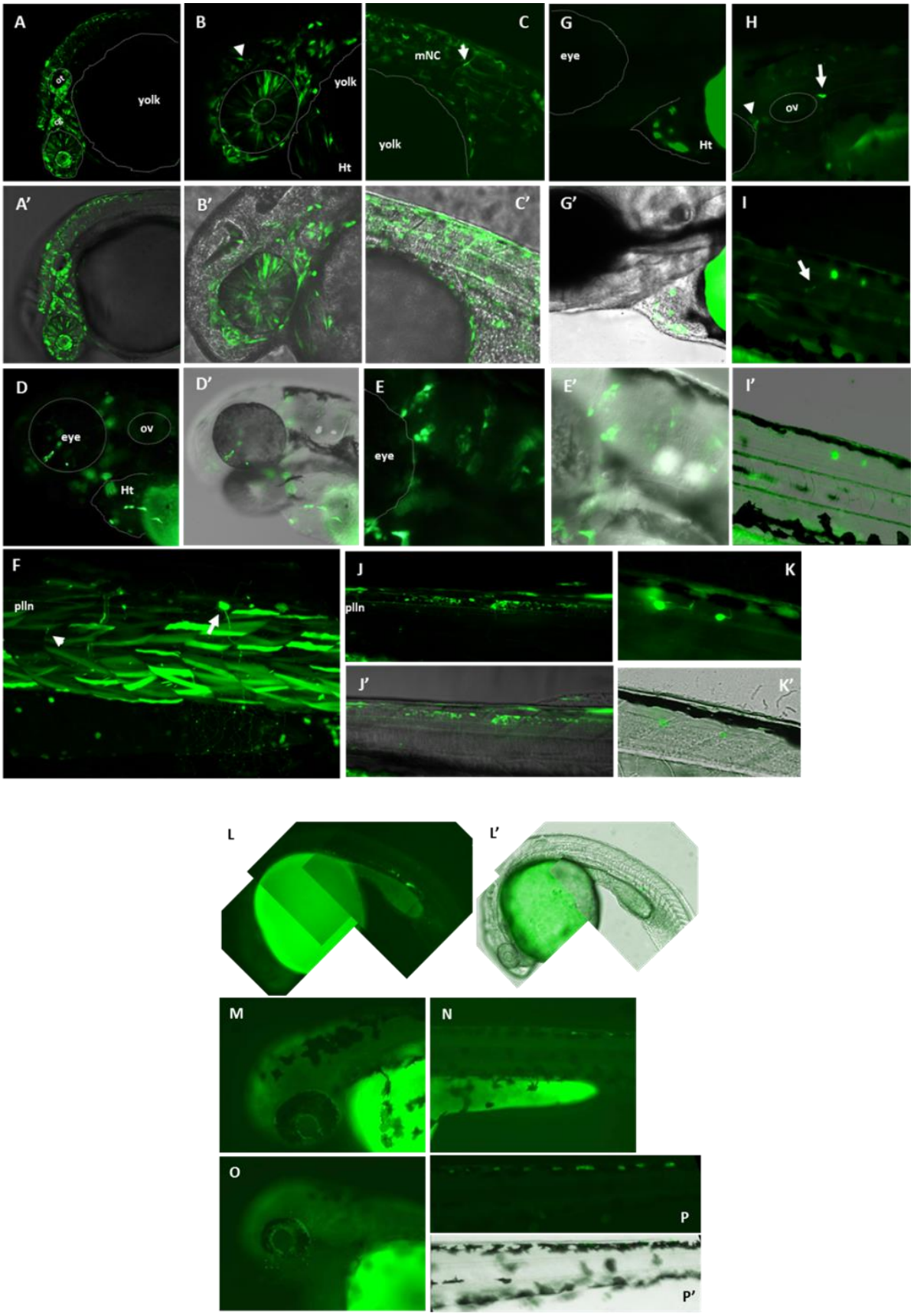

A
ot
yolk
ob
B
yolk
Ht
C
mNC
yolk
G
eye
Ht
H
ov
A'
B'
C'
G'
I
D
eye
ov
Ht
D'
E
eye
E'
I'
F
plln
J
plln
J'
K
K'
L
L'
M
N
O
P
P'

Figure 5.5 Transient injected embryos express GFP under control of *runx3* PP(-3930/-554):EGFP promoter construct. Brightfield and fluorescent images of living zebrafish embryos directed by (L-P) *runx3* PD(-5094/-17):EGFP and (A-K) *runx3* PP(-3930/-554):EGFP promoter constructs. (A-C, L) 1 dpf embryos; (D-F, M-N) 2 dpf embryos; (G-K, O-P) 3 dpf embryos. (A, B) confocal projection of the head of a 1 dpf embryo showing GFP expression in retinal neuroepithelial cells in the eye, otic epithelium (ot), cranial ganglia (cg) as well as migrating cranial neural crest (arrow in B) and in cells within the heart (ht). (C) confocal projection of the anterior trunk of a 1 dpf embryo showing GFP expression in migratory neural crest cells (mNCCs). (D, E) image of the head of a 2 dpf embryo showing GFP expression in a small fraction of retinal neuroepithelial cells in the eye, cells within the heart (ht) and cells in the brain. (F) confocal projection of the trunk of a 2 dpf embryo showing GFP expression in the Rohon-Beard cell (arrow), Schwann cells of the spinal nerves (arrowhead) and of the dorsal branch of the posterior lateral line nerve (plln), and expression in the muscle cells. At 3 dpf GFP expression can still be detected in (G) the cells within the heart, (H) in the trigeminal ganglion (arrowhead) and optic vesicle (arrow), (I) in growth cone of the nearby Rohon-Beard neuron (arrow), and (J, K) in interneurons. (L-P) The *runx3*-P1 transgenic embryos generally showed little or no GFP fluorescence. Note that signals in N and P are auto-reflection of iridophores. All are ventral images, dorsal is up and anterior is to the left. All panels contain fluorescent images from live embryos, overlaid on DIC images in (A'-F', G', I'-K', P') to show embryonic context for (A-F, G, I-K, P). Images were obtained using a Zeiss LSM710 confocal microscope and an Olympus IX81 fluorescence microscope coupled with a F-view II camera (Olympus). dpf - days post-fertilization.

5.4 Discussion

RUNX3 belongs to the family of Runt domain transcription factors, that also includes RUNX1 and RUNX2 and, as the name indicates, all three members share a conserved region of 128 amino acid termed the Runt domain, also conserved throughout evolution (Levanon and Groner, 2004). This region has been shown to associate with a broad range of proteins that can either be co-activators or repressors to direct biological outcomes in a context-dependent manner (reviewed in Downing, 1999). Although all three RUNX proteins bind and recognize the same DNA sequence, and even if in some tissues its expression overlaps, they have been shown to have a tissue-specific expression pattern, indicating a tight regulation as they are involved in different biological functions. RUNX3 is the less studied of the three member of the RUNX family, and although in the last years it was extensively investigated for its capacity to regulate transcription of a variety of target genes, data related to transcriptional regulation of *RUNX3* itself are scarce. However, the little information available shows that *RUNX3* promoter activity is regulated by its family RUNX members

(Drissi et al., 2000; Spender et al., 2005). Similar to what is described for other species, in zebrafish three splice variants of *runx3* were identified, derived from two alternative promoters, and encoding two protein isoforms, Runx3-MASN and Runx3-MHIPV (Kataoka et al, 2000). But since there are few functional studies, it is still unclear how each isoform is regulated and the relative importance of the different isoforms remains unknown. We combined *in silico*, *in vitro* and *in vivo* analysis to gain insights into the regulation of the zebrafish *runx3* gene 5' flanking regions.

Since the primary regulatory elements of many genes are encoded in proximal sequences upstream of the transcriptional start site and in the 5' UTR region, our first analysis focused on the regulatory regions located 1 kb upstream the translating starting codons of the zebrafish *runx3* gene. The *in silico* analysis using the MATCH software, that uses a library of positional weight matrices from TRANSFAC^{R} Public 6.0, predicted in each promoter several cis-elements described as important for transcriptional activity. Both promoters contain two CAAT boxes in the proximity of the TSSs located in *runx3*-P1 at positions -476 and -587 bp and in *runx3*-P2 at positions -634 and -665 bp upstream the translation start codon (**Figure 5.1A** and **5.1B**), in agreement to what has been described for other species, where no TATA box was detected, but a consensus CAAT box was predicted (Bangsow et al, 2001; Ng et al, 2007; Nah et al, 2014). However, in the zebrafish *runx3*-P2 promoter region a TATA box was predicted at a consensus distance from the P2-TSS (position -43 bp), and it would be interesting to see if this TATA box is functional, which could indicate a divergence in the regulation of this gene. It has been shown that 65% of the human genes contain TATA-less promoters and only the remaining 35% contain a minimal canonical TATA box (3.6%), or have a TATA box with one or two mismatch (4.7% and 27%, respectively) (Moshonov et al, 2008). Promoters that lack a TATA box can still direct transcription initiation from a specific nucleotide, but this feature has been associated with genes with multiple transcription initiation start sites (Moshonov et al., 2008), a characteristic of all *Runx* genes, including the zebrafish *runx*3. Two TSSs were reported for zebrafish *runx3*-P1 transcripts located at positions -130 and -460 bp (Kataoka et al, 2000) and we identified one more located at position -641 (chapter II of this thesis) from the start codon (**Figure 5.1A**). Furthermore, genes with TATA-less promoters that are regulated by CAAT boxes are usually housekeeping genes (Dynan and Tjian, 1985; Roy and Lee, 1995). Our *in silico* analysis failed to predict an

Sp1 binding site, which is usually associated with TATA-less promoters, in the *runx3*-P1 promoter analysed, but one Sp1 binding site was predicted in *runx3*-P2 region located at position -88 upstream the P2 TSS, in the proximity of the TATA and CAAT boxes (**Figure 5.1A** and **5.1B**). Several other TFBSs were predicted in the 1 kb upstream region of each zebrafish *runx3* promoter, such as Ap1, Ikaros, C-Ets, C/ebp, Nf-y, Runx, Nfat, Nkx, Sry, FoxD3, MyoD, among others (**Figure 5.1** and data not shown).

One interesting feature of the *Runx3* promoter regions is the identification of two very well conserved putative RUNX binding sites in the proximal region of the P1 promoter that are present in all *Runx* genes. This conservation has been seen not only among the different *Runx* genes, but also through evolution (Drissi et al, 2000; Bangsow et al, 2001; Nah et al, 2014). Our *in silico* analysis (**Figure 5.1A**) indicate initially the presence of five putative Runx binding sites in the *runx3*-P1 proximal promoter region, but from those, only two were still detected after our comparative analysis with the fugu *runx3*-P1 region (**Table 3.1**, in the chapter III of this thesis). Although zebrafish and fugu *runx3*-P1 proximal upstream regions still retain two conserved putative Runx binding sites, that feature is slightly different to what has been seen in mammals, where these sequences are located in tandem separated only by three nucleotides. However, a comparative analysis between 1 kb of the zebrafish and human *runx3*-P1 upstream regions show that at least one putative RUNX binding site is still conserved between the two species (**Figure 5.S5**), suggesting a similar regulation by Runx transcription factors as observed for the other species.

To the best of our knowledge, so far there have been no studies in the transcriptional regulation of zebrafish *runx3* isoforms. In this study, we cloned 5 kb and 4 kb from the translation start codon of 5' upstream regions of the zebrafish Runx3-MASN and Runx3-MHIPV isoforms, respectively, and performed *in vitro* analysis by transfection into different cell lines and *in vivo* analysis by microinjection into zebrafish embryos, to test if these putative promoter regions are transcriptionally functional. We confirmed that both fragments are active both *in vitro* (**Figure 5.2**, **5.4** and **Figure 5.S3**) and *in vivo* (**Figure 5.S1**). So, in order to identify regulatory regions in both promoters, we performed functional *in vitro* analysis by transfection of different cell lines from different origins with various deletion promoter constructs (**Figure 5.2** and **5.4**). This analysis confirmed that both

promoter constructs are functional in all cells tested, with P2-promoter constructs being more highly expressed than the P1- constructs when compared to the promoter-less control vector (**Figure 5.2** and **5.4**). Combining the transfection results from all cell lines tested, it was possible to identify multiple regulatory regions in *runx3*-P1 and *runx3*-P2 promoters (**Figure 5.2** and **5.4** and **Table 5.1**). In the 5 kb region of *runx3*-P1 promoter region tested, we identified four regions, two potential activating regions [(A1) from -5094 to -1766 and (A2) from -948 to -786] and two potential repressing regions [(R1) from -1766 to -948 and (R2) from -662 to -559] (**Figure 5.2A** and **5.4A** and **Table 5.1A**). It was also observed that the PD(-948/-662)LuC construct, that comprises the potential activating region A2 of *runx3*-P1 promoter, gave the highest relative expression in all cells tested, except in the U2OS cell line (this construct was not tested in Neuro2a cell line; **Figure 5.2A** and **5.4A** and **Table 5.1A**). Various putative TFBSs described as important regulators of transcriptional activity of promoters are detected in this regions (e.g. CAAT-box (V$CAAT), E-box (V$EBOX), Inre (O$INRE), Sal2 (V$SAL2), Miz1 (V$MIZ1); **Table 5.2**). In the 4 kb region of *runx3*-P2 promoter tested, two activating regulatory regions [(A1) from -1232 to -699 and (A2) from -554 to -474] and one repressing regulatory region [(R1) from -3693 to -1232] were observed (**Figure 5.2B** and **5.4B** and **Table 5.1B**). We observed that, besides these regions, none of the constructs used showed significant differences of relative expression between each other in all the cell lines tested, with the exception of the construct PP(-1232/-713)LuC that showed low activity in all cells tested (**Figure 5.2B** and **5.4B** and **Table 5.1B**). Interestingly, those potential activating/repressing regions identified in our study both in *runx3*-P1 and P2 promoters contain putative TFBSs, such as Creb (family V$CREB), Mitf ((family V$MITF), Brn3a (family V$BRNF), Rbp-j (family V$RBPF), Scl/Tal1 (family V$HAND), Stat (family V$STAT) and Runx (family V$HAML; **Table 5.2**), described in the literature as regulating the *RUNX3* gene and also identified by us in our *in silico* analysis (Chapter III of these thesis).

Our *in vitro* analysis also allowed us to identify the essential regions for the *runx3* promoter basal activity. For the *runx3*-P1 promoter, the essential region to drive basal activity is between -701 and -17, as the deletion of this region in PD(-5094/-701)LuC construct completely abolished the expression of luciferase to the level of the promoter-less control vector (**Figure 5.2A** and **5.4A**). Although, our *in silico* analysis predicted two putative CAAT box in this region, our *in vitro* assays indicated that probably these CAAT boxes are not

essential to drive the *runx3*-P1 promoter activity, since deletion of both CAAT boxes in the construct PD(-786/-662)LuC did not prevent activation of luciferase expression at levels significantly higher than the promoter-less vector (**Figure 5.2A** and **5.4A**). This hypothesis was confirmed in the cathepsin L gene by Jean and co-workers (2006) that have demonstrated that the CCAAT motif and GC boxes were not directly involved in the regulation of the high promoter activity observed in melanoma cells (Jean et al, 2006). More experiments will be needed to define exactly the region responsible for the basal activity of this promoter. Regarding the *runx3*-P2 promoter, we observed that the region from -713 to -554 is essential to drive basal activity, as deletion of that region in the construct PP(-1232/-713)LuC abolished the promoter activity (**Figure 5.2B** and **5.4B**). Although our comparative analysis did not detect any conserved blocks between zebrafish and fugu in the -950 bp region upstream the start codon, our *in silico* analysis on the zebrafish *runx3*-P2 promoter region predicted various putative TFBSs known as important regulators of transcription in this region (*e.g*. TATA box, CAAT boxes, Sp1, Ap1 and C/ebp) (**Figure 5.1B**). It would be interesting to further analyse if the identified TATA box is responsible for the basal activity of this promoter as, so far, no TATA box has been identified in *runx3* promoters in other species. Overall, these results indicate that the identified regions contain transcription factors or enhancer binding sites important for *runx3* transcription, confirming the identity of the regions of *runx3*-P1 and P2 promoters important for basal transcription and providing a first basis to correlate expression of distinct Runx3 variants with specific transcription factors affecting P1 or P2 promoter activity.

As described above, all *runx* genes have two conserved Runx binding sites in its P1 promoter region, a feature also observed in zebrafish *runx3*-P1 promoter region (**Figure 5.1** and **Table 5.2**). Furthermore, it has been confirmed in mammals that *RUNX* promoters are indeed cross-regulated by RUNX proteins (Brady el al, 2009; Levanon and Groner, 2004; Spender et al, 2005). In addition, both P1 and P2 promoters of *RUNX* genes have been shown to contain several dispersed putative RUNX binding sites (Ghozi et al, 1996; Ducy et al, 1999; Drissi et al, 2000; Bangsow et al, 2001; Otto et al, 2003). Based on those studies and on our *in silico* analysis of zebrafish *runx3* promoter regions, (**Figure 5.1**, **Table 5.1** and Chapter III of this thesis), we have evaluated both (i) the possible cross-regulation of zebrafish *runx3* promoters by Runx2 isoforms; (ii) the putative auto-regulation of *runx3* promoter by its own

isoforms; and (iii) if the Cbfβ co-regulator positively enhances that regulation, as shown for a variety of Runx target genes (Miller et al, 2002). The co-transfection analyses were performed in HEK293 cells, as these cells do not express any of the RUNX family members (Zheng et al, 2007; Bone et al, 2010), so the interference by endogenous RUNX proteins is eliminated. Our co-transfection results indicate that both zebrafish *runx3*-P1 and -P2 promoter fragments tested are regulated by at least one of the Runx2/3 isoforms (**Figure 5.3**). For *runx3*-P1 promoter only the Runx3-MASN isoform, in the presence of Cbfβ, induced a significant repression of the promoter activity, providing evidence that at least one of the Runx putative TFBSs is functional (**Figure 5.3A**). For *runx3*-P2 promoter a significant effect was observed following co-transfection with all Runx2 and Runx3 isoforms in the presence of the Cbfβ co-activator partner (except with the Runx3-MHIPV isoform, where an effect was observed only in the absence of Cbfβ) (**Figure 5.3B**) also indicating the functionality of at least one Runx binding site. However, further experiments are needed since these results represent only two independent experiments. Also, these experiments needs to be optimized as the co-transfection of *runx3*-P1 with the empty pCMX vector showed a significant decrease in the promoter activity (**Figure 5.3A**).

Finally, we have evaluated *in vivo* the expression pattern of each zebrafish *runx3* promoter. We have microinjected into zebrafish embryos the *runx3*-P1 construct PD(-5094/-17):EGFP and the *runx3*-P2 construct PP(-3930/-554):EGFP and evaluated their expression pattern by analysing the EGFP expression in transient transgenics (**Figure 5.5**). The transient *in vivo* analysis showed that the *runx3*-P2 fragment used to drive the EGFP expression reproduce, in part, the expression pattern observed for the endogenous *runx3* by RNA *in situ* hybridization (chapter II of this thesis; Kataoka et al, 2000; Kalev-zylinska et al, 2003).

Although the analysis of the transient embryos should be done in more detail to identify the exact cells expressing the EGFP, we tried to identify these cells by its location and morphology. The *runx3*-P2 promoter fragment induces EFGP expression in retinal neuroepithelial cells in the eye, otic epithelium, cranial ganglia, Rohon-Beard cells, Schwann cells of the spinal nerves and of the dorsal branch of the posterior lateral line nerve (plln), and migratory Neural Crest cells (**Figure 5.5**), a profile of expression similar to the endogenous *runx3* expression pattern (chapter II of this thesis). The expression in cells

within the heart and in muscle, need further confirmation. Our *in vitro* analysis showed that *runx3*-P2 is highly active in all cells tested, suggesting the presence of an enhancer region promoting expression in a variety of tissues. Accordingly, we have previously shown by *runx3* isoform-specific qPCR that *runx3*-P2 transcripts are expressed in all zebrafish tissues tested (chapter II of this thesis).

Regarding the transient *in vivo* analysis in embryos injected with the *runx3*-P1 construct, few or no cells were detected expressing GFP (**Figure 5.5**). This may indicate that (i) this construct possible lacks an important enhancer to be able to drive its expression up to detectable levels of EGFP *in vivo*, or (ii) this promoter fragment is driving EGFP expression in a very strict tissue-specific way, and the method used to visualise the GFP expression was not sensitive enough to detect positive cells.

In conclusion, we performed an extensive analysis of the *runx3*-P1 and *runx3*-P2 upstream regions, identifying for each *runx3* promoter the region responsible for its basal activity. Furthermore, we identified different putative positive and negative regulatory regions for this gene and a variety of putative TFBSs that can be responsible for the regulation of those regulatory regions. We showed that the Runx2 and Runx3 isoforms are involved in the regulation of both *runx3* promoter regions and that this regulation can be enhanced by their co-operation with Cbfβ, although these results need to be confirmed. We also confirmed that the *runx3*-P2 major fragment is functionally active *in vivo*, and drives expression of EGFP, recapitulating at least partially, the endogenous *runx3* expression pattern. On the other hand, we did not detect EGFP expression using the *runx3*-P1 major fragment. Nonetheless, we cannot exclude the possibility that this fragment drives EGFP in a tissue-specific way in a very small subset of cells that were not detected in our transient analysis, which is usually mosaic. This result is in agreement with (i) our *in vitro* assays, where this promoter region showed, in general, a low activity, (ii) our mRNA *in situ* hybridization studies where we could not detect any signal when a riboprobe specific for P1-derived transcripts was used (chapter II of this thesis). For this reason, the generation of stable zebrafish transgenic lines is an essential tool to clearly identify sites of expression. This experiment is ongoing.

Acknowledgments

This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Program and national funds through FCT – Foundation for Science and Technology, under the project "PEst-C/MAR/LA0015/2011. NC and BS are supported, respectively, by a post-doctoral and doctoral grant from FCT (SFRH/BPD/48206/2008 and SFRH/BD/38083/2007).

5.5 Supplementary Figures

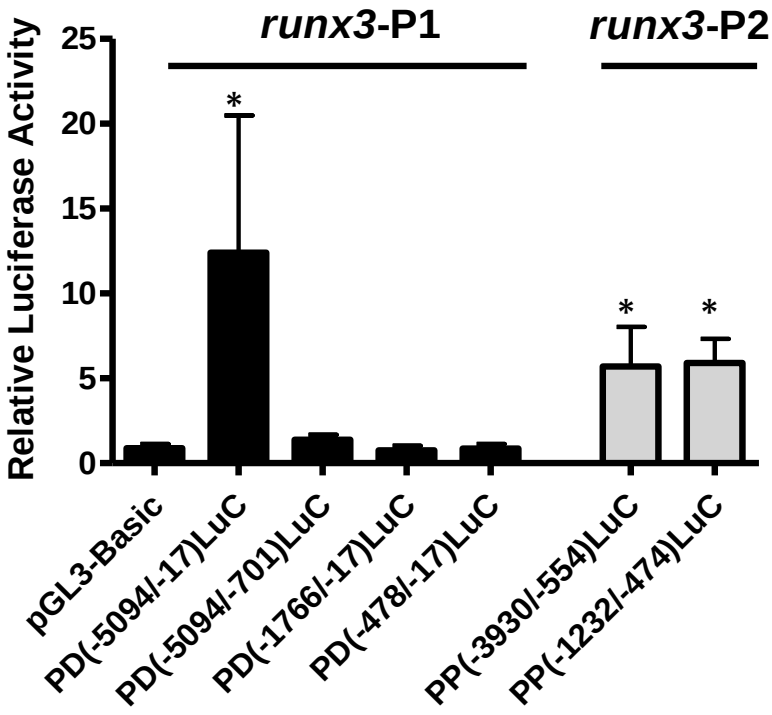


Figure 5.S1 Both *runx3*-P1 and *runx3*-P2 luciferase promoter constructs are functional *in vivo*. Relative luciferase activity obtained by microinjection of *runx3* promoter constructs into 1-cell stage zebrafish embryos. Firefly and renilla luciferase activities were measured 24 hours post injection. n≥4.

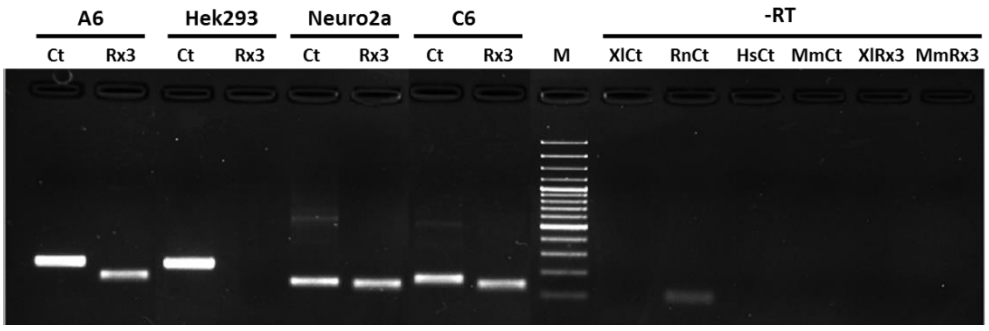


Figure 5.S2 Qualitative analysis of *Runx3* amplification by RT-PCR in different cell lines shows that *Runx3* is expressed in all cells tested, except in HEK293 cell line. Xenopus Actin and mouse, rat and human GAPDH were used as control (Ct) for sample integrity. Rx3 means Runx3. The right side of the gel corresponds to the negative controls (no cDNA). M corresponds to the marker (Thermo Scientific GeneRuler 100bp Plus DNA Ladder). Xl – *Xenopus laevis*; Rn – *Rattus norvegicus*; Mm – *Mus musculus*.

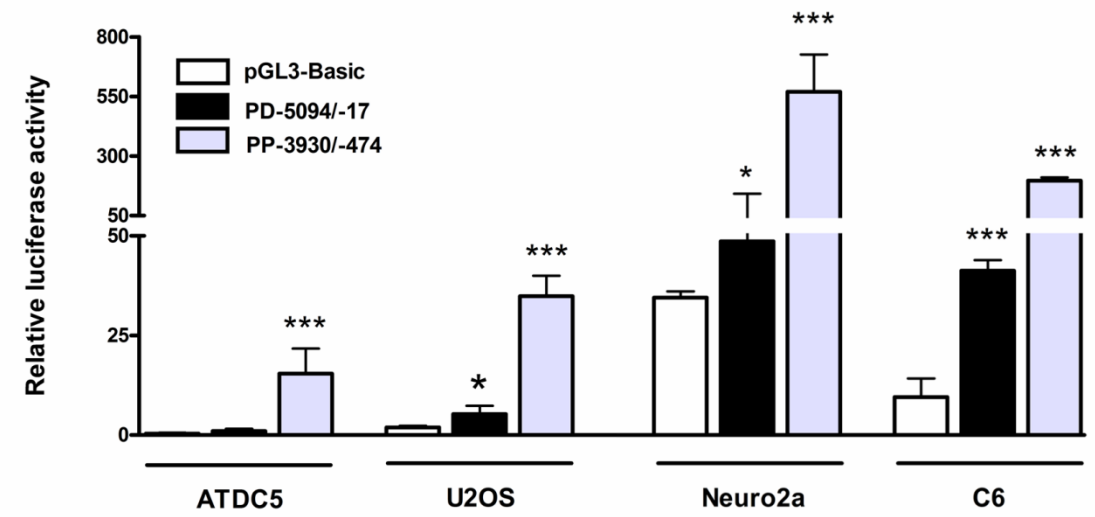


Figure 5.S3 Basal activity of zebrafish *runx3*-P1 and *runx3*-P2 promoter regions in various *runx3* expressing cells. Mouse chondrogenic ATDC5, human osteosarcoma U2OS, human neuroblastoma Neuro2a and rat glioma C6 cell lines were transiently transfected with *runx3*-P1 [PD(-5094/-17)] and *runx3*-P2 [PP(-3930/-474)] promoter constructs. The promoter-less pGL3-Basic vector was used as control and significance (*) compared to the baseline was determined by One Way ANOVA followed by Tukey's test. All samples were tested in duplicates at least three times. *P<0.05; **P<0.01; ***P<0.001.

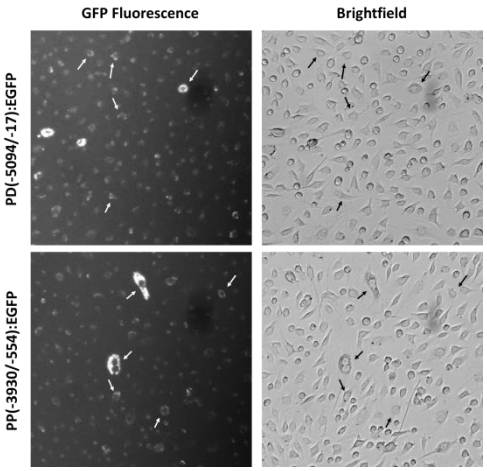


Figure 5.S4 Transfection of *runx3*-P1 and *runx3*-P2 promoter constructs driving EGFP results is GFP expressing cells. A6 cells were transfected with *runx3*-P1 [PD(-5094/-17):EGFP] and *runx3*-P2 [PP(-3930/-554):EGFP] promoter constructs and analysed for GFP expression 24 hours post-transfection under an inverted fluorescence microscope (Olympus IX81 fluorescence microscope coupled with an F-view II camera). Arrows indicates cells expressing GFP.

Alignment (DiAlign format):

Please note that only **upper-case** letters are considered to be **aligned**. For more information, please have a look at the user guide.

V$ZF35 V$MOKF V$NKXH V$AIRE V$ETSF V$E2FF V$CDEF V$EGRF O$XCPE V$AP4R V$IKRS V$MYOD
V$MYT1 V$PAX5 V$AP2F V$AP1R V$HAND V$STAT V$EBOX V$HAML V$HOXF V$SORY V$KLFS V$CEBP
V$DLXF V$HOMF V$BRNF V$PAXH V$CAAT V$HEAT V$FKHD V$RUSH V$GABF V$MIZ1 V$LEFF V$SP1F
V$ZF5F

```
zebrafish  agagttttgtg tatgagctgt atgttttttgt ggtccccTCT TCAAGAAGGT
human      ----------- ----------- ----------- ------- -TCT TCTAAAAGGC
                                                        *** **********

zebrafish  CCCATCATCt tct gggaaaac tcttcaaaaca CCAGCCAtgg gccaatcaga
human      CGTGACATCa cg gcccaggt gaccgcggc- CCAGCCAatg agccaaggcc
           ********* ** ******** ********* ******* *******

zebrafish  tgagtgagaa agg-------- ----------- ----------- --CACTGGGA
human      gcgagcaggc ttctcgcatc ctgtgagctg aggttgggtt gaCACTGGGA
                                                        ********

zebrafish  AGaaaaggag aactt gaatg catgcaagcg agggagagag agaacgtcat
human      AGgcctggtc cctca accac agaaccacaa ggccaggccc ttgccgcctc
           **                                                    *

zebrafish  gtgactggag ttccacctat cagatcatgg gattggcgag cagcagtgat
human      cagggccctg ----------- ----------- ----------- -----g-ccc

zebrafish  acCGCGGGAG CTGtgacctc g--------- ----------- ----------
human      tgCGCGGGAG CTGgttggct cctggtgctc ccccaccccg gccgccctcg
           ********* *** *                                      

zebrafish  ---------- ----------- ----------- ----------- ----------
human      taccaccag agcctgggct ctgtcaaggg taagcctcat tcattcattc
```

Figure 5.S5. TFBS families detected by DiAlignTF software that are conserved in the alignment of the *runx3*-P1 promoter regions between zebrafish and human. The list of the TFBSs obtained by DiAlignTF and the alignment showing the TFBSs conserved is represented. The colour code for each specific TFBS is shown above the alignment.

5.6 Supplementary Tables

Table 5.S1 PCR primers used to amplify the zebrafish *runx3* promoter regions.

Sense primers	Primer sequence (5' to 3')[#]
GWZfRunx3_F1	ATCCATCCATTCATTCACCCATCCATCC
DrRunx3PD_F2	TGGCTGTCCAGGGGAGAGGGCTCTGAGC
DrRunx3PP_F4	TTTCAGGGCTTCTCATTGGCCACGT
DrRunx3PD_KpnI_F1	CCGGAGGGTACCATAGGCCGGCCGCAACCATGTCAAAC
DrRunx3PD_KpnI_F2	CCGGAGGGTACCCCCCAAAACCACCAAAACACACTCG
DrRunx3PD_KpnI_F3	CCGGAGGGTACCGGATTGGCGAGCAGCAGTGATACCG
DrRunx3PD_KpnI_F5	AGGGTACCTTTGGAAAAAGTCAGAAAAGTGGGGGTG
Runx3_PPF1_KpnI	CCGGAGGGTACCTATTTTGTCTGTGGTTTTAAAGGTT
DrRunx3PPKpnI_F3	AGGGTACCGCGACTGTAACCGTAGAAACTGCG
Antisense primers	**Primer sequence (5' to 3')[#]**
GWZfRunx3_R1	GCGGTATCACTGCTGCTCGCCAATCCCA
GWZfRunx3_R3	GTCAAGCGTTCTTTCCAAATGAATGCGTAA
GWZfRunx3_R4	GCGAGTTCCCCACTGACTTTAGCCACACAC
GWZfRunx3_R5	CCCAGTGCCTTTCTCACTCATCTGATTG
DrRunx3II&III_R2	CGCAGCAAACTGGGCGAGTAGCTAGAGA
DrRunx3_R4	CGTCTGCTCGTGCTCGGGTCT
DrRunx3PD_XhoI_R1	CACGCCTCGAGGGAAAAGAGCTGGCCTCTCCGC
DrRunx3PD_XhoI_R2	CACGCCTCGAGGCTCATACACAAACTCTTGAGTAAA
DrRunx3PD_XhoI_R5	GCCTCGAGATGAAGGATAGCTGAGGATTATTAGGGA
Runx3IntExR1_XhoI	CACGCCTCGAGCGCGAACGCCGCGCTCTGCACTCT
Runx3PP_R1_XhoI	CACGCCTCGAGGAGTCCTCGTGAAAAGCTCGGGCA
Universal primers	**Primer sequence (5' to 3')**
Adaptor Primer 1 (AP1)	GTAATACGACTCACTATAGGGC
Adaptor Primer 2 (AP2)	ACTATAGGGCACGCGTGGT
pGL3_FW2	TAGCAAAATAGGCTGTCCCC
oligo(dT)-adapter primer	ACGCGTCGACCTCGAGATCGATGTTTTTTTTTTTTT

[#] indicates the recognition sequence for the restriction enzyme named on the primer.

Table 5.S2 Genes and respective set of primers used for the RT-PCR performed in the cell lines.

Cell line	Specie	Gene	Primers	Primer sequence (5' to 3')
A6	*Xenopus laevis*	**Runx3**	XlRunx3_F	CTGCGGAAAGATGGGAGAGAACAACG
			XlRunx3_R	TGGGACGGCAGAACTGAGCACAA
		Actin	xAct_F	ATGCTCCCCGTGCTGTTTTCCCATCTAT
			xAct_R	TTCTGGTGCCACTCGCAGTTCATTGTAG
HEK293	*Homo sapiens*	**Runx3**	MmRunx3_F	CACAATCACCGTGTTCACCAACCCTA[#]
			MmRunx3_R	GGCCTTGGTCTGGTCTTCTATCTTCTGC[#]
		GAPDH	Hs_GAPDH_F	TCAACGGATTTGGTCGTATTGGGCG
			Hs_GAPDH_R	CTCGCTCCTGGAAGATGGTGATGGG
Neuro2a	*Mus musculus*	**Runx3**	MmRunx3_F	CACAATCACCGTGTTCACCAACCCTA
			MmRunx3_R	GGCCTTGGTCTGGTCTTCTATCTTCTGC
		GAPDH	MmGAPDH_F	CCTTCCGTGTTCCTACCCCCAATGT
			MmGAPDH_R	AGTGTAGCCCAAGATGCCCTTCAGT
C6	*Rattus norvegicus*	**Runx3**	MmRunx3_F	CACAATCACCGTGTTCACCAACCCTA[#]
			MmRunx3_R	GGCCTTGGTCTGGTCTTCTATCTTCTGC[#]
		GAPDH	RnGAPDH_F	CGGCAAGTTCAACGGCACAGTCAAG
			RnGAPDH_R	GAAGACGCCAGTAGACTCCACGACAT

The mouse primers were used as the sequences are very similar. The different nucleotides are underlined. GAPDH - glyceraldehydes-3-phosphate dehydrogenase.

5.7 References

Blyth K, Cameron ER, Neil JC, 2005. The runx genes: gain or loss of function in cancer. Nat Rev Cancer. 5:376-387.

Bone KR, Gruper Y, Goldenberg D, Levanon D, Groner Y, 2010. Translation regulation of Runx3. Blood Cells, Molecules, and Diseases. 45:112-116.

Brady G, Whiteman HJ, Spender LC, Farrell PJ, 2009. Downregulation of RUNX1 by RUNX3 requires the RUNX3 VWRPY sequence and is essential for Epstein-Barr virus-driven B-cell proliferation. J Virol. 83:6909-6916.

Burns CE, DeBlasio T, Zhou Y, Zhang J, Zon L, Nimer SD, 2002. Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. Exp Hematol. 30:1381-1389.

Chen AI, de Nooij JC, Jessell TM, 2006. Graded activity of transcription factor Runx3 specifies the laminar termination pattern of sensory axons in the developing spinal cord. Neuron. 49:395-408.

Coffman JÁ, 2003. Runx transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol Int. 27:315-324.

Downing JR, 1999. The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance. Br J Haematol. 106:296-308.

Drissi H, Luc Q, Shakoori R, Lopes SCS, Choi JY, Terry A, Hu M, Jones S, Neil JC, Lian JB, Stein JL, Van Wijnen AJ, Stein GS, 2000. Transcriptional autoregulation of the bone related CBFA1/RUNX2 gene. J Cell Physiol. 184:341-350.

Dynan WS, Tjian R, 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature. 316:774-778.

Ito Y, 2004. Oncogenic potential of the RUNX gene family: 'overview'. Oncogene. 23:4198-4208.

Ito Y, 2008. RUNX genes in development and cancer: regulation of viral gene expression and the discovery of RUNX family genes. Adv Cancer Res. 99:33-76.

Jean D, Rousselet N, Frade R, 2006. Expression of cathepsin L in human tumor cells is under the control of distinct regulatory mechanisms. Oncogene. 10:1474-1484.

Lai CB, Mager DL, 2012. Role of Runt-related Transcription Factor 3 (RUNX3) in Transcription Regulation of Natura l Cytotoxicity Receptor 1 (NCR1/NKp46), an Activating Natural Killer (NK) Cell Receptor. J Biol Chem. 287:7324-7334.

Levanon D, Groner Y, 2004. Structure and regulated expression of mammalian RUNX genes. Oncogene. 23:4211-4219.

Kataoka H, Ochi M, Enomoto K, Yamaguchi A, 2000. Cloning and embryonic expression patterns of the zebrafish Runt domain genes, runxa and runxb. Mech Dev. 98:139-143.

Kawakami K, Koga A, Hori H, Shima A, 1998. Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio. Gene. 225:17-22.

Kawakami K, 2007. Tol2: a versatile gene transfer vector in vertebrates. Genome Biol. Suppl 1:S7.

Komori, T, 2005. Regulation of skeletal development by the Runx family of transcription factors. J Cell Biochem. 95:445-453.

Komori T, 2015. The functions of Runx family transcription factors and Cbfb in skeletal development. Oral Science International. 12:1-4

Kramer I, Sigrist M, de Nooij JC, Taniuchi I, Jessell TM, Arber S, 2006. A role for Runx transcription factor signaling in dorsal root ganglion sensory neuron diversification. Neuron. 49:379-393.

Miller J, Horner A, Stacy T, Lowrey C, Lian JB, Stein G, Nuckolls GH, Speck NA, 2002. The core-binding factor beta subunit is required for bone formation and hematopoietic maturation. Nat Genet. 32:645-649.

Moshonov S, Elfakess R, Golan-Mashiach M, Sinvani H, Dikstein R, 2008. Links between core promoter and basic gene features influence gene expression. BMC Genomics. 9:92.

Nah GS, Lim ZW, Tay BH, Osato M, Venkatesh B, 2014. Runx family genes in a cartilaginous fish, the elephant shark (*Callorhinchus milii*). PLoS One. 9:e93816.

Ng CE, Osato M, Tay BH, Venkatesh B, Ito Y, 2007. cDNA cloning of Runx family genes from the pufferfish (Fugu rubripes). Gene. 399:162-173.

Otto F, Lübbert M, Stock M, 2003. Upstream and downstream targets of RUNX proteins. J Cell Biochem. 89:9-18.

Roy B, Lee AS, 1995. Transduction of calcium stress through interaction of the human transcription factor CBF with the proximal CCAAT regulatory element of the grp78/BiP promoter. Mol Cell Biol. 15:2263-2274.

Spender LC, Whiteman HJ, Karstegl CE, Farrell PJ, 2005. Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells. Oncogene. 24:1873-1881.

van Wijnen AJ, Stein GS, Gergen JP, Groner Y, Hiebert SW, Ito Y, Liu P, Neil JC, Ohki M, Speck N, 2004. Nomenclature for Runt-related (RUNX) proteins. Oncogene. 23:4209-4210.

Wigner NA, Soung do Y, Einhorn TA, Drissi H, Gerstenfeld LC, 2013. Functional role of Runx3 in the regulation of aggrecan expression during cartilage development. J Cell Physiol. 228:2232-2242.

Westendorf JJ, Hiebert SW, 1999. Mammalian runt-domain proteins and their roles in hematopoiesis, osteogenesis, and leukemia. J Cell Biochem Suppl. 32-33:51-58.

Westerfield M, 2000. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene.

Zheng L, Iohara K, Ishikawa M, Into T, Takano-Yamamoto T, Matsushita K, Nakashima M, 2007. Runx3 negatively regulates Osterix expression in dental pulp cells. Biochem J. 405:69-75.

Chapter 6

General conclusions and future perspectives

6.1 General conclusions and future perspectives

The RUNX family of transcription factors has been extensively studied over the last years, due to the involvement of its members in a variety of biological processes and diseases, and each have been shown to be regulated by two alternative promoters (P1 and P2) that generate different transcript variants; these are also subject to alternative splicing events generating multiple RUNX protein isoforms (Coffman, 2003; Komori, 2005; Cohen, 2009; Wang et al, 2010; Chuang et al, 2013; among others).

In an attempt to clone the *runx3* open reading frames (ORFs), we identified and cloned multiple transcript variants. Besides the three different transcripts previously described for zebrafish *runx3* (Kataoka et al, 2000), we have cloned 10 additional variants obtained by alternative splicing. From the new variants cloned, four are P1-derived (isoforms 2 to 5) and six are P2-derived (isoforms 7 to 12). Although this study focused only on the analysis of the two major Runx3 variants, it would be interesting to further analyse if the different protein isoforms generated by the splicing variants are functional, since some isoforms lack certain domains but contain others. This may suggest that these isoforms have additional functions different from those previously described, such as dominant negative or constitutively active proteins. This hypothesis is supported by the findings of Runx1 C-terminal deletion mutants in mouse and grass carp, which lack one or more domains, and have been shown to have different transactivation potentials and different roles in biological processes (Tanaka et al, 1995; Kawazu et al, 2005; Yao et al, 2014). For example, in mouse, one variant that lacks the TAD acts as a natural negative isoform repressing the function of the non-mutated form of Runx1 (Tanaka et al, 1995). The existence of several alternatively spliced isoforms has been shown to increase cellular and functional complexity in many species, due to the different functions of the splicing variants (Reviewed in Kelemen et al, 2013). We also observed two alternatively spliced *runx3*-P1 5' UTR variants (isoforms 13 and 14). Both *runx3*-P1 and *runx3*-P2 5' UTR regions contain multiple upstream ATG triplets that may generate several putative upstream ORFs (uORFs) that vary drastically in size. These features highlight the complexity of both the molecular structure and the regulatory mechanism of the *runx3* gene. The functional significance of *runx3* regulatory regions has not been elucidated, but different authors have tested the functionality of the uORFs and have shown that their

presence is implicated in the regulation of the translation of the main ORF (van der Velden and Thomas, 1999; Wang and Rothnagel, 2004) and that the size of these uORFs can affect differently the translation of the protein, with longer uORFs having a stronger effect in preventing ribosome reinitiation (Kozak, 2001).

We then investigated the expression patterns of the *runx3*-P1 and *runx3*-P2 derived variants by qPCR in different stages of zebrafish development and in different adult tissues, concluding that both variants are differentially expressed. The *runx3*-P2 variant seems to be ubiquitously expressed and zygotically specific, in contrast to the *runx3*-P1 variant that is maternally expressed and continues to be expressed zygotically, but its expression seem to be tissue-specific.

To identify the spatial-temporal expression of the *runx3* transcripts we performed ISH at different zebrafish embryonic developmental stages, using RNA specific probes that detect i) the common region of all variants, and ii) specifically the 5' UTR of each variant. Our results with a probe targeting the common 3' region of the transcripts confirmed the *runx3* expression pattern in neural and cartilaginous tissues as previously reported (Kataoka et al, 2000; Burns et al, 2002; Kalev-Zylinkska et al, 2003; Flores et al, 2006), although we could not detect expression in hematopoietic tissues as observed in a previous report (Kalev-Zylinkska et al, 2003). The zebrafish *runx3* expression patterns seem to be conserved thoughout evolution, since the expression of *Runx3* in hematopoietic, neural and cartilagineous tissues was also observed in other lower vertebrates, such as the xenopus (Park and Saint-Jeannet, 2010), and in higher vertebrates, such as the mouse (Levanon et al, 2001). However, our ISH results using isoform specific probes were not conclusive about the isoform specific expression patterns and their sensitivity needs to be improved, using for example more powerful tools that would allow us to amplify the hybridisation signal and decrease the background. For this, multiple transcript detection by RNAscope technology, a new ISH technique recently optimized for zebrafish (Wang et al, 2012), could be a great tool to help us determine the exact cells where each variant is expressed.

Our preliminary results of the functional analysis targeting the Runx3 isoforms using translational morpholinos, together with results published by other groups, strongly suggest that each isoform has a different function that is cell-type specific. Morpholinos have been

extensively used in zebrafish to study the effects of the knockdown of a gene, but recently some authors have shown that sometimes morpholinos results do not completely reproduce the knockout results (Eisen and Smith, 2008; Kok et al, 2015). This, plus the unexpected nature of some of the defects we have observed, raised clear questions over the validity of the preliminary morpholino data we obtained. Since we performed those experiments, a new knockout technique was developed (the CRISPR-Cas9 system) and shown to be a simple and highly efficient method to mutate a gene of interest (Hwang et al, 2013a; 2013b; Xiao et al, 2013; Jao et al, 2013; among others). The CRISPR-Cas9 system requires the co-expression of a target-specific RNA (CRISPR-guide RNA) to recognize the endogenous target and a Cas9 protein that unwinds the DNA duplex and cleaves both strands. Thus, this novel technique provides a method that could be readily applied to target the endogenous *runx3* locus and test our morpholino results to determine the resulting phenotype of each isoform knockout.

All three RUNX proteins bind and recognize the same DNA sequence, and while in some tissues their expression overlaps, they have all been shown to have distinct tissue-specific expression patterns, indicating that their regulation must be tightly controlled. RUNX3 has been extensively investigated in the last years for its capacity to regulate transcription of a variety of target genes, however, data related to transcriptional regulation of *RUNX3* itself remains scarce. To gain insight into the regulatory mechanism of the *runx3* gene, an *in silico* comparative analyses of the *runx3*-P1 and *runx3*-P2 promoter regions was performed, identifying numerous regions in each promoter that contain putative TFBSs that may serve as targets for sequence-specific enhancer/silencer transcription factors likely to contribute to regulation of the *runx3* transcription. Our *in silico* analysis successfully predicted binding sites for TFs already known from work in mammals as transcriptional regulators of *Runx3*, but also predicted novel TFs potentially with a role in *runx3* regulation. We then cloned these promoter regions and tested their transcriptional activity, confirming that both regions are functionally active *in vitro* and *in vivo*. Together, those results allowed us to define positive and negative regulatory regions in both *runx3* promoters, and identify the putative TFBSs in those regions predicted by our *in silico* analysis, providing a powerful tool to guide future dissection of *runx3* transcriptional regulation *in vitro* and/or *in vivo*.

One interesting feature of the *RUNX3* promoter regions is the identification of two evolutionarily conserved putative Runx binding sites in the proximal region of the P1

promoter that are present in all *Runx* genes (Drissi et al, 2000; Bangsow et al, 2001; Nah et al, 2014). CBFβ is known to heterodimerize with all Runx family members and enhance their transcriptional regulatory effect upon target genes. In this study, we cloned and characterized 10 novel zebrafish *cbfβ* isoforms, which are generated by alternative splicing, and structural conservation during evolution from fish to mammals was confirmed by a comparative analysis between zebrafish *cbfβ* gene and protein and its orthologs in different species. The expression of four of the *cbfβ* splicing variants (isoforms 1 to 4) was analysed by RT-PCR in different zebrafish developmental stages and in adult tissues showing that *cbfβ* is widely expressed, being detected in all samples analysed. However, our data demonstrate that *cbfβ* isoforms 1 and 2 are maternally expressed, but not isoforms 3 and 4, suggesting variances in the biological function of Cbfβ splice variants. The transcriptional activity of these four *cbfβ* splicing isoforms that differ in the presence or absence of the exons 5a and 5b, was evaluated by luciferase reporter assay, using a fragment of the zebrafish collagen type Xα1 promoter previously shown to be regulated by Runx2-MASN isoform (Simões et al, 2006). The results showed that zebrafish Cbfβ isoforms carrying the exon 5b (isoforms 1 and 2) have a higher capacity to promoteinduction of *ColX* promoter by Runx2-MASN isoform, compared to the isoforms lacking exon 5b (isoforms 3 and 4). Although we confirmed by immunoprecipitation the direct interaction of Runx2 with Cbfβ isoforms, further experiments are needed to clarify the biological importance of these findings. For example, mass spectrometry (MS) can be used for each evaluated isoform (1-4) in order to conclude if these different amino acids in the C-terminal region can be interacting with different co-activators involved in the transcription complex. For this, specific oligos are tagged, anneled and incubated with cell nuclear extracts. Then, the binding mixture is coupled to magnetic beads to alow for efficient affinity pull-down of the putative binding proteins. After purification, the sample containing the DNA binding proteins from the cell nuclear extract is digested, usually with trypsin, and the resulting product is analysed by MS for protein identification.

Our comparative *in silico* analysis predicted putative RUNX binding sites in both *runx3*-P1 and *runx3*-P2 promoter regions and *in vitro* co-transfection analysis with Runx2 and Runx3 isoforms showed that both *runx3* promoter regions tested are regulated by at least one of the Runx2/3 isoforms. Furthermore, this regulation is further enhanced by the presence of

the co-factor Cbfβ. These results showed that zebrafish *runx3* promoter regions are cross- and auto-regulated by Runx2 and Runx3 isoforms, respectively, and indicate the functionality of at least one Runx TFBS predicted in both regions. Further work should be done, for example using deletion mutants of the promoters regions, to identify the region responsible for the Runx binding. Then by site-directed mutagenesis, it would be possible to mutate the Runx binding site and confirm the exact location responsible of the binding. Chromatin immunoprecipitation (ChIP) assays could also be used to investigate the interaction between the Runx2/3 and *runx3* promoter regions in the cell.

In conclusion, in this work we developed new molecular tools and described new findings that contribute to better understand the complex regulation and function of the zebrafish *runx3* gene. Although more work needs to be performed to clarify the specific function and contribution of each Runx3 isoform to the embryonic development, important progress has been made.

6.2 References

Bangsow C, Rubins N, Glusman G, Bernstein Y, Negreanu V, Goldenberg D, Lotem J, Ben-Asher E, Lancet D, Levanon D, Groner Y, 2001. The RUNX3 gene-sequence, structure and regulated expression. Gene. 279:221-232.

Burns CE, DeBlasio T, Zhou Y, Zhang J, Zon L, Nimer SD, 2002. Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. Exp Hematol. 30:1381-1389.

Chuang LSH, Ito K, Ito Y, 2013. RUNX family: Regulation and diversification of roles through interacting proteins. Int J Cancer. 132:1260-1271.

Coffman JA, 2003. Runx transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol Int. 27:315-324.

Cohen MM Jr, 2009. Perspectives on RUNX genes: An update. Am J Med Genet Part A 149A:2629-2646.

Drissi H, Luc Q, Shakoori R, Chuva De Sousa Lopes S, Choi JY, Terry A, Hu M, Jones S, Neil JC, Lian JB, Stein JL, Van Wijnen AJ, Stein GS, 2000. Transcriptional autoregulation of the bone related CBFA1/RUNX2 gene. J Cell Physiol. 184:341-350.

Eisen JS, Smith JC, 2008. Controlling morpholino experiments: don't stop making antisense. Development. 135:1735-1743.

Flores MV, Lam EYN, Crosier P, Crosier K, Fisher S, 2006. A Hierarchy of Runx transcription factors modulate the onset of chondrogenesis in craniofacial endochondral bones in zebrafish. Dev Dyn. 235:3166-3176.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK, 2013a. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 31:227-229.

Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, Yeh JR, 2013b. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS One. 8:e68708.

Jao LE, Wente SR, Chen W, 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci USA. 110:13904-13909.

Kalev-Zylinska ML, Horsfield JA, Flores MV, Postlethwait JH, Chau JY, Cattin PM, Vitas MR, Crosier PS, Crosier KE, 2003. Runx3 is required for hematopoietic development in zebrafish. Dev Dyn. 228:323-336.

Kataoka H, Ochi M, Enomoto K, Yamaguchi A, 2000. Cloning and embryonic expression patterns of the zebrafish Runt domain genes, runxa and runxb. Mech Dev. 98:139-143.

Kawazu M, Asai T, Ichikawa M, Yamamoto G, Saito T, Goyama S, Mitani K, Miyazono K, Chiba S, Ogawa S, Kurokawa M, Hirai H, 2005. Functional domains of Runx1 are differentially required for CD4 repression, TCRbeta expression, and CD4/8 double-negative to CD4/8 double-positive transition in thymocyte development. J Immunol. 174:3526-3533.

Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S, 2013. Function of alternative splicing. Gene. 514:1-30.

Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, Kirchmaier BC, Peterson-Maduro J, Kourkoulis G, Male I, DeSantis DF, Sheppard-Tindell S, Ebarasi L, Betsholtz C, Schulte-Merker S, Wolfe SA, Lawson ND, 2015. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell. 32:97-108.

Komori T, 2005. Regulation of Skeletal Development by the Runx Family of Transcription Factors. J Cell Biochem. 95:445-453.

Kozak M, 2001. Constraints on reinitiation of translation in mammals. Nucleic Acids Res. 29:5226-5232.

Levanon D, Brenner O, Negreanu V, Bettoun D, Woolf E, Eilam R, Lotem J, Gat U, Otto F, Speck N and Groner Y, 2001. Spatial and temporal expression pattern of Runx3 (Aml2) and Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis. Mech Dev. 109:413-417.

Nah GS, Lim ZW, Tay BH, Osato M, Venkatesh B, 2014. Runx family genes in a cartilaginous fish, the elephant shark (Callorhinchus milii). PLoS One. 9:e93816.

Park B-Y, Saint-Jeannet J-P, 2010. Expression analysis of Runx3 and other Runx family members during Xenopus development. Gene Expr Patterns. 10: 159-166.

Simões B, Conceição N, Viegas CSB, Pinto JP, Gavaia PJ, Hurst LD, Kelsh RN, Cancela ML, 2006. Identification of a promoter element within the zebrafish colXα1 gene responsive to Runx2 Isoforms Osf2/Cbfa1 and til-1 but not to pebp2aA2. Calcif Tissue Int. 79:230-244.

Tanaka T, Tanaka K, Ogawa S, Kurokawa M, Mitani K, Nishida J, Shibata Y, Yazaki Y, Hirai H, 1995. An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. EMBO J. 14:341-350.

van der Velden AW, Thomas AA, 1999. The role of the 5' untranslated region of an mRNA in translation regulation during development. Int J Biochem Cell Biol. 31:87-106.

Wang CQ, Jacob B, Nah GS, Osato M, 2010. Runx family genes, niche, and stem cell quiescence. Blood Cells Mol Dis. 44:275-286.

Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ, Luo Y, 2012. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J Mol Diagn. 14:22-29.

Wang XQ, Rothnagel JA, 2004. 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Res. 32:1382-1391.

Xiao A, Wang Z, Hu Y, Wu Y, Luo Z, Yang Z, Zu Y, Li W, Huang P, Tong X, Zhu Z, Lin S, Zhang B, 2013. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic Acids Res. 41:e141.

Yao F, Liu Y, Du L, Wang X, Zhang A, Wei H, Zhou H, 2014. Molecular identification of transcription factor Runx1 variants in grass carp (*Ctenopharyngodon idella*) and their responses to immune stimuli. Vet Immunol Immunopathol. 160:201-208.