View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Bath Research Portal

Citation for published version:
Simões, B, Conceição, N, Matias, AC, Bragança, J, Kelsh, RN & Cancela, ML 2015, 'Molecular characterization of cbf gene and identification of new transcription variants: Implications for function', Archives of Biochemistry and Biophysics, vol. 567, pp. 1-12. https://doi.org/10.1016/j.abb.2014.12.023

DOI:
10.1016/j.abb.2014.12.023

Publication date:
2015

Document Version
Peer reviewed version

Link to publication

Publisher Rights
CC BY-NC-ND

Published version available via: http://dx.doi.org/10.1016/j.abb.2014.12.023

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. May. 2019

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

B. Simões[1,2,3,#], N. Conceição[3,#], A.C. Matias[1,5], J. Bragança[1,5], R.N. Kelsh[4], M.L. Cancela[1,3]*

[1]Department of Biomedical Sciences and Medicine/DCBM, University of Algarve, Faro, Portugal

[2]PhD program in Biomedical Sciences, University of Algarve, Faro, Portugal

[3]Centre of Marine Sciences, University of Algarve, Faro, Portugal

[4]Department of Biology and Biochemistry and Centre for Regenerative Medicine, University of Bath, Claverton Down, United Kingdom

[5]Centre for Molecular and Structural Biomedicine, University of Algarve, Faro, Portugal

*Corresponding author:

Leonor Cancela, University of Algarve/DCBM-CCMAR, Campus de Gambelas, 8005-139

Faro, Portugal

Fax: (+351)289800076; Phone: (+351)289800971; E-mail: lcancela@ualg.pt

# These authors contributed equally to this work.

Summary

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 regulator of bone-related gene transcription that contains a conserved DNA-binding 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 zebrafish *cbfβ* gene transcripts, one resulting in the known Cbfβ protein (with 187 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 (resulting in a protein with 188 aa). Functional analysis of these isoforms provides insight into their role in regulating gene transcription. From the other variants two are premature termination Cbfβ forms, while the others show in-frame exon-skipping causing changes in the Cbfβ domain that may affect its function.

Keywords

Transcription factor; Alternative splicing; cbfβ; Runx2; Functional analysis; Zebrafish

1. 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; 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 to 4 in various adult tissues and at different embryonic developmental stages.

MATERIALS AND METHODS

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 1**).

Zebrafish *cbfβ* cDNA cloning using RT-PCR

The primer sequences used for cloning are shown in **Table 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 24hpf 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).

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).

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).

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.

Isoform expression profile

To determine the presence of the *Cbfβ* alternative transcripts (isoforms 1 to 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 1). Then, 1µl 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 1) and the CBFβ isoforms 1 and 3 specific primers (CBFβ_F4 and CBFβ_R3; Table 1). The zebrafish *gadph* was used as control (Gapdh_F and Gapdh_R; Table 1). The RT-PCR amplification was performed with the DreamTaq PCR Master Mix (Thermo Scientific) in a 25µl reaction for 40 cycles.

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 to 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 to 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).

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 37ºC 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.

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 (20 mM Tris pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 0.05% Tween-20) and the bound material was eluted with 200 µg/ml solution of flag peptide (Sigma), for 30 minutes at 4ºC. Samples were subjected to western blot analysis.

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.

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 $p<0.001$.

3. Results

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 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 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 2**).

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 1**). All the *cbfβ* isoforms 1 to 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 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 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 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 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 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.

Translation potential of the *cbfβ* spliced variants

The predicted protein sequences translated from these alternative spliced transcripts are summarized in **Figure 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.

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 3**). All splice junctions follow the GT/AG rule (Breathnach and Chambon, 1981). The zebrafish *cbfβ* isoform 1 contains all eight exons (exons 1 to 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.

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**), confirming the existence of domains in the protein that are highly conserved. Interestingly, we found four different protein isoforms (labelled A to D) that differ only in the C-terminal region (**Figure 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 to 6 (excluding exons 5a and 5b). 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** 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** 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 S1**), even if we take in account the C-terminal differences observed in the alignment.

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 5** and **Table S2**). Interestingly, from this list of genes only *noc4* and *kiaa1049* are present downstream side of *cbfβ* gene in zebrafish chromosome 18 (**Figure 5**). This syntenic conservation supports the identification of *cbfβ* as ortholog to human *CBFβ*.

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 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 6A**). The *cbfβ* mRNA was widely distributed. In a first PCR round the amplification of all four possible splicing variants (isoforms 1 to 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 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 6A**). 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 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 6B**).

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 to 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 7**). Furthermore, Cbfβ isoforms 3 and 4 seem to have lost the ability to regulate Runx2 (**Figure 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 7**).

Co-immunoprecipitation of Cbfβ splicing variants and runx2

To assess the heteromeric assembly of zebrafish Cbfβ protein isoforms 1 to 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 to 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 to 4 and Flag-Runx2 (**Figure S1**). Flag-Runx2 was specifically co-immunoprecipitated with HA-tagged Cbfβ isoforms 1, 2 and 4, but not HA-Cbfβ isoform 3 (**Figure 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. 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 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 to 4) in all the tissues analyzed (**Figure 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 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 S3**). A search of the human dbEST and non-redundant data bases identified three more exons in the human *CBFβ* gene (**Figure 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 and Figure 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 leukaemia 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 to 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 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 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 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, 2000). 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 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 function(s). 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.

Figure Legends:

Figure 1. Schematic representation of zebrafish *cbfβ* transcripts. The different transcripts originated by alternative splicing are indicated as isoforms 1 to 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.

Figure 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.

Figure 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.

Figure 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*).

Figure 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/.

Figure 6. Identification of the expression profile of zebrafish cbfβ splicing variants (isoform 1 to 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 to 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).

Figure 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 ($p<0.001$).

Figure S1. Preparation of fusion proteins. Whole cell extracts from untransfected HEK293T cells or transiently expressing the indicated proteins were analyzed 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 S2. Runx2 binds to isof 1, isof 2 and isof 4 but not to isof 3 of Cbfβ. Whole cell extracts from HEK293T cells transiently expressing Ha-Cbfβ isoforms 1, 2, 3 and 4 (isof 1, isof 2, isof 3 and isof 4) alone or in combination 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, western blot. The position of non-specific proteins (*) is indicated.

Figure 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 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 to 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.

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

Table 1: Oligonucleotides used for PCR amplification.

Table 2: Splice boundaries of the partial exon-skipping *cbfβ* mRNA variants.

Table 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).

Table S2: Zebrafish-human ortholog gene pairs

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).

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, Baselga J, Beroukhim R, Polyak K, Sgroi DC, Richardson AL, Jimenez-Sanchez G, Lander ES, Gabriel SB, Garraway LA, Golub TR, Melendez-Zajgla J, Toker A, Getz G, Hidalgo-Miranda A, Meyerson M, 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.

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, Oberkfell B, Du F, Demeter R, Vickery TL, Elhammali A, Piwnica-Worms H, McDonald S, Watson M, Dooling DJ, Ota D, Chang LW, Bose R, Ley TJ, Piwnica-Worms D, Stuart JM, Wilson RK, Mardis ER, 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. Cbf beta 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.

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.

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.

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. U S A. 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.

Vijayakumar P, Laizé V, Cardeira J, Trindade M, Cancela ML, 2013. Development of an in vitro cell system from zebrafish suitable to study bone cell differentiation and extracellular matrix mineralization. Zebrafish. 10: 500-509.

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.

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

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. U S A. 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. doi:10.1371/journal.pone.0033495.

Table 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

Table 2

Transcripts	Partial exon-skipping	Splice donor sites	Splice aceptor sites	Deletion caused by alternative splicing
Isoform 5	pΔ3-4	TTCATGGGG**at**cagcggc	agtatgtg**tg**ATCTGGAGAG	98 bp
Isoform 6	pΔ1-3	GTTCGAG**aa**cgagga	cgcccacccg**ag**AATATGTGG	211 bp
Isoform 7	pΔ1-4	GTTCGAG**aa**cgaggag	tatgatcc**tg**AACGGAGTATG	270 bp
Isoform 7	pΔ5-5a	GCACAG**gt**acagcaaaat	gatggcac**ag**CTCAATCAT	114 bp
Isoform 8 and 10	pΔ6	GGACCAGG**gt**ctgtctcc	cccagccc**ag**GCGACAGCAG	5 bp
Isoform 9 and 10	pΔ5a	TTGGAG**gt**gagagct	catttagc**ag**ATGCCGATG	3 bp
Isoform 11	pΔ3-6	GATCAG**cg**gcaggcg	gcgct**ag**CGGCATTCAC	423 bp

Figure 1

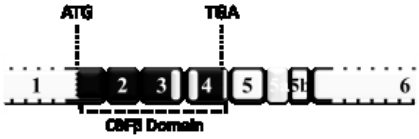
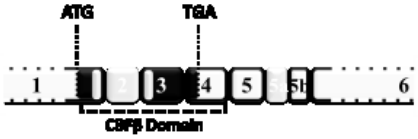
Isoforms	Description	Schematic diagram of cbfβ splice forms	mRNA* (bp)	Predicted amino acids	GenBank accession nº
Isoform 1	WT		684	188	KF709194
Isoform 2	Δ5a		642	174	KF709195
Isoform 3	Δ5b		636	201	KF709196
Isoform 4	Δ5a,5b		594	187	KF709197
Isoform 5	pΔ3; pΔ4; Δ5a		544	94	KF709198
Isoform 6	pΔ1; Δ2; pΔ3; Δ5a		431	26	KF709199
Isoform 7	pΔ1; Δ2,3; pΔ4; Δ5; pΔ5a		300	60	KF709200
Isoform 8	pΔ6		781	188	KJ704807
Isoform 9	pΔ5a		783	187	KJ704808
Isoform 10	Δ2; pΔ5a; pΔ6		587	158	KJ704809
Isoform 11	pΔ3; Δ4,5,5a,5b; pΔ6		261	>84	KJ704810

Figure 2

```
            1                                                                   70
 Isoform 1  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
 Isoform 2  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
 Isoform 3  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFMATGTNLSLQFFP
 Isoform 4  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
 Isoform 5  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
 Isoform 6  MPRVVPDQRSKFENMWTSSGRRERCT-------------------------------------------
 Isoform 7  MPRVVPDQRSKFEN----------------------------------------------------------
 Isoform 8  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
 Isoform 9  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP
Isoform 10  MPRVVPDQRSKFENEEFFRKLSRECE----------------------------AFVATGTNLSLQFFP
Isoform 11  MPRVVPDQRSKFENEEFFRKLSRECEIKYTGFRDRPHEERQARFQNACRDGRSEIAFVATGTNLSLQFFP

            71                                                                  140
 Isoform 1  ANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
 Isoform 2  ANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
 Isoform 3  ANLHGDQRQAPAREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
 Isoform 4  ANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
 Isoform 5  ANLHGDLERLARSSPSGWHGLSGI-----------------------------------------------
 Isoform 6  ----------------------------------------------------------------------
 Isoform 7  ---------------------------------GVCVIWRGWLDLHRLDGMGCLEYDDER----------
 Isoform 8  ANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
 Isoform 9  ANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
Isoform 10  ANLHGDQRQAPTREYVDFERETGKVYLKAPMILNGVCVIWRGWLDLHRLDGMGCLEYDDERAQHEDALAQ
Isoform 11  ANLHGDQRHSPKWA----------------------------------------------------------

            141                                                                 210
 Isoform 1  AAFEEARRRTRDFEDRDRSHREDLEQMPMAQLNHLITQEDPVASKIWD----------------------
 Isoform 2  AAFEEARRRTRDFEDRDRSHREDLE--------------DPVASKIWD----------------------
 Isoform 3  AAFEEARRRTRDFEDRDRSHREDLEQMPMAQLNHLITQE---------PRRQQDPSPGSNMGNTDDHKMR
 Isoform 4  AAFEEARRRTRDFEDRDRSHREDLE----------------------PRRQQDPSPGSNMGNTDDHKMR
 Isoform 5  ----------------------------------------------------------------------
 Isoform 6  ----------------------------------------------------------------------
 Isoform 7  ----------------------------AQLNHLITQEDPVASKIWD----------------------
 Isoform 8  AAFEEARRRTRDFEDRDRSHREDLEQMPMAQLNHLITQEDPVASKIWD----------------------
 Isoform 9  AAFEEARRRTRDFEDRDRSHREDLE-MPMAQLNHLITQEDPVASKIWD----------------------
Isoform 10  AAFEEARRRTRDFEDRDRSHREDLE-MPMAQLNHLITQEDPVASKIWD----------------------
Isoform 11  ----------------------------------------------------------------------
```

Figure 3

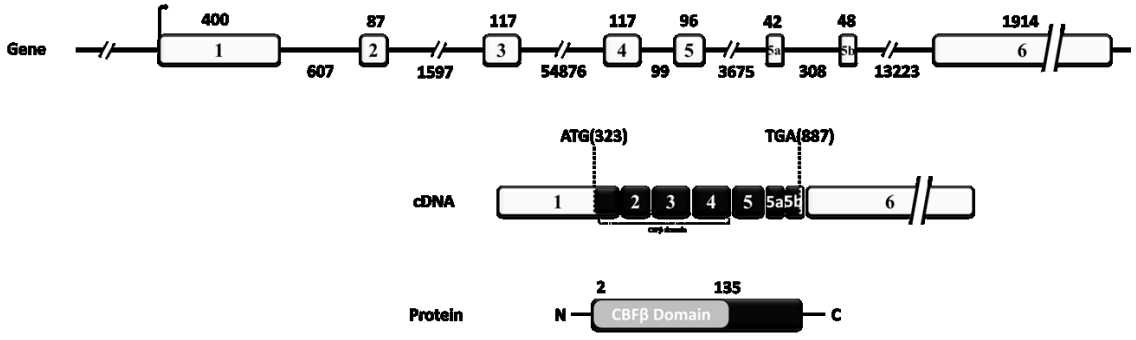
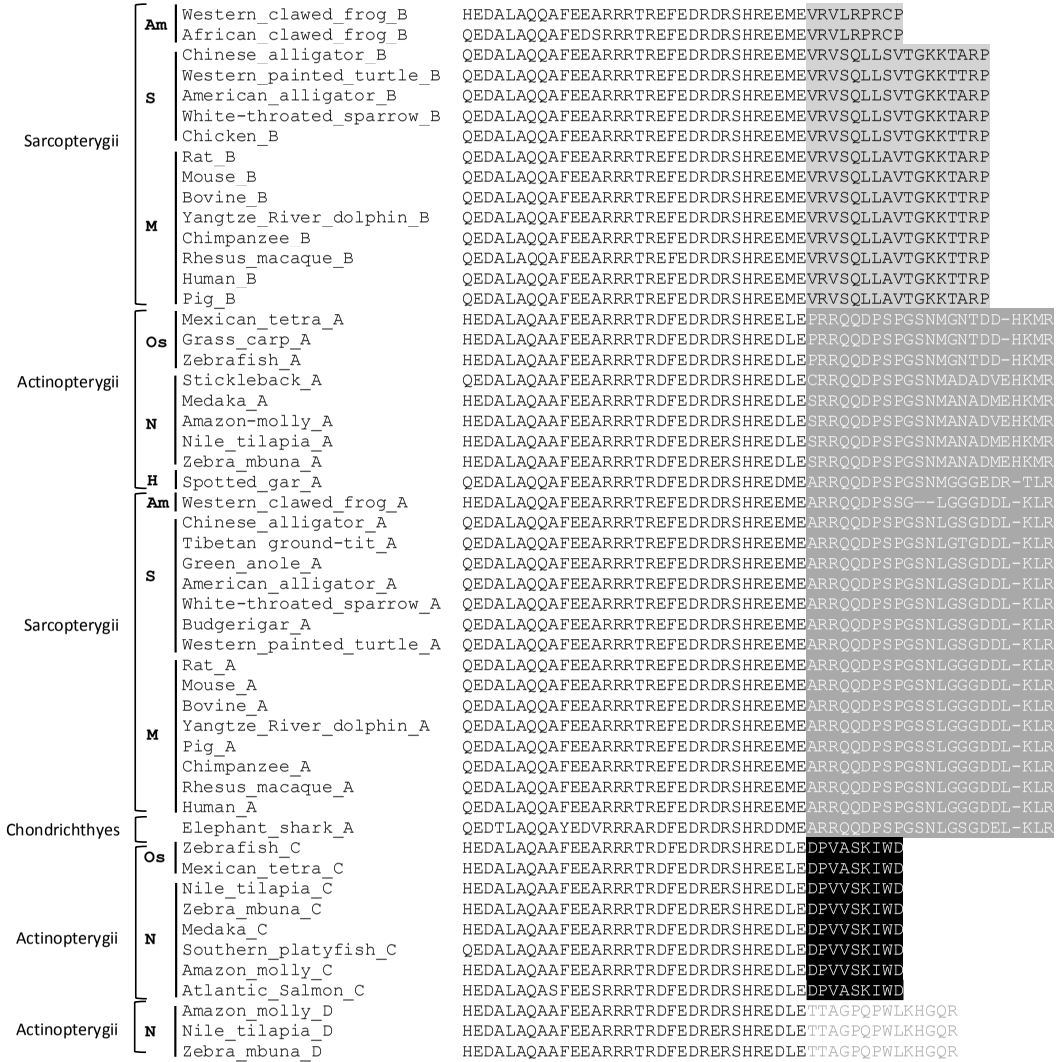


Figure 4


Sarcopterygii
Actinopterygii
Sarcopterygii
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

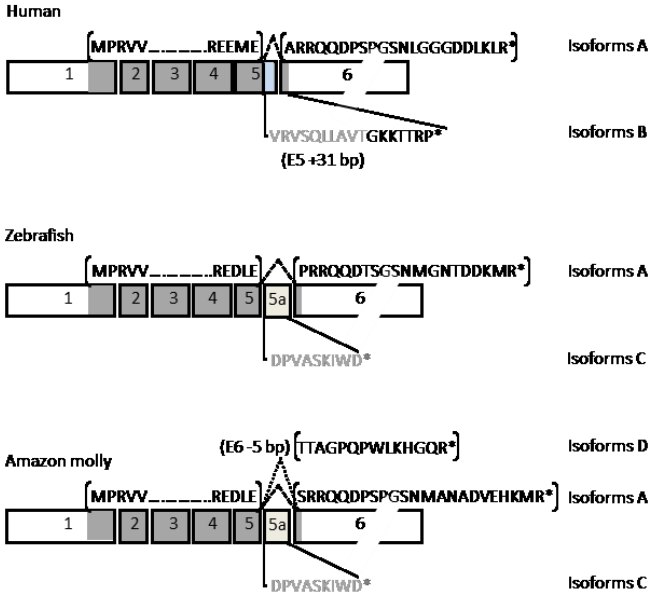

Human
MPRVV_____REEME
ARRQQDPSPGSNLGGGDDLKLR*
Isoforms A
1
2
3
4
5
6
VRVSQLLAVTGKKTTRP*
Isoforms B
(E5 +31 bp)
Zebrafish
MPRVV_____REDLE
PRRQQDTSGSNMGNTDDKMR*
Isoforms A
1
2
3
4
5
5a
6
DPVASKIWD*
Isoforms C
(E6 -5 bp)
TTAGPQPWLKHGQR*
Isoforms D
Amazon molly
MPRVV_____REDLE
SRRQQDPSPGSNMANADVEHKMR*
Isoforms A
1
2
3
4
5
5a
6
DPVASKIWD*
Isoforms C

Figure 5

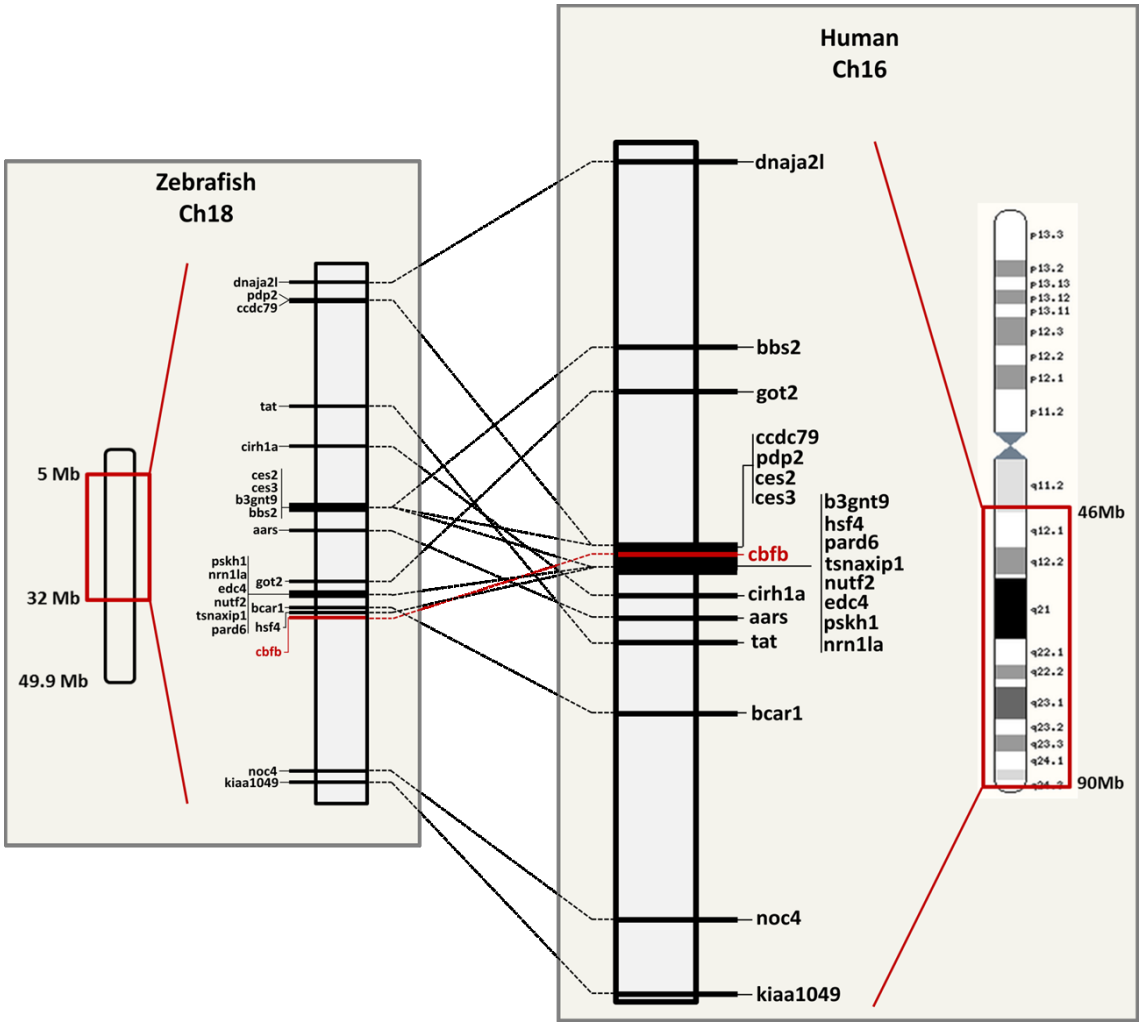

Zebrafish
Ch18
5 Mb
32 Mb
49.9 Mb
dnaja2l
pdp2
ccdc79
tat
cirh1a
ces2
ces3
b3gnt9
bbs2
aars
pskh1
nrn1la
edc4
got2
nutf2
bcar1
tsnaxip1
hsf4
pard6
cbfb
noc4
kiaa1049
Human
Ch16
dnaja2l
bbs2
got2
ccdc79
pdp2
ces2
ces3
b3gnt9
hsf4
pard6
cbfb
tsnaxip1
nutf2
cirh1a
edc4
aars
pskh1
tat
nrn1la
bcar1
noc4
kiaa1049
46Mb
90Mb

Figure 6

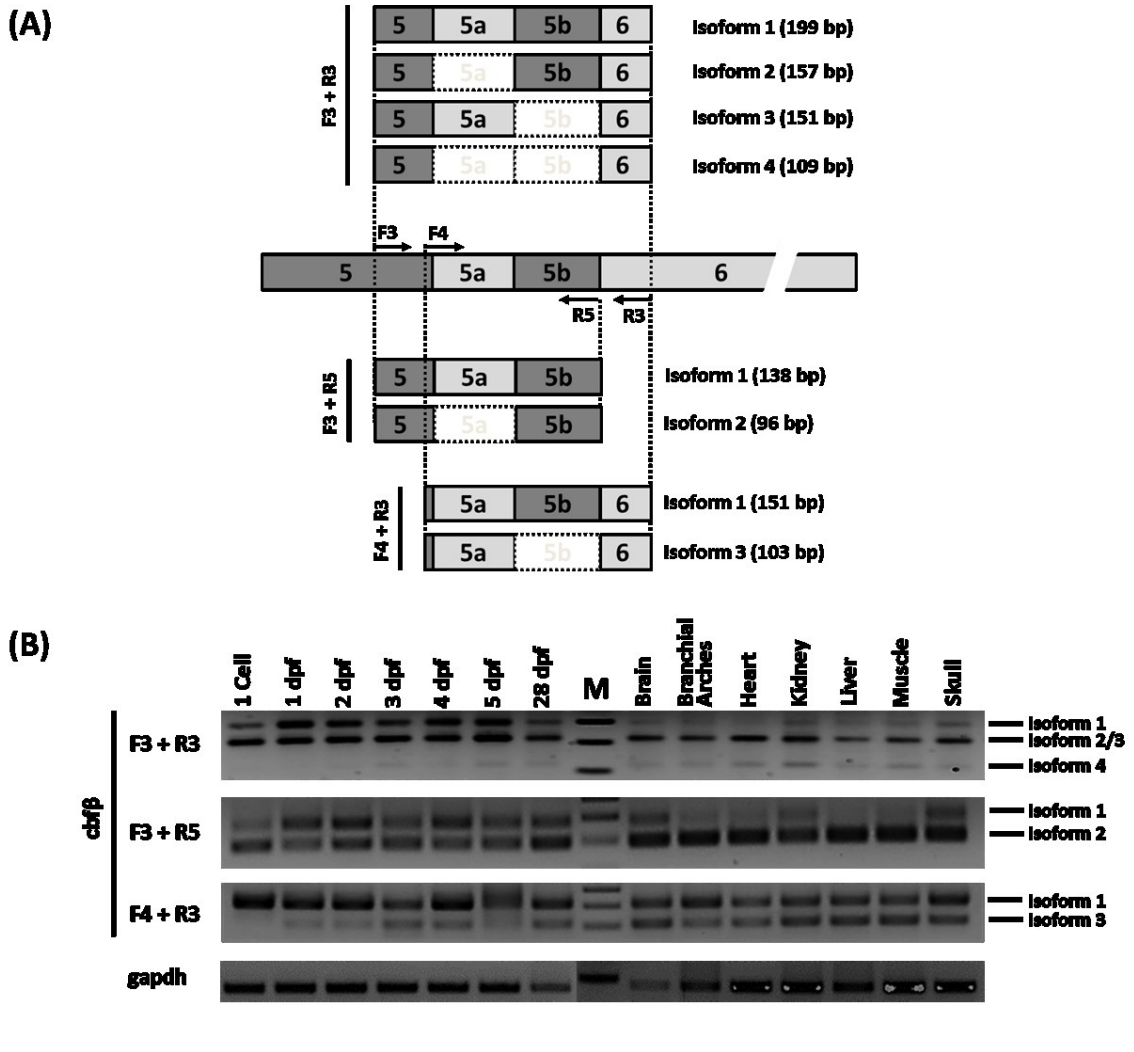


Figure 7

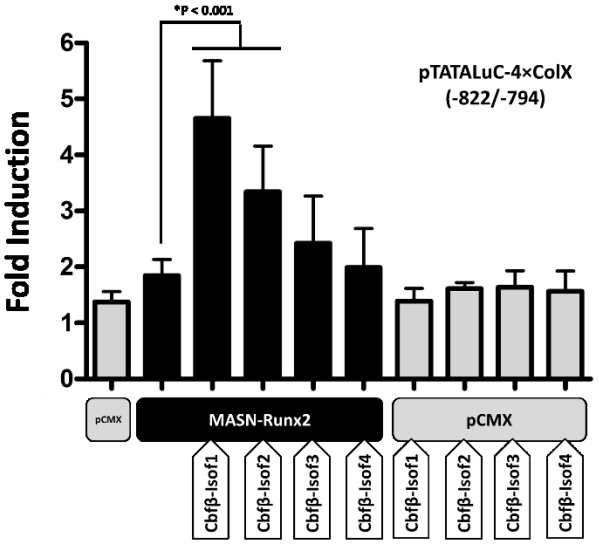


Table S1

A

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	
	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

B

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	
	Ami	Asi	Cpi	Gga	Zal	Xla	Xtr	Bta	Hsa	Lve	Mmu	Mmus	Ptr	Rno	Ssc

C

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	
	Ame	Dre	Mze	Ola	Oni	Pfo	Ssa	Xma

D

Mze			
Oni	100		
Pfo	97	97	
	Mze	Oni	Pfo

Table S2

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)

Figure S1

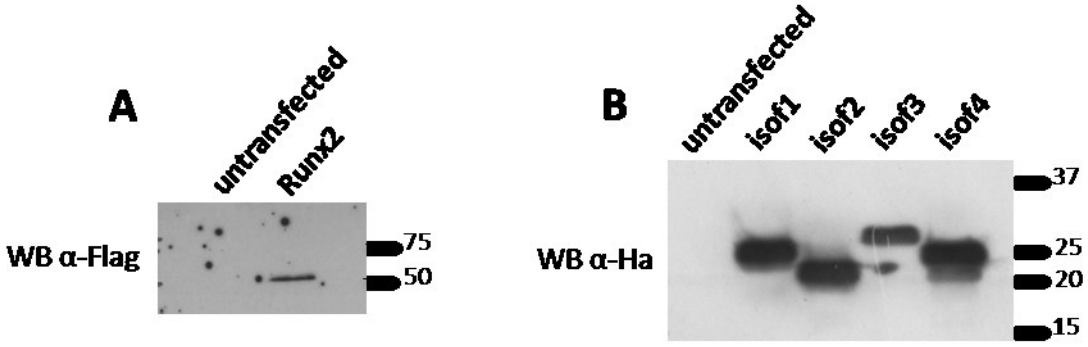


Figure S2

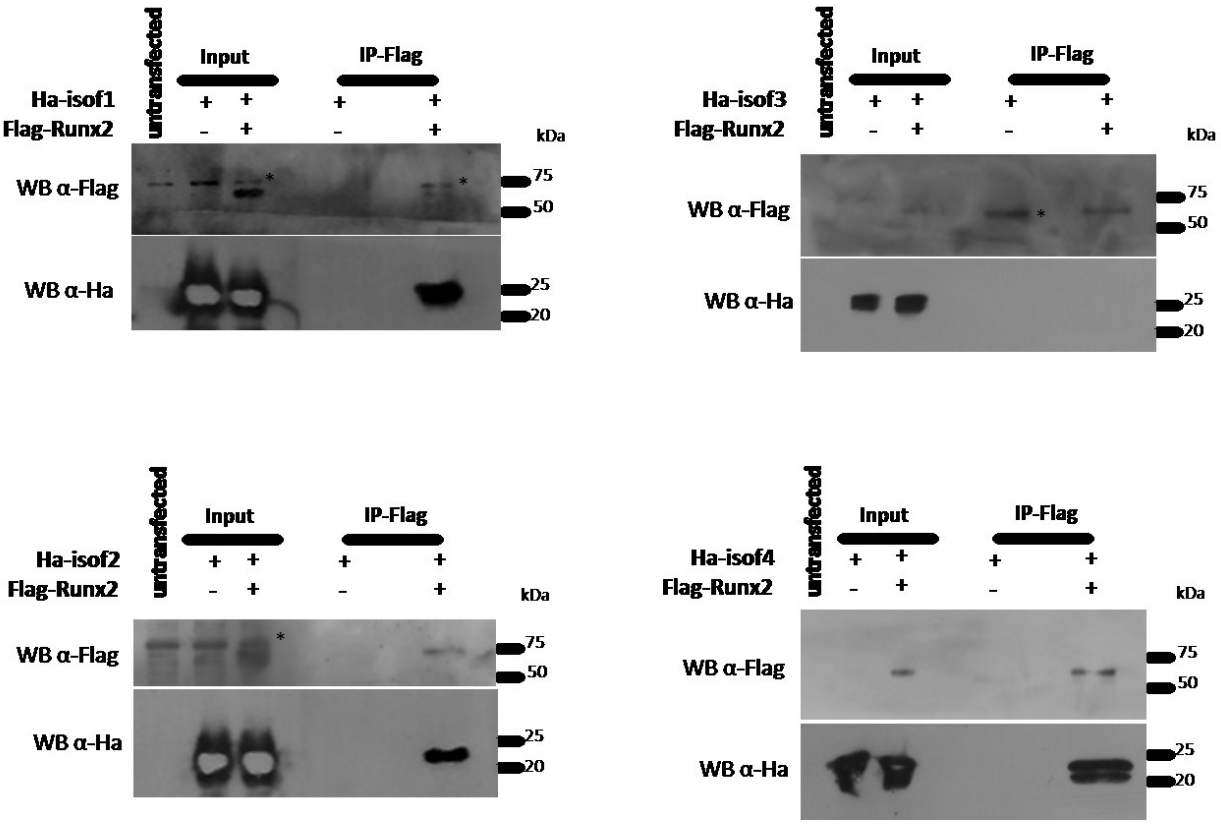


Figure S3

```
                          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 S4

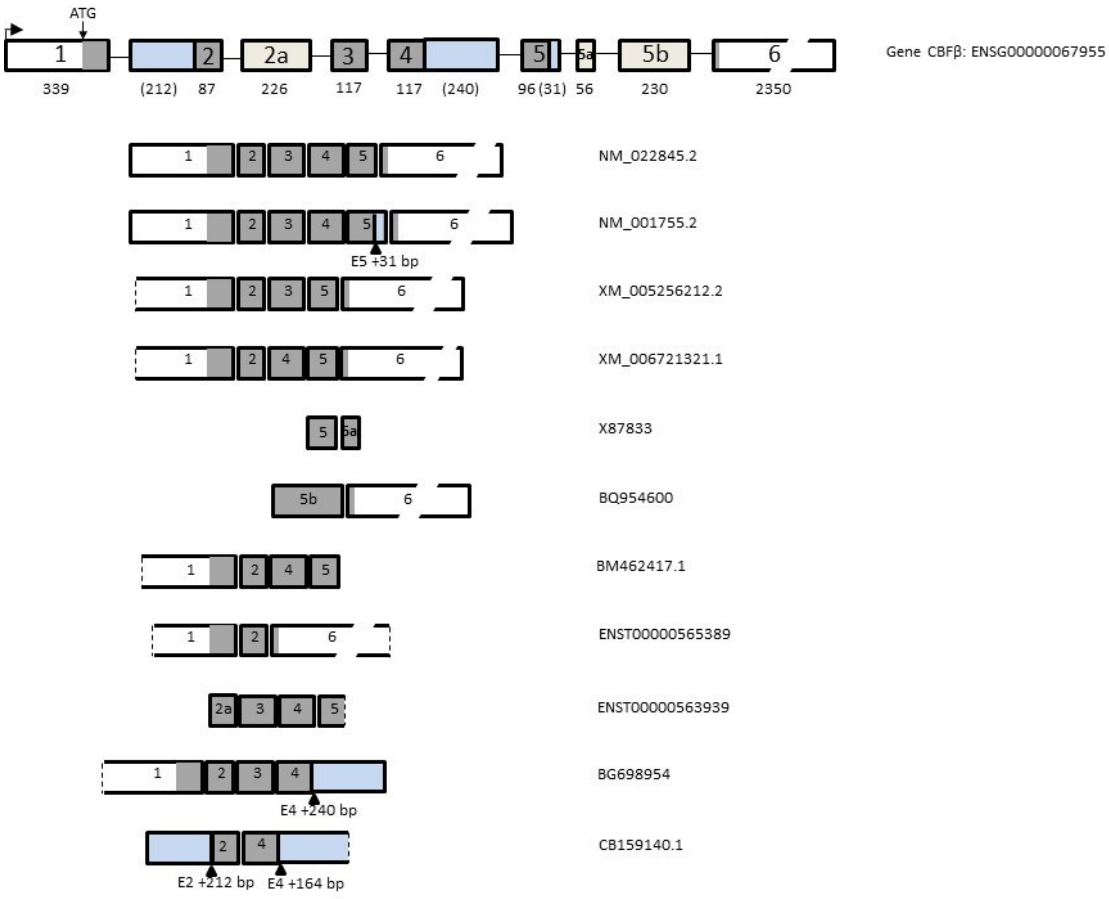


Figure S5

```
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
```