Differential expression of arid5b isoforms in Xenopus laevis pronephros

Ronan Le Bouffant, Anne-Claire Cunin, Isabelle Buisson, Jérôme Cartry, Jean-François Riou, Muriel Umbhauer

Developmental Expression Pattern

Title : Differential expression of *arid5b* isoforms in *Xenopus laevis* pronephros

Authors : Ronan Le Bouffant, Anne-Claire Cunin, Isabelle Buisson, Jérôme Cartry, Jean-François Riou and Muriel Umbhauer.

Institutional affiliations :
Sorbonne Universités, UPMC Univ Paris 06, UMR7622 Developmental Biology, F- 75005 Paris, France.
CNRS, UMR7622 Developmental Biology, F-75005 Paris, France.

Short running title : *arid5b* expression in *Xenopus* embryo

Key words: *arid5b*, *Xenopus*, pronephros

Abbreviations :
RA : retinoic acid
ARID : A-T Rich Interaction Domain
aa : amino acid
nt : nucleotide

Corresponding author : Muriel Umbhauer
Laboratoire de Biologie du Développement UMR7622, CNRS, UPMC, 9 quai Saint-Bernard, case 24, 75005, Paris France.
muriel.umbhauer@upmc.fr
phone : (33) 1 44 27 39 18 : fax : (33) 1 44 27 34 45

Author's E-mail addresses :
Ronan Le Bouffant : ronan.le-bouffant@snv.jussieu.fr
Anne-Claire Cunin : aclairecunin@gmail.com
Isabelle Buisson : isabelle.buisson@upmc.fr
Jérôme Cartry : jeromecartry@hotmail.com
Jean-François Riou : jean-francois.riou@upmc.fr
Muriel Umbhauer : muriel.umbhauer@upmc.fr

GenBank accession numbers: HG518326, HG518327

Abstract

Arid5b belongs to the ARID family of transcription factors characterised by a helix-turn-helix motif- based DNA-binding domain called ARID (A-T Rich Interaction Domain). In human, alternative splicing leads to a long and a short isoforms (isoform1 and 2, respectively) that differ in their N-terminal part. In this study, we report the cloning and expression pattern of *Xenopus laevis arid5b*. We have isolated a full length cDNA that shows homology with the human *arid5b isoform1*. Furthermore, 5'RACE experiments revealed the presence of a shorter isoform equivalent to the human isoform2. Temporal expression analysis by RT-qPCR indicated that *X. laevis arid5b isoform1* and *isoform2* are differentially expressed during development. *Isoform1* is strongly expressed maternally while *isoform2* expression is essentially restricted to tailbud stages. Spatial expression analysis by whole mount *in situ* showed that *arid5b* is predominantly expressed in the developing pronephros. *Arid5b* mRNAs are detected in the antero-dorsal part of the pronephros anlage at the early tailbud stage and later on, in the proximal part of the pronephric tubule. RT-qPCR analyses with primers that allow to discriminate *isoform1* from *isoform2* showed that the latter is enriched in the pronephros anlage. In agreement with a specific pronephric signature of the *isoform2*, we also observed that *isoform2* but not *isoform1* is upregulated in animal caps induced to form pronephric tissue in response to activin A and retinoic acid. These results indicate that the two *arid5b* isoforms are differentially expressed and likely play different roles during early *Xenopus* development.

Introduction

The ARID (A-T Rich Interaction Domain) is a helix–turn–helix motif-based DNA-binding domain, conserved in eukaryotes, that defines the ARID family of transcription factors. The human ARID family can be divided into seven subfamilies (ARID1, ARID2, ARID3, ARID4, ARID5, JARID1 and JARID2) based both on degree of homology within the ARID domain, as well as similarity between highly variable non-ARID domain structures. The founding members, murine *Bright (ARID3A)* and *Drosophila dead ringer (Dri),* were independently cloned on the basis of their ability to selectively bind to AT-rich DNA sequences but this behaviour is not an intrinsic feature of the ARID domain since the majority of ARID subfamilies bind DNA without obvious sequence preference (Patsialou *et al.*, 2005). ARID-encoding genes are involved in a variety of biological processes including regulation of cell cycle, gene expression, differentiation, embryonic development, transcriptional regulation and chromatin-remodeling. The ARID protein Osa has been shown to associate with the SWI/SNF

complex in both *Drosophila* and humans, suggesting that the function of ARID proteins in chromatin remodelling has been conserved through evolution (Collins *et al.*, 1999, Kozmik *et al.*, 2001).

Arid5b, also called *Desrt* or *MRF-2* (modulator recognition factor-2), was originally cloned thanks to its ability to bind sequences in the transcriptional modulator of the human cytomegalovirus major immediate- early promoter (Lubon *et al.*, 1989). Its ARID domain, whose three- dimensional structure has been solved, binds preferentially to an AT-rich core sequence (Whitson *et al.*, 1999). Arid5b has been found to associate with the jmjC demethylase PHF2. Assembly of the PHF2–Arid5b complex, its recruitment to target promoters, and its H3H9Me2 demethylase activity are dependent on protein kinase A activity (Baba *et al.*, 2011). During mouse organogenesis, *arid5b* displays a complex and highly dynamic pattern of expression. It is first expressed in the intermediate mesoderm and subsequently in the nephrogenic cords of the urogenital ridges. *Arid5b* is also detected in the limbs, the myotomes, the oro-naso- pharyngeal ectoderm and the underlying mesenchyme, the otic vesicles, the gut and its derivatives, and transiently in the liver. *Arid5b* mutant mice generated by gene targeting have reduced viability, pronounced growth retardation, and a high incidence of abnormalities of the reproductive organs (Lahoud *et al.*, 2001). They also show significant reductions in lipid accumulation and weight gain in postnatal and adult life (Whitson *et al.*, 2003). Arid5b is required for adipogenesis and to maintain normal functions in mature adipocytes. Knockdown of Arid5b in mature 3T3-L1-derived adipocytes activates both lipolysis and triglyceride synthesis, and causes a significant increase in the ratio of glycerol release to free fatty acid release (Yamakawa *et al.*, 2010, Yamakawa *et al.*, 2008). *Arid5b i*s highly expressed in the cardiovascular system and is believed to play essential roles in smooth muscle cell differentiation and proliferation (Watanabe *et al.*, 2002). In homozygous *arid5b* mutant mice, kidneys are small showing often degraded glomeruli with defects in smooth muscle cell number and location. Skeletal abnormalities, including defects in the patterning of the ribs and sternum, have also been described (Schmahl *et al.*, 2007).
Although the temporal and spatial pattern of expression of *arid5b* during embryogenesis have been described in mouse (Ristevski *et al.*, 2001), there is no detailed expression data available for non mammalian vertebrates. We report the cloning of two *arid5b* isoforms in *Xenopus laevis* and have examined their expression patterns during development.

Results and discussion

Molecular cloning of *X. lævis arid5b*

In order to clone a full coding sequence of *arid5b* in *Xenopus*, we started from a partial IMAGE clone sequence (no 686,6480), and obtained the missing 5' sequence by RACE PCR. A 3570 nucleotides (nt) clone was amplified by end-to-end PCR (GenBank accession no. HG518326). Sequence analysis revealed an open reading

frame encoding a predicted 1187 amino acids (aa) protein (Fig.S1). This protein displays 87.6% identity with a predicted *X. tropicalis* protein sequence deduced from gene models (accession no. XM_002939542), 57.5% identity (82.2% similarity) with the long human Arid5b isoform1, and 56.7% identity (80.7% similarity) with the murine long isoform α. The ARID domain (aa 324-418) is highly conserved, with more than 90% identity observed with other vertebrate Arid5b sequences (Fig.1). The conserved sequence includes a lysine residue (lys-342) at a position homologous to the lysine of the long murine isoform α (lys-336) whose demethylation by PHF2 promotes recruitment of PHF2-Arid5b complex to promoters (Baba et al., 2011). Blast analysis on the *X. lævis* genome 6.0 scaffolds shows that the long isoform is encoded by two genes located on scaffolds 9729 and 48311, respectively, which probably represent pseudoalleles resulting from *X. lævis* allotetraploidy. In a similar way to human and murine *arid5b*, the long *X. lævis* isoform is encoded by ten different exons (table S1,S2).

We further investigated whether a shorter isoform homologous to the short human *arid5b* isoform was also detectable. Human *arid5b isoform2* is generated by alternative splicing resulting in the replacement of phe-244 by a start methionine. Using 5' RACE-PCR, we were able to clone a partial *X. lævis arid5b* sequence of 368nt containing an ORF encoding a 91 aa polypeptide where phe-244 is replaced by a methionine. The following aa are identical to those of the long isoform (ala-245-leu-334). The 95 nt sequence located upstream to this ATG codon does not contain any other in frame ATG codon, but three stop codons indicating that it probably encodes the 5'UTR of a shorter isoform mRNA. Blast analysis on the *X. lævis* genome revealed that this sequence is encoded by a novel exon (exon 4b, table S1,S2), while ala-245-leu334 are encoded by exons 5 and 6, supporting the idea of an isoform generated by alternative splicing. Nested 3'RACE-PCR was carried out to clone the full sequence encoding this putative short isoform. The first primer was located 61nt upstream of the putative start ATG, and the nested primer 1307nt downstream of this ATG. A 2163nt sequence was cloned containing 1529nt of putative coding sequence, and 634nt of 3'UTR. It is identical to the sequence of the long isoform, in line with human *arid5b isoform2* sequence data, which only differs from isoform1 at the start ATG. Using different sets of primers, we have then tried to amplify the entire short isoform by end-to-end PCR but failed to amplify a full cDNA. Using forward primers corresponding to exon 4b sequence, and reverse primers at different levels of the sequence obtained by 3' RACE-PCR, we could only amplify a partial cDNAs encoding the first 476 aa of the short isoform. This cDNA corresponds to exons 4b, 5-9 and part of exon 10 (accession number HG518327). Whether the short isoform is ending at the same stop codon as the long one therefore remains unclear. Nonetheless, a 3' sequence for the short isoform mRNA distinct from that of the long isoform would imply an alternative splicing within exon10, that is not occurring with human *arid5b isoform2*. Together, these data show that the short isoform1 lacks the first 243 N-terminal amino acids of the long isoform2 (Fig.S2). Using InterProSCan software and performing an extensive analysis of the litterature on Arid5 family members, we could not find any known domain in this region, precluding to identify any functional difference

between the two isoforms.

Genomic synteny and molecular phylogeny

In order to further confirm the identity of the *Xenopus* ortholog of *arid5b*, we have examined the synteny maps from human, mouse and *X. tropicalis* genomes. Synteny maps were obtained from Ensembl genome browser (release 74, December 2013) and JGI genome browser (X. tropicalis version 7.1). Synteny maps are shown in Fig.1B. Flanking genes are partially conserved between *arid5b* genes in human, mouse and *X. tropicalis* genomes. The conserved syntenic region flanks the 3' end of *arid5b*. It comprises *rkn2*, *znf365*, *ado* and *egr2* genes. Genes flanking the 5' end of *arid5b* in *X. tropicalis* genome totally differ from those flanking human or mouse *arid5b*. Phylogenetic analysis indicates that *Xenopus arid5b* is relatively distant from the zebrafish ortholog. Tetrapod orthologs appear to be more closely related, with the chicken ortholog being the most closely related to *Xenopus arid5b* (Fig.1C).

Spatial and temporal expression of *arid5b* during *X. laevis* development

Temporal expression of *ari5b* during early *X. laevis* development was examined by RT-qPCR. Specific primers were designed in order to amplify either both *arid5b* isoforms (total *arid5b*), *isoform1* or *isoform2* (Fig. S3). *Isoform1* was strongly expressed during cleavage stages; its expression declined during gastrulation and neurulation, then it increased at tailbud stages (fig.2B). In contrast, *isoform2* transcripts were scarcely detected at the pretailbud stages, but became detectable during organogenesis and persisted at least up to the late tailbud stage (stage 28) (fig.2C). Notably, the temporal profile of *isoform1* is almost identical to the profile obtained for total *arid5b*, suggesting that *isoform1* is the main isoform to be expressed during embryonic development (fig.2A,B). In agreement with this idea, when normalized to total *arid5b,* relative expression of *isoform1* was constant during embryonic development while relative mRNA expression of *isoform2* increased during tailbud stages (fig.2D,E). Thus, *isoform1* and *isoform2* are differentially expressed during development. *Isoform1* is strongly expressed maternally and is the main isoform expressed during embryonic development. *Isoform2* expression is essentially restricted to tailbud stages.

Next, we studied the spatial expression pattern of *arid5b* by whole mount *in situ* hybridization using two probes : the first encompassing nucleotide 775 to 1612 of *isoform1* (which corresponds to nucleotides 46 to 853 of *isoform2)*; the second corresponding to nucleotide 529 to 3564 of *isoform1*. Both probes gave the same expression pattern (fig.4 and data not shown). At cleavage, gastrula and neurula stages, embryos were uniformly stained indicating that *arid5b* mRNAs were ubiquitously distributed (not shown). RT-qPCR analysis of dissected explants from early gastrula confirmed this observation and further showed that none of the two isoforms showed a regionalized expression (fig.3). At the early tailbud stage, a specific signal was detected in

the pronephric region by *in situ* hybridization (fig.4A). At stage 25, *arid5b* was strongly expressed in the antero-dorsal part of the pronephric anlage. In comparison with *delta1* expression which is restricted to the most antero-dorsal part of the anlage ((Rones *et al.*, 2000) et fig.4C,F), *arid5b* is expressed in a larger domain that extends more ventrally and medially (fig.4B,E). Expression of *arid5b* remained restricted to the anterior part of the developing pronephros at tailbud stages (fig.4D). At tadpole stage 35/36, *arid5b* mRNAs were localized in the proximal part of the pronephric tubule (fig.4H,J,I,L). *Arid5b* expression domain is adjoining to that of *scl12a1* which marks the intermediate and the first distal segments (Raciti *et al.*, 2008) (fig.4H,I,L). Thus, from the early tailbud stage onward, *arid5b* expression is restricted to the developing proximal part of the pronephros. Since our *in situ* hybridization experiments do not allow to distinguish between the two isoforms, we performed RT-qPCR to analyse the expression levels of each isoform in the pronephric anlage in comparison to the whole embryo and to different embryonic regions. The results clearly showed that *isoform2* is more strongly expressed in the pronephric anlage than in the other tissues at tailbud stage (fig.5). No such enrichment was observed for *isoform1* although it is slightly more expressed in the head in comparison to other tissues. As expected, the kidney marker *pax8* was found to be strongly expressed in the pronephric anlage (fig.5). It has been previously shown that treatment of blastula animal cap ectoderm with activin A and retinoic acid (RA) results in the formation of pronephric tubules at high frequency (Ariizumi and Asashima, 2001). We studied whether this treatment could upregulate *arid5b* isoforms expression. We showed that *isoform2* but not *isoform1* expression is upregulated in response to activin A and RA (fig.6). This result is in agreement with a specific pronephric signature of the *isoform2* .

In summary, we have cloned the *X. laevis* ortholog of *arid5b*, identified two isoforms and examined their expression pattern during embryonic development. The two isoforms are differentially expressed : *isoform1* is strongly expressed maternally, while *isoform2* is specifically expressed in the pronephric anlage at tailbud stage. These results indicate that the two *arid5b* isoforms likely play different roles during early *Xenopus* development.

Materials and Methods

Molecular cloning and Bioinformatic analyses

The partial IMAGE clone 696,6480 was obtained from RZPD ImaGenes. RACE-PCR was performed with the SMARTer™ RACE cDNA amplification kit (Clontech). End-to-end PCR was carried out according to manufacturer instructions with Advantage 2 polymerase (Clontech), and stage 35/36 embryo cDNA prepared as described (Le Bouffant *et al.*, 2012). Amino acid sequence comparison were performed with MultAlin software (http://multalin.toulouse.inra.fr/multalin/) (Corpet, 1988) and CLUSTAL W (version 1.83).

Developmental Expression Pattern

Embryos, explants dissection, animal cap assay

Xenopus lævis were purchased from the CNRS *Xenopus* breeding Center (Rennes, France). Embryos were obtained after artificial fertilization, and were raised in modified Barth's solution (MBS). Stages were according to the normal table of *Xenopus lævis* (Nieuwkoop and Faber, 1967). Dissections were all performed in 1X MBS on 1% agar-coated dishes. Presumptive ectoderms (animal cap) were isolated from blastula stage embryos (stages 8–9) and immediately transferred into 1X MBS, 0.1% BSA in the presence of recombinant human activin A (10 ng/mL, Sigma) and all-trans retinoic acid (10^{-4} M, Sigma) or DMSO alone (1/250). The animal caps were incubated for 3 hrs, after which they were washed twice in 1X MBS and further cultured for 48 hrs. Explants comprising the pronephric anlage were dissected from early tailbud embryos (stage 21). Using platinum loop and wire, stage 25 embryos were dissected into several pieces : the head, the tail, the truncal dorsal part (essentially somites, neural tube and notochord) and the truncal ventral part (mainly endoderm, ventral and lateral mesoderm). Pronephric anlagen were isolated from somitic and lateral plate mesoderm and separated from the underlying endoderm. The overlying ectoderm was kept. Explants were immediately processed for RT-qPCR.

Real-time quantitative PCR

RT-PCR analyses were carried out as previously reported in (Le Bouffant *et al.*, 2012). Sequences of oligonucleotides used are the following : *arid5b :* forward: 5'TATGTTTCAAGCTGCGCAAAA3', reverse: 5'CCATTGCCTCCGTGCAGTA3' ; *arid5b isoform1* forward : 5'CCCAGAAGATACCCCCAAGG3', reverse : 5'ACTTCATGCTCTCCGTGGCT3' ; *arid5b isoform2* : forward : 5'TGCTCTGTGGCGTTCATGAG3' reverse : 5'TCGACTAGCATCTGTCTCGTTTGT3' ; *pax8* : forward : 5'CAGCAATTTCAATATAGGTCACGG3', reverse 5'TCCATTCACAAAAGCCCCAC3' ; *ODC* : forward : 5'GGGCAAAGGAGCTTAATGTGG3', reverse 5'TGCCAACATGGAAACTCACAC3'. The Comparative Ct method was used to determine the relative quantities of mRNA, using *ODC* mRNA as the endogenous reporter except for figure 2D,E for which *arid5b* was used . Same results were obtained using *β-actin* mRNA as the endogenous reporter instead of *ODC* (data not shown). Each RNA sample was analysed in duplicate. Each data point represents the mean ± SEM of at least three independent experiments. Data were analysed using R Commander (R software) by paired Student's t-test.

In situ hybridization

Whole mount *in situ* hybridization for *arid5b*, *delta1* (Rones *et al.*, 2000), and *slc12a1* (Raciti *et al.*, 2008) were carried out as previously reported (Cartry *et al.*, 2006). The antisense and control sense RNA probes for *arid5b* were generated from linearized plasmids containing cDNA sequences from nucleotide 529 to 3564 of

196 *isoform1* and from nucleotide 775 to 1612 of *isoform1* (which corresponds to nucleotides 46 to 853 of
197 *isoform2)*. The *arid5b* RNA probes were subjected to limited alkaline hydrolysis in two volumes of carbonate
198 buffer (60 mM Na_2CO_3, 40 mM $NaHCO_3$, pH 10.2) for 5 min at 60°C to reduce its size and increase its access to
199 tissues. The hydrolysis was terminated by adding an equal volume of neutralizing solution (1 M Tris-HCl, pH 8.0,
200 containing 1.5 M NaCl). Hydrolyzed fragments were precipitated with ethanol.

201

202 **Acknowledgements**:
203 We thank S. Autier and E. Manzoni for animal care, E. Jones and E. Bellefroid for plasmids. This work was
204 supported by grants from CNRS and from Universitý Pierre et Marie Curie. We acknowledge funding from
205 Emergence-UPMC-2009 research program.

206

207

208

209 **Figure Legends**

210 **Fig. 1 :** Characterization of a *X. laevis* ortholog of *arid5b*.

211 **(A**) Predicted ARID domain amino acid sequence comparison: *G. gallus* (ac Q5ZJ69); *H. sapiens* (ac Q14865); *X.*
212 *tropicalis* (F6QQ73); *D. rerio* (E7F888). **(B**) Synteny blocks containing *arid5b* genes in *H. sapiens*; *M. musculus*
213 and *X. tropicalis* genomes. Genes organization in the human *arid5b* gene region was used as basis for
214 comparison. Chromosomal localization is indicated. The scaffold number is given for *X. tropicalis*. Relative
215 spacing between the genes is not shown. (**C**) Phylogenetic tree of *arid5b* genes from various vertebrate species
216 constructed using the neighbour-joining method. Accession numbers used are shown.

217

218 **Fig. 2 : Temporal expression of *arid5b* during *Xenopus* embryonic development**

219 Expression of *arid5b* analysed by RT-qPCR at cleavage (stages 5 and 9), gastrula (stage 11), neurula (stages 14)
220 and tailbud stages (stages 22 and 28). Primers were designed in order to amplified either both isoforms (total
221 *arid5b*) (**A**), *isoform1* (**B,D**) or *isoform2* (**C,E**). The relative quantities of mRNA were determined with *ODC*
222 (**A,B,C**) or total *arid5b* (**D,E**) mRNA as the endogenous reporter. *Isoform1* is strongly maternally expressed;
223 *isoform2* is mainly expressed during tailbud stages. Average values from three independent experiments.

224 **Fig. 3 : Expression of *arid5b* at the early gastrula stage**

225 RT-qPCR analyses for total *arid5b*, *isoform1* and *isoform2* were performed on dissected explants from early
226 gastrula stage embryo (stage10.5). Embryos were dissected either into ventral and dorsal halves or into dorsal

marginal zone (DMZ), ventral marginal zone (VMZ), lateral marginal zone (LMZ) ectoderm and endoderm. Both isoforms are ubiquitously expressed at the early gastrula stage. Average values from three independent experiments.

Fig. 4 : Spatial expression of *arid5b* during *Xenopus* development

In situ hybridization of whole (**A-D, G-L**) or transverse fractured embryos (**E, F**) at the indicated stages of development with antisense probe for *arid5b* (nucleotide 775 to 1612 of *isoform1*) (**A,B,D,E,H,J,L,I**), *delta1* (**C,F**), *slc12a1* (**I,K,L**) and with control sense probe for *arid5b* (**G**). Lateral views with anterior to the right (**A-D,G,H, J-L**). Transverse section at the level of the proximal pronephric tubule. In I and L, *arid5b* is revealed in light blue and *slc12a1* in purple. *Arid5b* mRNAs are detected in the anterior pronephric anlage during tailbud stages in a broader domain than *delta1*. At tadpole stage 35/36, *arid5b* expression is restricted to the proximal part of the tubule and does not overlap with *slc12a1* expression which is specific for the intermediate and the first segment of the distal tubule. Scale bars are 0.3 mm.

Fig. 5 : *Arid5b isoform2* is specifically expressed in the pronephros

Expression of *isoform1*, *isoform2* and *pax8* analysed by RT-qPCR in different embryonic regions (see materials and methods) dissected at tailbud stage 25. *Isoform2* as well as *pax8* are strongly expressed in pronephric anlage in comparison to other embryonic tissues (statistically significant for all). *Isoform1* is slightly more expressed in the head in comparison to other tissues. Average values from three independent experiments. ∗ $P<0.05$

Fig. 6 : *Arid5b isoform2* is specifically induced in animal caps treated with activin A and RA

RT-qPCR analysis of *arid5b isoform1* and *isoform2*, as well as *pax8* expression in induced blastula animal caps. Animal caps were dissected at blastula stage 9, incubated for 3 hours in presence of RA and activin A, or mock solution for the control group. Animal caps were further cultured in 1X MBS for 48h hours and processed for RT-qPCR. A significant increase of *isoform2* expression is observed in response to RA and activin A. Average values from three independent experiments. ∗ $P<0.05$∗∗ $P < 0.005$

Legends to Supplementary materials

Fig. S1 : Comparison of Arid5b isoform1 amino acid sequence between vertebrates

Predicted amino acid sequence comparison of Arid5b: *X. tropicalis* (F6QQ73) isoform1; *G. gallus* (Q5ZJ69); *H. sapiens* (Q14865); *M. musculus* (Q8BM75); *D. rerio* (E7F888). Red boxes indicate amino acid residues

conserved in all species. The green arrow indicates phe-244 of isofrom1 that is replaced by a start methionine in isoform2.

Fig. S2 : Amino acid comparison of the two Arid5b isoforms. The ARID domain is highlighted in yellow. Regions specific to isoform1 or isOform2 are indicated in red letters. The sequence in blue corresponds to the predicted C-terminal sequence of isoform2 observed in 3'RACE PCR experiment but that we could not amplify by end-to-end PCR.

Fig. S3 : Nucleotide position of PCR primers used for RT-qPCR. Nucleotide sequences of the N-terminal region of *arid5b isoform1* and *isoform2* are aligned. Arrows indicate the primer sequences that were used to amplify both isoforms (in blue), *isoform1* (green) and *isoform2* (brown). Conserved nucleotides are in red.

Table S1 : Exon-intron organization of the *arid5b* gene in *Xenopus laevis* on scaffold 9729.

Exon sequences are indicated by uppercase letters and intron sequences by lowercase letters. Splice donor and acceptor sites are underlined. Exon and intron size are reported as base pairs.

Table S2 : Exon-intron organization of the *arid5b* gene in *Xenopus laevis* on scaffold 48311.

Exon sequences are indicated by uppercase letters and intron sequences by lowercase letters. Splice donor and acceptor sites are underlined. Exon and intron size are reported as base pairs.

Developmental Expression Pattern

References

ARIIZUMI, T. and ASASHIMA, M. (2001). In vitro induction systems for analyses of amphibian organogenesis and body patterning. *Int J Dev Biol* 45: 273-9.

BABA, A., OHTAKE, F., OKUNO, Y., YOKOTA, K., OKADA, M., IMAI, Y., NI, M., MEYER, C.A., IGARASHI, K., KANNO, J. *et al.* (2011). PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. *Nat Cell Biol* 13: 668-75.

CARTRY, J., NICHANE, M., RIBES, V., COLAS, A., RIOU, J.F., PIELER, T., DOLLE, P., BELLEFROID, E.J. and UMBHAUER, M. (2006). Retinoic acid signalling is required for specification of pronephric cell fate. *Developmental Biology* 299: 35-51.

COLLINS, R.T., FURUKAWA, T., TANESE, N. and TREISMAN, J.E. (1999). Osa associates with the Brahma chromatin remodeling complex and promotes the activation of some target genes. *EMBO J* 18: 7029-40.

CORPET, F. (1988). Multiple Sequence Alignment with Hierarchical-Clustering. *Nucleic Acids Res* 16: 10881-10890.

KOZMIK, Z., MACHON, O., KRALOVA, J., KRESLOVA, J., PACES, J. and VLCEK, C. (2001). Characterization of mammalian orthologues of the Drosophila osa gene: cDNA cloning, expression, chromosomal localization, and direct physical interaction with Brahma chromatin-remodeling complex. *Genomics* 73: 140-8.

LAHOUD, M.H., RISTEVSKI, S., VENTER, D.J., JERMIIN, L.S., BERTONCELLO, I., ZAVARSEK, S., HASTHORPE, S., DRAGO, J., DE KRETSER, D., HERTZOG, P.J. *et al.* (2001). Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res* 11: 1327-34.

LE BOUFFANT, R., WANG, J.H., FUTEL, M., BUISSON, I., UMBHAUER, M. and RIOU, J.F. (2012). Retinoic acid-dependent control of MAP kinase phosphatase-3 is necessary for early kidney development in Xenopus. *Biol Cell* 104: 516-32.

LUBON, H., PITTIUS, C.W. and HENNIGHAUSEN, L. (1989). Invitro Transcription of the Mouse Whey Acidic Protein Promoter Is Affected by Upstream Sequences. *Febs Letters* 251: 173-176.

NIEUWKOOP, P.D. and FABER, J. (1967). *Normal table of Xenopus lævis (Daudin)*. North-Holland Publishing Company, Amsterdam.

PATSIALOU, A., WILSKER, D. and MORAN, E. (2005). DNA-binding properties of ARID family proteins. *Nucleic Acids Res* 33: 66-80.

RACITI, D., REGGIANI, L., GEFFERS, L., JIANG, Q., BACCHION, F., SUBRIZI, A.E., CLEMENTS, D., TINDAL, C., DAVIDSON, D.R., KAISSLING, B. *et al.* (2008). Organization of the pronephric kidney revealed by large-scale gene expression mapping. *Genome Biol* 9: R84.

RISTEVSKI, S., TAM, P.P.L., KOLA, I. and HERTZOG, P. (2001). Desrt, an AT-rich interaction domain family transcription factor gene, is an early marker for nephrogenic mesoderm and is expressed dynamically during mouse limb development. *Mechanisms of Development* 104: 139-142.

RONES, M.S., MCLAUGHLIN, K.A., RAFFIN, M. and MERCOLA, M. (2000). Serrate and Notch specify cell fates in the heart field by suppressing cardiomyogenesis. *Development* 127: 3865-76.

SCHMAHL, J., RAYMOND, C.S. and SORIANO, P. (2007). PDGF signaling specificity is mediated through multiple immediate early genes. *Nature Genetics* 39: 52-60.

WATANABE, M., LAYNE, M.D., HSIEH, C.M., MAEMURA, K., GRAY, S., LEE, M.E. and JAIN, M.K. (2002). Regulation of smooth muscle cell differentiation by AT-rich interaction domain transcription factors Mrf2 alpha and Mrf2 beta. *Circ Res* 91: 382-389.

WHITSON, R.H., HUANG, T. and ITAKURA, K. (1999). The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. *Biochem Biophys Res Commun* 258: 326-31.

WHITSON, R.H., TSARK, W., HUANG, T.H. and ITAKURA, K. (2003). Neonatal mortality and leanness in mice lacking the ARID transcription factor Mrf-2. *Biochem Biophys Res Commun* 312: 997-1004.

YAMAKAWA, T., SUGIMOTO, K., WHITSON, R.H. and ITAKURA, K. (2010). Modulator recognition factor-2 regulates triglyceride metabolism in adipocytes. *Biochem Biophys Res Commun* 391: 277-281.

YAMAKAWA, T., WHITSON, R.H., LI, S.L. and ITAKURA, K. (2008). Modulator recognition factor-2 is required for adipogenesis in mouse embryo fibroblasts and 3T3-L1 cells. *Molecular Endocrinology* 22: 441-453.

Developmental Expression Pattern

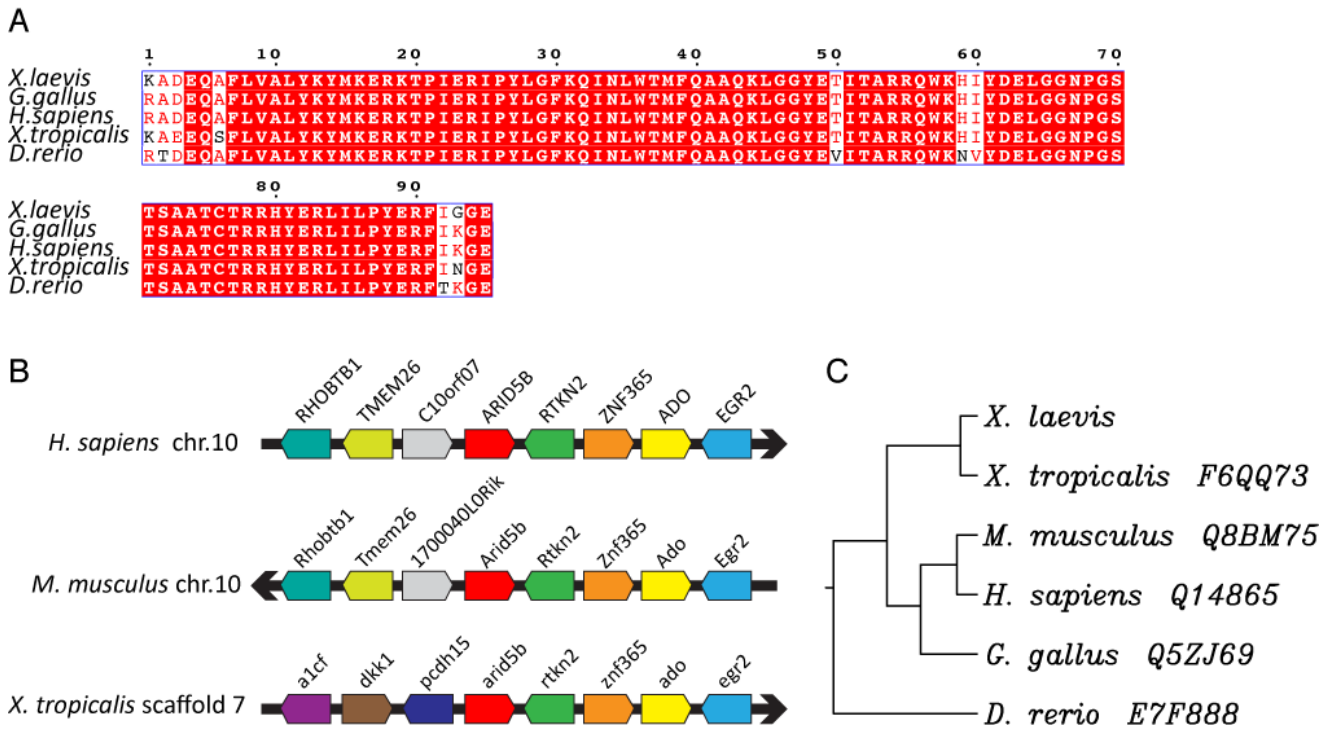


Figure 1

333

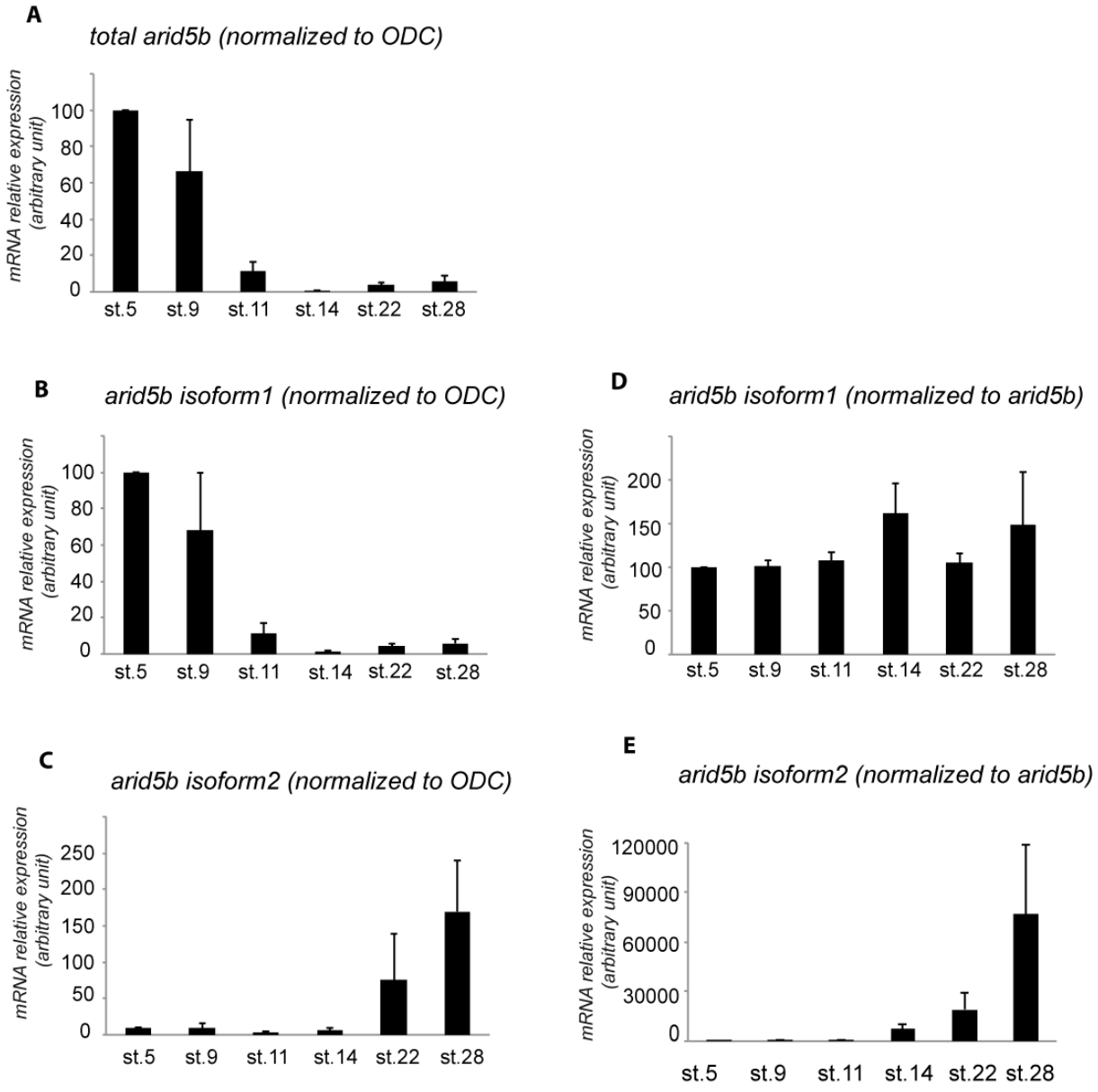

A
total arid5b (normalized to ODC)
mRNA relative expression (arbitrary unit)
100
80
60
40
20
0
st.5
st.9
st.11
st.14
st.22
st.28
B
arid5b isoform1 (normalized to ODC)
mRNA relative expression (arbitrary unit)
100
80
60
40
20
0
st.5
st.9
st.11
st.14
st.22
st.28
C
arid5b isoform2 (normalized to ODC)
mRNA relative expression (arbitrary unit)
250
200
150
100
50
0
st.5
st.9
st.11
st.14
st.22
st.28
D
arid5b isoform1 (normalized to arid5b)
mRNA relative expression (arbitrary unit)
200
150
100
50
0
st.5
st.9
st.11
st.14
st.22
st.28
E
arid5b isoform2 (normalized to arid5b)
mRNA relative expression (arbitrary unit)
120000
90000
60000
30000
0
st.5
st.9
st.11
st.14
st.22
st.28

Figure 2

Developmental Expression Pattern

334

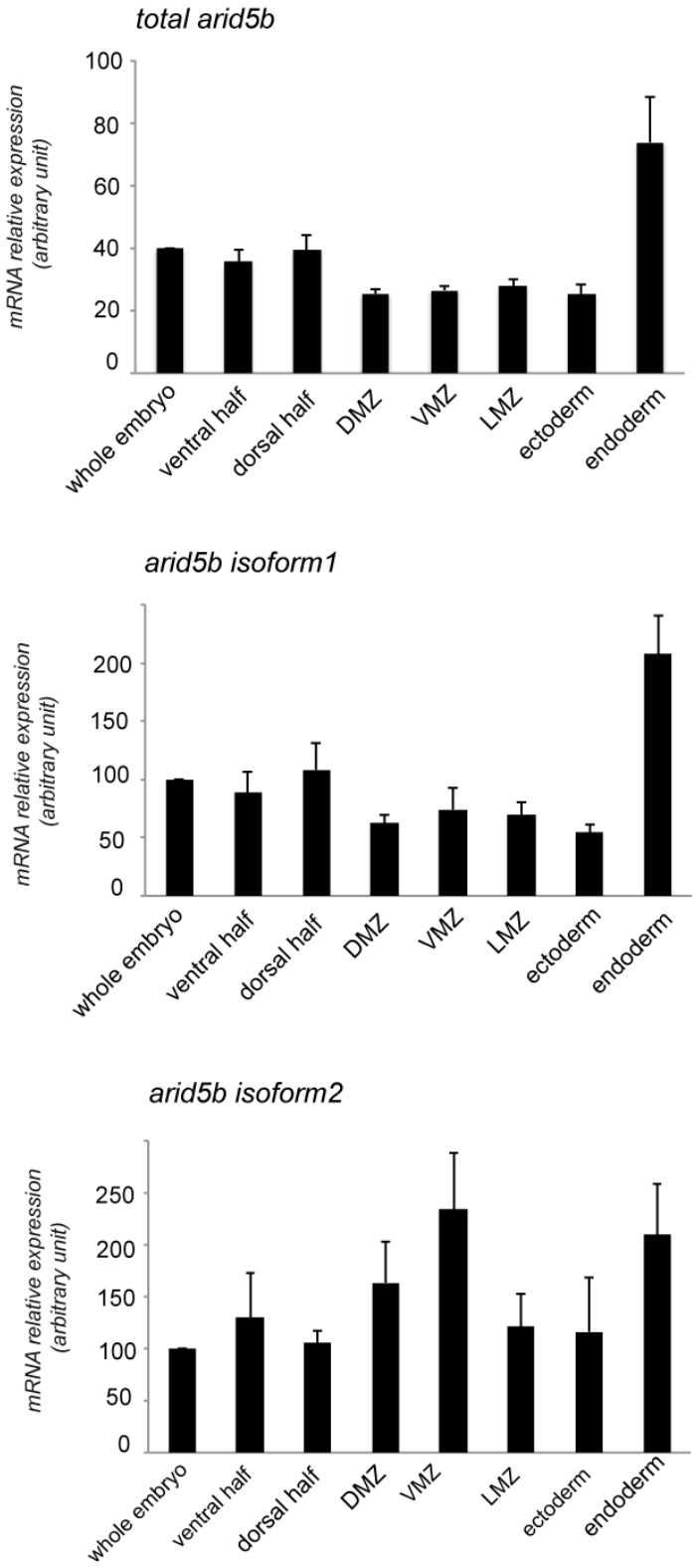


figure 3

Developmental Expression Pattern

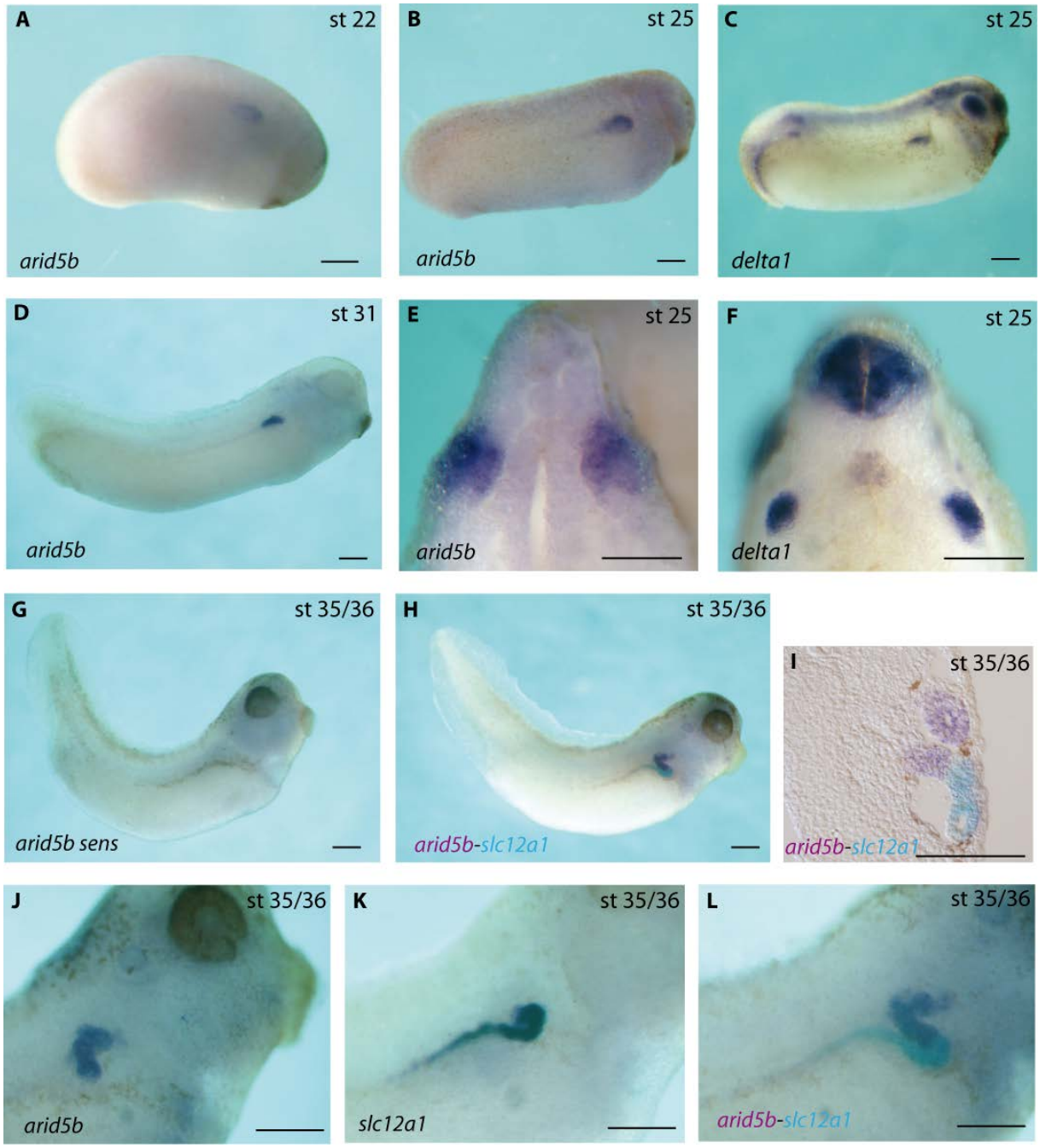


Figure 4

Developmental Expression Pattern

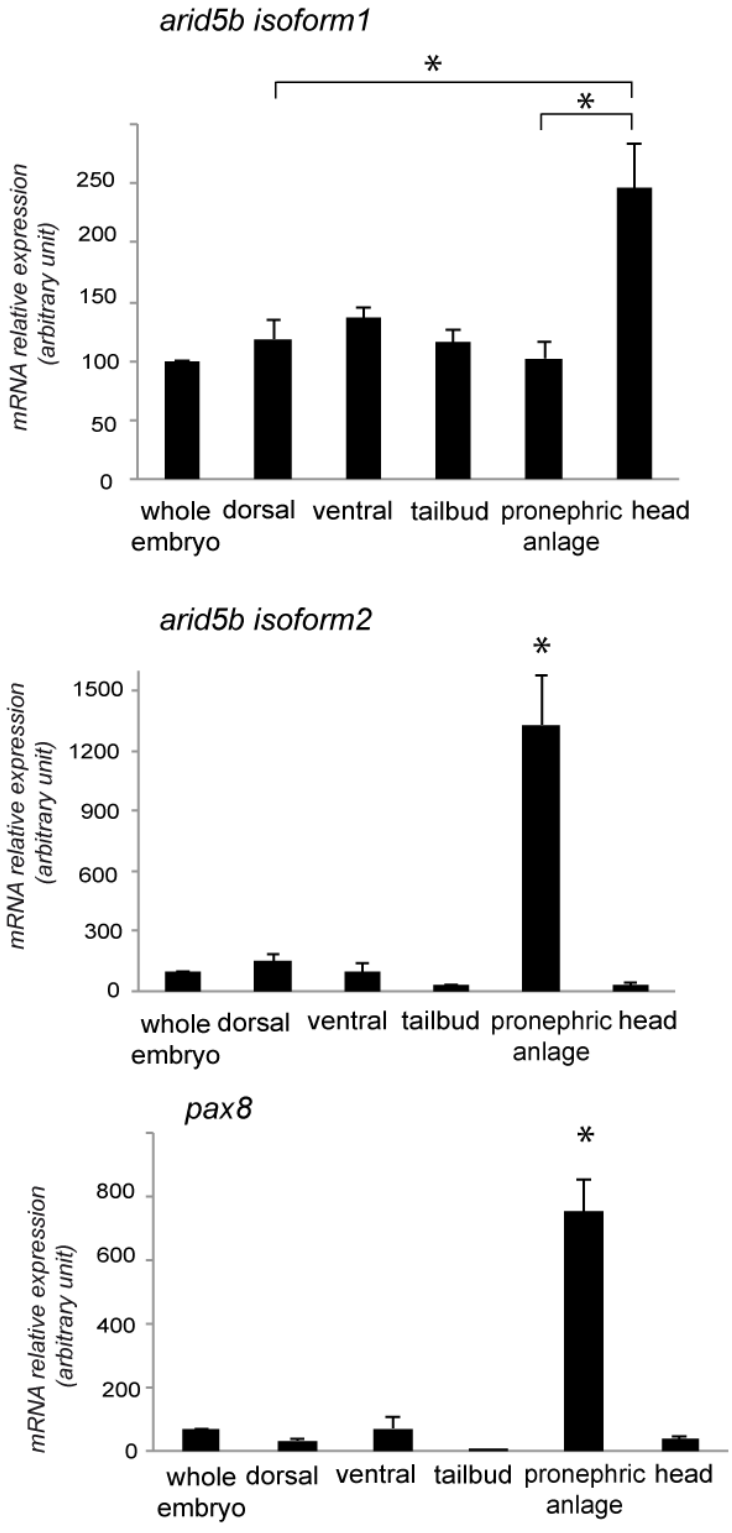


Figure 5

Developmental Expression Pattern

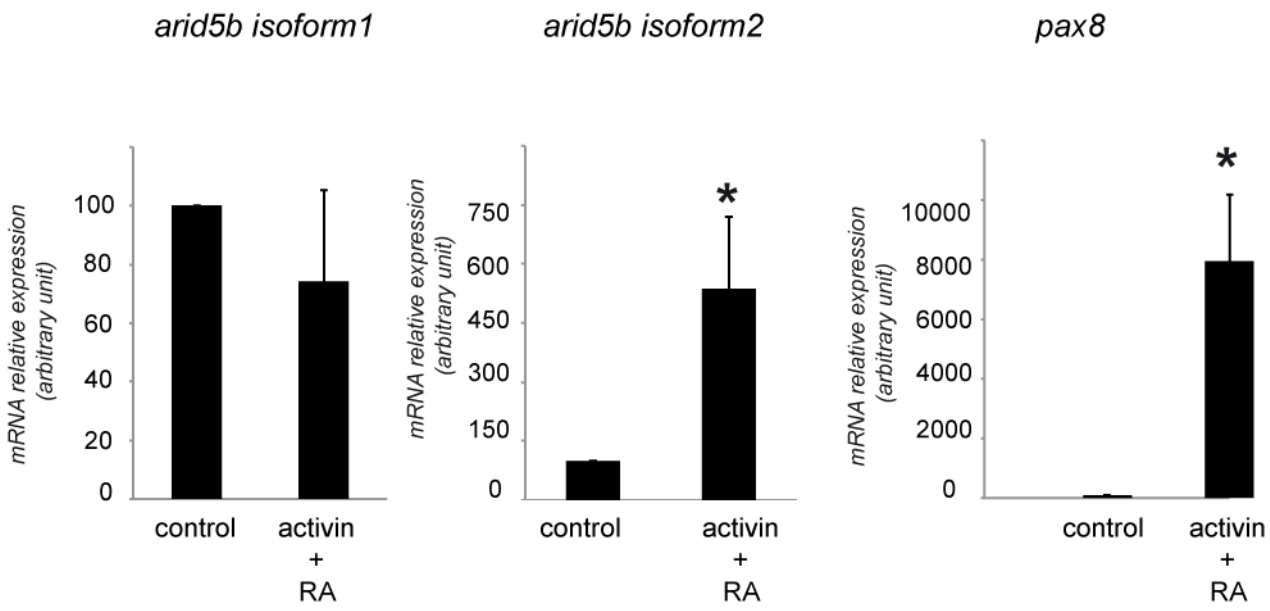
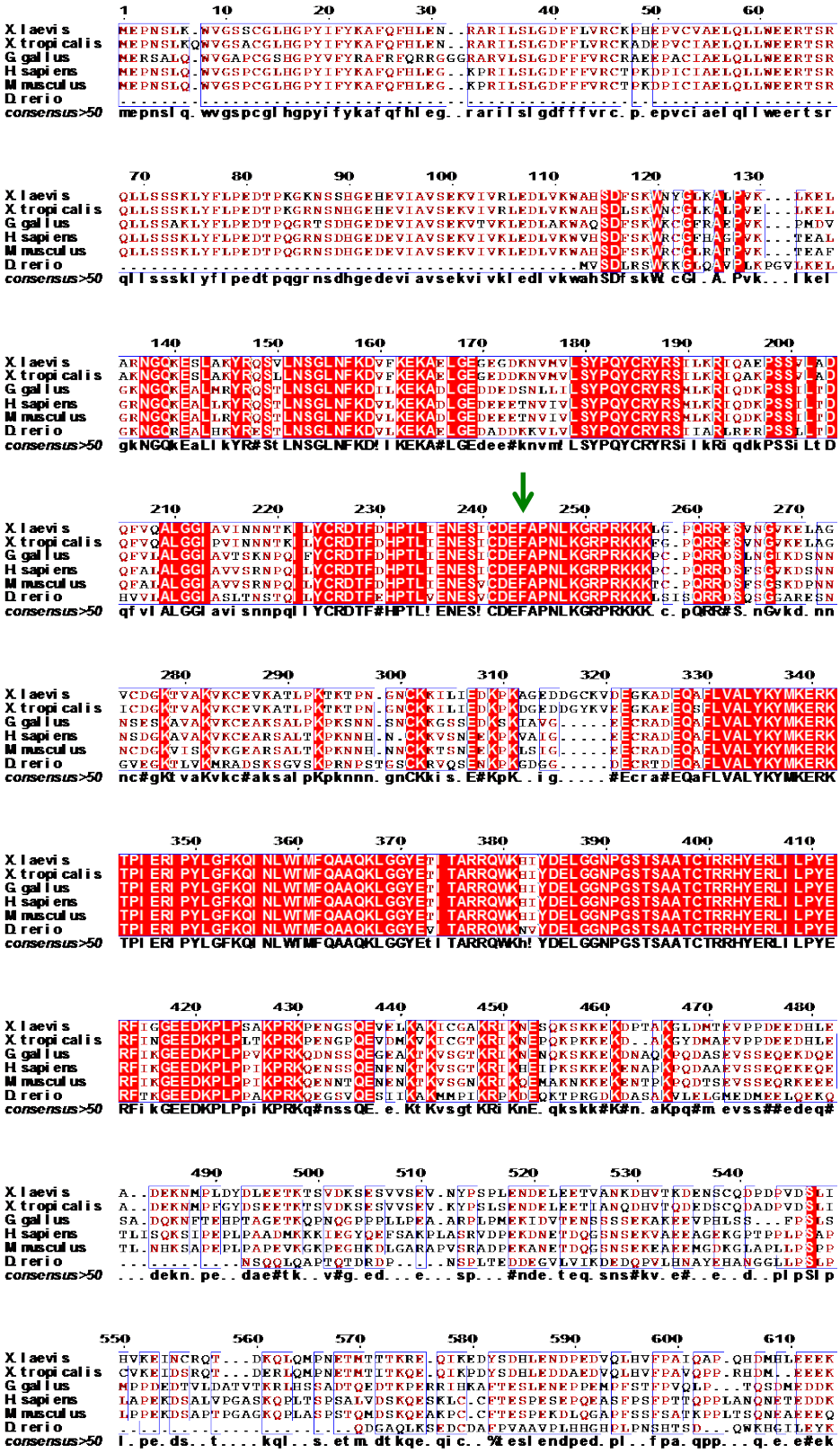


Figure 6

Developmental Expression Pattern



Developmental Expression Pattern

Figure S1

Developmental Expression Pattern

```
        1                                                                                 80
Isoform1 MEPNSLKWVG SSCGLHGPYI FYKAFQFHLE NRARILSLGD FFLVRCKPHE PVCVAELQLL WEERTSRQLL SSSKLYFLPE
Isoform2 .......... .......... .......... .......... .......... .......... .......... ..........
        81                                                                               160
Isoform1 DTPKGKNSSH GEHEVIAVSE KVIVRLEDLV KWAHSDFSKW NYGLKALPVK LKELARNGQK ESLAKYRQSV LNSGLNFKDV
Isoform2 .......... .......... .......... .......... .......... .......... .......... ..........
       161                                                                               240
Isoform1 FKEKAELGEG EGDKNVMVLS YPQYCRYRSI LKRIQAEPSS VLADQFVQAL GGIAVINNNT KILYCRDTFD HPTLIENESI
Isoform2 .......... .......... .......... .......... .......... .......... .......... ..........
        241                                                                              320
Isoform1 CDEFAPNLKG RPRKKKLGPQ RRESVNGVKE LAGVCDGKTV AKVKCEVKAT LPKTKTPNGN CKKILIEDKP KAGEDDGCKV
Isoform2 ...MPNLKG  RPRKKKLGPQ RRESVNGVKE LAGVCDGKTV AKVKCEVKAT LPKTKTPNGN CKKILIEDKP KAGEDDGCKV
        321                                                                              400
Isoform1 DEGKADEQAF LVALYKYMKE RKTPIERIPY LGFKQINLWT MFQAAQKLGG YETITARRQW KHIYDELGGN PGSTSAATCT
Isoform2 DEGKADEQAF LVALYKYMKE RKTPIERIPY LGFKQINLWT MFQAAQKLGG YETITARRQW KHIYDELGGN PGSTSAATCT
        401                                                                              480
Isoform1 RRHYERLILP YERFIGGEED KPLPSAKPRK PENGSQEVEL KAKICGAKRI KNESQKSKKE KDPTAKGLDM TEVPPDEEDH
Isoform2 RRHYERLILP YERFIGGEED KPLPSAKPRK PENGSQEVEL KAKICGAKRI KNESQKSKKE KDPTAKGLDM TEVPPDEEDH
        481                                                                              560
Isoform1 LEADEKNMPL DYDLEETKTS VDKSESVVSE VNYPSPLEND ELEETVANKD HVTKDENSCQ DPDPVDSLIH VKEINCRQTD
Isoform2 LEADEKNMPL DYDLEETKTS VDKSESVVSE VNYPSPLEND ELEETVANKD HVTKDENSCQ DPDPVDSLIH VKEINCRQTD
        561                                                                              640
Isoform1 KQLQMPNETM TTTKREQIKE DYSDHLENDP EDVQLHVFPA IQAPQHDMHL EEEKLPDMPD YIANCTVKVD PLGSNDLKNP
Isoform2 KQLQMPNETM TTTKREQIKE DYSDHLENDP EDVQLHVFPA IQAPQHDMHL EEEKLPDMPD YIANCTVKVD PLGSNDLKNP
        641                                                                              720
Isoform1 LDSNLLQNAL KQNPKVYFVQ TLDMLSDEKD TSASMNDDSS FSYTPLLYSR GNPGIMSPLA KKKLLSQVSG ASQPGNLPYG
Isoform2 LDSNLLQNAL KQNPKVYFVQ TLDMLSDEKD TSASMNDDSS FSYTPLLYSR GNPGIMSPLA KKKLLSQVSG ASQPGNLPYG
        721                                                                              800
Isoform1 SPPPLISKKK LSSKGEVSPS LLQTHHSSNS ESAAINRPSV IQHVQSFKQK SPEEKKTVND HYKNSMFGKV DSYCCDFAKH
Isoform2 SPPPLISKKK LSSKGEVSPS LLQTHHSSNS ESAAINRPSV IQHVQSFKQK SPEEKKTVND HYKNSMFGKV DSYCCDFAKH
        801                                                                              880
Isoform1 HQSVLADSYA LKSCVQECKE KMAEKRAASN SNVPSFVAEF YSSPHLHRLY RQAEHHLHNE NSAKFHSREM FRDLENVSSH
Isoform2 HQSVLADSYA LKSCVQECKE KMAEKRAASN SNVPSFVAEF YSSPHLHRLY RQAEHHLHNE NSAKFHSREM FRDLENVSSH
        881                                                                              960
Isoform1 KHHYHASLHQ HDKQNLHDDV DDQPTDLSLP KSLHKLSTKI PGSSICHQPV QQDSKSHNPF QTPNSKTLGL DCNPKACRVS
Isoform2 KHHYHASLHQ HDKQNLHDDV DDQPTDLSLP KSLHKLSTKI PGSSICHQPV QQDSKSHNPF QTPNSKTLGL DCNPKACRVS
        961                                                                             1040
Isoform1 PMTMPISKRH MDSIQRPSKT VKPDTLRKVE GLVHPFSIGK TNTHNFGAPR SLKRNLEDMD NPLTDKKIRA VSPLHLPKEM
Isoform2 PMTMPISKRH MDSIQRPSKT VKPDTLRKVE GLVHPFSIGK TNTHNFGAPR SLKRNLEDMD NPLTDKKIRA VSPLHLPKEM
        1041                                                                            1120
Isoform1 SGKDTFVGQD GESSKSVHDI HSGSMIESHK YPLSSPFFPG MYPGSLCGGL SSRIPTAYSH PLQYLKNQTA LSPLMQPLAL
Isoform2 SGKDTFVGQD GESSKSVHDI HSGSMIESHK YPLSSPFFPG MYPGSLCGGL SSRIPTAYSH PLQYLKNQTA LSPLMQPLAL
        1121                                                        1187
Isoform1 HTFMMQRQYL TNSTNSQQLY RQIASHAPVG SSYGDLLHSS IYPLTAINPQ SPFPSSQMSS VYPSTKL
Isoform2 HTFMMQRQYL TNSTNSQQLY RQIASHAPVG SSYGDLLHSS IYPLTAINPQ SPFPSSQMSS VYPSTKL
```

Figure S2

Developmental Expression Pattern

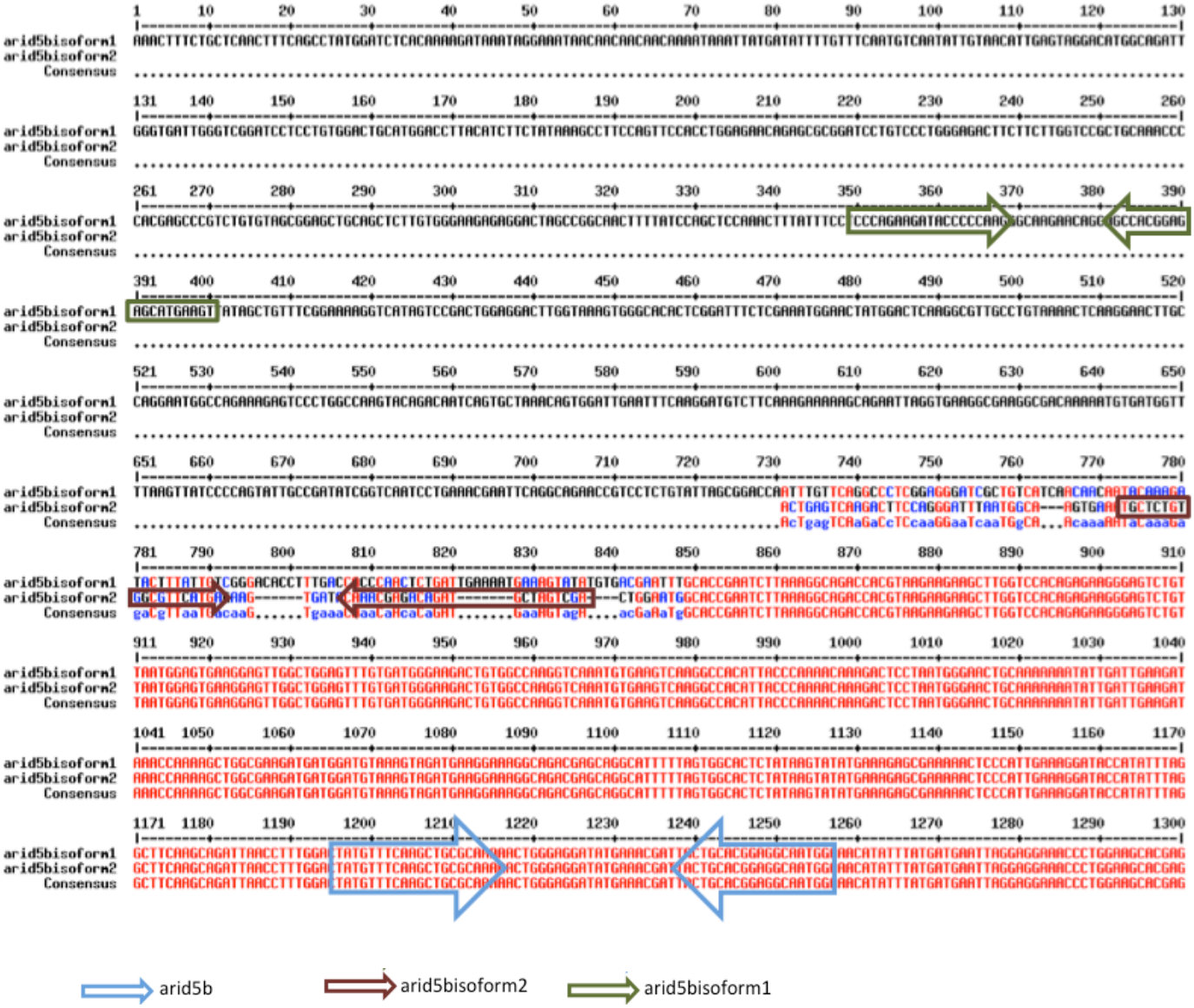


Figure S3

Developmental Expression Pattern

Table S1

Exon n°	Exon size (bp)	Sequence at exon-intron junction: 5' splice donor	Sequence at exon-intron junction: 3' splice acceptor	Intron size (bp)	Amino acid interupted
1	>27	ACTCAAG**gt**attattagctg	tcttgc**ag**TGGGTCGG	328	Trp-8
2*	>68	?	attttc**ag**CATGAAGT	>39437	Glu-92?
3	226	GAATTAG**gt**aaccatggttc	aaccac**ag**TGAAGGCG	40399	His-93
4	231	AATTTGG**gt**aagttccaatc	ttttac**ag**CACCGAAT	66835	Gly-168
5	113	GGCCAAG**gt**aaatccttttt	tctttc**ag**GTCAAATG	1316	Lys-282
6	220	AAGCAGA**gt**aagtagacact	tctaac**ag**TTAACCTT	4458	Ile-356
7	53	TGAAACG**gt**aagtgctttta	tataac**ag**ATTACTGC	3545	Ile-374
8	98	ACGAAAG**gt**aggataatcta	ccccac**ag**ATTAATCC	33496	Arg-406
9	199	GACTGAG**gt**aaattggggca	ttgttt**ag**GTTCCACC	5264	Val473
10	>2712				
4b	>99	GGAATGG**gt**aaccgcttcag	tctttc**ag**GTCAAATG	6493	Ala-2 (short isoform)

Scaffold 9729. * sequence nt89-275 mRNA ORFmissing in genomic sequence

Table S2

Exon n°	Exon size (bp)	Sequence at exon-intron junction: 5' splice donor	Sequence at exon-intron junction: 3' splice acceptor	Intron size (bp)	Amino acid interupted
1*	>27	ACTCAAG**gt**attattagctg	?	?	Trp-8
2	236	CGGAGAG**gt**gacatccctac	ttttc**ag**CATGAAGT	53173	His-93
3	226	GAACTAG**gt**aaccatgattc	aaccac**ag**GTGAAGGC	53008	Gly-168
4	231	GAATTTG**gt**aagttccaatt	ttttc**ag**CACCAAAT	77336	Ala-245
5	113	TGCCAAG**gt**aaatcggtttt	tgtttt**ag**GTCAAATG	1264	Val-283
6	220	AAGCAGA**gt**aagtagacatt	tctaac**ag**TTAACCTT	4033	Ile-356
7	53	TGAAACG**gt**aagtgcttttt	tgtaac**ag**ATTACTGC	4690	Ile-374
8	98	ACGAAAG**gt**aggataatctg	ccccac**ag**ATTAATCC	33865	Arg-406
9	199	GGCTGAG**gt**atattggggca	ttgttc**ag**GTTCCACC	4311	Val-473
10	>2705				
4b	>96	GGAATGG**gt**aaccgctttat	ttttc**ag**CACCAAAT	5945	Ala-2 (short isoform)

Scaffold 48311 * sequence nt21-39 mRNA ORFmissing in genomic sequence