HMGA1 zebrafish co-orthologue hmga1b can modulate p53-dependent cellular responses but is unable to control the alternative splicing of *psen1*

Seyed Hani Moussavi Nik [1,†], Morgan Newman [1], Amanda Lumsden [2], Tanya Jayne [1] and Michael Lardelli [1]

Affiliations:

1- Alzheimer's Disease Genetics Laboratory, Center for Molecular Pathology, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia.
2- College of Medicine and Public Health, and Center for Neuroscience, Flinders University, Adelaide, South Australia, Australia.

† Corresponding Author:
Dr. Seyed Hani Moussavi Nik
Alzheimer's Disease Genetics Laboratory, Center for Molecular Pathology, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia.
Email: Seyyed.moussavinik@adelaide.edu.au

ABSTRACT

The HIGH MOBILITY GROUP AT-HOOK 1 (HMGA1) family of chromatin-binding proteins plays important roles in cellular responses to low oxygen. HMGA1 proteins regulate gene activity both in the nucleus and within mitochondria. They are expressed mainly during embryogenesis and their upregulation in cancerous cells indicates poor prognosis. The human HMGA1a isoform is upregulated under hypoxia via oxidative stress-dependent signalling and can then bind nascent transcripts of the familial Alzheimer's disease gene PSEN2 to regulate alternative splicing to produce the truncated PSEN2 protein isoform PS2V. Zebrafish where hmga1a expression is induced by hypoxia to control splicing of the psen1 gene to produce the PS2V-equivalent isoform PS1IV. Zebrafish possess a second gene with apparent HMGA1 orthology, hmga1b. Here we investigate the predicted structure of Hmga1b protein and demonstrate it to be co-orthologous to human HMGA1 and most similar in structure to human isoform HMGA1c. We show that forced over-expression of either hmga1a or hmga1b mRNA can suppress the action of the cytotoxin hydroxyurea in stimulating cell death and transcription of the genes mdm2 and cdkn1a that, in humans, are controlled by p53. Our experimental data support an important role for HMGA1 proteins in modulation of p53-dependent responses and illuminate the evolutionary subfunctionalisation.

INTRODUCTION

The protein HIGH MOBILITY GROUP AT-HOOK 1 (HMGA1) is notable for numerous reasons. It is the only protein to be exported from the nucleus to act in mitochondria [1,2] and its presence in mitochondria changes during the cell cycle [2]. The protein is induced by the oxidative stress produced by hypoxia [3,4] and so is present during early embryogenesis in many vertebrate species. It can also be upregulated during cancer progression and is regarded as an indicator of poor prognosis [5].

Cancerous tumours present hypoxic environments in which HMGA1 may be activated. Therefore, there has been considerable interest in the possible interaction of this protein with the tumour suppressor p53, that is very commonly lost during oncogenesis [6]. An inhibitory interaction between the HMGA1 and the pro-apoptotic functions of p53 has been reported although retraction of a paper central to this research area has led to uncertainty over these ideas [7].

The human *PRESENILIN* genes, *PSEN1* and *PSEN2* are loci for approximately 80% of the mutations causing early onset, dominantly-inherited, familial Alzheimer's disease (on a population frequency basis [8]). HMGA1 not only binds chromatin but has also been demonstrated to bind to *PSEN2* transcripts to control hypoxia-induced alternative splicing of this gene in humans. Binding of HMGA1 protein to *PSEN2* transcripts results in exclusion of exon 5 sequence leading to a frameshift and premature termination of the coding sequence [9-11]. These alternative "PS2V" spliced transcripts are not completely removed by nonsense-mediated decay and so can be translated to produce a truncated form of PSEN2 protein that suppresses the unfolded protein response (UPR) to reduce the cell death that can be induced by hypoxia. PS2V is thus part of a negative feedback mechanism that limits UPR activation [11,12]. We have previously demonstrated that this mechanism of HMGA1 action exists also in zebrafish but acts, instead, on the transcripts of the zebrafish's *PSEN1*-orthologous gene, *psen1*, rather than its *PSEN2* orthologue. This is strong evidence that the hypoxia-induced, HMGA1-dependent alternative splicing of *PRESENILIN* transcripts predates the duplication and divergence of the common ancestor of the *PSEN1* and *PSEN2* genes early in vertebrate evolution [3,11]. The subfunctionalisation of *PSEN1* and *PSEN2* that occurred after the duplication event presumably occurred after the divergence of the teleost and tetrapod evolutionary lineages over 400 million years ago [13] The HMGA1-controlled splicing mechanism was lost from the *psen2* gene lineage of teleosts and the *PSEN1* lineage of the tetrapods, including humans.

We have previously shown that, in zebrafish, the *hmga1a* gene is orthologous to human *HMGA1*, is upregulated in embryos and adult brains by hypoxia, and can drive formation of the zebrafish PS2V-equivalent isoform, PS1IV. However, zebrafish possess a second "co-orthologue" of *HMGA1* named *hmga1b* that has not been characterised. In this paper we analyse the putative structure of the *hmga1b* protein product, its phylogenetic relationship to other *HMGA1* genes, and its embryonic expression. *hmga1b* transcription is regulated by hypoxia but, unlike *hmga1a*, the protein product of *hmga1b* cannot regulate alternative splicing of *psen1* transcripts to produce isoform PS1IV. Nevertheless, and similar to *hmga1a*, it appears able to suppress the regulatory effects of p53 on target genes so as to inhibit cell death. Thus, HMGA1 proteins act at multiple levels to suppress the cell death caused by hypoxia and oxidative stress.

Materials and Methods

Ethics
This work was conducted under the auspices of The Animal Ethics Committee of The University of Adelaide and in accordance with EC Directive 86/609/EEC for animal experiments and the Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

Zebrafish husbandry and experimental procedures
Danio rerio were bred and maintained at 28°C on a 14 h light/10 h dark cycle. Embryos were collected from natural mating, grown in embryo medium (E3) and staged [14]. Treatment of embryos at 48 hours post fertilisation (hpf) and adult zebrafish with low oxygen and mimicry of hypoxia using sodium azide (NaN_3) has been described previously [3]. For treatment with hydroxyurea (HU), embryos were divided into two groups after 24 hpf and hydroxyurea was added to one of the two groups to a concentration of 250 mM. Incubation of both groups then continued until the developmental stage equivalent to that of untreated embryos at 48 hpf (at 28.5°C). Embryos were then processed for further analysis.

Phylogenetic analyses
Multiple alignment of HMGA-related proteins in human and zebrafish was done using ClustalW with the gap opening penalty of 10.0 and gap extension penalty of 3.0 [15]. Phylogenetic analysis was conducted using MrBayes [16], the program was run under the JC69 (Juke-Cantor) model, with rate variation gamma and unconstrained branch lengths of 10.

Injection of mRNAs
Embryos were collected from natural mating, grown in embryo medium (E3) and staged. *In vitro*-synthesised mRNAs were injected into one-cell stage embryos as previously described [17].

Quantitative real-time PCR assay
Total RNA was extracted from adult brain or embryos using the QIAGEN RNeasy mini kit (QIAGEN, GmbH, Hilden, Germany). cDNA synthesis, quantitative real-time PCR (qPCR) assay, and data analysis has been described previously [11]. Gene specific primers used for qPCR are given in Supplementary Data File 2.

Western immunoblot analyses
At 48 hpf embryos were dechorionated, deyolked and placed in lithium dodecyl sulfate (LDS) sample buffer (NuPAGE® LDS sample buffer 4X, NuPAGE® Reducing reagent 10X), heated immediately at 80°C for 25 min and then stored at −80°C prior to separation on a 4-12% NuPAGE® Gel. Western blotting and transfer of protein onto nitrocellulose membranes was performed as previously described [17]. For incubation with anti-Cleaved Caspase 3 (Asp175) antibody (Cell Signaling Technology Inc., Danvers, MA), primary antibody was diluted 1 in 1000 in TBST containing 2% w/v skim milk. The membranes were then washed in TBST and incubated with a goat anti-rabbit IgG secondary antibody (Cell Signaling Technology Inc., Danvers, MA.) diluted 1 in 3000 in TBST containing 2% w/v skim milk. After incubation with the secondary antibody the membranes were washed and visualised as previously described [17]. Incubation and detection of anti-β-tubulin antibodies (Antibody E7, Developmental Studies Hybridoma Bank, The University of Iowa, IA, USA) was done as previously described [17].

Assessment of relative cell death
To visualise and quantify the relative cell death produced by hydroxyurea treatment in 48 hpf embryos, acridine orange (AO) assays were performed as previously described [3] except that quantification was performed on the tail region of 6 embryos per treatment (to avoid a signal

from the developing lens).

Promoter analysis

Genomatix Gene2Promoter software (Genomatix AG, Munich, Germany) was used to extract the promoter sequences of human *HMGA1*, mouse *Hmga1*, bovine *HMGA1*, chicken *HMGA1*, and zebrafish *hmga1a* and *hmga1b*. Sequences consisted of 1000 bp upstream of the first transcription start site (TSS) to 100 bp downstream of the last TSS in each analysed promoter region. Sequences were examined using the *CommonTFs* function of MatInspector software (Genomatix AG [18]) to identify transcription factor matrix families (Vertebrate and General Core Promoter Elements) common to *HMGA1* gene orthologues from human, mouse, cow and chicken, and at least one zebrafish paralog. Default stringency conditions were used (Minimum Matrix Core similarity, 0.075; Minimum matrix similarity, 'Optimized'). Matches with a significant frequency of occurrence (p-value less than 0.05, based on expected frequency in promoter sequences) were extracted and presented, with position within the extracted sequence (in the DNA strand phase of the gene) given.

Results

Confirmation of zebrafish *hmga1b* as a co-orthologue of human *HMGA1*

We previously identified a zebrafish orthologue of the human *HMGA1* gene, at that time named *hmga1* [3]. However, subsequent examination of zebrafish genome sequence led to putative identification of a second *HMGA1* "co-orthologous" gene, a common situation in zebrafish genetics [19]. This second gene is now named *hmga1b* while the previously characterised gene is now *hmga1a*. The *hmga1b* gene is located on zebrafish chromosome 6 (RefSeq NM_ 001077276.1).

The human *HMGA1* gene expresses three protein isoforms (derived from alternative transcript splicing), HMGA1a, HMGA1b, and HMGA1c. These differ in the spacing and number of their DNA-binding AT-hook domains (see Figure 1A, B). Overall, the amino acid (aa) residue identity between the human HMGA1 and the putative protein sequence of zebrafish Hmga1b is 59 %. The two most N-terminal AT-hook domains in Hmga1b share identical aa sequence with human HMGA1 and zebrafish Hmga1a. However, while the most C-terminal AT-hook domain is also identical between human HMGA1 and zebrafish Hmga1a, this domain shows divergence in zebrafish Hmga1b. Hmga1b also apparently lacks the acidic tail domain that is conserved in zebrafish Hmga1a and human HMGA1a and HMGA1b (although not in isoform HMGA1c, see [20], Figure 1A, B).

To confirm the orthology relationships between human *HMGA1* and zebrafish *hmga1a* and *hmga1b*, we conducted a phylogenetic analysis comparing the zebrafish genes with DNA sequences of the *HMGA1* orthologues of human, mouse, bovine and chicken, as well as the most closely related *HMGA* gene family, the *HMGA2* genes, of these species (Figure 1C). (See Table S1 in Supplementary Data File 1 for sequence accession numbers and description of the exact sequences from the multiple cDNA sequence alignment used in the analysis). The close association between zebrafish *hmga1a* and *hmga1b* strongly supports that these two genes are co-orthologues of human *HMGA1*.

Expression of zebrafish *hmga1b*

The temporal expression of zebrafish *hmga1b* transcripts was observed by qRT-PCR analysis (relative to expression of housekeeping gene *eef1a1l1*) using three independent biological replicates for each developmental time point (the qPCR raw data and statistical analysis is available in Supplementary Data File 2). Particular adult tissues were also examined. Zebrafish *hmga1b* is present in embryos at two hours post fertilisation (hpf), before the mid-blastoderm transition when the zygotic genome becomes transcriptionally active [21]. This indicates a maternal contribution of *hmga1b* mRNA to the embryo. We observed a peak of *hmga1b* mRNA levels at 48 hpf when embryogenesis ends with hatching, followed by a decline over subsequent days (Figure 2). Expression of *hmga1b* transcripts in adult brain, liver and kidney appeared lower than in whole animals up to 120 hpf (Figure 2).

Common hypoxia-related regulatory elements in promoters of HMGA1 homologs

The genes *hmga1a* (on chromosome 23) and *hmga1b* (on chromosome 6) probably arose in the teleost lineage-specific whole genome duplication that occurred over 300 million years ago [22]. This is supported by the synteny of each gene with other genes that show similarity between the two chromosome regions. For example, the genes *nudt3a* and *pacsin1a* lie close downstream of *hmga1a* while genes *nudt3b* and *pacsin1b* are close downstream of *hmga1b* (Ensembl release 90, zebrafish GRCz10 [23]).

After a gene duplication event, subfunctionalisation may occur where one of the duplicates uniquely retains particular activities of the ancestral gene while these are lost from the other duplicate copy (and vice-versa [24]) The duplicate genes may also show some redundant function. Human *HMGA1* is both an important regulator of other genes and shows strong regulatory responses itself. For example, *HMGA1* is commonly upregulated in oncogenesis and by hypoxia [25]. To identify transcription factor binding sites (TFBSs) that may be functionally relevant in the regulation of expression of HMGA1 homologs, we performed a comparative promoter analysis to find TFBS motifs common to a range of mammalian and avian *HMGA1* orthologues (human, mouse, bovine, chicken) that were also present in one or both of the zebrafish paralogs *hmga1a* and *hmga1b*. Promoter regions spanning 1000 nucleotides (nt) upstream of the first transcription start site (TSS) to 100 nt downstream of the last TSS in the analysed region were compared using the *CommonTFs* function of the program Genomatix MatInspector [18] to identify common transcription regulatory motifs. 28 matches were identified with a statistically significant frequency within the sequences (*p*-value <0.05; Table 2). All sequence positions and match scores of the regulatory motifs are presented in Supplementary Data File 2. Amongst the transcription factor binding motifs common to all the analysed homologs sites were found for factors involved in the hypoxic response [26] such as hypoxia inducible factor HIF family members (V$HIFF), and nuclear factor kappa B factors (V$NFKB) (Figure 3).

Testing regulation by hypoxia of zebrafish *hmga1b* transcript levels

In humans, hypoxia increases expression of *HMGA1* transcripts in neural tissue through induction of oxidative stress [9,27] and this explains the induction by hypoxia of the PS2V alterative transcript isoform from the *PSEN2* gene [9]. We previously demonstrated that this mechanism is conserved in zebrafish with increased expression of *hmga1a* transcripts causing induction of the PS2V-cognate isoform PS1IV from the zebrafish *psen1* gene [11]. In humans, the HMGA1a isoform of HMGA1 shows a greater ability to induce PS2V formation than HMGA1b [10] while the activity of isoform HMGA1c has not been tested in this regard. Structurally, the putative zebrafish Hmga1b protein resembles somewhat human HMGA1c in that Hmga1b possesses only two highly conserved AT-hook domains and lacks the long acidic C-terminal domain seen in HMGA1a, HMGA1b and zebrafish Hmga1a. Therefore, we examined experimentally whether hypoxia might also regulate expression of *hmga1b* and, if so, whether *hmga1b* expression could (like *hmga1a*) influence alternative splicing of *psen1* to form PS1IV.

Hypoxia (or chemical mimicry of hypoxia using NaN_3) increases the expression of zebrafish *hmga1a* transcripts both in whole embryos and adult brain tissue [3]. Therefore, we compared the relative levels of *hmga1b* mRNA in 48 hpf embryos under normoxia and chemical mimicry of hypoxia using qRT-PCR. Similarly, we examined the relative levels of *hmga1b* transcripts in the brains of adult zebrafish that had been exposed to normoxia or hypoxia. Under chemical mimicry of hypoxia or real hypoxia, relative induction of *hmga1b* (compared to transcript levels of the housekeeping gene *eef1a1l1*) is observed in 48 hpf embryos (Figure 4A) and adult brains (Figure 4B) respectively, to a degree similar to that of *hmga1a*.

Foreced over-expression of zebrafish *hmga1b* has no effect on zebrafish *psen1* alternative splicing

Forced embryonic expression of zebrafish *hmga1a* after injection of its mRNA into zebrafish zygotes causes increased alternative splicing of endogenous *psen1* transcripts to form PS1IV [11]. Therefore, we compared this ability of *hmga1a* mRNA with that of *hmga1b* mRNA to test whether the latter can also control *psen1* alternative splicing. We used qPCR to compare levels of the PS1IV form of *psen1* mRNA in the presence of an injected, non-translated mRNA (as a negative control) or *hmga1a* mRNA or *hmga1b* mRNA. Only injection of

hmga1a mRNA increased the level of PS1IV (Figure 5). Thus, only *hmga1a* appears able to regulate PS1IV production.

Zebrafish *hmga1b* decreases P53-dependent apoptosis in embryos

Upregulation of HMGA1 expression is a marker of poor prognosis in cancer [28]. Research into the promotion of cancer progression by HMGA1 has identified that this protein may inhibit, by various means, the function of the tumour suppressor p53 (e.g. [29]) although the research on this topic has been afflicted by controversy (e.g. see [7]). Therefore, we examined whether the *hmga1a* and *hmga1b* genes of zebrafish might show activity supporting an interaction with p53-mediated cell death. The external development of zebrafish embryos makes them highly amenable to studying the effects of drug exposure and the cytotoxin hydroxyurea inhibits DNA synthesis and can induce apoptosis via a p53-dependent pathway [30,31]. Therefore, we exposed embryos to 250 mM hydroxyurea to induce cell death and observed whether forced expression of *hmga1a* or *hmga1b* mRNA could reduce this.

Zygotes were injected with either an untranslatable mRNA (as a negative control) or with mRNA encoding proteins Hmga1a or Hmga1b before exposure to hydroxyurea at 24 hpf. At 48 hpf, embryos were then stained with acridine orange to visualise and quantify cell death (Figure 6A). Protein lysates from similarly-treated embryos were examined by western immunoblotting to quantify the levels of cleaved Caspase3 protein that is a marker of apoptosis (Figure 6B). We observed that both forced *hmga1a* expression and forced *hmga1b* expression could suppress hydroxyurea-driven cell death with high statistical significance.

In the retracted paper by Pierantoni et al.[7] it was reported that increasing dosage of a transgene expressing human isoform HMGA1b could increase p53-stimulated transcription of the gene *MDM2* and suppress p53-stimulated transcription of gene *CDKN1A* (previously known as *WAF1* or p*21*) in a transfected cell culture system. Consistent with this, we found that forced expression of either Hmga1a or Hmga1b by mRNA injection into zebrafish zygotes was able to suppress hydroxyurea-driven induction of transcription of the p53-activated gene *cdkn1a* and to increase the induction by hydroxyurea of transcription of the gene *mdm2* (Figure 6C). These data support the idea of an intimate interaction between Hmga1a/b and p53 in the regulation of cell death in vertebrates.

DISCUSSION

The human *HMGA1* gene has been studied substantially in relation to tumour progression and other disease. To gain better insight into *HMGA1* function and dysfunction, various animal models have been used. The zebrafish and its embryos are a highly manipulable vertebrate model for genetic studies due to a combination of advantageous features including easy embryo accessibility, large zygote size, relatively constant embryo size during its rapid development, short generation time and a well-characterised genome [32]. We have previously identified and studied a zebrafish co-orthologue of this gene, *hmga1a* [3].

The zebrafish *hmga1a* and *hmga1b* genes are co-orthologues of human *HMGA1*. They are duplicate genes that arose from an ancestral gene through the whole-genome duplication early in the evolution of the teleost lineage [22]. The structure of the putative protein coded by *hmga1b* differs significantly from that of *hmga1a* in the third (downstream) nucleic acid-binding AT-hook domain. This has diverged significantly from the consensus sequence preserved in both human HMGA1a and HMGA1b isoforms and zebrafish Hmga1a. In that sense, and in the lack of the C-terminal acidic tail domain, the putative protein product of *hmga1b* resembles somewhat the human HMGA1c isoform. This suggests subfunctionalisation of *hmga1b* to maintain functions similar to those of HMGA1c and

specific to this isoform. However, we note that the C-terminal of the Hmga1b protein consists of two acidic glutamine residues so we cannot yet state with confidence that its function in transcription control differs greatly from the acidic tail domain of Hmga1a.

The lack of conservation of the most C-terminal AT-hook domain in Hmga1b protein affects it ability to control alternative splicing of *psen1* transcripts to produce the PS1IV isoform that is equivalent to the PS2V isoform produced from human PSEN2. Unlike Hmga1a, the Hmga1b protein does not appear able to promote PS1IV formation. This implies that the HMGA1c isoform may not affect alternative splicing of *PSEN2* transcripts to generate PS2V in human cells, although this has not been tested experimentally. However, the structural difference between the zebrafish Hmga1a and Hmga1b co-orthologous proteins does not appear to affect their influence on the p53-regulated genes *cdkn1* and *mdm2*. This is also consistent with the data in the retracted Pierantoni et al paper [7] where a truncated form of HMGA1b lacking the downstream, third AT-hook domain retained the ability to bind to p53.

HMGA1 orthologous genes are expressed during vertebrate embryogenesis [3,33-35] and are also commonly upregulated during cancer progression [28]. They are upregulated in response to hypoxia via a mechanism dependent on reactive oxygen species [9,27]. We observed a peak of *hmga1b* transcript levels at 48 hpf, the time at which zebrafish embryogenesis ends with hatching from the chorion. This is consistent with a possible restriction of oxygen flow caused by the presence of the chorion during fish embryo development [36].

The differences and similarities in structure and function between the two zebrafish co-orthologues of human *HMGA1* are an excellent illustration of how such ancient duplicate genes can be serendipitous tools for dissection of gene/protein activity. This, combined with the ease of exposure of zebrafish embryos to hypoxia and drugs and their facility as an oncogenesis model [37] justify more extensive investigation of HMGA1 functions in zebrafish.

Acknowledgements

Authors Contributions

Designed the experiments and supervised the study (S.H.M.N and M.L). Performed the experiments (S.H.M.N., M.N., A.L., and T.J). Wrote the paper (M.L., and S.H.M.N). Helped review the manuscript (M.L., M.N, and A.L).

Competing Interests
The authors declare that they have no competing interests.

References

1   Reeves, R. Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* **277**, 63-81 (2001).

2   Dement, G. A., Treff, N. R., Magnuson, N. S., Franceschi, V. & Reeves, R. Dynamic mitochondrial localization of nuclear transcription factor HMGA1. *Exp Cell Res* **307**, 388-401, doi:S0014-4827(05)00159-X [pii]

10.1016/j.yexcr.2005.04.004 (2005).

3   Moussavi Nik, S. H., Newman, M. & Lardelli, M. The response of HMGA1 to changes in oxygen availability is evolutionarily conserved. *Exp Cell Res* **317**, 1503-1512, doi:S0014-4827(11)00139-X [pii]

10.1016/j.yexcr.2011.04.004 (2011).

4   Yanagita, T. *et al.* Possible involvement of the expression and phosphorylation of N-Myc in the induction of HMGA1a by hypoxia in the human neuroblastoma cell line. *Neuroscience letters* **374**, 47-52, doi:S0304-3940(04)01291-1 [pii]

10.1016/j.neulet.2004.10.039 (2005).

5   Huso, T. H. & Resar, L. M. The high mobility group A1 molecular switch: turning on cancer - can we turn it off? *Expert Opin Ther Targets* **18**, 541-553, doi:10.1517/14728222.2014.900045 (2014).

6   Powell, E., Piwnica-Worms, D. & Piwnica-Worms, H. Contribution of p53 to metastasis. *Cancer Discov* **4**, 405-414, doi:10.1158/2159-8290.CD-13-0136 (2014).

7   Pierantoni, G. M. *et al.* High Mobility Group A1 (HMGA1) proteins interact with p53 and inhibit its apoptotic activity. *Cell Death Differ* **13**, 1554-1563, doi:10.1038/sj.cdd.4401839 (2006).

8   Lanoiselée, H. M. *et al.* APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. *PLoS Med* **14**, e1002270, doi:10.1371/journal.pmed.1002270 (2017).

9   Manabe, T. *et al.* HMGA1a: sequence-specific RNA-binding factor causing sporadic Alzheimer's disease-linked exon skipping of presenilin-2 pre-mRNA. *Genes Cells* **12**, 1179-1191, doi:10.1111/j.1365-2443.2007.01123.x (2007).

10   Manabe, T. *et al.* Induced HMGA1a expression causes aberrant splicing of Presenilin-2 pre-mRNA in sporadic Alzheimer's disease. *Cell Death Differ* **10**, 698-708, doi:10.1038/sj.cdd.4401221 (2003).

11   Moussavi Nik, S. H. *et al.* Alzheimer's disease-related peptide PS2V plays ancient, conserved roles in suppression of the unfolded protein response under hypoxia and stimulation of γ-secretase activity. *Hum Mol Genet* **24**, 3662-3678, doi:10.1093/hmg/ddv110 (2015).

12   Katayama, T. *et al.* Induction of neuronal death by ER stress in Alzheimer's disease. *J Chem Neuroanat* **28**, 67-78, doi:10.1016/j.jchemneu.2003.12.004 (2004).

13   Broughton, R. E., Betancur-R, R., Li, C., Arratia, G. & Ortí, G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLoS Curr* **5**, doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e (2013).

14   Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 5th Edition*. (University of Oregon Press, Eugene, 2007).

15   Thompson, J. D., Higgins, D. G. & Gibson, T. J. 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 (1994).

16   Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).

17   Newman, M. *et al.* Differential, dominant activation and inhibition of Notch signalling and APP cleavage by truncations of PSEN1 in human disease. *Human molecular genetics* **23**, 602-617, doi:10.1093/hmg/ddt448 (2014).

18   Cartharius, K. *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933-2942, doi:10.1093/bioinformatics/bti473 (2005).

19   Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. & Van de Peer, Y. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* **13**, 382-390, doi:10.1101/gr.640303 (2003).

20   Nagpal, S. *et al.* Retinoid-dependent recruitment of a histone H1 displacement activity by retinoic acid receptor. *J Biol Chem* **274**, 22563-22568 (1999).

21   Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol* **30**, 581-613, doi:10.1146/annurev-cellbio-100913-013027 (2014).

22   Glasauer, S. M. & Neuhauss, S. C. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics* **289**, 1045-1060, doi:10.1007/s00438-014-0889-2 (2014).

23   Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res* **44**, D710-716, doi:10.1093/nar/gkv1157 (2016).

24   Mazet, F. & Shimeld, S. M. Gene duplication and divergence in the early evolution of vertebrates. *Curr Opin Genet Dev* **12**, 393-396 (2002).

25   Resar, L. M. The high mobility group A1 gene: transforming inflammatory signals into cancer? *Cancer Res* **70**, 436-439, doi:10.1158/0008-5472.CAN-09-1212 (2010).

26   Kenneth, N. S. & Rocha, S. Regulation of gene expression by hypoxia. *Biochem J* **414**, 19-29, doi:10.1042/BJ20081055 (2008).

27   Sato, N. *et al.* A novel presenilin-2 splice variant in human Alzheimer's disease brain tissue. *J Neurochem* **72**, 2498-2505 (1999).

28   Sumter, T. F. *et al.* The High Mobility Group A1 (HMGA1) Transcriptome in Cancer and Development. *Curr Mol Med* **16**, 353-393 (2016).

29   Frasca, F. *et al.* HMGA1 inhibits the function of p53 family members in thyroid cancer cells. *Cancer Res* **66**, 2980-2989, doi:10.1158/0008-5472.CAN-05-2637 (2006).

30   El Husseini, N., Schlisser, A. E. & Hales, B. F. Editor's Highlight: Hydroxyurea Exposure Activates the P53 Signaling Pathway in Murine Organogenesis-Stage Embryos. *Toxicol Sci* **152**, 297-308, doi:10.1093/toxsci/kfw089 (2016).

31   Gao, L. *et al.* TopBP1 Governs Hematopoietic Stem/Progenitor Cells Survival in Zebrafish Definitive Hematopoiesis. *PLoS Genet* **11**, e1005346, doi:10.1371/journal.pgen.1005346 (2015).

32   Newman, M., Ebrahimie, E. & Lardelli, M. Using the zebrafish model for Alzheimer's disease research. *Front Genet* **5**, 189, doi:10.3389/fgene.2014.00189 (2014).

33   Ruyter-Spira, C. P. *et al.* Nucleotide sequence of the chicken HMGI-C cDNA and expression of the HMGI-C and IGF1 genes in autosomal dwarf chicken embryos. *Biochim Biophys Acta* **1399**, 83-87 (1998).

34   Macrì, S. *et al.* Expression and functional characterization of Xhmg-at-hook genes in Xenopus laevis. *PLoS One* **8**, e69866, doi:10.1371/journal.pone.0069866 (2013).

35   Chiappetta, G. *et al.* High level expression of the HMGI (Y) gene during embryonic development. *Oncogene* **13**, 2439-2446 (1996).

36   Matschak, T. W., Stickland, N. C., Crook, A. R. and Hopcroft, T. Is physiological hypoxia the driving force behind temperature effects on muscle development in embryonic Atlantic salmon (*Salmo salar* L.)? *Differentiation* **59**, doi:10.1046/j.1432-0436.1995.5920071.x (1995).

37   Idilli, A. I., Precazzini, F., Mione, M. C. & Anelli, V. Zebrafish in Translational Cancer Research: Insight into Leukemia, Melanoma, Glioma and Endocrine Tumor Biology. *Genes (Basel)* **8**, doi:10.3390/genes8090236 (2017).

Figure 1. A) Amino acid residue sequence alignment of the zebrafish Hmga1b putative protein sequence and human HMGA1 performed using ClustalW. B) Schematic illustration comparing the structural differences between different HMGA1 proteins in Human (*Hs*) and zebrafish (*Dr*). The dark grey shadinged areas indicate the AT-hook DNA binding domains (DBD) characterised by the sequence **K/R(X)RGRP**. The black shaded regions indicate the acidic tails of the proteins. A less conserved AT-hook-like domain near the C-terminal end of Hmga1b is indicated in stippled light grey. Additional protein sequence from an additional exon included in HMGA1a and HMGA1c is indicated by *. Sequence divergence between HMGA1c and the other HMGA1 isoforms due to alternative splicing is indicated by dotted lines and +. Numbers indicated amino acid residues. C) Phylogenetic tree of the high mobility group A protein family generated using MrBayes. Sequences used in the phylogenetic analysis are shown in Table S1 in Supplementary Data File 1.

Figure 2. Temporal and tissue variation in expression of zebrafish *hmga1b* transcripts. Analysis of *hmga1b* expression by qPCR at various developmental time-points and in various adult tissues relative to housekeeping gene *eef1a1l1*. Means and standard error of the means are shown for three biological replicates.

Figure 3. Location of transcription factor binding sites related to hypoxia response, within a selection of vertebrate *HMGA1* orthologue promoters. Genomatix schematic output modified for clarity. Genomic sequence coordinates and associated transcripts are presented in Table 1.

Figure 4. Response of *hmga1b* expression to (**A**) mimicry of hypoxia in embryos and (**B**) hypoxia in adult brain compared to *hmga1a* expression responses. qPCR analysis of *hmga1b* and *hmga1a* expression relative to eukaryotic translation factor *eef1a1l1*. Only relative responses to hypoxia, not absolute expression levels, can be compared between the two genes. Means and standard error of the mean are shown for three biological replicates (n=3). *P*-values for two-tailed T-tests (unequal variances) for bracketed comparisons are shown. N, Normoxia. H, Hypoxia.

Figure 5. Injection of embryos with *hmga1a* mRNA, but not *hmga1b* mRNA, induces the formation of the *psen1* transcript isoform PS1IV under normoxia. qPCRs detecting, specifically, full-length *psen1* RNA or the alternatively-spliced PS1IV transcript isoform indicate induction of PS1IV by forced expression under normoxia of zebrafish Hmga1a, but not Hmga1b. Note that only comparisons of responses to forced expression of the same transcripts (i.e. full-length or PS1IV) are valid. Expression data is not comparable between full-length *psen1* and PS1IV. Expression of *psen1* and PS1IV is shown relative to eukaryote translation factor gene *eef1a1l1*. Means and standard error of the means are shown for three biological replicates in each case (n=3). *P*-values for two-tailed T-tests (unequal variances) for bracketed comparisons are shown. n.s., not statistically significant. Control TAA mRNA is an untranslatable control mRNA. No significant difference in levels of full-length *psen1* mRNA was observed for any treatment.

Figure 6. Injection of embryos with *hmga1a* or *hmga1b* mRNA decreases p53-dependent apoptosis. A) Embryos injected with *hmga1a* or *hmga1b* mRNA and treated with 250 mM hydroxyurea show decreased acridine orange staining indicating decreased cell death. B) Embryos injected with *hmga1a* or *hmga1b* mRNA and treated with 250 mM hydroxyurea show decreased cleaved Caspase-3 levels by western immunoblotting in comparison to injection with non-translatable Control TAA mRNA. C) Embryos injected with *hmga1a* and

hmga1b mRNA and treated with 250 mM hydroxyurea show altered expression of genes related to p53-dependent apoptosis (*cdkn1a* and *mdm2*) when assayed by qPCR. *P*-values for two-tailed T-tests (unequal variances) for bracketed comparisons are shown. n.s., not statistically significant.

Homolog	Genome build	Chromosome & contig ID	Coordinates (length)	Strand	TSS	Transcripts
Human *HMGA1*	GRCh38	Chr. 6 NC_000006	34235800-34237304 (1505 bp)	+	1001	NM_001319082 NM_145899 NM_001319080 NM_002131 NM_001319078 NM_145903 NM_001319079 NM_001319077 NM_001319081
					1074	NM_145901 NM_145902 AK096863 ENST00000347617 ENST00000401473 ENST00000311487
					1075	AK301434
					1104	AK308245
					1405	ENST00000478214
Mouse *Hmga1*	NCBI build 38	Chr. 17 NC_000083	27555574-27556768 (1195 bp)	+	1001	NM_001166535 NM_001166543 NM_001166540 NM_001039356 NM_001166539 NM_001166545 NM_001025427 NM_001166544 NM_001166546 NM_001166541 NM_001166536 NM_001166542
					1031	NM_016660 NM_001166537
					1048	AK171334
					1050	AK155661
					1068	ENSMUST00000119486
					1079	AK010617
					1080	AK169783
					1081	AK154957
					1085	ENSMUST00000118599
					1095	ENSMUST00000117254 ENSMUST00000118570
Bovine *HMGA1*	Bos taurus UMD 3.1.1	Chr. 23 AC_000180	8257735-8258850 (1116 bp)	+	1001	XM_010818062
					1002	XM_010818061
					1016	XM_005223268 NM_001076523 ENSBTAT00000048910
Chicken *HMGA1*	Gallus gallus 5.0	Chr. 26 NC_006113	4421556-4422656 (1101 bp)	-	1001	NM_204369
Zebrafish *hmga1a*	GRCz10	Chr. 23 NC_007134	3815634-3817789 (2156 bp)	-	1001	ENSDART00000028885
					1182	NM_001289906
					1232	ENSDART00000137707
					1257	ENSDART00000136394
					1348	ENSDART00000132205
					1818	NM_001289907
					1869	ENSDART00000137826
					1977	ENSDART00000139408
					2056	ENSDART00000131536
Zebrafish *hmga1b*	GRCz10	Chr. 6 NC_007117	54133090-54134338 (1249 bp)	+	1001	ENSDART00000165322
					1117	ENSDART00000156554
					1149	NM_001077276

Table 1: Genomic co-ordinates and associated transcripts of the promoter sequences of *HMGA1* homologs used for comparative analysis of transcription factor binding sites.

	Matrix Family	Description
a)	O$XCPE	Activator-, mediator- and TBP-dependent core promoter element for RNA polymerase II transcription from TATA-less promoters
	V$AHRR	AHR-arnt heterodimers and AHR-related factors
	V$CREB	cAMP-responsive element binding proteins
	V$DLXF	Distal-less homeodomain transcription factors
	V$EBOX	E-box binding factors
	V$ETSF	Human and murine ETS1 factors
	V$EVI1	EVI1-myleoid transforming protein
	V$GZF1	GDNF-inducible zinc finger gene 1
	V$HIFF	Hypoxia inducible factor, bHLH/PAS protein family
	V$HOMF	Homeodomain transcription factors
	V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D
	V$HUB1	HTLV-I U5 repressive element-binding protein 1
	V$KLFS	Krueppel like transcription factors
	V$MEF2	MEF2, myocyte-specific enhancer binding factor
	V$MYBL	MYB-like proteins
	V$NFKB	Nuclear factor kappa B/c-rel
	V$NKX6	NK6 homeobox transcription factors
	V$NKXH	NK-family of homeodomain transcription factors
	V$NR2F	Nuclear receptor subfamily 2 factors
	V$OCT1	Octamer binding protein
	V$PAX3	PAX-3 binding sites
	V$RXRF	RXR heterodimer binding sites
	V$SNAP	snRNA-activating protein complex
	V$SORY	SOX/SRY-sex/testis determining and related HMG box factors
	V$STAT	Signal transducer and activator of transcription
b)	V$CDEF	Cell cycle regulators: Cell cycle dependent element
	V$MIZ1	Myc-interacting Zn finger protein 1
	V$PAX9	PAX-9 binding sites

Table 2: Transcription Factor Matrices present in the *HMGA1* promoter from human, mouse, bovine and chicken, that are also present in **a)** zebrafish *hmga1a* and *hmga1b* promoter sequences, or **b)** zebrafish *hmga1a* but not *hmga1b*. In this analysis, no sites common to human, mouse, bovine and chicken *HMGA1* were detected in *hmga1b* that were not also found in *hmga1a*. .

A

Acridine orange assay in 48 hpf emrbryos treated with 250 mM hydroxyurea

n.s.

0.002

0.003

Median relative brightness

Un-injected | Control TAA mRNA | *hmga1a* mRNA | *hmga1b* mRNA
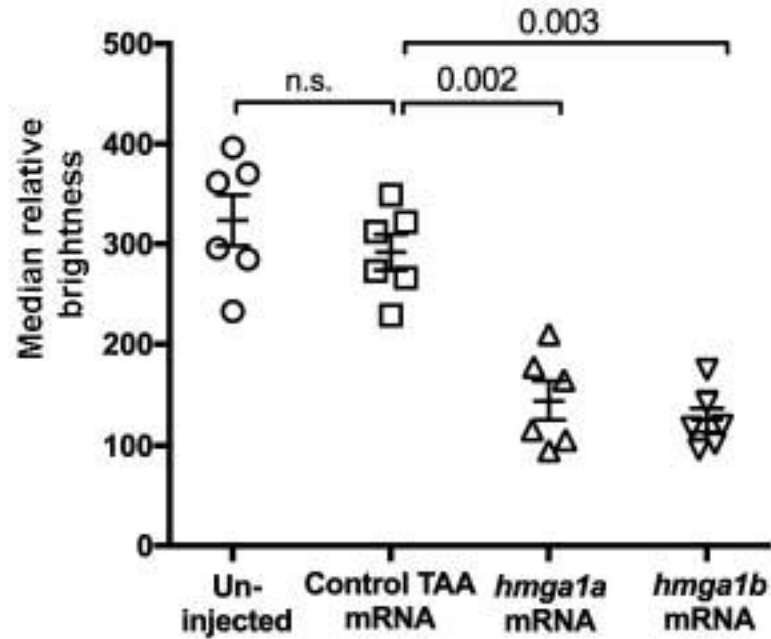
B

Cleaved Caspase3 in 48 hpf embryos treated with 250 mM hydroxyurea

0.068

0.007

n.s.

0.004

expression relative to β-Tublin

Un-injected | Control TAA mRNA | *hmga1a* mRNA | *hmga1b* mRNA

C

48 hpf embryos treated with 250 mM hydroxyurea

○ *cdkn1a*
□ *mdm2*

0.0004 0.0008 n.s.

0.004 0.01 n.s.

expression relative to *eef1a1l1*

Control TAA mRNA | Control TAA mRNA | *hmga1a* mRNA | *hmga1b* mRNA | Control TAA mRNA | Control TAA mRNA | *hmga1a* mRNA | *hmga1b* mRNA
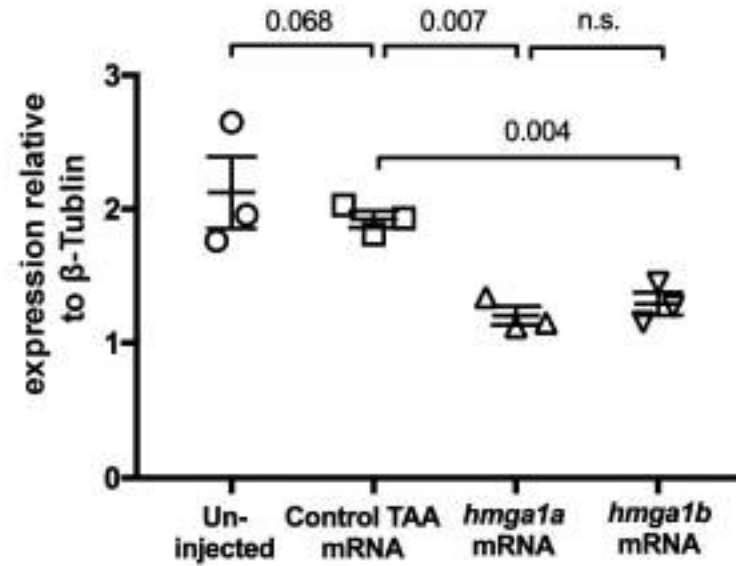
Treated with 250 mM hydroxyurea

Treated with 250 mM hydroxyurea
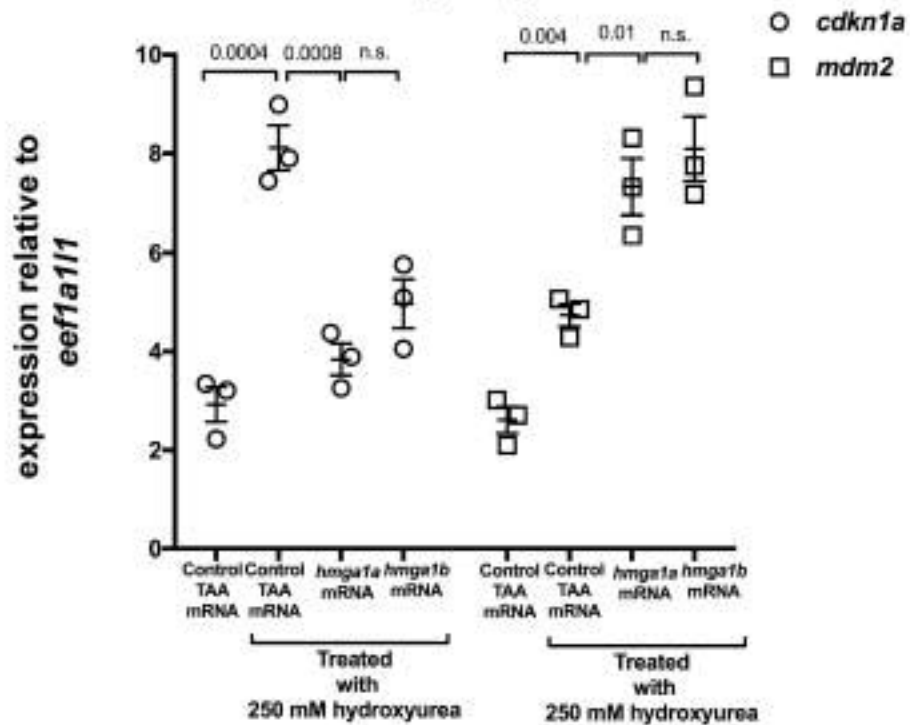
48 hpf embryos injected with
100 ng/µL of mRNA
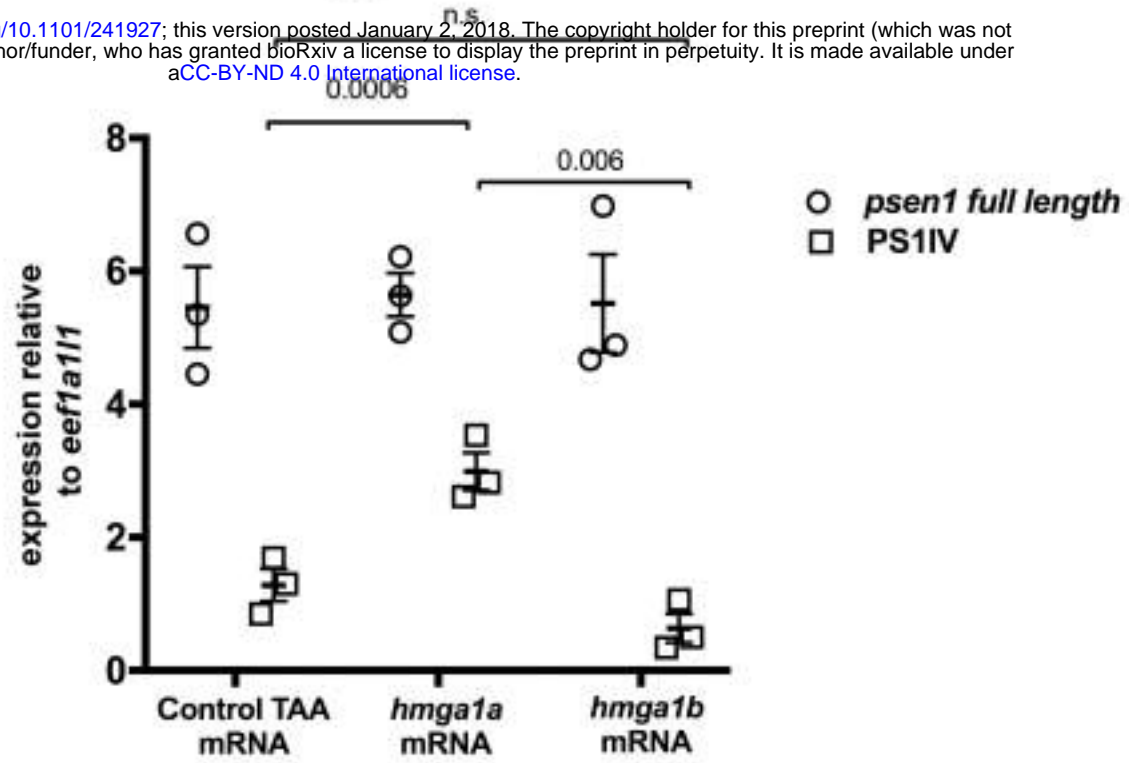
Figure 5

A

48hpf embryos under normoxia and mimicry of hypoxia
using100 μM sodium azide (NaN$_3$)
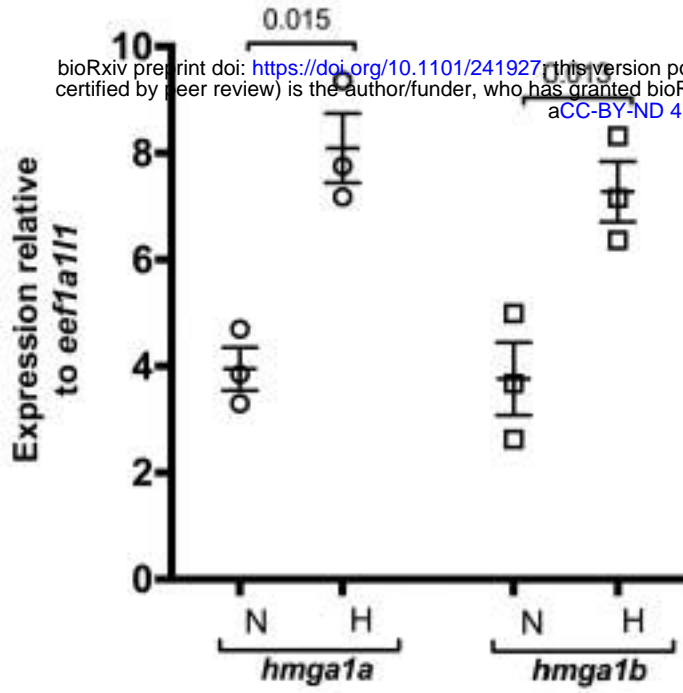
B

Adult brain under normoxia and hypoxia
(1.16±0.5 mg/L oxygen)
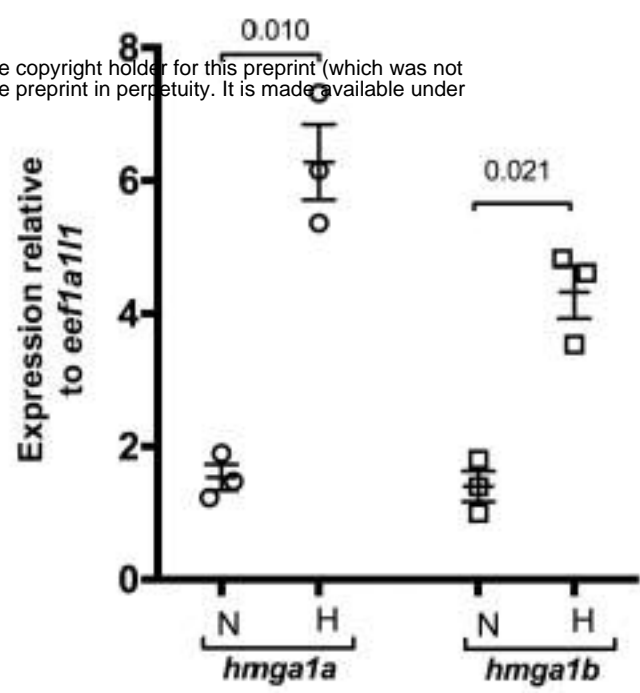
Figure 4.

Human *HMGA1a*
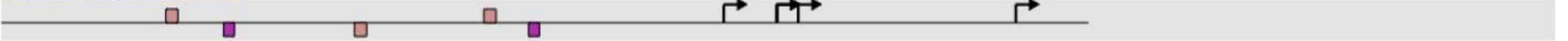
Mouse *HMGA1*

Bovine *HMGA1*
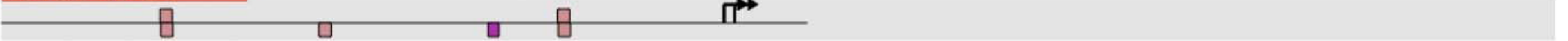
Chicken *HMGA1*
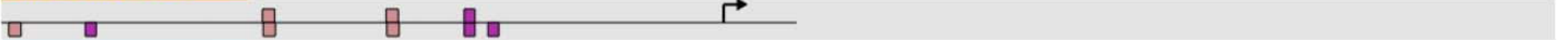
Zebrafish *hmga1a*
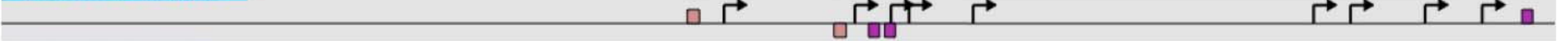
Zebrafish *hmga1b*

500 bp   1000 bp   1500 bp   2000 bp

Matrix families:

V$HIFF   V$NFKB

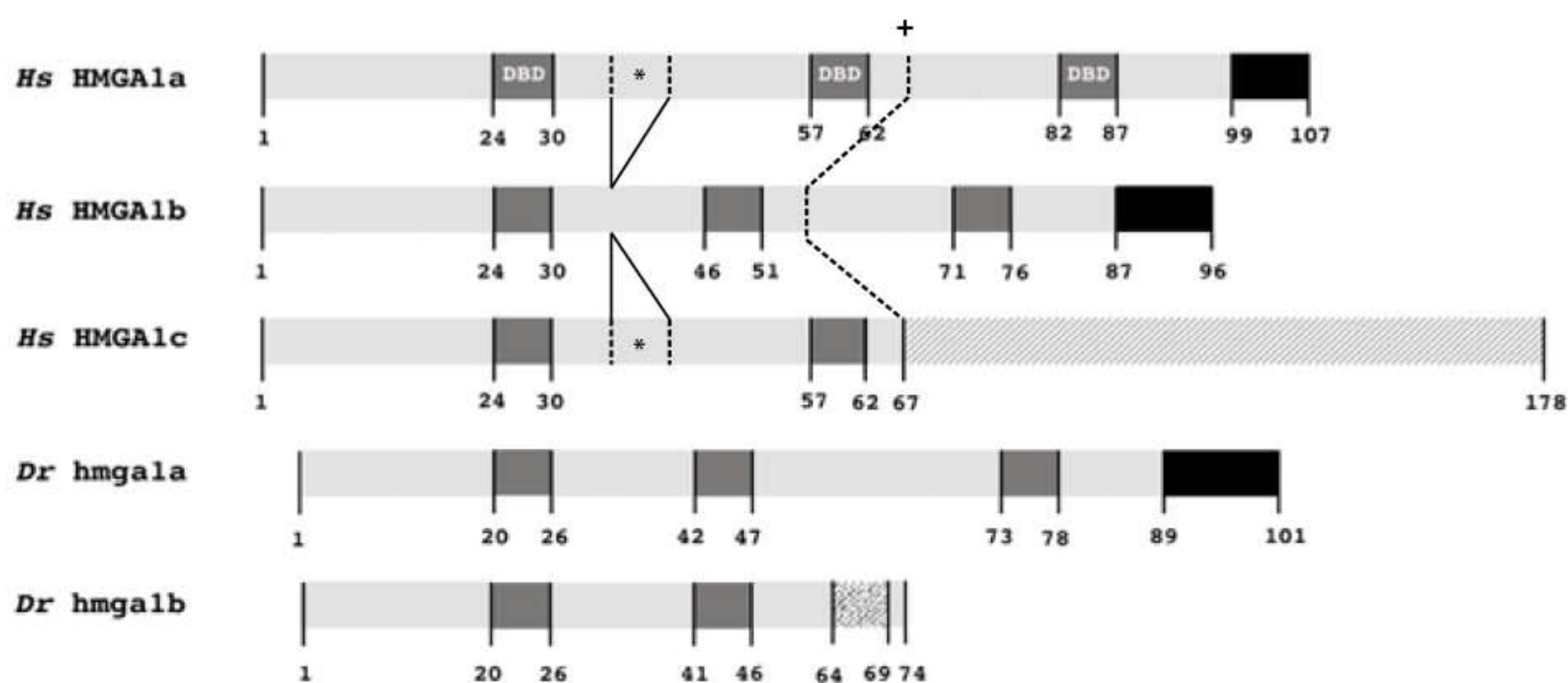## Relative expression levels of *hmga1b* mRNA at different developmental stages
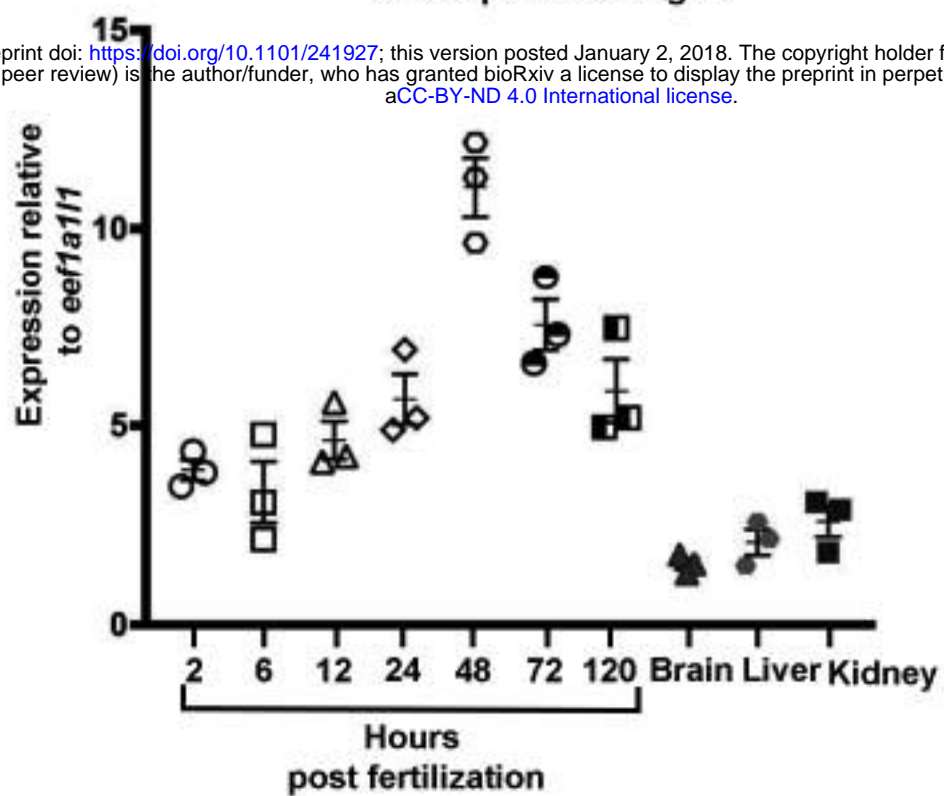
Figure 2.

A

```
Hs HMGA1a   MSESSSKSSQPLASKQEKDGTEKRGRGRPRKQPPVSPGTALVGSQKEPSEVPTPKRPRGR
Hs HMGA1b   MSESSSKSSQPLASKQEKDGTEKRGRGRPRKQPP-----------KEPSEVPTPKRPRGR
Dr hmga1a   MSDSGKDT----VATKEKDGAEKRGRGRPRKHPQ-----------QEASGSPTPKRPRGR
Dr hmga1b   MSDSEKQT----VSLKDKDGVEKRGRGRPRKHP------------KESSGSPSAKKPRGR
```

                          +
```
Hs HMGA1a   PKGSKNKGA---AKTRK---TTTTPGRKPRGRPKKLEKEEEEGISQESS---EEEQ
Hs HMGA1b   PKGSKNKGA---AKTRK---TTTTPGRKPRGRPKKLEKEEEEGISQESS--EEEQ
Dr hmga1a   PKGSKNKPSSTVSRGKKAAAASAPAGTKRRGRPKKAEKEEQPSKSSEEEEEEEEEQ
Dr hmga1b   PKGSKNKGP---SKRK-----SSTSGSKAKGKPKK--------------EQQ
```

B

C