PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Enhancer turnover and conserved regulatory function in vertebrate evolution

Sabina Domené, Viviana F. Bumaschny, Flávio S. J. de Souza, Lucía F. Franchini, Sofía Nasif, Malcolm J. Low and Marcelo Rubinstein

Phil. Trans. R. Soc. B 2013 **368**, 20130027, published 11 November 2013

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Research

Cite this article: Domené S, Bumaschny VF, de Souza FSJ, Franchini LF, Nasif S, Low MJ, Rubinstein M. 2013 Enhancer turnover and conserved regulatory function in vertebrate evolution. Phil Trans R Soc B 368: 20130027. http://dx.doi.org/10.1098/rstb.2013.0027

One contribution of 12 to a Theme Issue 'Molecular and functional evolution of transcriptional enhancers in animals'.

Subject Areas:
bioinformatics, developmental biology, evolution, genetics, genomics, neuroscience

Keywords:
transgenic zebrafish, proopiomelanocortin, hypothalamus, mammals, teleosts, comparative genomics

Author for correspondence:
Marcelo Rubinstein
e-mail: mrubins@dna.uba.ar

†These authors contributed equally to this study.

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2013.0027 or via http://rstb.royalsocietypublishing.org.

Royal Society Publishing
Informing the science of the future

Enhancer turnover and conserved regulatory function in vertebrate evolution

Sabina Domené[1,†], Viviana F. Bumaschny[1,2,†], Flávio S. J. de Souza[1,3], Lucía F. Franchini[1], Sofía Nasif[1], Malcolm J. Low[4] and Marcelo Rubinstein[1,3]

[1]Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, Consejo Nacional de Investigaciones Científicas y Técnicas, C1428ADN Buenos Aires, Argentina
[2]Departamento de Fisiología y Biofísica, Facultad de Medicina, Universidad de Buenos Aires, C1121ABG Buenos Aires, Argentina
[3]Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA Buenos Aires, Argentina
[4]Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, MI 48105, USA

Mutations in regulatory regions including enhancers are an important source of variation and innovation during evolution. Enhancers can evolve by changes in the sequence, arrangement and repertoire of transcription factor binding sites, but whole enhancers can also be lost or gained in certain lineages in a process of turnover. The proopiomelanocortin gene (*Pomc*), which encodes a prohormone, is expressed in the pituitary and hypothalamus of all jawed vertebrates. We have previously described that hypothalamic *Pomc* expression in mammals is controlled by two enhancers—nPE1 and nPE2—that are derived from transposable elements and that presumably replaced the ancestral neuronal *Pomc* regulatory regions. Here, we show that nPE1 and nPE2, even though they are mammalian novelties with no homologous counterpart in other vertebrates, nevertheless can drive gene expression specifically to POMC neurons in the hypothalamus of larval and adult transgenic zebrafish. This indicates that when neuronal *Pomc* enhancers originated de novo during early mammalian evolution, the newly created *cis*- and *trans*-codes were similar to the ancestral ones. We also identify the neuronal regulatory region of zebrafish *pomca* and confirm that it is not homologous to the mammalian enhancers. Our work sheds light on the process of gene regulatory evolution by showing how a locus can undergo enhancer turnover and nevertheless maintain the ancestral transcriptional output.

1. Introduction

The understanding of organismal evolution at the molecular level is closely related to the study of gene regulation, as mutations in regulatory regions in the DNA are considered one of the main drivers of biological evolution [1]. Transcriptional enhancers are *cis*-acting DNA elements that work as arrays of transcription factor binding sites (TFBSs) and control the transcriptional activity of genes located nearby. The identification and study of enhancers is a difficult task, as there are no general, fixed rules regarding the location or sequence composition of such elements. Compared with coding regions, which are generally highly constrained by protein structure and function, enhancers are freer to evolve, because TFBSs are small (6–10 nucleotides long) and accept much higher levels of sequence divergence than the genetic code, i.e. each transcription factor can recognize a variety of binding site sequences. In addition, although some examples of rigid enhancer organization exist [2,3], the order of TFBSs on a given enhancer can be quite flexible [4]. Finally, enhancers act at a distance and can even be embedded within neighbouring loci [5–9], indicating a large degree of positional freedom of enhancers in relation to the genes they control. Thus, while new protein coding genes mostly appear by

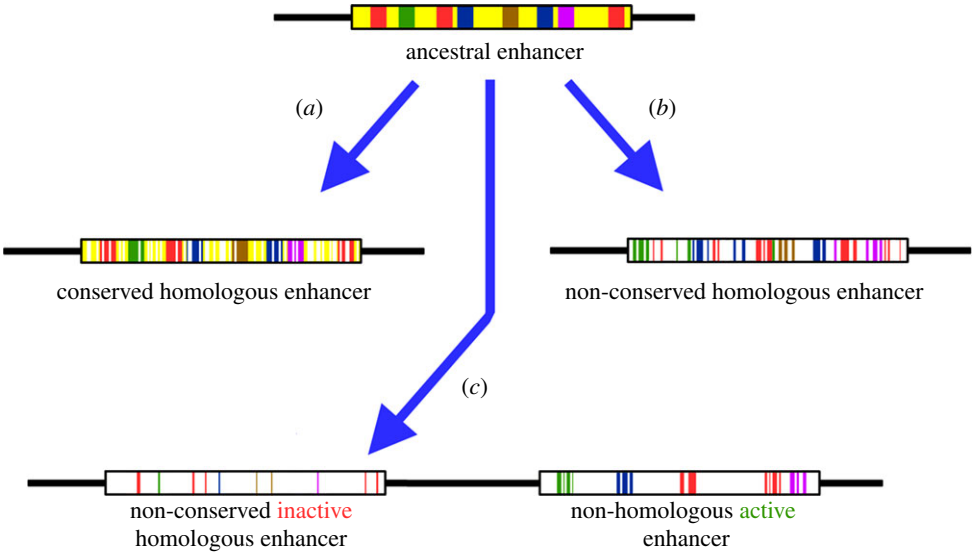

Figure 1. Scheme of possible evolutionary paths of enhancers. Top: ancestral enhancer (box) has a given arrangement of transcription factor binding sites (TFBSs, coloured squares) in a background of near-neutral sequence (yellow). (*a*) As evolutionary time goes by, the enhancer can accumulate relatively few mutations (indicated as white areas in box) and give rise to homologous enhancers that can be aligned (i.e. are 'conserved') between species. (*b*) If TFBSs undergo much turnover and the accumulation of mutations is too high, homologous enhancers may no longer be alignable (i.e. be 'non-conserved'). (*c*) The ancestral enhancer becomes inactive by the accumulation of mutations, but this is compensated by the appearance of a new enhancer elsewhere in the locus that fulfils the same regulatory function. In the path to (*c*), both functional enhancers coexist for some time until one of them becomes non-functional.

duplication of pre-existing genes [10], DNA sequences within an enhancer can change quickly over evolutionary time and new enhancers can appear in new locations and replace ancient ones.

(a) Enhancer evolution in vertebrates

Traditionally, enhancers were found by the patient testing of different DNA regions coming from the locus under study in reporter gene assays in cell culture and in transgenic mice. With the advent of whole-genome sequencing, it was quickly recognized that enhancers and other regulatory regions might be identified by the sequence alignments of genomes of different organisms, the rationale being that non-coding, functional DNA regions would show signs of purifying selection (i.e. constraint) and have a higher level of sequence conservation than non-functional, neutrally evolving regions [11,12]. Thus, early alignments of human and mouse genomes allowed for the identification of candidate regulatory regions and, as more genomes were sequenced, comparisons of mammalian genomes with those of teleost fishes such as fugu (*Takifugu rubripes*) and zebrafish (*Danio rerio*) allowed for the identification of thousands of conserved non-coding elements (CNEs) that have maintained their function throughout vertebrate evolution [13–15]. In fact, many of these highly conserved CNEs behave as enhancers when tested in transgenic mice and/or zebrafish [5,14–16], indicating that enhancer function can be conserved during the 450 million years that have passed since the divergence of the teleost and tetrapod lineages [17].

But, can all enhancers in a given genome be simply identified by phylogenetic footprinting? Because enhancers can evolve relatively fast, the answer to this question is not easy, as many circumstances can be envisaged that would lead to failure in the identification of enhancers by sequence conservation between two given lineages (figure 1). Thus, (i) orthologous enhancers (i.e. originated from a common ancestral sequence) can mutate redundant nucleotides within TFBSs and/or reshuffle their TFBSs during evolution (TFBS turnover), but may still be alignable, something that facilitates their identification, as in the case of CNEs. However, (ii) it is also possible that extensive mutations can occur that maintain enhancer function but erase alignability, preventing a straightforward identification by sequence conservation. Alternatively, (iii) current enhancers performing the same regulatory function in two different lineages may be non-homologous (analogous), being rather the result of the de novo appearance of a new enhancer in a different location of the locus in one or both of the lineages, and the ancestral enhancer may be lost in at least one of them (enhancer turnover). Of course, it is also possible that enhancers are gained or lost independently of the fate of the ancestral enhancers, again leading to lineage-specific differences in enhancer repertoire. As for the regulatory consequences of TFBS and enhancer turnover, three non-mutually exclusive alternatives exist: (i) either reshuffling and turnover may happen without altering the ancestral transcriptional activity of the controlled gene, i.e. 'things change to stay the same'; (ii) enhancer turnover occurs modifying quantitative or qualitative traits of the spatio-temporal gene expression domain that may produce phenotypic changes, i.e. longer limbs, darker colour, etc.; or (iii) enhancer turnover originates completely novel patterns of expression in different lineages, leading to regulatory innovation with reflections in the morphology and physiology of organisms.

Hence, given enough time, it is expected that TFBS and enhancer reshuffling and turnover will erase sequence similarity in different lineages, and the power to detect functional sequence by comparative approaches will depend on the distance of the species analysed, meaning that as the phylogenetic distance increases, the proportion of CNEs should decrease. For instance, Prabhakar *et al*. [18] identified 170 000 CNEs in a human–rodent comparison, but this number dropped to 40 000 in human–chicken and 5700 in human–fugu comparisons. When tested in transgenic mouse embryos, 27% of human–rodent CNEs (mammalian-specific) were

found to have enhancer activity [18], a percentage that is similar to that obtained with human–fugu CNE (29% [16]). Mammalian-specific and vertebrate-specific enhancers seem to be distributed non-homogeneously among genes, as enhancers conserved between mammals and fish are more common near genes involved in embryonic development, suggesting that genes with complex spatio-temporal expression patterns are less flexible in their enhancer organization and architecture compared with genes with other functions [19,20]. In conclusion, the proportion of enhancers in the human genome that will be identifiable by phylogenetic footprinting is still unclear. The recent sequencing of 29 mammalian genomes should be an invaluable tool to help identify enhancers common to mammals or, at least, most mammalian groups [21]. In the case of enhancers with more restricted distribution, such as primate- or even human-specific enhancers, their identification might only be possible by conventional functional assays or aided by epigenetic marks typical for enhancers in the locus under study [22,23]. Concerning the latter strategy, many investigations in recent years have used genome-wide chromatin immunoprecipitation to map epigenetic marks that are typical for enhancers and, in some cases, to test the elements in transgenic mouse analysis (reviewed in [24,25]). Such studies have detected a large proportion of enhancers that are conserved among several mammals but also a significant fraction of enhancers without apparent evolutionary constraint [22,26,27]. It remains to be tested, however, the relative importance of conserved versus weakly- or non-conserved enhancers in the regulation of the genes they control. Although there are well-characterized examples of *cis*-acting regulatory regions restricted to particular mammalian lineages (e.g. [28,29]), several disease-associated mutations mapping to non-coding regions in humans have turned out to reside within phylogenetically conserved enhancers [6,30–36].

(b) Zebrafish as a tool to investigate enhancer evolution

When it comes to studying human enhancers, researchers have traditionally studied reporter gene expression in transgenic mice as a gold standard. Being mammals, mice are suitable models to test the activity of human enhancers, with the invaluable advantage over cell culture systems that enhancer activity can be analysed in a spatio-temporal way and simultaneously in all tissues and cell types. There are, however, other attractive vertebrate models that can be used to study enhancers by transgenesis, including the teleost zebrafish, which allows for quicker and cheaper transgenic assays [37]. Mammals and zebrafish share thousands of CNEs that can work as enhancers, and zebrafish transgenesis plays an important role in the analysis of such conserved genetic elements [14,15,38–41].

As discussed earlier, the few thousand enhancers that are conserved between mammals and teleosts should constitute only a fraction of the total [18], as TFBS and enhancer turnover tend to erase sequence similarity during evolution. Recent high-throughput studies participating in the mouse ENCODE project detected 230 000 elements with enhancer-type signatures based on the genomic distribution of certain epigenetic marks in murine cells [42]. Can zebrafish transgenesis be used to study mammalian enhancers that are not conserved in teleosts? The answer to this question is yes, as it has been found in some experiments that mammalian enhancers without obvious teleost orthologues can nevertheless drive specific transgene expression in zebrafish embryos [39,43,44], indicating that at least part of the underlying *cis*- (TFBS composition or order) as well as *trans*- (transcription factors) elements are conserved in mammals and teleosts in many diverged enhancers. It is likely that many of these non-conserved enhancers do possess orthologues that cannot be readily identified owing to the turnover of neutral and redundant nucleotide sites within the enhancers in each lineage (TFBS turnover). Indeed, using the genome of the frog *Xenopus tropicalis*, an evolutionary intermediate reference between mammals and teleosts, 1500 CNEs derived from human and zebrafish genomes have been found that do not align to each other but do align with elements in the frog [45]. When analysed in detail, such elements show evidence of TFBS conservation that cannot be detected in whole-genome alignments and, when tested in transgenic zebrafish, many elements display enhancer activity, indicating that the lack of mammal–fish identity of these orthologous enhancers was caused by the accumulation of mutations within the ancestral sequences during the evolution of each lineage [45]. Apart from turnover of TFBS sequence within enhancers, evolution can also lead to the appearance of non-orthologous novel enhancers and, eventually, to enhancer turnover with complete functional loss of the ancestral sequence (figure 1). In practice, however, it is difficult to be certain whether the lack of overt sequence identity of such non-conserved mammal–fish enhancers is the result of TFBS sequence turnover in orthologous enhancers or due to an enhancer appearing de novo, unless a detailed knowledge of the evolutionary history of the enhancer in at least one of the two lineages is found [46,47]. Here, we show that two mouse enhancers of the proopiomelanocortin gene (*Pomc*) that certainly appeared in the mammalian lineage direct appropriate reporter gene expression to homologous neurons in the ventromedial hypothalamus of transgenic zebrafish.

(c) Evolution of *Pomc* transcriptional expression

The *Pomc* gene encodes a prohormone found in all vertebrates that is processed to several bioactive peptides to orchestrate the stress response. *Pomc* is expressed in the pituitary gland and brain, mainly in the hypothalamus, and its derived peptides are involved in processes as diverse as glucocorticoid release, stress-induced analgesia, energy balance control and, in fishes and amphibians, background colour adaptation [48–50]. Mammals, birds and other tetrapods have only one *Pomc* copy, but zebrafish and other teleosts, including fugu and *Tetraodon nigroviridis*, have two *Pomc* paralogues—*pomca* and *pomcb*—that date from the ancient whole-genome duplication in the teleost lineage around 320 Ma [51]. In zebrafish, *pomca* is expressed in the anterior and intermediate lobes of the pituitary as well as in the ventral hypothalamus, just like the tetrapod gene [52,53]. The expression pattern of zebrafish *pomcb* is unknown, but the *Tetraodon* homologue is expressed in the preoptic area of the brain and pituitary [51].

Pituitary and hypothalamic *Pomc* expression are controlled by different regulatory regions. In the mammalian pituitary, there is a large body of evidence that *Pomc* expression is regulated by several TFBSs present in the proximal promoter [54–56], as well as by a mammalian-conserved distal enhancer [57]. In the zebrafish, similar to mammals, it has been found that a construct carrying the proximal

promoter and first intron of the *pomca* gene can drive green fluorescent protein (GFP) expression to the anterior and intermediate lobes of the pituitary, but not to ventral hypothalamic neurons [52]. A similar construct carrying 500 bp of the *Xenopus laevis Pomc* gene promoter directs GFP expression to cells of the intermediate lobe, but not the hypothalamus, of transgenic frog embryos [58], suggesting that the function of the proximal *Pomc* promoter in driving pituitary expression is ancestral in vertebrates.

In the mouse hypothalamus, *Pomc* expression depends on a distal module that we identified in the pre-genomic era following a tedious chromosome walking and deletional analysis in transgenic mice [59]. Upon release of the first human genome draft, we were able to perform a human–mouse phylogenetic footprinting that revealed the presence of two highly conserved sequences, later also found in other mammalian genomes, which proved to drive reporter gene expression to POMC hypothalamic neurons [60]. These *n*euronal *P*omc *E*nhancers, that we named nPE1 and nPE2, are located at −12 and −10 kb relative to the mouse transcription start site (TSS) [60]. While nPE2 is found in monotremes (egg-laying mammals) and marsupials, as well as placental mammals, nPE1 is only found in placentals [46,47]. No sequences similar to either nPE enhancer can be identified in non-mammalian genomes. A detailed phylogenetic analysis of both enhancers revealed that nPE1 and nPE2 are derived from unrelated ancient transposable elements (TEs) that were independently co-opted—or exapted—at different times during early mammalian evolution [46,47,61]. nPE2 was derived from a CORE-SINE retroposon before the mammalian radiation [46], whereas nPE1 is derived from an MaLR retroposon inserted at the time of the placental mammal diversification [47]. Transgenic mouse studies showed that either nPE1 or nPE2 is able to drive authentic reporter gene expression to POMC hypothalamic neurons, whereas the simultaneous absence of both enhancers from a 4 kb mouse neuronal distal module completely inactivates this function [60]. More recently, we reported a first case of convergent evolution of cell-specific transcriptional enhancers by demonstrating that nPE1 and nPE2 control a fully overlapping spatio-temporal expression domain in the ventral hypothalamus [47,61].

While nPE1 and nPE2 explain neuronal *Pomc* expression in mammals, nothing is known about the regulatory regions controlling hypothalamic *Pomc* expression in all other vertebrate classes. Because nPE1 and nPE2 are derived from TEs exapted in the mammalian lineage, they should be non-orthologous to *Pomc* regulatory regions in other vertebrates (figure 1). With this in mind, here we attempt to investigate the following aspects of *Pomc* regulatory evolution: (i) can the mammalian enhancers be recognized by the *trans*-acting machinery of the teleost hypothalamic neurons? (ii) What is/are the enhancer/s controlling *Pomc* expression in the zebrafish hypothalamus? (iii) Are the zebrafish neuronal *Pomc* enhancer/s conserved in genomes from other teleost suborders?

2. Material and methods

(a) Transgenic zebrafish

Adult stocks of strain AB wild-type zebrafish (*D. rerio*) obtained from the Zebrafish International Resource Center (University of Oregon, USA) were maintained at 28.5°C on a 14 L : 10 D cycle. All embryos were collected by natural spawning and were raised at 28°C in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.3 mM $MgSO_4$ and 0.1% methylene blue) in Petri dishes. Transposase mRNA was synthesized *in vitro* using mMESSAGE mMACHINE Sp6 kit (Ambion, Inc.) from pCS-TP. One nanolitre of a DNA/RNA solution containing 25 ng μl^{-1} of circular DNA (transgene vectors) and 35 ng μl^{-1} of transposase mRNA were co-injected into one cell fertilized egg. Quantification and selection of positive transgenic zebrafish was carried out in tricaine-anaesthetized embryos and viewed in an Olympus BX41 fluorescent microscope coupled to an Olympus DP71 digital camera. Patterns of neuronal enhancer activity were confirmed by transient injections (300–1000 embryos injected for each construct in at least three different days). Five-day-old transgenic embryos were screened for enhanced green fluorescent protein (EGFP) expression in the pituitary (used as a positive control for the transgene activity) and in the hypothalamus. Around 80–90% of injected embryos express EGFP in the pituitary. Only those embryos expressing EGFP in both the pituitary and hypothalamus were considered to be positive for neuronal enhancer activity. All procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals, United States Public Health Services.

(b) DNA constructs

The plasmid pT2KXIG containing *Tol2* sequences for transposase recognition (gift from Koichi Kawakami) was modified to replace the pEF1α promoter and the intron of rabbit β-globin (cut by *Xho*I and *Bam*HI) with a 991 bp (−562/+428) fragment, including the zebrafish *pomca* promoter, the entire first exon and intron, and 22 bp of the second exon, to yield transgene 1 (Tg1). An approximately 4 kb DNA fragment containing mouse neuronal enhancer elements nPE1 and nPE2 located at −13 and −9 from mouse *Pomc* was subcloned into Tg1 upstream of the zebrafish promoter using *Xho*I and an introduced *Xma*I site to yield Tg2. Additionally, the following constructs with different deletions (Δ) of *Pomc* neuronal enhancers were subcloned into Tg1 using *Xho*I, and then characterized for correct orientation: ΔnPE1 (Tg3), ΔnPE2 (Tg4) and ΔnPE1–ΔnPE2 (Tg5), all in the −13/−9 mouse *Pomc* context (see details in de Souza *et al*. [60]). All fragments were first PCR amplified with Pfx DNA polymerase (Invitrogen), subcloned into pZErO-2 (Invitrogen) and sequenced before subcloning into Tg1 vector.

Several zebrafish fragments located upstream of *pomca* were amplified using a bacterial artificial chromosome (BAC) clone (RP71–36D5) as template. First, two consecutive fragments of 4.2 (−10.6/−6.4) and 5.8 kb (−6.4/−0.56) were amplified by PCR (*TaKaRa Ex Taq*), linked together at a natural *Bam*HI site located at −6.4, and then subcloned into Tg1 in artificially added *Xho*I and *Xma*I sites (Tg6). Additionally, three fragments encompassed in the −10.6/−0.56 zebrafish region included in Tg6 were PCR amplified with Pfx DNA polymerase (Invitrogen), subcloned into pZErO-2 and finally subcloned with *Xho*I upstream of the *pomca* promoter in Tg1 to generate Tg7 (−3/−0.6), Tg8 (−6.9/−4) and Tg9 (−10.6/−6.9). Further dissection of the 3.6 kb Tg9 was performed by restriction digestion using either *Mlu*I and *Eco*RV (eliminating a 1273 bp fragment), followed by fill-in reaction and blunt ligation, to analyse the 5′ region of 2327 bp (Tg10, −10.6/−8.2), or digested with *Apa*I and *Eco*RV (eliminating a 2327 bp fragment), followed by fill-in reaction and blunt ligation to analyse the 3′ 1273 bp fragment (Tg11, −8.2/−6.9). Three sites of Tg11 (indicated in figure 4*d*) were mutated from TAAT to TGGT by PCR using overlapping megaprimers bearing the mutations. The resulting mutated 1273 bp fragment was subcloned into pZErO-2, sequenced and then subcloned into Tg1 using *Xho*I sites present in the primers to yield Tg12. All plasmid

constructs were verified by restriction digest mapping and direct DNA sequencing.

(c) Whole-mount mRNA *in situ* hybridization

Whole-mount *in situ* hybridization was performed essentially as described by Thisse *et al*. [62] with the following modifications: larvae were grown in a buffer containing 0.2 mM 1-phenyl-2-thiourea from 24 h post-fertilization onwards to remove pigmentation. After fixation in 4% paraformaldehyde, 5 days-post fertilization (dpf) larvae were permeabilized in proteinase K (10 µg ml^{-1}) for 30 min. Riboprobe was synthesized with DIG RNA labelling kit (Roche) according to manufacturer's instructions using a zebrafish *pomca* template plasmid which included a 623 bp sequence from exon 3 cloned into pGEM-T easy (Promega). The template plasmid was digested with *Apa*I and transcribed with Sp6 to generate digoxigenin-labelled antisense riboprobe. Pre-hybridization and hybridization were performed at 65°C. After stringent washing, incubation in blocking solution (1× PBS/Tween, 2% sheep serum, 2 mg ml^{-1} bovine serum albumin) was performed for 1 h at room temperature followed by incubation with alkaline phosphatase-coupled anti-digoxigenin antibody (Roche) by shaking at 1 : 2000 for 2 h at room temperature. The reaction was developed with NBT/BCIP (Roche) in the dark. Larvae were cleared in 70% glycerol and mounted for photography on an Olympus BX41 fluorescent microscope coupled to an Olympus DP71 digital camera.

(d) Immunohistochemistry on cryosections

Five dpf larvae or adult zebrafish brains were fixed for 2 h in 4% paraformaldehyde in PBS at 4°C. They were rinsed three times in PBS, cryoprotected in 10% sucrose in PBS, immersed in 10% gelatin/10% sucrose/PBS, fast frozen in isopentane and stored at −80°C until use. Cryosections (16 µm) were incubated overnight at 4°C with rabbit polyclonal anti-human ACTH–IC-1 (1 : 700; National Hormone and Pituitary Program, National Institutes of Health) and mouse monoclonal anti-EGFP (1 : 700; MAB3580, Millipore) in KPBS (0.9% NaCl, 16 mM K_2HPO_4, 3.6 mM KH_2PO_4)/0.3% Triton X-100, 2% normal goat serum. Following several washes with KPBS, primary antibodies were detected using Alexa-555 anti-rabbit immunoglobulins and Alexa-488 anti-mouse immunoglobulins conjugated secondary antibodies (Invitrogen), diluted 1 : 1000 in KPBS/0.3% Triton X-100. Images were taken either with an Olympus BX41 fluorescent microscope or, for confocal images, with a Nikon Eclipse E800.

(e) Bioinformatic analysis

Comparative analysis of mouse nPE2 and zebrafish 1.2 zfnPE sequences to detect TFBSs was performed using the matrix-based program MatInspector (www.genomatix.de), which uses a large and updated library of TFBS weight matrices [63]. Multiple sequences were compared using the PipMaker program (http://bio.cse.psu.edu/pipmaker).

3. Results: functional conservation of neuronal *Pomc* enhancers in mammals and zebrafish

(a) Mammalian neuronal enhancers drive reporter gene expression to zebrafish POMC hypothalamic neurons

The mammalian *Pomc* gene is expressed in the pituitary and the arcuate nucleus of the ventral hypothalamus [60]. In the zebrafish and other teleosts, apart from the pituitary, the *pomca* gene is expressed in the *nucleus lateralis tuberis* (NLT) [51,64], considered a homologous structure to the mammalian arcuate nucleus [65]. Hence, expression of *Pomc* in the ventral hypothalamus is a primitive feature of the common ancestor of ray-finned fishes and tetrapods. However, enhancers nPE1 and nPE2, which control hypothalamic *Pomc* expression in the mammalian brain, are mammalian novelties, as they resulted from the exaptation of TEs at the base of the mammalian radiation [61]. Consistent with this, we used the PipMaker program [66] to perform multiple local alignments between the mouse and human *Pomc* loci and the region encompassing approximately 40 kb around the *pomca* locus in zebrafish chromosome 17 (contig AL590149.9, Zv9), but no non-coding regions could be aligned between these species (data not shown).

To check whether the mammalian enhancers, in spite of their relatively recent evolutionary origin, could be recognized by the transcriptional machinery of zebrafish hypothalamic POMC neurons, we created transgenic zebrafish carrying the mouse distal neuronal *Pomc* enhancer module using *Tol2*-transposon-based constructs. A transgene series was built using an approximately 1 kb region of the zebrafish *pomca* gene that encompasses 451 bp of the upstream promoter, exon 1 (85 bp), intron 1 and 22 bp of exon 2 that were fused to EGFP coding sequences (figure 2*b*). This region has been shown by Liu *et al*. [52] to drive robust EGFP expression to the distal and intermediate lobes of transgenic zebrafish, but not to the hypothalamus, a result that we reproduced here (figure 2*b*, Tg1). Next, we inserted the mouse distal neuronal *Pomc* enhancer module corresponding to −13 and −9 kb from the mouse *Pomc* TSS upstream of the zebrafish promoter construct (figure 2*b*, Tg2). Interestingly, constructs carrying this module drove EGFP expression to the zebrafish embryonic hypothalamus at 5 dpf (figure 2*b*, Tg2). In whole-mount embryos, positive hypothalamic neurons are seen as a small group of cells located anteriorly to the pituitary. This latter tissue also expresses EGFP, because the construct carries the zebrafish proximal promoter (figure 2*b*, Tg2). In total, of 717 embryos displaying EGFP expression in the pituitary, 179 (25%) also showed EGFP cells in the ventromedial hypothalamic area, with very little, if any, ectopic expression in the brain, retina or other body areas. To confirm the identity of EGFP-expressing neurons, we checked for co-expression of EGFP and POMC by immunofluorescence in sections of 5 dpf-transgenic embryos using an antibody against the human adrenocorticotropic hormone (ACTH) peptide, which is very similar to POMCa-derived zebrafish ACTH [51,67] and efficiently labels zebrafish POMC neurons. Approximately 30 ACTH-positive neurons arranged as two arcs around the midline are detected per embryo and, of these, approximately 50% are also EGFP-positive. We have also detected transgenic expression in a few neurons located within the array of ACTH-immunoreactive (ir) neurons but that were not labelled with the ACTH antibody. It remains to be determined whether these neurons show ectopic transgenic expression or the human ACTH antibody fails to label the entire population of zebrafish POMCa-expressing neurons.

To check whether the 4-kb mouse regulatory region is also functional in the adult fish brain, we grew and mated Tg2 F0 transgenic zebrafish and created F1 transgenic lines. From one of these lines, we collected many 5 dpf F2 embryos and observed that all of them expressed EGFP in the pituitary and in approximately 50% of ACTH-ir neurons of the hypothalamus (figure 3*a*). EGFP-positive neurons were also observed in the NLT of the ventral hypothalamus of three-month-old F2 transgenic adults in which, again, approximately 50% of human ACTH-ir neurons coexpressed EGFP (figure 3*b*).

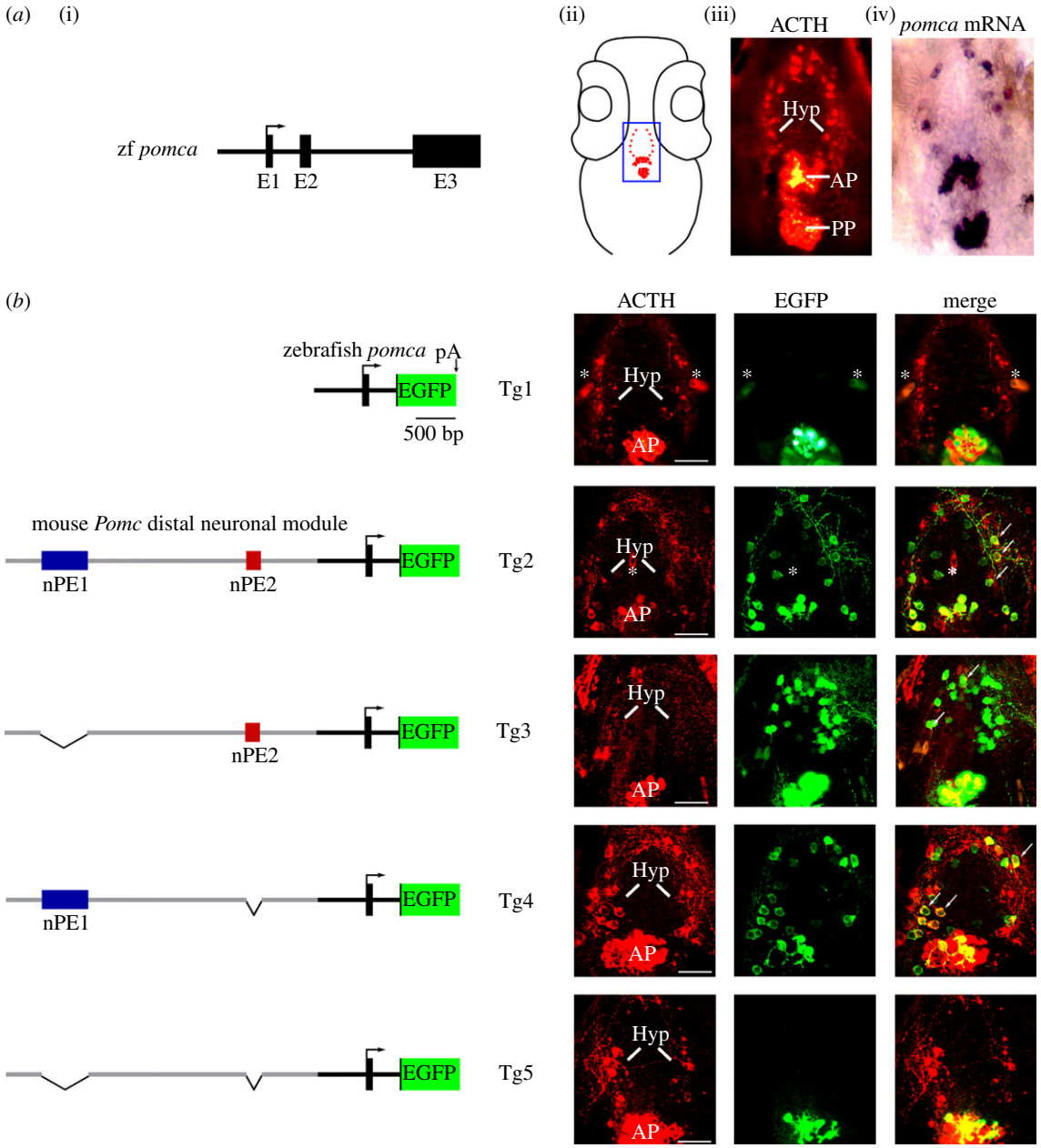

Figure 2. Mouse neuronal *Pomc* enhancers drive EGFP expression to POMC hypothalamic neurons in transgenic zebrafish. (*a*) (i) Zebrafish *pomca* gene. Black boxes represent exons. (ii) Drawing of a 5 dpf zebrafish head in a horizontal plane showing the distribution of *pomca*-expressing cells in the pituitary and hypothalamus (red dots). The blue rectangle depicts the representative area shown in the other panels. (iii) ACTH-ir cells on a horizontal section of a 5 dpf zebrafish brain. Hyp, NLT of the hypothalamus; AP and PP, anterior and posterior pituitary lobes, respectively. (iv) Horizontal view of a whole-mount *in situ* hybridization for *pomca*. (*b*) Expression analysis of transgenes driving EGFP under the control of zebrafish proximal *pomca* sequences (−0.6 to +0.4 kb; in black) and a neuronal mouse *Pomc* distal module (grey line) carrying enhancers nPE1 (blue box) and nPE2 (red box) in transgenic zebrafish. Pictures show horizontal cryosections of 5 dpf zebrafish brains subjected to double immunohistochemistry against ACTH (left column, in red) and EGFP (central column, green). Right superimposed images show coexpression of POMC and EGFP (white arrows). Asterisk: non-specific signal. Transgenes are described in scale on the left, numbered in the centre (TgX) and expression analysis shown on the right.

Together, these results show that a mammalian *Pomc* regulatory region drives EGFP expression with similar efficiency in embryos and adult zebrafish POMC neurons.

The results above show that a distal neuronal regulatory region of mouse *Pomc* can be recognized by the zebrafish transcriptional machinery despite the fact that these two vertebrate lineages split 450 Ma. This 4 kb distal module carries the enhancer analogues nPE1 and nPE2, and previous studies showed that either nPE1 or nPE2 is able to drive reporter gene expression to POMC hypothalamic neurons in transgenic mice [60]. Thus, we asked whether expression in transgenic zebrafish depended on one or both mammalian enhancers. To this end, we tested deletion constructs in which nPE1 (ΔnPE1), nPE2 (ΔnPE2) or both enhancers simultaneously (ΔnPE1/ΔnPE2) were removed (figure 2*b*; Tg3–5). As previously found in transgenic mice [60], ΔnPE1 and ΔnPE2 constructs still drove EGFP expression to the zebrafish hypothalamus (figure 2*b*, Tg3 and Tg4), with 51/250 and 55/260 5 dpf embryos showing hypothalamic EGFP expression relative to transgenic signal in the pituitary, respectively. Immunofluorescence experiments showed similar levels of ACTH/EGFP colocalization for Tg3 and Tg4 to that of Tg2 (figure 2*b*). By contrast, simultaneous deletion of both enhancers led to a complete loss of EGFP expression

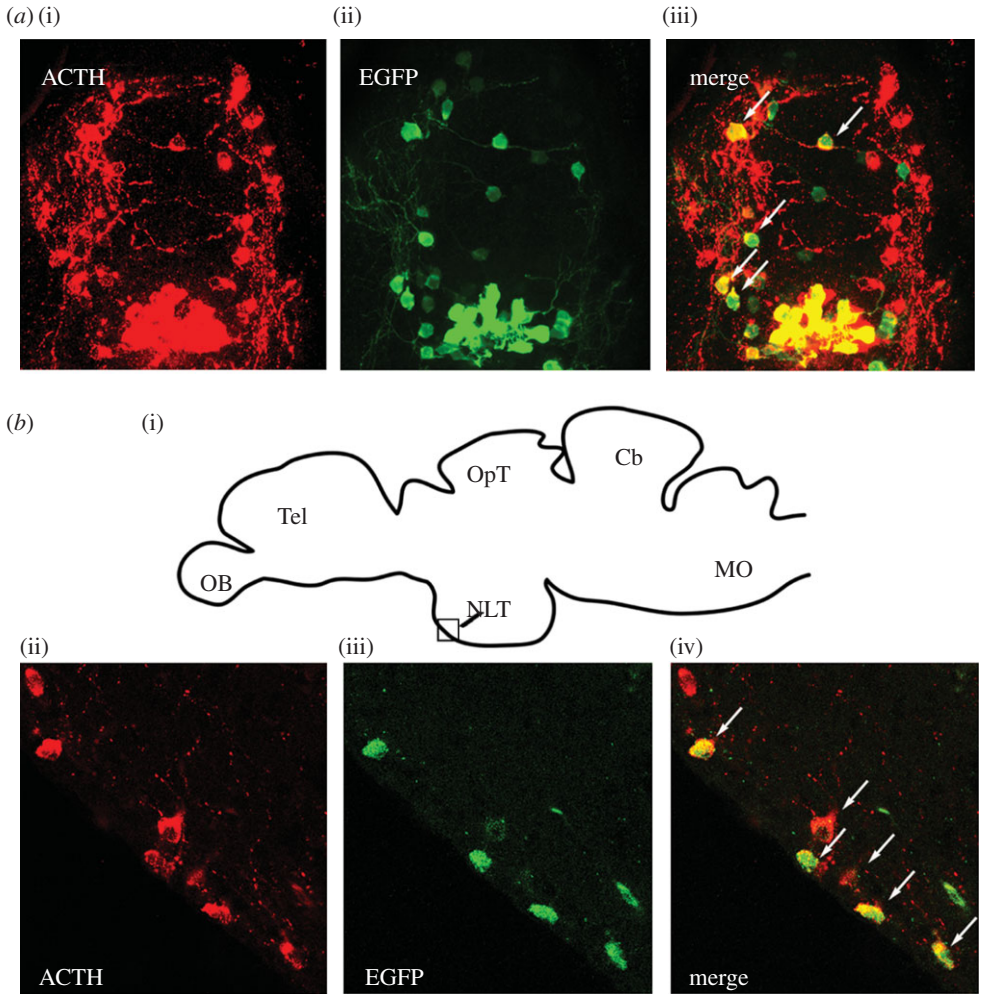

Figure 3. Functional conservation of the mouse neuronal *Pomc* enhancer module in a transgenic zebrafish line. (*a*) Horizontal sections of F2 zebrafish (5 dpf) carrying Tg2 (figure 2). Pictures show horizontal cryosections of a zebrafish brain subjected to double immunohistochemistry against (i) ACTH, (ii) EGFP and (iii) superimposed images. (*b*) (i) A diagram of a sagittal section of an adult brain in which the area shown in (ii–iv) is indicated by a square. OB, olfactory bulb; NLT, *nucleus lateralis tuberis* of the hypothalamus; Tel, telencephalon; OpT, optic tectum; Cb, cerebellum; MO, medulla oblongata. (ii–iv) Magnifications of a sagittal cryosection of the NLT area of an adult zebrafish of the transgenic line Tg2 subjected to double immunohistochemistry against (ii) ACTH, (iii) EGFP and (iv) superimposed images. White arrows indicate representative neurons with colocalization of the ACTH and EGFP signals.

in the zebrafish hypothalamus (figure 2*b*; Tg5). Thus, we conclude that the ability of the 4 kb mouse regulatory region to drive reporter gene expression to the zebrafish hypothalamus depends on the presence of either nPE1 or nPE2, which seem here to play overlapping regulatory roles, as observed previously in transgenic mice [47,60].

(b) Identification and characterization of a neuronal regulatory region of zebrafish *pomca*

Comparative studies of *Pomc* transcriptional regulation between mammals and fishes would greatly profit from the characterization of the zebrafish *cis*-elements that control hypothalamic *pomca* expression. Bioinformatic analyses failed to detect sequences similar to nPE1 or nPE2 in the zebrafish genome. Likewise, phylogenetic footprinting studies performed between the *pomca* zebrafish locus and sequences available from other teleost fish genomes failed to detect conserved non-coding regions. Therefore, in order to search for a putative zebrafish neuronal *Pomc* enhancer (zfnPE), we returned to a classical deletional analysis in transgenic zebrafish, similar to that we did 15 years ago to detect the mouse distal neuronal module [59] but, in this case, aided by an assembled zebrafish BAC genomic library. Using a combination of DNA primers and the BAC clone RP71–36D5, we assembled a transgene similar to Tg1 but with approximately 11 kb of 5′ flanking sequences of zebrafish DNA (figure 4*b*, Tg6). Tg6 led to the expression of EGFP in the ventromedial hypothalamus in 67 (17%) out of 394 transgenic embryos (5 dpf) showing EGFP-positive expression in the pituitary, and immunofluorescence analyses showed that hypothalamic EGFP was expressed in POMC neurons (figure 4*c*), indicating that a zebrafish neuronal regulatory module resides within the 11 kb genomic region. To narrow down this element further, we created a series of three deletions of the 11 kb upstream region (figure 4*b*; Tg7–Tg9) and observed that only the most distal, 3.6 kb segment, located between −10.6 and −6.9 kb of *pomca* exon 1, could drive EGFP expression to the hypothalamus (figure 4*c*; Tg9), with a hypothalamus/pituitary proportion of 57/398 (14%). By contrast, constructs Tg7 and Tg8, covering the region between −6.9 to −0.6 kb from *pomca* exon 1, were expressed only in the pituitary. We further divided the 3.6 kb region into a 5′ (−10.6 to −8.2 kb; Tg10) and a 3′ portion (−8.2 to −6.9 kb; Tg11) and produced transgenic zebrafish. While the 3′ part was as efficient as the original Tg6, showing a 77/391

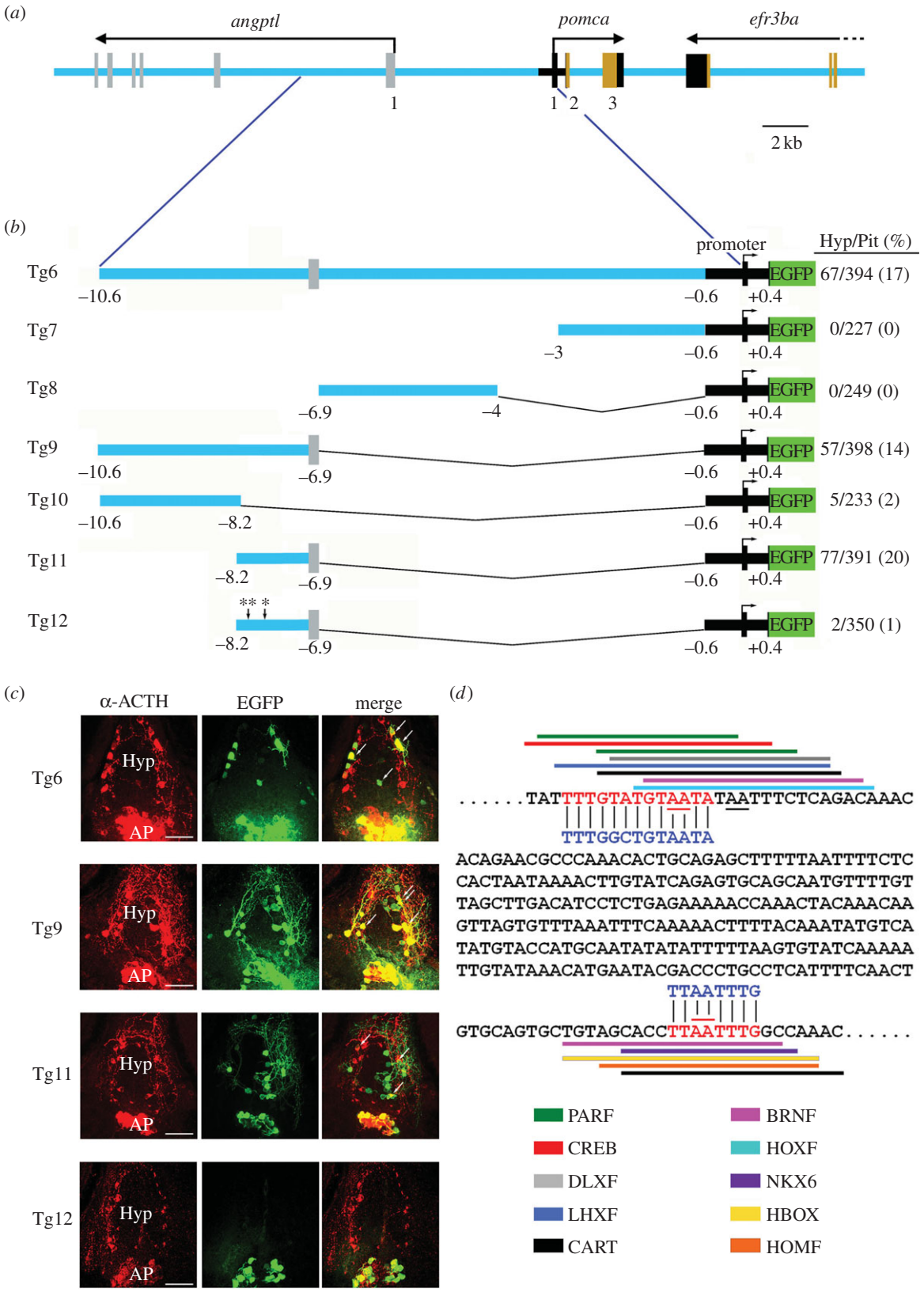

Figure 4. Identification and characterization of a zebrafish *pomca* neuronal enhancer. (*a*) Genomic structure of the zebrafish *pomca* locus. The 5′ flanking angiopoietin-like (*angpl*) and 3′ *efr3ba* genes are indicated. Exons are depicted in boxes. *Pomca* exons and *angpl* exon1 are numbered. (*b*) Set of transgenes (Tg) that carry variable lengths of 5′ flanking regions of zebrafish *pomca* fused to EGFP coding sequences used for the identification of a *pomca* neuronal zebrafish enhancer. Arrows in Tg12 point to the location of nucleotide substitutions indicated with asterisks. Number and percentage of positive EGFP embryos in the hypothalamus relative to embryos expressing EGFP in the pituitary (Hyp/Pit) are indicated on the right. (*c*) Horizontal brain cryosections of transgenic zebrafish (5 dpf) subjected to double immunohistochemistry against ACTH (left), EGFP (centre) or superimposed images that show coexpression (right, white arrows). Hyp, NLT of the hypothalamus; AP, anterior lobe of the pituitary. (*d*) Sequence of Tg11 mutated in Tg12. Homeobox conserved elements are shown in red (zebrafish sequence) or blue (orthologous sequence in mammalian nPE2). Mutated A to G nucleotides are underlined including a nearby TAAT site. Coloured lines indicate putative transcription factor families that could bind to the mutated sequences.

hypothalamic/pituitary ratio (20%), the 5′ part drove EGFP expression very inefficiently with only a 5/233 hypothalamic/pituitary ratio (2%), suggesting that a zfnPE resides in the 1.2 kb region present in Tg11 (figure 4*c*).

Because the zebrafish enhancer zfnPE and the mammalian enhancers nPE1 and nPE2 drive EGFP expression to the same array of hypothalamic POMC neurons in transgenic zebrafish, it is tempting to speculate that they share common DNA motifs recognized by the unique set of transcription factors present in these neurons. To begin pinpointing these regulatory elements, we made an exhaustive BLAST search of the 1.2 kb zebrafish region (chr17 : 33 424 669–33 425 868; Zv9/danRer7) to find DNA elements also present in mammalian enhancers nPE1 (586 bp) and nPE2 (180 bp). We found 10 short motifs (between 16 and 18 bp) of more than 85% identity between mouse nPE2 and the 1.2 kb zebrafish sequence containing zfnPE. We selected two of these motifs because they are embedded within a 50 bp fragment that was shown to be essential for nPE2 enhancer function in transgenic mice [46]. These two motifs contain TAAT sequences normally found in homeodomain TFBSs, that we substituted for mutant variants carrying TGGT (figure 4*c*, Tg12 and figure 4*d*). Strikingly, transgenic zebrafish embryos bearing Tg12 failed to express EGFP in the brain while retaining the ability to express the reporter gene in the pituitary (figure 4*c*), indicating that these two elements play a critical role in *pomca* neuronal expression within the 1.2 kb region. Because transcription factors controlling hypothalamic *Pomc* expression remain unknown, we searched for common TFBSs present in the mutated regions of zfnPE and in mammalian nPE2. We found 10 candidate transcription factor families (figure 4*d*) that could help identify transcription factors that drive neuronal-specific *Pomc* expression. A complete list of common TFBSs shared by mammalian nPE1 and nPE2 as well as by zebrafish zfnPE and a tetraodon upstream element (see below) is shown in the electronic supplementary material, figure S1.

The mutated TAAT sites present in zfnPE are included in a region annotated in the UCSC Genome Browser (www.genome.ucsc.edu) as a hAT Charlie DNA transposon (DNA-8–14_DR), a family of DNA transposons that is highly abundant in the *D. rerio* genome [68]. One peculiarity found in the recently published zebrafish genome is the unusually large proportion of sequences derived from DNA transposons relative to those originated from retrotransposons, in clear contrast to what has been found in most other vertebrate genomes, including the other four available Acanthopterygii fishes [68]. In this regard, our finding that the *pomca* locus of medaka, stickleback, tetraodon and fugu is devoid of sequences derived from a hAT Charlie strongly suggests that insertion of this DNA transposon upstream of zebrafish *pomca* is a relatively recent event. Future genome projects will help us to determine whether exaptation of zfnPE from a hAT Charlie DNA transposon occurred solely in zebrafish, in Cypriniformes or in the lineage leading to Ostariophysi.

In addition, the 1.2 kb zebrafish enhancer module maps within a predicted transcriptional unit of a gene that encodes for a protein similar to angiopoietin-like (*angpl*). This gene has a conserved position in relation to *Pomc* that dates back to the bony fish ancestor of teleosts and tetrapods, although it is not found in placental mammals [51]. The 1.2 kb module particularly contains the putative first exon and part of the first intron of the gene, and further studies will be necessary to establish whether zfnPE includes an active or extinct coding exon as has been reported for other vertebrate enhancers [69,70]. Although a paralogue of putative *angpl1* exon 1 is also found upstream of *pomcb*, the fragment carrying the two mutated TAAT sites of zfnPE seems to be absent from the *angpl1/pomcb* locus. Moreover, we failed to detect hAT Charlie derived-sequences in the *pomcb* locus, suggesting that insertion of this DNA transposon upstream of *pomca* occurred after the whole-genome duplication that took place in teleosts around 320 Ma [51].

(c) Identification and characterization of a neuronal regulatory region of tetraodon *pomca*

Similar to placental mammals, some members of the teleost superorder Acantomorpha—including tetraodon, fugu and medaka—also lack the *angpl* locus in the 5′ region of *pomca*, suggesting that these fishes might use a neuronal enhancer different from zfnPE to express *pomca* in the ventromedial hypothalamus. In fact, comparisons of the zebrafish *pomca* locus with those of other vertebrates did not yield any conserved non-coding element. However, when the 5′ flanking region of *pomca* from fugu was used in a phylogenetic footprinting analysis together with orthologous genomic sequences taken from other teleosts, we detected an 80 bp region that is conserved in representative species of the superorder Acanthomorpha, including tetraodon (*T. nigroviridis*), three-spined stickleback (*Gasterosteus aculeatus*) and medaka (*Oryzias latipes*) but not in zebrafish *pomca* or *pomcb*, a teleost that belongs to the superorder Ostariophysi (figure 5*a*). To check whether this conserved element might be an enhancer, we cloned a 426 bp fragment of the tetraodon *pomca* locus upstream of the zebrafish *pomca* promoter (figure 5*b*; Tg13), and the construct was tested in transgenic zebrafish. We observed that Tg13 is expressed in the ventromedial hypothalamus of zebrafish (25 embryos positive in the hypothalamus among 202 positive in the pituitary), and that a few cells colocalize with ACTH (figure 5*c*). This result indicates that the 426 bp sequence located in the 5′ flanking region of tetraodon *pomca* harbours a hypothalamic *pomca* enhancer. The presence of an 80 bp conserved element in the four Acanthomorpha species used in this study suggests that this enhancer could be active in all members of this superorder. In the same hypothalamic area where we observed coexpression of EGFP and POMC, we have also observed other EGFP-positive neurons that were not colabelled with the anti-human ACTH antibody. These hypothalamic neurons may represent another neuronal population to be identified.

4. Discussion

(a) Conservation of function by non-homologous enhancers in vertebrate evolution

Here, we show that mouse *Pomc* enhancers nPE1 and nPE2 can drive expression to POMC neurons of the zebrafish hypothalamus, even though DNA sequences similar to these elements cannot be found in the genomes of any non-mammalian vertebrate. Indeed, nPE2 is present in all mammalian groups, whereas nPE1 can be found only in placental mammals ([46,47] and figure 6). Evidence that nPE enhancers are mammalian novelties comes from detailed

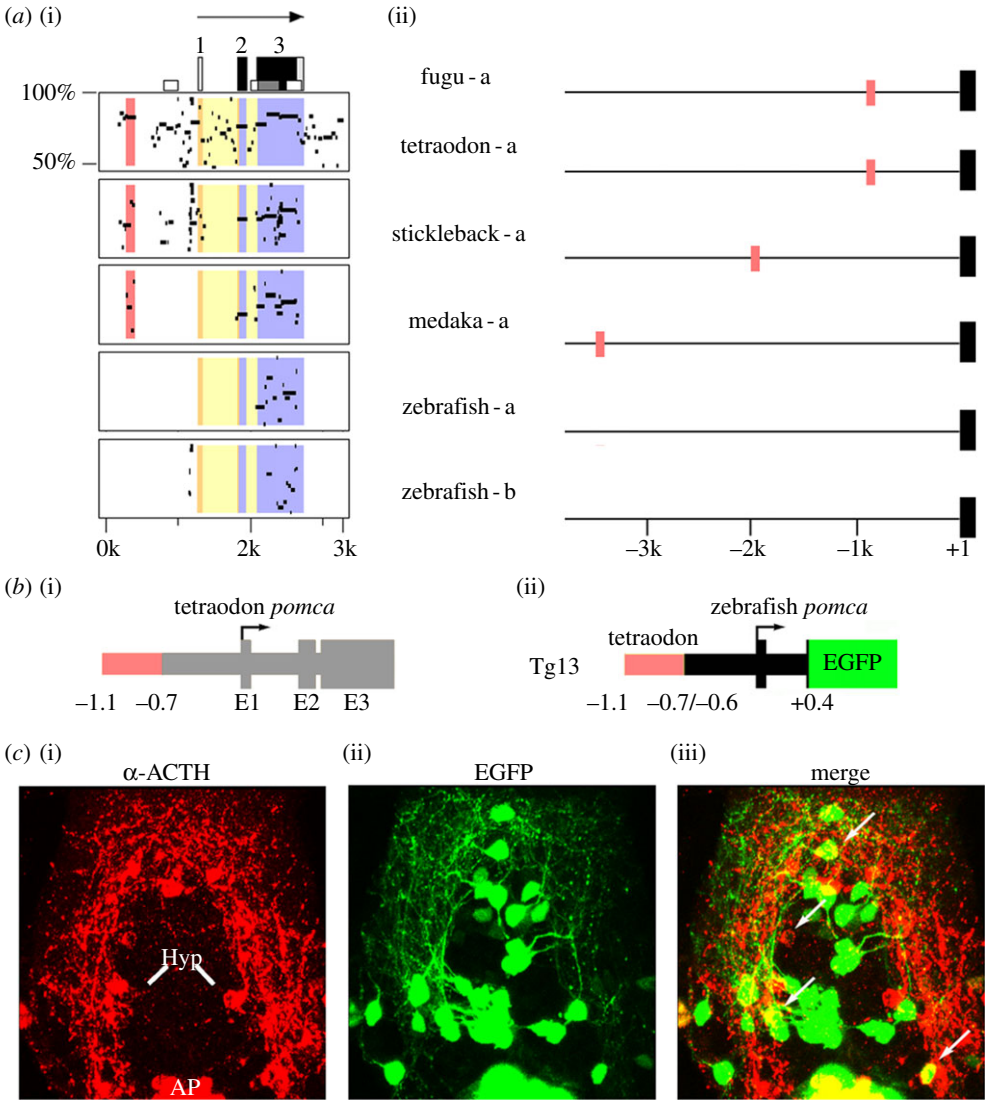

Figure 5. Identification of a conserved sequence in Acanthomorph fishes which drives EGFP expression to hypothalamic POMC neurons. (*a*) (i) Phylogenetic footprinting PipMaker analysis of fugu *pomca* and 5′ flanking sequences in comparison with homologous DNA taken from tetraodon, stickleback, medaka and zebrafish *pomc a* and *b*. A conserved block is shown in pink. (ii) Location of the conserved sequence is indicated with a pink box in the *pomca* 5′ flanking regions of each species relative to exon 1 (black box). (*b*) (i) Scheme of *tetraodon pomca* gene. Exons are numbered. Pink area corresponds to the segment included in Tg13 shown on the right that maintains the basic structure of previous transgenes. (*c*) Horizontal brain cryosections of zebrafish (5 dpf) carrying Tg13 subjected to double immunohistochemistry against (i) ACTH, (ii) EGFP or (iii) superimposed images. Coexpression is indicated by white arrows. Hyp, NLT of the hypothalamus; AP, anterior lobe of the pituitary.

genomic comparisons showing that both nPEs are derived from TE sequences: nPE2 is most similar to TEs of the CORE-SINE subfamily [46], which are very abundant in mammalian genomes [71,72], whereas nPE1 is most similar to a long terminal-repeat retrotransposon of the MaLR subfamily [47], which is also restricted to mammals [73]. The limited phylogenetic distribution of nPEs, and the fact that they are more similar to TEs of mammals, strongly indicates that these enhancers are mammalian novelties (figure 6). As a consequence, nPEs are predicted to be analogous, but not homologous (orthologous), to the neuronal *Pomc* regulatory regions of other vertebrate groups.

Other works have indicated that mammalian enhancers can drive expression to zebrafish embryos in the absence of overt sequence conservation [39,43,74,75], but in these studies it was not clear whether the mammalian enhancers had originated de novo in the mammalian lineage or if they were divergent orthologues to those of zebrafish. As indicated by the work of Taher *et al*. [45], which used NCEs of the amphibian *Xenopus* as an intermediate species to find divergent mammalian and zebrafish NCEs, orthologous vertebrate enhancers can become so dissimilar by extensive mutations in TFBSs and intervening sequences that computer algorithms may fail to identify them. Thus, the defined timing of origin of nPEs makes them, to the best of our knowledge, the first bona fide non-homologous vertebrate enhancers that show conserved function in mammals and teleosts. Conclusive de novo generation of vertebrate enhancers has also been shown in the teleost fish medaka, in which three enhancers that are only active in medaka are derived from degenerating exonic sequences, but in these cases the enhancers were shown to be active only in medaka itself [70].

Even mammalian enhancers that have zebrafish homologues can fail to drive specific, eutopic expression in transgenic fish. In a comparative enhancer study, Ritter *et al*. [40] observed that of 13 orthologous human and zebrafish enhancers tested, only four (31%) yielded similar expression patterns in transgenic zebrafish embryos. A larger study

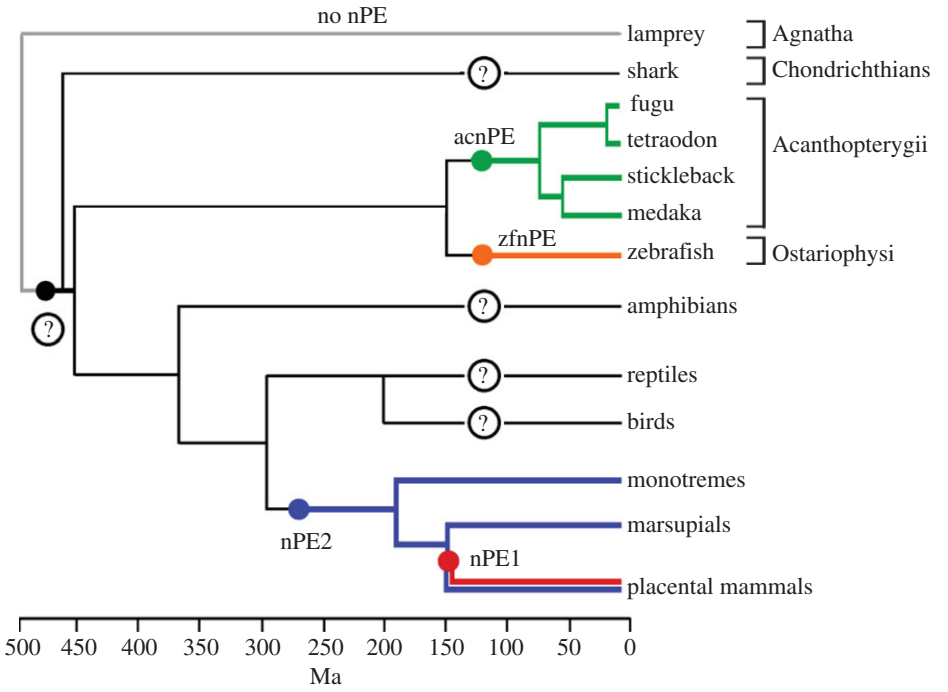

Figure 6. Phylogenetic tree showing the hypothetical turnover of neuronal *Pomc* enhancers (nPE) along vertebrate evolution. Mammalian nPE1 (red) and nPE2 (blue), zebrafish zfnPE (orange) and a putative enhancer of Acanthopterygii (green) seem to be evolutionary unrelated functional analogues. The coloured circles indicate the probable time of origin as functional enhancers. The black circle indicates the origin of a putative ancestral nPE and the question marks depict the current lack of identity for nPEs in various vertebrate classes or lineages that express *Pomc* in the ventromedial hypothalamus.

testing 47 human enhancers with zebrafish orthologues [44] found that 36% displayed radically different expression patterns in mouse and zebrafish transgenesis. Given this context, it is particularly striking that nPEs, which are not homologous to enhancers in zebrafish, nevertheless gave specific expression in zebrafish hypothalamic POMC neurons.

NCEs conserved between mammals and teleosts are typically located close to genes encoding transcription factors and receptors involved in embryonic development, probably because these genes require tight temporal and spatial regulation [19,20]. Although it is expressed early during organogenesis, hypothalamic POMC has no known developmental function. Instead, it is necessary for fitness in adulthood, but even then, hypothalamic *Pomc* mRNA levels can vary widely without gross physiological effects [76]. Thus, non-developmental genes such as *Pomc* may be particularly prone to enhancer turnover.

The fact that the mammalian enhancers direct reporter gene expression to zebrafish POMC neurons indicates that the specific set of transcription factors that defines a *Pomc*-expressing hypothalamic neuron is present in both mammalian and teleost POMC neurons. Thus, the *cis*-acting motifs controlling *Pomc* neuronal expression in the zebrafish should be very similar to those present in the mouse enhancers, meaning that the *cis*/*trans* code of neuronal *Pomc* expression has been conserved since the times of the bony fish ancestor. The fact that nPEs are mammalian novelties indicates that their *cis*-elements had to be created de novo before the new enhancers eventually replaced the ancestral *Pomc* regulatory elements during early mammalian evolution (figure 1). We observed, however, that nPEs reproducibly drove EGFP expression to only around 50% of zebrafish POMC neurons, and there was EGFP expression in some non-POMC-ir neurons within the zebrafish hypothalamus. This apparent lack of perfect targeting precision, found even in transgenic studies with CNEs [40,44], is probably due to species-specific differences in *Pomc* regulation and is hardly surprising considering the evolutionary distance between mammals and teleosts.

The mammalian arcuate nucleus and the teleost NLT are generally considered to be homologous anatomical structures since they are both located in the ventromedial hypothalamus and express a highly similar gene set, such as agouti-related peptide (*agrp*) [64,77,78], somatostatin (*sst*) [79], growth hormone releasing hormone (*ghrh*) [80], the leptin receptor (*lepr*) [81] and *Pomc* [51,64]. The NLT and the arcuate also show conserved function. In mammals, POMC-derived melanocortin peptides produced by arcuate neurons play a crucial role in the control of energy balance, and are antagonized by Agouti-related protein (AgRP), a peptide produced by a separate population of arcuate neurons [50]. In fish, there is abundant evidence that POMC melanocortin peptides and AgRP also control energy balance [77,78,82–85]. Not everything is conserved, though: while the mouse POMC neurons coexpress the neuropeptide CART, and AgRP is coexpressed with NPY, in zebrafish neither CART nor NPY is expressed in the NLT [86,87].

Currently, little is known about transcription factors regulating neuronal *Pomc* expression in any animal model. Our results suggest that the transcription factor repertoires of POMC neurons in the arcuate nucleus and the NLT are similar and define, at least to some extent, a unique regulatory code, consistent with the idea that the neurosecretory hypothalamus is an ancient metazoan structure [88]. As nPEs are functional both during zebrafish embryonic development and adulthood, those common transcription factors are present throughout the life of zebrafish POMC neurons. Given the relative ease and low cost of genetic manipulation of zebrafish compared with mice, our neuronal *Pomc*–EGFP transgenic zebrafish line could provide a new tool to search

for transcription factors controlling hypothalamic *Pomc* expression, similar to the use of a zebrafish pituitary *Pomc*–EGFP line to screen for mutations altering pituitary development [89]. Genes affecting nPE1 and nPE2 function in transgenic zebrafish are likely to be the same as those that control the enhancers in mammals.

(b) Zebrafish *pomca* neuronal regulatory region

Pomc is not found in basal chordates but exists in all vertebrate groups, including jawless fish (agnathans) and cartilaginous fish [90]. In the agnathan lamprey, *Pomc* paralogues *Poc* and *Pom* are expressed in the pituitary, but not in the brain [91], whereas cartilaginous fish do express *Pomc* in the hypothalamus [92]. Thus, it seems that *Pomc* was originally a pituitary gene that became transcriptionally active in the hypothalamus in an ancestor to jawed fish (figure 6). In teleosts, amphibians and mammals, the *Pomc* proximal promoter is responsible for expression in the pituitary [52,54,56,58], but mammalian nPEs are the only known neuronal regulatory regions of the gene.

The identification of *Pomc* neuronal *cis*-elements in other vertebrates might help understanding the various paths that enhancers can take during evolution. In a first step in this direction, we sought to identify the regulatory elements that control zebrafish *Pomc* in the hypothalamus. Comparative genomics was of no use in this case, because non-coding conserved regions could not be found around the zebrafish *pomca* gene, even when comparing it with the *pomca* loci of other teleosts. The reason for this is likely to be the distant evolutionary relationship between the superorder Ostariophysi, to which zebrafish belongs, and the superorder Acanthomorpha, to which other sequenced teleosts such as fugu, tetraodon, medaka and three-spined stickleback belong [93]. The efficient use of fish comparative genomics to find regulatory regions in zebrafish would greatly benefit from the sequencing of the genomes of closer relatives such as the channel catfish (*Ictalurus punctatus*), the common carp (*Cyprinus carpio*) or the goldfish (*Carassius auratus*).

From an initial 10 kb region upstream of the *pomca* promoter, we narrowed the *pomca* neuronal regulatory region down to 1.2 kb. This distal fragment drove efficient EGFP expression to ACTH-ir neurons of the zebrafish NLT. Thus, we consider that this region harbours a bona fide neuronal *Pomc* enhancer that we named zfnPE. Interestingly, the 1.2 kb region is embedded within the locus immediately 5′ of *pomca*, encompassing exon 1 and part of intron 1 of a gene similar to *angpl*. We have not yet determined whether the neuronal regulatory region is located within the intron or the exon, but mutagenesis of three TAAT sites within the intron abolished EGFP expression in the hypothalamus. These critical sites are embedded within DNA sequences derived from a hAT Charlie DNA transposon, a family of TEs present in high copy number in the zebrafish genome. Further work is necessary to determine the birth date and phylogeny of this exaptation event (figure 6). Establishing the precise *cis*-elements controlling zebrafish *pomca* expression might help pinpoint the elements regulating mammalian *Pomc*, because zfnPE, nPE1 and nPE2 are functional analogues.

5. Concluding remarks

The de novo origin of enhancers is presumably an important feature in the evolution of gene regulation. The appearance of *Pomc* nPE enhancers early in the mammalian lineage offers a clear example of how new enhancers can arise and eventually replace ancestral ones while conserving the underlying *cis*- and *trans*-regulatory codes (figure 6). This was possibly facilitated by the fact that *Pomc* is not a gene involved in development, making its regulation easier to be 'tinkered' with by evolution. Zebrafish is an ideal model for such studies, because its phylogenetic distance to mammals coupled with the ease to perform transgenesis offers many possibilities to discover what is ancestral and what is derived in mammalian regulatory evolution. At the same time, the search for the regulatory regions in other species, as we did here with zebrafish and tetraodon *pomca*, will illuminate the evolutionary paths that regulatory regions can take in different lineages during evolution.

Acknowledgements. We thank Nora Calcaterra and K. Kawakami for providing valuable resources.

Funding statement. This work was supported by an NIH grant DK068400 (M.J.L., M.R.), International Research Scholar Grant of the HHMI (M.R.), Agencia Nacional de Promoción Científica y Tecnológica, Argentina (M.R., F.J.S.) and Universidad de Buenos Aires (M.R., V.F.B.).

References

1. Carroll SB. 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36. (doi:10.1016/j.cell.2008.06.030)
2. Carey M. 1998 The enhanceosome and transcriptional synergy. *Cell* **92**, 5–8. (doi:10.1016/S0092-8674(00)80893-4)
3. Swanson CI, Evans NC, Barolo S. 2010 Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18**, 359–370. (doi:10.1016/j.devcel.2009.12.026)
4. Ludwig MZ. 2002 Functional evolution of noncoding DNA. *Curr. Opin. Genet. Dev.* **12**, 634–639. (doi:10.1016/S0959-437X(02)00355-6)
5. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. 2003 Scanning human gene deserts for long-range enhancers. *Science* **302**, 413. (doi:10.1126/science.1088328)
6. Lettice LA *et al*. 2003 A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735. (doi:10.1093/hmg/ddg180)
7. Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T. 2005 Elimination of a long-range *cis*-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development* **132**, 797–803. (doi:10.1242/dev.01613)
8. Irimia M *et al*. 2012 Extensive conservation of ancient microsynteny across metazoans due to *cis*-regulatory constraints. *Genome Res.* **22**, 2356–2367. (doi:10.1101/gr.139725.112)
9. Marinic M, Aktas T, Ruf S, Spitz F. 2013 An integrated holo-enhancer unit defines tissue and gene specificity of the *fgf8* regulatory landscape. *Dev. Cell* **24**, 530–542. (doi:10.1016/j.devcel.2013.01.025)
10. Kaessmann H. 2010 Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326. (doi:10.1101/gr.101386.109)
11. Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA. 2000 Identification of a coordinate regulator of interleukins 4, 13, and 5

by cross-species sequence comparisons. *Science* **288**, 136–140. (doi:10.1126/science.288.5463.136)

12. Wasserman WW, Sandelin A. 2004 Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* **5**, 276–287. (doi:10.1038/nrg1315)
13. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. 2004 Ultraconserved elements in the human genome. *Science* **304**, 1321–1325. (doi:10.1126/science.1098119)
14. Woolfe A *et al*. 2005 Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7. (doi:10.1371/journal.pbio.0030007)
15. Shin JT, Priest JR, Ovcharenko I, Ronco A, Moore RK, Burns CG, MacRae CA. 2005 Human–zebrafish non-coding conserved elements act *in vivo* to regulate transcription. *Nucleic Acids Res.* **33**, 5437–5445. (doi:10.1093/nar/gki853)
16. Pennacchio LA *et al*. 2006 *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502. (doi:10.1038/nature05295)
17. Hedges SB. 2002 The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838–849. (doi:10.1038/nrg929)
18. Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, Pennacchio LA. 2006 Close sequence comparisons are sufficient to identify human *cis*-regulatory elements. *Genome Res.* **16**, 855–863. (doi:10.1101/gr.4717506)
19. Plessy C, Dickmeis T, Chalmel F, Strahle U. 2005 Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes. *Trends Genet.* **21**, 207–210. (doi:10.1016/j.tig.2005.02.006)
20. Woolfe A, Elgar G. 2008 Organization of conserved elements near key developmental regulators in vertebrate genomes. *Adv. Genet.* **61**, 307–338. (doi:10.1016/S0065-2660(07)00012-0)
21. Lindblad-Toh K *et al*. 2011 A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482. (doi:10.1038/nature10530)
22. Visel A *et al*. 2009 ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858. (doi:10.1038/nature07730)
23. Heintzman ND *et al*. 2009 Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112. (doi:10.1038/nature07829)
24. Zentner GE, Scacheri PC. 2012 The chromatin fingerprint of gene enhancer elements. *J. Biol. Chem.* **287**, 30 888–30 896. (doi:10.1074/jbc.R111.296491)
25. Buecker C, Wysocka J. 2012 Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284. (doi:10.1016/j.tig.2012.02.008)
26. Blow MJ *et al*. 2010 ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810. (doi:10.1038/ng.650)
27. May D *et al*. 2011 Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* **44**, 89–93. (doi:10.1038/ng.1006)
28. Lunyak VV *et al*. 2007 Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248–251. (doi:10.1126/science.1140871)
29. Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J, Tuan D. 2010 Long-range function of an intergenic retrotransposon. *Proc. Natl Acad. Sci. USA* **107**, 12 992–12 997. (doi:10.1073/pnas.1004139107)
30. Loots GG, Kneissel M, Keller H, Baptist M, Chang J, Collette NM, Ovcharenko D, Plajzer-Frick I, Rubin EM. 2005 Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Res.* **15**, 928–935. (doi:10.1101/gr.3437105)
31. Rahimov F *et al*. 2008 Disruption of an AP-2alpha binding site in an *IRF6* enhancer is associated with cleft lip. *Nat. Genet.* **40**, 1341–1347. (doi:10.1038/ng.242)
32. Jeong Y *et al*. 2008 Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* **40**, 1348–1353. (doi:10.1038/ng.230)
33. Benko S *et al*. 2009 Highly conserved non-coding elements on either side of *SOX9* associated with Pierre Robin sequence. *Nat. Genet.* **41**, 359–364. (doi:10.1038/ng.329)
34. Ghiasvand NM, Rudolph DD, Mashayekhi M, Brzezinski JAt, Goldman D, Glaser T. 2011 Deletion of a remote enhancer near *ATOH7* disrupts retinal neurogenesis, causing NCRNA disease. *Nat. Neurosci.* **14**, 578–586. (doi:10.1038/nn.2798)
35. Sur IK *et al*. 2012 Mice lacking a *Myc* enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363. (doi:10.1126/science.1228606)
36. Zhou X *et al*. 2012 Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum. Mol. Genet.* **21**, 1325–1335. (doi:10.1093/hmg/ddr569)
37. Pashos EE, Kague E, Fisher S. 2008 Evaluation of *cis*-regulatory function in zebrafish. *Brief Funct. Genomic Proteomic* **7**, 465–473. (doi:10.1093/bfgp/eln045)
38. de la Calle-Mustienes E, Feijoo CG, Manzanares M, Tena JJ, Rodriguez-Seguel E, Letizia A, Allende ML, Gomez-Skarmeta JL. 2005 A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate *Iroquois* cluster gene deserts. *Genome Res.* **15**, 1061–1072. (doi:10.1101/gr.4004805)
39. McGaughey DM, Vinton RM, Huynh J, Al-Saif A, Beer MA, McCallion AS. 2008 Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at *phox2b*. *Genome Res.* **18**, 252–260. (doi:10.1101/gr.6929408)
40. Ritter DI, Li Q, Kostka D, Pollard KS, Guo S, Chuang JH. 2010 The importance of being *cis*: evolution of orthologous fish and mammalian enhancer activity. *Mol. Biol. Evol.* **27**, 2322–2332. (doi:10.1093/molbev/msq128)
41. Chatterjee S, Bourque G, Lufkin T. 2011 Conserved and non-conserved enhancers direct tissue specific transcription in ancient germ layer specific developmental control genes. *BMC Dev. Biol.* **11**, 63. (doi:10.1186/1471-213X-11-63)
42. Shen Y *et al*. 2012 A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120. (doi:10.1038/nature11243)
43. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. 2006 Conservation of *RET* regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276–279. (doi:10.1126/science.1124070)
44. Ariza-Cosano A, Visel A, Pennacchio LA, Fraser HB, Gomez-Skarmeta JL, Irimia M, Bessa J. 2012 Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression. *BMC Genomics* **13**, 713. (doi:10.1186/1471-2164-13-713)
45. Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega MA, McCallion AS, Ovcharenko I. 2011 Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res.* **21**, 1139–1149. (doi:10.1101/gr.119016.110)
46. Santangelo AM, de Souza FS, Franchini LF, Bumaschny VF, Low MJ, Rubinstein M. 2007 Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. *PLoS Genet.* **3**, 1813–1826. (doi:10.1371/journal.pgen.0030166)
47. Franchini LF, Lopez-Leal R, Nasif S, Beati P, Gelman DM, Low MJ, de Souza FJ, Rubinstein M. 2011 Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. *Proc. Natl Acad. Sci. USA* **108**, 15 270–15 275. (doi:10.1073/pnas.1104997108)
48. Raffin-Sanson ML, de Keyzer Y, Bertagna X. 2003 Proopiomelanocortin, a polypeptide precursor with multiple functions: from physiology to pathological conditions. *Eur. J. Endocrinol.* **149**, 79–90. (doi:10.1530/eje.0.1490079)
49. Jenks BG, Kidane AH, Scheenen WJ, Roubos EW. 2007 Plasticity in the melanotrope neuroendocrine interface of *Xenopus laevis*. *Neuroendocrinology* **85**, 177–185. (doi:10.1159/000101434)
50. Millington GW. 2007 The role of proopiomelanocortin (POMC) neurones in feeding behaviour. *Nutr. Metab. (Lond.)* **4**, 18. (doi:10.1186/1743-7075-4-18)
51. de Souza FS, Bumaschny VF, Low MJ, Rubinstein M. 2005 Subfunctionalization of expression and peptide domains following the ancient duplication of the proopiomelanocortin gene in teleost fishes. *Mol. Biol. Evol.* **22**, 2417–2427. (doi:10.1093/molbev/msi236)
52. Liu NA, Huang H, Yang Z, Herzog W, Hammerschmidt M, Lin S, Melmed S. 2003 Pituitary corticotroph ontogeny and regulation in transgenic zebrafish. *Mol. Endocrinol.* **17**, 959–966. (doi:10.1210/me.2002-0392)
53. Herzog W, Zeng X, Lele Z, Sonntag C, Ting JW, Chang CY, Hammerschmidt M. 2003 Adenohypophysis formation in the zebrafish and its dependence on sonic hedgehog. *Dev. Biol.* **254**, 36–49. (doi:10.1016/S0012-1606(02)00124-0)
54. Bumaschny VF, de Souza FS, Lopez Leal RA, Santangelo AM, Baetscher M, Levi DH, Low MJ,

Rubinstein M. 2007 Transcriptional regulation of pituitary POMC is conserved at the vertebrate extremes despite great promoter sequence divergence. *Mol. Endocrinol.* **21**, 2738–2749. (doi:10.1210/me.2006-0557)

55. Liu B, Hammer GD, Rubinstein M, Mortrud M, Low MJ. 1992 Identification of DNA elements cooperatively activating proopiomelanocortin gene expression in the pituitary glands of transgenic mice. *Mol. Cell Biol.* **12**, 3978–3990.
56. Jenks BG. 2009 Regulation of proopiomelanocortin gene expression: an overview of the signaling cascades, transcription factors, and responsive elements involved. *Ann. N.Y. Acad. Sci.* **1163**, 17–30. (doi:10.1111/j.1749-6632.2008.03620.x)
57. Langlais D, Couture C, Sylvain-Drolet G, Drouin J. 2011 A pituitary-specific enhancer of the POMC gene with preferential activity in corticotrope cells. *Mol. Endocrinol.* **25**, 348–359. (doi:10.1210/me.2010-0422)
58. Jansen EJ, Holling TM, van Herp F, Martens GJ. 2002 Transgene-driven protein expression specific to the intermediate pituitary melanotrope cells of *Xenopus laevis*. *FEBS Lett.* **516**, 201–207. (doi:10.1016/S0014-5793(02)02523-1)
59. Young JI, Otero V, Cerdan MG, Falzone TL, Chan EC, Low MJ, Rubinstein M. 1998 Authentic cell-specific and developmentally regulated expression of pro-opiomelanocortin genomic fragments in hypothalamic and hindbrain neurons of transgenic mice. *J. Neurosci.* **18**, 6631–6640.
60. de Souza FS, Santangelo AM, Bumaschny V, Avale ME, Smart JL, Low MJ, Rubinstein M. 2005 Identification of neuronal enhancers of the proopiomelanocortin gene by transgenic mouse analysis and phylogenetic footprinting. *Mol. Cell. Biol.* **25**, 3076–3086. (doi:10.1128/MCB.25.8.3076-3086.2005)
61. Franchini LF, de Souza FS, Low MJ, Rubinstein M. 2012 Positive selection of co-opted mobile genetic elements in a mammalian gene: if you can't beat them, join them. *Mob. Genet. Elements* **2**, 106–109. (doi:10.4161/mge.20267)
62. Thisse C, Thisse B, Schilling TF, Postlethwait JH. 1993 Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development* **119**, 1203–1215.
63. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. 2005 MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933–2942. (doi:10.1093/bioinformatics/bti473)
64. Forlano PM, Cone RD. 2007 Conserved neurochemical pathways involved in hypothalamic control of energy homeostasis. *J. Comp. Neurol.* **505**, 235–248. (doi:10.1002/cne.21447)
65. Lohr H, Hammerschmidt M. 2011 Zebrafish in endocrine systems: recent advances and implications for human disease. *Annu. Rev. Physiol.* **73**, 183–211. (doi:10.1146/annurev-physiol-012110-142320)
66. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. 2000 PipMaker: a web server for aligning two genomic DNA sequences. *Genome Res.* **10**, 577–586. (doi:10.1101/gr.10.4.577)
67. Gonzalez-Nunez V, Gonzalez-Sarmiento R, Rodriguez RE. 2003 Identification of two proopiomelanocortin genes in zebrafish (*Danio rerio*). *Brain Res. Mol. Brain Res.* **120**, 1–8. (doi:10.1016/j.molbrainres.2003.09.012)
68. Birnbaum RY *et al*. 2012 Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22**, 1059–1068. (doi:10.1101/gr.133546.111)
69. Howe K *et al*. 2013 The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503. (doi:10.1038/nature12111)
70. Eichenlaub MP, Ettwiller L. 2011 De novo genesis of enhancers in vertebrates. *PLoS Biol.* **9**, e1001188. (doi:10.1371/journal.pbio.1001188)
71. Gilbert N, Labuda D. 2000 Evolutionary inventions and continuity of CORE-SINEs in mammals. *J. Mol. Biol.* **298**, 365–377. (doi:10.1006/jmbi.2000.3695)
72. Gilbert N, Labuda D. 1999 CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl Acad. Sci. USA* **96**, 2869–2874. (doi:10.1073/pnas.96.6.2869)
73. Smit AF. 1993 Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* **21**, 1863–1872. (doi:10.1093/nar/21.8.1863)
74. Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I. 2010 Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.* **20**, 565–577. (doi:10.1101/gr.104471.109)
75. Narlikar L, Sakabe NJ, Blanski AA, Arimura FE, Westlund JM, Nobrega MA, Ovcharenko I. 2010 Genome-wide discovery of human heart enhancers. *Genome Res.* **20**, 381–392. (doi:10.1101/gr.098657.109)
76. Bumaschny VF, Yamashita M, Casas-Cordero R, Otero-Corchon V, de Souza FS, Rubinstein M, Low MJ. 2012 Obesity-programmed mice are rescued by early genetic intervention. *J. Clin. Invest.* **122**, 4203–4212. (doi:10.1172/JCI62543)
77. Song Y, Cone RD. 2007 Creation of a genetic model of obesity in a teleost. *FASEB J.* **21**, 2042–2049. (doi:10.1096/fj.06-7503com)
78. Song Y, Golling G, Thacker TL, Cone RD. 2003 Agouti-related protein (AGRP) is conserved and regulated by metabolic state in the zebrafish, *Danio rerio*. *Endocrine* **22**, 257–265. (doi:10.1385/ENDO:22:3:257)
79. Canosa LF, Cerda-Reverter JM, Peter RE. 2004 Brain mapping of three somatostatin encoding genes in the goldfish. *J. Comp. Neurol.* **474**, 43–57. (doi:10.1002/cne.20097)
80. Castro A, Becerra M, Manso MJ, Tello J, Sherwood NM, Anadon R. 2009 Distribution of growth hormone-releasing hormone-like peptide: immunoreactivity in the central nervous system of the adult zebrafish (*Danio rerio*). *J. Comp. Neurol.* **513**, 685–701. (doi:10.1002/cne.21977)
81. Liu Q, Chen Y, Copeland D, Ball H, Duff RJ, Rockich B, Londraville RL. 2010 Expression of leptin receptor gene in developing and adult zebrafish. *Gen. Comp. Endocrinol.* **166**, 346–355. (doi:10.1016/j.ygcen.2009.11.015)
82. Cerda-Reverter JM, Agulleiro MJ, RG R, Sanchez E, Ceinos R, Rotllant J. 2011 Fish melanocortin system. *Eur. J. Pharmacol.* **660**, 53–60. (doi:10.1016/j.ejphar.2010.10.108)
83. Cerda-Reverter JM, Peter RE. 2003 Endogenous melanocortin antagonist in fish: structure, brain mapping, and regulation by fasting of the goldfish agouti-related protein gene. *Endocrinology* **144**, 4552–4561. (doi:10.1210/en.2003-0453)
84. Cerda-Reverter JM, Schioth HB, Peter RE. 2003 The central melanocortin system regulates food intake in goldfish. *Regul. Pept.* **115**, 101–113. (doi:10.1016/S0167-0115(03)00144-7)
85. Schjolden J, Schioth HB, Larhammar D, Winberg S, Larson ET. 2009 Melanocortin peptides affect the motivation to feed in rainbow trout (*Oncorhynchus mykiss*). *Gen. Comp. Endocrinol.* **160**, 134–138. (doi:10.1016/j.ygcen.2008.11.003)
86. Mukherjee A, Subhedar NK, Ghose A. 2012 Ontogeny of the cocaine- and amphetamine-regulated transcript (CART) neuropeptide system in the brain of zebrafish, *Danio rerio*. *J. Comp. Neurol.* **520**, 770–797. (doi:10.1002/cne.22779)
87. Yokobori E, Azuma M, Nishiguchi R, Kang KS, Kamijo M, Uchiyama M, Matsuda K. 2012 Neuropeptide Y stimulates food intake in the zebrafish, *Danio rerio*. *J. Neuroendocrinol.* **24**, 766–773. (doi:10.1111/j.1365-2826.2012.02281.x)
88. Tessmar-Raible K, Raible F, Christodoulou F, Guy K, Rembold M, Hausen H, Arendt D. 2007 Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell* **129**, 1389–1400. (doi:10.1016/j.cell.2007.04.041)
89. Kurrasch DM, Nevin LM, Wong JS, Baier H, Ingraham HA. 2009 Neuroendocrine transcriptional programs adapt dynamically to the supply and demand for neuropeptides as revealed in NSF mutant zebrafish. *Neural Dev.* **4**, 22. (doi:10.1186/1749-8104-4-22)
90. Dores RM, Baron AJ. 2011 Evolution of POMC: origin, phylogeny, posttranslational processing, and the melanocortins. *Ann. N.Y. Acad. Sci.* **1220**, 34–48. (doi:10.1111/j.1749-6632.2010.05928.x)
91. Ficele G, Heinig JA, Kawauchi H, Youson JH, Keeley FW, Wright GM. 1998 Spatial and temporal distribution of proopiomelanotropin and proopiocortin mRNA during the life cycle of the sea lamprey: a qualitative and quantitative *in situ* hybridisation study. *Gen. Comp. Endocrinol.* **110**, 212–225. (doi:10.1006/gcen.1998.7071)
92. Chiba A. 2001 Marked distributional difference of alpha-melanocyte-stimulating hormone (alpha-MSH)-like immunoreactivity in the brain between two elasmobranchs (*Scyliorhinus torazame* and *Etmopterus brachyurus*): an immunohistochemical study. *Gen. Comp. Endocrinol.* **122**, 287–295. (doi:10.1006/gcen.2001.7628)
93. Chen WJ, Orti G, Meyer A. 2004 Novel evolutionary relationship among four fish model systems. *Trends Genet.* **20**, 424–431. (doi:10.1016/j.tig.2004.07.005)