Origin and evolutionary landscape of *Nr2f* transcription factors across Metazoa

Ugo Coppola¹ and Joshua S. Waxman^{1,2,*}

¹Molecular Cardiovascular Biology Division and Heart Institute, Cincinnati Children's
Research Foundation, Cincinnati, OH 45229, USA.

²Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH
45229, USA.

*Corresponding author

E-mail: joshua.waxman@cchmc.org

Short title: Evolution of Nr2f transcription factors

Abstract

Background: Nuclear Receptor Subfamily 2 Group F (Nr2f) orphan nuclear hormone transcription factors (TFs) are fundamental regulators of many developmental processes in invertebrates and vertebrates. Despite the importance of these TFs throughout metazoan development, previous work has not clearly outlined their evolutionary history.

Results: We integrated molecular phylogeny with comparisons of intron/exon structure, domain architecture, and syntenic conservation to define critical evolutionary events that distinguish the *Nr2f* gene family in Metazoa. Our data indicate that a single ancestral pre-metazoan *Nr2f* gene, we have termed *Nr2f1/2/5/6*, predated six main Bilateria subfamilies, which include a single *Nr2f1/2/5* homolog that is present throughout protostomes and invertebrate deuterostomes, *Nr2f1/2* homologs in agnathans, and *Nr2f1*, *Nr2f2*, *Nr2f5*, *Nr2f6* orthologs that are found in gnathostomes. The three *Nr2f1/2* members in agnathans are due to independent expansions not found in gnathostomes, while the vertebrate *Nr2f1*, *Nr2f2*, *Nr2f5* members arose from whole-genome duplications (WGDs). However, *Nr2f6* members are the most divergent subfamily, likely originating from an ancient duplication, and are only retained by gnathostomes. Interestingly, *Nr2f5* TFs have been independently lost in both cartilaginous fish and amniotes, such as humans. Furthermore, our analysis shows there are differential expansions and losses of *Nr2f* genes in teleosts following their additional rounds of WGDs.

Conclusion: Overall, our evolutionary genomic analysis of Nr2f proteins helps to reveal the origins and previously unrecognized relationships of this ancient transcription factor family, which may allow for greater insights into the conservation of Nr2f functions that shape Metazoan body plans.

Introduction

Nuclear hormone receptors (NRs) form a large, ancient superfamily of transcription factors (TFs) found in all Metazoa [1]. While NR functions are often dictated by interactions with specific ligands, including steroids, thyroid hormones, and retinoids [2,3], the ligands for many NRs, called orphan NRs, are still not known [4]. Nuclear Receptor Subfamily 2 Group F Members (Nr2fs), initially named Chicken ovalbumin upstream promoter-transcription factors (Coup-TFs) due to their ability to bind the COUP element of the ovalbumin gene [5–7], are some of the most highly studied orphan NRs. Despite an overall expansion of the NR superfamily [1,2], invertebrate phyla appear to have predominantly retained a single *Nr2f* gene. Only one *Nr2f* member is present in the protostome *Drosophila melanogaster* (fly), early-branching deuterostome *Strongylocentrotus purpuratus* (sea urchin) [8,9], and invertebrate chordates *Branchiostoma floridae* (amphioxus) and *Ciona robusta* (sea squirt) [10,11]. However, the number of *Nr2f* genes in early-branching metazoans is presently less clear. In cnidarians, one *Nr2f* has been reported in *Hydra*, while multiple have been reported in *Nematostella* [12,13]. In contrast to most invertebrates, vertebrates have exhibited a significant expansion of the *Nr2f* family, with gnathostomes having multiple *Nr2f* genes. Furthermore, teleosts possess additional *Nr2f* Ohnologs (duplicates originating from whole-genome duplication (WGD)) [14], most likely reflecting the additional WGDs that have occurred in the teleost lineage [15,16].

Nr2f proteins are highly conserved at the sequence level throughout Metazoa (Giguere, 1999). From the N-terminus to the C-terminus, all Nr2f proteins have six domains (Fig 1): an A/B domain, which contains the activating function-1 (AF-1) domain; the C domain, which contains the DNA-binding domain (DBD); the D domain (a linker); the E domain, which is comprised of the ligand-binding domain (LBD) and an AF-2 domain; the F domain (C-terminal) [17]. While the A/B domains are the most divergent in sequence, strikingly, the

DBDs and LBDs of Nr2f members even from distantly related species (e.g. fly, sea urchin, frog, zebrafish, mouse, and human) are ~94% identical [18]. The extremely high degree of conservation among several species implies the preservation of critical roles for Nr2f in development and differentiation [18,19]. Requirements for *Nr2f* genes have been found in organs of all three germ layers during embryogenesis [12,19]. For instance, the *Drosophila* *Nr2f* homolog, called *seven up* (*svp*) is required for retinal, dorsal vessel, and liver development [20,21]. Furthermore, Nr2f transcription factors in vertebrates appear to both have acquired diverse and retained redundant functions. For instance, in mice, *Nr2f1* is predominantly required for neural development with a role in regulation of premigratory and migratory neural crest cells in the developing hindbrain [22,23]. However, the mouse *Nr2f2* gene is required for differentiation of mesodermal derivatives, including atrial cardiomyocytes of the heart and venous endothelial cells [19,24,25]. An example of redundancy are zebrafish *nr2f1a* and *nr2f2*, which are both required for proper ventricular cardiomyocyte and cranial muscle specification [26].

Fig 1. Schematic of conserved domain architecture of Nr2f transcription factors.

A/B (N-terminal variable domain with transactivating AF-1 domain), C (DBD, which contains two Zinc finger (Znf) motifs), D (a linker domain), E (LBD plus transactivating AF-2 domain), and F (C-terminal).

While Nr2f proteins were initially identified as transcriptional activators of chicken ovalbumin gene [5], they have since been shown to function directly as both transcriptional activators and repressors in several developmental contexts [19,27,28]. Nr2fs can bind a range of different response elements [29–31] and in signaling reporter assays can compete with and inhibit retinoic acid receptors [32]. *In vivo* they bind numerous targets that reflect their various

requirements in the specific tissues. For instance, *Nr2f1* KO mice also have inner ear defects [33]. In the mouse inner ear, direct targets of *Nr2f1* include fatty acid binding protein 7 (FABP7), cellular retinoic acid binding protein 1 (CRABP1)[34], microRNA-140 (mi-R140), and Kruppel-like 9 (Klf9) [34,35]. In adipogenesis, *Nr2f2* directly represses peroxisome proliferator-activated receptor γ (PPAR γ) downstream of canonical Wnt/ β -catenin signaling [36]. In the mammalian heart, *Nr2f2* is thought to directly orchestrate a regulatory network that facilitates atrial cardiomyocyte identity through concurrently promoting *Tbx5* and repressing *Irx4* and *Hey2*, the latter of which promote ventricular cardiomyocyte identity [37]. Thus, *Nr2fs* can activate and repress a range of direct targets related to their functions in specific tissues.

Despite the conservation and clear importance of this gene family to numerous developmental processes in Metazoa, we still do not completely understand the evolution of *Nr2f* TFs. Here, we investigated *Nr2f* family evolution through a combination of phylogenetic, domain architecture, intron/exon structure, and genomic synteny analyses. Our data show that the single *Nr2f1/2/5/6* gene in cnidarians and placozoans, represents the ancestral *Nr2f* to those found in protostomes and deuterostomes. Importantly, a single *Nr2f* homolog, which we have named *Nr2f1/2/5*, is present in the majority of invertebrates, while most vertebrate genomes contain *Nr2f1*, *Nr2f2*, *Nr2f5* and *Nr2f6* orthologs. Interestingly, the invertebrate *Nr2f1/2/5* and agnathan *Nr2f1/2* homologs have retained the greatest similarity with vertebrate *Nr2f1* and *Nr2f2* paralogs, while *Nr2f5* has been independently lost in the lineages of cartilaginous fish and amniotes. The vertebrate-specific *Nr2f6* subfamily, the most divergent *Nr2f* subfamily, predates the vertebrate radiation, suggesting a different evolutionary relationship to the other *Nr2f* genes. Overall, our data clarify the relationships among *Nr2f* genes within Metazoa and define the expansion, divergence, and independent loss events of extant *Nr2f* genes in vertebrates, which will allow us to make meaningful inferences about the conserved developmental functions of this family that have helped mold animal body plans.

Results

Phylogenetic reconstruction of *Nr2f* evolution in animals

Although previous work has investigated the homology of some *Nr2f*s within metazoans, these analyses were primarily focused on their relationship to other NRs and were limited by the comparatively little genomic information at the time [1,12,38]. To garner a better understanding of how the *Nr2f* family has evolved in animals, we performed a phylogenetic analysis using 130 *Nr2f* proteins with representatives from placozoans to mammals (Fig 2). Pre-metazoan models *Amphimedon queenslandica* (sponge) and *Mnemiopsis leidyi* (ctenophore) were not included, as we did not find putative *Nr2f* orthologs based on current databases. A maximum-likelihood (ML) phylogenetic tree showed the existence of distinct *Nr2f* subfamilies (Fig 2), which were also supported using a Bayesian model selection (S1 Fig). The placozoan *Trichoplax adhaerens* and cnidarian *Nematostella vectensis* *Nr2f*s, which we now call the *Nr2f1/2/5/6*, were at the root of all the metazoan *Nr2f* proteins (Fig 2). Interestingly, the tree indicated that the three distinct *Nr2f1/2/5/6* members in *N. vectensis* that were previously reported are likely the result of gene duplications [13]. The protostome and deuterostome *Nr2f* sequences clustered into six subfamilies, which we have called *Nr2f1/2/5*, *Nr2f1/2*, *Nr2f1*, *Nr2f2*, *Nr2f5*, and *Nr2f6*. The *Nr2f1/2/5* subfamily genes, which are highly conserved, yet evolutionary divergent from the *Nr2f1/2/5/6* genes present in early-branching animals, were found in invertebrate protostomes, invertebrate deuterostomes (hemichordates, echinoderms), and invertebrate chordates (amphioxus, tunicates). This *Nr2f1/2/5* group is more closely related to the branch that includes *Nr2f1/2*s from the agnathan vertebrate (lamprey) and vertebrate *Nr2f1* and *Nr2f2* proteins (Fig 2). The clustering of invertebrate *Nr2f1/2/5* proteins with *Nr2f1*, *Nr2f2*, and *Nr2f5* of gnathostomes suggests these proteins arose from distinct

duplicative events during vertebrate evolution. Furthermore, the three agnathan Nr2f proteins (Fig 2), which we have called Nr2f1/2A, Nr2f1/2B, and Nr2f1/2C, diverge and cluster together at the base of the vertebrate-specific Nr2f1 and Nr2f2 proteins (Fig 2), suggesting that the duplications leading to these proteins in agnathans were distinct from those that gave rise to the Nr2f paralogs in gnathostomes.

Fig 2. Evolutionary reconstruction of Nr2f proteins in metazoans.

Phylogenetic tree of Nr2f members demonstrating the existence of six protein classes: Nr2f1/2/5/6 (violet box), Nr2f1/2/5 (light blue box), Nr2f1/2 (yellow box), Nr2f1 (pink box), Nr2f2 (green box), Nr2f5 (blue box), Nr2f6 (orange box). Values at the branches indicate replicates obtained using the aLRT method.

Our analysis shows that all gnathostomes have retained Nr2f1, Nr2f2 and Nr2f6, whereas Nr2f5 has been lost by cartilaginous fish (chimaerae, sharks, skates) and amniotes (reptiles, avians, mammals), suggesting these were independent losses. Within the vertebrate Nr2f1, Nr2f2, and Nr2f5 cluster, the branching and distances from our phylogenetic tree indicate that Nr2f1 and Nr2f2 are more highly conserved (Fig 2; S1 Fig). Remarkably, Nr2f6 is the most divergent vertebrate Nr2f subfamily and occupies a distinct position in the tree, suggesting a different duplication event for its origin that occurred prior to gnathostomes. Moreover, to analyze the impact of WGDs that took place in teleosts [15,16], and specifically, in salmonids [39], we surveyed the Nr2f proteins of 12 teleost species (Fig 2). Consistent with the WGDs in these species, there was a tremendous expansion of the *Nr2f* family in this clade, although it was accompanied by differential *Nr2f* paralog losses in some species (Fig 2). To further interrogate the evolution of the Nr2f proteins, we examined alignments of the highly conserved zinc-fingers (Znf) within their DNA-binding domains (DBDs) using representatives

from each subfamily (Fig 3). Although there is a high degree of conservation in all the examined Nr2fs, the amino acid changes in the DBDs parallels the phylogenetic results of the whole proteins. The ancient Nr2f1/2/5/6 of early-branching metazoans showed a high degree of variability and multiple differences with respect to Nr2f1/2/5 DBDs of other invertebrates and the Nr2f DBDs in agnathans and gnathostomes. There is high similarity between Nr2f1/2, Nr2f1 and Nr2f2 DBDs in agnathans and gnathostomes, whereas Nr2f5 and Nr2f6 DBDs of gnathostomes exhibited specific changes that are consistent with their positions in the phylogenetic trees (Fig 2; Fig 3). Thus, our phylogenetic reconstruction of *Nr2f* genes in metazoans generally shows the presence of a single ortholog in invertebrates and significant expansion of the family in vertebrates that is punctuated with independent losses of *Nr2f5* in cartilaginous fishes and amniotes.

Fig 3. Zinc finger (Znf) motifs within the DBD of the Nr2f family.

Alignments of first (I) and second (II) Znfs. Yellow represents conserved amino acids conserved during evolution. White indicates amino acids that are not conserved. Turquoise and blue indicate the amino acid changes that are conserved within Znf I of Nr2f5 and -6, respectively. Magenta and green indicate the amino acid that are conserved within Znf II of Nr2f5 and -6, respectively.

***Nr2f* genes have conserved intron codes**

To complement the phylogenetic analysis of *Nr2f* genes, we first analyzed the conservation of *Nr2f* intron/exon structure [40–42]. Intron/exon junctions from early-branching metazoans and vertebrates matching the transcripts and the translated proteins were mapped and given a score for the intron phases (S1 File), with 0, 1 and 2 introns falling before the first, second and third bases of a codon, respectively. Then, we mapped the introns on a

protein alignment comprising the highly conserved Nr2f DBDs and LBDs of Nr2f proteins (S1 File). We found that two “phase 1” introns (one before and one within the LBD) are preserved in all the extant *Nr2f* subfamilies (Fig 4A; S1 File). However, *Nr2f6* genes exhibited a “phase 2” intron inside the second zinc-finger belonging to the nuclear receptor DNA-binding site (Fig 4A; S1 File). The conservation of intron/exon junctions in the examined *Nr2f* genes allows two groups to be distinguished: one constituted by *Nr2f1/2/5/6*, *Nr2f1/2/5*, *Nr2f1/2*, *Nr2f1*, *Nr2f2*, *Nr2f5*, and one comprising only *Nr2f6* (Fig 4B). Thus, our analysis of intron/exon boundaries demonstrates the existence of a highly conserved intron code throughout *Nr2f* family members that also robustly supports the divergence of *Nr2f6* genes.

Fig 4. Intron code of the *Nr2f* family in metazoans.

(A) Protein alignment showing conservation of intron/exon structures in Nr2f members, with a portion comprising the second DBD (underlined) and LBD (red). Asterisks indicate 100% amino acid conservation. Colons indicate high levels (>90%) amino acid conservation. Periods indicate moderate levels (50-89%) of amino acid conservation. Phase1-introns (green) and the *Nr2f6*-specific Phase 2-intron (turquoise). (B) Schematization of distinct intron/exon structures of *Nr2f1/2/5/6*, *Nr2f1/2/5*, *Nr2f1/2*, *Nr2f1*, *Nr2f2*, *Nr2f5* genes and of *Nr2f6*. Black boxes represent the zinc-fingers motifs within the DBD, while pink bars indicate the LBD. Small vertical bars were used to portray the ultra-conserved Phase1-introns and the *Nr2f6*-specific Phase 2-intron.

Synteny analysis defines differential duplications and losses in the *Nr2f* family

In order to confirm the specific homologies indicated from the phylogenetic analysis, we next carried out an extensive examination of synteny within the *Nr2f* genomic environments. With respect to the more ancient *Nr2f1/2/5/6* genes, we did not find evidence of synteny between placozoans and cnidarians. However, the location of the three *N. vectensis* genes indicates they were likely derived from an ancient duplication event followed by a tandem duplication (S2 Fig). With respect to *Nr2f1/2/5* genes in invertebrates, we only found preservation of the *Nr2f1/2/5* loci between two slow-evolving deuterostomes: the amphioxus (*B. belcheri*) [43] and the hemichordate (*S. kowalevskii*) [44] (S2 Fig), corroborating the existence of the *Nr2f1/2/5* cluster shown in the phylogenetic trees (Fig 2; S1 Fig). The only remaining synteny between *Nr2f1/2/5* invertebrates and vertebrate orthologs appears to be the linkage between *UNC45A-NR2F2* of primates and *Unc45a-Nr2f1/2/5* of the tunicate *C. robusta* (S3 Fig), which is considered the closest living relative of vertebrates [45]. Focusing on *Nr2f1* and *Nr2f2* in the genomes of gnathostomes, including cartilaginous fish, coelacanth, spotted gar, zebrafish, chicken, and human, we found a high degree of synteny for *Nr2f1* and *Nr2f2* loci and conservation of the location of flanking genes among these taxa (Fig 5). Specifically, *Nr2f1* and *Nr2f2* genes exhibited a remarkably conserved syntenic environment, clustering with putative orthologs belonging to other families. In fact, *Lysmd3*, *Arrdc3*, *Mctp1* and *Mef2a* flank *Nr2f1* orthologs, while *Nr2f2* orthologs are flanked by *Lysmd4*, *Arrdc4*, *Mctp2* and *Mef2c* paralogs. Furthermore, in teleosts like zebrafish, two *Nr2f1* Ohnologs (*nr2f1a* and *nr2f1b*) also shared significant conservation of paralogous genes, which is consistent with an origin from the teleost-specific genome duplication (TSGD) [15,16]. However, the *nr2f1b* gene has been lost by several teleost species (Fig 2; S1 Fig). Although somewhat fragmented, orthologs of flanking genes found in gnathostome *Nr2f1* and *Nr2f2*, such as *Arrdc2/3*, *Lysmd3*, *Fam172a*, were also found near each of the three lamprey *Nr2f1/2* genes (S4 Fig), which is consistent with these genes arising from genome duplication(s) within the agnathan lineage [46].

Together, these results suggest that *Nr2f1* and *Nr2f2* of gnathostomes have a common origin and are derived from a WGD event [47,48].

Fig 5. Synteny analysis of *Nr2f1* and *Nr2f2* genes.

Schematization of conserved genome environments of gnathostome *Nr2f1* and *Nr2f2* genes (red rectangles) in selected species with relative chromosomes/scaffolds. Flanking orthologous genes are represented employing rectangles of the same color. Black arrows indicate transcription orientation.

Examining *Nr2f5* loci in the same representative gnathostomes showed a high degree of conservation in species that have retained the gene. However, the synteny is not shared with gnathostome *Nr2f1* and *Nr2f2* orthologs (Fig 6). The *Nr2f5* loci in coelacanth and amphibians have retained *Txnip* (Fig 6), which is also named *Arrdc6*. *Arrdc* family members flank both *Nr2f1* and *Nr2f2* genes (Fig 5). Interestingly, despite having lost *Nr2f5*, amniotes have largely preserved the flanking genomic loci present in fish in amphibians (Fig 6). In contrast, the absence of *Nr2f5* in cartilaginous fish, such as *C. milii*, correlates with the lack of the entire locus. Within the *Actinopterygii* (ray-finned fishes), the synteny of genes has been lost only on one side of the *Nr2f5* loci. With respect to the lamprey, its *Nr2f1/2C* ortholog is flanked by a *Bola1* ortholog as well as orthologs of gnathostome *Nr2f1* and *Nr2f2* (S4 Fig; Fig 5; Fig 6), which suggests ancestral linkage with the single *Nr2f1/2/5* genes (Fig 2; S1 Fig). As might be expected, the *Nr2f6* subfamily did not share common elements with the other *Nr2f* loci in gnathostomes (Fig 7). However, within gnathostomes the *Nr2f6* loci were highly conserved from cartilaginous fish to mammals, although there were significant gene losses surrounding *Nr2f6* loci in zebrafish and human. Furthermore, the presence of conserved orthologs (*ano8a* and *ano8b*, *plvapa* and *plvapb*) flanking *nr2f6a* and *nr2f6b* zebrafish genes suggested that they

originated from the TSGD. Therefore, these findings show that despite the divergence of the *Nr2f5* and *Nr2f6* within vertebrates the genomic environments surrounding *Nr2f5* and *Nr2f6* loci are highly conserved within gnathostomes.

Fig 6. Synteny analysis of *Nr2f5* genes.

Schematization of conserved genome environments of gnathostome *Nr2f5* genes (red rectangles) in selected species with relative chromosomes/scaffolds. Flanking orthologous genes are represented using rectangles of the same color. Black arrows indicate transcription orientation.

Fig 7. Synteny analysis of *Nr2f6* genes.

Schematization of conserved genomic environments of gnathostome *Nr2f5* genes (red rectangles) in selected species with relative chromosomes/scaffolds. Flanking orthologous genes are represented using the same color code. Black arrows indicate transcription orientation.

Effects of TSGD on the *Nr2f* gene repertoire

We next wanted to measure the impact of the series of additional WGDs that have occurred in teleosts on *Nr2f* gene number (Fig 2; S1 Fig). For this comparison, we examined all the *Nr2f* loci in zebrafish, the Asian arowana (*S. formosus*), which is documented to retain duplicates [49], and the Atlantic salmon (*S. salar*), which has a salmonid-specific genome duplication (SSGD) [39]. We found that each of these teleosts retained two *Nr2f1* Ohnologs (Fig 8), suggesting they either were not duplicated or that one pair of Ohnologs was lost in salmonids. Zebrafish lost one *nr2f2* Ohnolog, maintaining only the *nr2f2a* ortholog, while Asian arowana retained two *Nr2f2* Ohnologs. Salmonids have 3 *Nr2f2* genes (*Nr2f2a1*,

Nr2f2a2, and *Nr2f2b1*), due to a loss of the one of the *Nr2f2b* Ohnologs following their additional WGD. With respect to *Nr2f5*, only Atlantic salmon showed two copies, implying these were generated during the SSGD event, as suggested by the presence of two *Nr2f5* Ohnologs in other salmonids (*Oncorhynchus spp.*). Finally, zebrafish and Asian arowana each possess two *Nr2f6* genes, while Atlantic salmon has 3 similar to what is found in the *Nr2f2* subfamily (Fig 8). Inspecting other teleost *Nr2f* gene family repertoires (Fig 2; S1 Fig), we found that the Channel catfish (*Ictalurus punctatus*), Red-bellied piranha (*Pygocentrus nattereri*), cavefish (*Astyanax mexicanus*) and Sheepshead minnow (*Cyprinodon variegatus*) all retained only *Nr2f6b*. The Sheepshead minnow and Princess cichlid (*Neolamprologus brichardi*) also lost *Nr2f5*. However, other cichlids like Blue tilapia (*Oreochromis aureus*) and Zebra mbuna (*Maylandia zebra*) did not lose *Nr2f5* (data not shown). Intriguingly, the Monterrey platifish (*Xiphophorus couchianus*) is the only gnathostome without any *Nr2f1* paralogs, differing from its sibling species, the common platifish (*X. maculatus*), which possesses *Nr2f1a*. Therefore, teleosts show an expansion of *Nr2f* genes following TSGD and SSGD, which were followed by high variability in species-specific losses of *Nr2f* Ohnologs.

Fig 8. Synteny analysis of *Nr2f* genes in teleosts.

Comparison of *Nr2f* genome environments in selected teleosts (zebrafish, Asian arowana, Atlantic salmon) with relative chromosomes/scaffolds. Rectangles of the same color represent flanking orthologous genes. Black arrows indicate transcription orientation.

Discussion

We have performed an extensive examination of *Nr2f* gene evolution in metazoans. Our analysis corroborates previous work showing that *Nr2f* genes are present in some early-

branching metazoans (cnidarians, placozoans) [13,50], but that they are absent in other early-branching metazoans, i.e. sponges and ctenophores. Importantly, our data support a model in which a single *Nr2f* gene (*Nr2f1/2/5/6*), which is present in extant cnidarians and placozoans, predated the six Bilateria subfamilies that include *Nr2f1/2/5* (found in invertebrates), *Nr2f1/5* (found in agnathans), and *Nr2f1*, *Nr2f2*, *Nr2f5*, and *Nr2f6* (found in vertebrates; Fig 9). Single, conserved *Nr2f1/2/5* genes are predominantly found throughout invertebrate protostomes and deuterostomes and have even been retained in species traditionally considered gene losers, such as the tunicates [42,51,52]. There has been significant expansion and retention of *Nr2fs* in gnathostomes, particularly in teleosts. Although initial analysis in lampreys suggested they may possess only one *Nr2f* gene [53], our evolutionary assessment shows that the agnathans have three *Nr2f* members, which appear to have originated in part from an agnathan WGD event [46]. Interestingly, the single *Nr2f1/2/5* proteins in invertebrates are also highly conserved at the sequence level and cluster with the *Nr2f1/2* proteins in agnathans and *Nr2f1* and *Nr2f2* proteins in gnathostomes. Within gnathostomes, *Nr2f1* and *Nr2f2* orthologs have retained significant synteny of their loci, consistent with similar requirements shown across vertebrate models [14,18,19]. According to our data, *Nr2f1*, *Nr2f2*, and *Nr2f5* descended from the same ancestor, as limited synteny exists between them with members of the *Arrdc* family. However, *Nr2f6* orthologs do not share syntenic genomic loci with *Nr2f1*, *Nr2f2*, and *Nr2f5* genes. They have only retained synteny with their respective orthologs and have a distinct intron/exon structure. Together with the *Nr2f6* proteins being the most divergent *Nr2f* with respect to other gnathostome *Nr2fs*, these data suggest the *Nr2f6* origin predated the gnathostome radiation.

Fig 9. Evolutionary model for the *Nr2f* family in Metazoa.

Nr2f1/2/5/6 of placozoans and cnidarians (violet) is the ancestor of all other extant *Nr2fs*. An ancient duplicative event (green circle) generated *Nr2f1/2/5* (light blue) that is present from protostomes to invertebrate chordates and *Nr2f6* (orange), which has been retained only by gnathostomes and lost in invertebrates. WGDs in vertebrates (red circle) originated *Nr2f1/2* (yellow), *Nr2f1* (red), *Nr2f2* (green) and *Nr2f5* (blue), which has been lost in cartilaginous fish and amniotes. All these subfamilies show extra-duplicates: agnathan-specific (magenta circle) TSGDs (blue circle) and SSGDs (yellow circle).

Integrating our phylogenetic assessment with current expression and functional analyses of the *Nr2f1/2/5/6* members in evolutionary distant animals [12,54] supports a hypothesis that regulation of neuronal differentiation is the most ancient *Nr2f* function and that *Nr2fs* have gained additional functions in the development of mesodermal and endodermal-derived tissues. The *Nr2f* member (*Nr2f1/2/5/6*) of the diploblastic cnidarian *Hydra* only appears to be expressed in neurons and has a requirement in neurogenesis [12,54]. Furthermore, the function of *Nr2f1/2/5* orthologs of protostome nematodes and flies have been extensively studied in neural sensory cell differentiation [8,55], [56]. In invertebrate deuterostomes, the single *Nr2f1/2/5* orthologs are expressed in neural tissue of sea urchin (*Strongylocentrotus purpuratus*), amphioxus, and sea squirt embryos [11,57,58], [59]. Interestingly, recent functional analysis of the Mediterranean sea urchin (*Paracentrotus lividus*) *Nr2f1/2/5* shows that it is maternally inherited and required for the development of neural and ectodermal derivatives [60]. A *Nr2f1/2/5* ortholog from the agnathan river lamprey (*Lampetra fluviatilis*) is also expressed in the developing nervous system [53]. However, our identification of three *Nr2f1/2* members in agnathans suggests that additional expression and potentially functional analysis should be performed in the sea lamprey (*Petromyzon marinus*) and/or hagfish (*Eptatretus burgeri*) to understand the conservation of the different agnathan paralogs

compared to *Nr2fs* in vertebrates. In addition to the conservation of *Nr2f1/2/5/6*, *Nr2f1/2/5*, and *Nr2f1/2* expression in neural tissues from cnidarians to agnathans, *Nr2f* genes have also evolved critical new expression and functionalities, such as copulation control in nematodes [61] or heart vessels specification in flies [20,21]. The recent work with the Mediterranean sea urchin suggests that it is required for the development of mesendodermal derivative [60], supporting that ancestral *Nr2fs* also acquired requirements in all 3 germ layers within the invertebrate lineage. However, the conservation of these requirements in the different germ layers throughout different metazoan phyla is not yet as clear, with present analyses suggesting functions in mesendodermal tissues are newly derived functions.

Similar to the expression in invertebrates, all the vertebrate *Nr2f* genes are expressed in neural tissue during development. Furthermore, it is interesting to consider the conservation of expression and functions of *Nr2f1* and *Nr2f2* genes in vertebrates compared to the analysis of the invertebrate *Nr2f* genes, since the functions of *Nr2f1* and *Nr2f2* have been intensely investigated in vertebrates and they are required for proper human development [19,28,62]. Our evolutionary analysis strongly supports paralogy between *Nr2f1* and *Nr2f2* of gnathostomes that is derived from one of the two rounds of WGD events in the vertebrate lineage [47,48]. Consistent with the common origin for *Nr2f1* and *Nr2f2* genes in vertebrates, they share overlapping expression patterns and functions, but also appear to have acquired distinct developmental roles during evolution. Both *Nr2f1* and *Nr2f2* orthologs share overlapping central nervous system (CNS) expression in mouse and zebrafish [14,18,63], further supporting the conserved expression of *Nr2f* genes throughout neural tissue of metazoans. However, *nr2f1a* and *nr2f2* are both expressed more extensively in neural tissue of zebrafish embryos, following overlapping expression in early mouse embryos *Nr2f1* and *Nr2f2* become predominantly expressed in neural and mesendodermal tissues, respectively. Functional analysis of these *Nr2f* genes in mice and zebrafish support the functional divergence

of these proteins. Murine *Nr2f1* KOs have glial differentiation defects [64], while *Nr2f2* is required for proper development of many mesendodermal-derived tissues, including atrial chamber and arterial-venous differentiation [37,65]. Intriguingly, mouse *Nr2f2* and zebrafish *nr2f1a* are functional homologs with respect to heart development, as both are required for atrial differentiation [66], further supporting the common evolutionary origins of these paralogs. While zebrafish *nr2f2* is not required for early atrial or vein development [26], NR2F1 and NR2F2 proteins do appear to have redundant requirements, for instance promoting atrial cardiomyocytes differentiation in human embryonic stem cells [67,68]. It is interesting that the single *Nr2f1/2/5* homolog of flies (*svp*) is also required for dorsal vessel (heart) development [20]. However, if these similar roles in heart tissues reflect conserved requirements within metazoans for cardiac differentiation requires functional studies from many additional model organisms, though the analysis in sea urchins hints at conservation of more ancestral requirements in all germ layers [60]. With respect to the expansion of *Nr2f1* and *Nr2f2* in teleosts, *Nr2f1b* has been lost in the majority of surveyed teleosts. *Nr2f1b* zebrafish mutants are viable [69] and do not exhibit redundancy with *nr2f1a* in atrial cardiomyocyte differentiation [26], but do exhibit some redundancy with multiple other *Nr2f* genes in neural crest cells that promote jaw development [69]. Virtually all the analyzed gnathostome genomes have a single *Nr2f2* gene, excluding the teleosts *S. formosus* (2) and *S. salar* (3), implying there may be some dosage sensitivity that favors the retention of single orthologs in vertebrates. Otherwise, the presence of additional duplicates in these species could suggest new functionalities.

With respect to *Nr2f5* orthologs, our analysis shows they are the smallest *Nr2f* subfamily in vertebrates. It has been maintained in coelacanth, ray-finned fish, and amphibians, but independently lost in cartilaginous fishes and mammals (Fig 9). Despite the loss of *Nr2f5*, the synteny of organisms that both have and have not retained *Nr2f5* suggests a common

evolutionary origin with *Nr2f1* and *Nr2f2* paralogs. Similar to invertebrate *Nr2f* genes, *Nr2f1* and *Nr2f2* orthologs, *Nr2f5* is expressed in neural tissue [63], including in the eyes of zebrafish and newts [70,71]. Like *nr2f1b*, zebrafish *nr2f5* mutants are viable, yet function redundantly with other *nr2f* genes for a proper upper-jaw development [69]. While expression and functional analysis from other organisms that have retained *Nr2f5* (coelacanth, spotted gar, and frog) may provide insights into conserved functions of *Nr2f5* orthologs, the loss of *Nr2f5* lineage in cartilaginous fish and amniotes, as well as the lack of overt requirements alone in zebrafish, suggests that *Nr2f5* orthologs likely have retained singular developmental requirements within the vertebrate lineage.

In contrast to *Nr2f5*, the divergent *Nr2f6* subfamily has been retained by all the evaluated gnathostomes and exhibit high sequence conservation, suggesting pressure to maintain these genes in vertebrates. Intriguingly, the high degree of divergence of the *Nr2f6* subfamily evokes a pre-gnathostome origin for this gene, as has been proposed from more limited analysis [72]. Despite the divergence to other *Nr2f* subfamilies, *Nr2f6* genes have conserved expression within the central nervous system of mammals [14], as well as both zebrafish *nr2f6* orthologs. Murine *Nr2f6* KO mice have forebrain defects. Specifically, a loss of neurons that regulate the circadian clock genes [73]. *Nr2f6* also has a critical role in lymphocyte differentiation and T-cell mediated tumor surveillance, suggesting derived functions in adaptive immunity [74,75]. Thus, our data support a pre-gnathostome origin for the *Nr2f6* subfamily during metazoan evolution and that it has been selectively retained in gnathostomes.

In examining the evolution of the *Nr2f* transcription factors, it is also worthwhile to note that there is conserved responsiveness to retinoic acid (RA) signaling that exists in early-branching metazoans through invertebrate chordates and gnathostomes, implying this relationship may form the core of an ancient gene regulatory network. *Nr2f* genes from

placozoans [50] and the invertebrate chordates *Ciona* and amphioxus are all RA-responsive [11,57]. Furthermore, in vertebrates, where the earliest requirement for RA is posteriorization of the embryo [76], virtually all the *Nr2f* genes have been shown to be responsive to RA signaling in developmental contexts involving all three germ layers. Specifically, RA signaling has been shown to positively regulate all the *Nr2fs* in zebrafish in developing zebrafish endoderm [77], the CNS [63], and anterior lateral plate mesoderm (ALPM) [26]. RA signaling also positively regulates *Nr2f1*, *Nr2f2* and *Nr2f6* in mice [78,79], and *NR2F1* and *NR2F2* in humans [80,81]. *Nr2fs* can inhibit RA signaling in some contexts, suggesting it may form a negative feedback loop. One role *Nr2fs* may play is through direct competition with retinoic acid receptors (RARs) in binding retinoic acid response elements (RAREs) [18]. Moreover, it has been shown that the cnidarian *Nr2f1/2/5/6* possesses the ability to inhibit RA signaling in *in vitro* signaling assays [12]. Thus, the responsiveness of the *Nr2f* family to RA may have evolved very early and has been highly maintained through the diversification of multiple vertebrate *Nr2f* genes, implying there is high selection to maintain this relationship.

Conclusions

Overall, our evolutionary assessment sheds new light on the events that have shaped the extant *Nr2f* family in Metazoa. The phylogenetic analysis clearly defines the individual *Nr2f* subfamilies and their relationships across metazoan phyla, which complements available expression and functional data that together support an origin of their requirements in the development of neural tissue. Interestingly, the functions of *Nr2f* proteins are found to regulate development of virtually all germ layers of invertebrates and vertebrates. The detailed evolutionary understanding of the *Nr2f* gene family we now have will allow us to infer more meaningful conclusions about the origins and conserved requirements of *Nr2f* genes in normal metazoan development and their role in sculpting diverse body plans.

Methods

Ethics Statement

Ethical approval is not required. No animals were used in this study.

Genome database searches and phylogenetic reconstruction

Homo sapiens NR2F protein sequences were employed as queries in BLASTp and tBLASTn in genome databases of selected species (NCBI, Ensembl, Ensembl Metazoa, SkateBase [82], ANISEED [83]). The entire dataset of protein sequences for domain architecture was analyzed by using the domain database provided by Expasy, named PROSITE [84] and then, manually annotated. All the surveyed sequences were verified to be Nr2f proteins through analysis of DBDs and LBDs (S2 File). The analysis was weighted with 30 species from agnathans to primates to take into account the impact of multiple WGDs in vertebrates [47,48] and in teleosts [15,16]. Orthology of the *Nr2f* members was initially assessed by using a reciprocal best blast hit (RBBH) approach employing default parameters and corroborated by phylogenetic analyses. Protein alignment for phylogeny was generated using L-INS-i (accurate; for alignment of <200 sequences) on MAFFT [85,86] (S3 File). The phylogenetic reconstruction of Fig 2 was performed on the entire protein sequences and based on maximum-likelihood (ML) inferences calculated with PhyML 3.0 [87], employing automatic Akaike Information Criterion (AIC) by Smart Model Substitution (SMS) [88], which selected the JTT+G+F model employing discrete gamma distribution in categories. All parameters (gamma shape = 0.692; proportion of invariants (fixed) = 0.000) were established from the dataset. Branch support was provided by aLRT [89]. The phylogeny of S1 Fig was carried out employing Bayesian Information Criterion (BIC) by SMS, which sorted the JTT+G+F model using discrete gamma

distribution in categories. All parameters (gamma shape = 0.690; proportion of invariants (fixed) = 0.000) were established from the dataset, with branch support calculated employing aBayes method [90]. Accession numbers and protein sequences used for phylogenetic tree reconstructions are provided in S4 File, while those excluded for their divergence are listed in S5 File. Common and Latin names for species used in this study are listed in S6 File.

Analysis of intron/exon structures and phases

Gene structures were deduced after merging the genomic sequences with ESTs when available, as previously described [40–42]. Introns were classified as phase 0, phase 1, and phase 2, according to their positions with respect to the protein-reading frame. The amino-acid residues with the conserved introns were manually mapped on a ClustalX alignment [91] of selected Nr2f proteins (S7 File).

Evaluation of synteny

We evaluated the presence/absence of synteny examining the chromosomes on public genome databases (NCBI, Ensembl, Ensembl Metazoa, ANISEED [83]). We verified the existence of duplicates using Genomicus [92] and Vertebrate Ohnologs [93]. The window considered for the locus analyses was twenty flanking genes. Genes that were not conserved were excluded from the analysis. All the genes were represented employing colored rectangles, using the same color for all *Nr2f* genes (red).

Acknowledgements

Not applicable

492

493 **Availability of Data and Materials**

494 All the data used for the analysis in this article are available within the article and in its
495 online supplementary material.

496

497 **Competing interests**

498 The authors declare they have no competing interests.

499

500 **Funding**

501 Work in the manuscript was supported by National Institutes of Health grants R01 HL141186
502 and R01 HL137766 to JSW.

503

504 **Authors' contribution**

505 UC performed the analyses, prepared the figures and wrote the initial manuscript. JSW
506 provided feedback for the analyses, revised figures, and wrote the manuscript.

507

508

509 **References**

- 510 1. Laudet V. Evolution of the nuclear receptor superfamily: early diversification from an ancestral
511 orphan receptor. J Mol Endocrinol. 1997;19: 207–226.
- 512 2. Escriva H, Safi R, Hänni C, Langlois MC, Saumitou-Laprade P, Stehelin D, et al. Ligand binding
513 was acquired during evolution of nuclear receptors. Proc Natl Acad Sci U S A. 1997;94: 6803–
514 6808.

- 515 3. Escriva H, Bertrand S, Laudet V. The evolution of the nuclear receptor superfamily. Essays in
516 Biochemistry. 2004. pp. 11–26. doi:10.1042/bse0400011
- 517 4. Mangelsdorf DJ, Evans RM. The RXR heterodimers and orphan receptors. Cell. 1995. pp. 841–
518 850. doi:10.1016/0092-8674(95)90200-7
- 519 5. Pastorcic M, Wang H, Elbrecht A, Tsai SY, Tsai MJ, O'Malley BW. Control of transcription
520 initiation in vitro requires binding of a transcription factor to the distal promoter of the
521 ovalbumin gene. Molecular and Cellular Biology. 1986. pp. 2784–2791.
522 doi:10.1128/mcb.6.8.2784
- 523 6. Sagami I, Tsai SY, Wang H, Tsai MJ, O'Malley BW. Identification of two factors required for
524 transcription of the ovalbumin gene. Mol Cell Biol. 1986;6: 4259–4267.
- 525 7. Wang LH, Tsai SY, Cook RG, Beattie WG, Tsai MJ, O'Malley BW. COUP transcription factor
526 is a member of the steroid receptor superfamily. Nature. 1989;340: 163–166.
- 527 8. Mlodzik M, Hiromi Y, Weber U, Goodman CS, Rubin GM. The Drosophila seven-up gene, a
528 member of the steroid receptor gene superfamily, controls photoreceptor cell fates. Cell.
529 1990;60: 211–224.
- 530 9. Chan SM, Xu N, Niemeyer CC, Bone JR, Flytzanis CN. SpCOUP-TF: a sea urchin member of
531 the steroid/thyroid hormone receptor family. Proc Natl Acad Sci U S A. 1992;89: 10568–10572.
- 532 10. Langlois MC, Vanacker JM, Holland ND, Escriva H, Queva C, Laudet V, et al. Amphicoup-TF,
533 a nuclear orphan receptor of the lancelet Branchiostoma floridae, is implicated in retinoic acid
534 signalling pathways. Dev Genes Evol. 2000;210: 471–482.
- 535 11. Ishibashi T, Usami T, Fujie M, Azumi K, Satoh N, Fujiwara S. Oligonucleotide-based
536 microarray analysis of retinoic acid target genes in the protochordate, Ciona intestinalis. Dev
537 Dyn. 2005;233: 1571–1578.

12. Gauchat D, Escriva H, Miljkovic-Licina M, Chera S, Langlois M-C, Begue A, et al. The orphan COUP-TF nuclear receptors are markers for neurogenesis from cnidarians to vertebrates. *Dev Biol.* 2004;275: 104–123.
13. Reitzel AM, Tarrant AM. Nuclear receptor complement of the cnidarian *Nematostella vectensis*: phylogenetic relationships and developmental expression patterns. *BMC Evolutionary Biology.* 2009. p. 230. doi:10.1186/1471-2148-9-230
14. Bertrand S, Thisse B, Tavares R, Sachs L, Chaumot A, Bardet P-L, et al. Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. *PLoS Genet.* 2007;3: e188.
15. Taylor JS. Genome Duplication, a Trait Shared by 22,000 Species of Ray-Finned Fish. *Genome Research.* 2003. pp. 382–390. doi:10.1101/gr.640303
16. Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature.* 2004;431: 946–957.
17. Germain P, Staels B, Dacquet C, Spedding M, Laudet V. Overview of nomenclature of nuclear receptors. *Pharmacol Rev.* 2006;58: 685–704.
18. Tsai SY, Tsai MJ. Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): coming of age. *Endocr Rev.* 1997;18: 229–240.
19. Pereira FA, Tsai MJ, Tsai SY. COUP-TF orphan nuclear receptors in development and differentiation. *Cell Mol Life Sci.* 2000;57: 1388–1398.
20. Lo PC, Frasch M. A role for the COUP-TF-related gene seven-up in the diversification of cardioblast identities in the dorsal vessel of *Drosophila*. *Mech Dev.* 2001;104: 49–60.
21. Ryan KM, Hendren JD, Helander LA, Cripps RM. The NK homeodomain transcription factor

- 561 Tinman is a direct activator of seven-up in the *Drosophila* dorsal vessel. *Dev Biol.* 2007;302:
562 694–702.
- 563 22. Qiu Y, Tsai SY, Tsai M-J. COUP-TF an orphan member of the steroid/thyroid hormone receptor
564 superfamily. *Trends in Endocrinology & Metabolism.* 1994. pp. 234–239. doi:10.1016/1043-
565 2760(94)p3081-h
- 566 23. Pereira FA, Qiu Y, Tsai MJ, Tsai SY. Chicken ovalbumin upstream promoter transcription factor
567 (COUP-TF): expression during mouse embryogenesis. *J Steroid Biochem Mol Biol.* 1995;53:
568 503–508.
- 569 24. Pereira FA, Qiu Y, Zhou G, Tsai M-J, Tsai SY. The orphan nuclear receptor COUP-TFII is
570 required for angiogenesis and heart development. *Genes & Development.* 1999. pp. 1037–1049.
571 doi:10.1101/gad.13.8.1037
- 572 25. Xie X, Qin J, Lin S-H, Tsai SY, Tsai M-J. Nuclear receptor chicken ovalbumin upstream
573 promoter-transcription factor II (COUP-TFII) modulates mesenchymal cell commitment and
574 differentiation. *Proc Natl Acad Sci U S A.* 2011;108: 14843–14848.
- 575 26. Dohn TE, Ravisankar P, Tirera FT, Martin KE, Gafranek JT, Duong TB, et al. Nr2f-dependent
576 allocation of ventricular cardiomyocyte and pharyngeal muscle progenitors. *PLoS Genet.*
577 2019;15: e1007962.
- 578 27. Kliewer SA, Umesono K, Noonan DJ, Heyman RA, Evans RM. Convergence of 9-cis retinoic
579 acid and peroxisome proliferator signalling pathways through heterodimer formation of their
580 receptors. *Nature.* 1992;358: 771–774.
- 581 28. Tang K, Tsai SY, Tsai M-J. COUP-TFs and eye development. *Biochim Biophys Acta.*
582 2015;1849: 201–209.
- 583 29. Cooney AJ, Tsai SY, O'Malley BW. Chicken ovalbumin upstream promoter transcription factor
584 (COUP-TF) dimers bind to different GGTCa response elements, allowing COUP-TF to repress

hormonal Molecular and cellular. 1992. Available:

<https://mcb.asm.org/content/12/9/4153.short>

30. Schröder M, Danielsson C, Wiesenberg I, Carlberg C. Identification of Natural Monomeric Response Elements of the Nuclear Receptor RZR/ROR. THEY ALSO BIND COUP-TF HOMODIMERS. Journal of Biological Chemistry. 1996. pp. 19732–19736. doi:10.1074/jbc.271.33.19732
31. Pipaón C, Tsai SY, Tsai MJ. COUP-TF upregulates NGFI-A gene expression through an Sp1 binding site. Mol Cell Biol. 1999;19: 2734–2745.
32. Tran P, Zhang XK, Salbert G, Hermann T, Lehmann JM, Pfahl M. COUP orphan receptors are negative regulators of retinoic acid response pathways. Mol Cell Biol. 1992;12: 4666–4676.
33. Tang LS, Alger HM, Pereira FA. COUP-TFI controls Notch regulation of hair cell and support cell differentiation. Development. 2006;133: 3683–3693.
34. Montemayor C, Montemayor OA, Ridgeway A, Lin F, Wheeler DA, Pletcher SD, et al. Genome-wide analysis of binding sites and direct target genes of the orphan nuclear receptor NR2F1/COUP-TFI. PLoS One. 2010;5: e8910.
35. Chiang DY, Cuthbertson DW, Ruiz FR, Li N, Pereira FA. A coregulatory network of NR2F1 and microRNA-140. PLoS One. 2013;8: e83358.
36. Okamura M, Kudo H, Wakabayashi K-I, Tanaka T, Nonaka A, Uchida A, et al. COUP-TFII acts downstream of Wnt/ -catenin signal to silence PPAR gene expression and repress adipogenesis. Proceedings of the National Academy of Sciences. 2009. pp. 5819–5824. doi:10.1073/pnas.0901676106
37. Wu S-P, Cheng C-M, Lanz RB, Wang T, Respress JL, Ather S, et al. Atrial identity is determined by a COUP-TFII regulatory network. Dev Cell. 2013;25: 417–426.

38. van der Wees J, Matharu PJ, de Roos K, Destre'e OHJ, Godsave SF, Durston AJ, et al. Developmental expression and differential regulation by retinoic acid of *Xenopus* COUP-TF-A and COUP-TF-B. *Mech Dev.* 1996;54: 173–184.
39. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon genome provides insights into rediploidization. *Nature.* 2016;533: 200–205.
40. D'Aniello S, Irimia M, Maeso I, Pascual-Anaya J, Jiménez-Delgado S, Bertrand S, et al. Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. *Mol Biol Evol.* 2008;25: 1841–1854.
41. Roy SW, Irimia M. When good transcripts go bad: artifactual RT-PCR “splicing” and genome analysis. *Bioessays.* 2008;30: 601–605.
42. Coppola U, Ristoratore F, Albalat R, D'Aniello S. The evolutionary landscape of the Rab family in chordates. *Cell Mol Life Sci.* 2019;76: 4117–4130.
43. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature.* 2008;453: 1064–1071.
44. Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, et al. Hemichordate genomes and deuterostome origins. *Nature.* 2015;527: 459–465.
45. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature.* 2006;439: 965–968.
46. Simakov O, Marlétaz F, Yue J-X, O'Connell B, Jenkins J, Brandt A, et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol.* 2020;4: 820–830.
47. Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. Evidence of en bloc duplication in vertebrate genomes. *Nat Genet.* 2002;31: 100–105.
48. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS*

631 Biol. 2005;3: e314.

632 49. Li J, Bian C, Hu Y, Mu X, Shen X, Ravi V, et al. A chromosome-level genome assembly of the
633 Asian arowana, *Scleropages formosus*. Scientific Data. 2016. doi:10.1038/sdata.2016.105

634 50. Novotný JP, Chughtai AA, Kostrouchová M, Kostrouchová V, Kostrouch D, Kaššák F, et al.
635 *Trichoplax adhaerens* reveals a network of nuclear receptors sensitive to 9-cis-retinoic acid at the
636 base of metazoan evolution. PeerJ. 2017. p. e3789. doi:10.7717/peerj.3789

637 51. Martí-Solans J, Belyaeva OV, Torres-Aguila NP, Kedishvili NY, Albalat R, Cañestro C.
638 Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a
639 Chordate. Mol Biol Evol. 2016;33: 2401–2416.

640 52. Coppola U, Kamal AK, Stolfi A, Ristoratore F. The Cis-Regulatory Code for Kelch-like 21/30
641 Specific Expression in *Ciona robusta* Sensory Organs. Front Cell Dev Biol. 2020;8: 569601.

642 53. Guérin A, d'Aubenton-Carafa Y, Marrakchi E, Da Silva C, Wincker P, Mazan S, et al.
643 Neurodevelopment genes in lampreys reveal trends for forebrain evolution in craniates. PLoS
644 One. 2009;4: e5374.

645 54. Duffy DJ, Frank U. Modulation of COUP-TF expression in a cnidarian by ectopic Wnt signalling
646 and allorecognition. PLoS One. 2011;6: e19443.

647 55. Kanai MI, Okabe M, Hiromi Y. seven-up Controls Switching of Transcription Factors that
648 Specify Temporal Identities of *Drosophila* Neuroblasts. Developmental Cell. 2005. pp. 203–213.
649 doi:10.1016/j.devcel.2004.12.014

650 56. Zhou HM, Walthall WW. UNC-55, an orphan nuclear hormone receptor, orchestrates synaptic
651 specificity among two classes of motor neurons in *Caenorhabditis elegans*. J Neurosci. 1998;18:
652 10438–10444.

653 57. Escriva H, Holland ND, Gronemeyer H, Laudet V, Holland LZ. The retinoic acid signaling

654 pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a
655 chordate lacking neural crest. *Development*. 2002;129: 2905–2916.

656 58. Devine C, Hinman VF, Degnan BM. Evolution and developmental expression of nuclear
657 receptor genes in the ascidian *Herdmania*. *Int J Dev Biol*. 2002;46: 687–692.

658 59. Kalampoki LG, Flytzanis CN. Cis-regulatory control of the nuclear receptor Coup-TF gene in
659 the sea urchin *Paracentrotus lividus* embryo. *PLoS One*. 2014;9: e109274.

660 60. Tsironis I, Paganos P, Gouvi G, Tsimpos P, Stamopoulou A, Arnone MI, et al. Coup-TF: A
661 maternal factor essential for differentiation along the embryonic axes in the sea urchin
662 *Paracentrotus lividus*. *Dev Biol*. 2021. doi:10.1016/j.ydbio.2020.12.012

663 61. Shan G, Walthall WW. Copulation in *C. elegans* males requires a nuclear hormone receptor. *Dev*
664 *Biol*. 2008;322: 11–20.

665 62. Yang X, Feng S, Tang K. COUP-TF Genes, Human Diseases, and the Development of the
666 Central Nervous System in Murine Models. *Curr Top Dev Biol*. 2017;125: 275–301.

667 63. Love CE, Prince VE. Expression and retinoic acid regulation of the zebrafish nr2f orphan nuclear
668 receptor genes. *Dev Dyn*. 2012;241: 1603–1615.

669 64. Yamaguchi H, Zhou C, Lin S-C, Durand B, Tsai SY, Tsai M-J. The nuclear orphan receptor
670 COUP-TFI is important for differentiation of oligodendrocytes. *Dev Biol*. 2004;266: 238–251.

671 65. Cui X, Lu YW, Lee V, Kim D, Dorsey T, Wang Q, et al. Venous Endothelial Marker COUP-
672 TFII Regulates the Distinct Pathologic Potentials of Adult Arteries and Veins. *Sci Rep*. 2015;5:
673 16193.

674 66. Duong TB, Ravisankar P, Song YC, Gafranek JT, Rydeen AB, Dohn TE, et al. Nr2f1a balances
675 atrial chamber and atrioventricular canal size via BMP signaling-independent and -dependent
676 mechanisms. *Dev Biol*. 2018;434: 7–14.

67. Devalla HD, Schwach V, Ford JW, Milnes JT, El-Haou S, Jackson C, et al. Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. *EMBO Mol Med*. 2015;7: 394–410.
68. Schwach V, Verkerk AO, Mol M, Monshouwer-Kloots JJ, Devalla HD, Orlova VV, et al. A COUP-TFII Human Embryonic Stem Cell Reporter Line to Identify and Select Atrial Cardiomyocytes. *Stem Cell Reports*. 2017;9: 1765–1779.
69. Barske L, Rataud P, Behizad K, Del Rio L, Cox SG, Gage Crump J. Essential Role of Nr2f Nuclear Receptors in Patterning the Vertebrate Upper Jaw. *Developmental Cell*. 2018. pp. 337–347.e5. doi:10.1016/j.devcel.2017.12.022
70. Fjose A, Nornes S, Weber U, Mlodzik M. Functional conservation of vertebrate seven-up related genes in neurogenesis and eye development. *EMBO J*. 1993;12: 1403–1414.
71. Sousounis K, Looso M, Maki N, Ivester CJ, Braun T, Tsonis PA. Transcriptome Analysis of Newt Lens Regeneration Reveals Distinct Gradients in Gene Expression Patterns. *PLoS ONE*. 2013. p. e61445. doi:10.1371/journal.pone.0061445
72. Bertrand S, Brunet FG, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi M. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. *Mol Biol Evol*. 2004;21: 1923–1937.
73. Warnecke M, Oster H, Revelli JP. Abnormal development of the locus coeruleus in Ear2 (Nr2f6)-deficient mice impairs the functionality of the forebrain clock and affects nociception. *Genes* . 2005. Available: <http://genesdev.cshlp.org/content/19/5/614.short>
74. Ichim CV, Dervovic D, Zuniga-Pflucker JC, Wells RA. The Orphan Nuclear Receptor NR2F6 Is a Novel Negative Regulator of T-Cell Development. *Blood*. 2009. pp. 915–915. doi:10.1182/blood.v114.22.915.915
75. Hermann-Kleiter N, Baier G. Orphan nuclear receptor NR2F6 acts as an essential gatekeeper of

Th17 CD4 T cell effector functions. *Cell Communication and Signaling*. 2014. p. 38.
doi:10.1186/1478-811x-12-38

76. Duester G. Retinoic acid synthesis and signaling during early organogenesis. *Cell*. 2008;134: 921–931.

77. Kinkel MD, Prince VE. On the diabetic menu: zebrafish as a model for pancreas development and function. *Bioessays*. 2009;31: 139–152.

78. Jonk LJ, de Jonge ME, Pals CE, Wissink S, Vervaart JM, Schoorlemmer J, et al. Cloning and expression during development of three murine members of the COUP family of nuclear orphan receptors. *Mech Dev*. 1994;47: 81–97.

79. Fjose A, Weber U, Mlodzik M. A novel vertebrate svp-related nuclear receptor is expressed as a step gradient in developing rhombomeres and is affected by retinoic acid. *Mech Dev*. 1995;52: 233–246.

80. Zaitseva M, Vollenhoven BJ, Rogers PAW. Retinoids regulate genes involved in retinoic acid synthesis and transport in human myometrial and fibroid smooth muscle cells. *Hum Reprod*. 2008;23: 1076–1086.

81. Laursen KB, Mongan NP, Zhuang Y, Ng MM, Benoit YD, Gudas LJ. Polycomb recruitment attenuates retinoic acid-induced transcription of the bivalent NR2F1 gene. *Nucleic Acids Res*. 2013;41: 6430–6443.

82. Wyffels J, King BL, Vincent J, Chen C, Wu CH, Polson SW. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. *F1000Res*. 2014;3: 191.

83. Dardaillon J, Dauga D, Simion P, Faure E, Onuma TA, DeBiasse MB, et al. ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates. *Nucleic Acids Res*. 2020;48: D668–D675.

84. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 2006;34: W362–5.
85. Katoh K, Kuma K-I, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005;33: 511–518.
86. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019;20: 1160–1166.
87. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59: 307–321.
88. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol.* 2017;34: 2422–2424.
89. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006;55: 539–552.
90. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* 2011;60: 685–699.
91. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics.* 2002;Chapter 2: Unit 2.3.
92. Nguyen NTT, Vincens P, Roest Crollius H, Louis A. Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Res.* 2018;46: D816–D822.
93. Singh PP, Isambert H. OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates. *Nucleic Acids Res.* 2020;48: D724–D730.

748

749 **Supporting Information**

750 **S1 Fig. Phylogenetic tree of the Nr2f family, using Bayesian Information Criterion (BIC).**

751 The same color code as Fig 2 is used. Values at the branches indicate replicates obtained
752 employing the aBayes method.

753

754 **S2 Fig. Synteny analysis of *Nr2f* genes in invertebrates.**

755 **A)** Schematic of *Nr2f1/2/5/6* gene duplications in *N. vectensis*. **B)** Schematic of conservation
756 of *Nr2f1/2/5* loci between *S. kowalevskii* and *B. belcheri*. Black arrows indicate transcription
757 orientation.

758

759 **S3 Fig. *Nr2f-Unc45* gene duplet preservation.**

760 Schematic of *Nr2f-Unc45* duplet conservation in genomes of ascidians (*Ciona*) and primates.
761 The duplet is absent in other vertebrate models, including zebrafish and mouse.

762

763 **S4 Fig. Synteny analysis of *Nr2f1/2* lamprey genes.**

764 Schematic of lamprey (*P. marinus*) *Nr2f1/2* loci with relative chromosomes. Same color code
765 of Figs. 3-6 is used. Flanking genes are in common with gnathostomes, with *Arrdc2* and -3
766 (green) that form a conserved duplet with *Nr2f1/2B* and *Nr2f1/2C*. Black arrows indicate
767 transcription orientation.

768

769 **S1 File. (docx) Intron/exon structure of *Nr2f* genes in Metazoa.**

770 Alignment of specific and conserved intron/exon boundaries within the Zinc finger motifs of
771 DBD (underlined) and LBDs (red). The intron phases have been depicted using color code:
772 Phase 0 (yellow), Phase 1 (green), Phase 2 (turquoise).

773

774 **S2 File. (docx) Nr2f domain architectures during metazoan evolution.**

775 Sequences used in analysis with DBDs (yellow) and LBDs (green) domains in metazoan Nr2f
776 proteins indicated. the Zinc-finger motifs within the DBDs are underlined.

777

778 **S3 File. (txt) MAFFT alignment of protein sequences used for phylogenetic analysis of**
779 **Fig 2 and S1 Fig.**

780

781 **S4 File. (txt) List of all protein sequences employed in Nr2f phylogenetic tree with**
782 **accession numbers.**

783

784 **S5 File. (txt) List of protein sequences excluded from Nr2f phylogenetic tree due to their**
785 **high degree of divergence.**

786

787 **S6 File. (xlsx) List of species used for our evolutionary analyses with their common names,**
788 **Latin names, and phyla.**

789

790 **S7 File. (docx) Selected Nr2f transcripts and translations used for analysis with the**
791 **positions and phases of intron/exon boundaries indicated.**

Figure 1

Figure 2

Znf I

Znf II

<i>N. vectensis</i>	Nr2f1/2/5/6a	CAVCGDKSSGKHYGVYTCEGC	CRGVRNCPVDIQNRNQCCQYC
<i>N. vectensis</i>	Nr2f1/2/5/6b	CAVCGDKSTGKHYGVSTCEGC	CRGQNTCAIDRNSRSRCPSC
<i>N. vectensis</i>	Nr2f1/2/5/6c	CAVCGDKSSGKHYGVFTCEGC	CRASRDCPIDQHHRNQCCQYC
<i>T. adhaerens</i>	Nr2f1/2/5/6	CLICGDRSNGRHYGVISCEGC	CTCSANCKITKANRNQCCQFC
<i>D. melanogaster</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQFTCEGC	CRGSRNCPIDQHHRNQCCQYC
<i>C. elegans</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQFSCEGC	CRATKNCAIDVQHRNQCCQYC
<i>L. gigantea</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQYTCEGC	CRGNKNCPIDQHHRNQCCQYC
<i>S. kowalevskii</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>S. purpuratus</i>	Nr2f1/2/5	CVVCHDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>C. robusta</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQYTCEGC	CRGNRNCPIDQHHRNQCCQYC
<i>O. dioica</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQFTCEGC	CRGNRSCPIDQHHRNQCCQYC
<i>B. belcheri</i>	Nr2f1/2/5	CVVCGDKSSGKHYGQFTCEGC	CRGNRTCPIDQHHRNQCCQYC
<i>P. marinus</i>	Nr2f1/2A	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>P. marinus</i>	Nr2f1/2B	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>P. marinus</i>	Nr2f1/2C	CVVCGDKSSGKHYGQFTCEGC	CRAARACPIDQHHRNQCCQYC
<i>E. burgeri</i>	Nr2f1/2A	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>E. burgeri</i>	Nr2f1/2B	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>E. burgeri</i>	Nr2f1/2C	CVVCGDKSSGKHYGQFTCEGC	CRAARNCPIDQHHRNQCCQYC
<i>C. milii</i>	Nr2f1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>L. oculatus</i>	Nr2f1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>D. rerio</i>	Nr2f1a	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>D. rerio</i>	Nr2f1b	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPVDQHHRNQCCQYC
<i>L. chalumnae</i>	Nr2f1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>X. tropicalis</i>	Nr2f1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>F. albicollis</i>	Nr2f1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>G. japonicus</i>	Nr2f1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>H. sapiens</i>	NR2F1	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>C. milii</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>L. oculatus</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>D. rerio</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>L. chalumnae</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>X. tropicalis</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>F. albicollis</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>G. japonicus</i>	Nr2f2	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>L. oculatus</i>	Nr2f5	CMVCGDKSSGKHYGQFTCEGC	CRGNRDCPIDQHHRNQCCQYC
<i>D. rerio</i>	Nr2f5	CMVCGDKSSGKHYGQFTCEGC	CRGNRDCPIDQHHRNQCCQYC
<i>L. chalumnae</i>	Nr2f5	CMVCGDKSSGKHYGQFTCEGC	CRGNRDCPIDQHHRNQCCQYC
<i>X. tropicalis</i>	Nr2f5	CLVCGDKSSGKHYGQFTCEGC	CRGNRDCPIDQHHRNQCCQYC
<i>C. milii</i>	Nr2f6	CVVCGDRASGKHYGQFTCEGC	CRSNRDCPIDQHHRNQCCQYC
<i>L. oculatus</i>	Nr2f6	CVVCGDKSSGKHYGVFTCEGC	CRSNRDCQIDQHHRNQCCQYC
<i>D. rerio</i>	Nr2f6a	CVVCGDKSSGKHYGVFTCEGC	CRSNRDCQIDQHHRNQCCQYC
<i>D. rerio</i>	Nr2f6b	CVVCGDKSSGKHYGVFTCEGC	CRSNRECQIDQHHRNQCCQYC
<i>L. chalumnae</i>	Nr2f6	CVVCGDKSSGKHYGAFTCEGC	CRSNRDCQIDQHHRNQCCQYC
<i>X. tropicalis</i>	Nr2f6	CVVCGDKSSGKHYGQFTCEGC	CRANRNCPIDQHHRNQCCQYC
<i>F. albicollis</i>	Nr2f6	CVVCGDKSSGKHYGVFTCEGC	CRSNRDCQIDQHHRNQCCQYC
<i>G. japonicus</i>	Nr2f6	CVVCGDKSSGKHYGVFTCEGC	CRSNRDCQIDQHHRNQCCQYC
<i>H. sapiens</i>	NR2F6	CVVCGDKSSGKHYGVFTCEGC	CRSNRDCQIDQHHRNQCCQYC

Figure 3

DBD

1101.2021.07.22.453436; this version posted July 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Nr2f1/2/5/6	<i>Trichoplax adhaerens</i>
Nr2f1/2/5	<i>Ciona robusta</i>
Nr2f1/2/2A	<i>Petromyzon marinus</i>
Nr2f1/2C	<i>Petromyzon marinus</i>
Nr2f1a	<i>Danio rerio</i>
Nr2f1b	<i>Danio rerio</i>
NR2F1	<i>Homo sapiens</i>
Nr2f2	<i>Danio rerio</i>
NR2F2	<i>Homo sapiens</i>
Nr2f5	<i>Danio rerio</i>
Nr2f6a	<i>Danio rerio</i>
Nr2f6b	<i>Danio rerio</i>
NR2F6	<i>Homo sapiens</i>

AKDETSLSKELCCLICGDRSNGRHVVISCEGCKGFFKRSVRRNMKYACTCSANACKIT
GNTSSTSEKPQIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRGN-RNCPID
holder for this morph
eprint in perpetuity. It is
GDAQ----QSNVECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRAN-RNCPID
---TPIASGQNAECVVCGDKSSGKHYGQLTCEGCKSFFKRSVRRNLSYTCRAN-RACPID
QNSGQS-Q-QHIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRAN-RNCPID
KAQIAAQNQHQHVECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRAN-RNCPVD
GPPGSGQSQQHIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRAN-RNCPID
QSQSEKQQPQHIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRAN-RNCPID
PGSDKQQQQQHIECVVCGDKSSGKHYGVFTCEGCKSFFKRSVRRNLSYTCRAN-RNCPID
ALSG--DKIPNVDCMVCVKDKSSGKHYGQFTCEGCKSFFKRSVRRNLSYTCRGN-RDCPID
EAGEDD-KGCVVDCVVCGDKSSGKHYGVFTCEGCKSFFKRSVRRNLSYTCRAN-RDCPID
EGGEED-KACVVDCVVCGDKSSGKHYGVFTCEGCKSFFKRSVRRNLSYTCRAN-RDCPID
EPGDEERPGLVDCVVCGDKSSGKHYGVFTCEGCKSFFKRSVRRNLSYTCRAN-RDCPID

Nr2f1/2/5/6	<i>Trichoplax adhaerens</i>
Nr2f1/2/5	<i>Ciona robusta</i>
Nr2f1/2A	<i>Petromyzon marinus</i>
Nr2f1/2B	<i>Petromyzon marinus</i>
Nr2f1/2C	<i>Petromyzon marinus</i>
Nr2f1a	<i>Danio rerio</i>
Nr2f1b	<i>Danio rerio</i>
NR2F1	<i>Homo sapiens</i>
Nr2f2	<i>Danio rerio</i>
NR2F2	<i>Homo sapiens</i>
Nr2f5	<i>Danio rerio</i>
Nr2f6a	<i>Danio rerio</i>
Nr2f6b	<i>Danio rerio</i>
NR2F6	<i>Homo sapiens</i>

KANRNQCQFCRLQKCFKVGMRKE-----AVQKERHTSTIRADRN-SGK-----
QHHRNQCOYCRNLNKC VKIGMRRE-----AVQGRMPPSQPHT--TGQ-----Y
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRLLPPSQHPNP-----A-QY
QHHRNQCOYCRLLKKCLKVGMRREAL--VPFPAVQGRLLAPTQHPNP-----ALQY
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRMPSSMQHPATPGGAYALQPNGGD
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRMPPTQP-N---PGQ-----Y
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRMPPN-QPN---PSH-----Y
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRMPPTQP-N---PGQ-----Y
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRMPPT-QPH---HGQ-----F
QHHRNQCOYCRLLKKQAVWGMRRRE-----AVQGRMPPT-QPT---HGQ-----F
QHHRNQCOYCRLLKKCLKVGMRRE-----AVQGRMSNSQS---SPGQ-----Y
QHHRNQCOYCRLLKKCFRVGMRKE-----AVQGRIPPSHSSLS PSTTP--VGGNAGG
QHHRNQCOYCRLLKKCFRVGMRKE-----AVQGRIPPSHAGISPASMV-GAGGDVGG
QHHRNQCOYCRLLKKCFRVGMRKE-----AVQGRIPHSLPGAVAASSG-SPPGSALA
: . . . * * * * :

LBD

Nr2f1/2/5/6	<i>Trichoplax adhaerens</i>
Nr2f1/2/5	<i>Ciona robusta</i>
Nr2f1/2A	<i>Petromyzon marinus</i>
Nr2f1/2B	<i>Petromyzon marinus</i>
Nr2f1/2C	<i>Petromyzon marinus</i>
Nr2f1a	<i>Danio rerio</i>
Nr2f1b	<i>Danio rerio</i>
NR2F1	<i>Homo sapiens</i>
Nr2f2	<i>Danio rerio</i>
NR2F2	<i>Homo sapiens</i>
Nr2f5	<i>Danio rerio</i>
Nr2f6a	<i>Danio rerio</i>
Nr2f6b	<i>Danio rerio</i>
NR2F6	<i>Homo sapiens</i>

Y-----LTHPQEVMMHHICLFQEAIVKLKKRFIDTTEFSCSLKALILFNPDPVIRGLVNPNYVE
HTS-MSADRVMTFMDHIRIFQEQVERLKS LHVDSAEYSCLKAIVLFTADSHGLSDMTHIE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTTDACGLSDAAHIE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDAAHIE
HASPMSAERVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDTAHVE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDAAHIE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDAAHIE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDAAHIE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDVAHVE
HASPMSADRVVAFMDHIRIFQEQVEKLKALHVDSEYSCIKAIVLFTSDACGLSDVAHVE
HASPMSAERVVAFMDHIRVFQEQVEKLKALQVDTAEYSCLSIVLFTSDAMGLSDVAHVE
HSSPMSAERVVSFMDQVRVFQDQVEKLTRLQVDSAEYSCIKAIALFSPDACGLTDPAHVE
HSSPMPADRVVSFMDQVRVFQDQVDKTRLQVDSVEYSCIKAIALFSPDACGLSDPAHVE
HAAPMAAERAVAFMDQVRAFQEQVDKLGRLQVDSAEYGCLKAIALFTP DACGLSDPAHVE
:
: * : : * : : * : : * : * : : * : : * : : *

B

DBD

LBD

Nr2f1/2/5/6	Nr2f1
Nr2f1/2/5	Nr2f2
Nr2f1/2	Nr2f5

Phase 1

Phase 1

Phase 2

Nr2f6

Figure 4

Figure 6

Figure 7

Figure 8

Figure 9