

RESEARCH

Open Access



Gene expression in notochord and *nuclei pulposi*: a study of gene families across the chordate phylum

Rahul Raghavan^{1†}, Ugo Coppola^{2,3†}, Yushi Wu^{4†}, Chibuike Ihewulezi^{4†}, Lenny J. Negrón-Piñeiro⁴, Julie E. Maguire⁴, Justin Hong¹, Matthew Cunningham^{5,6}, Han Jo Kim^{5,6}, Todd J. Albert^{5,6}, Abdullah M. Ali⁷, Jean-Pierre Saint-Jeannet⁴, Filomena Ristoratore², Chitra L. Dahia^{1,8*} and Anna Di Gregorio^{4*}

Abstract

The transition from notochord to vertebral column is a crucial milestone in chordate evolution and in prenatal development of all vertebrates. As ossification of the vertebral bodies proceeds, involutions of residual notochord cells into the intervertebral discs form the *nuclei pulposi*, shock-absorbing structures that confer flexibility to the spine. Numerous studies have outlined the developmental and evolutionary relationship between notochord and *nuclei pulposi*. However, the knowledge of the similarities and differences in the genetic repertoires of these two structures remains limited, also because comparative studies of notochord and *nuclei pulposi* across chordates are complicated by the gene/genome duplication events that led to extant vertebrates. Here we show the results of a pilot study aimed at bridging the information on these two structures. We have followed in different vertebrates the evolutionary trajectory of notochord genes identified in the invertebrate chordate *Ciona*, and we have evaluated the extent of conservation of their expression in notochord cells. Our results have uncovered evolutionarily conserved markers of both notochord development and aging/degeneration of the *nuclei pulposi*.

Keywords *Ciona*, *Coronin*, Human, Mouse, Notochord, *Nucleus pulposus*, *Phip*, *Rgm*, *Rgs*, *Tmod*, *Xenopus*, Zebrafish

[†]Rahul Raghavan, Ugo Coppola, Yushi Wu, Chibuike Ihewulezi contributed equally to this work.

⁸ Department of Cell and Developmental Biology, Weill Cornell Medicine, Graduate School of Medical Science, New York, NY 10065, USA

*Correspondence:

Chitra L. Dahia

dahiac@hss.edu

Anna Di Gregorio

adg13@nyu.edu

¹ Hospital for Special Surgery, Orthopedic Soft Tissue Research Program, New York, NY 10021, USA

² Stazione Zoologica 'A. Dohrn', Villa Comunale 1, 80121 Naples, Italy

³ Present Address: Molecular Cardiovascular Biology Division and Heart Institute, Cincinnati Children's Research Foundation, Cincinnati, OH 45229, USA

⁴ Department of Molecular Pathobiology, New York University College of Dentistry, New York, NY 10010, USA

⁵ Hospital for Special Surgery, New York, NY 10021, USA

⁶ Weill Cornell Medical College, New York, NY 10065, USA

⁷ Department of Medicine, Columbia University Irving Medical Center, New York, NY 10032, USA



Introduction

In all chordate embryos, the notochord is the main source of axial support and patterning cues [1–5]. In invertebrate chordates, *i.e.* tunicates and cephalochordates, the notochord persists throughout embryogenesis as the main supporting structure for the developing body. In vertebrates, the notochord undergoes segmentation as it is gradually replaced by the vertebral column, and its remnants are incorporated in the intervertebral discs (IVDs), where they form the centrally located *nuclei pulposi* (NP). The IVDs allow movement in between vertebrae, and evenly spread the mechanical loading on the vertebral bodies. The NP, with their gelatinous consistency, are critical for the function of the IVDs, and their alteration is intimately connected to IVD deterioration [6–8]. Because they are an integral component of the NP, notochord cells have emerged in recent years as a therapeutic avenue for human IVD regeneration [9].

Members of the tunicate subphylum are considered the invertebrate chordates most closely related to vertebrates [10], and indeed the tunicate notochord has been shown to share distinctive morphological and molecular traits with the vertebrate notochord (reviewed in [4, 11]). In the tunicate *Ciona*, the notochord consists of only 40 cells, arranged in a single row (*e.g.*, [12]). In the teleost zebrafish (*Danio rerio*), the notochord consists of two layers (inner and outer) of vacuolated cells [13]. In *Xenopus*, mouse, and humans, the notochord is composed of several cells that develop large vacuoles as this structure differentiates [1, 14–17]. In all these species, the notochord cells secrete abundant amounts of extracellular matrix proteins, which form a thick notochordal sheath [1, 15, 18–20].

Tunicate genomes have not undergone the extensive duplication events that have shaped the genomes of vertebrates [21, 22] and a considerable fraction of the genes that in vertebrates have originated composite gene families appear in single copy in *Ciona* and other tunicates, even though relevant examples of lineage-specific gene duplications have been reported [23–27]. Therefore, tunicates can serve as an informative point of reference for reconstructing the evolutionary origins of complex vertebrate gene families and for studies of conservation/divergence of gene expression patterns. We have previously assessed the conservation of notochord gene expression between two divergent tunicates, *Ciona robusta* and *Oikopleura dioica* [28]; in addition to that, we have identified counterparts of *Ciona* notochord genes in the mouse genome, and determined that the notochord expression observed in *Ciona* is mirrored by the majority of the mouse genes that we analyzed [29]. More recently, we have demonstrated that the cross-regulatory relationship that we uncovered in *Ciona* between two notochord

transcription factors, *Brachyury* and *Xbp1*, is conserved in *Xenopus* [30].

Our previous studies show that hallmark notochord genes, such as *Brachyury* and *Sonic hedgehog*, are expressed by the postnatal NP cells of mouse IVDs [31, 32], providing a rationale for examining the expression of other genes that may be conserved and possibly have functional significance in the maintenance of these cells.

Here we first report the notochord expression of five *Ciona* genes, which were identified as part of our ongoing effort to characterize the *Ciona* notochord genetic toolkit. We used these five *Ciona* genes as a starting point to survey the expression of notochord genes that are present in single copy in tunicates and, through duplication events, have given rise to multigenic families in vertebrates. We began this analysis with a gene, *Pleckstrin-homology domain interacting protein* (*Phip*), which has remained in single copy in all the chordates analyzed here, and we expanded this study to genes that in vertebrates are part of increasingly larger families. *Phip* encodes for a protein that binds to the insulin receptor substrate 1 (*Irs1*) and is hypothesized to act as a link between *Irs1* and the insulin receptor, thus modulating the insulin pathway [33, 34]; mice lacking *Phip1*, the main isoform of *Phip*, develop hypoglycemia and have a short lifespan [35]. In humans, a nonsense mutation in the *PHIP* gene has been recently linked to Chung-Jansen syndrome, which is characterized by obesity and developmental delay [36]. Another gene we analyzed, *repulsive guidance molecule* (*Rgm*), is present in single copy in *Ciona*, while in vertebrates is part of a multigenic family with four distinct subfamilies (*RgmA–D*), whose components, originally identified as key players in neuronal growth dynamics, have been associated with the development of numerous tissues and structures, and with their respective pathologies [37]. A larger gene family of interest for this study is the *regulators of G-protein signaling* (*Rgs*), because of their high degree of conservation across divergent chordates and their requirement for the timely inactivation of G-proteins [38]. The protein encoded by the *Ciona robusta* *Rgs6/7* gene is equally related to members of the *Rgs6* and *Rgs7* subfamilies, which in vertebrates are involved, in particular, in the control of phototransduction [39, 40].

In addition to the aforementioned genes, we followed across the chordate spectrum the expression of the orthologs of a member of the *Coronin* gene family, *Ciona Coronin-7* (*Coro7*). Coronins are actin-binding regulators of cytoskeletal remodeling, vesicle trafficking and cell motility [41, 42] and belong to the WD-repeat superfamily [43]; we also analyzed two *tropomodulin/leiomodin*-related *Ciona* genes, which provided insights into lineage-specific changes in notochord gene expression.

Lastly, in an effort to use information gathered in *Ciona* to shed light on gene expression during the evolutionary and developmental transition from notochord to NP, and to guide the identification of genetic markers of disco-pathies, we have analyzed the expression of orthologs of these *Ciona* notochord genes in both aging and degenerating mammalian NP.

Results

Identification of notochord genes in *Ciona*

We have previously reported the expression of orthologs of vertebrate notochord genes in *Ciona* [30, 44–47]. As a starting point for this study, we selected five genes expressed in the *Ciona* notochord for which the information on the expression in the vertebrate notochord was either fragmentary or lacking altogether.

Whole-mount in situ hybridization experiments revealed that expression of *Ciona Phip* (Fig. 1A–E; gene model: KH.C12.267; Table S1) is rather diffuse, and is detected throughout most of the embryo until gastrulation, resembling the expression pattern usually attributed to maternal transcripts [48, 49]. However, the notochord and neural precursors show a more intense signal (Fig. 1A,B), which at the tailbud stages becomes more localized to CNS and mesenchyme, while persisting at low levels in the notochord (Fig. 1C–E).

The *Ciona robusta* genome contains one identifiable member of the *Rgm* gene family, which encodes for a protein equally related to vertebrate Rgm proteins A–D, and has been therefore designated here as *RgmA/B/C/D* (Fig. 1F–J; gene model: KH.L170.10; Table S1). Starting from early developmental stages, this gene is specifically expressed in neural precursors (Fig. 1F) and by late gastrulation becomes detectable at high levels in muscle precursors as well (Fig. 1G). Expression in both territories is still detected in late neurulae (inset in Fig. 1H), however, in initial tailbuds, muscle expression begins to fade while expression in the CNS remains strong and becomes localized to different regions of the sensory vesicle (Fig. 1H). Finally, in later tailbud stages, expression in the CNS persists in the ventral region of the sensory vesicle, and weakens in muscle cells, while becoming detectable in the notochord (Fig. 1I,J). The expression of this gene in muscle and sensory vesicle is also reported in the Aniseed expression database (<https://www.aniseed.fr>) [52].

We also selected for this analysis two *Ciona* genes related to vertebrate *tropomodulin* genes, which we indicate here as *Tropomodulin1/2/3/4* (*Tmod1/2/3/4*) and *Leiomodin* (*Lmod*). The reason for using this nomenclature is that we observed that although the predicted proteins for both genes are related to tropomodulins from other species, the putative *Ciona* Lmod protein contains in its C-terminal a WASP-Homology 2

(WH2) actin-binding domain and a short proline-rich region, which are characteristic features of leiomodins from other species [53, 54].

Tmod1/2/3/4 (Fig. 1K–O; gene model: KH.L161.1; Table S1) is first detected in invaginating notochord precursors (Fig. 1K) and is transiently expressed in neural precursors during late gastrulation (Fig. 1L); the hybridization signal increases considerably at the time of neurulation (inset in Fig. 1M) and remains intense and notochord-specific throughout the tailbud stages (Fig. 1M–O). On the other hand, *Lmod* (Fig. 1P,Q; gene model: KH.C14.251; Table S1), is detectable by WMISH only in muscle cells of late tailbuds (Fig. 1P,Q); this result is consistent with reports of muscle-specific or muscle-predominant expression of leiomodins [55].

Next, we analyzed the expression of a member of the *Rgs* gene family, which we named *Rgs6/7* because it is orthologous to both *Rgs6* and *Rgs7* genes identified in vertebrates (see below). In early embryos, at the 16-cell stage, expression of *Ciona Rgs6/7* (gene model: KH.C2.958; Table S1) had been reported in all blastomeres [56]. We found that although *Rgs6/7* remains widely expressed during early embryogenesis, in a pattern that is suggestive of maternal expression (Fig. 1R), by late gastrulation its expression begins to fade from the precursors of the epidermis (Fig. 1S). In initial tailbuds, *Rgs6/7* is detected in notochord and mesenchyme cells, as well as in the sensory vesicle (Fig. 1T). Around the mid-tailbud stage, expression in the notochord decreases, persisting mainly in cells of the secondary notochord, while expression in the CNS increases in intensity and broadens to encompass the entire sensory vesicle and the nerve cord, which extends throughout the length of the tail (Fig. 1U). In late tailbuds, expression in notochord cells falls below detection, while expression in the CNS increases further (Fig. 1V).

Ciona Coronin7 (*Coro7*, Fig. 1W–AA; gene model: KH.C2.819; Table S1) displays a diffuse expression at early and late gastrula stages (Fig. 1W,X); by the initial tailbud stage, the hybridization signal becomes refined to sensory vesicle and mesenchyme, and persists in intercalating notochord cells, although at lower levels compared to the other territories (Fig. 1Y). Expression is reduced but still detectable in notochord cells throughout the mid-tailbud stages (Fig. 1Z,AA).

Even taking into account possible differences in the experimental conditions that influence embryonic staging (e.g., temperature, incubation time), the results of our WMISH are mostly in agreement with published single-cell RNA-sequencing (scRNA-Seq) data [50] (Fig. 1AB). In particular, both results indicate that while *Tmod1/2/3/4* is strongly expressed in notochord cells at all stages analyzed (Fig. 1K–O, 1AB),

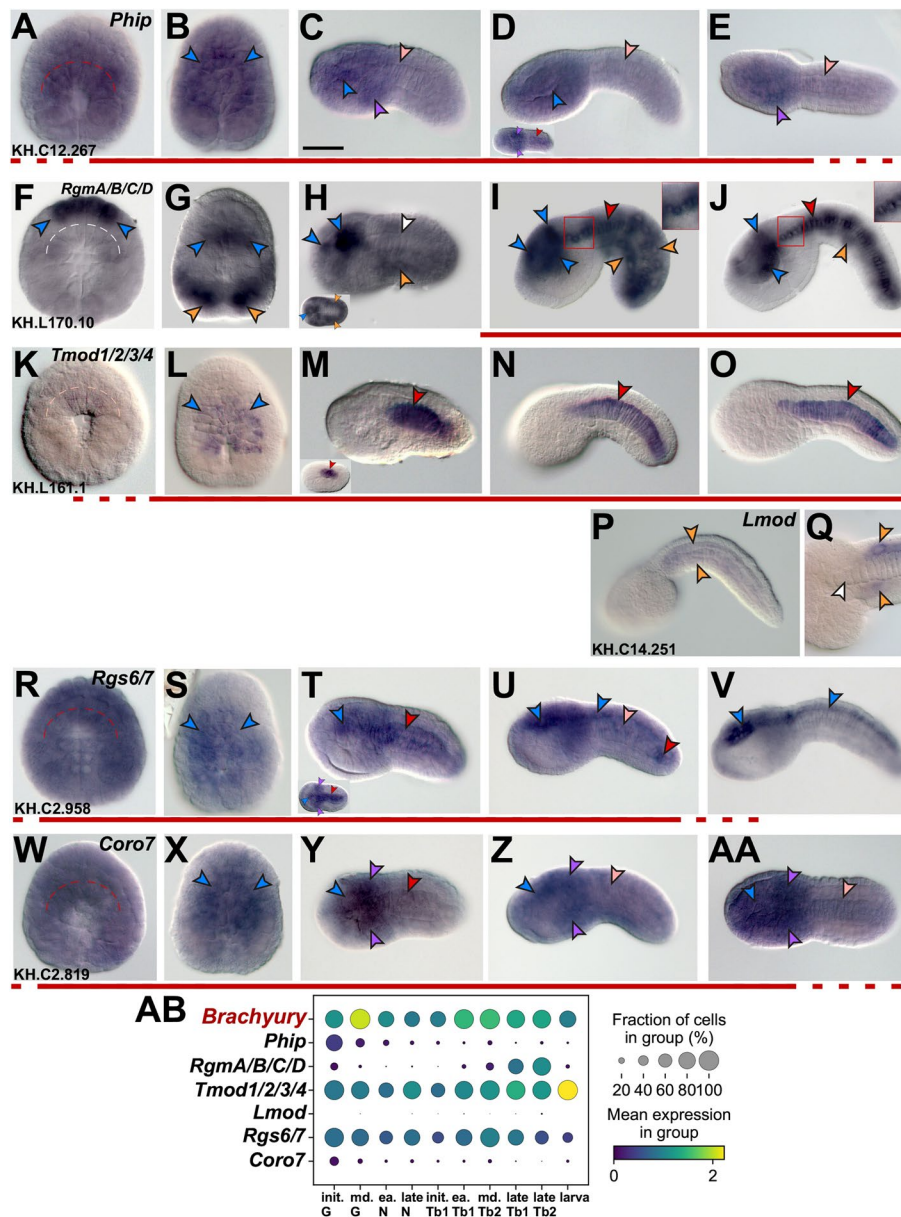


Fig. 1 Identification of genes expressed during notochord development in *Ciona robusta*. *Ciona robusta* embryos ranging from early gastrula to late tailbud, hybridized in situ with antisense RNA probes specific for the genes reported on top of panels **A, F, K, P, R, W** (Table S1). Gene models are indicated in the bottom left corners. Insets in **(D, H, M, T)** show embryos at different developmental stages. Insets in **(I, J)** show a higher magnification view of the regions boxed in red in the main panels, to display staining in notochord cells, which is clearer in the trunk because it is not obscured by the staining in muscle cells. Stained territories are denoted by arrowheads, color-coded as follows: red, notochord; light pink, fading notochord staining; white, no detectable notochord staining; blue, CNS; purple, mesenchyme; orange, muscle. Red lines underneath the panels indicate the approximate time span of notochord expression throughout development. In **(A, F, K, R, W)** dashed curved lines delineate the location of notochord precursors. Scale bar: 50 μ m. **(AB)** Dot plot summary of the published scRNA-Seq data [50] available for the genes in **(A-AA)**, compared to the scRNA-Seq data available for *Ciona Brachyury*, which was used as a reference for notochord expression [51]. In the dot plot, each dot represents two values: the mean expression of each gene (visualized by color) and the fraction of cells expressing each gene (visualized by the size of the dot). The embryonic stages used by Cao et al. [50] to generate the scRNA-Seq dataset reflect only approximately the stages that we used for WMISH. Abbreviations: init., initial; md., middle; ea., early; G, gastrula; N, neurula; Tb, tailbud

expression of *Lmod* is either undetectable or negligible (Fig. 1P,Q,AB).

Identification of vertebrate orthologs of *Ciona* notochord genes through phylogenetic analyses

In order to identify vertebrate orthologs of the *Ciona* notochord genes selected for this study (Fig. 1), we carried out phylogenetic analyses for the *Phip*, *Rgm*, *Tmod*, and *Coronin* families, employing a manually curated database of protein sequences (Supplementary files 1–5).

Phip genes (also known as *Brdw3*) have remained in single copy in the species analyzed in this study. We assessed their conservation by aligning *Phip* protein sequences selected from chordates and other metazoan taxa (Fig. S1; Supplementary file S1). As expected, the highest degree of sequence conservation was found in the WD-repeat region and in the bromodomain (highlighted in Fig. S1).

It was previously reported that invertebrate chordates possess a single-copy *Rgm* gene [57]; the maximum likelihood (ML) phylogenetic tree that we obtained for the *Rgm* family shows the relationship between the single-copy *Rgm* gene present in *Ciona* and in the non-chordate invertebrates selected for this analysis, which we termed *RgmA/B/C/D* (Fig. 2). Most vertebrate genomes contain three paralogous groups of *Rgm* genes: *RgmA*, *RgmB* (also known as *Dragon*) and *RgmC* (also known as *hemojuvelin/hjv*, or *hfe2*). Our phylogenetic reconstruction indicates that a fourth paralogous group of this gene family, *RgmD*, which thus far has only been reported in teleosts [58] and in cartilaginous fish (*C. milii*, Fig. 2), is closely related to the *RgmB* genes of tetrapods (Fig. 2). The close relationship between *RgmA* and *RgmB* paralogs had been previously reported, and is corroborated by the synteny analysis of their genomic surroundings; in particular, in syntenic regions of different chordates *RgmB* is linked to *Dnajb5* (*Hsp40*) [58]. Remarkably, the linkage of *Rgm* to *Dnajb5* is present in *Ciona* as well as in *Gasterosteus aculeatus* (stickleback), *Xenopus* and mouse [58]. We updated this earlier report using the latest version of the *Ciona robusta* genome [59, 60], traced these markers to the latest version of the human genome, and identified additional conserved markers in vertebrate genomes relevant to this study (Fig. S2). Of note, we found that the clustering of a *RgmA/B/C/D* gene with genes encoding chromodomain helicase DNA-binding proteins (*chd*), which are conserved neighbors of vertebrate *RgmA* and *RgmB* genes [58], is present in the genome of the hemichordate *Saccoglossus kowalevskii* (Fig. S2). We also found additional genes that are maintained in the proximity of *RgmA* and *RgmB* genes of vertebrates (Fig. S2, S3), a finding that reinforces the close relationship between *RgmA* and *RgmB* paralogs. Accordingly, our

synteny analysis of vertebrate *RgmC* genes indicates a considerable difference between the genomic surroundings of these paralogs and those of the *RgmA* and *RgmB* genes (Fig. S3). None of the conserved neighbors of the *RgmA-C* genes seems to be present in the proximity of the *RgmD* paralogs in any of the species that we surveyed (Fig. S4). Most of the genes neighboring *RgmD* in fish genomes are maintained on the same chromosomes in the tetrapods analyzed here, despite the absence of *RgmD* in tetrapods (Fig. S4).

A phylogenetic reconstruction of the *tropomodulin* family is shown in Fig. 3, and indicates that the *Tmod2* and *Tmod3* subfamilies are closely related (Fig. 3), which is consistent with previous reports and with the clustered arrangement that these genes have in mammals and chickens [61]. The zebrafish genome contains only one of these two paralogs, which is commonly referred to as *tmod2* (Fig. 3), although it has been argued to be more closely related to *tmod3* [61]. We carried out a comparative study of the genomic surroundings of *Tmod2* and *Tmod3* genes across vertebrates, and found that *Callorhynchus milii* (elephant shark), *Latimeria chalumnae* (coelacanth), *Lepisosteus oculatus* (spotted gar) and stickleback also contain *Tmod2/Tmod3* genes in single copy, and that these genes are all located within genomic contexts comparable to those of the *Tmod2* and *Tmod3* genes found in tetrapods (Fig. S5).

At least 21 *Rgs* genes have been reported in mouse [62], and in the human genome individual and/or tandem duplications of several members of this family have generated 20 canonical RGS proteins and 19 RGS-like proteins [63]. A complete phylogenetic study of the *Rgs* proteins found in *Ciona robusta* was already available [38], and *Rgs6*, *Rgs7*, *Rgs9* and *Rgs11* had been suggested to share a common evolutionary origin [38, 62]; therefore, we focused our phylogenetic analysis on members of these groups (Fig. 4). In particular, the presence of two *rgs7* orthologs in zebrafish suggests that they have stemmed from the teleost-specific WGD event [64, 65], and prompted us to analyze the expression of *Danio rgs7* genes.

The evolutionary origins of *coronin* genes can be traced back to either a single gene [66] or two related genes [67], which expanded to give rise to at least 723 coronin proteins grouped into different classes [43]. Our ML phylogeny highlighted the relationships among the members of four subfamilies, *Coro1*, *Coro2*, *Coro6* and *Coro7* (Fig. 5). Although class-4 coronins have been identified from different eukaryotic species, they have undergone lineage-specific events that caused gain/loss of their PH and gelsolin protein domains [67]. This high variability, along with the paucity of coronin-4 homologs and the changes in nomenclature

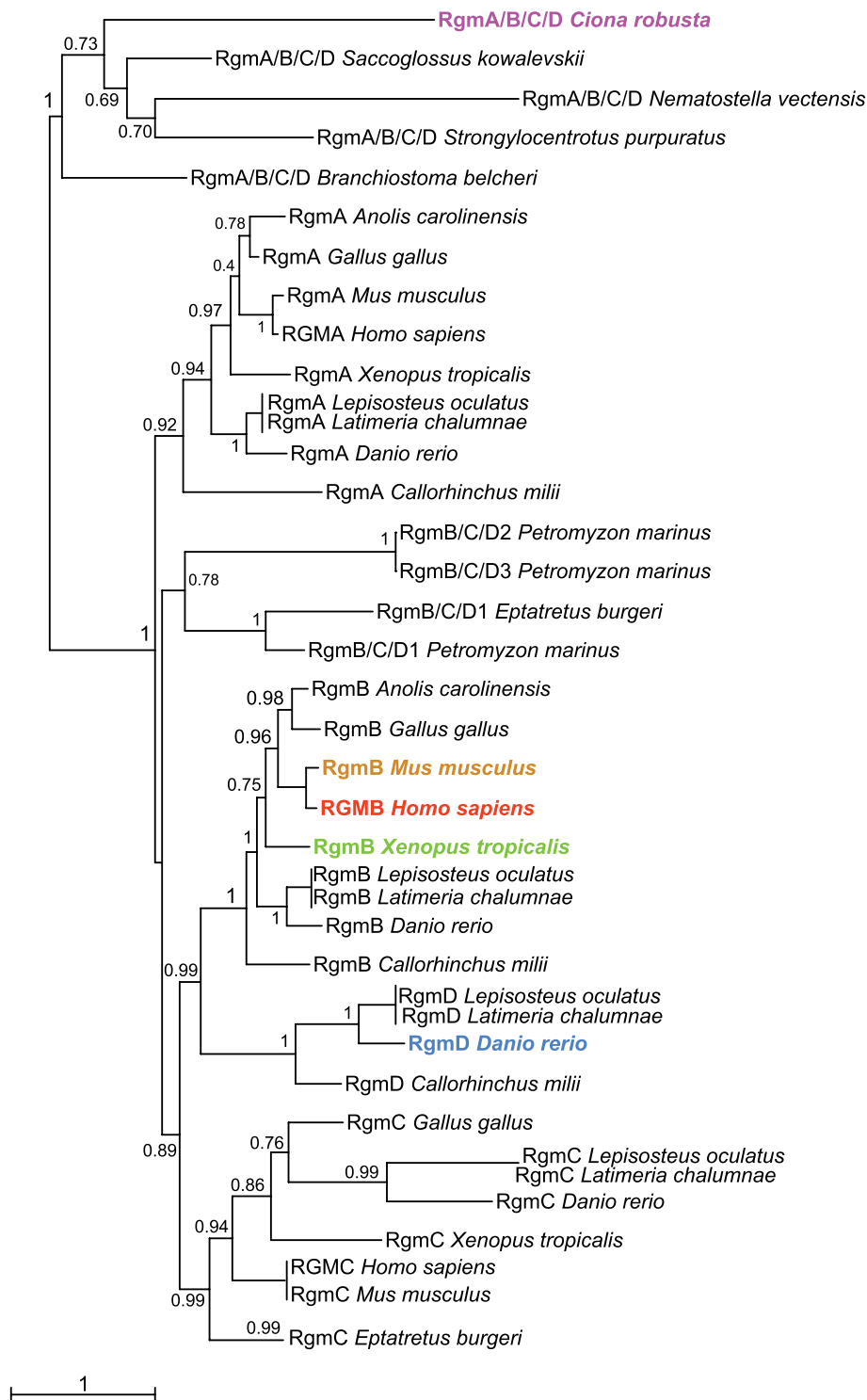


Fig. 2 Phylogenetic reconstruction of the evolutionary relationships within the *Rgm* gene family. ML phylogenetic tree showing the relationships among members of different *Rgm* classes. Proteins encoded by genes present in invertebrates in single copy and equally related to the RgmA–D classes have been indicated as RgmA/B/C/D. RgmA, RgmB, and RgmC were found in all the vertebrates analyzed in this study, while *RgmD* genes have only been reported, thus far, in teleosts and cartilaginous fish. Distinct colors highlight the family members analyzed in this study. Values reported at the branching points indicate replicates obtained using the aLRT method

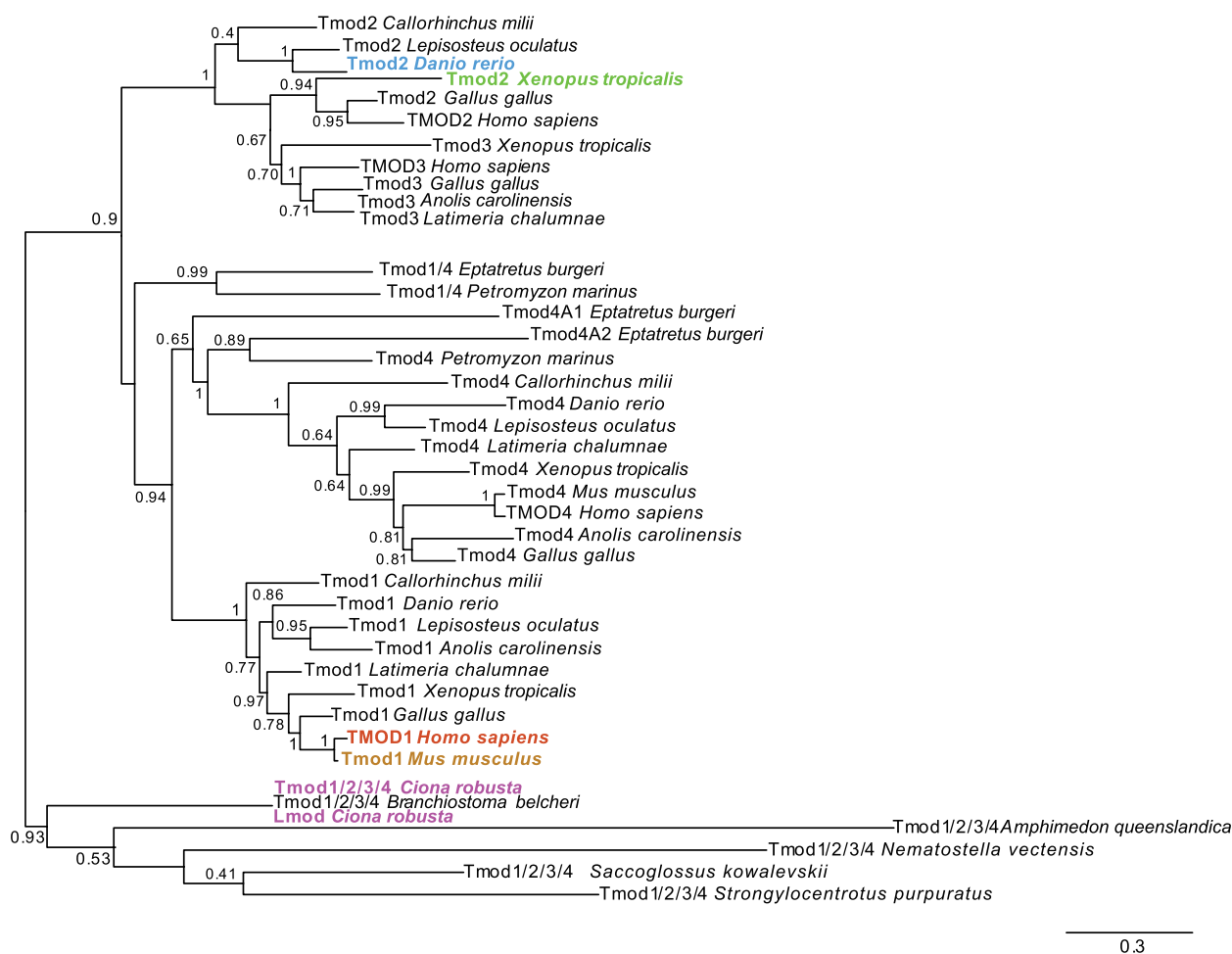


Fig. 3 Phylogenetic reconstruction of the evolutionary relationships within the *Tmod* gene family. ML phylogenetic tree displaying the relationships among members of different *Tmod* classes. Proteins encoded by genes present in invertebrates in single copy and equally related to the 1–4 classes have been tentatively indicated as *Tmod*1/2/3/4. *Tmod*1, *Tmod*2, *Tmod*3 and *Tmod*4 have been found in all the vertebrates analyzed here. Distinct colors highlight the family members analyzed in this study. Values reported at the branching points indicate replicates obtained using the aLRT method

of some of these genes, prevented the inclusion of this subfamily in this analysis. With the notable exception of the hemichordate *S. kowalevskii*, members of the *Coro7* subfamily seem to have remained in single copy, while members of the other subfamilies are duplicated in vertebrates (Fig. 5). Remarkably, our synteny analysis showed that one of the *Coro7* genes of *S. kowalevskii* is linked to *Dnaja3*; in *Ciona robusta*, this gene is not adjacent to *Coro7*, but is located on the same chromosome (Fig. S6). One of the introns of *Ciona Coro7* contains a coding region whose predicted product is a transmembrane protein related to both fibronectin leucine-rich transmembrane protein 1 (*flrt1*) and vasorin (*vasn*). Vasorin acts as a negative regulator of TGF-beta signaling [68] and appears embedded within the *Coro7* transcription unit in all the genomic loci that we analyzed, except in the case of the *Danio* genome

(Fig. S6). Other genes found on the same *Ciona* chromosome as *Coro7*, and conserved in the corresponding vertebrate loci, are *Pam16* and *Hmox* (Fig. S6). Interestingly, in human and in other vertebrate genomes *Coro7* and *Pam16* are located in close proximity and arranged in tandem orientation, and form a naturally occurring read-through transcription unit [69]. We found that this arrangement is not present in *Saccoglossus* and *Ciona*; *Saccoglossus Pam16* is located on a sequence scaffold different from those of *Coro7a* and *Coro7b*, while in *Ciona* these genes are very distant, even though they are located on the same chromosome (Fig. S6). In spotted gar, zebrafish, and all vertebrates analyzed here, *Coro7* and *Pam16* are adjacent and in tandem orientation (Fig. S6), which suggests that this read-through transcription unit formed early during vertebrate evolution.

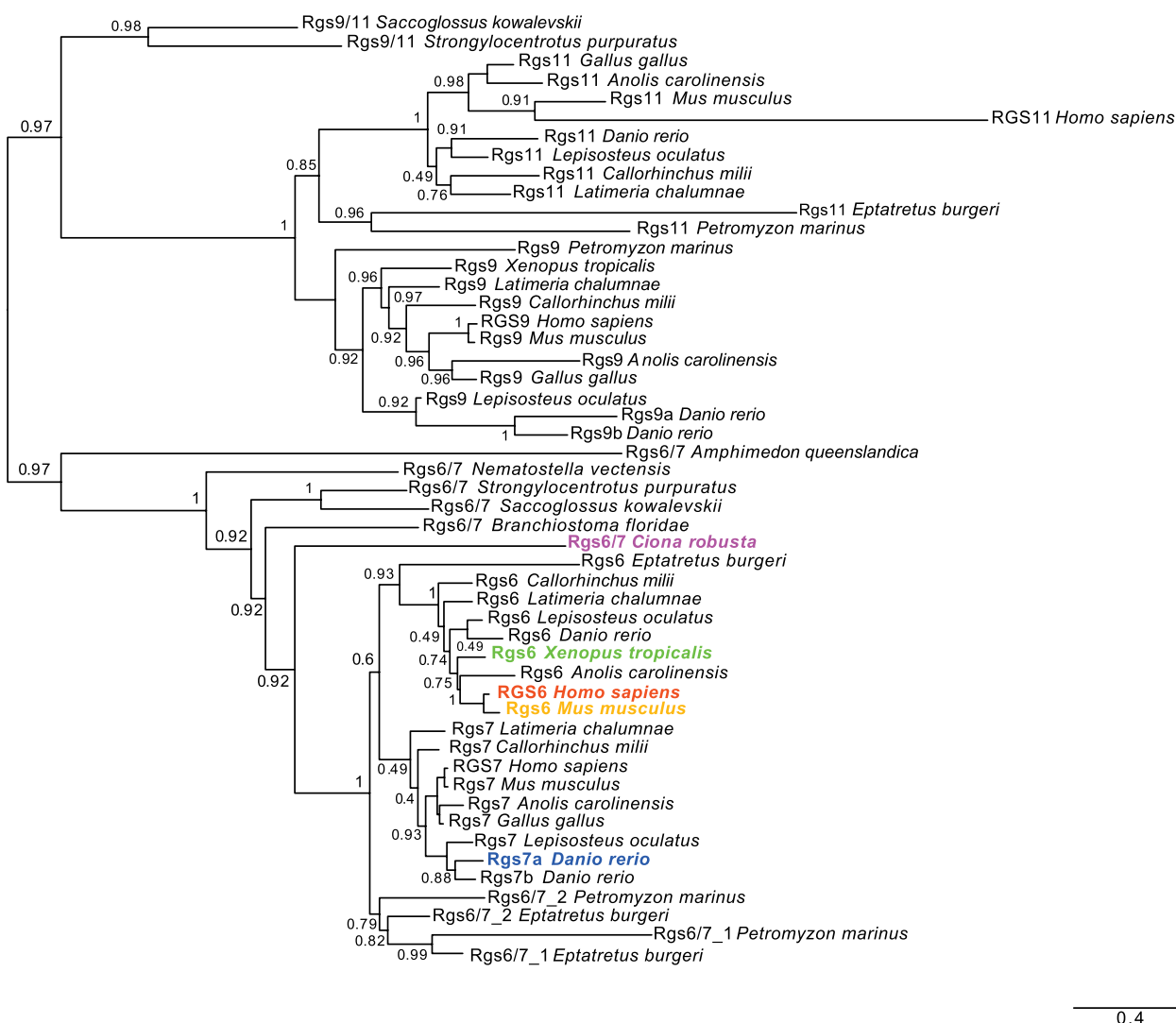


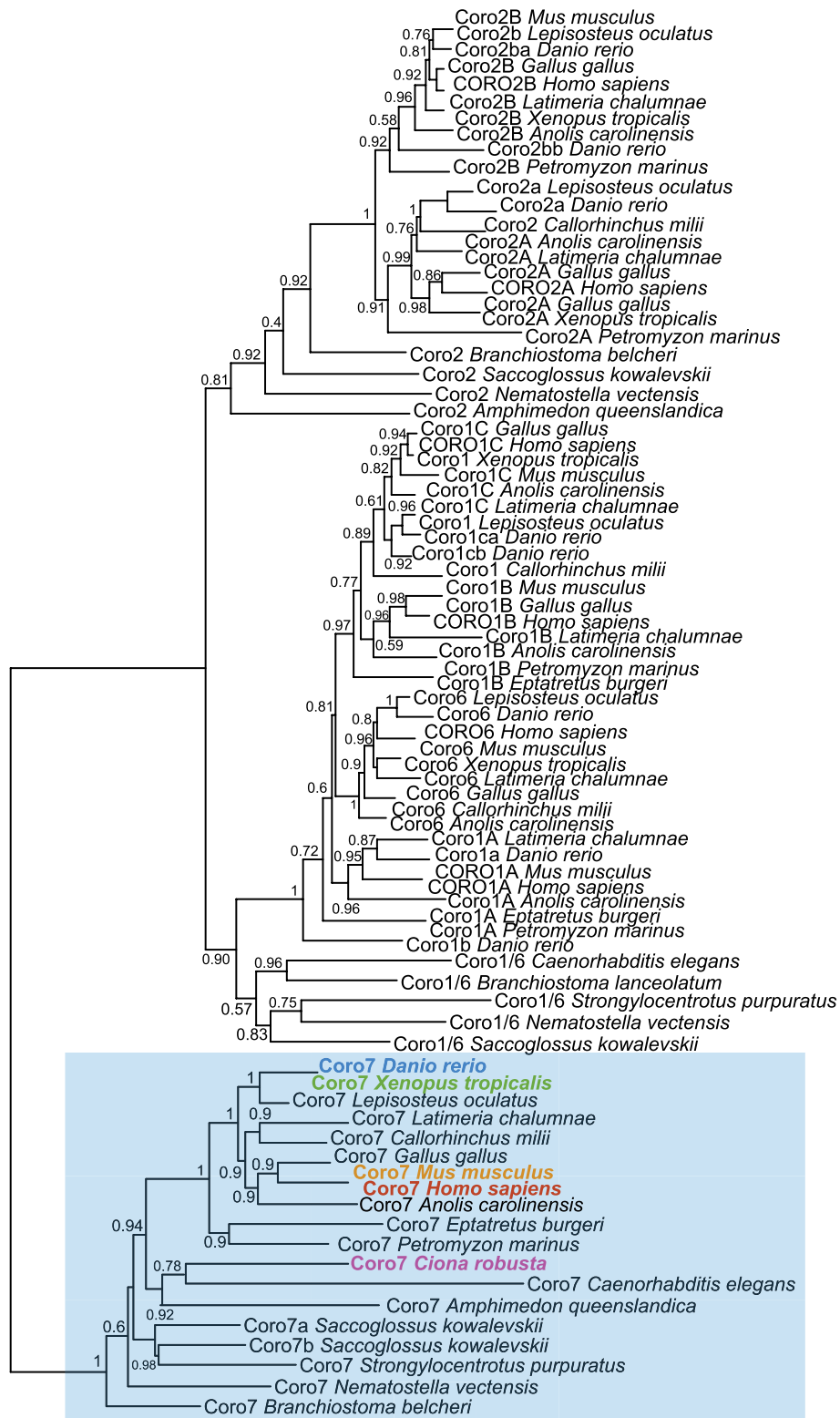
Fig. 4 Phylogenetic reconstruction of the evolutionary relationships between proteins of the Rgs6/7 and Rgs9/11 subfamilies. A global phylogenetic study of the Rgs proteins found in *Ciona robusta* is available [38]. This ML phylogenetic tree is centered on the Rgs6/7 subfamily, since the expression of some of its members has been studied here, and on the Rgs9/11 subfamily, because of its close relationship to the Rgs6/7 subfamily. Distinct colors highlight the family members whose expression was analyzed in this study. Values at the branches indicate replicates obtained using the aLRT method

Expression of zebrafish orthologs of *Ciona* notochord genes

To assess the conservation of notochord expression of vertebrate genes related to the *Ciona* notochord genes described above, we analyzed *Danio* genes that were selected on the basis of phylogenetic analyses and database searches. We found that *Danio phip* is expressed in notochord and nervous system at 30 hpf (Fig. 6A,A') and 36 hpf (Fig. 6A'').

The *rgm* family in zebrafish consists of four genes, *rgma*, *rgmb*, *rgmc* and *rgmd* (Fig. 2). While *rgma* is expressed predominantly in floor plate, developing

midbrain and hindbrain, skeletal muscle and notochord [70–73], *rgmb* is expressed in several domains of the nervous system and in developing muscle, with faint expression in the notochord during early developmental stages [57, 74], and *rgmC* is strongly expressed in the notochord and its flanking somites [72]. Hence, here we investigated the poorly characterized zebrafish *rgmd*. According to expression patterns retrieved from the Zfin database (zfin.org; [75]), expression of this gene is widespread throughout the embryo from 10–13 somites to 14–19 somites stage [74]. Our WMISH results show an unlocalized signal in the head and a similar diffuse



0.5

Fig. 5 Phylogenetic reconstruction of the evolutionary relationships within the *Coronin* gene family. ML phylogenetic tree of the *Coronin* (Coro) family. The Coro7 subfamily is highlighted by a light blue box. Distinct colors indicate the family members whose expression was analyzed in this study. Values reported at the branching points indicate replicates obtained using the aLRT method

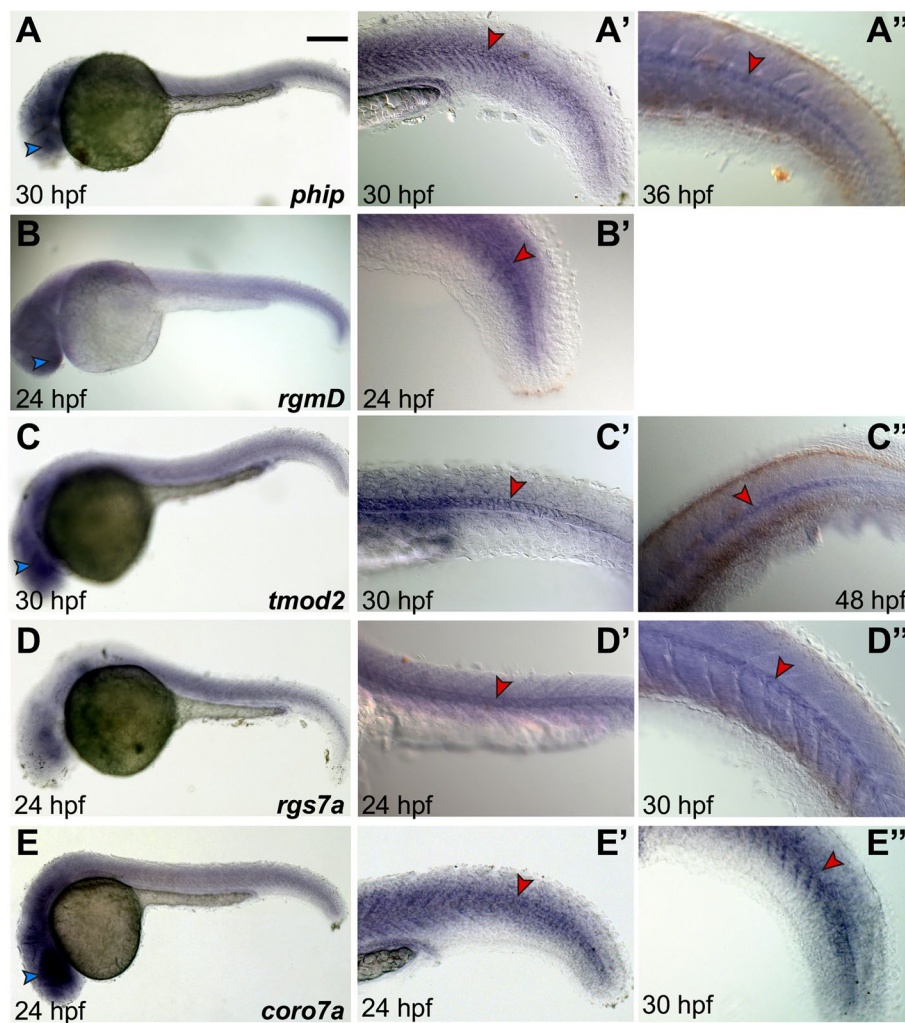


Fig. 6 Expression of *phip*, *rgmD*, *tmod2*, *rgs7a* and *coro7a* in *Danio rerio*. **A-E** Whole-mount zebrafish embryos at the stages indicated at the bottom of each panel, hybridized in situ with probes specific for the genes indicated on the bottom right of panels A-E. (**A'-E''**) Close-ups of the tails of stained embryos, either magnified from (A-E) or acquired from representative embryos from the same batch as those in (A-E). Arrowheads are color-coded as follows: blue, nervous system; red, notochord. All panels show lateral views, anterior to the left. Scale bar: 150 μ m

expression throughout the tail (Fig. 6B). However, by 24 hpf, we detected a considerable increase in notochord expression (Fig. 6B,B').

With respect to the *tmod* family, zebrafish only contains three of the four *tmod* genes reported in mammals, *tmod1*, *tmod2* and *tmod4* [76] (Fig. 3). Since *tmod1* and *tmod4* are well-characterized genes expressed in muscle cells and involved in muscle development [76, 77], here we analyzed *tmod2* and found that this gene is expressed in notochord cells at 30 hpf and 48 hpf (Fig. 6C-C').

The zebrafish genome contains three orthologs of *Ciona Rgs6/7*: *rgs6*, *rgs7a* and *rgs7b* (Fig. 4). Expression of both *rgs6* and *rgs7a* had been reported in various regions of the nervous system, and expression of *rgs7a*

had also been reported in notochord cells (zfin.org; [75]). Using specific probes for each gene, we detected expression of *rgs7a* in notochord at both 24 hpf and 30 hpf (Fig. 6D-D'). No detectable hybridization signal was observed for *rgs7b* at the stages that we analyzed; accordingly, RNA-Seq data indicate that the expression levels of this gene during early embryogenesis peak at the 128-cell and 1k-cell blastula stage, drop at the dome stage, and remain low until the late larval stages [78].

Danio coro7 is strongly expressed in various structures of the head, and in particular in the developing nervous system (Fig. 6E); expression becomes detectable in the notochord by 24 hpf (Fig. 6E') and persists at later stages (30 hpf; Fig. 6E'').

Expression of *Xenopus* orthologs of *Ciona* notochord genes

To follow the expression of the genes of interest in an additional vertebrate clade, we cloned the *Xenopus laevis* orthologs of *Ciona* notochord genes and analyzed their expression by WMISH. With the exception of one brief report describing the diffuse expression of *tmod2* at NF stage 30 [79], the expression of *Xenopus* *phip*, *rgmb*, *tmod2*, *rgs6* and *coro7* had not been previously reported. We found that at the tailbud stage (NF stage 28), *Xenopus* *phip* is enriched dorsally in the trunk, in the head region, the optic vesicles and the branchial arches (Fig. 7A), whereas *rgmb* shows strong signal in the optic and otic vesicles, with weaker expression in the branchial arches (Fig. 7B). Additionally, *tmod2* is expressed around the developing eye, in a region corresponding to the prospective trigeminal nerve (Fig. 7C). *rgs6* is detected in the brain and the otic vesicles and shows more diffuse expression in the somites (Fig. 7D), while *coro7* is expressed dorsally in the developing somites, the pronephros, the optic vesicles and the branchial arches (Fig. 7E).

To assess the expression of these genes in deeper tissues we performed transverse sections on NF stage 32 embryos, and we found that all five genes are expressed in the *Xenopus* notochord (Fig. 7A'-E').

Expression of mouse and human orthologs of *Ciona* notochord genes

To test the conservation of the expression of *Ciona* notochord genes in the notochord and in notochord descendant NP cells in mouse and human, we used the results of the phylogenetic analyses (Figs. 2, 3, 4 and 5) to identify

candidate orthologs of *Ciona* notochord genes in the *Mus musculus* and *Homo sapiens* genomes. With the exception of *Rgmd* genes, which thus far have only been described in cartilaginous fish and teleosts, all members of the *Rgm* family were identified in mouse (*Rgma*, *Rgmb*, and *Rgmc/Hjv/Hfe2*) and human (*RGMA*, *RGMB*, and *RGMC/HJV/HFE2*) (Fig. 2). Similarly, the *Tmod* family is fully represented in both mouse (*Tmod1-4*) and humans (*TMOD1-4*) (Fig. 3), and conserved members of the *Rgs* subfamilies 6, 7, 9 and 11 (Fig. 4), and *Coronin* subfamilies, including *Coronin7* orthologs (Fig. 5), are present in both mouse and humans. Previously, expression of mouse *Coro7*, also known as *Pod-1*, had been reported in the developing brain and in parts of the immune system [80].

Next, we analyzed the expression of the genes selected for this study in various stages and tissues, during embryonic and post-natal development. Published single-cell RNAseq data obtained in mouse embryos at different stages of gastrulation (E-MTAB-6967; [81]) at E6.5 (primitive streak stage), E7.25 (node/notochord formation), and E8.5 (fully formed notochord) indicated that *Phip*, *Coro1b*, *Coro1c*, and *Tmod3* were ubiquitously expressed in all embryonic structures, including primitive streak and notochord; moreover, expression of *Rgmb* and *Tmod2* was particularly robust in the notochord of E8.5 embryos (Fig. S7).

Next, using published bulk-RNAseq data from murine E12.5 notochord and post-natal day 0 (P0) NP cells (GSE100934; [82]), we found expression of all the orthologs of *Ciona* notochord genes at both stages

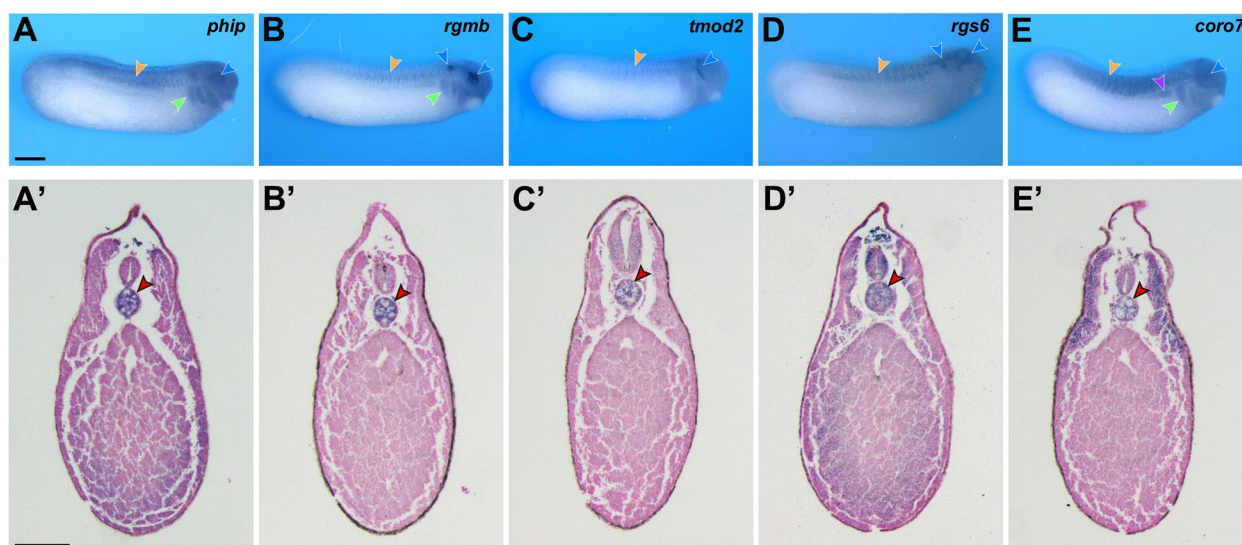


Fig. 7 Expression of *phip*, *rgmb*, *tmod2*, *rgs6* and *coro7* in *Xenopus laevis*. **A-E** Whole-mount in situ hybridization of NF stage 28 embryos. Lateral views, anterior to the right, dorsal on top. **(A'-E')** Transverse sections of NF stage 32 embryos. Dorsal is on top. Arrowheads color code: red, notochord; blue, nervous system; orange, somites; green, branchial arches; violet, pronephros. Scale bars: 500 μ m (A); 200 μ m (A')

(Fig. 8A), with the exception of *Coro1b*, which had zero reads in the database that we used. Expression of *RgmC/Hjv* (reported as *RgmC* in Fig. 2) is considerably reduced in the transition from notochord to NP cells (Fig. 8A). A reduction of the expression in NP compared to the signal in notochord cells is also seen in the case of *Rgma*, *Rgmb*, and *Coro6*, while *Rgs7*, *Rgs11*, *Coro7* and in particular *Coro1a*, display the opposite trend, being detected at higher levels in NP than in notochord (Fig. 8A). Previous studies have shown that expression of notochord genes, including *Brachyury* and *Sonic hedgehog*, is higher in neonatal mouse NP cells, but decreases with age [32, 83–85];

IVD pathologies, including those in the NP cells, become evident around 2 years of age [83, 84, 86, 87]. Therefore, we validated the expression of select members from each gene family in NP cells microdissected from the lumbar discs of neonatal (one-week), middle-age (one-year), and very aged (two-year) wild-type mice in Friend leukemia virus B (FVB) background by multiplex qPCR analysis, using gene-specific TaqMan probes, and monitoring *Gapdh* expression as a control. We observed a significant increase in expression of *Phip* from one-week to one-year old mice (Fig. 8B). *Rgmb* expression also significantly increased from one-week to one-year old (Fig. 8C). On

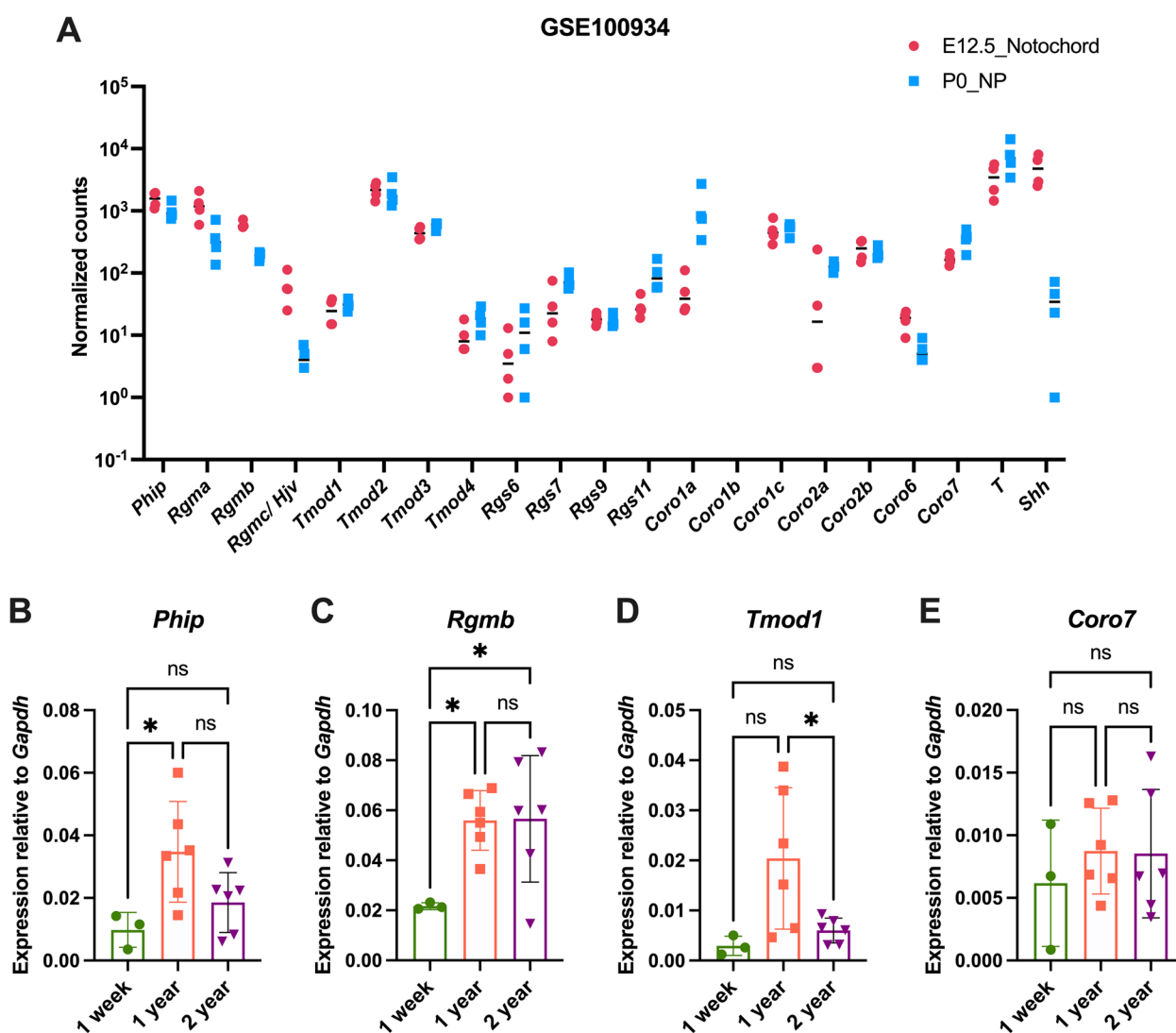


Fig. 8 Gene expression analysis in mouse *nuclei pulposi*. **A** Normalized read counts in E12.5 notochord and P0 NP cells from the GSE100934 dataset [82]. **B-E** Multiplex qPCR analysis showing the expression of *Phip* (**B**), *Rgmb* (**C**), *Tmod1* (**D**), and *Coro7* (**E**) relative to *Gapdh* in one-week ($n=3$), one-year ($n=6$) and two-year ($n=6$) old mouse NP cells from lumbar IVDs of wild-type mice. Results are presented as scatter dot plots with mean and SD for each cohort. The qPCR results plotted in (**B-E**) were analyzed by ordinary one-way ANOVA followed by Tukey’s multiple comparisons test. * $p < 0.05$

the other hand, *Tmod1* expression showed a significant decline from one-year to two-year old mice (Fig. 8D). We also tested the expression of *Rgs6*, however we did not detect this gene at any of the stages analyzed. *Coro7* expression was detected at all ages, with no significant changes in its expression (Fig. 8E).

To study the expression of the genes of interest in the human notochord and NP, we scanned the dataset obtained from microarray analyses of human notochord at 7.5, 8.5, 12 and 14 weeks post-conception (E-MTAB-6868; [88]). The normalized chip signals show that different members of the gene families analyzed here are detected at different levels in the human notochord (Fig. 9A). We next traced the expression of these genes using the bulk RNA-seq dataset from human NP cells collected from the lumbar discs of patients with two types of disc pathologies: disc herniation (DH) and degenerative spondylolisthesis (DS) (GSE146904; [89]). The normalized counts plotted in Fig. 9B show that the expression of most of the genes of interest is maintained in the postnatal NP cells of human IVDs. Of note, *RGMC/HJV/HFE2* (annotated as *RGMC/HFE2* in Fig. 9B) was expressed at very low levels.

Next, to validate the expression of our genes of interest during the progression of disc pathologies, we performed multiplex qPCR analysis on NP cells collected from human lumbar IVDs at early stage of IVD degeneration (Grade 1–3) and moderate to severe degeneration (Grade 4–5), using gene-specific TaqMan probes. The results indicate that expression of *PHIP* was significantly reduced with increased IVD pathologies severity (Fig. 9C). Expression of *RGMB*, *TMOD1*, and *RGS6* was detected in both cohorts, although the expression of *RGMB* was very low (Fig. 9D–F). As in the case of *PHIP*, the expression of *CORO7* showed a significant decline that correlated with the increased severity of IVD pathologies (Fig. 9G).

Discussion

The notochord is a vital structure conserved throughout half a billion years of chordate evolution. During this time, the number of genes and the composition of the gene families expressed in the notochord and in other chordate hallmarks have been shaped by the two rounds of whole-genome duplication (WGD) seen in vertebrates, by an independent third WGD event specific to the teleost lineage, and by isolated lineage-specific gene duplication events, often counteracted by lineage-specific gene losses [64, 65, 90–92]. These events complicate the elucidation of the evolutionarily conserved complement of genes that confer to each structure its distinctive features. In this study, we sought to use the information on notochord genes that we had gained through a survey of

notochord genes in *Ciona* to identify genes that could potentially be expressed in the notochord of different chordates and/or in the notochord remnants that compose the NP of murine and human IVDs. As a proof of principle, we have followed the evolutionary trajectories of five *Ciona* notochord genes in different vertebrates. We selected for this study one gene, *Phip*, which is not part of a multigenic family, and four genes that in vertebrates have become part of multigenic families. In *Ciona*, *Phip* is expressed in notochord, in mesenchymal cells, which are a population of small cells that give rise to most of the post-metamorphic tissues, and in the sensory vesicle, the region where the ocellus, a photoreceptive structure, and the otolith, a statocyst, are located. Interestingly, in addition to being expressed in the notochord, both *Danio* and *Xenopus phip* genes are detected in the developing eyes, and mouse *Phip* is reportedly expressed in the notochord, developing retina and inner ear [93]. Human *PHIP* is expressed in both notochord and NP.

The single-copy *Rgma/B/C/D* gene found in invertebrates has duplicated in vertebrates to give rise to four paralogs, *Rgma*, *Rgmb*, *Rgmc* and *Rgmd*; however, only cartilaginous fish and teleosts seem to have the *Rgmd* paralog [58], as confirmed by our synteny analysis. In *Ciona*, neural precursors are the earliest site of expression of *Rgma/B/C/D*, and it is only at the late tailbud stage that expression in notochord cells becomes detectable. On the other hand, expression in the sensory vesicle remains strong during early embryogenesis, and widens as the embryos progress through the tailbud stages. Interestingly, in addition to the expression in neurons during CNS development, notochord expression had been reported for three of the four zebrafish *rgm* genes, and the present study indicates that also the fourth zebrafish *rgm* gene, *rgmd*, is expressed in this structure. Similarly, *Xenopus rgmb* is strongly expressed in notochord, and also in the optic and otic vesicles; hence, the expression patterns of these genes in these vertebrates recapitulate the expression of the single-copy *Ciona* gene. *Rgmb* is also expressed in mouse and human notochord and NP, indicating conservation of expression in pre- and post-natal notochord cells of mammals.

The *Tmod* gene family is limited to a single member in most invertebrates, and has expanded to four components in vertebrates. In amphioxus, which is considered more distant than *Ciona* from vertebrates [10], there is only one *Tmod* gene, expressed in striated muscle [94]. Hence, the presence of a *Tmod* and a *Lmod* gene in *Ciona* suggests that the *leiomodin* genes were formed from *tropomodulin* genes via a duplication that occurred in the common ancestor of Olfactores (tunicates and vertebrates) [95]. Together with the reported expression of both *Tmod* and *Lmod* genes in the notochord of *Xenopus*

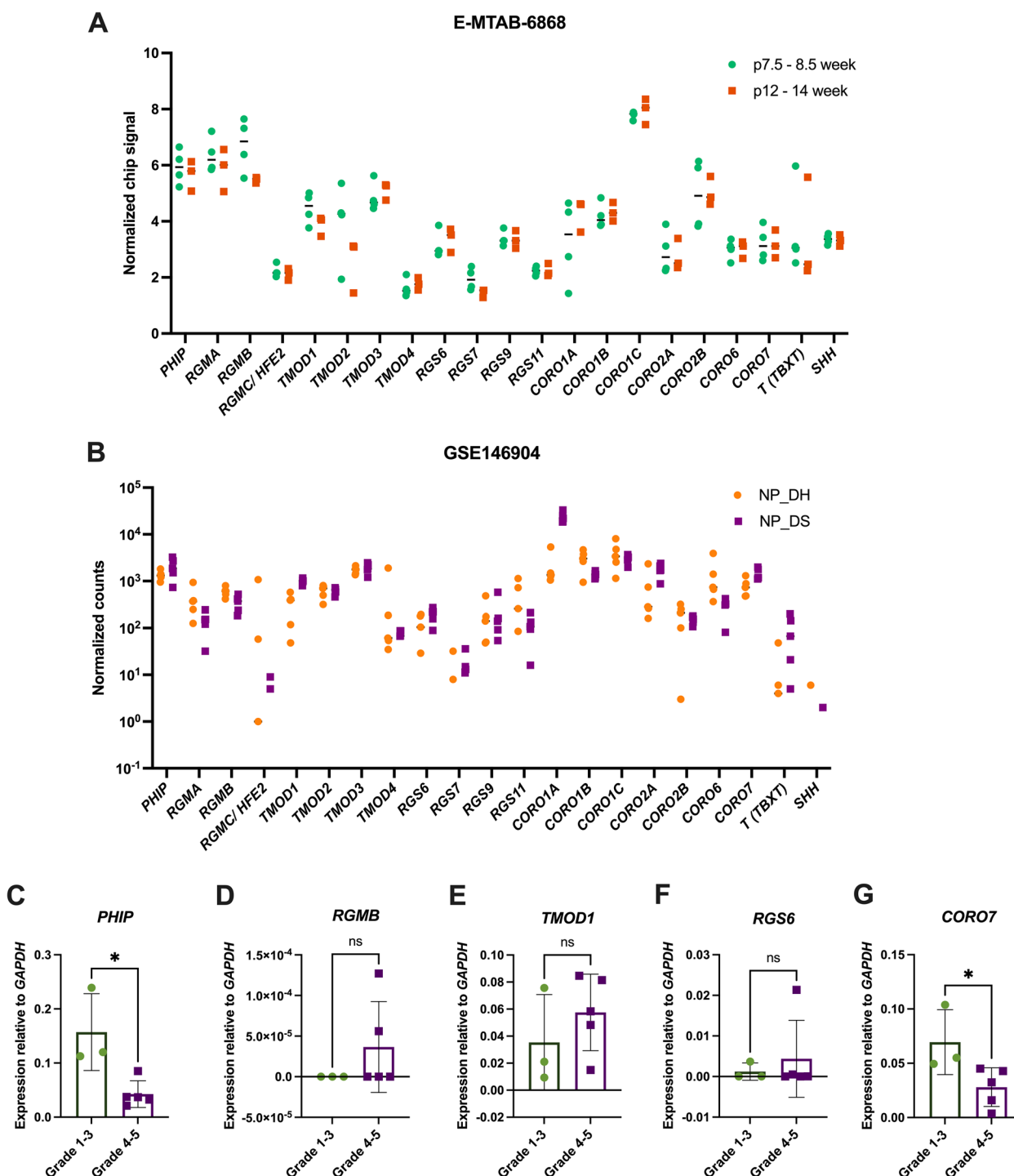


Fig. 9 Expression of the genes of interest in human *nuclei pulposi*. **A** Normalized chip signals plotted as gene expression measures from microarray data (E-MTAB-6868; [88]) obtained from notochordal cells from human embryonic (7.5–8.5 weeks post-conception, $n=3$) and fetal (12–14 weeks post-conception, $n=2$) stages. **B** Log₂ normalized read counts plotted as gene expression measures from the RNA-seq data of NP samples from lumbar disc herniation (DH, $n=5$), and lumbar disc spondylolisthesis (DS, $n=5$) obtained from the GEO database (GSE146904; [89]). Expression of PHIP (**C**), RGMB (**D**), TMOD1 (**E**), RGS6 (**F**), and CORO7 (**G**) relative to GAPDH, measured by qPCR analysis in the NP tissue collected from less degenerated (Grade 1–3) or moderately to severely degenerated (Grade 4–5) lumbar disc from human male and female of various age groups (see [Methods](#)). Results are presented as scatter dot plots with mean and standard deviation for each cohort. The qPCR results plotted in (C–G) were analyzed by unpaired *t*-test. * = $p < 0.05$. p, post-conception

[79], our results suggest that the notochord expression might have been an ancestral feature shared by *Tmod* and *Lmod* genes, which has been lost by *Ciona Lmod*. To study this gene family in vertebrates, we analyzed *Tmod2* paralogs from *Danio* and *Xenopus*, and found that both are expressed in notochord cells. When we tested expression of *Tmod1* in mouse and human NP, we found that also in these mammals *Tmod1* orthologs are expressed in these structures at high levels.

Rgs6, *Rgs7*, *Rgs9* and *Rgs11* are highly related to one another, and are predominantly expressed in neurons [39]. In particular, *Rgs6* is reportedly the only member of the *Rgs* family able to inhibit the function of different receptors involved in neurotransmission, and for this reason its deletion is associated with phenotypes ranging from problems with the parasympathetic regulation of the heart rate to neuropsychiatric disorders [96–98]; loss-of-function mutations in this gene can also lead to cancer growth, due to the inactivation of its tumor-suppressing ability [99]. Studies carried out in human mesenchymal cells indicate that *RGS5*, *RGS7* and *RGS10* promote chondrogenesis, while *RGS4* inhibits this process [100]. Our phylogenetic reconstruction suggests that *Ciona Rgs6/7* is more closely related to vertebrate *Rgs6* and *Rgs7* compared to *Rgs6/7* genes from amphioxus and non-chordate invertebrates. Together with the role of *RGS7* in chondrogenesis and the evolutionary and histological relationships between notochord and cartilage [101], these findings suggest that *Ciona Rgs6/7* might contribute to notochord formation. Through this comparative study we elucidated the expression of *Xenopus Rgs6* in notochord, brain, and otic vesicles, which is reminiscent of the expression of *Ciona Rgs6/7*. Although published bulk RNA-Seq and microarray data suggest that *Rgs6* orthologs are expressed in the mouse and human notochord, respectively, we were able to validate expression of *RGS6* in human NP but not in mouse NP, which might be indicative of a difference in the molecular composition of the NP during the postnatal stages of these two species.

Coronins have been associated with a plethora of cellular processes, including auto-immunity, neuronal development and cancer progression [43]; however, a direct involvement of any member of this large family in notochord formation is yet to be reported. *Coronin 1A*, one of the best-characterized members of this family, is reportedly expressed in osteoclasts, where it functions as a regulator of bone resorption [102]. Recent studies have determined that in both *Drosophila* and mammalian cells, *Coro7* interacts with core components of the Hippo pathway, and is required for its activation in response to various stimuli, including cell–cell contact [103]. We found that the closely related *Coro7* genes of *Danio* and *Xenopus* are both expressed in notochord

cells, akin to the *Ciona* gene, and that their mouse and human counterparts are expressed in both notochord and postnatal NP.

Except for the case of *Ciona Rgs6/7*, which is downregulated in the primary notochord during the tailbud stages, we did not detect differences in gene expression along the rostral-caudal axis of the notochord in any of the species analyzed here. Regional differences in the expression of notochord genes appeared early in the chordate evolutionary timeline, having been reported in both tunicates [104, 105] and cephalochordates [106, 107], and they have been proposed to regulate regional morphogenesis in the mouse notochord [108]. Since we relied on RNA-Seq data in the case of mouse and human embryos, it remains to be determined whether any of the genes from the current study displays regional differences in its expression in developing mammals.

Interestingly, our analysis of differential gene expression in mouse NP uncovered a significant increase in the expression of *Phip* and *Rgmb* associated with aging. Since *Phip* proteins are modulators of the insulin pathway [33, 34], and *Rgm* proteins are co-receptors and regulators of the BMP signaling pathway [109], these results could be interpreted as part of a compensatory mechanism aimed at counteracting NP senescence. On the other hand, the decline of *Tmod1* expression could represent a read-out of the aging process. We also found that in human NP clinical samples, *PHIP* and *CORO7* expression significantly declined as NP degeneration increased; while the reasons and consequences of these changes in gene expression remain to be addressed, these findings open the possibility that these genes could become novel candidate diagnostic markers of human NP degeneration.

In conclusion, the results of this study add new candidate components to the notochord genetic toolkit shared by divergent chordates, and have uncovered changes in the expression of some of these genes that might be associated either with the transition from notochord to NP or with NP aging and degeneration.

Materials and methods

Ciona robusta embryo culture and whole-mount in situ hybridization (WMISH)

Adult *Ciona robusta* were purchased from Marine Research and Educational Products (M-REP; Carlsbad, CA) and kept at 16 °C in recirculating artificial seawater. Embryo cultures, fixation and staining were carried out as previously described [104]. Embryos were staged following the developmental timeline established in Hotta et al. [110]. Digoxigenin-labeled antisense RNA probes were synthesized in vitro using as templates *Ciona robusta* EST clones [111–113] (Table S1) linearized through appropriate restriction enzymes. WMISH experiments

were performed as previously described [46, 104], using a hybridization temperature of 42 °C. After signal detection was satisfactorily completed (~4–48 h.), embryos were rinsed in 100% ethanol, washed briefly in xylenes, mounted in Permount (ThermoFisher Scientific, Waltham, MA) and photographed using a Leica DMR microscope (Leica Microsystems Inc., Buffalo Grove, IL).

Scanpy analysis of single-cell RNA-Seq datasets

We utilized the single-cell RNA sequencing (scRNA-Seq) dataset available for *Ciona robusta* developing embryos from gastrula to larva stages [50]. We performed this analysis on the normalized and log-transformed notochord gene expression matrix available in the Gene Expression Omnibus (GEO) under accession number GSE131155, using the Scanpy (Single-cell analysis in Python <https://scanpy.readthedocs.io/en/stable/>) toolkit for the visualization of single-cell gene expression data [114].

Evolutionary analyses

The protein sequences used for the phylogenetic surveys were retrieved from the NCBI and Ensembl databases using *Ciona robusta* proteins as initial queries for tBLASTn searches of the genomes of the organisms included in Figs. 2, 3, 4 and 5 [115]; reciprocal BLAST searches were performed using the Aniseed/WashU *Ciona robusta* genome browser (<https://www.aniseed.fr>) [52]. The sequences selected for phylogenetic analyses and their corresponding accession numbers are listed in Supplementary files 1–5. Sequence orthology was initially assessed using the reciprocal best BLAST hit approach, utilizing default parameters, and was later corroborated by phylogenetic analyses. The protein sequences were aligned by ClustalW using default parameters [116]. Phylogenetic trees were computed with the Maximum Likelihood (ML) inferences using PhyML 3 [117], employing automatic Akaike Information Criterion (AIC) by Smart Model Substitution (SMS) [118], which selected the Jones-Taylor-Thornton (JTT) substitution model, 0.4 as the proportion of invariable sites (I) and 4 as the gamma distribution parameter (γ) [119]. Branch support was provided by aLRT (approximate likelihood ratio test) [120]. Domain analyses were carried out employing the PROSITE database [121] and InterPro software [122].

First-pass synteny analyses were carried out using the Genomicus genome browser [123] (<https://www.genomicus.bio.ens.psl.eu/genomicus-109.01/cgi-bin/search.pl>), using a window of twenty genes. The results were cross-referenced and detailed using species-specific UCSC (<https://genome.ucsc.edu>) and Ensembl (<https://www.ensembl.org>) genome browsers.

Zebrafish handling, probe synthesis and WMISH

Zebrafish (*Danio rerio*) embryos were obtained from natural spawning of wild-type animals. The embryos were fixed overnight in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS), washed three times in pre-chilled 1×PBT (PBS/0.1% Tween), then three times in cold methanol, and stored in methanol at -20 °C until use. All the protocols for handling of zebrafish and experiments that involve non-feeding larvae were approved by the local review panel. The sequences of the zebrafish genes of interest were retrieved from the NCBI database using the corresponding *C. robusta* coding regions as queries for BLASTn searches [124]. ESTs were found only for *rgs7a* and *tmod2*; templates for RNA probe synthesis for the remaining genes were cloned using the oligos listed in Table S2. PCR-amplified gene fragments were cloned into the pGEM[®]-T Easy vector (Promega, Madison WI) and 500 ng of purified template DNA were used for in vitro transcription of digoxigenin-labeled RNA probes with SP6 and T7 RNA polymerases (Roche, Indianapolis, IN). All RNA probes were purified using 4 M lithium chloride and stored in formamide at -80 °C until use.

WMISH was carried out as previously described [74, 125]. In short, embryos were re-hydrated and permeabilized through digestion with Proteinase K (10 µg/ml), followed by five washes in PBT. After 1 h of post-fixation at RT in 4% PFA dissolved in PBS, embryos were rinsed with PBT four times and hybridized overnight at 65 °C in hybridization buffer [125]. After the hybridization solution was removed, embryos were washed several times in maleic acid buffer and incubated overnight at 4 °C with anti-digoxigenin-AP antibody (Roche, Indianapolis, IN). The staining reaction was performed at room temperature employing BM Purple (Roche, Indianapolis, IN). Images of stained embryos were captured using a Zeiss Axio Imager M1.

Xenopus laevis handling, probe synthesis, WMISH and histology

Xenopus laevis embryos were staged according to Nieuwkoop and Faber [126] and raised in 0.1X NAM (Normal Amphibian Medium; [127]). All the procedures used for these experiments were approved by the New York University Institutional Animal Care and Use Committee (IACUC animal protocol #150,201).

Xenopus laevis *phip.S*, *coro7.S*, *tmod2.L*, *rgmb.S* and *rgs6.S* were amplified by PCR (S100 Thermal Cycler; Bioprad, Hercules, CA) from NF stage 11.5 (*phip.S*, *coro7.S* and *rgmb.S*) or NF stage 25 (*tmod2.L* and *rgs6.S*) cDNA with the primer sets described in Table S3, using Illustra PuReTaq[™] Ready-To-Go[™] PCR Beads (GE Healthcare, Chicago, IL). The PCR conditions were as follows:

denaturation at 95 °C (30 s), annealing at 60 °C (60 s) and extension at 72 °C (90 s) for 35 cycles. The PCR products recovered were cloned into pGEM[®]-T Easy (Promega, Madison, WI), sequenced, and linearized to generate sense and antisense *in situ* hybridization probes.

Embryos at the appropriate developmental stages (NF stage 28 and 32) were fixed in MEMFA (0.1 M 3-N-Morpholino-propanesulfonic acid pH 7.4, 2 mM EGTA, 1 mM MgSO₄ and 3.7% formaldehyde), and processed for *in situ* hybridization. For each gene, sense and antisense digoxigenin-labeled probes (Genius kit; Roche, Indianapolis, IN) were synthesized using the corresponding linearized pGEM[®]-T Easy construct. WMISH was performed as described [128, 129]. For histology, NF stage 32 stained embryos were embedded in Paraplast+ (Sigma-Aldrich, St. Louis MO), sectioned (12 µm) on a rotary microtome (Cut4060; Olympus, Center Valley, PA), counterstained with Eosin Y (Sigma-Aldrich, St. Louis MO) and mounted in Permount (ThermoFisher Scientific, Waltham, MA). Embryos and sections were imaged on a Leica M165 Stereomicroscope (Leica Microsystems Inc., Buffalo Grove, IL). Staining was confirmed on four different batches of embryos.

Mouse nucleus pulposus cell collection

Wild-type female and male in FVB background mice used in these studies were maintained in a temperature-controlled facility with equal light–dark cycle and food and water provided *ad libitum*, in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. All experiments were carried out in accordance with institutional guidelines under IACUC approved at Weill Cornell Medical College (WCMC) under the IACUC protocol number 2016–0026.

FVB mice at one week of age ($n=3$), one-year ($n=6$), and two-years ($n=6$) of age were used to analyze the expression of *Phip*, *Rgmb*, *Tmod1*, *Rgs6*, and *Coro7* genes in NP cells. The NP cells were microdissected from the IVDs of lumbar spine in cold PBS underneath a Nikon bright-field stereomicroscope (Nikon, Japan) as we described recently [130]. The NP cells from each biological replicate were directly collected in RNAlater[™] (Invitrogen by Thermo Fisher Scientific, Lithuania, AM7024) and stored at 4 °C for 24 h.

Human nucleus pulposus tissue collection

NP tissue was collected under the Hospital for Special Surgery (HSS) Institutional Review Board (IRB) approved research study and protocol number 2016–933, all in compliance with the applicable requirements of the FDA regulations and HSS regulations. Patients who were recruited for this study were undergoing spine surgery

due to prior medical diagnosis and treatment. Informed consent was obtained to collect the NP tissue, which otherwise would have been discarded after the surgical procedure. The T2-weighted MRI image taken prior to surgery was used to assess the Pfirrmann grade of the disc. Pfirrmann grading [131] is the gold standard for the quantification of disc health based on water content. Grade 1–2 IVDs are considered healthy; Grade 3 IVDs show early sign of disc degeneration and Grade 4–5 IVDs are moderately to severely degenerated. A total of eight samples were analyzed in the current study from males and females between 19–79 years of age. In the present study, we combined NP samples from Grade 1–3 ($n=3$) in one cohort, and compared them to NP samples from Grade 4–5 ($n=5$). Following surgery, the samples were immediately stored on ice and delivered to the lab, where they were weighed, washed three times in PBS, and stored in RNAlater[™] (Invitrogen by Thermo Fisher Scientific, Lithuania, AM7024) at 4 °C for 24 h.

Mouse and human nucleus pulposus RNA isolation and qPCR analysis

Total RNA was isolated by extraction in TRI-reagent (Sigma-Aldrich, St. Louis MO), followed by purification and elution using Qiagen RNeasy Kit (Qiagen, Germany; cat. #74,004 for mouse, and cat. #74,707 for human cells), as previously described [130]. RNA concentration was quantified in duplicates using a NanoDrop[™] One Micro-volume UV–Vis Spectrophotometer (Thermo Scientific, USA, AZY1601393). The RNA was immediately converted into cDNA using the SuperScript[™] IV First-Strand Synthesis System (Invitrogen by Thermo Fisher Scientific, Lithuania, 18,091,050). Multiplex qPCR was performed using the CFX96 Touch[™] Real-Time PCR Detection System (Bio-Rad, Singapore, 1,855,195). Each reaction used 8 ng of cDNA, iQ[™] Multiplex Powermix (Bio-Rad, USA, 1,725,849) master mix, gene- and species-specific TaqMan[™] probes (Table S4) conjugated to FAM-MGB (Thermo Scientific, USA), and an internal control (*Gapdh*) conjugated to VIC-MGB (Thermo Scientific, USA). The data are presented as median and relative to the reference gene. The Cq values obtained for each gene were subtracted from the Cq values of *Gapdh* to obtain the delta Cq values. The logarithm of the -deltaCq value was calculated to plot the expression of each gene relative to *Gapdh*. The qPCR results in mouse NP cells compared three age points (one week, one and two year; see above) and were analyzed by ordinary one-way ANOVA followed by Tukey's multiple comparisons test to determine statistical differences in gene expression among different age groups. The qPCR results from human NP cells collected from Grade 1–3 and Grade 4–5 were analyzed using unpaired *t*-test.

Analysis of gene expression using existing mouse and human datasets

Single-cell data were obtained as SingleCellExperiment objects from the Bioconductor package [132]. The raw sequencing data in the MouseGastrulation-Data package were acquired from ArrayExpress (accession E-MTAB-6967), analyzed using Seurat [133] and processed as described in Pijuan-Sala et al., 2019 [81].

RNA sequencing (RNA-seq) data from mouse notochord cells at E12.5 and NP cells at P0 were obtained from the Gene Expression Omnibus (GEO) database (GSE100934) [82]; read counts were normalized using the DESeq2 method [134], and the normalized values are plotted as gene expression measures. Microarray data (E-MTAB-6868; [88]) obtained from notochordal cells from human embryonic (7.5–8.5 week, $n = 3$) and fetal (12–14 weeks, $n = 2$) stages post-conception were downloaded and the intensity values were extracted and normalized. The normalized gene expression values were plotted as gene expression measures. RNA-seq data from 10 human NP samples of lumbar degenerated discs (DH, herniated and DS, spondylolisthesis; $n = 5$ for each tissue) were obtained from the GEO database (GSE146904; [89]), and Log2 normalized read counts were plotted as gene expression measure.

Abbreviations

aLRT	Approximate likelihood ratio test
ANOVA	Analysis of variation
cDNA	Complementary DNA
CNS	Central nervous system
EST(s)	Expressed Sequence Tag(s)
FVB	Friend leukemia virus B
hpf	Hours post fertilization
hr(s)	Hour(s)
kb	Kilobase(s), or 1000 base pairs
ML	Maximum likelihood
NAM	Normal Amphibian Medium
NF	Nieuwkoop Faber
NP	<i>nucleus pulposus/nuclei pulposi</i>
nt	Nucleotide(s)
ORF	Open reading frame
(q)PCR	(Quantitative) Polymerase Chain Reaction
P0	Post-natal day zero
PFA	Paraformaldehyde
PBS	Phosphate-buffered saline
PBT	PBS/0.1% Tween
WGD	Whole-genome duplication
WMISH	Whole-mount in situ hybridization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-023-02167-1>.

Additional file 1: Figure S1. Alignment of Phip protein sequences.

Additional file 2: Figure S2. Comparative view of the genomic context of the single-copy *RgmA/B/C/D* genes found in hemichordates and tunicates and of *RgmB* genes of vertebrates.

Additional file 3: Figure S3. Comparative view of the genomic context of vertebrate *RgmA* and *RgmC* genes.

Additional file 4: Figure S4. Comparative view of the genomic context of the fish-specific *RgmD* genes.

Additional file 5: Figure S5. Comparative view of the genomic context of vertebrate *Tmod2* and *Tmod3* genes.

Additional file 6: Figure S6. Comparative view of the genomic context of hemichordate, tunicate and vertebrate *Coro7* genes.

Additional file 7: Figure S7. Expression of members of the gene families analyzed in this study during select stages of mouse gastrulation.

Additional file 8: Table S1. *Ciona robusta* gene models and ESTs.

Additional file 9: Table S2. *Danio rerio* gene models, ESTs and cloning primers.

Additional file 10: Table S3. *Xenopus laevis* gene models and cloning primers.

Additional file 11: Table S4. Mouse and human gene models and probe IDs.

Additional file 12: Table S5. ENSEMBL gene models for the mouse genes in Fig. S7.

Additional file 13: Supplemental file 1. Phip protein sequences used in Fig. S1, with their respective accession numbers.

Additional file 14: Supplemental file 2. Rgm protein sequences used in Fig. 2, with their respective accession numbers.

Additional file 15: Supplemental file 3. Tmod protein sequences used in Fig. 3, with their respective accession numbers.

Additional file 16: Supplemental file 4. Rgs protein sequences used in Fig. 4, with their respective accession numbers.

Additional file 17: Supplemental file 5. Coronin protein sequences used in Fig. 5, with their respective accession numbers.

Acknowledgements

This work benefited from the support of Xenbase (<http://www.xenbase.org/>—RRID: SCR_003280).

Authors' contributions

RR, U.C., Y.W., C.I. contributed equally to the research. RR, U.C., Y.W., C.I., L.J.N.-P., J.E.M., J.H., A.M.A. carried out experiments, interpreted results and prepared tables and figures. M.C., H.J.K., T.J.A., provided human samples. A.M.A. performed computational analysis of human and mouse high-throughput gene expression datasets including single-cell, bulk RNAseq and microarray. J.-P.S.-J., F.R., C.L.D. and A.D.G. directed research, supervised experiments, interpreted results, and wrote parts of the manuscript. A.D.G. wrote the manuscript's draft and prepared some of the figures. The author(s) read and approved the final manuscript.

Funding

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R03HD098395 and R03HD098395-02S1 to ADG, and by a pilot grant to A.D.G. and J.P.S.-J. from the New York University Center for Skeletal and Craniofacial Biology, which was established by NIH grant 1P30DE020754. This research was also supported by the Department of Molecular Pathobiology Accelerator Award B01 2020 to A.D.G. U.C. was supported by American Heart Association (AHA) postdoctoral fellowship 831018. L.J.N.-P. was supported in part by NIH training grant T32HD007520. J.E.M. was supported by NIH postdoctoral fellowship 1F32GM123700.

Research reported in this publication was supported by the National Institute Of Arthritis And Musculoskeletal And Skin Diseases of the National Institutes of Health under Award Number R01AR077145, National Institute On Aging of the National Institutes of Health under award number R01AG070079, and Office of the Director of the National Institutes of Health under award number S10OD026763 to C.L.D. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

All data generated during this study are included in this published article and its supplementary information files. Data sharing is not applicable to this article as no datasets were generated during the current study.

Declarations**Ethics approval and consent to participate**

All methods were performed in accordance with the relevant guidelines and regulations.

All the protocols for handling of zebrafish and experiments that involve non-feeding larvae were approved by the local review panel.

The work involving *Xenopus laevis* was performed in accordance with the recommendations of the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, and was approved by the Institutional Animal Care and Use Committee of New York University, protocol #IA16-00052.

All experiments involving mice were carried out in accordance with institutional guidelines under Institutional Animal Care and Use Committee (IACUC) approved at Weill Cornell Medical College (WCMC) under the IACUC protocol number 2016-0026.

Human NP tissue was collected under the Hospital for Special Surgery (HSS) Institutional Review Board (IRB) approved research study and protocol number 2016-933, all in compliance with the applicable requirements of the FDA regulations and HSS regulations. Patients who were recruited for this study were undergoing spine surgery due to prior medical diagnosis and treatment. Informed consent was obtained to collect the NP tissue, which otherwise would have been discarded after the surgical procedure.

Consent for publication

N/A.

Competing interests

The authors declare no competing interests.

Received: 4 June 2023 Accepted: 8 August 2023

Published online: 27 October 2023

References

- Adams DS, Keller R, Koehl MA. The mechanics of notochord elongation, straightening and stiffening in the embryo of *Xenopus laevis*. *Development*. 1990;110(1):115–30.
- Annona G, Holland ND, D'Aniello S. Evolution of the notochord. *EvoDevo*. 2015;6:30.
- Cleaver O, Krieg PA. Notochord patterning of the endoderm. *Dev Biol*. 2001;234(1):1–12.
- Di Gregorio A. The notochord gene regulatory network in chordate evolution: Conservation and divergence from *Ciona* to vertebrates. *Curr Top Dev Biol*. 2020;139:325–74.
- Stemple DL. Structure and function of the notochord: an essential organ for chordate development. *Development*. 2005;132(11):2503–12.
- Lawson L, Harfe BD. Notochord to Nucleus Pulposus Transition. *Curr Osteoporos Rep*. 2015;13(5):336–41.
- Matta A, Erwin WM. Current status of the instructional cues provided by notochordal cells in novel disc repair strategies. *Int J Mol Sci*. 2021. 23(1).
- Mohanty S, Dahia CL. Defects in intervertebral disc and spine during development, degeneration, and pain: New research directions for disc regeneration and therapy. *Wiley Interdiscip Rev Dev Biol*. 2019;8(4):e343.
- Bach FC, Poramba-Liyanage DW, Riemers FM, Guicheux J, Camus A, Iatridis JC, Chan D, Ito K, Le Maitre CL, Tryfonidou MA. Notochordal cell-based treatment strategies and their potential in intervertebral disc regeneration. *Front Cell Dev Biol*. 2021;9:780749.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*. 2006;439(7079):965–8.
- Satoh N, Tagawa K, Takahashi H. How was the notochord born? *Evol Dev*. 2012;14(1):56–75.
- Developmental biology of ascidians. Satoh N, eds. New York: Cambridge University Press. 1994.
- Ellis K, Hoffman BD, Bagnat M. The vacuole within: how cellular organization dictates notochord function. *BioArchitecture*. 2013;3(3):64–8.
- Babić MS. Development of the notochord in normal and malformed human embryos and fetuses. *Int J Dev Biol*. 1991;35(3):345–52.
- Paavola LG, Wilson DB, Center EM. Histochemistry of the developing notochord, perichordal sheath and vertebrae in Danforth's short-tail (sd) and normal C57BL/6 mice. *J Embryol Exp Morphol*. 1980;55:227–45.
- Peacock A. Observations on the prenatal development of the intervertebral disc in man. *J Anat*. 1951;85(3):260–74.
- Peacock A. Observations on the postnatal structure of the intervertebral disc in man. *J Anat*. 1952;86(2):162–79.
- Gansner JM, Gitlin JD. Essential role for the alpha 1 chain of type VIII collagen in zebrafish notochord formation. *Dev Dyn*. 2008;237(12):3715–26.
- Götz W, Osmers R, Herken R. Localisation of extracellular matrix components in the embryonic human notochord and axial mesenchyme. *J Anat*. 1995;186 (Pt 1)(Pt 1):111–21.
- Miyamoto DM, Crowther RJ. Formation of the notochord in living ascidian embryos. *J Embryol Exp Morphol*. 1985;86:1–17.
- Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. Evidence of en bloc duplication in vertebrate genomes. *Nat Genet*. 2002;31(1):100–5.
- Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*. 2005;3(10):e314.
- Anno C, Satou A, Fujiwara S. Transcriptional regulation of *ZicL* in the *Ciona intestinalis* embryo. *Dev Genes Evol*. 2006;216(10):597–605.
- Coppola U, Kamal AK, Stolfi A, Ristoratore F. The cis-regulatory code for Kelch-like 21/30 specific expression in *Ciona robusta* sensory organs. *Front Cell Dev Biol*. 2020;8:569601.
- Coppola U, Ristoratore F, Albalat R, D'Aniello S. The evolutionary landscape of the Rab family in chordates. *Cell Mol Life Sci*. 2019;76(20):4117–30.
- Di Gregorio A, Spagnuolo A, Ristoratore F, Pischetola M, Aniello F, Branno M, Cariello L, Di Lauro R. Cloning of ascidian homeobox genes provides evidence for a primordial chordate cluster. *Gene*. 1995;156(2):253–7.
- Kugler JE, Gazdoui S, Oda-Ishii I, Passamaneck YJ, Erives AJ, Di Gregorio A. Temporal regulation of the muscle gene cascade by *Macho1* and *Tbx6* transcription factors in *Ciona intestinalis*. *J Cell Sci*. 2010;123(Pt 14):2453–63.
- Kugler JE, Kerner P, Bouquet JM, Jiang D, Di Gregorio A. Evolutionary changes in the notochord genetic toolkit: a comparative analysis of notochord genes in the ascidian *Ciona* and the larvacean *Oikopleura*. *BMC Evol Biol*. 2011;11:21.
- Capellini TD, Dunn MP, Passamaneck YJ, Selleri L, Di Gregorio A. Conservation of notochord gene expression across chordates: insights from the Leprecan gene family. *Genesis*. 2008;46(11):683–96.
- Wu Y, Devotta A, José-Edwards DS, Kugler JE, Negrón-Piñero LJ, Braslavskaya K, Addy J, Saint-Jeannet JP, Di Gregorio A. *Xbp1* and *Brachyury* establish an evolutionarily conserved subcircuit of the notochord gene regulatory network. *Elife*. 2022;11.
- Dahia CL, Mahoney EJ, Durrani AA, Wylie C. Postnatal growth, differentiation, and aging of the mouse intervertebral disc. *Spine (Phila Pa 1976)*. 2009;34(5):447–55.
- Dahia CL, Mahoney EJ, Durrani AA, Wylie C. Intercellular signaling pathways active during intervertebral disc growth, differentiation, and aging. *Spine (Phila Pa 1976)*. 2009;34(5):456–62.
- Farhang-Fallah J, Randhawa VK, Nimnual A, Klip A, Bar-Sagi D, Rozakis-Adcock M. The pleckstrin homology (PH) domain-interacting protein couples the insulin receptor substrate 1 PH domain to insulin signaling pathways leading to mitogenesis and GLUT4 translocation. *Mol Cell Biol*. 2002;22(20):7325–36.
- Farhang-Fallah J, Yin X, Trentin G, Cheng AM, Rozakis-Adcock M. Cloning and characterization of PHIP, a novel insulin receptor substrate-1 pleckstrin homology domain interacting protein. *J Biol Chem*. 2000;275(51):40492–7.
- Li S, Francisco AB, Han C, Pattabiraman S, Foote MR, Giesy SL, Wang C, Schimenti JC, Boisclair YR, Long Q. The full-length isoform of the mouse pleckstrin homology domain-interacting protein (PHIP) is required for postnatal growth. *FEBS Lett*. 2010;584(18):4121–7.

36. Kaur H, Panigrahi I. Chung-Jansen syndrome with obesity. *Obes Res Clin Pract.* 2021;15(3):303–5.
37. Siebold C, Yamashita T, Monnier PP, Mueller BK, Pasterkamp RJ. RGMs: structural insights, molecular regulation, and downstream signaling. *Trends Cell Biol.* 2017;27(5):365–78.
38. Prasobh R, Manoj N. The repertoire of heterotrimeric G proteins and RGS proteins in *Ciona intestinalis*. *PLoS ONE.* 2009;4(10):e7349.
39. Anderson GR, Posokhova E, Martemyanov KA. The R7 RGS protein family: multi-subunit regulators of neuronal G protein signaling. *Cell Biochem Biophys.* 2009;54(1–3):33–46.
40. Jayaraman M, Zhou H, Jia L, Cain MD, Blumer KJ. R9AP and R7BP: traffic cops for the RGS7 family in phototransduction and neuronal GPCR signaling. *Trends Pharmacol Sci.* 2009;30(1):17–24.
41. Chan KT, Creed SJ, Bear JE. Unraveling the enigma: progress towards understanding the coronin family of actin regulators. *Trends Cell Biol.* 2011;21(8):481–8.
42. Shina MC, Noegel AA. Invertebrate coronins. *Subcell Biochem.* 2008;48:88–97.
43. Liu X, Gao Y, Lin X, Li L, Han X, Liu J. The coronin family and human disease. *Curr Protein Pept Sci.* 2016;17(6):603–11.
44. José-Edwards DS, Kerner P, Kugler JE, Deng W, Jiang D, Di Gregorio A. The identification of transcription factors expressed in the notochord of *Ciona intestinalis* adds new potential players to the brachyury gene regulatory network. *Dev Dyn.* 2011;240(7):1793–805.
45. José-Edwards DS, Oda-Ishii I, Nibu Y, Di Gregorio A. Tbx2/3 is an essential mediator within the Brachyury gene network during *Ciona* notochord development. *Development.* 2013;140(11):2422–33.
46. Kugler JE, Passamaneck YJ, Feldman TG, Beh J, Regnier TW, Di Gregorio A. Evolutionary conservation of vertebrate notochord genes in the ascidian *Ciona intestinalis*. *Genesis.* 2008;46(11):697–710.
47. Kugler JE, Wu Y, Katikala L, Passamaneck YJ, Addy J, Caballero N, Oda-Ishii I, Maguire JE, Li R, Di Gregorio A. Positioning a multifunctional basic helix-loop-helix transcription factor within the *Ciona* notochord gene regulatory network. *Dev Biol.* 2019;448(2):119–35.
48. Imai KS, Hino K, Yagi K, Satoh N, Satou Y. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development.* 2004;131(16):4047–58.
49. Satou Y, Takatori N, Yamada L, Mochizuki Y, Hamaguchi M, Ishikawa H, Chiba S, Imai K, Kano S, Murakami SD, Nakayama A, Nishino A, Sasakura Y, Satoh G, Shimotori T, Shin IT, Shoguchi E, Suzuki MM, Takada N, Utsumi N, Yoshida N, Saiga H, Kohara Y, Satoh N. Gene expression profiles in *Ciona intestinalis* tailbud embryos. *Development.* 2001;128(15):2893–904.
50. Cao C, Lemaire LA, Wang W, Yoon PH, Choi YA, Parsons LR, Matese JC, Wang W, Levine M, Chen K. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature.* 2019;571(7765):349–54.
51. Corbo JC, Levine M, Zeller RW. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, *Ciona intestinalis*. *Development.* 1997;124(3):589–602.
52. Brozovic M, Martin C, Dantec C, Dauga D, Mendez M, Simion P, Percher M, Laporte B, Scornavacca C, Di Gregorio A, Fujiwara S, Gineste M, Lowe EK, Piette J, Racioppi C, Ristoratore F, Sasakura Y, Takatori N, Brown TC, Delsuc F, Douzery E, Gissi C, McDougall A, Nishida H, Sawada H, Swalla BJ, Yasuo H, Lemaire P. ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. *Nucleic Acids Res.* 2016;44(D1):D808–18.
53. Boczkowska M, Rebowski G, Kremneva E, Lappalainen P, Dominguez R. How Leiomodin and Tropomodulin use a common fold for different actin assembly functions. *Nat Commun.* 2015;6:8314.
54. Dominguez R. The WH2 Domain and Actin Nucleation: Necessary but Insufficient. *Trends Biochem Sci.* 2016;41(6):478–90.
55. Fowler VM, Dominguez R. Tropomodulins and leiomodins: actin pointed end caps and nucleators in muscles. *Biophys J.* 2017;112(9):1742–60.
56. Nishikata T, Yamada L, Mochizuki Y, Satou Y, Shin-i T, Kohara Y, Satoh N. Profiles of maternally expressed genes in fertilized eggs of *Ciona intestinalis*. *Dev Biol.* 2001;238(2):315–31.
57. Jorge EC, Ahmed MU, Bothe I, Coutinho LL, Dietrich S. RGMa and RGMb expression pattern during chicken development suggest unexpected roles for these repulsive guidance molecules in notochord formation, somitogenesis, and myogenesis. *Dev Dyn.* 2012;241(12):1886–900.
58. Camus LM, Lambert LA. Molecular evolution of hemojuvelin and the repulsive guidance molecule family. *J Mol Evol.* 2007;65(1):68–81.
59. Satou Y, Nakamura R, Yu D, Yoshida R, Hamada M, Fujie M, Hisata K, Takeda H, Satoh N. A Nearly complete genome of *Ciona intestinalis* Type A (*C. robusta*) reveals the contribution of inversion to chromosomal evolution in the genus *Ciona*. *Genome Biol Evol.* 2019;11(11):3144–57.
60. Satou Y, Tokuoka M, Oda-Ishii I, Tokuhiro S, Ishida T, Liu B, Iwamura Y. A manually curated gene model set for an Ascidian, *Ciona robusta* (*Ciona intestinalis* Type A). *Zool Sci.* 2022;39(3):253–60.
61. Yamashiro S, Gokhin DS, Kimura S, Nowak RB, Fowler VM. Tropomodulins: pointed-end capping proteins that regulate actin filament architecture in diverse cell types. *Cytoskeleton (Hoboken).* 2012;69(6):337–70.
62. Sierra DA, Gilbert DJ, Householder D, Grishin NV, Yu K, Ukidwe P, Barker SA, He W, Wensel TG, Otero G, Brown G, Copeland NG, Jenkins NA, Wilkie TM. Evolution of the regulators of G-protein signaling multigene family in mouse and human. *Genomics.* 2002;79(2):177–85.
63. Squires KE, Montañez-Miranda C, Pandya RR, Torres MP, Hepler JR. Genetic analysis of rare human variants of regulators of G protein signaling proteins and their role in human physiology and disease. *Pharmacol Rev.* 2018;70(3):446–74.
64. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biémont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigó R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quétiér F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature.* 2004;431(7011):946–57.
65. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res.* 2003;13(3):382–90.
66. Roadcap DW, Clemen SE, Bear JE. The role of mammalian coronins in development and disease. *Subcell Biochem.* 2008;48:124–35.
67. Eckert C, Hammesfahr B, Kollmar M. A holistic phylogeny of the coronin gene family reveals an ancient origin of the tandem-coronin, defines a new subfamily, and predicts protein function. *BMC Evol Biol.* 2011;11:268.
68. Bonnet AL, Chaussain C, Broutin I, Rochefort GY, Schrewe H, Gaucher C. From vascular smooth muscle cells to folliculogenesis: what about vasorin? *Front Med (Lausanne).* 2018;5:335.
69. Mehawej C, Delahodde A, Legeai-Mallet L, Delague V, Kaci N, Desvignes JP, Kibar Z, Capo-Chichi JM, Chouery E, Munnich A, Cormier-Daire V, Mégarbané A. The impairment of MAGMAS function in human is responsible for a severe skeletal dysplasia. *PLoS Genet.* 2014;10(5):e1004311.
70. Bian YH, Xu C, Li J, Xu J, Zhang H, Du SJ. Development of a transgenic zebrafish model expressing GFP in the notochord, somite and liver directed by the hfe2 gene promoter. *Transgenic Res.* 2011;20(4):787–98.
71. Brown S, Jayachandran P, Negesse M, Olmo V, Vital E, Brewster R. Rgma-Induced Neo1 Proteolysis Promotes Neural Tube Morphogenesis. *J Neurosci.* 2019;39(38):7465–84.
72. Gibert Y, Lattanzi VJ, Zhen AW, Vedder L, Brunet F, Faasse SA, Babitt JL, Lin HY, Hammerschmidt M, Fraenkel PG. BMP signaling modulates hepcidin expression in zebrafish embryos independent of hemojuvelin. *PLoS ONE.* 2011;6(1):e14553.
73. Kudoh T, Tsang M, Hukriede NA, Chen X, Dedekian M, Clarke CJ, Kiang A, Schultz S, Epstein JA, Toyama R, Dawid IB. A gene expression screen in zebrafish embryogenesis. *Genome Res.* 2001;11(12):1979–87.
74. Thisse B, Thisse C. Fast release clones: a high throughput expression analysis. ZFIN Direct Data Submission. 2004.
75. Thisse C, Thisse B. High throughput expression analysis of ZF-models consortium clones. ZFIN Direct Data Submission. 2005; Available from: <http://zfin.org>.

76. Berger J, Tarakci H, Berger S, Li M, Hall TE, Arner A, Currie PD. Loss of Tropomodulin4 in the zebrafish mutant *träge* causes cytoplasmic rod formation and muscle weakness reminiscent of nemaline myopathy. *Dis Model Mech*. 2014;7(12):1407–15.
77. Mazelet L, Parker MO, Li M, Arner A, Ashworth R. Role of active contraction and tropomodulins in regulating actin filament length and sarcomere structure in developing zebrafish skeletal muscle. *Front Physiol*. 2016;7:91.
78. White RJ, Collins JE, Sealy IM, Wali N, Dooley CM, Digby Z, Stemple DL, Murphy DN, Billis K, Hourlier T, Füllgrabe A, Davis MP, Enright AJ, Busch-Nentwich EM. A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife*. 2017;6.
79. Nworu CU, Kraft R, Schnurr DC, Gregorio CC, Krieg PA. Leiomodlin 3 and tropomodulin 4 have overlapping functions during skeletal myofibrillogenesis. *J Cell Sci*. 2015;128(2):239–50.
80. Rybakin V, Rastetter RH, Stumpf M, Uetrecht AC, Bear JE, Noegel AA, Clemens CS. Molecular mechanism underlying the association of Coronin-7 with Golgi membranes. *Cell Mol Life Sci*. 2008;65(15):2419–30.
81. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, Mulas C, Ibarra-Soria X, Tyser RCV, Ho DLL, Reik W, Srinivas S, Simons BD, Nichols J, Marioni JC, Göttgens B. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*. 2019;566(7745):490–5.
82. Peck SH, McKee KK, Tobias JW, Malhotra NR, Harfe BD, Smith LJ. Whole Transcriptome Analysis of Notochord-Derived Cells during Embryonic Formation of the Nucleus Pulposus. *Sci Rep*. 2017;7(1):10504.
83. Mohanty S, Pinelli R, Pricop P, Albert TJ, Dahia CL. Chondrocyte-like nested cells in the aged intervertebral disc are late-stage nucleus pulposus cells. *Aging Cell*. 2019;18(5):e13006.
84. Winkler T, Mahoney EJ, Sinner D, Wylie CC, Dahia CL. Wnt signaling activates Shh signaling in early postnatal intervertebral discs, and re-activates Shh signaling in old discs in the mouse. *PLoS ONE*. 2014;9(6):e98444.
85. Dahia CL, Mahoney E, Wylie C. Shh signaling from the nucleus pulposus is required for the postnatal growth and differentiation of the mouse intervertebral disc. *PLoS ONE*. 2012;7(4):e35944.
86. Melgoza IP, Chenna SS, Tessier S, Zhang Y, Tang SY, Ohnishi T, Novais EJ, Kerr GJ, Mohanty S, Tam V, Chan WCW, Zhou CM, Zhang Y, Leung VY, Brice AK, Séguin CA, Chan D, Vo N, Risbud MV, Dahia CL. Development of a standardized histopathology scoring system using machine learning algorithms for intervertebral disc degeneration in the mouse model—An ORS spine section initiative. *JOR Spine*. 2021;4(2):e1164.
87. Vincent K, Mohanty S, Pinelli R, Bonavita R, Pricop P, Albert TJ, Dahia CL. Aging of mouse intervertebral disc and association with back pain. *Bone*. 2019;123:246–59.
88. Rodrigues-Pinto R, Ward L, Humphreys M, Zeef LAH, Berry A, Hanley KP, Hanley N, Richardson SM, Hoyland JA. Human notochordal cell transcriptome unveils potential regulators of cell function in the developing intervertebral disc. *Sci Rep*. 2018;8(1):12866.
89. Bydon M, Moinuddin FM, Yolcu YU, Wahood W, Alvi MA, Goyal A, Elminawy M, Galeano-Garces C, Dudakovic A, Nassr A, Larson AN, van Wijnen AJ. Lumbar intervertebral disc mRNA sequencing identifies the regulatory pathway in patients with disc herniation and spondylolisthesis. *Gene*. 2020;750:144634.
90. Holland PW, Garcia-Fernández J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Dev Suppl* 1994: 125–33.
91. Postlethwait JH. The zebrafish genome in context: ohnologs gone missing. *J Exp Zool B Mol Dev Evol*. 2007;308(5):563–77.
92. Taylor JS, Van de Peer Y, Braasch I, Meyer A. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci*. 2001;356(1414):1661–79.
93. Gray PA, Fu H, Luo P, Zhao Q, Yu J, Ferrari A, Tenzen T, Yuk DI, Tsung EF, Cai Z, Alberta JA, Cheng LP, Liu Y, Stenman JM, Valerius MT, Billings N, Kim HA, Greenberg ME, McMahon AP, Rowitch DH, Stiles CD, Ma Q. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science*. 2004;306(5705):2255–7.
94. Bao Y, Kake T, Hanashima A, Nomiya Y, Kubokawa K, Kimura S. Actin capping proteins, CapZ (β -actinin) and tropomodulin in amphioxus striated muscle. *Gene*. 2012;510(1):78–86.
95. Inoue J, Satoh N. Deuterostome genomics: lineage-specific protein expansions that enabled chordate muscle evolution. *Mol Biol Evol*. 2018;35(4):914–24.
96. Posokhova E, Ng D, Opel A, Masuho I, Tinker A, Biesecker LG, Wickman K, Martemyanov KA. Essential role of the m2R-RGS6- IKACh pathway in controlling intrinsic heart rate variability. *PLoS ONE*. 2013;8(10):e76973.
97. Stewart A, Maity B, Fisher RA. Two for the Price of One: G Protein-Dependent and -Independent Functions of RGS6 In Vivo. *Prog Mol Biol Transl Sci*. 2015;133:123–51.
98. Yang J, Huang J, Maity B, Gao Z, Lorca RA, Gudmundsson H, Li J, Stewart A, Swaminathan PD, Ibeawuchi SR, Shepherd A, Chen CK, Kutschke W, Mohler PJ, Mohapatra DP, Anderson ME, Fisher RA. RGS6, a modulator of parasympathetic activation in heart. *Circ Res*. 2010;107(11):1345–9.
99. Yang J, Platt LT, Maity B, Ahlers KE, Luo Z, Lin Z, Chakravarti B, Ibeawuchi SR, Askeland RW, Bondaruk J, Czerniak BA, Fisher RA. RGS6 is an essential tumor suppressor that prevents bladder carcinogenesis by promoting p53 activation and DNMT1 downregulation. *Oncotarget*. 2016;7(43):69159–72.
100. Appleton CT, James CG, Beier F. Regulator of G-protein signaling (RGS) proteins differentially control chondrocyte differentiation. *J Cell Physiol*. 2006;207(3):735–45.
101. Wang F, Zhang C, Shi R, Xie ZY, Chen L, Wang K, Wang YT, Xie XH, Wu XT. The embryonic and evolutionary boundaries between notochord and cartilage: a new look at nucleus pulposus-specific markers. *Osteoarthritis Cartil*. 2018;26(10):1274–82.
102. Ohmae S, Noma N, Toyomoto M, Shinohara M, Takeiri M, Fuji H, Takemoto K, Iwaisako K, Fujita T, Takeda N, Kawatani M, Aoyama M, Hagiwara M, Ishihama Y, Asagiri M. Actin-binding protein coronin 1A controls osteoclastic bone resorption by regulating lysosomal secretion of cathepsin K. *Sci Rep*. 2017;7:41710.
103. Park J, Jun K, Choi Y, Yoon E, Kim W, Jang YG, Chung J. CORO7 functions as a scaffold protein for the core kinase complex assembly of the Hippo pathway. *J Biol Chem*. 2021;296:100040.
104. Oda-Ishii I, Di Gregorio A. Lineage-independent mosaic expression and regulation of the *Ciona* multidom gene in the ancestral notochord. *Dev Dyn*. 2007;236(7):1806–19.
105. Reeves W, Thayer R, Veeman M. Anterior-posterior regionalized gene expression in the *Ciona* notochord. *Dev Dyn*. 2014;243(4):612–20.
106. Albuixech-Crespo B, López-Blanch L, Burguera D, Maeso I, Sánchez-Arrones L, Moreno-Bravo JA, Somorjai I, Pascual-Anaya J, Puelles E, Bovolenta P, Garcia-Fernández J, Puelles L, Irimia M, Ferran JL. Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain. *PLoS Biol*. 2017;15(4):e2001573.
107. Ferran JL, Puelles L. Lessons from amphioxus bauplan about origin of cranial nerves of vertebrates that innervates extrinsic eye muscles. *Anat Rec (Hoboken)*. 2019;302(3):452–62.
108. Yamanaka Y, Tamplin OJ, Beckers A, Gossler A, Rossant J. Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. *Dev Cell*. 2007;13(6):884–96.
109. Halbrooks PJ, Ding R, Wozney JM, Bain G. Role of RGM coreceptors in bone morphogenetic protein signaling. *J Mol Signal*. 2007;2:4.
110. Hotta K, Mitsuhashi K, Takahashi H, Inaba K, Oka K, Gobjori T, Iikeo K. A web-based interactive developmental table for the ascidian *Ciona intestinalis*, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev Dyn*. 2007;236(7):1790–805.
111. Gilchrist MJ, Sobral D, Khoueiry P, Daian F, Laporte B, Patrushev I, Matsumoto J, Dewar K, Hastings KE, Satou Y, Lemaire P, Rothbächer U. A pipeline for the systematic identification of non-redundant full-ORF cDNAs for polymorphic and evolutionary divergent genomes: Application to the ascidian *Ciona intestinalis*. *Dev Biol*. 2015;404(2):149–63.
112. Satou Y, Yamada L, Mochizuki Y, Takatori N, Kawashima T, Sasaki A, Hamaguchi M, Awazu S, Yagi K, Sasakura Y, Nakayama A, Ishikawa H, Inaba K, Satoh N. A cDNA resource from the basal chordate *Ciona intestinalis*. *Genesis*. 2002;33(4):153–4.
113. Satou Y, Mineta K, Ogasawara M, Sasakura Y, Shoguchi E, Ueno K, Yamada L, Matsumoto J, Wasserscheid J, Dewar K, Wiley GB, Macmill SL, Roe BA, Zeller RW, Hastings KE, Lemaire P, Lindquist E, Endo T, Hotta K, Inaba K. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. *Genome Biol*. 2008;9(10):R152.

114. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15.
115. Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.* 2006;4:41.
116. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673–80.
117. Guindon S, Dufayard JF, 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(3):307–21.
118. Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol.* 2017;34(9):2422–4.
119. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9.
120. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006;55(4):539–52.
121. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 2006;34(Web Server issue):W362–5.
122. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. InterPro in 2022. *Nucleic Acids Res.* 2023;51(D1):D418–d427.
123. Nguyen NTT, Vincens P, Dufayard JF, Roest Crollius H, Louis A. Genomicus in 2022: comparative tools for thousands of genomes and reconstructed ancestors. *Nucleic Acids Res.* 2022;50(D1):D1025–d1031.
124. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
125. Coppola U, Annona G, D'Aniello S, Ristoratore F. Rab32 and Rab38 genes in chordate pigmentation: an evolutionary perspective. *BMC Evol Biol.* 2016;16:26.
126. Normal table of *Xenopus laevis* (Daudin), Nieuwkoop PD, Faber J eds. Amsterdam: North Holland Publishing Company. 1967
127. Slack JM, Forman D. An interaction between dorsal and ventral regions of the marginal zone in early amphibian embryos. *J Embryol Exp Morphol.* 1980;56:283–99.
128. Harland RM. In situ hybridization: an improved whole-mount method for *Xenopus* embryos. *Methods Cell Biol.* 1991;36:685–95.
129. Saint-Jeannet JP. Whole-Mount In Situ Hybridization of *Xenopus* Embryos. *Cold Spring Harb Protoc.* 2017;2017(12):pdb.prot097287.
130. Piprode V, Mohanty S, Bonavita R, Loh S, Anbazhagan R, Saini C, Pinelli R, Pricop P, Dahiya CL. An optimized step-by-step protocol for isolation of nucleus pulposus, annulus fibrosus, and end plate cells from the mouse intervertebral discs and subsequent preparation of high-quality intact total RNA. *JOR Spine.* 2020;3(3):e1108.
131. Pfirrmann CW, Metzendorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa).* 2001;26(17):1873–8.
132. Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, Marini F, Rue-Albrecht K, Risso D, Sonesson C, Waldron L, Pagès H, Smith ML, Huber W, Morgan M, Gottardo R, Hicks SC. Orchestrating single-cell analysis with Bioconductor. *Nat Methods.* 2020;17(2):137–45.
133. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573–3587.e29.
134. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

