



# Gain of gene regulatory network interconnectivity at the origin of vertebrates

Alejandro Gil-Gálvez<sup>a,1</sup>, Sandra Jiménez-Gancedo<sup>a,1</sup>, Alberto Pérez-Posada<sup>a</sup>, Martin Franke<sup>a</sup>, Rafael D. Acemel<sup>a</sup>, Che-Yi Lin<sup>b</sup>, Cindy Chou<sup>b</sup>, Yi-Hsien Su<sup>b</sup>, Jr-Kai Yu<sup>b,c</sup>, Stephanie Bertrand<sup>d</sup>, Michael Schubert<sup>e</sup>, Héctor Escrivá<sup>d</sup>, Juan J. Tena<sup>a,2</sup>, and José Luis Gómez-Skarmeta<sup>a,3</sup>

Edited by Denis Duboule, Université de Genève, Genève, Switzerland; received August 13, 2021; accepted January 21, 2022

Signaling pathways control a large number of gene regulatory networks (GRNs) during animal development, acting as major tools for body plan formation [A. Pires-daSilva, R. J. Sommer, *Nat. Rev. Genet.* 4, 39–49 (2003)], although only a few of these pathways operate during this period [J. J. Sanz-Ezquerro, A. E. Münsterberg, S. Stricker, *Front. Cell Dev. Biol.* 5, 76 (2017)]. Moreover, most of them have been largely conserved during metazoan evolution [L. S. Babonis, M. Q. Martindale, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, 20150477 (2017)]. How evolution has generated a vast diversity of animal morphologies with such a limited number of tools is still largely unknown. Here, we show that gain of interconnectivity between signaling pathways and the GRNs they control may have critically contributed to the origin of vertebrates. We perturbed the retinoic acid, Wnt, FGF, and Nodal signaling pathways during gastrulation in the invertebrate chordate amphioxus and zebrafish and compared the effects on gene expression and *cis*-regulatory elements (CREs). We found that multiple developmental genes gain response to these pathways through vertebrate-specific CREs. Moreover, in contrast to amphioxus, many of these CREs responded to multiple pathways in zebrafish, which reflects their high interconnectivity. Furthermore, we found that vertebrate-specific cell types are more enriched in highly interconnected genes than in tissues with more ancient origin. Thus, the increase of CREs in vertebrates integrating inputs from different signaling pathways probably contributed to gene expression complexity and to the formation of new cell types and morphological novelties in this lineage.

evolution | regulatory network | signaling pathway | amphioxus | zebrafish

During embryonic development, thousands of genes are expressed in a coordinated and tightly regulated manner. This coordination is facilitated by complex hierarchical relationships between genes (1, 2), where the expression of a particular effector gene triggers the transcription of many other genes in a multilevel cascade that can involve hundreds of different targets (3). Signaling pathways control most of these genetic cascades, interconnecting many genes and playing pivotal roles in more complex gene regulatory networks (GRNs). As a consequence, they are key substrates for the generation of morphological diversity during evolution (4). However, the complexity of networks is not only dictated by the number of nodes but also by the number and patterns of interactions among them. In this context, effectors of signaling pathways constitute hubs in GRNs.

It has already been demonstrated that after the vertebrate-specific whole genome duplications (WGDs), many duplicated developmental genes were maintained in this group (5). Furthermore, it is also known that regulatory landscapes in general, and especially those of developmental genes, have been expanded in the vertebrate lineage (6). Yet, it remains unclear how these features interplay to generate the highly complex body plans of adult vertebrates. Here, we investigated the contribution of key signaling pathways in the evolutionary transition from invertebrate chordates to vertebrates. To study this question, we compared the effects of interfering with the retinoic acid (RA), Wnt, fibroblast growth factor (FGF), and Nodal pathways during gastrulation in both amphioxus (a cephalochordate) and zebrafish embryos. To do so, we used pharmacological compounds known to act either as agonists of the RA and Wnt or antagonists of the FGF and Nodal pathways (7). We then examined the impact of these treatments on global gene expression by RNA sequencing (RNA-seq) in both species (Fig. 1A and *SI Appendix*, Fig. S1). Additionally, we performed assay for transposase accessible chromatin using sequencing (ATAC-seq) to identify open chromatin regions (8), including enhancers and promoters, affected by these manipulations (Fig. 1A and *SI Appendix*, Fig. S1).

## Significance

In this manuscript, we address an essential question in developmental and evolutionary biology: How have changes in gene regulatory networks contributed to the invertebrate-to-vertebrate transition? To address this issue, we perturbed four signaling pathways critical for body plan formation in the cephalochordate amphioxus and in zebrafish and compared the effects of such perturbations on gene expression and gene regulation in both species. Our data reveal that many developmental genes have gained response to these signaling pathways in the vertebrate lineage. Moreover, we show that the interconnectivity between these pathways is much higher in zebrafish than in amphioxus. We conclude that this increased signaling pathway complexity likely contributed to vertebrate morphological novelties during evolution.

Author contributions: J.J.T. and J.L.G.-S. designed research; A.G.-G., S.J.-G., A.P.-P., M.F., R.D.A., C.-Y.L., C.C., Y.-H.S., J.-K.Y., S.B., H.E., and J.J.T. performed research; M.S. contributed new reagents/analytic tools; S.B. and H.E. contributed biological material; A.G.-G., S.J.-G., A.P.-P., R.D.A., and J.J.T. analyzed data; and J.J.T. and J.L.G.-S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

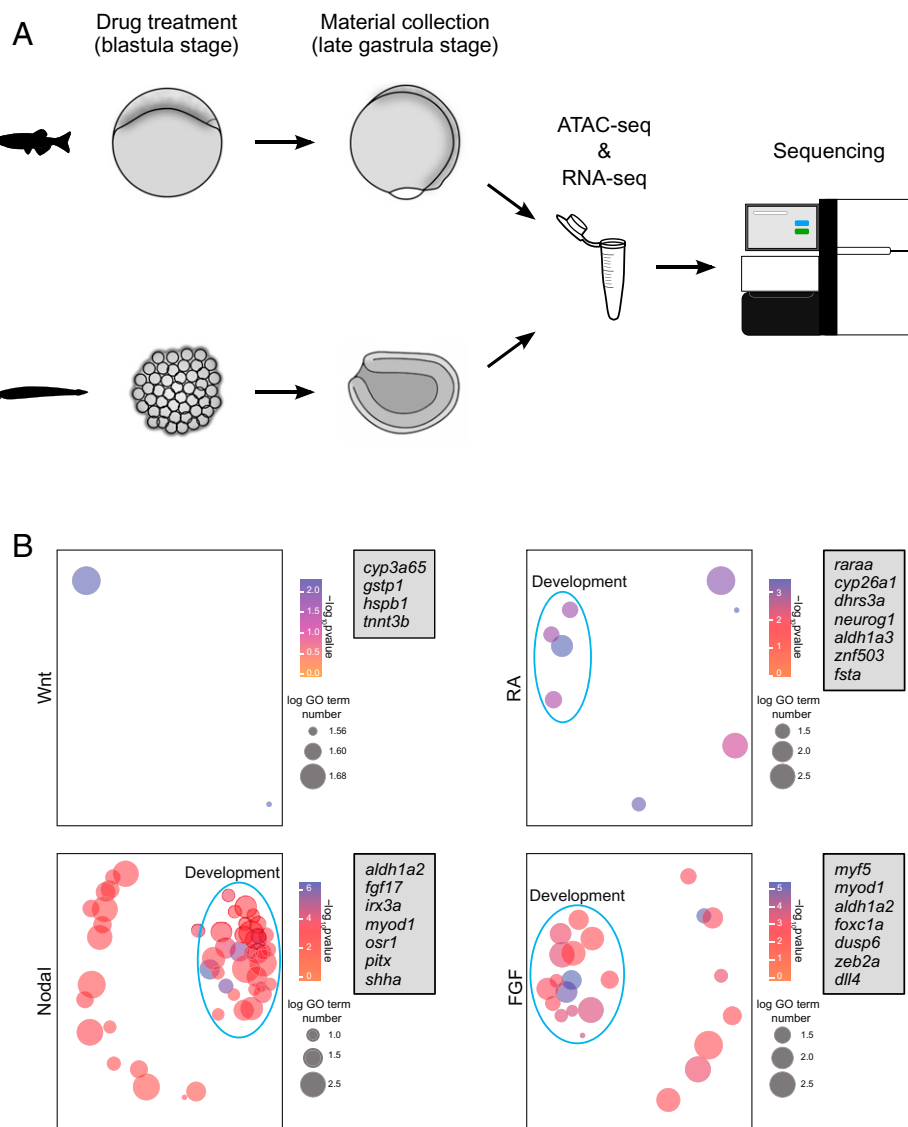
<sup>1</sup>A.G.-G. and S.J.-G. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [jjtenagu@upo.es](mailto:jjtenagu@upo.es).

<sup>3</sup>Deceased September 16, 2020.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2114802119/-/DCSupplemental>.

Published March 9, 2022.



**Fig. 1.** Experimental design and differential analyses. (A) Overall design of the experiment: zebrafish and amphioxus embryos were treated with four different compounds at the blastula stage and dissociated at the late gastrula stage for RNA-seq and ATAC-seq library preparations. (B) GO term enrichment analysis for orthologous genes that are commonly affected in both amphioxus and zebrafish. Each panel shows GO enrichment visualization obtained with the REVIGO tool for one of the treatments, taking into account only genes significantly modified upon the indicated treatment. Circles represent GO terms, and the X and Y axes map the semantic space: the closer the terms appear in the plot, the more related they are. In order to facilitate direct comparisons, a blue line surrounds GO terms associated with developmental processes. Gray boxes at the right of each panel show some of the developmental genes represented in the plot.

The WGDs at the base of the vertebrate lineage (5, 9) and the additional WGD in the teleost lineage (10) resulted in gene number imbalances between amphioxus and zebrafish. To overcome this limitation in our analysis, we used previously published data (6) to retrieve all the vertebrate gene family members (i.e., paralogy groups) corresponding to each amphioxus gene affected by the treatments. For zebrafish, we used only affected genes.

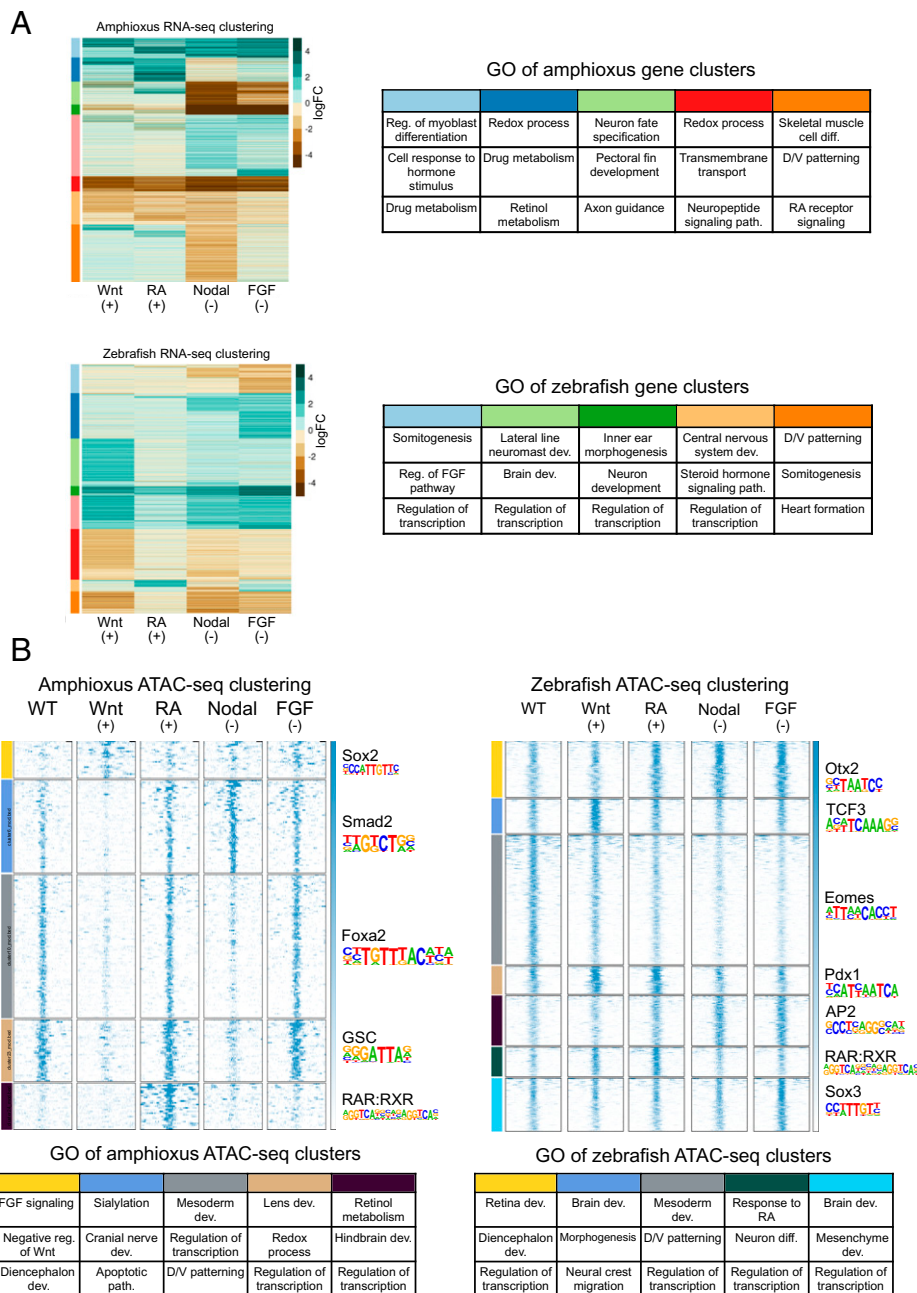
The RNA-seq analyses revealed hundreds of differentially expressed genes following the different treatments in both amphioxus and zebrafish (SI Appendix, Fig. S1). Interestingly, we found transcripts of orthologous genes similarly altered in both species upon the same treatment (Fig. 1B). Gene ontology (GO) analysis for these commonly affected transcripts confirmed that they are highly associated with developmental processes (e.g., mesoderm, endoderm, and hindbrain development), which are known to be regulated by the tested signaling pathways (7, 11, 12) (Fig. 1B and Dataset S1). Surprisingly, only a few genes were similarly affected upon Wnt activation in

both species (Fig. 1B and Dataset S1). We then examined all the genes affected by each of these treatments, and we observed that the number of genes perturbed in zebrafish is higher than that in amphioxus and, proportionally, more strongly associated with development and signaling terms. Moreover, confirming our experimental approach, the genes altered by these treatments and their corresponding GO terms clearly associated with developmental processes known to be regulated by these pathways (7, 11, 12) (SI Appendix, Fig. S2 and Dataset S2).

Regarding the ATAC-seq regions modified by the treatments, we observed a very different genomic distribution in both species, with a big proportion of peaks near gene promoters and very few distant enhancers in amphioxus and the opposite trend in zebrafish (SI Appendix, Fig. S3), results that were expected from our previous work (6). We next analyzed the ATAC-seq data by searching for motifs enriched in peaks either more accessible after treatment with RA or Wnt agonists or less accessible upon treatment with Nodal or FGF inhibitors.

The binding sites for transcription factors (TFs) that mediate signaling of these pathways and/or well-known downstream TFs of these pathways were found for all treatments in both species (13–27) (*SI Appendix, Fig. S4* and *Dataset S3*). Overall, these results confirmed that the pharmacological treatments of amphioxus and zebrafish embryos indeed perturbed the targeted pathways *in vivo*. Using previously published ATAC-seq data at different developmental stages in zebrafish and amphioxus (6), we calculated the ATAC-seq signal dynamics during embryonic development surrounding the differentially accessible regions upon treatments. We observed that, in general, these regions were active during development and showed a dynamic pattern of accessibility (*SI Appendix, Fig. S5*).

To better classify the genes that responded to perturbation of the different signaling pathways, we performed a clustering analysis of gene expression, which resulted in groups of genes with similar transcriptional behavior (*Fig. 2A*). We then carried out GO enrichment analyses of these groups. The RNA-seq-derived clusters in zebrafish were mostly associated with GO terms related to embryonic development, while in amphioxus, we also detected many terms related to metabolism and cell homeostasis (*Fig. 2A* and *Dataset S4*). In some cases, the association of a single pathway with a specific function (as defined by the GO terms) was well established. For example, in amphioxus, retinol metabolism appeared as a main function in genes up-regulated by RA treatment (dark blue cluster) and



**Fig. 2.** Clustering of RNA-seq and ATAC-seq data. (A) RNA-seq-derived gene expression FC in amphioxus (*Upper*) and zebrafish (*Lower*). FC values were calculated for each condition versus control samples by means of differential gene expression analyses. Tables on the right show three representative GO terms enriched in the cluster marked with the same color. (B) Clustering of ATAC-seq data in amphioxus (*Left*) and zebrafish (*Right*). The top DNA-binding motif found in each cluster is presented on the right, and three representative GO terms enriched in the clusters are shown in the tables below. dev., development; diff., differentiation; path., pathway; Reg., regulation; D/V, dorsal/ventral.

muscle cell differentiation in genes down-regulated by Nodal inhibition (orange cluster). In zebrafish, heart formation and somitogenesis were associated with Nodal and FGF inhibition, respectively (dark orange and light blue clusters), while brain development was associated with Wnt pathway activation (light green cluster) (Fig. 2A). This cluster analysis also revealed that in both amphioxus and zebrafish, the majority of developmental processes were influenced by several of the studied signaling pathways.

In the case of ATAC-seq peak clustering, we assigned peaks to their putative target genes using the GREAT algorithm (28) in order to derive GO terms. In general, we observed a good correlation between the average ATAC-seq signal around the peaks in the different clusters, the GO terms of their putative target genes, and the TF-binding motifs identified within these peaks (Fig. 2B). Furthermore, open chromatin regions altered upon treatments were mainly associated with development, especially in zebrafish (Fig. 2B), which was consistent with our RNA-seq analysis. For example, the zebrafish cluster of ATAC-seq peaks that was, on average, more accessible upon Wnt stimulation (blue cluster) was associated with brain development (Fig. 2B). In addition, TCF3 motifs, usually bound by effectors of the Wnt pathway, were found within these peaks (Fig. 2B). Similarly, in both species, the clusters enriched following RA treatment (dark purple cluster in amphioxus, dark green cluster in zebrafish) were associated with response to RA and hindbrain development (Fig. 2B) and contained RAR:RXR motifs (Fig. 2B). Importantly, genes belonging to clusters that responded to the same pathways in both species shared similar functions, according to the GO terms associated with them (SI Appendix, Fig. S6 and Dataset S4). Nevertheless, clusters of ATAC-seq peaks in zebrafish generally responded to more treatments than in amphioxus, and for this reason, the similarities of clusters found in both species were limited (SI Appendix, Fig. S6).

The majority of ATAC-seq peak clusters contained peaks that changed in more than one signaling pathway, which was similar to what we observed for the RNA-seq clustering. Several signaling pathways thus seemed to act through the same regulatory elements, suggesting a certain level of interconnection between the different pathways, especially in zebrafish (SI Appendix, Fig. S7). The percentages of overlap between sets of ATAC-seq peaks affected by different pathways were variable in both species (from 0 to 72.4% in amphioxus and from 0 to 89.1% in zebrafish; SI Appendix, Fig. S7B). Nevertheless, this overlap was significantly higher in zebrafish than in amphioxus (SI Appendix, Fig. S7).

In order to combine the treatment effect at the transcriptomic and regulatory levels, we intersected the differentially regulated genes (RNA-seq) with the genes associated to differential ATAC-seq peaks for each treatment and species. The resulting intersected gene lists are thereafter referred to as double-selected genes (DSGs). In this analysis, we included both positively and negatively affected genes/regulatory elements. Although the level of transcription of a gene did not necessarily correlate with the degree of accessibility of its promoter, comparisons of the treatment effects at the transcriptomic and regulatory levels in both species showed a slight correlation between them (SI Appendix, Fig. S8). The intersection resulted in a total of 2,098 genes and 4,609 ATAC-seq peaks in zebrafish and 481 genes and 853 ATAC-seq peaks in amphioxus (Dataset S5). Although the lower numbers of treatment-affected genes and ATAC-seq peaks in amphioxus might be interpreted as a loss of regulatory information in this species, it is generally accepted that there was a general gain of regulatory input

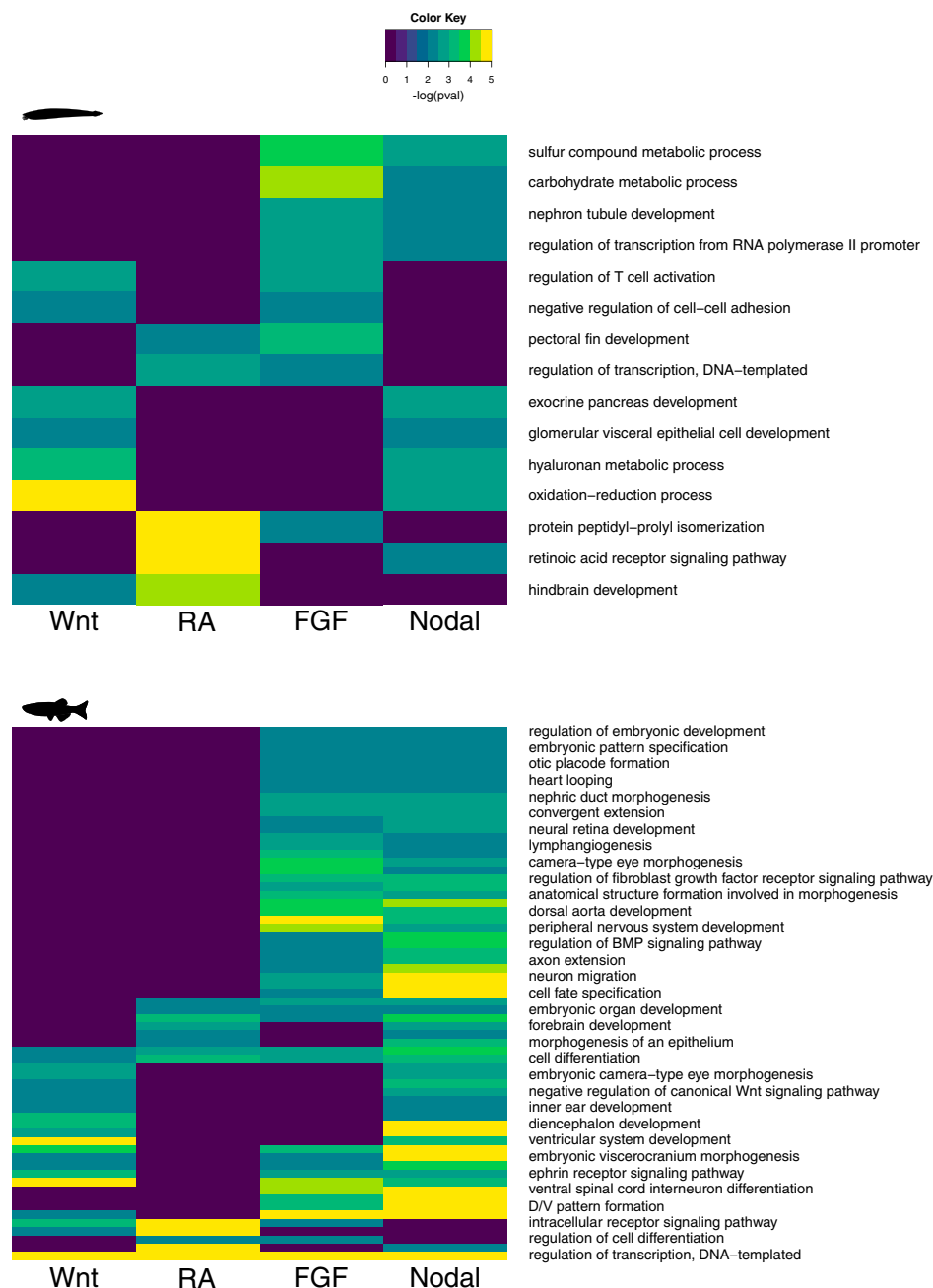
during the invertebrate chordate-to-vertebrate transition (6). It is thus very likely that this significant difference between zebrafish and amphioxus is indicative of a gain of response to these signaling pathways in vertebrates through the incorporation of *cis*-regulatory elements (CREs).

We then clustered the results of the GO enrichment analyses corresponding to DSGs in amphioxus and zebrafish (Fig. 3 and SI Appendix, Fig. S9). We found that the *P* values associated with the enriched GO terms were, in general, much lower in zebrafish than in amphioxus and that the number of GO terms significantly affected by the different treatments was much higher in zebrafish (Fig. 3). This suggests that the regulatory networks involved in developmental processes are more complex in vertebrates than in invertebrate chordates. The same effect was observed when we analyzed gene families, which were not affected by the overestimation of the number of genes in amphioxus due to the inclusion of whole families of orthologous genes, instead of *P* values corresponding to specific genes (SI Appendix, Fig. S9).

These results further supported the notion that there are more genes controlled by the studied signaling pathways in vertebrates than in invertebrate chordates. Among the DSGs identified in zebrafish, we found significant numbers of TFs and regulators of signaling pathways (Dataset S6). We also found that some GO terms in zebrafish were significantly enriched for signaling pathways (e.g., targeting the FGF and Nodal pathways resulted in a significant overlap of development-related terms; Fig. 3). These two facts indicated that, in agreement with the results of our previous clustering analysis, different signaling pathways are interconnected and that the degree of connectivity is higher in vertebrates than in invertebrate chordates. To directly test this hypothesis, we counted the number of genes that were affected by one, two, three, or four different signaling pathways in zebrafish and amphioxus. Interestingly, independently of considering only RNA-seq data or combining RNA-seq and ATAC-seq information, we found that the interconnection between pathways was higher in zebrafish than in amphioxus, with a higher proportion of genes responding to two, three, or four perturbations in zebrafish (Fig. 4A and SI Appendix, Fig. S10A). Moreover, we observed an increase of connectivity in vertebrates that was independent of the number of genes retained after the WGD events (SI Appendix, Fig. S10B). Nevertheless, we also found that the more gene copies were retained, the higher the gain of connectivity (SI Appendix, Fig. S10C).

To address the question of whether this gain of connectivity is species-specific or a general distinction of vertebrates and invertebrates, we performed experiments in two additional species, the invertebrate hemichordate *Ptychodera flava* and another vertebrate, the amphibian *Xenopus tropicalis*. Embryos from both species were treated during gastrulation with the same antagonists of the FGF and Nodal pathways (SI Appendix, Fig. S11). The results of these experiments suggest that increased connectivity is a general tendency of vertebrates compared to invertebrates, since we observed a similar distribution of connectivity in *P. flava* and amphioxus (*P* value 0.545,  $\chi^2$  test), and a higher degree of connectivity in *X. tropicalis* (*P* values 0.006 and 0.048 against *P. flava* and amphioxus, respectively,  $\chi^2$  test). Although the connectivity in *X. tropicalis* was lower than in zebrafish, in which the complexity of GRNs might have been further increased due to the extra round of WGD characterizing all teleost, this difference was not statistically significant (*P* value 0.196,  $\chi^2$  test).

The gain of complexity observed in the vertebrate species could be the product of the addition of new TF binding sites to the available regulatory regions during evolution or the generation of new regulatory regions. In order to shed light on this

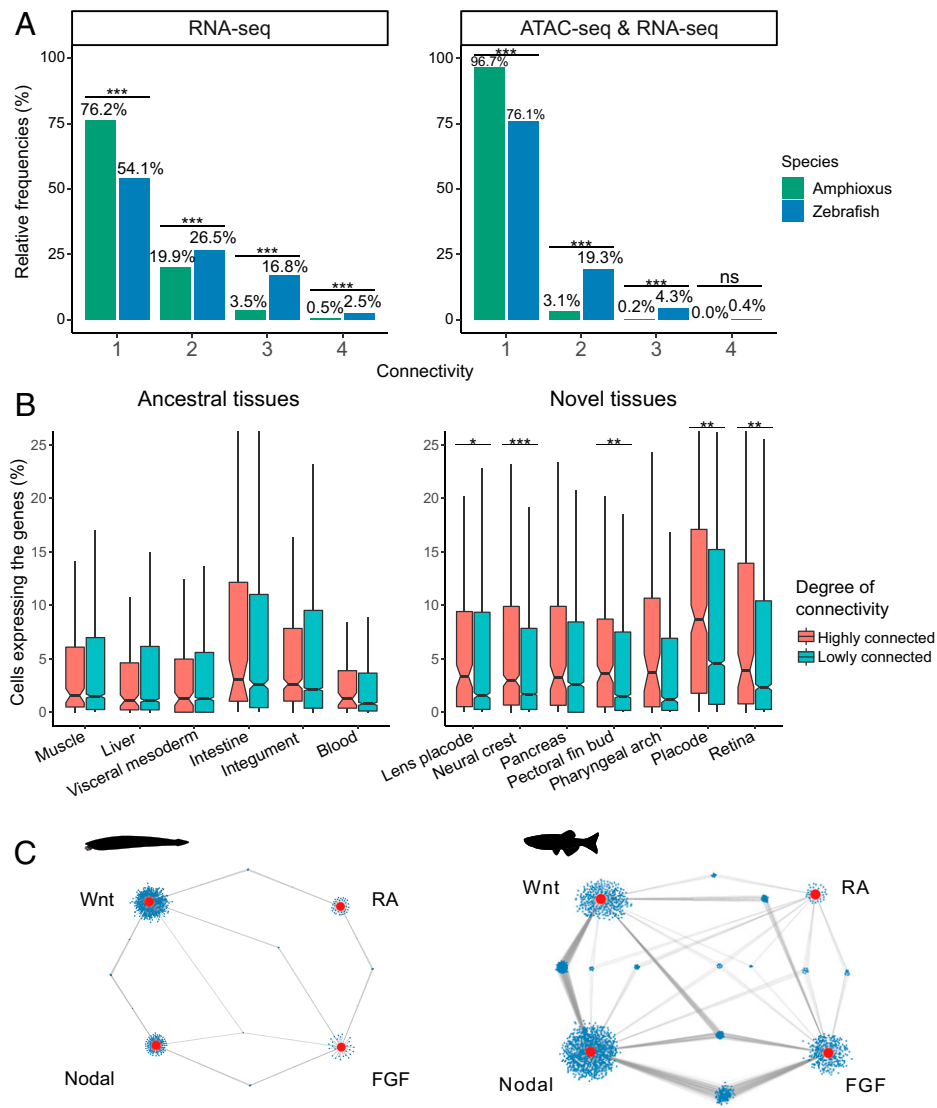


**Fig. 3.** Clustering of GO terms. GO terms enriched in differentially regulated genes at the transcriptomic level (based on RNA-seq analysis) that are also associated with differential ATAC-seq peaks, clustered by  $P$  values (pval) in amphioxus (*Upper*) and zebrafish (*Lower*). BMP, bone morphogenetic protein; D/V, dorsal/ventral.

issue, we interrogated the seven regulatory regions known to be conserved between vertebrates and invertebrates (29), searching for TF-binding sites in the orthologous sequences of both zebrafish and amphioxus. We were unable to identify a significant increase in the number and/or complexity of the binding sites in the sequences of both species (*SI Appendix, Figs. S12 and S13*). Pointing in the same direction, we found that newer regulatory regions presented a higher level of connectivity in zebrafish than more ancient enhancers (*SI Appendix, Fig. S14*). Thus, according to these results, the gain in regulatory complexity in vertebrates might have been driven by the WGD-dependent increase in the number of regulatory regions. This hypothesis is coherent with the results of a previous study reporting a higher number of enhancers per gene in zebrafish when compared to amphioxus (6).

Using already available single-cell RNA-seq data (30) in zebrafish, we observed that, interestingly, tissues that appeared as novelties during the invertebrate chordate-to-vertebrate transition, such as neural crest or sensory placodes, showed a statistically significant enrichment in the expression of highly connected genes, while in more evolutionary ancient tissues such as muscles or intestine, expression of highly and lowly connected genes was very similar (Fig. 4*B* and *SI Appendix, Fig. S15*). Finally, we used the Cytoscape tool (31) to visualize all connections between genes and the different signaling pathways in amphioxus and zebrafish (Fig. 4*C*). This plot clearly showed that developmental GRNs associated with the four targeted signaling pathways were more interconnected in vertebrates.

By comparing epigenomic and transcriptomic data in amphioxus, zebrafish, and other vertebrates, we recently showed that



**Fig. 4.** Gene connectivity to different signaling pathways. (A) Percentage of genes that responded to one, two, three, or four different pathways in amphioxus (green bars) and zebrafish (blue bars), only at the RNA-seq level (Left) or at both the RNA-seq and ATAC-seq levels (Right). (B) Percentage of cells expressing highly connected genes (connectivity  $\geq 3$ , pink bars) and lowly connected genes (connectivity = 1, light blue bars), according to single-cell RNA-seq experiments carried out in zebrafish (30) in ancient tissues (and thus also present in amphioxus) (Left) and in vertebrate-specific tissues (Right). Boxes represent the first, second, and third quartiles, lines represent the range. (C) Cytoscape plot showing connectivity networks in amphioxus (Left) and zebrafish (Right). Small blue dots represent genes, and big red dots represent signaling pathways. Gray lines mark the responsiveness of each gene to the connected signaling pathway. ns, not significant. \* $P$  value < 0.05; \*\* $P$  value < 0.01; \*\*\* $P$  value < 0.001.

the invertebrate chordate-to-vertebrate transition was associated with an increase of regulatory information in the latter lineage and that this increase likely contributed to the spatial and functional specialization of some duplicated genes (6). Here, we go one step farther and demonstrate that some of the vertebrate CREs contribute to a more complex interconnection between the RA, Wnt, FGF, and Nodal signaling pathways. An increased interconnection between these four key developmental signaling pathways likely facilitated the restriction of the expression domains of some duplicated developmental genes, which, in turn, contributed to the increment of tissue complexity required to generate morphological novelties in vertebrates.

**Methods**

**Animal Husbandry and Treatment of Embryos.** Zebrafish (*Danio rerio*) embryos were manipulated following the protocols that have been approved by the Ethics Committee of the Andalusian Government (license number 182-41106) and the national and European regulations established. All

experiments with zebrafish were carried out in accordance with the principles of 3Rs (replacement, reduction, and refinement). Zebrafish embryos at 30% epiboly were treated by adding different drugs to the medium. In the case of SB505124, embryos were treated at the 256-cell stage (SB505124 is an inhibitor of transforming growth factor (TGF)- $\beta$  type I receptors that blocks signaling through Nodal and other related TGF- $\beta$  molecules; we will refer to this compound as Nodal inhibitor for simplicity). The final concentrations of drugs were all-*trans* retinoic acid (ATRA) 0.1  $\mu$ M (cat no. R2625, Sigma-Aldrich Merck), SU5402 20  $\mu$ M (cat no. SML0444, Sigma-Aldrich Merck), BIO 14  $\mu$ M (cat no. B1686, Sigma-Aldrich Merck), and SB505124 30  $\mu$ M (cat no. S4696, Sigma-Aldrich Merck). These molecules were dissolved in dimethyl sulfoxide (DMSO) (cat no. D2438, Sigma-Aldrich Merck). As a control experiment, the same volume of DMSO was added to the medium. Drugs were removed by changing the medium several times when embryos were at 80% epiboly stage. After that, embryos were carefully transferred to a glass Petri dish, and the chorion was removed using pronase.

Amphioxus (*Branchiostoma lanceolatum*) adults were collected at the Racou beach in Argelès-sur-Mer (France). Spawning was induced as previously described (32, 33), and the fertilization of eggs was carried out in vitro. Amphioxus embryos were manipulated in filtered seawater unless otherwise specified. After fertilization, embryos were dechorionated in a Petri dish covered with agarose (0.8% agarose

in filtered seawater) by pipetting them toward the border of the dish and then gently transferred to a small Petri dish. Drugs were dissolved in DMSO, and the following final concentrations were used: ATRA 1  $\mu\text{M}$  (cat no. R2625, Sigma-Aldrich Merck), SU5402 25  $\mu\text{M}$  (cat no. 572630, Sigma-Aldrich Merck), BIO 1  $\mu\text{M}$  (cat no. B1686, Sigma-Aldrich Merck), and SB505124 50  $\mu\text{M}$  (cat no. S4696, Sigma-Aldrich Merck). SB505124 was added to filtered seawater 3 h after fertilization, while the other drugs were added 5 h after fertilization. Control embryos were treated with 0.1% DMSO. At 15 h after fertilization, treated embryos were washed several times with filtered seawater to remove the drugs from the medium.

Western clawed frog (*X. tropicalis*) embryos were treated with the antagonists for the FGF and Nodal pathways from late blastula (stage 9) to late gastrula stage (stage 12). Drugs were dissolved in DMSO, and the following final concentrations were used: 5  $\mu\text{M}$  SU5402 (cat no. SML0444, Sigma-Aldrich Merck) and 100  $\mu\text{M}$  SB505124 (cat no. S4696, Sigma-Aldrich Merck). Control embryos were treated with 0.1% DMSO. After treatment, embryos were washed with medium and subsequently used for total RNA extraction.

*P. flava* adults were collected from Penghu Islands, Taiwan. Methods for spawning induction and embryo culture were described previously (34). *P. flava* embryos were treated with DMSO (control), SU5402 (20  $\mu\text{M}$ ) (35), or SB505124 (10  $\mu\text{M}$ ) at the late blastula stage (18 hpf) and collected at the late gastrula stage (43 hpf) for RNA-seq.

**ATAC-seq.** ATAC-seq assays were performed in at least two biological replicates. Forty-five amphioxus embryos and 20 zebrafish embryos were dissociated into individual cells. After the number of cells was counted, around 70,000 cells were transferred to another tube to perform the experiment. ATAC-seq experiments were performed as previously described (6, 8). ATAC-seq data analyses were performed using standard pipelines (6, 8). Reads were aligned with Bowtie2 using GRCz10 (danRer10) and B171 (6) assemblies for zebrafish and amphioxus samples, respectively. Reads that were separated by more than 2 kb were filtered out of the analysis. The exact position of the Tn5 cutting site was determined as the position  $-4$  (reverse strand) or  $+5$  (forward strand) from the read start, and this position was extended 5 kb in both directions. BED files were transformed into BigWig using the wigToBigWig University of California, Santa Cruz (UCSC) tool. Reads were extended 100 bp in order to visualize the data in the UCSC Genome Browser (36). Macs2 (37) software was used to perform the peak calling in each sample, using the parameters  $-\text{nomodel}$ ,  $-\text{shift } 45$ , and  $-\text{extsize } 100$ . The different called peaks of each sample were merged into a unique set of peaks, taking into account the replicates (two per sample). Using Bedtools (38), the number of reads per called peak and per sample in both treatment and control conditions was subsequently computed, and a differential analysis was performed using DESeq2 v1.18.0 in R 3.4.3 (39). A corrected  $P$  value  $<0.05$  was set as cutoff for statistical significance of the differential accessibility of the chromatin in ATAC-seq peaks. Motif enrichment was calculated using the program FindMotifsGenome.pl from the Homer tool suite (40).  $k$ -means clustering of the ATAC-seq signal was performed using Deeptools 2.0 (41) and seqMiner (42). The assignment of ATAC-seq peaks to genes was done using the GREAT tool (43), with default parameters for basal plus extension regions calculation. We verified that the results were robust regardless of the peak-to-gene association method by comparing the GREAT association using the basal plus extension with an association to the closest genes. The overlap of genes associated by both methods was high (SI Appendix, Fig. S16), and their functions were similar. Indeed, we calculated a GO enrichment overlap score of 0.94 out of 1 using the GOSemSim R package. GO analyses were carried out with the TopGO R package, using the *elim* test for selecting the most specific GO-enriched terms.

**RNA-seq.** RNA-seq experiments were performed in three biological replicates for each species. RNA samples were extracted from 15 zebrafish, 100 amphioxus, or 20 western clawed frog embryos following previously published protocols (6). For the data analysis, reads were aligned against GRCz10 (danRer10) and B171 assemblies using STAR v2.5.3a (44) and were assigned to genes using the HTSeq toolkit v0.11.2 (45). In the case of *P. flava* and *X. tropicalis*, reads were pseudoaligned to the set of longest isoforms of the currently available transcriptome of *P. flava* (46) and to the set of primary coding DNA sequences of the XenTrop v10.0 transcriptome (47), respectively, using Kallisto with standard parameters (48). Differential gene expression analyses were performed using DESeq2 v1.18.0 (39). A corrected  $P$  value  $<0.05$  and an absolute  $\log_2$  fold

change (FC)  $>1$  were used as thresholds for calling differential genes. The enrichment of biological process GO terms was calculated using the TopGO R package. Gene clustering was performed using the Pheatmap 1.0.12 R package, with  $k$ -means as clustering method. The integration of differential ATAC-seq peaks and differential expressed genes was performed at the gene level in both species. The gene lists that resulted from these analyses were intersected for a given treatment to identify genes with a differential RNA-seq signal and one or more associated differential ATAC-seq peaks. Since there is currently no functional annotation available for the amphioxus genes, the orthologous zebrafish genes were used for completing the GO term enrichment analysis. Amphioxus versus zebrafish orthologous genes were already available from previous studies (6) (Dataset S7).

**Connectivity Analyses.** In order to categorize genes that respond to more than one treatment, we computed a connectivity score based exclusively on genes responding at both the ATAC-seq and the RNA-seq levels. Each gene was assigned a discrete score ranging from 1 to 4 and corresponding to the number of different treatments that it responded to. Cytoscape (31) networks were generated to better represent the connectivity of responsive genes. The connectivity tables are provided in Dataset S7. Statistical analyses were performed using the programming language R.

**Single-Cell RNA-seq (scRNA-seq) Analysis.** We selected genes with high connectivity ( $\geq 3$ ) and low connectivity ( $=1$ ) from the previously published scRNA-seq zebrafish developmental atlas (30) and explored their expression levels in a set of ancient and vertebrate-specific novel tissues. A gene was defined as expressed in a cell if its normalized expression in that cell was greater than 0. For a given tissue, we calculated the proportions of cells that expressed a certain gene and aggregated these proportions for genes with high or low connectivity scores in a boxplot. To determine which genes were expressed in each tissue, we defined a threshold of a minimum of 5% of the cells of that specific tissue. Subsequently, for a direct comparison between a set of ancient tissues (present in both zebrafish and amphioxus) and a set of vertebrate-specific tissues, we computed—and represented in barplots—the ratios between genes with high connectivity scores and genes with low connectivity scores expressed in a given tissue. Wilcoxon test ( $P < 0.05$ ) was used in order to look for statistically significant differences.

**Data Availability.** Data sets presented in this study have been deposited in the Gene Expression Omnibus (GEO) database (accession no. GSE148783) (49, 50).

**ACKNOWLEDGMENTS.** We would like to acknowledge the impact on our research from our mentor and friend J.L.G.-S. We thank all members of the J.L.G.-S. laboratory for fruitful discussions and the Universidad Pablo de Olavide supercomputing cluster C3UPO for high-performance computing support. We especially thank José María Santos-Pereira, Christina Paliou, Ignacio Maeso, and Manuel Irimia for helpful comments on the manuscript. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant agreement no. 740041) and the Spanish Ministerio de Economía y Competitividad (Grants BFU2016-74961-P and PID2019-103921GB-I00 to J.L.G.-S. and J.J.T.). This work was also supported by the institutional grant Unidad de Excelencia María de Maeztu (Grant MDM-2016-0687 to the Department of Gene regulation and morphogenesis of Centro Andaluz de Biología del Desarrollo). M.F. was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement [#800396]. M.S., H.E., and S.B. were supported by the CNRS, and H.E. and S.B. additionally by Agence Nationale de la Recherche (ANR) CHORELAND (Grant ANR-16-CE12-0008-01) and the Institut Universitaire de France. Y.-H.S. and J.-K.Y. are supported by intramural funds from Academia Sinica and grants from Ministry of Science and Technology, Taiwan (Grants 110-2326-B-001-006 to Y.-H.S. and 110-2621-B-001-001-MY3 to J.-K.Y.).

Author affiliations: <sup>a</sup>Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas-Universidad Pablo de Olavide-Junta de Andalucía, 41013 Seville, Spain; <sup>b</sup>Institute of Cellular and Organismic Biology, Academia Sinica, 11529 Taipei, Taiwan; <sup>c</sup>Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica, 26242 Yilan, Taiwan; <sup>d</sup>Biologie Intégrative des Organismes Marins, Observatoire Océanologique, Sorbonne Université, CNRS, 66650 Banyuls-sur-Mer, France; and <sup>e</sup>Laboratoire de Biologie du Développement de Villefranche-sur-Mer, Institut de la Mer de Villefranche, Sorbonne Université, CNRS, 06230 Villefranche-sur-Mer, France

1. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabási, Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
2. A. L. Barabási, Z. N. Oltvai, Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
3. E. Azpeitia *et al.*, The combination of the functionalities of feedback circuits is determinant for the attractors' number and size in pathway-like Boolean networks. *Sci. Rep.* **7**, 42023 (2017).
4. A. Pires-daSilva, R. J. Sommer, The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* **4**, 39–49 (2003).
5. N. H. Putnam *et al.*, The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
6. F. Marlétaz *et al.*, Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 (2018).
7. S. Bertrand, Y. Le Petillon, I. M. L. Somorjai, H. Escriva, Developmental cell-cell communication pathways in the cephalochordate amphioxus: Actors and functions. *Int. J. Dev. Biol.* **61**, 697–722 (2017).
8. J. D. Buenostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
9. P. Dehal, J. L. Boore, Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
10. A. Christoffels *et al.*, Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **21**, 1146–1151 (2004).
11. C. Kiecker, T. Bates, E. Bell, Molecular specification of germ layers in vertebrate embryos. *Cell. Mol. Life Sci.* **73**, 923–947 (2016).
12. F. B. Tuazon, M. C. Mullins, Temporally coordinated signals progressively pattern the anteroposterior and dorsoventral body axes. *Semin. Cell Dev. Biol.* **42**, 118–133 (2015).
13. R. T. Böttcher, C. Niehrs, Fibroblast growth factor signaling during early vertebrate development. *Endocr. Rev.* **26**, 63–77 (2005).
14. K. M. Cadigan, M. L. Waterman, TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb. Perspect. Biol.* **4**, a007906 (2012).
15. R. M. Charney, K. D. Paraiso, I. L. Blitz, K. W. Y. Cho, A gene regulatory program controlling early *Xenopus* mesendoderm formation: Network conservation and motifs. *Semin. Cell Dev. Biol.* **66**, 12–24 (2017).
16. S. S. Bian *et al.*, Clock1a affects mesoderm development and primitive hematopoiesis by regulating Nodal-Smad3 signaling in the zebrafish embryo. *J. Biol. Chem.* **292**, 14165–14175 (2017).
17. S. Jia, Z. Ren, X. Li, Y. Zheng, A. Meng, smad2 and smad3 are required for mesendoderm induction by transforming growth factor- $\beta$ /nodal signals in zebrafish. *J. Biol. Chem.* **283**, 2418–2426 (2008).
18. R. A. S. Kjolby, M. Truchado-Garcia, S. Iruvanti, R. M. Harland, Integration of Wnt and FGF signaling in the *Xenopus* gastrula at TCF and Ets binding sites shows the importance of short-range repression by TCF in patterning the marginal zone. *Development* **146**, dev179580 (2019).
19. T. Tian, A. M. Meng, Nodal signals pattern vertebrate embryos. *Cell. Mol. Life Sci.* **63**, 672–685 (2006).
20. J. R. Friedman, K. H. Kaestner, The Foxa family of transcription factors in development and metabolism. *Cell. Mol. Life Sci.* **63**, 2317–2328 (2006).
21. N. B. Ghyselinck, G. Duester, Retinoic acid signaling pathways. *Development* **146**, dev167502 (2019).
22. P. A. Hoodless *et al.*, FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev.* **15**, 1257–1271 (2001).
23. P. Joshi, A. J. Darr, I. Skromne, CDX4 regulates the progression of neural maturation in the spinal cord. *Dev. Biol.* **449**, 132–142 (2019).
24. R. A. S. Kjolby, R. M. Harland, Genome-wide identification of Wnt/ $\beta$ -catenin transcriptional targets during *Xenopus* gastrulation. *Dev. Biol.* **426**, 165–175 (2017).
25. D. Aldea *et al.*, Genetic regulation of amphioxus somitogenesis informs the evolution of the vertebrate head mesoderm. *Nat. Ecol. Evol.* **3**, 1233–1240 (2019).
26. T. Onai, Canonical Wnt/ $\beta$ -catenin and Notch signaling regulate animal/vegetal axial patterning in the cephalochordate amphioxus. *Evol. Dev.* **21**, 31–43 (2019).
27. Y. Yasuoka, Y. Tando, K. Kubokawa, M. Taira, Evolution of cis-regulatory modules for the head organizer gene *gooseoid* in chordates: Comparisons between *Branchiostoma* and *Xenopus*. *Zoological Lett.* **5**, 27 (2019).
28. C. Y. McLean *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
29. J. L. Royo *et al.*, Transphyletic conservation of developmental regulatory state in animal evolution. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14186–14191 (2011).
30. D. R. Farnsworth, L. M. Saunders, A. C. Miller, A single-cell transcriptome atlas for zebrafish development. *Dev. Biol.* **459**, 100–108 (2020).
31. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
32. M. Fuentes *et al.*, Preliminary observations on the spawning conditions of the European amphioxus (*Branchiostoma lanceolatum*) in captivity. *J. Exp. Zool. B Mol. Dev. Evol.* **302**, 384–391 (2004).
33. M. Fuentes *et al.*, Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B Mol. Dev. Evol.* **308**, 484–493 (2007).
34. C. Y. Lin, C. H. Tung, J. K. Yu, Y. H. Su, Reproductive periodicity, spawning induction, and larval metamorphosis of the hemichordate acorn worm *Ptychodera flava*. *J. Exp. Zool. B Mol. Dev. Evol.* **326**, 47–60 (2016).
35. T. P. Fan, H. C. Ting, J. K. Yu, Y. H. Su, Reiterative use of FGF signaling in mesoderm development during embryogenesis and metamorphosis in the hemichordate *Ptychodera flava*. *BMC Evol. Biol.* **18**, 120 (2018).
36. J. Casper *et al.*, The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.* **46** (D1), D762–D769 (2018).
37. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
38. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
39. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
40. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
41. F. Ramírez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–5 (2016).
42. T. Ye *et al.*, seqMINER: An integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).
43. M. Hiller *et al.*, Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: Application to zebrafish. *Nucleic Acids Res.* **41**, e151 (2013).
44. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
45. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
46. O. Simakov *et al.*, Hemichordate genomes and deuterostome origins. *Nature* **527**, 459–465 (2015).
47. J. D. Fortriede *et al.*, Xenbase: Deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. *Nucleic Acids Res.* **48** (D1), D776–D782 (2020).
48. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
49. A. Gil-Gálvez *et al.*, Gain of gene regulatory network interconnectivity at the origin of vertebrates. Gene Expression Omnibus (GEO) Database. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148783>. Deposited 16 April 2020.
50. M. Franke *et al.*, Functional genomic and transcriptomic analysis of ptychodera and the origin of vertebrate genomic traits. Gene Expression Omnibus (GEO) Database. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE189808>. Deposited 29 November 2021.