Insights into deuterostome evolution from the biphasic transcriptional programmes of hemichordates

Alberto Perez-Posada*[1,2], Che-Yi Lin[3], Ching-Yi Lin[3], Yi-Chih Chen[3], José Luis Gómez Skarmeta[†,1], Jr-Kai Yu[3,4], **Yi-Hsien Su***[3], **Juan J. Tena***[1]

[1] Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas-Universidad Pablo de Olavide-Junta de Andalucía, 41013 Seville, Spain

[2] Department of Biological and Medical Sciences, Oxford Brookes University, Headington Campus, Gipsy Lane, OX3 0BP Oxford, United Kingdom

[3] Institute of Cellular and Organismic Biology, Academia Sinica, 11529 Taipei, Taiwan

[4] Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica, 26242 Yilan, Taiwan

[†] Deceased September 16, 2020

* Corresponding authors

Abstract

Evolutionary history of deuterostomes has remained unsolved and is intimately related to how chordates originated. Within the clade of deuterostomes, hemichordates and echinoderms (together called Ambulacraria) are sister groups of chordates; comparative studies among the three are insightful for understanding deuterostome evolution. Indirect-developing hemichordates, such as *Ptychodera flava*, develop into planktonic larvae, which are morphologically similar to echinoderm larvae, before metamorphosing into the adult body plan that retains an anteroposterior polarity homologous to that of chordates. Therefore, deciphering developmental programmes of indirect-developing hemichordates may provide a key for understanding the evolution of deuterostomes and chordate origins. Here, we characterise transcriptomes and chromatin accessibility across multiple developmental stages of *P. flava* and uncover biphasic developmental programmes controlled by different sets of transcription factors and their corresponding cis-regulatory elements. Transcriptome age and network analyses reveal that the gastrula transcriptome is relatively ancient and highly connected. By comparing

developmental transcriptomes of hemichordate, sea urchin and amphioxus, a high conservation of gene expression during gastrulation is identified and extended to neurula stages of amphioxus, in addition to the highly similar larval transcriptomes among the three species. Moreover, *P. flava* possesses conserved interactions of transcription factors required for the development of echinoderm endomesoderm and chordate axial mesoderm. These results suggest a deuterostome phylotypic stage during gastrulation (corresponding to amphioxus gastrulation and neurulation) controlled by gene regulatory networks with conserved cis-regulatory interactions, and support the hypothesis that the deuterostome ancestor is an indirect-developer.

Introduction

The deuterostome lineage comprises diverse body plans in such different animals like echinoderms and tetrapods. Studies of specific groups located in key positions of the deuterostome phylogenetic tree can help us to understand the complex evolution of these animals. One of these groups is formed by the hemichordates, a sister group of echinoderms that share important developmental and anatomical features with chordates, which place them in a strategic position for comparative studies.

Hemichordates share bilateral symmetry with chordates, yet with a distinctive tripartite body plan comprising a proboscis, a collar and a trunk (Urata and Yamaguchi 2004; Swalla 2007). They show conserved structures such as pharyngeal gill slits, an ancestral character that support the position of hemichordates within deuterostomes (Simakov et al. 2015; Lowe et al. 2015). Importantly, a number of hemichordate structures utilise highly conserved patterning mechanisms for their development, including the overall anteroposterior (AP) axis regionalization (Lowe et al. 2003) and the AP patterning of the ectoderm in the proboscis and collar region by similar signalling organising centres to those of the vertebrate central nervous system (ANR, ZLI and Iso) (Pani et al. 2012), and the formation of coelomic cavities from mesoderm pouches deriving from an endomesoderm region (Green et al. 2013; Fan et al. 2018; Kaul-Strehlow and Stach 2013). These are complemented by numerous expression patterns of neural and developmental marker genes similar to that of vertebrates (Lowe 2021). The patterning mechanism of the dorsoventral (DV) polarity is also similar, but chordates

appear to invert their DV axis (Lowe et al. 2006; Su et al. 2019). Other functionally analogous structures with uncertain homology include a collar nerve cord that follows a morphogenesis process similar to chordate neurulation (Miyamoto and Wada 2013; Nakajima et al. 2004) but employing divergent genes for neuronal patterning (Kaul-Strehlow et al. 2017). Despite remarkable efforts in elucidating the genomic complexity and patterning mechanisms in hemichordates, a comprehensive understanding of the hemichordate developmental programs and how the genome is temporally regulated is still lacking. Additionally, questions remain as to whether a comparable developmental period can be recognized in deuterostomes, and to what extent the developmental gene regulatory networks (GRNs) are conserved among different deuterostome groups.

To address these questions, we characterised developmental transcriptomes and chromatin accessibility of *Ptychodera flava*, an indirect-developing hemichordate. We identified and compared clusters of stage-specific modules of coexpressed genes controlled at the cis-regulatory level by likewise stage-specific classes of transcription factors (TFs). Integration of chromatin and gene expression data revealed cis-regulatory gene networks that recapitulate gene-gene interactions conserved throughout deuterostomes, including core interactions of the GRN controlling axial mesoderm in chordates. Comparisons with other deuterostomes revealed similar programmes of gene family and transcription factor usage throughout development, with conserved expression of gene families at early and larval stages. Despite the morphological differences between deuterostome phyla, functional genomics of hemichordates showcase similarities in the control of gene expression during deuterostome development, as well as cis-regulatory interactions pre-dating the emergence of chordates.

Results

Two major transcriptional programmes controlled by different TF classes

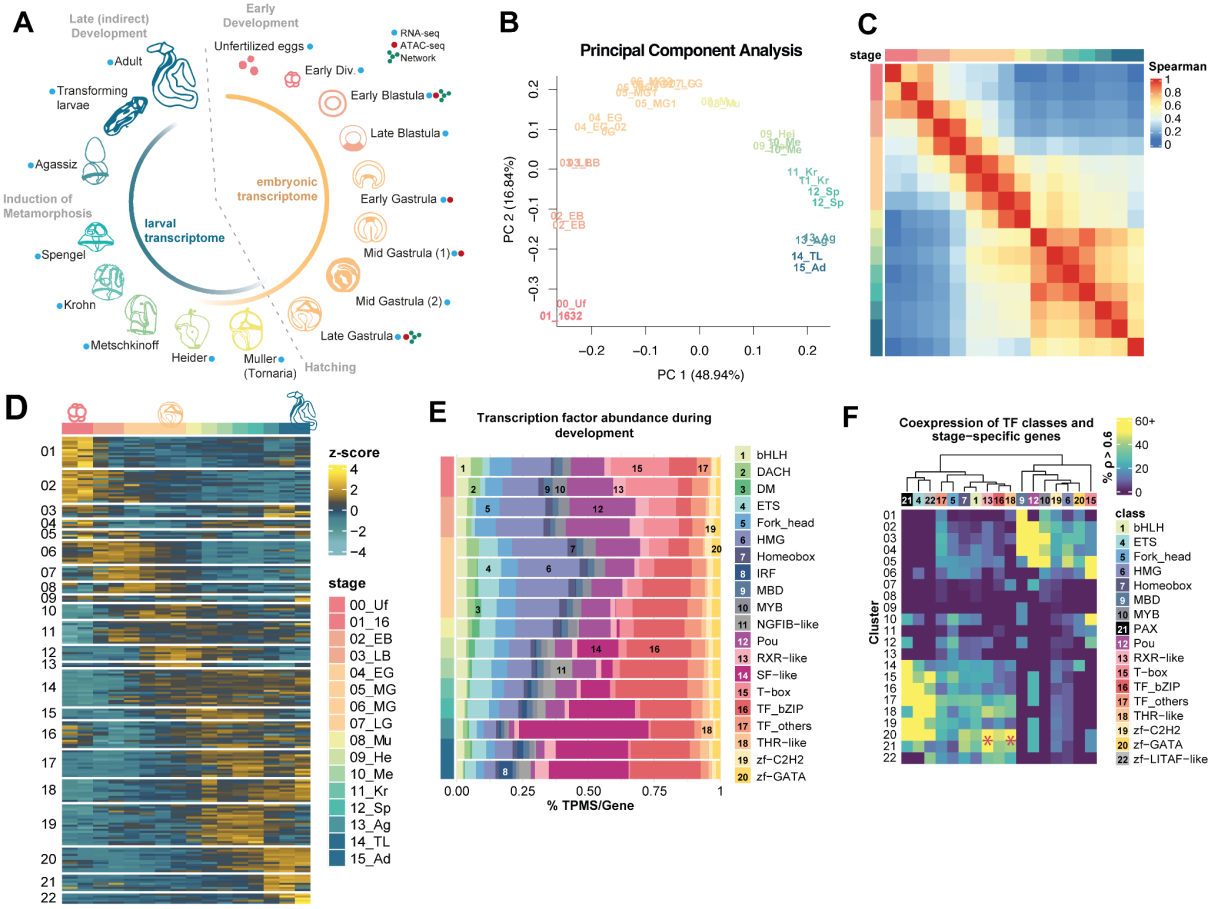

Figure 1: Transcriptomics of *P. flava* during development. A: Life cycle of *P. flava* and sample collection. Grey dashed line represents the time of hatching, and inner coloured lines represent the duration of the embryonic and larval transcriptome. B: PCA analysis of the developmental transcriptomes. C: Two major transcriptional programmes (phases) during development: embryonic and larval development. Pairwise Spearman correlation of transcriptomes of each stage (quantile-normalised, averaged replicates). D: 22 clusters of dynamically expressed genes among the 16 stages . Hierarchical clustering of stage-specific genes by peak expression throughout development (for visualisation, a random subset of 200 genes). E-F: The two phases are orchestrated by distinct sets of TF classes E: Relative expression (transcripts per million per gene) of each major expressed class during development. F: Coexpression of TFs and stage-specific genes. For a given TF class X and cluster Y, fraction of X genes with a spearman correlation above 0.6 with the average z-score profile of genes in cluster Y.

Indirect-developing hemichordates undergo several distinct phases of developmental processes. After 2 days of embryogenesis, the late gastrula of *P. flava* hatch and undergo a series of planktonic, tornaria larval stages that stay in the water column

for months before metamorphosing into a worm-like body plan (Fig. 1A). Given that changes in gene expression represent one of the most important driving forces of phenotypic changes and animal evolution, we measured gene expression at sixteen stages, covering the entire *P. flava* life cycle, to understand the dynamics of gene usage during its development (Fig. 1A). Principal Component Analysis (PCA) of these transcriptomes reflects the chronological order of the developmental stages from which they were obtained (Fig. 1B, Supplementary Fig. S1A), validating the quality of this data series. Overall, we determined expression levels of 28413 genes across all stages (Supplementary Table 1). Similarities at the whole-transcriptome level suggest two major transcriptional programmes of gene expression: early (embryonic) versus late (larval) that run independently (Fig. 1C). Differential gene expression analysis throughout development identified 23726 dynamically expressed genes (hereafter referred to as stage-specific genes), which represent 83% of all genes detected and can be grouped into twenty-two clusters based on their temporal gene expression profiles (Fig. 1D, Supplementary Data 1). Several of these gene clusters clearly correspond to their deployment at different developmental phases (e.g. clusters 1-2 match to maternal factors used in early embryogenesis; clusters 6-8 are activated during embryogenesis; clusters 15-19 are generally deployed during the larval phase; clusters 20-22 are predominantly utilised during metamorphosis). We noted that genes expressed at the Muller stage (the newly hatched tornaria) include clusters deployed during embryonic (clusters 12-13) and later larval stages (clusters 15-17 and 19). This observation suggests that Muller stage represents the hinge between embryonic and larval transcriptional states, since its transcriptomic profile is essentially employing embryonic-specific genes, but starting to shift to a 'larval transcriptomic programme'. Gene Ontology (GO) enrichment analysis of these clusters revealed major developmental processes taking place at the time of peak expression (Table 1). For example, we found GO terms associated with cell cycle divisions and activation of transcription during early divisions, morphogenetic pathways during gastrulation, and formation of body/larval-specific structures at late development clusters (Supplementary figure S2, Supplementary Data 2). These results corroborate our understanding on the temporal controls of transcriptional programmes orchestrating different developmental processes.

Table 1: Description of differentially expressed, stage-specific gene clusters.

Cluster ID	Cluster size	Time of expression	Description based on GO terms
1	1671	Early Divisions	Protein (de)ubiquitination & phosphorylation, cell response to starvation, DNA methylation
2	1352	Early Divisions	RNA processing, histone modification, DNA repair, sister chromatid cohesion, nuclear mRNA surveillance, initiation of transcription
3	879	Early Div., Agassiz, Adult	Phospholipid catabolic process, regulation of developmental growth, transport & catabolism of biomolecules
4	696	Early Divisions, Larval stages	Lipid homeostasis, vesicle and endosome organisation, sterol biosynthesis
5	508	Early Div., Early Blastula	Positive regulation of Wnt pathway, protein polyubiquitination, spindle assembly, membrane fusion, mitochondrial fission, regulation of gene expression, histone acetylation
6	1436	From Early Div. to Mid Gastula	mRNA splicing, RNA processing, DNA replication, centromere assembly, mitosis
7	1204	From Early Blast. to Mid Gastr.	RNA processing, gene expression, histone acetylation, patterning formation
8	1247	From Late Blast. To Mid Gastr.	Telomere maintenance, protein folding, cilium assembly, mRNA splicing, IMP biosynthesis. Vesicle fusion, histone exchange
9	226	Early Div, Mid Gastr., Adult	Response to nitrogen compound, DNA integration
10	1038	Mid Gastrula to Larval stages	glycolipid, IMP, ATP biosynthesis, protein glycosylation, organ morphogenesis & development, cellular respiration
11	1296	Early Blast. To Larval stages	Protein degradation, DNA repair, Wnt pathway, vesicle formation, patterning, respiration
12	582	Late Gastrula to Metschkinoff	Muscle contraction, calcium transport, lipid metabolism
13	321	Late Gastrula, Tornaria	Sphingolipid biosynthesis, cilium assembly, cell projection, sodium transport. Potentially related to axon formation and/or larval cillia
14	1539	Mid Gastrula to Adult	Ribosomal assembly, vesicle transport, exocitosis, actin filament processes.
15	943	Late Gastrula to Agassiz	Axoneme assembly, mesenchyme development, cilium formation. Potentially related to axon formation and/or larval cillia
16	1276	Early Div., Tornaria to Adult	Cilia motility, axon formation and guidance, Smoothened pathway, protein transport
17	1512	Heider to Agassiz, Adult	Signal transduction, ERK cascade, ion transport, actin nucleation
18	895	Heider to Adult	Muscle development, Notch and TLR pathways, thyroid hormone metabolism, proteolysis, DNA integration, apoptosis
19	1440	Heider to Spengel	Ion and phospholipid transport, telomere organisation, glycoside metabolism, netrin pathway
20	1379	Krohn to Adult	Ion transport, signal transduction, nervous system process, cell communication, gluconeogenesis
21	1242	Agassiz to Adult	Transmembrane transport, trans-synaptic signalling, regulation of growth, protein glycosylation, nucleotide biosynthesis,

			catecholamine metabolism, thyroid hormone metabolism
22	1044	Transforming larvae, Adult	Apoptosis, regeneration, immune system development, SMAD transduction, steorid biosynthesis, hematopoiesis or lymphoid organ development, TLR pathway,

To understand how the transcriptional programmes are controlled, we focused on the usage of transcription factors (TFs) during *P. flava* development. For an approach to identify the underlying gene expression regulatory machinery, we classified the transcription factors into major classes (following animalTFDB and automated orthology relationships with vertebrates) and looked at their expression throughout development (Supplementary Data 2, Supplementary Data 3). We observed two distinct programmes of TF composition, with a high prevalence of T-box, Pou and High Mobility Group (HMG, which include Sox) that are maternally deposited and used at early stages, and SF-like and bZIP classes taking over at larval stages (Fig. 1E), and with a higher number of Homeobox TFs expressed at later stages) (Supplementary Fig S3A,D). The timing of these TF classes generally correspond to their roles in development in other species. For example, the Pou and T-box (including Brachyury) classes, mediate pluripotency and cell fate specification in early stages and gastrulation (Nichols et al. 1998; Niwa, Miyazaki, and Smith 2000; Babaie et al. 2007; Osorno and Chambers 2011), while SF-like classes related to steroidogenic signalling and gonad maturation function at later stages (Mangelsdorf et al. 1992) Consistently, these trends roughly match similar observations in sea urchin and amphioxus (Supplementary Figure S3B,C,E-J), and published data in ecdysozoa (Adryan and Teichmann 2010), with the early and late deployment of HMG and bZIP factors respectively. Additionally, several TF classes exhibit high agreement in coexpression with stage-specific gene clusters throughout development, supporting that the two distinct developmental programmes of embryonic and larval development are controlled by different TF classes (Fig.1F). Notably, RXR (class 13) and Thyroid hormone (THR) (class 18) pathways are involved in metamorphosis in a variety of animals (X. Wang, Matsuda, and Shi 2008; Shao et al. 2017), and their co-expression with clusters 20-21 (red asterisks in Fig. 1F), genes predominantly expressed during metamorphosis, suggests the similar role in *P. flava*. These results indicate that the two major developmental programmes of *P. flava* are controlled by two sets of TFs to perform stage-specific processes known in other deuterostomes.

Gastrulation retains the oldest transcriptome of *P. flava*

To further delineate the evolutionary history of the different transcriptional programmes, we sought to investigate the evolutionary conservation of the gene expression programme underlying *P. flava* development. To this end, we firstly characterised the gene age of the whole annotated genome of *P. flava*. We assigned groups of orthology (orthogroups) using OMA (Altenhoff et al. 2021) and a comprehensive set of metazoan species, with an emphasis on deuterostomes and ambulacraria (Fig. 2A, Supplementary Data 4). In turn, we assigned a gene age to every orthogroup by Dollo parsimony (Csurös 2010), resulting in ancient genes present in all species and younger genes found only in *P. flava* (see Methods) (Fig. 2A).

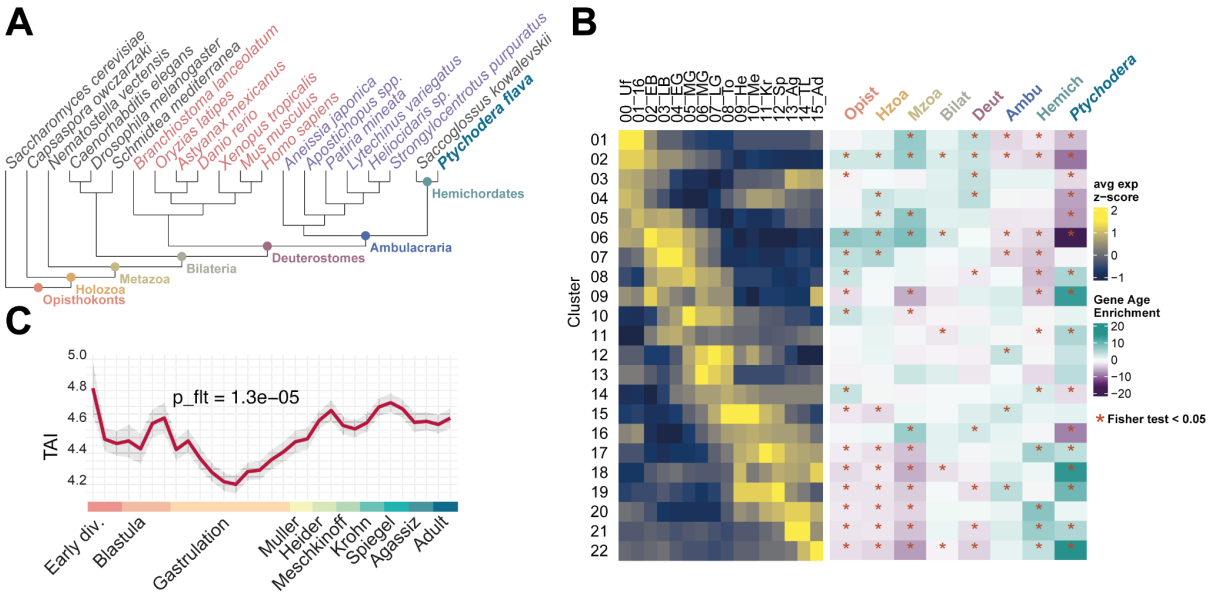

Figure 2: Gene Age of the developmental transcriptional programme in *P. flava*. A: Tree of species used for the orthology and gene age assignment analysis. B: Age of the 22 clusters: (Left) Schematics of average gene expression for each stage-specific cluster (summary of Fig. 1D); and (Right) Gene age enrichment of stage-specific clusters. C: Transcriptomic Age Index (TAI) throughout the development of *P. flava.*

We first calculated gene age enrichment for every cluster of stage-specific genes, by comparison with gene age of all annotated *P. flava* genes. We found that clusters of early stage genes tended to be enriched in genes from Opisthokonta all the way up to Bilateria, but depleted in ambulacraria- or *Ptychodera*-specific genes, and an

opposite trend was found in clusters of later stages with high enrichments in ambulacraria-, hemichordate- or *Ptychodera*-specific genes (Fig. 2B). This result suggests that older genes are used for embryogenesis, while younger genes are deployed for larval development and the metamorphosis process leading to the adult body plan. This matches reports of similar trends in echinoderms (Li et al. 2020). To further evaluate the transcriptome age among different developmental stages, we used the same approach to calculate the Transcriptomic Age Index (TAI) (Domazet-Lošo and Tautz 2010). We identified that the transcriptome of mid-to-late gastrula is the oldest among all developmental stages (Fig. 2C), suggesting that gastrulation represents a more conserved period, relative to the other developmental stages.

Cis-regulation and gene regulatory networks during hemichordate gastrulation

For a better understanding of how the gene expression programme is controlled at the cis-regulatory level before and during the conserved gastrulation period, we identified open chromatin regions in the genome using Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) (Buenrostro et al. 2013) at four time points: early blastula, early-, mid-, and late gastrula (Fig. 1A, indicated with the red dots). We identified 82807 open chromatin regions (OCRs) that are potential cis-regulatory elements (REs) throughout gastrulation in *P. flava*, mostly located in the ±20kb vicinity of genes, specifically within intergenic regions or in introns, followed by within promoters (Fig. S4A,B Supp, Supplementary Data 5,6). Binding motifs for distal regulators such as developmental TFs appear more frequently outside promoters and in distal elements (Figure S5 Supp). Nearly eighty percent of the annotated genes have at least one OCR associated, with more than fifty percent associated with two or more OCRs (Fig. S4C Supp). We observed a trend of RE openness in accordance with the expression level of their associated gene (Fig. S4D Supp), suggesting cis-regulation generally occurs as transcriptional activation. Genes annotated as trans-developmental (with GO terms in transcription or development, as similarly defined in (Marlétaz et al. 2018), see Methods) tend to be associated to cis-regulatory elements, with more OCRs than housekeeping genes (Fig. S4E,F,G Supp), agreeing with former observations of trending towards

complexity of cis-regulatory architecture in the genes orchestrating metazoan development (Irimia et al. 2012; Marlétaz et al. 2018; Bogdanović et al. 2012).

Differential accessibility analysis between pairs of embryonic stages identified 15946 dynamically accessible OCRs, which were categorised into six clusters (Fig. 3A). These six clusters were enriched with specific DNA binding motifs of major TF classes that correspond to their roles in controlling gene expression during development in other metazoan lineages (Fig. 3B, Figure S6 Supp). Motifs of blastula-specific OCRs (01_EB) show little overlap with gastrula-specific OCRs (03_EG_MLG to 06_LG) (Fig. 3B). For example, motifs of T-box transcription factors (including Brachyury, Eomes, Tbr and Tbx) are enriched in the OCRs at the early blastula and early gastrula stages. On the other hand, Sox binding motif is enriched in OCRs after gastrulation occurs. The dynamic of motifs enriched in stage-specific OCRs is consistent with the transcriptome data, in which T-box and HMG (including Sox) factors are highly expressed during embryogenesis, with T-box predominantly present before gastrulation and HMG more abundant during gastrulation (Fig. 1E). Other examples of stage-specific OCRs include binding motifs of E2F and the p53/63/73 family that are related to cell divisions, and of Dlx, Fox and Gsc that are known to express in different germ layers during *P. flava* gastrulation (Fig. S6 Supp) (Su et al. 2019; Harada et al. 2001; E. Röttinger and Martindale 2011; Taguchi et al. 2000). Although only ATAC-seq data of embryonic stages is available, assuming that some of the enhancers are used at multiple stages (Sebé-Pedrós et al. 2018, 2016), we crossed ATAC-seq data with stage-specific genes of all stages and observed two phases of motif usages during *P. flava* development (Fig. 3C). This is consistent with the two transcriptional programmes shown in Fig. 1, with binding motifs of T-box, E2F and bHLH enriched in the OCRs of early-expressed genes and of bZIP, Homeobox, Sox (HMG) and NR (nuclear receptor) at later stages (Fig. 3C). These results showcase a gradual shift of chromatin opening throughout *P. flava* development.

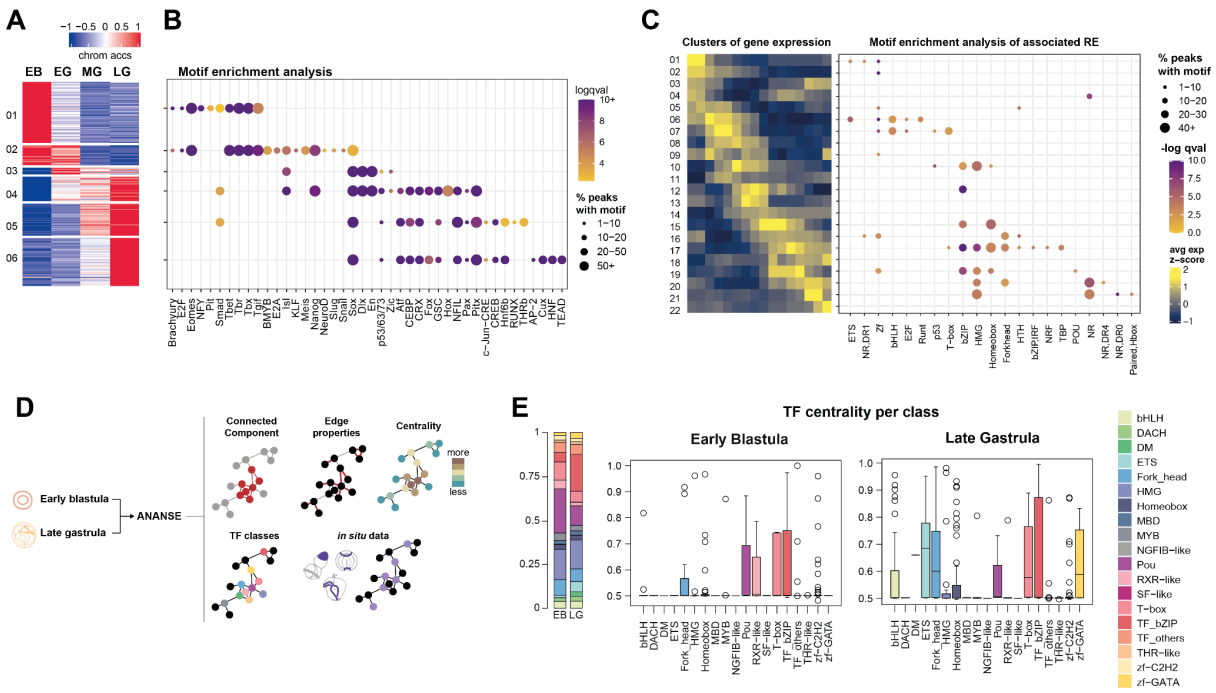

Figure 3: Cis-regulation of *P. flava* during gastrulation. A: Clustering of OCRs over time. B: Motif enrichment analysis of each cluster of OCRs over time. Dot plot for HOMER logqvalue (colour) and fraction of peaks with motif (dot size). C: Motif enrichment analysis of OCRs associated with stage-specific genes: (Left) Schematics of average gene expression for each stage-specific cluster (summary of Fig. 1D); (Right) Dot plot for HOMER -log(p.adj) (colour) and fraction of peaks with motif (dot size). D: Schematic of the ANANSE analysis for networks of cis-regulatory interactions. E: Left barplots: composition of major expressed TF classes at EB and LG stages; Right: distribution of network centrality for the common major expressed TF classes in EB and LG.

Processes in development do not result from independent actions of individual genes, but rather are driven by complex interactions between many genes, constituting the developmental GRNs. To better assess putative GRNs in *P. flava* we used ANANSE (Xu et al. 2021), a computational tool that uses cis-regulatory and gene expression data to predict GRNs and key TFs involved. These networks are composed of nodes (the TFs and target genes) and edges (the interactions between TFs and genes, whose values are calculated using gene expression, OCR accessibility, distance between OCR and the transcription start site, and motif enrichments within the OCR). We generated two independent networks, one of early blastula (EB) and one of late gastrula (LG) (Fig. 1A, 3D) (Supplementary Data 7,8). The EB network comprised 14000 genes and the LG 12000 genes (edge values above top 10%), with a higher interconnectivity at LG illustrated by larger size of the main connected component (CC) and the number of connections per target gene

(Fig S7A-C Supp). In these main CCs, 97 TFs are common between the two networks (Fig S7D Supp), and are enriched in terms of differentiation and organogenesis, while stage-exclusive TFs of each GRN are respectively enriched in cell cycle (EB network) and cell fate-specification (LG network), corresponding to the known developmental processes at these two stages (Fig. S7F-H Supp). Top-connected target genes exclusive for each network are enriched in similar terms of transcriptional regulation, with the LG targets enriched in Wnt pathway and cell fate commitment (Fig. S7E,I,J Supp).

To check the composition of the GRNs, we measured the centrality (Borgatti 2005) (how much a given gene is embedded within each GRN, related to the number of connections to other genes) of different genes in each network. We observed that trans-developmental genes seem to play especially important roles within the LG network, as illustrated by a higher number of these genes at central positions in the LG GRN, and by the presence of many post-gastrulation, stage-specific genes in the gastrula GRN (Fig. S7K-L Supp). We further inspected the centrality of different TF classes in each network. Overall, TF classes showing higher centrality correspond to the dominantly expressed TFs at the respective stage (Fig. 3E and Fig. S7M-N Supp). For example, T-box and Pou classes show high centrality in the EB network, and bZIP, ETS, GATA and Homeobox factors start to show higher centrality in the LG network (Fig. 3E). This suggests a different set of core TFs controlling the GRNs at each stage. Our results also show that compared to the EB GRN, more genes activated in the LG GRN are likely pleiotropic and continue to be deployed in subsequent larval stages.

To further inspect whether the GRNs can be subdivided into circuits that mediate germ layer formation, we investigated the position of *P. flava* genes that are known to show germ layer-specific expression at the LG stage in the LG network (Supplementary Data 9) (Fig. S8 Supp). Within those circuits, we notice that TF genes showing high centrality tend to express broadly, while low centrality genes often display restricted expression patterns in the respective germ layer. For example, xbp1 (Eric Röttinger et al. 2015), soxb2 and otx are the most connected genes within the ectoderm circuit (Fig. S9A Supp) and are known to express in broad ectodermal domains at the LG stage (Fig. S10A,B Supp). Similarly, genes show high

centrality in the mesoderm network (e.g., foxg and six3) and endoderm circuit (e.g., soxb1a and foxa) (Fig. S9B,C Supp) are expressed broadly in the mesoderm-derived protocoel and endoderm-derived archenteron of the LG embryos, respectively (Fig. S10F-G and S10K-L Supp). Genes that are known to express in restricted regions in the corresponding germ player, such as xlox (posterior ectoderm; (Ikuta et al. 2013)), foxq2 (anterior ectoderm; (E. Röttinger and Martindale 2011)), msx (dorsal mesoderm; (Eric Röttinger et al. 2015)), and tsg (anterior endoderm; (Eric Röttinger et al. 2015)) show low centrality in the respective circuit (Fig. S9 Supp). Other examples of TF genes with low centrality displaying restricted germ layer-specific expression patterns are shown in Fig. S10 Supp. The correlation between spatial expression patterns of TF genes and their hierarchical positions in GRNs coincides with the expected roles of TFs in specifying germ layers and the subdomains constituting the specific germ layer.

Altogether these results suggest a more interconnected GRN at late gastrula by addition of factors and targets to the blastula GRN. TF genes exhibiting high centrality correspond to observations at the gene expression level and their spatial expression patterns in specific germ layers. Relative to those at EB, trans-dev genes at LG are highly connected, which may impose more constraints on the gastrulation stage and prohibit its evolvability. This is consistent with the observation that more ancient genes are utilised during *P. flava* gastrulation (Fig. 2).

Comparative transcriptomics between *P. flava* and other deuterostomes

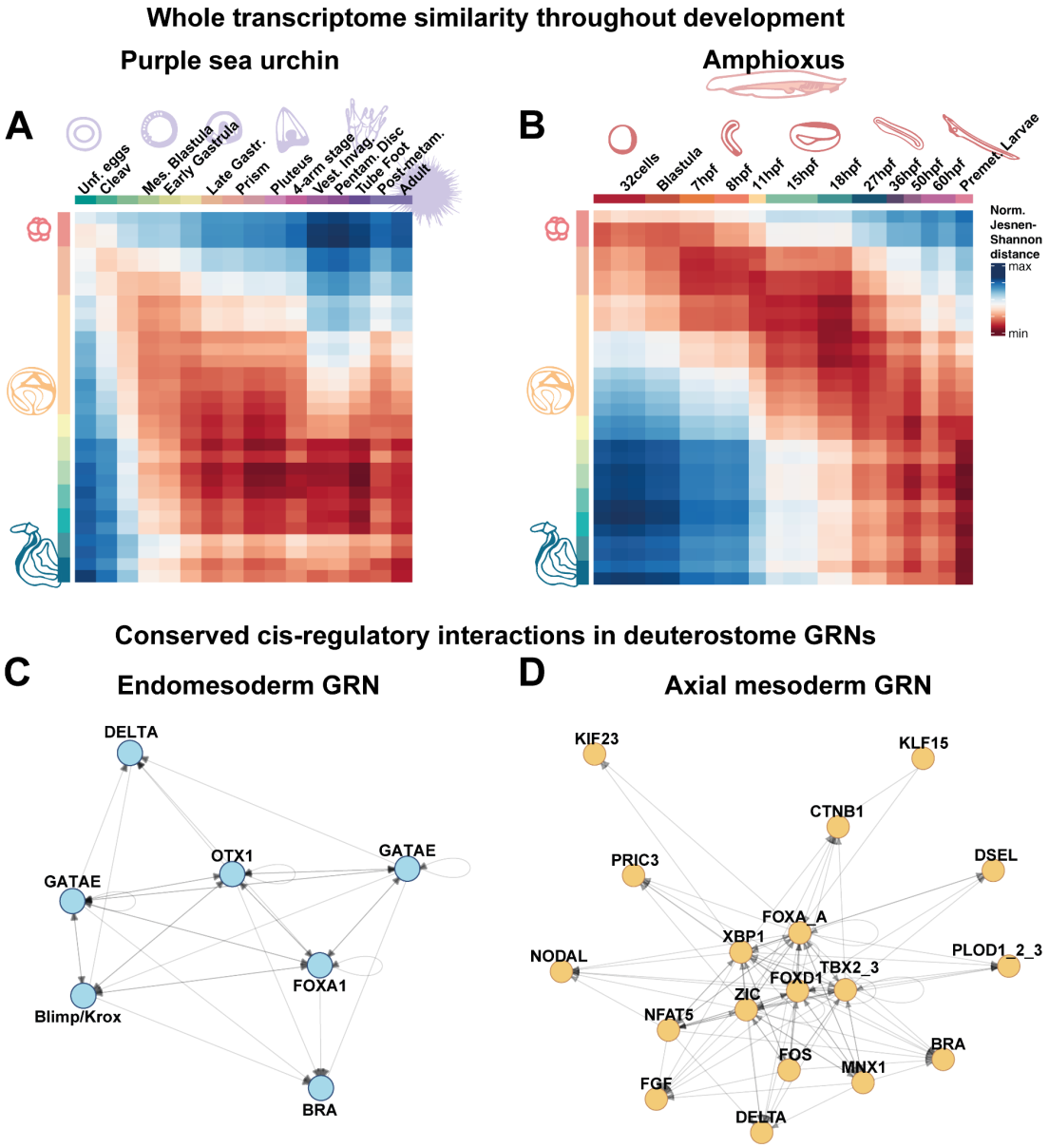

Figure 4: Gene age and evolutionary transcriptomics of *P. flava* under a deuterostome framework. A: Jensen Shannon Distance of pairwise-comparisons between *P. flava*and sea urchin developmental transcriptomes. B: Same analysis as D for *P. flava* and Amphioxus developmental transcriptomics. C: Cis-regulatory interactions of the echinoderm endomesoderm kernel in *P. flava*. D: Cis-regulatory interactions of the chordate axial mesoderm GRN in *P. flava*.

To globally assess the degree of conservation in developmental programs among the three major deuterostome groups, we performed pairwise comparisons of developmental transcriptomes between *P. flava* and two other invertebrate

deuterostome species: the purple sea urchin *Strongylocentrotus purpuratus*, and the lancelet (amphioxus) *Branchiostoma lanceolatum*. Whole transcriptome, pairwise comparisons between *P. flava* and sea urchin (Tu, Cameron, and Davidson 2014) revealed high similarities at the gastrula stages (Fig. 4A); noted that the hemichordate early gastrula transcriptome bears high resemblance to that of the sea urchin mesenchyme blastula, which represents the initial gastrulation stage when the primary mesenchyme cells ingress into the blastocoel of sea urchin embryos (Hardin and Cheng 1986). We also observed the highest similarity that occurred at larval stages of both species, supporting the hypothesis that hemichordate and echinoderm larvae are homologous and the dipleura-type larva is an ancestral trait of ambulacraria (Tagawa 2016). These results point to a similar direction as Fig. 2B, supporting the idea that the last common ancestor of ambulacraria (LCA) possessed larval development. The pairwise comparison between *P. flava* and amphioxus (Marlétaz et al. 2018) also showed resemblance of the gastrula transcriptomes (Fig. 4B). Intriguingly, this similarity extends to the amphioxus neurula stages. The observation that amphioxus neurula stages are transcriptomically similar to hemichordate gastrulae likely reflects the exact timings of comparable developmental events; for example, separation of mesoderm and endoderm occurs at the mid-gastrula in *P. flava* and during neurulation in amphioxus (Lin et al. 2016; Carvalho et al. 2021). As the transcriptomes of amphioxus neurula stages also show the highest similarity when compared to those of several vertebrates (Marlétaz et al. 2018), this period (gastrula stages of ambulacrarians and gastrula+neurula stages of chordates) may represent the putative phylotypic stage of deuterostomes. An additional unexpected observation is the transcriptomic similarity between hemichordate and amphioxus larval stages, when both species start to feed. This result suggests that despite the morphological disparity in their larval forms, similar transcriptional programs are utilised by hemichordate and amphioxus larvae.

Given that genes do not work alone, and that our results shown above point to conserved gene expression between different groups, we further tested whether this conservation extends beyond individual genes to the GRN level. We examined interactions among *P. flava* genes whose orthologs constitute well described GRNs in other deuterostome species (Supplementary Data 10). Using the cis-regulatory networks of *P. flava*, especifically the behavior of genes of central position in these

networks and their conservation across deuterostomes, we were able to retrieve cis-regulatory interactions between core genes of two important GRN kernels, the endomesoderm kernel of echinoderms and the GRN of axial mesoderm in chordates. We observed interactions for the core regulatory genes in both networks (Fig. 4C,D), hinting at stable subcircuits that underpin developmental programmes (Davidson 2011, 2010). This result suggests that the conservation of the endomesoderm kernel could be pushed further to the common ancestor of Ambulacraria, reinforcing the notion that part of this kernel already existed in the bilaterian ancestor (Boyle, Yamaguchi, and Seaver 2014). The identification of the cis-regulatory interactions for specifying chordate axial mesoderm in *P. flava* suggests that the genetic circuit predates the emergence of novel structures in chordates, such as the axial mesoderm-derived notochord.

Discussion

In this work, we shed light on the origin of chordates by comparisons with their closest living relatives, echinoderms and hemichordates, at the transcriptomic and cis-regulatory level. This study is the first comprehensive analysis of gene expression and cis-regulation throughout development in *P. flava*, an indirect-developing hemichordate species, ecompassing early and late development. The cis-regulatory landscape of *P. flava* shares essential aspects with cephalochordates, which suggests that many of the cis-regulatory mechanisms present in the stem lineage that led to vertebrates were already present in the last common ancestor of deuterostomes. The two phases of the global transcriptomes in this species hint at a separate developmental programme of controlling gene expression in this species. This is further supported by differential expression of specific TF classes and their binding motifs detected in regions of open chromatin nearby stage-specific genes, together constitutingGRNs at play during development. This differential prevalence of TF class expression is also observed throughout deuterostomes and there is also evidence that this is the case in protostomes (J. Wang et al. 2020; Adryan and Teichmann 2010), hinting at deep homologies in TF usage throughout metazoan development. More data on *P. flava* and other deuterostomes, preferably at the tissue or cell type-level and spanning further into metamorphic development, are required to explore the significance of these observations in order to better identify and determine the conservation degree of

modules of gene expression in hemichordates, and to compare adult tissues with homologous structures in echinoderms and cephalochordates (Marlétaz et al. 2018). These could potentially hint at the degree of gene module usage and the conservation of homologous structures and mechanisms such as the nerve cord and neurulation.

The identification of gastrulation as the stage with more ancient genes expressed, and late larval stages having relatively younger genes, suggests the potential of a pan-bilateria mechanism in the control of gastrulation process. This is further reinforced by the detection of stable, co-regulated kernels of GRNs important in other deuterostome lineages. Of note is to mention that, since hemichordates lack a notochord homologous structure (Eric Röttinger and Lowe 2012; Lowe 2021), new questions arise as to which structure deploys the axial mesoderm GRN in hemichordates. Potential candidates include the endodermal region and protocoel, since some of the orthologous genes within the network are known to express in these embryonic territories. More validation is needed using *in situ* hybridisation to corroborate the co-expression or co-localization of genes belonging to the same GRNs. Likewise, the full scale of the GRNs has not been elucidated because our methods are based on cis-regulatory interactions and not protein-protein interactions (such as ligands and receptors), a known limitation of this analysis.

In summary, this work establishes the first steps and opens new hyptoheses of deuterostome evolution by bringing a global transcriptomic characterization of hemichordate embryo/larval development, and by the generation of the first catalogue of cis-regulatory information in this phylum.

Materials and Methods

Sample collection and animal husbandry

Ptychodera flava adults were collected from Penghu Islands, Taiwan. Methods for spawning induction and embryo culture were described previously (Lin et al. 2016). For RNA-seq, total RNA of each developmental stage was isolated from one to three biological replicates using the RNeasy Micro Kit (Qiagen). Library preparation for stranded polyA-based mRNA was then performed using the TruSeq Stranded mRNA Library Prep Kit (Illumina). The multiplexed cDNA libraries were subjected to sequencing with 151bp paired-end read length on a HiSeq 2500 sequencer (Illumina). The average sequencing depth was 24,861,133 paired reads per sample. For ATAC-seq, embryos at different stages were collected to reach 75,000 cells per sample (~700 early blastulae, ~150 early gastrulae, ~100 mid gastrulae and ~50 late gastrulae). The collected embryos were treated with 0.1% N-acetylcysteine for 10 sec to remove the jelly coat, followed by incubating with 8.5 mM of DTT in seawater containing 5.7 mM of NaOH. The cells were then lysed for the transposition reaction and PCR amplification, and the purified libraries were sequenced as described (Marlétaz et al. 2018).

Time-series analysis

RNA-seq analysis

To quantify gene expression, reads were pseudo aligned against the set of longest transcripts per isoform per gene using Kallisto (Bray et al., n.d.). RNA-seq counts were corrected for batch effect and quantile normalised using edgeR (Robinson et al., n.d.; McCarthy, Chen, and Smyth 2012; Chen, Lun, and Smyth 2016) and RUVSeq (Risso et al. 2014) taking the 10000 bottom-ranked genes in edgeR as in-silico empirical, invariable genes. Pairwise Spearman correlation of replicates was done and visualised in R (Ripley 2001) using base and ComplexHeatmap (Gu, Eils, and Schlesner 2016) packages. Differential Gene Expression Analysis (DGEA) was performed using DESeq2 (Love, Huber, and Anders 2014) using likelihood-ratio-test and RUV-correction as reduced formula. Genes with an adjusted p-value below 0.05 (negative binomial LRT) were taken as differentially expressed. For every replicate, variance-stabilised counts were averaged to generate a time-series dataset of every differentially expressed gene throughout development. For every gene, expression

data was scaled on a z-score (standard deviation over mean). Stage-specific clusters were then generated using hierarchical clustering on the scaled dataset. Data visualisation was done using R base (https://www.r-project.org/), ggplot2 (Wickham 2011), RColorBrewer (Neuwirth 2022), colorspace (Zeileis et al. 2019), pheatmap (RRID:SCR_016418), and ComplexHeatmap (Gu, Eils, and Schlesner 2016) packages under R 4.0.3.

ATAC-seq analysis

ATAC-seq data analyses were performed as described in (Gil-Gálvez et al. 2022), using standard pipelines (Marlétaz et al. 2018; Buenrostro et al. 2013). Briefly, reads were aligned with Bowtie2 using a de novo sequenced, HiC-guided assembly of the genome of *Ptychodera flava* (Simakov et al. 2015). We generated a dataset of peaks using Macs2 (Zhang et al. 2008) for peak calling and Bedtools (Quinlan and Hall 2010) to quantify the accessibility of each chromatin region. Differential analysis was performed using DESeq2 v1.18.0 in R 3.4.3 (Love, Huber, and Anders 2014) using likelihood-ratio test (LRT). An adjusted p-value = 0.05 was set as a cutoff for statistical significance of the differential accessibility of the chromatin in ATAC-seq peaks. Motif enrichment was calculated using the programme FindMotifsGenome.pl from the Homer tool suite (Heinz et al. 2010). k-means clustering of the ATAC-seq signal was performed using R (Ripley 2001) and visualised using ComplexHeatmap (Gu, Eils, and Schlesner 2016). Open Chromatin Regions were assigned to the closest transcription start site (TSS) using bedtools closestbed and standard parameters. Categorisation of OCRs into distal, proximal, intronic etc. was done using the Homer tool suite (Heinz et al. 2010).

<u>Gene annotation</u>

Gene Ontology (GO) enrichment analysis

For every cluster, GO enrichment analysis was performed using the topGO R package (Alexa, Rahnenführer, and Lengauer 2006; “topGO” n.d.) using the whole gene set of *Ptychodera* as gene universe and the *elim* method (Alexa, Rahnenführer, and Lengauer 2006). GO terms were generated using a combination of BLAST2GO and eggNOG annotation software, using RefSeq non-redundant proteins (NR) Database (Pruitt, Tatusova, and Maglott 2007) and eggNOG metazoa (Huerta-Cepas et al. 2019) as reference databases respectively.

Trans-developmental and housekeeping classification

A gene was classified as trans-developmental if annotated with any of the following GO terms: GO:0044430, GO:0006412, GO:0003735, GO:0005840, GO:0005842, GO:0030529, GO:0006414, GO:0009119, GO:0009062, GO:0044262, GO:0006725; but none of the GO terms for housekeeping genes. A gene was classified as housekeeping if annotated with any of the following GO terms: GO:0009790, GO:0030154, GO:0043565, GO:0007267, GO:0008380, GO:0003700, GO:0140297, GO:0010467, GO:0010468, GO:0016070; but none of the GO terms for trans-developmental genes.

Transcription Factor analysis

We generated groups of orthology using *Ptychodera*, Purple Sea Urchin, Japanese Feather Star, amphioxus, and well-annotated vertebrate species (human, zebrafish) using OrthoFinder (Emms and Kelly 2015, 2019) (Supplementary Data 4). Using human and zebrafish queries for TFDB Metazoa (Hu et al. 2019), we transferred TF class annotation to their respective orthogroup and thus to every gene of the other species. For a given TF class X in a given species (*Ptychodera*, Amphioxus, Sea Urchin) at a given developmental stage Y, we computed i) the sum of genes of TF class X expressed (> 1 TPM) at stage Y; ii) the sum of TPMs (using kallisto, see above) for every expressed gene of class X at stage Y; and iii) the rate of ii divided by i. To measure correlation between classes and stage-specific gene expression, we first computed the average expression profile (z-scored) of every stage-specific cluster. Then, for a given TF class X, we computed the number of TFs of class X with a similar expression profile (z-scored) as that of the stage specific cluster (with a Spearman correlation above 0.6), divided by the number of TFs in class X. We did this for every TF class with every stage-specific cluster. For motif discovery analysis, we first extracted the open chromatin regions (differential or not) associated with genes of each stage-specific cluster of gene expression. Then, we performed motif enrichment analysis using the programme FindMotifsGenome.pl from the Homer tool suite (Heinz et al. 2010).

Cis-regulatory networks

We computed two networks of cis-regulation (early blastula and late gastrula) using ANANSE (Xu et al. 2021). We initially generated a custom database of TF-domain association using GimmeMotifs (Bruse and van Heeringen 2018) motif2factors, tailored for the genome annotation of *Ptychodera*. For each developmental stage, we first ran ananse binding to create a genome-wide profile of TF binding across regions of open chromatin using the .bam alignment of the ATAC-seq, and then ran ananse network to create a network of transcription factors (TFs) and target genes (TGs) using the binding profile and the RNA-seq expression data. The resulting network was imported to R and analysed using iGraph. We subsetted the network to retrieve edges of the top decile (ananse score value, 'prob', above quantile 0.9), and subsequently computed network size, main connected component size, number of connections per TG, and per-TFclass centrality using iGraph for R. Germ layer genes were retrieved from in situ hybridisation data on hemichordates in the scientific literature, and assigned to the newest annotation using BLAST best reciprocal hits. Genes from the notochord / paraxial mesoderm GRNs were retrieved from the scientific literature and used as query on a BLAST best-reciprocal hit against the newest annotation of *Ptychodera*.

In situ hybridisations

In situ hybridisations were performed using embryos of *Ptychodera flava* at the late gastrula stage, as previously described in (Ikuta et al. 2013; Luo and Su 2012; Lu, Luo, and Yu 2012).

Supplementary Data

All Supplementary Data information is publicly available at:
https://figshare.com/projects/Ptychodera_supplementary_data/141296.

Acknowledgements

We thank Yi-Cheng Chang, Yi-Chih Chen, Tzu-Pei Fan, Ching-Yi Lin, Chang-Tai Tsai and Cindy Chou for collecting *P. flava* samples for RNA-seq and ATAC-seq. We thank the NGS High Throughput Genomics Core, Biodiversity Research Center, Academia Sinica. This study was supported by the Ministry of Science and Technology, Taiwan (Grant MOST-110-2326-B-001-006 to Y.H.S. and Grant MOST-110-2621-B-001-001-MY3 to J.K.Y.) and Academia Sinica (AS-GC-111-L01 to Y.H.S. and J.K.Y.). J.J.T. and J.L.G.S. were supported by the European Research Council (ERC, grant agreement No 740041) and the Spanish Ministerio de Economía y Competitividad (grant PID2019-103921GB-I00). We thank Alejandro Gil Galvez and Panos Firbas for advice with the analyses, and Maarten Van der Sande, Siebren Frolich, and Simon Van Heeringen, for their advice with ANANSE. A.P.P. especially thanks J.L.G.S., J.J.T. and all members of the Skarmeta/Tena lab, plus Y.H.S. and J.K.Y., for their trust and support during the difficult times of this project.

Work by Authors

A.P.P. performed the RNA-seq, ATAC-seq, and Network analyses, and wrote the first version of the manuscript. Che Y.L. generated the genome annotation. Ching Y.L. and Y.C.C. performed de *In situ* hybridisations. J.L.G.S., Y.H.S. and J.K.Y. conceptualised the project. J.J.T., Y.H.S. and J.K.Y. supervised and provided feedback on the analysis and the writing. All authors edited the manuscript.

References

Adryan, Boris, and Sarah A. Teichmann. 2010. “The Developmental Expression Dynamics of Drosophila Melanogaster Transcription Factors.” *Genome Biology* 11 (4): R40.

Alexa, Adrian, Jörg Rahnenführer, and Thomas Lengauer. 2006. “Improved Scoring of Functional Groups from Gene Expression Data by Decorrelating GO Graph Structure.” *Bioinformatics* 22 (13): 1600–1607.

Altenhoff, Adrian M., Clément-Marie Train, Kimberly J. Gilbert, Ishita Mediratta, Tarcisio Mendes de Farias, David Moi, Yannis Nevers, et al. 2021. “OMA Orthology in 2021: Website Overhaul, Conserved Isoforms, Ancestral Gene Order and More.” *Nucleic Acids Research* 49 (D1): D373–79.

Babaie, Yasmin, Ralf Herwig, Boris Greber, Thore C. Brink, Wasco Wruck, Detlef Groth, Hans Lehrach, Tom Burdon, and James Adjaye. 2007. “Analysis of Oct4-Dependent Transcriptional Networks Regulating Self-Renewal and Pluripotency in Human Embryonic Stem Cells.” *Stem Cells* 25 (2): 500–510.

Bogdanović, Ozren, Ana Fernandez-Miñán, Juan J. Tena, Elisa de la Calle-Mustienes, Carmen Hidalgo, Ila van Kruysbergen, Simon J. van Heeringen, Gert Jan C. Veenstra, and José Luis Gómez-Skarmeta. 2012. “Dynamics of Enhancer Chromatin Signatures Mark the Transition from Pluripotency to Cell Specification during Embryogenesis.” *Genome Research* 22 (10): 2043–53.

Borgatti, Stephen P. 2005. “Centrality and Network Flow.” *Social Networks* 27 (1): 55–71.

Boyle, Michael J., Emi Yamaguchi, and Elaine C. Seaver. 2014. “Molecular Conservation of Metazoan Gut Formation: Evidence from Expression of Endomesoderm Genes in Capitella Teleta (Annelida).” *EvoDevo* 5 (October): 39.

Bray, Pimentel, Melsted, and Pachter. n.d. “Near-Optimal RNA-Seq Quantification with Kallisto.” *Nature Biotechnology*.

Bruse, Niklas, and Simon J. van Heeringen. 2018. “GimmeMotifs: An Analysis Framework for Transcription Factor Motif Analysis.” *bioRxiv*. https://doi.org/10.1101/474403.

Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. “Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position.” *Nature Methods* 10 (12): 1213–18.

Carvalho, João E., François Lahaye, Luok Wen Yong, Jenifer C. Croce, Hector Escrivá, Jr-Kai Yu, and Michael Schubert. 2021. “An Updated Staging System for Cephalochordate Development: One Table Suits Them All.” *bioRxiv*. https://doi.org/10.1101/2020.05.26.112193.

Chen, Yunshun, Aaron T. L. Lun, and Gordon K. Smyth. 2016. “From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the edgeR Quasi-Likelihood Pipeline.” *F1000Research*. https://doi.org/10.12688/f1000research.8987.2.

Csurös, Miklós. 2010. “Count: Evolutionary Analysis of Phylogenetic Profiles with Parsimony and Likelihood.” *Bioinformatics* 26 (15): 1910–12.

Davidson, Eric H. 2010. “Emerging Properties of Animal Gene Regulatory Networks.” *Nature* 468 (7326): 911–20.

———. 2011. “Evolutionary Bioscience as Regulatory Systems Biology.” *Developmental Biology* 357 (1): 35–40.

Domazet-Lošo, Tomislav, and Diethard Tautz. 2010. “A Phylogenetically Based Transcriptome Age Index Mirrors Ontogenetic Divergence Patterns.” *Nature*. https://doi.org/10.1038/nature09632.

Emms, David M., and Steven Kelly. 2015. “OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy.” *Genome Biology* 16 (August): 157.

———. 2019. “OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics.” *Genome Biology* 20 (1): 238.

Fan, Tzu-Pei, Hsiu-Chi Ting, Jr-Kai Yu, and Yi-Hsien Su. 2018. “Reiterative Use of FGF Signaling in Mesoderm Development during Embryogenesis and Metamorphosis in the Hemichordate Ptychodera Flava.” *BMC Evolutionary Biology* 18 (1): 120.

Gil-Gálvez, Alejandro, Sandra Jiménez-Gancedo, Alberto Pérez-Posada, Martin Franke, Rafael D. Acemel, Che-Yi Lin, Cindy Chou, et al. 2022. “Gain of Gene Regulatory Network Interconnectivity at the Origin of Vertebrates.” *Proceedings of the National Academy of Sciences of the United States of America* 119 (11): e2114802119.

Green, Stephen A., Rachael P. Norris, Mark Terasaki, and Christopher J. Lowe. 2013. “FGF Signaling Induces Mesoderm in the Hemichordate Saccoglossus Kowalevskii.” *Development* 140 (5): 1024–33.

Gu, Zuguang, Roland Eils, and Matthias Schlesner. 2016. “Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data.” *Bioinformatics* 32 (18): 2847–49.

Harada, Yoshito, Noko Okai, Shunsuke Taguchi, Eiichi Shoguchi, Kunifumi Tagawa, Tom Humphreys, and Nori Satoh. 2001. “Embryonic Expression of a Hemichordate Distal-Less Gene.” *Zoological Science* 18 (1): 57–61.

Hardin, Jeffrey D., and Louis Y. Cheng. 1986. “The Mechanisms and Mechanics of Archenteron Elongation during Sea Urchin Gastrulation.” *Developmental Biology* 115 (2): 490–501.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010. “Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities.” *Molecular Cell* 38 (4): 576–89.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. “eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” *Nucleic Acids Research* 47 (D1): D309–14.

Hu, Hui, Ya-Ru Miao, Long-Hao Jia, Qing-Yang Yu, Qiong Zhang, and An-Yuan Guo. 2019. “AnimalTFDB 3.0: A Comprehensive Resource for Annotation and Prediction of Animal Transcription Factors.” *Nucleic Acids Research* 47 (D1): D33–38.

Ikuta, Tetsuro, Yi-Chih Chen, Rossella Annunziata, Hsiu-Chi Ting, Che-Huang Tung, Ryo Koyanagi, Kunifumi Tagawa, et al. 2013. “Identification of an Intact ParaHox Cluster with Temporal Colinearity but Altered Spatial Colinearity in the Hemichordate Ptychodera Flava.” *BMC Evolutionary Biology* 13 (June): 129.

Irimia, Manuel, Juan J. Tena, Maria S. Alexis, Ana Fernandez-Miñan, Ignacio Maeso, Ozren Bogdanovic, Elisa de la Calle-Mustienes, Scott W. Roy, José L. Gómez-Skarmeta, and Hunter B. Fraser. 2012. “Extensive Conservation of Ancient Microsynteny across Metazoans due to Cis-Regulatory Constraints.” *Genome Research* 22 (12): 2356–67.

Kaul-Strehlow, Sabrina, and Thomas Stach. 2013. “A Detailed Description of the Development of the Hemichordate Saccoglossus Kowalevskii Using SEM, TEM, Histology and 3D-Reconstructions.” *Frontiers in Zoology* 10 (1): 53.

Kaul-Strehlow, Sabrina, Makoto Urata, Daniela Praher, and Andreas Wanninger. 2017. “Neuronal Patterning of the Tubular Collar Cord Is Highly Conserved among Enteropneusts but Dissimilar to the Chordate Neural Tube.” *Scientific Reports* 7 (1): 7003.

Lin, Ching-Yi, Che-Huang Tung, Jr-Kai Yu, and Yi-Hsien Su. 2016. “Reproductive Periodicity, Spawning Induction, and Larval Metamorphosis of the Hemichordate Acorn Worm Ptychodera Flava.” *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* 326 (1): 47–60.

Li, Yongxin, Akihito Omori, Rachel L. Flores, Sheri Satterfield, Christine Nguyen, Tatsuya Ota, Toko Tsurugaya, et al. 2020. “Genomic Insights of Body Plan Transitions from Bilateral to Pentameral Symmetry in Echinoderms.” *Communications Biology* 3 (1): 371.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12):

550.
Lowe, Christopher J. 2021. “Molecular Insights into Deuterostome Evolution from Hemichordate Developmental Biology.” *Current Topics in Developmental Biology* 141 (January): 75–117.
Lowe, Christopher J., D. Nathaniel Clarke, Daniel M. Medeiros, Daniel S. Rokhsar, and John Gerhart. 2015. “The Deuterostome Context of Chordate Origins.” *Nature* 520 (7548): 456–65.
Lowe, Christopher J., Mark Terasaki, Michael Wu, Robert M. Freeman Jr, Linda Runft, Kristen Kwan, Saori Haigo, et al. 2006. “Dorsoventral Patterning in Hemichordates: Insights into Early Chordate Evolution.” *PLoS Biology* 4 (9): e291.
Lowe, Christopher J., Mike Wu, Adrian Salic, Louise Evans, Eric Lander, Nicole Stange-Thomann, Christian E. Gruber, John Gerhart, and Marc Kirschner. 2003. “Anteroposterior Patterning in Hemichordates and the Origins of the Chordate Nervous System.” *Cell* 113 (7): 853–65.
Luo, Yi-Jyun, and Yi-Hsien Su. 2012. “Opposing Nodal and BMP Signals Regulate Left-Right Asymmetry in the Sea Urchin Larva.” *PLoS Biology* 10 (10): e1001402.
Lu, Tsai-Ming, Yi-Jyun Luo, and Jr-Kai Yu. 2012. “BMP and Delta/Notch Signaling Control the Development of Amphioxus Epidermal Sensory Neurons: Insights into the Evolution of the Peripheral Sensory System.” *Development* 139 (11): 2020–30.
Mangelsdorf, D. J., U. Borgmeyer, R. A. Heyman, J. Y. Zhou, E. S. Ong, A. E. Oro, A. Kakizuka, and R. M. Evans. 1992. “Characterization of Three RXR Genes That Mediate the Action of 9-Cis Retinoic Acid.” *Genes & Development* 6 (3): 329–44.
Marlétaz, Ferdinand, Panos N. Firbas, Ignacio Maeso, Juan J. Tena, Ozren Bogdanovic, Malcolm Perry, Christopher D. R. Wyatt, et al. 2018. “Amphioxus Functional Genomics and the Origins of Vertebrate Gene Regulation.” *Nature* 564 (7734): 64–70.
McCarthy, Davis J., Yunshun Chen, and Gordon K. Smyth. 2012. “Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation.” *Nucleic Acids Research* 40 (10): 4288–97.
Miyamoto, Norio, and Hiroshi Wada. 2013. “Hemichordate Neurulation and the Origin of the Neural Tube.” *Nature Communications* 4: 2713.
Nakajima, Yoko, Tom Humphreys, Hiroyuki Kaneko, and Kunifumi Tagawa. 2004. “Development and Neural Organization of the Tornaria Larva of the Hawaiian Hemichordate, Ptychodera Flava.” *Zoological Science* 21 (1): 69–78.
Neuwirth, Erich. 2022. “ColorBrewer Palettes [R Package RColorBrewer Version 1.1-3],” April. https://CRAN.R-project.org/package=RColorBrewer.
Nichols, J., B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, and A. Smith. 1998. “Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4.” *Cell* 95 (3): 379–91.
Niwa, H., J. Miyazaki, and A. G. Smith. 2000. “Quantitative Expression of Oct-3/4 Defines Differentiation, Dedifferentiation or Self-Renewal of ES Cells.” *Nature Genetics* 24 (4): 372–76.
Osorno, Rodrigo, and Ian Chambers. 2011. “Transcription Factor Heterogeneity and Epiblast Pluripotency.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 366 (1575): 2230–37.
Pani, Ariel M., Erin E. Mullarkey, Jochanan Aronowicz, Stavroula Assimacopoulos, Elizabeth A. Grove, and Christopher J. Lowe. 2012. “Ancient Deuterostome Origins of Vertebrate Brain Signalling Centres.” *Nature* 483 (7389): 289–94.
Pruitt, Kim D., Tatiana Tatusova, and Donna R. Maglott. 2007. “NCBI Reference Sequences (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins.” *Nucleic Acids Research* 35 (Database issue): D61–65.
Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42.
Ripley, B. D. 2001. “The R Project in Statistical Computing.” *MSOR Connections. The Newsletter of the LTSN Maths*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.449.6899&rep=rep1&type=p

df.
Risso, Davide, John Ngai, Terence P. Speed, and Sandrine Dudoit. 2014. “Normalization of RNA-Seq Data Using Factor Analysis of Control Genes or Samples.” *Nature Biotechnology* 32 (9): 896–902.
Robinson, McCarthy, Chen, and Smyth. n.d. “edgeR: Differential Expression Analysis of Digital Gene Expression Data.” *User’s Guide*. http://129.217.206.11/packages/2.9/bioc/vignettes/edgeR/inst/doc/edgeR.pdf.
Röttinger, E., and M. Q. Martindale. 2011. “Ventralization of an Indirect Developing Hemichordate by $NiCl_2$ Suggests a Conserved Mechanism of Dorso-Ventral (D/V) Patterning in Ambulacraria (hemichordates and Echinoderms).” *Developmental Biology* 354 (1): 173–90.
Röttinger, Eric, Timothy Q. DuBuc, Aldine R. Amiel, and Mark Q. Martindale. 2015. “Nodal Signaling Is Required for Mesodermal and Ventral but Not for Dorsal Fates in the Indirect Developing Hemichordate, Ptychodera Flava.” *Biology Open* 4 (7): 830–42.
Röttinger, Eric, and Christopher J. Lowe. 2012. “Evolutionary Crossroads in Developmental Biology: Hemichordates.” *Development* 139 (14): 2463–75.
Sebé-Pedrós, Arnau, Cecilia Ballaré, Helena Parra-Acero, Cristina Chiva, Juan J. Tena, Eduard Sabidó, José Luis Gómez-Skarmeta, Luciano Di Croce, and Iñaki Ruiz-Trillo. 2016. “The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity.” *Cell* 165 (5): 1224–37.
Sebé-Pedrós, Arnau, Baptiste Saudemont, Elad Chomsky, Flora Plessier, Marie-Pierre Mailhé, Justine Renno, Yann Loe-Mie, et al. 2018. “Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq.” *Cell* 173 (6): 1520–34.e20.
Shao, Changwei, Baolong Bao, Zhiyuan Xie, Xinye Chen, Bo Li, Xiaodong Jia, Qiulin Yao, et al. 2017. “The Genome and Transcriptome of Japanese Flounder Provide Insights into Flatfish Asymmetry.” *Nature Genetics* 49 (1): 119–24.
Simakov, Oleg, Takeshi Kawashima, Ferdinand Marlétaz, Jerry Jenkins, Ryo Koyanagi, Therese Mitros, Kanako Hisata, et al. 2015. “Hemichordate Genomes and Deuterostome Origins.” *Nature* 527 (7579): 459–65.
Su, Yi-Hsien, Yi-Chih Chen, Hsiu-Chi Ting, Tzu-Pei Fan, Ching-Yi Lin, Kuang-Tse Wang, and Jr-Kai Yu. 2019. “BMP Controls Dorsoventral and Neural Patterning in Indirect-Developing Hemichordates Providing Insight into a Possible Origin of Chordates.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (26): 12925–32.
Swalla, Billie J. 2007. “New Insights into Vertebrate Origins.” *Principles of Developmental Genetics*, 114–28.
Tagawa, Kuni. 2016. “Hemichordate Models.” *Current Opinion in Genetics & Development* 39 (August): 71–78.
Taguchi, S., K. Tagawa, T. Humphreys, A. Nishino, N. Satoh, and Y. Harada. 2000. “Characterization of a Hemichordate Fork head/HNF-3 Gene Expression.” *Development Genes and Evolution* 210 (1): 11–17.
“topGO.” n.d. Bioconductor. Accessed May 30, 2022. https://bioconductor.org/packages/release/bioc/html/topGO.html.
Tu, Qiang, R. Andrew Cameron, and Eric H. Davidson. 2014. “Quantitative Developmental Transcriptomes of the Sea Urchin Strongylocentrotus Purpuratus.” *Developmental Biology* 385 (2): 160–67.
Urata, Makoto, and Masaaki Yamaguchi. 2004. “The Development of the Enteropneust Hemichordate Balanoglossus Misakiensis KUWANO.” *Zoological Science* 21 (5): 533–40.
Wang, Jing, Lingling Zhang, Shanshan Lian, Zhenkui Qin, Xuan Zhu, Xiaoting Dai, Zekun Huang, et al. 2020. “Evolutionary Transcriptomics of Metazoan Biphasic Life Cycle Supports a Single Intercalation Origin of Metazoan Larvae.” *Nature Ecology & Evolution* 4 (5): 725–36.
Wang, Xuedong, Hiroki Matsuda, and Yun-Bo Shi. 2008. “Developmental Regulation and

Function of Thyroid Hormone Receptors and 9-Cis Retinoic Acid Receptors during Xenopus Tropicalis Metamorphosis." *Endocrinology* 149 (11): 5610–18.

Wickham, Hadley. 2011. "Ggplot2." *Wiley Interdisciplinary Reviews. Computational Statistics* 3 (2): 180–85.

Xu, Quan, Georgios Georgiou, Siebren Frölich, Maarten van der Sande, Gert Jan C. Veenstra, Huiqing Zhou, and Simon J. van Heeringen. 2021. "ANANSE: An Enhancer Network-Based Computational Approach for Predicting Key Transcription Factors in Cell Fate Determination." *Nucleic Acids Research* 49 (14): 7966–85.

Zeileis, Achim, Jason C. Fisher, Kurt Hornik, Ross Ihaka, Claire D. McWhite, Paul Murrell, Reto Stauffer, and Claus O. Wilke. 2019. "Colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes." *arXiv [stat.CO]*. arXiv. http://arxiv.org/abs/1903.06490.

Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology* 9 (9): R137.