Epigenomics of embryogenesis in turbot

Oscar Aramburu, Belen G Pardo, Ada Jimenez-Gonzalez, et al.

Genome Res. published online November 4, 2025
Access the most recent version at doi:10.1101/gr.280355.124

Epigenomics of embryogenesis in turbot

Oscar Aramburu[1], Belén G. Pardo[1], Ada Jimenez-Gonzalez[2], Andrés Blanco-Hortas[1], Daniel J. Macqueen[3], Carmen Bouza[1], Paulino Martínez[1]

[1]University of Santiago de Compostela, Campus Terra, Lugo, Spain, 27002

[2]Sheffield Institute for Translational Neuroscience (SITraN), Department of Neuroscience, University of Sheffield, UK, S10 2TN

[3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, UK, EH25 9RG

ABSTRACT

Embryogenesis is the foundational step of ontogeny, where a complex organism emerges from a single totipotent cell. This process is orchestrated by changes in transcriptional regulation, influenced by chromatin accessibility and epigenetic modifications, enabling transcription factor accessibility. Epigenomic regulation of embryogenesis has been studied in model fish, but little attention has been paid to farmed fish - where relevant traits to aquaculture rely on early developmental processes. This study reports a regulatory atlas of turbot (*Scophthalmus maximus*) embryogenesis. 14,560 active genes were identified in the embryonic transcriptome with > 90% showing differential expression across consecutive stages. By integrating multi-histone ChIP-seq with ATAC-seq, we built a genome-wide chromatin state model, defining promoter and enhancer activity across stages. Diverse transcription factor binding motifs were detected within regulatory elements showing differential accessibility at distinct developmental stages. Strong shifts in chromatin accessibility across stages, notably during the transition from shield to early segmentation, suggest profound chromatin reorganization underpinning somitogenesis and early organogenesis. Regardless, most changes in chromatin accessibility did not affect promoters of differentially expressed genes, suggesting that their accessibility precedes gene transcription changes. Comparative analyses with zebrafish revealed a global transcriptomic correlation of single-copy orthologs at matched stages. While conserved expression dynamics were revealed for many orthologous Hox genes, notable cross-species differences were identified from pre-ZGA leading up to hatching. This multi-omics investigation provides a novel atlas of non-coding regulatory elements controlling turbot development, with key applications for flatfish biology and sustainable aquaculture.

INTRODUCTION

Embryogenesis is a fundamental process in the life cycle of multicellular organisms. Insights into embryonic development can shed light on biological questions such as its genetic control, inter-species variation and evolution. Research on embryogenesis is not only limited to basic biological knowledge, but also crucial for understanding genetic and epigenetic mechanisms behind the establishment of cell lineages and tissue formation (Petratou et al., 2024). Such insights can help optimize breeding strategies and improve the management of commercially valuable species farmed in controlled environments.

Embryogenesis involves a dynamic succession of key processes such as cell differentiation, migration, patterning, organogenesis and apoptosis (Zernicka-Goetz, 2002; Voss and Strasser, 2020). These processes are particularly sensitive to environmental stressors such as pH, salinity, pathogens and temperature (Miranda et al., 2023), currently one of the most acute issues in aquaculture due to global warming (Skjærven et al., 2024). By understanding gene regulatory interactions underlying these processes, genetic variants linked to favourable traits can be identified and targeted by breeders.

Annotation of the genome of aquaculture species represents an essential resource for understanding the genetic and functional genomic mechanisms underlying complex traits such as growth, feed efficiency, muscle quality, reproduction, and disease resistance (Clark et al., 2020; Johnston et al., 2024). Efforts in functional annotation of finfish genomes have to date concentrated on coding and non-coding genes, with transcriptomic studies establishing detailed annotations in zebrafish (White et al., 2017; Lawson et al., 2020) and aquaculture species including Atlantic salmon (Bizuayehu et al., 2019; Abdellaoui and Kim, 2024), rainbow trout (Ma et al., 2019; Yang et al., 2023), Nile tilapia (Powell et al., 2021; Kumar Behera et al., 2024) and European sea bass (Papadaki et al., 2022), among others (Blasweiler et al., 2023; Bouza et al., 2024; Villareal et al., 2024).

In contrast, *cis*-regulatory elements (CRE), including promoters, enhancers, silencers, and insulators, remain largely unexplored in aquaculture species (Harvey et al., 2024), despite recent advances in model species such as zebrafish (Jimenez-Gonzalez et al., 2023) and terrestrial livestock (Summers et al., 2020; Pan et al., 2023). These elements are crucial for regulating gene expression, and their activity varies depending on ontogeny, tissue, cell, sex, age, and health status of the organism (van Mierlo et al., 2023). Therefore, annotation of these elements will help address fundamental biological questions on development and tissue differentiation (White et al., 2017; Pan et al., 2023), while also promoting understanding of phenotypic plasticity in response to environmental changes (Bouza et al., 2024; Hu et al., 2024).

Turbot (*Scophthalmus maximus*) is a high value fish species farmed primarily in Asia and Europe, with an annual production exceeding 100,000 tons largely concentrated in China (Gao et al., 2023), followed by Spain (APROMAR 2023). European turbot breeding is in its ninth generation of selection (Riaza, Stolt Sea

Farm SA, pers. comm.). Functional annotation of the turbot transcriptome has been conducted using high-quality genome assemblies (Figueras et al., 2016; Maroso et al., 2018; Martínez et al., 2021; Xu et al., 2022), but non-coding elements have received limited attention, aside from studies exploring non-coding RNAs (Robledo et al., 2017; Cai et al., 2023), and recently on epigenomic changes controlling immune response in response to viral and bacterial mimics (Aramburu et al., 2025). The morphological processes underlying the transformation of a fertilized egg into a free-living flatfish are well-established in turbot (Jones, 1972; Wang et al., 2017). However, few studies have focused on underpinning developmental changes in gene expression (Wu et al., 2024), with only one investigating regulation of chromatin accessibility during flatfish metamorphosis, using turbot as a study system (Guerrero-Peña et al., 2023).

In this study, we aimed to generate the first comprehensive functional annotation of turbot embryogenesis by performing RNA-seq, ATAC-seq and ChIP-seq on samples from multiple developmental stages covering early to late embryogenesis by integrating changes in the transcriptome with chromatin accessibility, and epigenetic states determined using functionally distinct (H3K4me3, H3K27ac and H3K27me3) histone marks. In addition, we performed a comparative transcriptomic analysis with zebrafish (*Danio rerio*) embryonic development (White et al., 2017), focusing on the Hox gene family as master regulator of body plans, to advance our understanding on the regulation of embryonic development within teleost. Together, these data offer a resource for examining the regulatory mechanisms involved in flatfish embryonic development and for exploring broader questions related to gene regulation, developmental transitions, and evolutionary variation in teleost.

RESULTS

114 multiomic libraries were produced across turbot embryogenesis in this study (Figure 1), which resulted in 107 datasets after quality control, including RNA-seq data for 12 developmental stages (36 libraries), along with ATAC-seq (18 libraries) and ChIP-seq data (47 libraries; 17 for H3K4me3, 15 for H3K27ac and 15 for H3K27me3, plus 6 ChIP-seq input libraries) for six key developmental stages. Full information on samples and metadata is shown in Supplemental tables S1 and S2.

Transcriptomic profiling of turbot embryogenesis

RNA-seq was used to compare transcriptomic profiles across different stages of turbot development. On average, 64,550,207 raw reads per library were produced, with 97.59% mapping rate (range: 97.21-98.20%), no 5’ nor 3’ bias, and high-quality standard metrics (Supplemental table S3A, Supplemental figure S1A). Principal component analysis (PCA) separated the 12 developmental stages following a smooth parable from the cleavage stage (64 cells) up to the end of embryogenesis (prehatch).

Hierarchical clustering grouped the samples into two major clusters: pre-zygotic genome activation (ZGA) stages (64-cells, morula and early blastula) grouped with ZGA (late blastula and embryonic shield) stages; and post-ZGA (neural plate, blastopore closure, early, mid and late segmentation, early pharyngula and prehatch) stages. The post ZGA stages were further separated into two sub-clusters split before and after early mid segmentation stages (Figure 2B). For all stages, the three replicates clustered closely in the PCA and heatmap (Figure 2A and B).

Waves of differential gene expression highlight key transitions at ZGA and segmentation stages

14,560 expressed genes (TPM > 5 cut-off; 42.51 % of the 34,252annotated genes in the turbot genome) were identified, constituting the embryonic transcriptome of turbot (Supplemental table S4). A trend of increasing transcriptomic complexity was observed from early (e.g. 8,640 genes at 64-cell stage) to late (e.g. 12,112 genes at prehatch stage) stages (Figure 2C). Correlation networks using BioLayout (Figure 3A; Theocharidis et al., 2009) were applied to visualize the expression changes across development. The correlation network shows a “pulley-arch” structure, where the clusters ordered from left to right of the graph represent genes expressed at early, mid and late development stages, respectively (Supplemental table S5). Groups are connected through shared genes that reached peaks of expression during gastrulation and segmentation stages, as well as by a small set of genes with high expression at early and late stages along the “string” of the bow-shaped graph (Figure 3A; Supplemental table S5). This wave-like expression pattern, with early-expressed genes gradually decreasing expression in parallel to mid- and late- expression genes being activated, was also observed by ordering genes temporarily by stage of maximum expression in 12 clusters represented in a heatmap (Figure 3B). Detailed expression profiles for a subset of relevant genes associated with maternal RNAs, ZGA, writing and reading of epigenetic marks, cellular differentiation and organogenesis are shown in Supplemental figure S2 (see discussion section).

Differential gene expression analysis was performed by contrasting the six main embryonic stages (Figure 1) sequentially, resulting in five comparisons. 14,060 differentially expressed genes (DEGs; p. adj < 0.05) were identified between the stage comparisons, representing 96.6% of the embryonic transcriptome (Figure 3C; Supplemental table S6). Overall, more DEGs were identified at early and mid-development (late blastula to shield transition; early to mid-segmentation transition) compared to later developmental stages (Figure 3C; Supplemental table S6; Supplemental figure S3). For each comparison, similar proportions of DEGs were found between consecutive stages (Figure 3C). Additionally, we compared all pre-ZGA samples (64-cells, morula and early blastula) to ZGA and post-ZGA ones (from late blastula to prehatch), generating a catalogue of 4,498 DEGs specific to pre-ZGA stages. This implies that 30.89% of the transcripts in the turbot developmental transcriptome are maternal in origin (Supplemental

table S6), with the caveat that a small proportion could be the first ZGA genes, particularly for the early blastula stage (Chen et al., 2019).

Core and stage-specific biological functions driving turbot embryogenesis

Biological processes occurring during embryogenesis were characterized by identifying enriched Gene Ontology (GO) terms for the following lists of genes (Supplemental table S7): i) the total expressed transcriptome (Figure 4A), ii) all non-expressed turbot genes during embryogenesis (Supplemental figure S4), iii) the stage-specific transcriptomes (12 stages; Supplemental figure S4), iv) DEGs between the five sequential-stage comparisons (Figure 4B; Supplemental figure S4), v) upregulated DEGs between the three pre-ZGA developmental stages versus the ZGA and post-ZGA stages, and vi) the genes detected in the early, mid and late-development stages, and early-late clusters identified with BioLayout.

Core cellular GO processes reflecting housekeeping functions, including terms related to transcription, translation, ribosome biogenesis and protein localization, were consistently active across all developmental stages and accordingly in the total active transcriptome (Figure 4A; Supplemental table S7; Supplemental figure S4). Additionally, terms associated with pattern specification processes and organogenesis were highly represented in the total transcriptome but following a more stage-specific gene expression pattern. This includes cell cycle regulation, detected across early developmental stages (especially from late blastula to early segmentation); reproductive process at the cleavage / early blastulation stages (64 cell, morula and early blastula); methylation-related terms (RNA, protein and macromolecule methylation) at all developmental stages except from late segmentation onwards; endoderm formation in late blastula and embryonic shield stages; and morphogenesis-related terms in early segmentation (mesenchymal tissues, epithelium and glial cells), late segmentation (actin organization), and early pharyngula (neuron differentiation) (Figure 4A, Supplemental figure S4, Supplemental table S7). Many mature, tissue-specific functions expected to be transcriptionally silent during embryogenesis were indeed only present among the non-expressed genes (19,893 annotated genes), and included enriched GO terms associated with postembryonic functions, such as immune processes, response to stimulus, presynapse organization, regulated exocytosis, feeding behaviour, cell adhesion and cell communication (Supplemental figure S4; Supplemental table S7).

RNA processing and transcription associated terms were enriched across all five pairwise stage comparisons (Figure 4B; Supplemental table S7). More stage-specific processes, including reproductive processes, were found in the late blastula to shield transition. Inspection of these genes revealed a variety of reproductive, oocyte germline genes (*zp3b*, *bmp15*, *figla*) seemingly upregulated in late blastula when compared to shield, likely reflecting maternal mRNA clearance. On the opposite end, different stem cell and pluripotency-related factors (*nanog*, *kdm6a*, *pum1*, *pum3*) were upregulated towards the

shield stage. Moreover, gastrulation, convergent extension, pattern specification and tissue (epithelium) morphogenesis were enriched in the transition from shield to early segmentation. Morphogenesis-related terms (particularly muscle development and nerve development / synapsis) were enriched in all the transitions across segmentation, while head, heart, blood, digestive system and gland development-related terms were prominently enriched in mid to late segmentation. Finally, terms related to locomotion and behavior were enriched in the transitions from both mid to late segmentation and late segmentation to prehatch, the latter also including response to stimulus (Figure 4B; Supplemental figure S4; Supplemental table S7).

GO term enrichment of the transcripts found in the pre-ZGA stages (64-cells, morula and early blastula) unveiled functions associated with reproductive processes such as egg coat formation, sexual reproduction, meiotic nuclear division, gamete generation and sperm-egg recognition, along with more generalist functions associated with cell-cell recognition, cell cycle progress, DNA replication and RNA processing.

Chromatin accessibility and histone mark mapping

Open chromatin regions (ATAC-seq) and those including histone epigenetic marks (ChIP-seq) for H3K4me3 (marking active promoter regions), H3K27ac (active enhancer and promoter regions) and H3K27me3 (Polycomb repressed regions) were identified throughout the genome across six key developmental stages in turbot. After sequencing, 75,600,169 (ATAC-seq), 121,913,111 (H3K4me3), 116,148,829 (H3K27ac) and 98,992,315 (H3K27me3) raw reads were produced on average, with a mapping rate of 94.80% and no obvious deviation from the expected fragment size, insert size, and sequence read length distributions (Supplemental table S3B and S3C, Supplemental figure S1B and S1C; Supplemental figure S5). On average, 50,004 peaks were identified for ATAC-seq data across all stages, while 18,466 (H3K4me3), 11,708 (H3K27ac) and 16,651 (H3K27me3) peaks were called for ChIP-seq data (Figure 5A; Supplemental figure S6A and B). PCA clustered each library type, with some overlap between H3K4me3 and H3K27ac (Figure 5B). ATAC-seq and ChIP-seq libraries were mostly segregated across PC1 (35%), while PC2 (19%) explained most of the variation between histone marks, particularly differentiating the repressor mark H3K27me3 from the nearby activation marks H3K4me3 and H3K27ac (Figures 5B and C). All four assays clearly differentiated the ZGA stages (late blastula and shield) from post-ZGA stages, and ATAC-seq showed the greatest resolution to differentiate each developmental stage (Figure 5C).

Epigenomic profiling of chromatin states reveals regulatory transitions during early development

To gain a comprehensive view of the regulatory landscape during turbot embryogenesis, genome-wide chromatin state predictions were obtained by integrating the ChIP-seq and ATAC-seq datasets using ChromHMM (Figure 6A, Supplemental table S8). A 10-state model was chosen as it maximized the

number of non-redundant biologically relevant states. This model represented four active promoter states proximal to the transcription start site (TSS), comprising two strictly active promoter states (STssA and TssA), one bivalent signal after the TSS (TssBiv), and one flanking active TSS without ATAC signal (TssFlnk). The remaining states consisted of one active enhancer state (EnhA), one weak active regulatory element state lacking ATAC signal (FlnkWk), one bivalent regulatory element state (Biv), one ATAC island state (open chromatin lacking histone marks; ATACi), one Polycomb repressed state (ReprPC) and one low signal state (LowSi). STssA, TssA, TssBiv and TssFlnk exhibited the highest signal around the TSS (±2 kb), corresponding to promoter regions and/or transcriptionally active regions.

The low signal state (LowSi) covered most of the turbot genome (average: 85.98%). ATACi was the most prominent open chromatin state (average: 5.37%), while active promoter (STssA, TssA, TssFlnk) and enhancer (EnhA) states represented on average 2.72% and 1.35% of the genome, respectively. Unspecific weak active and bivalent states (TssBiv, FlnkWk and Biv, respectively) represented on average 2.31% of the genome, bringing the total coverage of accessible, activation-marked regions to 6.37%, while the repression state (ReprPC) represented 3.41% of the genome (Figure 6B).

The global dynamics of each chromatin state across development was estimated by calculating the percentage of 200 bp bins of each chromatin state transitioning to a different state in the next developmental stage (Figure 6C; Supplemental figure S7B; Supplemental table S8A). Globally, genomic regions tended to show stable chromatin states between consecutive developmental stages, 58.47% on average (range: 9.99-96.63%), compared to those changing to a different state, 18.13% (range: 0.00-69.02%). The low signal state was the most static across development (average 94.87%, range 92.84-96.63%), while the other states showed variable dynamics: Transitions from EnhA to LowSi states appear the most frequent (average 44.61%, range: 36.22-57.79%), suggesting widespread silencing or deactivation once cell lineages are committed. This was followed by transitions like TssFlnk to FlnkWk (32.46%), FlnkWk to LowSi (32.46%), FlnkWk to LowSi (30.59%) and STssA to TssA (28.45%). The general pattern suggests that poised (Biv) and weak states (FlnkWk, TssFlnk, ATACi) tend to shift between each other, and towards ReprPC or LowSi states across development. Similarly, active promoter regions (STssA and TssA) appear to dynamically change between each other or towards weak states (FlnkWk, TssFlnk), but never directly to ReprPC or LowSi states (Figure 6C; Supplemental figure S7B; Supplemental table S8).

Each narrow promoter (-1000 to +100 bp of TSS) was classified into one of ten discrete categories according to their dominant chromatin state (Supplemental figures S7B and C and S8; Supplemental table S9A). Overall, genes with STssA dominant states showed the highest average expression across development, followed by genes with the TssA, TssFlnk, and EnhA states dominant states. As expected, genes with ReprPC and LowSi dominant states showed the lowest average expression levels, and dominant bivalent states (TssBiv and Biv) were

the most dynamic: TssBiv-dominated genes had very low expression level during late blastula and shield stages, but increased along segmentation; meanwhile Biv-dominated genes started at low levels during late blastula, increased expression (with high diversity among genes) during the shield stage and further along segmentation.

Additionally, GO enrichment analysis was used to search for enriched functions among the genes linked to each dominant chromatin state category (Supplemental table S9). For STssA, terms related with RNA processing, transcriptional regulation and metabolism were highly enriched, similar, but to a higher extent, to terms enriched in the TssA and TssFlnk-dominated genes. TssBiv, Biv and ATACi-dominated genes showed the highest variety of enriched GO terms, including cellular differentiation, patterning, epi/meso/endodermal development and organogenesis, particularly neurogenesis and heart, liver, pancreas, nephron and retina development. FlnkWk-dominated genes were enriched in terms associated with DNA repair, cell cycle and stress response, and EnhA-dominated genes only in transcriptional regulation. ReprPC-dominated genes were enriched in a variety of organ development terms also present in TssBiv, Biv and ATACi, particularly myogenesis and neurogenesis, but also terms associated with post-hatch, like swim-bladder, nose, kidney and immune system development, the latter also present in LowSi-dominated genes (Supplemental table S9).

Segmentation marks a major shift in chromatin accessibility at the onset of organogenesis

Consensus ATAC-seq peaks overlapping promoter regions (-2000, +600 bp) and intergenic / intron regions (putative enhancers) were intersected with the annotated chromatin states to characterize active cis-regulatory elements (CREs) at each stage (Supplemental figure S7A; Supplemental table S10). On average, 15,313 open chromatin regions, with active regulatory states, were found across the six stages, ranging from 13,127 in late segmentation to 17,814 in early segmentation. Promoter activity showed a steady increase from late blastula (13,817) to prehatch (15,859), while enhancer activity showed a rapid increase from shield (5,307) to early segmentation (7,770), decaying along mid segmentation and late segmentation to reach similar levels to early development in the prehatch stage (5,598; Supplemental figure S7C; Supplemental table S10). Detailed chromatin profiles for relevant genes associated with maternal expression, ZGA, writing and reading of histone marks, cellular differentiation and organogenesis are presented in Supplemental figure S9, showing a generally conserved accessibility of the promoter through the analyzed developmental stages.

Significantly differentially accessible regions (DARs, FDR-adjusted $p < 0.05$) between consecutive developmental stages were identified (Figure 6D, Supplemental figure S10; Supplemental table S11). The aggregated line plots and heatmaps of DAR signals across the six main developmental stages (Supplemental figure S11A) showed that, overall, most DARs across

segmentation were already highly accessible in early segmentation, whereas upregulated DARs in late blastula, shield and prehatch stages were mostly accessible only at their corresponding stage, a behavior observed for both transition-specific and all DARs. In fact, by far the highest number of DARs was detected in the transition from embryonic shield to early segmentation (28,686 DARs), while only 14 were observed in the transition from mid to late segmentation. On average 12.64% DARs were found within promoters and 20.86% in intergenic and intronic regions (Figure 6D; Supplemental table S11).

Limited overlap between DARs and differential gene expression is consistent with pre-priming of promoters

To identify genes that were both differentially expressed and had differential accessibility in their promoters through development, the intersection between DAR promoters and DEGs for each comparison was tested through hypergeometric distribution tests (p. adj < 0.05, Bonferroni correction, Supplemental table S12). Little correspondence was found between DARs in promoters and DEGs, except for the transition from shield to early segmentation. Here, 2,034 DEGs showed upregulated DAR promoters in early segmentation (p-value $6.48x\ 10^{-08}$; 33.29 % of the DEGs and 35.89 % of the promoter-DARs in that transition). We also observed that the transition from late segmentation to prehatch showed 42 DEGs with DAR promoters upregulated in prehatch (p-value $1.86\ x\ 10^{-10}$, 2.01 % of the DEGs and 36.52 % of the promoter-DARs in that transition), suggesting that promoters of DEGs are generally already accessible as part of pre-priming before the onset of transcription. To test the occurrence of more complex dynamics, the *cis*-regulatory logic of consensus ATAC peaks with their closest gene was tested adapting the approach of Li et al. (2020), which considers three regulatory logics: synchronization, repression and regulatory switch (Figure 7, Supplemental table S13).

Starting with synchronization, where the accessibility of assigned regulatory elements and the expression of their target gene is highly correlated, 6,016 high-confidence peak-gene pairs were identified nearby to 2,855 genes (Supplemental table S13). This was the case for *pou5f3* (Figure 7A), a pioneer transcription factor that guides the start of ZGA, and two upstream regulators (ur1 and ur2); a high synchronization between expression and accessibility was detected in each pair (Spearman's correlation above 0.94), where the highest values were detected in late blastula, followed by the shield stage. The next type, repression, where the accessibility of assigned regulatory elements and gene expression are inversely correlated, was detected in 1,824 peak-gene pairs, near to 1,260 genes (Supplemental table S13). An example of this dynamic was observed in *hoxc13a* (Figure 7B), a member of the HOXCa cluster involved in embryonic patterning, and one upstream regulator (ur1), where the reduced accessibility of ur1 correlates with the start of expression of *hoxc13a* in the early to mid-segmentation transition. Transcription factor binding motif analysis (TFBM) of the peaks with synchronization and repression dynamics unveiled 57 different TFBMs: 16 shared between synchronization and

repression peaks, mostly represented by members of the SOX (9) and KLF (4) families, one (ELF1) exclusive of repression and 40 exclusives of synchronization from various families (Figure 7E; Supplemental table S14A).

The third type, early opening, refers to cases where chromatin accessibility was established before the onset of target gene expression. For this type of regulatory logic, 6,606 peak-gene pairs were identified nearby to 5,060 genes (Supplemental table S13). An example is *foxa2* (Figure 7C), a regulator of chondrocyte differentiation that starts to be highly expressed at the shield stage, with its promoter and downstream regulator (dr1) highly accessible in the late blastula stage. Early opening events represent almost half (5,125, 46.97%) of the 10,911 peak-pairs genes where the ATAC-seq and RNA-seq half-max stages (Stage at which ATAC-promoter peaks or their target gene reach half its maximum accessibility signal or half its maximum level of expression, respectively) were compared (Figure 7F).

The fourth type, regulatory switch, represents cases with complex regulation, where gene expression is maintained across development but with varied dynamics that correlate with opposing dynamics of regulatory elements in its neighborhood. For regulatory switch dynamics, a total of 578 high-confidence peak-pairs were identified, comprised of 869 peaks nearby to 268 genes (Supplemental table S13). An example is found nearby to *wnt11* (Figure 7D), a gene involved in cell fate determination and body axis planning, where five downstream regulators (dr2, 3, 4, 6 and 10) seem to drive *wnt11* expression changes. *Wnt11* shows moderate expression in late blastula and shield stages. Coinciding with the opening of dr2, dr3, dr4, dr6 and dr10 in early segmentation (the latter opening at the earlier shield stage), *wnt11* reaches its highest expression. By mid-segmentation, all dr elements except dr4 greatly decrease in accessibility, and *wnt11* expression returns to moderate levels, slowly decreasing as dr4 loses accessibility signal towards the prehatch stage.

Stage-specific transcription factor motif dynamics highlights developmental transitions and lineage specification

To identify the main TFs associated with chromatin accessibility dynamics in candidate CREs regulated during embryogenesis, TFBM enrichment was conducted for all DARs associated with promoters and putative enhancers. Significant TFBM enrichments were observed for all transitions (Figure 8A; Supplemental table S14B). The highest number of significantly enriched TFBMs was detected in the shield to early segmentation transition (75 motifs) followed by early to mid-segmentation (38) and late blastula to shield (20), with mid to late segmentation and late segmentation to prehatch having only 1 and 2 enriched TFBM, respectively.

Out of the 91 enriched TFBMs identified among DARs, 12 were enriched in the three transitions that span from late blastula to mid segmentation stages; 8 were motifs for different SOX family TFs (Figure 8A; Supplemental table S14B). 17 TFBMs were enriched in the transitions from shield to early segmentation and early to mid-segmentation. These represented TFs involved in muscle

(MYOD, MYOG) and neural (NEUROD1, NEUROG2) differentiation, tissue growth and cell proliferation (TEAD4) and three HOX proteins (HOXA9, HOXA11 and HOXB13), although most of the HOX binding motifs detected were exclusive to the early to mid-segmentation transition (along with TEAD motifs). Finally, 45 binding motifs were exclusively enriched in the shield to early segmentation transition. This group of TFBM mainly comprised members of the ETS and FOX TF families, two of the largest families of transcriptional regulators in metazoans, involved in cell proliferation, differentiation and development, as well as immune response (Figure 8A; Supplemental table S14B).

To further investigate the functions of TFs involved in turbot embryogenesis, Metascape (Zhou et al., 2019) was applied to identify clustered functions associated with TFs explaining all enriched TFBM (Figure 8B). The only enriched function detected through all available transitions was the deactivation of the β-catenin transactivating component, involved in the Wnt signaling pathway. Clustering for the late blastula to shield transitions unveiled functions related to regionalization, nervous and sensory organ development and the BMP signaling pathway. During the shield to early segmentation transition, organ / tissue development functions arise (i.e. primitive hemopoiesis, digestive system development, pancreas development, adipogenesis, lateral line development), as well as hormonal and other signaling pathways like FGF. Organogenesis-related functions continue to appear towards the early to mid-segmentation transition with the addition of myogenesis, gland development and peripheral nervous system development. No significant clustering was detected for the TFs involved in the mid-to-late, and late segmentation to prehatch transitions.

Comparative developmental transcriptomics between turbot and zebrafish

The developmental transcriptome of turbot and zebrafish was compared considering the six stages profiled for turbot (late blastula, embryonic shield, early segmentation, mid segmentation, late segmentation and prehatch) versus matched stages from zebrafish defined by White et al. (2017): dome, gastrula shield, 1-4 somites, 14-19 somites, 20-24 somites and long pec, respectively.

14,221 orthogroups were established between turbot and zebrafish; 8,466 genes constituted single copy orthogroups, with 6,967 shared genes showing expression (cut-off TPM > 5) in any of the six stages (Supplemental figure S12A; Supplemental table S15). However, some of these genes showed expression only in one species, 658 for turbot and 817 for zebrafish (Supplemental figures S12B and C, respectively), with 5,491 expressed in both species (Supplemental figure S13; Supplemental table S15).

Hierarchical clustering after inter-species normalization of the expressed orthologs for each species (Spearman's correlation) showed moderate correlations between expression patterns, with some genes showing varying levels of asynchrony between both species (Supplemental figure S12A, B and C). Enriched GO terms were only detected in the zebrafish-specific genes, with

most functions associated with early development functions, like reproductive processes, gamete generation, cell signaling and adhesion (Supplemental figure S12D; Supplemental table S15). Although no significant GO enrichment was detected in the turbot-specific genes, some notable genes were identified such as dystrophin (*dmd*; muscle development), patched 1(*ptch1*; hedgehog signaling pathway), GATA binding protein 4 (*gata4;* heart and gut development), KIT proto-oncogene (*kita;* germ and blood cell development), nuclear factor interleukin 3 regulated member 6 (*nfil3-6;* immune system development), cysteine rich transmembrane BMP regulator 1 (*crim1;* blood and neural tissues) and interleukin 1β (*Il-1b*; bone development).

Evolutionary conservation and divergence of Hox genes between turbot and zebrafish.

To investigate the evolutionary conservation and divergence of Hox genes, key regulators of anterior-posterior axis patterning and segmental identity, we compared the expression profiles of the 46 and 47 annotated Hox family genes annotated in the turbot and zebrafish genomes, respectively. Each species harbours seven Hox clusters distributed in Chromosomes 1, 11, 14, 15, 18, 19 and 22 in turbot and Chromosomes 3, 9, 11, 12, 16, 19 and 23 in zebrafish, for a total of 55 Hox genes (38 orthologs, 8 and 9 genes exclusive to turbot and zebrafish, respectively). HoxCb and HoxDb clusters were missing in turbot and zebrafish, respectively, while HoxBa, HoxBb and HoxDa were partially missing in turbot compared with zebrafish, and HoxAa and HoxBb in zebrafish compared with turbot (Figure 9A; Supplemental table S12).

24 Hox family genes mapping in orthologous syntenic regions shared very similar expression patterns between zebrafish and turbot (Table 1; Figure 9A and B; Supplemental table S16; Supplemental figure S13). Eight, however, showed variable degrees of asynchrony, for example, with *hoxa2b* in zebrafish reaching its peak of expression during mid-segmentation instead of early segmentation in turbot. Twelve *hox* orthologs showed different expression patterns. Two were not expressed in turbot (despite the gene being annotated) and 9 and 8 *hox* genes were not present in the genome of turbot and zebrafish, respectively.

16 Hox family genes showed similar expression patterns between turbot paralogs. Among them, 6 were asynchronous and 4 showed different expressions between paralogs (Table 2; Figure 9A and B; Supplemental table S16; Supplemental figure S13; Supplemental figure S14, Supplemental figure S15).

DISCUSSION

In this study, a multiomic approach, capturing transcriptomic and epigenomic dynamics, was used to provide insights into the transcriptional regulation of turbot embryogenesis, representing an essential resource to understand the genetic basis of traits relevant to sustainable turbot aquaculture, as well as for comparative development in teleost.

Our transcriptomic analysis revealed pervasive differential gene expression across developmental stage transitions, highlighting the dynamic regulation associated with the specification and differentiation of diverse cell types comprising the turbot body plan. Chromatin state analyses provided further insights into changes in chromatin state and CRE activity after ZGA is established (Minnoye et al., 2021). Our 10-state chromatin model defined states associated with active promoters and enhancers, repressed Polycomb regions and ATAC islands, annotated in other species following the same approach (Heintzman et al., 2007; Barral and Déjardin, 2023). Two of the identified chromatin states (FlnkBiv and Biv) had a combination of repressor and activation signals, classifying them as bivalent / poised states, which act to prepare genes for rapid transcriptional responses to regulatory cues during embryo development and immune responses (El-Dahr and Saifudeen, 2019; Herrera-Uribe et al., 2020). Genes that had their promoters dominated by these bivalent states were enriched in biological processes associated with cell differentiation, patterning and organ development, which were in turn not present in genes dominated by strictly active states, where RNA processing, transcriptional regulation and metabolism were consistently enriched. Bivalent chromatin states have previously been proposed to prime differentiation and developmental genes for future lineage commitment and fine tune gene expression levels (Yu et al., 2023; Glancy et al., 2024). Along with the weakly active chromatin states (TssFlnk and FlnkWk) and the active enhancer state (EnhA), the bivalent states were among the most dynamic across development, showing the highest rates of conversion to repressive states (ReprPC and LowSi) in developmental stage transitions, especially compared to that of active promoter states (STssA and TssA), This is coherent with the general understanding that enhancers are typically more dynamic than promoters (Ernst et al., 2011) and the previously mentioned characteristics of poised promoters.

The Onset of Zygotic Transcription: Chromatin Dynamics and Pioneer Factor Activity During Early Embryogenesis

At the onset of embryogenesis, during the first synchronized mitotic divisions, the embryo is transcriptionally quiescent, and development is driven by maternal transcripts (Xu et al., 2024). This was captured by our transcriptomic hierarchical clustering, which depicted three main stages, one comprising the maternal RNAs prior to ZGA, along with a ZGA associated and post-ZGA associated cluster. Moreover, all chromatin assays showed similar clustering, with ZGA stages well separated from post-ZGA stages, particularly noticeable in ATAC-seq, H3K4me3- and H3K27me3; this is expected since ZGA requires major chromatin remodeling events (Schulz and Harrison, 2019). The appearance of GO terms associated with cell cycle regulation, reproductive process and methylation in early developmental stages but not in later ones, is consistent with the presence of maternal transcripts that decay after ZGA and the establishment of methylation patterning at early development (Geiman and Muegge, 2009). GO terms enriched in the pre-ZGA stages (64-cells, morula and early blastula) are coherent with what would be expected of maternal transcripts in early development that decay after ZGA. However, minor zygotic transcription

activation is well documented to happen at even the cleavage stages (Abe et al., 2018), and in the early blastula (Zhou and Heald, 2023), so we cannot exclude that some genes in the pre-ZGA set may be transcribed by the embryo, subject to more detailed analysis in the future.

ZGA begins when a specialized set of TFs called pioneer factors (such as those coded by *pou5f3, nanog* and *sox19b* genes) bind to nucleosomes at promoters and enhancers during blastulation, leading to recruitment of additional TFs, chromatin remodeling enzymes, and polymerases (Barral and Zaret, 2024). In turbot, the mRNA expression of these pioneer factors peaked at early developmental stages (usually early blastula), then decreasing rapidly as development progresses. As pioneer factors find their target sequences on nucleosomes, histones are also marked and redistributed across the genome (Sinha et al., 2023). As an example, we detected writer (*brd4*) and reader genes (*ep300a* and *ep300b*) for the H3K27ac histone modification (Chan et al., 2019) to steadily increase in expression until the shield stage, with H3K4me3 writer (*kmt2b*) and reader genes (*taf3*) showing similar expression patterns (Hyun et al., 2017). Given their role in ZGA activation, future studies on turbot embryogenesis would benefit from incorporating chromatin accessibility analyses on pre-ZGA stages. Previous studies in zebrafish have shown that maternally provided factors regulate chromatin accessibility as early as the 256-cell stage (Pálfy et al., 2020). Moreover, other studies show that DNA methylation and certain histone marks can be partially retained during pre-implantation development in mammals (Zheng et al., 2016; Liu et al., 2018). Although it was generally assumed that non-mammalian vertebrates do not escape this early reprogramming, similar behaviors have been recently described in the cleavage stage of medaka (Fukushima et al., 2023), with marks like H3K27ac and H3K27me3 escaping complete reprogramming. Thus, screening for these early developmental chromatin states will facilitate creating a more comprehensive regulatory atlas of turbot embryogenesis.

Gastrulation and cell fate definition

As the embryo approaches the late blastula stage, cells undergo different paths of differentiation and migration, eventually giving rise to different cell types with committed fates and functions (Pálfy et al., 2020). Consistent with this information, the late blastula is an appropriate starting point for chromatin profiling through ATAC-seq and ChIP-seq assays.

The late blastula to shield was the second most dynamic transition in terms of differential gene expression. Most enriched GO terms were consistent with the formation of the three main germ layers through convergent extension during gastrulation and the start of pattern specification that happens in the late blastula to shield transition (Fulton et al., 2020). Moreover, the first enriched GO terms associated with morphogenesis were associated with this stage, as discrete cell populations appear during the blastula stage. Expression of genes associated with the embryonic enveloping layer and yolk-syncytial layer, such as *krt18a.1* and *gata3* (Xu et al., 2024), were detected in our study from the

shield stage onwards. We detected a similar pattern for various members of the Wnt gene family (*wnt5b*, *wnt11*), involved in cell fate determination through the Wnt/β-catening pathway and body axis planning through the Wnt/PCP pathway (Liu et al., 2022). In fact, *wnt11* was one of the genes showing regulatory switch logic, suggesting a fine adjustment of its expression across embryogenesis (Andre et al., 2015). Finally, we also detected an expression peak during the shield stage for *aldh1a2* and *rdh10* genes, involved in the synthesis of retinoic acid (RA) as well as genes from the Rar family of RA receptors. RA regulates key genes involved in patterning like the Hox, Meis and Sox gene families. Recent findings suggest that RA signaling may rewire the chromatin landscape leading to CRE activation, and exogenous RA induces chromatin binding of the key TFs HOXB1B, MEIS2B and SOX3 TFs (Moreno-Oñate et al., 2024).

Cell fate is driven by interactions between TFs and CREs with accessible chromatin, allowing TF binding (Mircea and Semrau, 2021). However, the number of DARs in gene promoters between consecutive stages of turbot embryogenesis was lower than the number of DEGs, with a small, non-significant overlap according to the hypergeometric test. This suggests that most promoters were already accessible without major alterations in global chromatin accessibility between consecutive stages, pointing to the key regulatory role of enhancers on gene expression. This asynchronism between chromatin accessibility of CREs and gene expression has been reported previously (Ma et al., 2020; Wike et al., 2021; Lin et al., 2023). Our data adds weight to this notion, as displayed by the chromatin state dynamics in a catalogue of 29 DEGs throughout embryogenesis (Supplemental figures S7 and S15). Each DEG shows accessible promoter regions from the late blastula stage, irrespective of further changes in chromatin accessibility or chromatin state complexity as development progresses.

A previous study in zebrafish suggests that promoter ATAC peaks generally precede transcriptional activity, but in contrast, the majority of distal CREs (which include enhancers) become accessible after gene expression starts (Liu et al., 2024). This past work fits well with our findings, where 20.86% of the DARs across development overlapped enhancer regions, compared to 12.64% for promoter regions.

Stabilization of chromatin dynamics during segmentation and organogenesis

By the end of gastrulation, the embryo starts a sequential patterning process along its axis, laying the groundwork for the body plan and organogenesis (Schröter et al., 2008). Accordingly, we detected enriched terms relating to somitogenesis and the development of various tissues and organs in mid-to-late embryogenesis (Cheng et al., 2023) and in the transitions between consecutive stages across segmentation (Miao and Porquié, 2024), while terms associated with locomotion and behavior were enriched specifically in the prehatch stage (Pohl, 2019). During the shield to early segmentation transition and subsequent stages, we observed the expected expression changes in master vertebrate

regulatory genes related to patterning (*bmp2b*, *bmp4*), chondrogenesis (*foxa2*, *foxa3*) osteoblastogenesis (*sp7*, *dlx3b*, *dlx5a*), muscle development (*dag1*, *myod1*) and nervous system development (*meis2b*, *sox2*, *neurog1*), among others (Petratou et al., 2024).The diversity and complexity of developmental and regulatory pathways at the onset of somitogenesis was also reflected in the high number of DEGs and DARS during the shield to early segmentation transition.

The number of DEGs steadily decreased as segmentation progressed, while the number of DARs decreased in each transition after early segmentation, most notably from mid-to-late segmentation. This may result in comparatively reduced transcriptional dynamics once the body plan is established, as well as stabilization of the chromatin landscape, as shown in the aggregated lineplots (Supplemental figure S11A and B), with the last major changes in the epigenomic landscape occurring in early segmentation. In fact, we reported a high number of peak-gene pairs showing early opening *cis*-regulatory logic, consistent with previous findings in medaka and mammals (Chereji et al., 2019; Li et al., 2020)

As ontogeny continues post-hatching, several biological processes were not enriched among genes expressed during embryogenesis. We identified enrichment of terms related to the immune system and feeding behavior among genes not expressed in the embryonic transcriptome, a result that aligns with studies on zebrafish, where immune-like responses, such as phagocytic activity were not detected until 1-day post-fertilization (Lange et al., 2023) and adaptive immunity becomes active three weeks post hatching (van der Vaart et al., 2012). On the other hand, feeding behavior is not properly active until complete yolk absorption and the mouth is fully developed (Lange et al., 2023).

Tracing Hox gene expression evolution

Hox genes are fundamental in shaping the anterior-posterior body plan of diverse metazoan taxa, as well as governing spinal ossification and regulating organogenesis (Petratou et al., 2024). Moreover, variations in their expression are linked to body plan differences across species, and Hox expression disruptions often lead to developmental disorders (Cumplido et al., 2024). Comparative analysis of Hox gene expression can help better understand the genetic basis of congenital abnormalities, while also providing evidence of shared ancestry between species.

Hox genes are organized in clusters of up to 13 genes, with a spatial organization that correlates with their temporal and spatial expression (Duboule, 1994). Three whole genome duplication (WGD) events in teleost evolutionary history (Pasquier et al. 2017) resulted in up to eight Hox clusters in teleost. However, Hox clusters have been modified by gene loss and co-option during teleost evolution, resulting in significant variation among lineages (Ozernyuk and Schepetov, 2022). In both turbot and zebrafish, the b paralogs of each initial cluster (HOXAb, HOXBb, HOXCb, HOXDb) have been mostly affected by this loss, even resulting in the complete loss of some clusters, such as turbot

cluster HOXCb and zebrafish HOXDb. Gene loss and subfunctionalization however has not removed any of the 13 Hox genes from the collinear expression chain, as the combined sum of all clusters in each of the species retains at least one Hox gene expressed in anterior (*hox-1* and *hox-2*), antero-central (*hox-3*), central (*hox-4* to *hox-8*) and posterior (*hox-9* to *hox-13*) regions of the embryo.

Even with lineage-specific Hox configurations, we found that the majority of Hox orthologs followed a similar expression pattern, increasing from early segmentation and reaching maximum mRNA levels during late segmentation and pharyngula stages. Although we detected eight Hox orthologs with similar but slightly asynchronous patterns between turbot and zebrafish, we cannot discard that these differences might not be biologically relevant, as the compared studies were not designed to capture the same developmental stages. Overall, the ortholog Hox genes identified with similar temporal expression patterns point to evolutionarily conserved functions between distant teleost clades, but spatial expression and functional studies are needed to confirm this hypothesis.

Hox genes are in a dynamic equilibrium between evolutionary conservation and divergence (Struemph and Henderson, 2021), both needed for the specification of a functional, but species-specific body plan. In that regard, we found 12 Hox orthologs with notably distinct patterns that certainly do not arise from different methodological approaches and may thus impact species-specific body plan regulation. Among them, some early expressed Hox genes in turbot, specially *hoxb8a* and *hoxc5a*, showed relatively high expression at the late blastula stage (~20 and ~25 TPM, respectively), while *hoxb1b* showed highest expression in the shield stage (~75 TPM). Although early Hox gene expression is generally observed during gastrulation in vertebrates (Wacker et al., 2004), several Hox genes have been reported to be expressed in early embryos and even oocytes in bovine and mouse (Paul et al., 2011). Further validation of these findings is needed to confirm functional significance of differences in Hox ortholog expression.

Divergent expression of turbot Hox paralogs

WGD events play a major role in genome evolution of teleost fishes (Ozernyuk and Schepetov, 2022). WGD provides raw material for evolutionary innovation, but the long-term fate of duplicated genes can vary widely. Following duplication, one gene copy (or both; subfunctionalization) may retain the ancestral function, while the other is free to accumulate mutations that could lead to pseudogenization, or neofunctionalization (Ohno, 1970). In salmonids, it was shown that ohnolog (paralogs from WGD) expression levels tend to evolve asymmetrically (Gillard et al., 2021), diverting one copy of the gene towards lower expression levels and the possibility of complete loss of function. However, some teleost-specific paralog pairs in turbot with similar temporal expression profiles (e.g., *hoxb3, hoxb5, hoxd11*) suggest that retaining similar copies may be important for embryonic functions (Pasquier et al. 2017). Looking

at the turbot Hox teleost paralogs, 25 out of 47 have been reduced to a single copy, with the most apparent non-functionalization being the loss of cluster hoxCb, leaving 11 hoxCa unpaired genes.

We hypothesize that species-specific gene repertories, such as those for different Hox configurations among teleost, may contribute to the morphological diversity of this group and may benefit from functional annotation resources to enable targeted functional and gene editing analyses (Cumplido et al. 2024). Indeed, a proportion of Hox paralogs have diverged in their expression pattern, as illustrated by the differences observed between the *a* and *b* pairs of *hoxa11*, *hoxa13* and *hoxd9*, or the similar but asynchronous expression patterns for *hoxa2*, *hoxa10* and *hoxb1*. Although paralogous genes may conserve their ancestral expression (Hunter and Prince, 2022), they often diverge, implying specialized or distinct functional roles (McClintock et al., 2001), or altered constraints leading towards loss of function (Kuraku and Meyer, 2009), as happened with the HOXCb cluster. Our analysis on Hox gene expression profiles suggest that this loss of function seems to be happening with the *hoxa9* and *hoxd4* genes, where paralog b shows much lower expression levels. It is worth mentioning that even paralogous pairs that have highly similar expression properties in bulk RNA-seq may show spatially distinct expression in different cell types or locations in the embryo. Future works that incorporate spatial transcriptomics or single cell sequencing will build upon the developmental atlas of turbot generated in this study.

Concluding remarks

Our study presents the first comprehensive atlas of regulatory element activity dynamics involved in turbot embryogenesis, contributing to the functional annotation of animal genomes (FAANG) initiative, where farmed fish are now highly prioritized (Clark et al. 2020). Our findings offer insights into (epi)genomic processes driving turbot embryogenesis serving as a novel resource for turbot developmental biology with applications in aquaculture. Future research will benefit from integrating the regulatory annotations generated by this study with genetic variants identified from whole genome re-sequencing. Such work will aid in prioritizing causal genetic variants related to growth, developmental abnormalities and reproductive traits, which has great potential value for selective breeding in turbot aquaculture.

MATERIALS AND METHODS

Embryos

A standard batch of fertilized turbot eggs was provided by Stolt Sea Farm SA (Ribeira, Spain) 2 hours post fertilization (hpf). Eggs were housed in an aerated aquarium with saltwater at the facilities of the Department of Genetics in the Veterinary Faculty of the University of Santiago de Compostela (Spain) at 14.0ºC (±0.7ºC) until hatching. All samples were collected before hatching (90hpf).

Experimental Protocols

Detailed protocols for egg dechorionation, embryo preservation, RNA isolation and ATAC-seq and ChIP-seq procedures (including library preparation) are available in the FAANG repository (data.faang.org; URLs for protocols in Supplemental tables S1 and S2).

Genome annotation

All RNA-seq, ATAC-seq and ChIP-seq analyses have been carried with turbot Ensembl genome ASM1334776v1 (GCA_013347765.1, Martínez et al., 2021) as reference. All gene and protein names used in the manuscript adhere to their official symbols, and supplemental tables provide the turbot-specific Ensembl IDs of these genes.

Full egg preservation for RNA-seq

Three pools of 50 to 150 embryonated eggs were preserved for RNA extraction of 12 developmental stages, with the number of eggs in each pool depending on the sampled stage (Figure 1, Supplemental table S1). The desired number of eggs were collected in 2 ml safe-lock tubes from the surface of the water with a Pasteur pipette. Saltwater was extracted from the tubes using P1000 and P200 pipettes. Eggs were then resuspended in 1 ml of TRIzol reagent (Invitrogen) and frozen at -80ºC for long term storage.

RNA isolation and sequencing

After sample thawing, TRIzol was removed and substituted with 700µl of Qiazol Lysis Reagent (miRNeasy Mini Kit, QIAGEN). Two zirconia beads (2.8-3.3 mm) were added to each tube before disrupting the eggs for 2 min, 20Hz in TissueLyser II (QIAGEN). After disruption, total RNA was extracted and purified using the miRNeasy Mini Kit (QIAGEN), following a protocol for “Total RNA extraction for turbot eggs” (Supplemental table S2). RNA integrity and quantity were evaluated using a Bioanalyzer (Bonsai Technologies, Madrid, Spain) and NanoDrop® ND-1000 spectrophotometer (NanoDrop® Technologies Inc., Wilmington, DE, USA). The RNA integrity number (RIN) averaged 9.5 across all samples and was always > 7.7. RNA samples were delivered to Novogene (UK) for library preparation using NEBNext Ultra Directional RNA Library Prep Kits for Illumina and sequenced using an Illumina NovaSeq S4 platform to generate 150 bp paired-end reads.

RNA-seq data processing

RNA-seq data was processed using nf-core/rnaseq 3.10.1 (Patel et al., 2023) with default parameters. In brief, the pipeline evaluated quality of raw reads using FastQC (Andrews, 2010) and trimmed adapters and low-quality nucleotides using Trim Galore! (Martin et al., 2011). Reads were then mapped to the genome using STAR (Dobin et al., 2013).

Differential expression analysis

Normalized transcript read counts were obtained using RSEM (Li and Dewey, 2011). After nf-core processing, the resulting count tables were filtered to

remove genes with TPM < 5 and represented in less than two samples across all developmental stages. These genes constitute the total expressed transcriptome during turbot embryogenesis, while genes that didn't pass the filter will be referred as the total non-expressed transcriptome. These genes were clustered using BioLayout (Theocharidis et al., 2009) with the following parameters: 0.94 Pearson's correlation threshold; 2.1 cluster granularity, minimum 8 genes per cluster, Markov clustering (MCL) and Fruchterman-Reingold layout algorithm

Egg dechorionation for ATAC-seq and ChIP-seq procedures

Four Petri dishes, each with 200 eggs per dish, were prepared 3 h before the estimated sampling point for each stage (Figure 1, Supplemental table S2). Saltwater was removed from three Petri dishes using a Pasteur pipette, and eggs resuspended in a pronase solution (from *Streptomices griseus*, ROCHE, 2 mg/ml), leaving the fourth dish as a control. All four Petri dishes were transferred to an oscillator (medium velocity) at 19ºC (±1ºC) for 2.5 to 3.5 h depending on stage. The consistency of the chorion was checked by pipetting 10 embryos of each plate after 2.5 h every 20 min. Once the chorion was weakened, the pronase solution was discarded with a Pasteur pipette, and eggs washed twice with PBS. The chorion was then removed by pipetting the eggs up and down, with freed embryos collected in 1.5 LoBind tubes with PBS.

Cell isolation, library preparation and sequencing for ATAC-seq

Following a standard protocol for "Cellular isolation of turbot embryos" (Supplemental table S1), freshly dechorionated embryos were disaggregated with two distinct procedures, depending on embryo stage. Late-blastula embryos were disaggregated by pipetting up and down 10 times with a P1000 pipette, and vortexed at slow speed for 15 s. Samples for the other five stages were resuspended in 1 ml of Trypsin-EDTA and left under rotation (50 rpm, room temperature -RT-) for 5 min (embryonic shield and early somitogenesis), 10 min (mid and late somitogenesis) or 40 min (prehatch). The trypsin reaction was stopped with 500 µl of cold, heat-inactivated FBS, then centrifuged at 500 g for 6 min at 4ºC and resuspended in 1 ml cold PBS. Three replicates for each developmental stage were used for ATAC-seq (Figure 1, Supplemental table S2).

The number and integrity of nuclei were assessed with a haemocytometer (minimum of ~50,000 nuclei in a 16.5 µl suspension; ~3,000 nuclei/µl) before carrying out the Tn5 transposase reaction with Illumina Tagment DNA TDE1 enzyme (Illumina) using a MS-100 Thermo-shaker incubator (Hangzhou Allsheng Instruments) at 37°C, 30 min, 1000 rpm, following the standard "Omni-ATAC protocol" (Corces et al., 2017; Supplemental table S2). The resulting DNA was purified with a MinElute PCR purification kit (Qiagen), and DNA concentration assessed with a Qubit using the dsDNA HS kit (ThermoFisher Scientific). Library amplification was carried out with Integrated DNA Technologies (IDT) for Illumina UD Indexes (96x, Plate A, Set 1, Illumina). Library size selection was performed to remove fragments below 180 bp and

above 700 bp using AMPure XP beads (Beckman Coulter). Finally, DNA fragment size distribution was assessed with the Bioanalyzer High Sensitivity DNA Assay kit (Agilent Technologies). ATAC-seq libraries were delivered to Novogene (UK) and sequenced on an Illumina NovaSeq S4 platform generating 150 bp paired-end reads.

Embryo cross-linking for ChIP-seq

Three replicates were taken for each of the six stages (Figure 1, Supplemental table S2). The freshly prepared tubes with pools of dechorionated embryos were centrifuged at 500 g for 6 min at 4ºC. After removing the supernatant, embryos were resuspended and incubated in 875 µl of 1% formaldehyde solution on a rotating mixer device (Hula Mixer) at 50 rotations per minute for 15 min at room temperature. Next, 125 µl of 1M glycine solution was added to quench the reaction, incubated under rotation at 40 rotations per minute for 10 min at room temperature. Tubes were centrifuged at 500 g for 6 min at 4ºC, removing the supernatant, and resuspended in cold PBS before storage at -80ºC.

Library preparation and sequencing for ChIP-seq

The frozen crosslinked embryos were thawed on ice. The number and integrity of nuclei was assessed with a haemocytometer (minimum 100,000 nuclei). Nuclei were pelleted and resuspended in complete sonication buffer, and the chromatin sheared using a Covaris S2 focused ultrasonicator with the following parameters: 2% duty cycle, intensity 3, with 200 cycles per burst, at 4°C for 7 min (Supplemental table S2).

Due to the expected low number of recovered nuclei, the immunoprecipitation (IP) and library preparation steps were performed following a modified µChIPmentation protocol (Diagenode, Supplemental table S2) using Diagenode antibodies for immunoprecipitation (30 ng of chromatin per IP) of three histone marks: H3K4me3 (cat. No. C15410003; 1.3 µg/µl), H3K27ac (cat. No. C15410196; 2.8 µg/µl) and H3K27me3 (cat. No. C15410195; 1.1 µg/µl). Each of the three replicates per developmental stage were subjected to IP. Additionally, for each developmental stage, 4 µl of pooled chromatin (~1.34 µl per replicate) was used as input controls, i.e. without IP.

Before sequencing, the quantity and quality of purified libraries were assessed using the Qubit DNA HS kit (ThermoFisher Scientific) and the High Sensitivity DNA Assay kit (Agilent Technologies), respectively. A minimum of 60% of the chromatin was required to have a size distribution between 200-700 bp (centred around 350-400 bp). ChIP-seq libraries were delivered to Novogene (UK) for sequencing on an Illumina NovaSeq S4 platform generating 150 bp paired-end reads.

ATAC-seq and ChIP-seq data processing

ATAC-seq and ChIP-seq data were processed using the nf-core/atacseq v1.2.1. and nf-core/chipseq v1.2.2 pipelines, respectively (Ewels et al., 2023), run with

the narrow_peak option for the ATAC-seq, and H3K4me3 and H3K27ac ChIP-seq datasets, and with the broad peak option for H3K27me3. Other parameters were kept as default, including the removal of mitochondrial DNA from the dataset. No additional filtering of nucleosome-free regions was conducted. Briefly, quality assessment of the reads was carried out with FastQC (Andrews, 2010), and adapters and low-quality bases were trimmed with Trim Galore! (Martin et al., 2011). Reads were mapped to the turbot genome using BWA (Li and Durbin, 2009). Further filtering was done with SAMtools (Danecek et al., 2021), BAMtools (Barnett et al., 2011) and Pysam (Danecek et al., 2021). Genome-wide IP enrichment relative to controls was done with deepTools (Ramírez et al., 2014) and broad/narrow peaks were called using MACS2 (Zhang et al., 2008).

An ATAC-seq consensus peak list was generated using ATAC-seq peaks detected consistently in two or more biological replicates in at least one stage using the R/Bioconductor package DiffBind (Stark and Brown, 2011) with default settings. Correlation heatmaps were generated using peak and alignment data for each library. Additionally, a previous turbot ChIP-seq blacklist (Aramburu et al., 2025) of high signal (5.58% of the turbot genome) and low mappability regions (1.40%) was incorporated for downstream filtering.

Chromatin state inferences

Genome-wide chromatin states were modelled across turbot embryogenesis using ChromHMM (Ernst and Kellis, 2017) integrating the ChIP-seq and ATAC-seq data. ChromHMM models including 8 to 15 states were tested (200 bp bin resolution), with the aim of maximizing the number of non-redundant, biologically relevant chromatin states (Vu and Enst, 2022; Pan et al., 2023). Transition probabilities of chromatin states between consecutive developmental stages were estimated by calculating the percentage of 200 bp bins assigned to chromatin state X in developmental stage A, that changed to chromatin state Y in developmental stage B, for each pair of chromatin states. Additionally, the narrow promoters (-1000 to +100 bp of the TSS) of each annotated gene in the turbot genome were classified in discrete categories according to the dominant chromatin state within it (i.e., chromatin state that covered the highest percentage of sequence in the narrow promoter). Each chromatin state was given a different priority (1 to 10) for annotation of dominant states, in order to resolve instances where promoters harbored more than one possible dominant state.

The ATAC-seq consensus peak set was intersected and annotated with the chromatin-state map for each embryonic stage, before peaks were further annotated according to their overlap to: broad promoters (-2000 to +600 bp of each gene TSS), introns, exons or intergenic regions. ATAC-seq peaks annotated as promoters and intron/intergenic regions with active chromatin states were considered as active promoters and active putative enhancers, respectively.

Differential gene expression and chromatin accessibility analysis

DEGs between consecutive pairs of the six embryonic stages were identified using the R/Bioconductor package DESeq2 v1.44 (Love et al., 2014), resulting in five comparisons. Genes with false discovery rate (FDR) adjusted $p < 0.05$ were considered DEGs. Using the consensus ATAC-seq peaks, DARs (adjusted $p < 0.05$) between the same pairs of consecutive stages were identified using the R/Bioconductor package DiffBind (Stark and Brown, 2011), with default settings.

Cis-regulatory logic and correlation between DEGs and DARs

To investigate the relationship between gene expression and chromatin accessibility, two approaches were followed. First, for each comparison between sequential developmental stages, the overlap significance between DEGs with DARs in the matched promoter region was tested with a hypergeometric test ($p < 0.05$, after Bonferroni correction).

Secondly, each ATAC peak was assigned as a *cis*-regulatory element to their closest gene (nearest TSS), assuming a correlation between functional regulation and genomic distance (Yoshida et al., 2019). Then, each peak-gene relationship was classified into four groups of *cis*-regulatory logic described by Li et al. (2020): synchronization, repression, early opening and regulatory switch. The pipeline by Li et al. (2020) was adapted to compare the distribution of transcript abundance (TPM) and chromatin accessibility (RPKM) across six developmental stages.

To identify cases of synchronization and repression, Spearman's correlations were calculated for each peak-gene pair, and pairs with correlations above 0.8 or below -0.8, were classified as cases of synchronization and repression, respectively. To consider early opening of regulatory elements, first we identified expressed genes and assigned consensus ATAC peaks that showed a gradual increase in expression/accessibility in at least three successive developmental states. Then, for each pair we checked the developmental stage when the expression level of the gene and the accessibility of the peak reached 50% of their maxima (halfmax state). Those cases where the halfmax accessibility of the peak occurred at an earlier developmental stage than the halfmax expression level of the genes were tagged as cases of early opening. Finally, regulatory switching dynamics were identified by looking at peak-pairs assigned to the same gene that showed a negative Spearman's correlation below -0.8. Then, a linear regression was applied to fit gene expression level and peak accessibility of each element of the pair. Those cases with an F-test cutoff below 0.1 were binned into the regulatory switch category.

Gene Ontology analysis

Gene Ontology (GO) enrichment for the tested gene sets (see Results) was performed using ShinyGO v0.80 (Ge et al., 2020) for Biological Process (BP) terms, which were ranked by statistical significance (FDR-adjusted $p < 0.05$). All resulting datasets were summarized using Revigo v1.8.1 (Medium complexity, FDR enrichment used as scaling parameter, Supek et al., 2011)

Transcription factor motif analysis

For each list of DARs annotated as active promoters or enhancers, enriched TFBMs annotated by HOMER (Heinz et al., 2010) were identified using the findMotifsGenome.pl function (settings: -size given -mask-mset vertebrates). Random genomic regions with GC-content matching each input genomic list were used as the background for automatic motif analysis by HOMER. TFBMs with adjusted $p < 0.05$ after Bonferroni correction representing > 10% of target sequences were considered enriched. GO analysis of TFs predicted to bind enriched TFBMs was performed with Metascape (Zhou et al., 2019), producing ontology enrichment clusters annotated through the Molecular Complex Detection algorithm, using zebrafish (*Danio rerio*) as reference.

Comparative analyses with zebrafish

Orthology analysis between the predicted proteomes of turbot (assembly ASM1334776v1) and zebrafish (GRCz11) was performed with OrthoFinder 2.5.4 (Emms and Kelly, 2019), restricted to the canonical predicted amino acid sequences. The orthology trees were rooted with the proteomes of three other teleost species: *Cynoglossus semilaevis* (Cse_v1.0), *Oryzias latipes* (ASM223467v1) and *Lepisosteus oculatus* (LepOcu1).

Turbot and zebrafish genes included as single copy orthogroups (i.e. only one orthologous gene in all included species) were retrieved for comparative analysis. Expression profiles were compared between the six main embryonic stages studied in turbot (late blastula, embryonic shield, early segmentation, mid segmentation, late segmentation and prehatch) and the six homologous morphological stages studied by White et al. (2017): dome, gastrula shield, 1-4 somites, 14-19 somites, 20-24 somites and long-pec, respectively.

Graph-based comparison of expression patterns across the six main developmental stages of both species was accomplished with R by grouping the genes in our dataset by their stage of maximum expression. Then, intra-species normalization of TPM counts was carried out against the top value across all stages considered, within each species. Spearman's correlation coefficient was used to calculate a distance matrix by hierarchical clustering (Complete linkage).

DATA ACCESS

All raw sequencing data was submitted to the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/home) under accession numbers PRJEB47933, PRJEB47934 and PRJEB57784

The in-depth experimental protocols have been submitted to the FAANG repository (https://data.faang.org/home), sustained by EMBL-EBI, under the URLs found in supplemental tables S1 and S2.

COMPETING INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

OA: Methodology, Software, Formal analysis, Investigation, Data Curation, Visualization, Writing original draft, Writing – review & editing; **BGP:** Methodology, Investigation, Resources, Writing review & editing; **AJG:** Formal Analysis, Methodology, Supervision, Resources, Writing – review & editing; **ABH:** Software, Formal Analysis, Data Curation, Visualization; **DJM**: Funding Acquisition, Methodology, Supervision, Resources, Writing – review & editing; **CB**: Conceptualization, Formal Analysis, Resources, Writing – review & editing, Project Administration, Supervision; **PM**: Conceptualization, Methodology, Formal Analysis, Resources, Writing – review & editing, Project Administration, Supervision.

This study was funded by the AQUA-FAANG project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817923. Additional funding was provided by Xunta de Galicia local government (Spain) (ED431C 2022/33), which also supported the research fellowship of OA (refs. ED481A-2020/119). Additional support was provided by the Biotechnology and Biological Sciences Research Council (BBSRC), namely Institute Strategic Programme grants to the Roslin Institute (BBS/E/RL/230001B and BBS/E/D/10002070). We acknowledge the technical support and informatic resources provided by the Centro de Supercomputación de Galicia (CESGA). We greatly appreciate the technical support provided by Lucía Insua and Susana Sánchez.

MAIN TEXT FIGURE DESCRIPTIONS

Figure 1. Developmental stages for transcriptomic and epigenomic assays. Details include the time of sampling (hours post-fertilization, hpf), the number of embryos pooled per replicate, and number of replicates per assay (RNA-seq, ATAC-seq and ChIP-seq for H3K4me3, H3K27ac and H3K27me3) sequenced (denominator) and passing QC (numerator). Stage colors are maintained throughout all figures.

Figure 2. Transcriptome analysis of turbot embryogenesis. A) PCA and **B)** heatmap (Spearman's correlation) of RNA-seq data on the 12 developmental stages and replicates studied; **C)** number of expressed genes (transcripts per million, TPM > 5) for each developmental stage.

Figure 3. Dynamics of the turbot embryonic transcriptome A) BioLayout correlation networks show the transition of gene (dots) expression patterns across development. The diagram represents clustering produced by MCL on a network graph linking genes with a Pearson correlation coefficient > 0.94 and removing unassigned genes and clusters with less than six genes. Genes with similar expression patterns group together. The line plots show the averaged expression pattern of a selection of correlated gene clusters across development; **B)** Heatmap of normalized gene expression values across developmental stages (TPM > 5 in at least two replicates per stage). Genes are hierarchically ordered by the stage where the average maximum

expression occurs. **C)** Number of differentially expressed genes (DEGs) for each comparison between consecutive developmental stages.

Figure 4. Enriched biological processes at different stages of turbot embryogenesis. A) Treemap chart of enriched GO terms (Biological Process, BP) in the total expressed embryonic transcriptome. Terms related to similar functions are grouped by color and the area of each term is inversely proportional to its FDR p-value. **B)** UpSet plot showing the overlapping enriched GO terms between consecutive developmental stages; each colored box corresponds to the same colored barplot, summarizing BP terms. Reproduction and reproductive process GO terms enriched in the L. Blastula to Shield transition might reflect maternal mRNA clearance.

Figure 5. Overview of epigenomic assays across turbot embryogenesis. A) Peaks detected across assays, developmental stages, and replicates; **B)** PCAs for each assay and **C)** Spearman's correlation heatmap between developmental stages for each assay based on genome-wide chromatin signal across consensus peaks.

Figure 6. Chromatin state dynamics during turbot embryogenesis. A) 10-state chromatin model of the turbot genome: Heatmaps of the emission parameters for each annotated chromatin state (left colored cells) are shown for ATAC-seq and ChIP-seq data with three histone marks (H3K4me3, H3K27ac and H3K27me3) (left heatmap); the emission parameters of each chromatin state (± 2,000 bp) around the TSS (middle); and their overlap with genomic features (right). **B)** Average genomic coverage for each chromatin state (%) per developmental stage, excluding the low signal state. Each color corresponds to the colors in part A. **C)** Sankey diagram showing the dynamics of each chromatin state across development. Each colored flow represents a percentage of 200bp bins of chromatin state A that changed to chromatin state B in the next sequential developmental stage. Rare transitions representing less than 5% of the original state are not plotted and can be seen in Supplemental figure S7A. Detailed percentages for each transition are found in Supplemental table S8H. **D)** Number of DARs per comparison between consecutive developmental stages. TSS: Transcription start site; TES: Transcription end site; DAR: differentially accessible region.

Figure 7. Common types of *cis*-regulatory logic between genes and consensus ATAC peaks. Schematic diagrams and examples of: **A)** synchronous logic between upstream regulatory elements (ur1 and ur2) and gene expression of *pou5f3*; **B)** repression logic between an upstream regulatory element (ur1) and gene expression of *hoxc13a*; **C)** early opening logic of the promoter (prom) and a downstream regulatory element (dr1) of *foxa2*; **D)** regulatory switch logic between downstream regulatory elements (dr2, dr3, dr4, dr6, dr10) of *wnt11*. **E)** Comparison between transcription factor binding motif enrichments of consensus ATAC peaks classified under synchronous and repression regulatory logic categories with their putative target gene. **F)** Comparison between the developmental stages at which ATAC-promoter peaks reach half its maximum accessibility signal and the stages at which their target gene reach half its maximum level of expression. Green bubbles below the red dotted line represent peaks that show early opening logic compared to the expression of their target gene.

Figure 8. Transcription factor binding motif enrichment in active regulatory elements showing dynamic changes across consecutive developmental stages. A) UpSet plot showing the overlap of TFs with enriched TFBMs shared by the differentially active regulatory elements found across comparisons between sequential developmental stages; each colored box corresponds to the same-colored histogram bar, in the same order. **B**) Ontology enrichment graph clusters of TFs with TFBMs

enriched along differentially active regulatory elements found across comparisons between sequential developmental stages; each cluster is colored with the most statistically significant term among terms clustered together; the size of each term is given by -$\log_{10}$p and the stronger the similarity among nearby terms, the thicker the edges between them. No significant ontology enrichment was detected for the mid to late, and late segmentation to prehatch transitions.

Figure 9. Comparative Hox gene expression analysis. A) Heatmap of normalized expression profiles of Hox family genes across 6 matched developmental stages in turbot and zebrafish. **B**) Examples of normalized expression profiles for Hox family genes with i) similar expression between orthologs and paralogs (*hoxa9*), ii) different pattern of orthologs (*hoxc5*), and iii) different pattern between orthologs and paralogs (*hoxd9*).

MAIN TEXT TABLES

Table 1. Categorization of Hox family genes based on the expression pattern similarity between turbot and zebrafish orthologs.

Category	Number of *hox* genes	*hox* genes
Similar pattern between orthologs	16	*hoxa3a, hoxa9a, hoxa9b, hoxb2a, hoxb3a, hoxb4a, hoxb6a, hoxb8a, hoxb9a, hoxb13a, hoxc6a, hoxc8a, hoxc9a, hoxd3a, hoxd4a, hoxd12a*
Similar but asynchronous pattern between orthologs	8	*hoxa2b, hoxa5a, hoxa10b, hoxb1a, hoxb1b, hoxc12a, hoxc13a, hoxd10a*
Different patterns between orthologs	12	*hoxa11a, hoxa11b, hoxa13a, hoxb5a, hoxb5a, hoxc1a, hoxc4a, hoxc5a, hoxc10a, hoxc11a, hoxd9a, hoxd11a*
Not expressed in turbot	2	*hoxa1a, hoxa13b*
Not in turbot genome	9	*hoxb6b, hoxb7a, hoxb8b, hoxb10a, hoxc3a, hoxc6b, hoxc12b, hoxc13b, hoxd13a*
Not in zebrafish genome	8	*hoxa2a, hoxa4a, hoxa7a, hoxa10a, hoxb3b, hoxd4b, hoxd9b, hoxd11b*

Table 2. Categorization of Hox family genes based on the expression pattern and expression level similarities between turbot paralogs.

Category	Number of *hox* genes	*hox* genes
Similar pattern and	6	

expression levels between paralogs		*hoxb3a and b, hoxb5a and b, hoxd11a and b*
Similar pattern, different expression levels between paralogs	4	*hoxa9a and b, hoxd4a and b*
Similar but asynchronous pattern between paralogs	6	*hoxa2a and b, hoxa10a and b, hoxb1a and b*
Different pattern between paralogs	6	*hoxa11a and b, hoxa13a and b, hoxd9a and b*
Unpaired paralog genes	25	*hoxa1a, hoxa3a, hoxa4a, hoxa5a, hoxa7a, hoxb2a, hoxb4a, hoxb6a, hoxb8a hoxb9a, hoxb13a, hoxc1a, hoxc4a, hoxc5a, hoxc6a, hoxc8a, hoxc9a, hoxc10a, hoxc11a, hoxc12a, hoxc13a, hoxd3a, hoxd10a, hoxd12a, hoxd13a*

REFERENCES

Abdellaoui N, Kim MS (2024). Transcriptome Profiling of Gene Expression in Atlantic Salmon (*Salmo salar*) at Early Stage of Development. *Marine Biotechnology*, *26*(5), 964–974. doi:10.1007/s10126-024-10354-4.

Abe K, Funaya S, Tsukioka D, Kawamura M, Suzuki Y, Suzuki MG, Schultz RM, Aoki F (2018). Minor zygotic gene activation is essential for mouse preimplantation development. *Proceedings of the National Academy of Sciences*, *115*(29). doi: 10.1073/pnas.1804309115.

Andre P, Song H, Kim W, Kispert A, Yang Y (2015). *Wnt5a* and *Wnt11* regulate mammalian anterior-posterior axis elongation. *Development*. doi: 10.1242/dev.119065.

Andrews S (2010). FastQC: A Quality Control Tool for High Throughput sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

APROMAR (2023). La acuicultura en España 2023. Asociación Empresarial de Acuicultura de España, Cádiz (Online, accessed October 18, 2024 at www.apromar.es)

Aramburu O, Gómez-Pardo B, Rodríguez-Villamayor P, Lamas J, Dewari P, Perojil-Morata D,, Macqueen DJ, Bouza C, Martínez P (2025). Multiomics uncovers the epigenomic and transcriptomic response to viral and bacterial stimulation in turbot. *GigaScience, 14,* giaf077. doi:10.1093/gigascience/giaf077.

Barral A, Zaret KS (2024). Pioneer factors: roles and their regulation in development. *Trends in Genetics*, *40*(2), 134–148. doi:10.1016/j.tig.2023.10.007.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics, 27(12), 1691–1692. doi: 10.1093/bioinformatics/btr174.

Barral A, Déjardin J (2023). The chromatin signatures of enhancers and their dynamic regulation. *Nucleus*, *14*(1). doi: 10.1080/19491034.2022.2160551.

Bizuayehu TT, Mommens M, Sundaram AYM, Dhanasiri AKS, Babiak I (2019). Postovulatory maternal transcriptome in Atlantic salmon and its relation to developmental potential of embryos. *BMC Genomics*, *20*(1), 315. doi: 10.1186/s12864-019-5667-4.

Blasweiler A, Megens HJ, Goldman MRG, Tadmor-Levi R, Lighten J, Groenen MAM, Dirks RP, Jansen HJ, Spaink HP, David L, Boudinot P, Wiegertjes GF (2023). Symmetric expression of ohnologs encoding conserved antiviral responses in tetraploid common carp suggest absence of subgenome dominance after whole genome duplication. *Genomics*, *115*(6), 110723. doi: 10.1016/j.ygeno.2023.110723.

Bouza C, Losada AP, Fernández C, Álvarez-Dios JA, de Azevedo AM, Barreiro A, Costas D, Quiroga MI, Martínez P, Vázquez S (2024). A comprehensive coding and microRNA transcriptome of vertebral bone in postlarvae and juveniles of Senegalese sole (*Solea senegalensis*). *Genomics*, *116*(2), 110802. doi: 10.1016/j.ygeno.2024.110802.

Brantley SE, di Talia S (2021). Cell cycle control during early embryogenesis. *Development*, *148*(13). doi: 10.1242/dev.193128.

Cai X, Gao C, Lymbery AJ, Armstrong NJ, Ma L, Li C (2023). The immune-related circRNA-miRNA-mRNA ceRNA regulatory network in the liver of turbot (*Scophthalmus maximus L.*) induced by *Vibrio anguillarum*. *Fish & Shellfish Immunology*, *132*, 108506. doi: 10.1016/j.fsi.2022.108506.

Chan SH, Tang Y, Miao L, Darwich-Codore H, Vejnar CE, Beaudoin JD, Musaev D, Fernandez JP, Benitez MDJ., Bazzini AA, Moreno-Mateos MA, Giraldez AJ (2019). Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. *Developmental Cell*, *49*(6), 867-881.e8. doi: 10.1016/j.devcel.2019.05.037.

Chen H, Einstein LC, Little SC, Good MC (2019). Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. *Developmental Cell*, *49*(6), 852-866.e7. doi: 10.1016/j.devcel.2019.05.036.Cheng T, Xing YY, Liu C, Li YF, Huang Y, Liu X, Zhang YJ, Zhao GQ, Dong Y, Fu XX, Tian YM, Shu LP, Megason SG, Xu PF (2023). Nodal coordinates the anterior-posterior patterning of germ layers and induces head formation in zebrafish explants. *Cell Reports*, *42*(4), 112351. doi: 10.1016/j.celrep.2023.112351.

Chereji Rv, Eriksson PR, Ocampo J, Prajapati HK, Clark DJ (2019). Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Research*, *29*(12), 1985–1995. doi: 10.1101/gr.249326.119.

Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, Houston R D, Kühn C, Lien S, Macqueen DJ, Reecy JM, Robledo D, Watson M, Tuggle CK , Giuffra E (2020). From FAANG to fork: application of highly annotated genomes to improve farmed animal production. Genome Biology, 21(1), 285. doi: 10.1186/s13059-020-02197-8.

Cumplido N, Arratia G, Desvignes T, Muñoz-Sánchez S, Postlethwait JH, Allende ML (2024). *Hox* genes control homocercal caudal fin development and evolution. *Science Advances*, *10*(3). doi: 10.1126/sciadv.adj5991.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). doi: 10.1093/gigascience/giab008.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. doi: 10.1093/bioinformatics/bts635.

Duboule D (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development*, *1994*(Supplement), 135–142. doi: 10.1242/dev.1994.Supplement.135.

DuBuc TQ, Stephenson TB, Rock AQ, Martindale MQ (2018). Hox and Wnt pattern the primary body axis of an anthozoan cnidarian before gastrulation. *Nature Communications*, *9*(1), 2007. doi: 0.1038/s41467-018-04184-x.

El-Dahr SS, Saifudeen Z (2019). Epigenetic regulation of renal development. *Seminars in Cell & Developmental Biology*, *91*, 111–118. doi: 10.1016/j.semcdb.2018.08.014.

Emms DM, Kelly S (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. doi: 10.1186/s13059-019-1832-y.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, *473*(7345), 43–49. doi: 10.1038/nature09906.

Ernst J, Kellis M (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols*, *12*(12), 2478–2492. doi: 10.1038/nprot.2017.124.

Ewels P, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Ulysse-Garcia M, Di Tommaso P, Nahnsen S (2023). The nf-core framework for community-curated bioinformatics pipelines. *Zenodo*. doi: 10.5281/zenodo.3240506.

Figueras A, Robledo D, Corvelo A, Hermida M, Pereiro P, Rubiolo JA, Gómez-Garrido J, Carreté L, Bello X, Gut M, Gut IG, Marcet-Houben M, Forn-Cuní G, Galán B, García JL, Abal-Fabeiro JL, Pardo BG, Taboada X, Fernández C, … Martínez P (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; *Pleuronectiformes*): a fish adapted to demersal life. *DNA Research*, 23(3), 181–192. doi: 10.1093/dnares/dsw007.

Fukushima HS, Takeda H, Nakamura R (2023). Incomplete erasure of histone marks during epigenetic reprogramming in medaka early development. *Genome Research*, *33*(4), 572–586. doi: 10.1101/gr.277577.122.

Fulton T, Trivedi V, Attardi A, Anlas K, Dingare C, Arias AM, Steventon B (2020). Axis Specification in Zebrafish Is Robust to Cell Mixing and Reveals a Regulation of Pattern Formation by Morphogenesis. *Current Biology*, *30*(15), 2984-2994.e3. doi 10.1016/j.cub.2020.05.048.

Gao Y, Wang Q, Liu Y, Ma Y, Jin H, Liu J, Wang H, Yan Y, Li J (2023). Epidemiol ogy of turbot bacterial diseases in China between October 2016 and December 2019. Frontiers in Marine Science, 10. doi: 10.3389/fmars.2023.1145083.

Ge SX, Jung D, Yao R (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics, 36(8), 2628–2629. doi: 10.1093/bioinformatics/btz931.

Geiman TM, Muegge K (2010). DNA methylation in early development. *Molecular Reproduction and Development*, *77*(2), 105–113. doi: 10.1002/mrd.21118.

Gillard GB, Grønvold L, Røsæg LL, Holen MM, Monsen Ø, Koop BF, Rondeau E. B, Gundappa MK, Mendoza J, Macqueen DJ, Rohlfs R, Sandve SR, Hvidsten TR (2021). Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biology*, *22*(1), 103. doi: 10.1186/s13059-021-02323-0.

Glancy E, Choy N, Eckersley-Maslin MA (2024). Bivalent chromatin: a developmental balancing act tipped in cancer. *Biochemical Society Transactions*, *52*(1), 217–229. doi: 10.1042/BST20230426.

Guerrero-Peña L, Suarez-Bregua P, Gil-Gálvez A, Naranjo S, Méndez-Martínez L, Tur R, García-Fernández P, Tena JJ, Rotllant J (2023). Genome-wide chromatin accessibility and gene expression profiling during flatfish metamorphosis. *Scientific Data*, *10*(1), 196. doi: 10.1038/s41597-023-02111-4.

Harvey TN, Gillard GB, Røsæg LL, Grammes F, Monsen Ø, Vik JO, Hvidsten TR, Sandve SR (2024). The genome regulatory landscape of Atlantic salmon liver through smoltification. *PLOS ONE*, *19*(4), e0302388. doi: 10.1371/journal.pone.0302388.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–318. doi: 10.1038/ng1966.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell, 38(4), 576–589. doi: 10.1016/j.molcel.2010.05.004.

Herrera-Uribe J, Liu H, Byrne KA, Bond ZF, Loving CL, Tuggle CK (2020). Changes in H3K27ac at Gene Regulatory Regions in Porcine Alveolar Macrophages Following LPS or PolyIC Exposure. *Frontiers in Genetics*, *11*. doi: 10.3389/fgene.2020.00817.

Hu W, Cao Y, Liu Q, Yuan C, Hu Z (2024). Effect of salinity on the physiological response and transcriptome of spotted seabass (*Lateolabrax maculatus*). *Marine Pollution Bulletin*, *203*, 116432. doi: 10.1016/j.marpolbul.2024.116432.

Hunter MP, Prince VE (2002). Zebrafish Hox Paralogue Group 2 Genes Function Redundantly as Selector Genes to Pattern the Second Pharyngeal Arch. *Developmental Biology*, *247*(2), 367–389. doi: 10.1006/dbio.2002.0701.

Hyun K, Jeon J, Park K, Kim J (2017). Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*, *49*(4), e324–e324. doi: 10.1038/emm.2017.11.

seqJimenez-Gonzalez A, Baranasic D, Müller F (2023). Zebrafish regulatory genomic resources for disease modelling and regeneration. *Disease Models & Mechanisms*, *16*(8). doi: 10.1242/dmm.050280.

Johnston IA, Kent MP, Boudinot P, Looseley M, Bargelloni L, Faggion S, Merino GA, Ilsley GR, Bobe J, Tsigenopoulos CS, Robertson J, Harrison PW, Martinez P, Robledo D, Macqueen DJ, Lien S (2024). Advancing fish breeding in aquaculture through genome functional annotation. *Aquaculture*, *583*, 740589. doi: 10.1016/j.aquaculture.2024.740589.

Jones A (1972). Studies on egg development and larval rearing of turbot, *Scophthalmus Maximus* L., and Brill, *Scophthalmus Rhombus* L., in the laboratory. *Journal of the Marine Biological Association of the United Kingdom*, *52*(4), 965–986. doi: 10.1017/S0025315400040698.

Kumar Behera B, Nayak C, Kumar Rout A, Priyambada Pradhan S, Kumar Parida P, Jyoti Sarkar D, Kumar Das B, Rai A (2024). Transcriptome profiling of Nile tilapia (*Oreochromis niloticus*) identifies candidate genes in response to riverine pollution. *Current Research in Biotechnology*, *7*, 100180. doi: 10.1016/j.crbiot.2024.100180.

Kuraku S, Meyer A (2009). The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *The International Journal of Developmental Biology*, *53*(5–6), 765–773. doi: 10.1387/ijdb.072533km.

seqLange M, Granados A, VijayKumar S, Bragantini J, Ancheta S, Santhosh S, Borja M, Kobayashi H, McGeever E, Solak AC, Yang B, Zhao X, Liu Y, Detweiler AM, Paul S, Mekonen H, Lao T, Banks R, Kim YJ, … Royer LA (2023). *Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State-Transition Dynamics of Late-Vertebrate Pluripotent Axial Progenitors*. doi: 10.1101/2023.03.06.531398.

Lawson ND, Li R, Shin M, Grosse A, Yukselen O, Stone OA, Kucukural A, Zhu L (2020). An improved zebrafish transcriptome annotation for sensitive and comprehensive detection of cell type-specific genes. *ELife*, *9*. doi: 10.7554/eLife.55792.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760. doi: 10.1093/bioinformatics/btp324.

Li B, Dewey CN (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics, 12(1), 323. doi: 10.1186/1471-2105-12-323.

Li Y, Liu Y, Yang H, Zhang T, Naruse K, Tu Q (2020). Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis. *Genome Research*, *30*(6), 924–937. doi: 10.1101/gr.258871.119.

Lin X, Yang X, Chen C, Ma W, Wang Y, Li X, Zhao K, Deng Q, Feng W, Ma Y, Wang H, Zhu L, Sahu SK, Chen F, Zhang X, Dong Z, Liu C, Liu L, Liu C (2023). Single-nucleus chromatin landscapes during zebrafish early embryogenesis. *Scientific Data*, *10*(1), 464. doi: 10.1038/s41597-023-02373-y.

Liu G, Wang W, Hu S, Wang X, Zhang Y (2019). Corrigendum: Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. *Genome Research*, *29*(3), 520–520. doi: 10.1101/gr.248732.119.

Liu J, Xiao Q, Xiao J, Niu C, Li Y, Zhang X, Zhou Z, Shu G, Yin G (2022). Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduction and Targeted Therapy*, *7*(1), 3. doi: 10.1038/s41392-021-00762-6.

Liu J, Castillo-Hair SM, Du LY, Wang Y, Carte AN, Colomer-Rosell M, Yin C, Seelig G, Schier AF (2024). Dissecting the regulatory logic of specification and differentiation during vertebrate embryogenesis. *bioRxiv (Cold Spring Harbor Laboratory)*. doi 10.1101/2024.08.27.609971.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550. doi: 10.1186/s13059-014-0550-8.

Ma H, Martin K, Dixon D, Hernandez AG, Weber GM (2019). Transcriptome analysis of egg viability in rainbow trout, *Oncorhynchus mykiss*. *BMC Genomics*, *20*(1), 319. doi: 10.1186/s12864-019-5690-5.

Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, Law T, Lareau C, Hsu YC, Regev A, Buenrostro JD (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell*, *183*(4), 1103-1116.e20. doi: 10.1016/j.cell.2020.09.056.

Maroso F, Hermida M, Millán A, Blanco A, Saura M, Fernández A, Dalla-Rovere G, Bargelloni L, Cabaleiro S, Villanueva B, Bouza C, Martínez P (2018). Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. *DNA Research*, 25(4), 439–450. doi: 10.1093/dnares/dsy015.

Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17(1), 10. doi: 10.14806/ej.17.1.200.

Martínez P, Robledo D, Taboada X, Blanco A, Moser M, Maroso F, Hermida M, Gómez-Tato A, Álvarez-Blázquez B, Cabaleiro S, Piferrer F, Bouza C, Lien S, Viñas AM (2021). A genome-wide association study, supported by a new chromosome-level genome assembly, suggests sox2 as a main driver of the undifferentiatiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). *Genomics*, 113(4), 1705–1718. doi: 10.1016/j.ygeno.2021.04.007.

McClintock JM, Carlson R, Mann DM, Prince VE (2001). Consequences of Hox gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes. *Development*, *128*(13), 2471–2484. doi: 10.1242/dev.128.13.2471.

Miao Y, Pourquié O (2024). Cellular and molecular control of vertebrate somitogenesis. *Nature Reviews Molecular Cell Biology*, *25*(7), 517–533. doi: 10.1038/s41580-024-00709-z.

Minnoye L, Marinov GK, Krausgruber T, Pan L, Marand AP, Secchia S, Greenleaf WJ, Furlong EEM, Zhao K, Schmitz RJ, Bock C, Aerts S (2021). Chromatin accessibility profiling methods. Nature Reviews Methods Primers, 1(1), 10. doi: 10.1038/s43586-020-00008-9.

Miranda S, Santos L, Chaves S, Lima B, Rodrigues J, Rosa-Silva M, Tercya H, Jesus P, Albuquerque E, Maximino C, Siqueira-Silva D (2023). The effects of water temperature and hybridization on embryonic development and gametogenesis of two species of Amazonian tetra. *Theriogenology Wild*, *3*, 100051. doi: 10.1016/j.therwi.2023.100051.

Mircea M, Semrau S (2021). How a cell decides its own fate: a single-cell view of molecular mechanisms and dynamics of cell-type specification. *Biochemical Society Transactions*, *49*(6), 2509–2525. doi: 10.1042/BST20210135.

Moreno-Oñate M, Gallardo-Fuentes L, Martínez-García PM, Naranjo S, Jiménez-Gancedo S, Tena JJ, Santos-Pereira JM (2024). Rewiring of the epigenome and chromatin architecture by exogenously induced retinoic acid signaling during zebrafish embryonic development. *Nucleic Acids Research*, *52*(7), 3682–3701. doi: 10.1093/nar/gkae065.

Ohno S (1970). *Evolution by Gene Duplication*. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-86659-3.

Ozernyuk N, Schepetov D (2022). HOX-Gene Cluster Organization and Genome Duplications in Fishes and Mammals: Transcript Variant Distribution along the Anterior–Posterior Axis. *International Journal of Molecular Sciences*, *23*(17), 9990. doi: 10.3390/ijms23179990.

Pálfy M, Schulze G, Valen E, Vastenhouw NL (2020). Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLOS Genetics*, *16*(1), e1008546. doi: 10.1371/journal.pgen.1008546.

Pan Z, Wang Y, Wang M, Wang Y, Zhu X, Gu S, Zhong C, An L, Shan M, Damas J, Halstead MM, Guan D, Trakooljul N, Wimmers K, Bi Y, Wu S, Delany ME, Bai X, Cheng HH, … Zhou H (2023). An atlas of regulatory elements in chicken: A resource for chicken genetics and genomics. *Science Advances*, 9(18). doi: 10.1126/sciadv.ade1204.

Papadaki M, Kaitetzidou E, Papadakis IE, Sfakianakis DG, Papandroulakis N, Mylonas CC, Sarropoulou E (2022). Temperature-Biased miRNA Expression Patterns during European Sea Bass (*Dicentrarchus labrax*) Development. *International Journal of Molecular Sciences*, *23*(19), 11164. doi: 10.3390/ijms231911164.

Pasquier J, Braasch I, Batzel P, Cabau C, Montfort J, Nguyen T, Jouanno E, Berthelot C, Klopp C, Journot L, Postlethwait JH, Guiguen Y, Bobe J (2017). Evolution of gene expression after whole-genome duplication: New insights from the spotted gar genome. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *328*(7), 709–721. doi: 10.1002/jez.b.22770.

Patel H, Ewels P, Peltzer A, Botvinnik O, Sturm G, Moreno D, Vemuri P, Garcia MU, Morins S, Pantano L, Binzer-Panchal M, Syme R, Zepper M, Kelly G, Hanssen F, Fellows Yates JA, Cheshire C, Espinosa-Carrasco J, Miller E, Talbot A, Zhou P, Guinchard S, Hörtenhuber M, Gabernet G, Mertes C, Straub D, Di Tommaso P (2023). nf-core/rnaseq: nf-core/rnaseq v3.10.1 – Plastered Rhodium Rudolph. *Zenodo*. doi: 10.5281/zenodo.7505987.

Paul D, Bridoux L, Rezsöhazy R, Donnay I (2011). *HOX* genes are expressed in bovine and mouse oocytes and early embryos. *Molecular Reproduction and Development*, *78*(6), 436–449. doi: 10.1002/mrd.21321.

Petratou K, Stehling M, Müller F, Schulte-Merker S (2024). Integration of ATAC and RNA-sequencing identifies chromatin and transcriptomic signatures in

classical and non-classical zebrafish osteoblasts and indicates mechanisms of *entpd5a* regulation. *eLife* 13:RP100230. doi: 10.7554/eLife.100230.1.

Pohl J (2019). Zebrafish (*Danio rerio*) embryo-larvae locomotor activity data analysis: Evaluating anxiolytic effects of the antidepressant compound citalopram. *Data in Brief*, *27*, 104812. doi: 10.1016/j.dib.2019.104812.

Powell D, Thoa NP, Nguyen NH, Knibb W, Elizur A (2021). Transcriptomic responses of saline-adapted Nile tilapia (*Oreochromis niloticus*) to rearing in both saline and freshwater. *Marine Genomics*, 60, 100879. doi: 10.1016/j.margen.2021.100879.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Research, 42(W1), W187–W191. doi: 10.1093/nar/gku365.

Robledo D, Martin AP, Álvarez-Dios JA, Bouza C, Pardo BG, Martínez P (2017). First characterization and validation of turbot microRNAs. *Aquaculture*, 472, 76–83. do: 10.1016/j.aquaculture.2016.05.002.

Schröter C, Herrgen L, Cardona A, Brouhard GJ, Feldman B, Oates AC (2008). Dynamics of zebrafish somitogenesis. *Developmental Dynamics*, *237*(3), 545–553. doi: 10.1002/dvdy.21458.

Schulz KN, Harrison MM (2019). Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics*, *20*(4), 221–234. doi: 10.1038/s41576-018-0087-x.

Sinha KK, Bilokapic S, Du Y, Malik D, Halic M (2023). Histone modifications regulate pioneer transcription factor cooperativity. *Nature*, *619*(7969), 378–384. doi:10.1038/s41586-023-06112-6.

Skjærven KH, Alix M, Kleppe L, Fernandes JMO, Whatmore P, Nedoluzhko A, Andersson E, Kjesbu OS (2024). Ocean warming shapes embryonic developmental prospects of the next generation in Atlantic cod. *ICES Journal of Marine Science*, *81*(4), 733–747. doi: 10.1093/icesjms/fsae025.

Stark R, Brown G (2011). DiffBind: differential binding analysis of ChIP-seq peak data. Bioconductor. doi: 10.18129/B9.bioc.DiffBind.

Summers KM, Bush SJ, Wu C, Su AI, Muriuki C, Clark EL, Finlayson HA, Eory L, Waddell LA, Talbot R, Archibald AL, Hume DA (2020). Functional Annotation of the Transcriptome of the Pig, *Sus scrofa*, Based Upon Network Analysis of an RNAseq Transcriptional Atlas. Frontiers in Genetics, 10. doi: 10.3389/fgene.2019.01355.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, *6*(7), e21800. doi: 10.1371/journal.pone.0021800.

Theocharidis A, van Dongen S, Enright AJ, Freeman TC (2009). Network visualization and analysis of gene expression data using BioLayout Express3D. *Nature Protocols*, *4*(10), 1535–1550. doi: 10.1038/nprot.2009.177.

van der Vaart M, Spaink HP, Meijer AH (2012). Pathogen Recognition and Activation of the Innate Immune Response in Zebrafish. *Advances in Hematology*, *2012*, 1–19. doi: 10.1155/2012/159807.

van Mierlo G, Pushkarev O, Kribelbauer JF, Deplancke B (2023). Chromatin modules and their implication in genomic organization and gene regulation. *Trends in Genetics*, *39*(2), 140–153. doi: 10.1016/j.tig.2022.11.003.

Voss AK, Strasser A (2020). The essentials of developmental apoptosis. *F1000Research*, *9*, 148. doi: 10.12688/f1000research.21571.1.

Villarreal F, Burguener GF, Sosa EJ, Stocchi N, Somoza GM, Turjanski AG, Blanco A, Viñas J, Mechaly AS (2024). Genome sequencing and analysis of black flounder (*Paralichthys orbignyanus*) reveals new insights into Pleuronectiformes genomic size and structure. *BMC Genomics*, *25*(1), 297. doi: 10.1186/s12864-024-10081-z

Vu H, Ernst J (2022). Universal annotation of the human genome through integration of over a thousand epigenomic datasets. Genome Biology, 23(1), 9. doi: 10.1186/s13059-021-02572-z.

Wang H, Wang B, Liu X, Liu Y, Du X, Zhang Q, Wang X (2017). Identification and expression of piwil2 in turbot *Scophthalmus maximus*, with implications of the involvement in embryonic and gonadal development. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *208–209*, 84–93. doi: 10.1016/j.cbpb.2017.04.007

White RJ, Collins JE, Sealy IM, Wali N, Dooley CM, Digby Z, Stemple DL, Murphy DN, Billis K, Hourlier T, Füllgrabe A, Davis MP, Enright AJ, Busch-Nentwich EM (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. *ELife*, *6*. doi: 10.7554/eLife.30860.

Wike CL, Guo Y, Tan M, Nakamura R, Shaw DK, Díaz N, Whittaker-Tademy AF, Durand NC, Aiden EL, Vaquerizas JM, Grunwald D, Takeda H, Cairns BR (2021). Chromatin architecture transitions from zebrafish sperm through early embryogenesis. *Genome Research*, *31*(6), 981–994. doi: 10.1101/gr.269860.120.

Wu L, Li J, Liu F, Song Z, Song C, Xu S, Yue X, Li X (2024). Comparative transcriptome analyses reveal changes of gene expression in turbot (*Scophthalmus maximus*) embryos exposed to different LED light spectra and potential photosensitive function of non-visual opsins. *Aquaculture Reports*, *35*, 101948. doi: 10.1016/j.aqrep.2024.101948.

Xu X, Zheng W, Meng Z, Xu W, Liu Y, Chen S (2022). Identification of stress-related genes by co-expression network analysis based on the improved turbot genome. *Scientific Data*, *9*(1), 374. doi: 10.1038/s41597-022-01458-4.

Xu Q, Zhang Y, Xu W, Liu D, Jin W, Chen X, Hong N (2024). The chromatin accessibility dynamics during cell fate specifications in zebrafish early embryogenesis. *Nucleic Acids Research*, *52*(6), 3106–3120. doi: 10.1093/nar/gkae095.

Yang Q, Yang XD, Liu MQ, Zeng C, Zhao HK, Xiang KW, Hou ZS, Wen HS, Li JF (2023). Transcriptome analysis of liver, gill and intestine in rainbow trout (*Oncorhynchus mykiss*) symptomatically or asymptomatically infected with *Vibrio anguillarum*. *Fish & Shellfish Immunology*, *135*, 108643. doi: 10.1016/j.fsi.2023.108643.

Yu Y, Li X, Jiao R, Lu Y, Jiang X, Li X (2023). H3K27me3-H3K4me1 transition at bivalent promoters instructs lineage specification in development. *Cell & Bioscience*, *13*(1), 66. doi: 10.1186/s13578-023-01017-3.

Zernicka-Goetz M (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development*, *129*(4), 815–829. doi: 10.1242/dev.129.4.815.

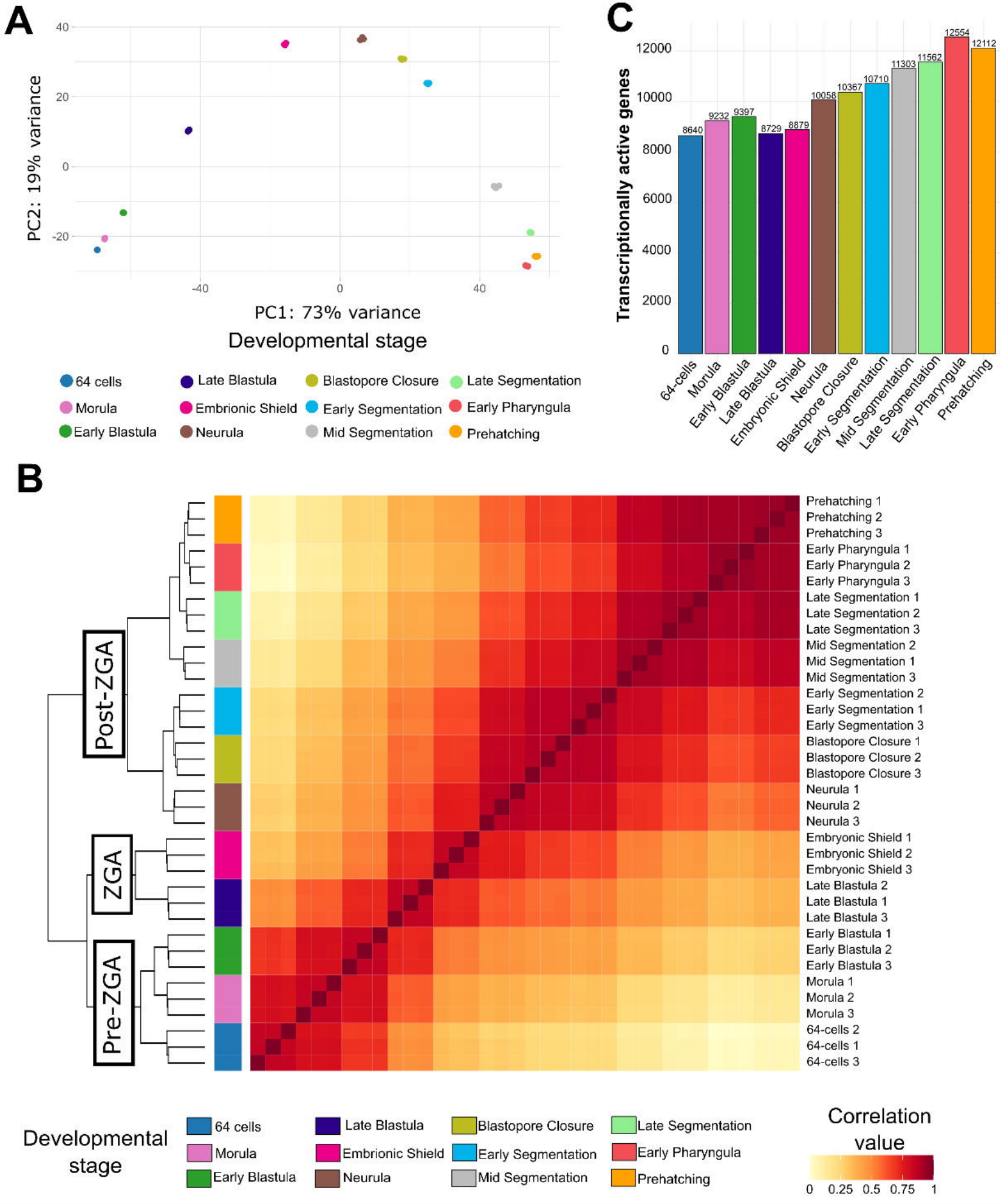
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008). Model-based Analysis of ChIP-seq (MACS). *Genome Biology*, *9*(9), R137. doi: 10.1186/gb-2008-9-9-r137.

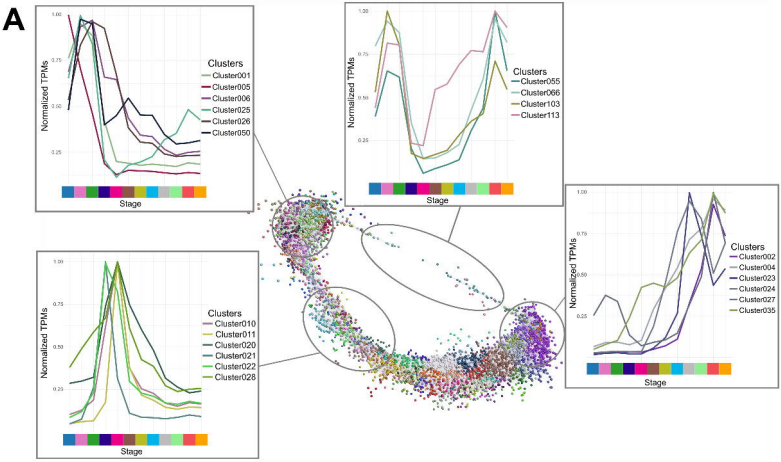
Zheng H, Huang B, Zhang B, Xiang Y, Du Z, Xu Q, Li Y, Wang Q, Ma J, Peng X, Xu F, Xie W (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell*, *63*(6), 1066–1079. doi: 10.1016/j.molcel.2016.08.032.

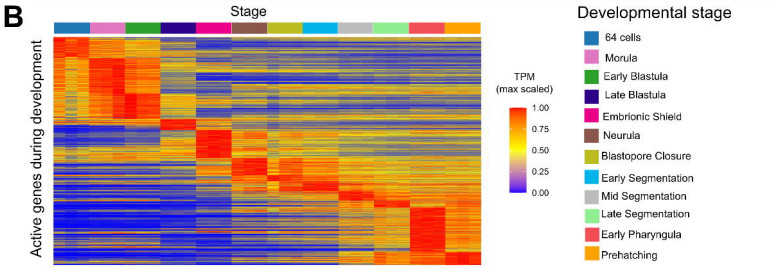
Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications, 10(1), 1523. doi: 10.1038/s41467-019-09234-6.

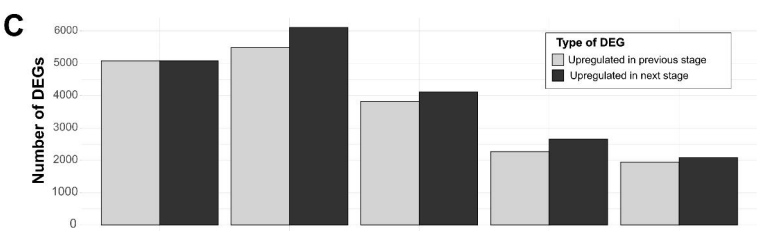
Zhou CY, Heald R (2023). Principles of genome activation in the early embryo. *Current Opinion in Genetics & Development*, *81*, 102062. doi: 10.1016/j.gde.2023.102062.

Colour codes												
Stage name	64-cells	Morula	Early Blastula	Late Blastula	Shield	Neurula	Blastopore closure	Early Seg.	Mid Seg.	Late Seg.	Early Pharyngula	Prehatch
Hours post-fertilization	6 hpf	9 hpf	10 hpf	12 hpf	24 hpf	30 hpf	35 hpf	40 hpf	57 hpf	68 hpf	73 hpf	90 hpf
Embryos per pool	150	150	150	100	100	100	100	80	80	80	50	50
RNA-seq replicates	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3
ATAC-seq replicates	-	-	-	3/3	3/3	-	-	3/3	3/3	3/3	-	3/3
H3K4me3 ChIP-seq replicates	-	-	-	3/3	3/3	-	-	3/3	2/3	3/3	-	3/3
H3K27ac ChIP-seq replicates	-	-	-	3/3	3/3	-	-	3/3	2/3	2/3	-	1/3
H3K27me3 ChIP-seq replicates	-	-	-	3/3	3/3	-	-	3/3	1/3	3/3	-	2/3

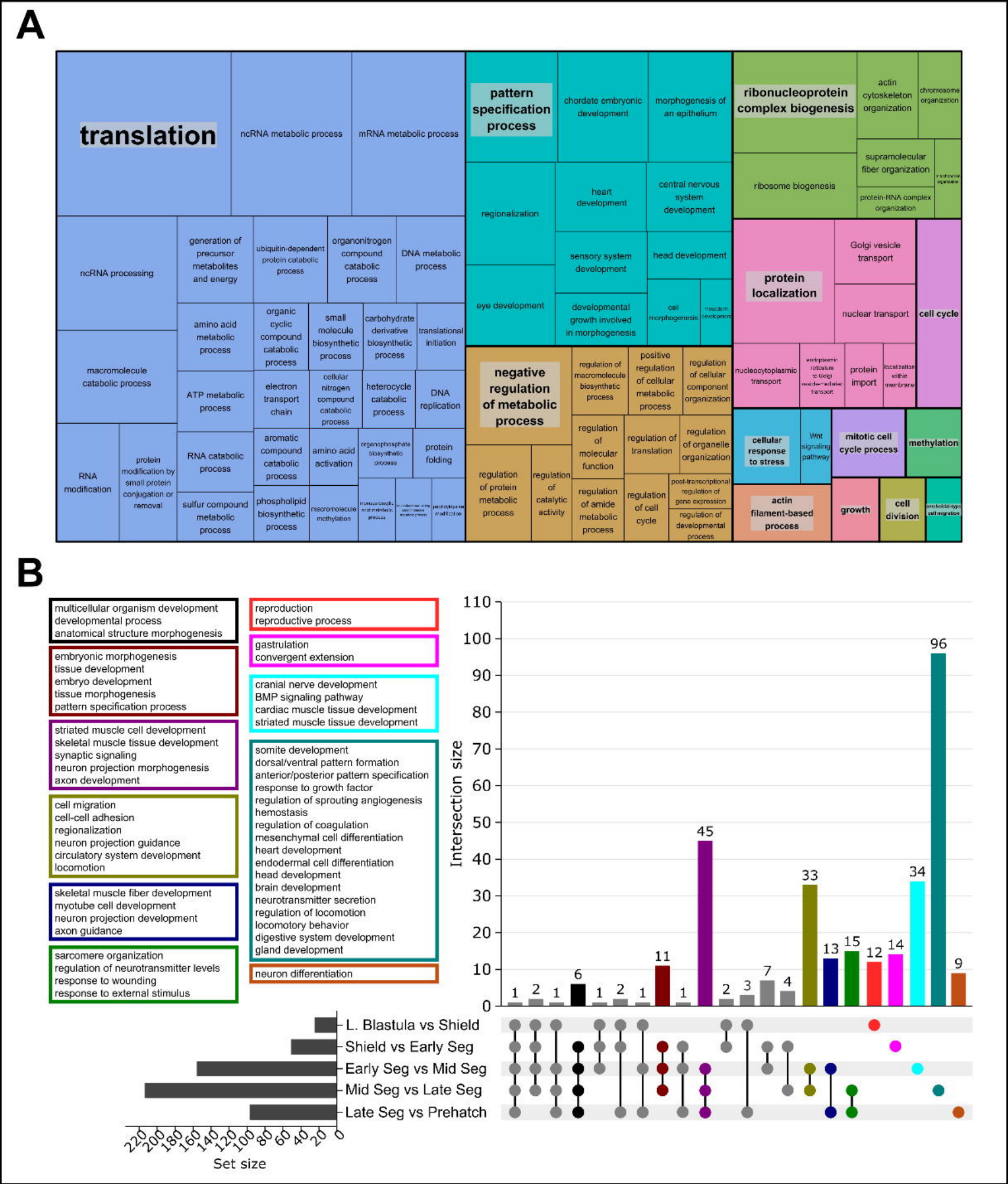

A
PC2: 19% variance
PC1: 73% variance
Developmental stage
64 cells
Morula
Early Blastula
Late Blastula
Embrionic Shield
Neurula
Blastopore Closure
Early Segmentation
Mid Segmentation
Late Segmentation
Early Pharyngula
Prehatching
C
Transcriptionally active genes
8640
9232
9397
8729
8879
10058
10367
10710
11303
11562
12554
12112
64-cells
Morula
Early Blastula
Late Blastula
Embryonic Shield
Neurula
Blastopore Closure
Early Segmentation
Mid Segmentation
Late Segmentation
Early Pharyngula
Prehatching
B
Post-ZGA
ZGA
Pre-ZGA
Prehatching 1
Prehatching 2
Prehatching 3
Early Pharyngula 1
Early Pharyngula 2
Early Pharyngula 3
Late Segmentation 1
Late Segmentation 2
Late Segmentation 3
Mid Segmentation 2
Mid Segmentation 1
Mid Segmentation 3
Early Segmentation 2
Early Segmentation 1
Early Segmentation 3
Blastopore Closure 1
Blastopore Closure 2
Blastopore Closure 3
Neurula 1
Neurula 2
Neurula 3
Embryonic Shield 1
Embryonic Shield 2
Embryonic Shield 3
Late Blastula 2
Late Blastula 1
Late Blastula 3
Early Blastula 1
Early Blastula 2
Early Blastula 3
Morula 1
Morula 2
Morula 3
64-cells 2
64-cells 1
64-cells 3
Developmental stage
64 cells
Morula
Early Blastula
Late Blastula
Embrionic Shield
Neurula
Blastopore Closure
Early Segmentation
Mid Segmentation
Late Segmentation
Early Pharyngula
Prehatching
Correlation value
0
0.25
0.5
0.75
1

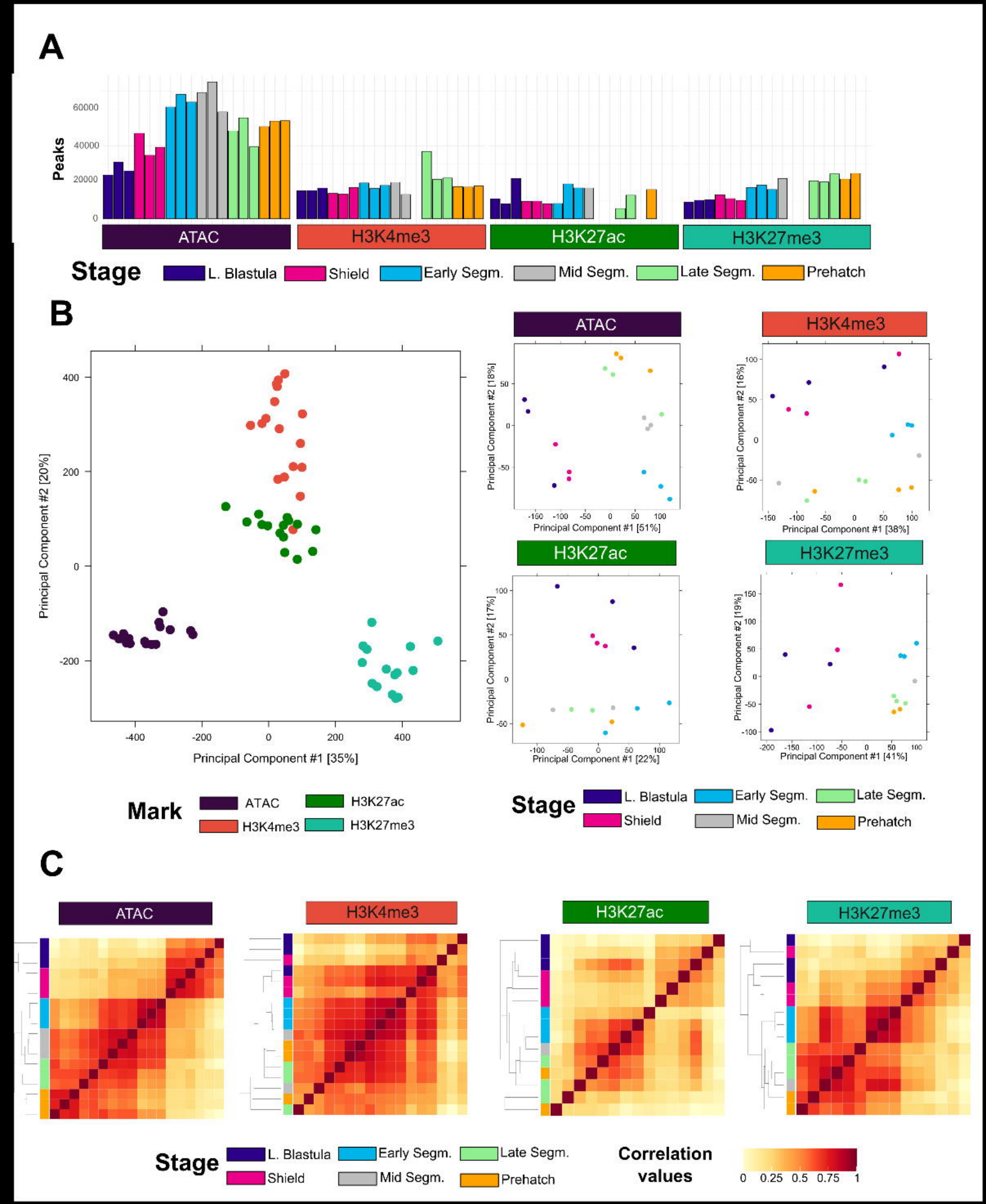


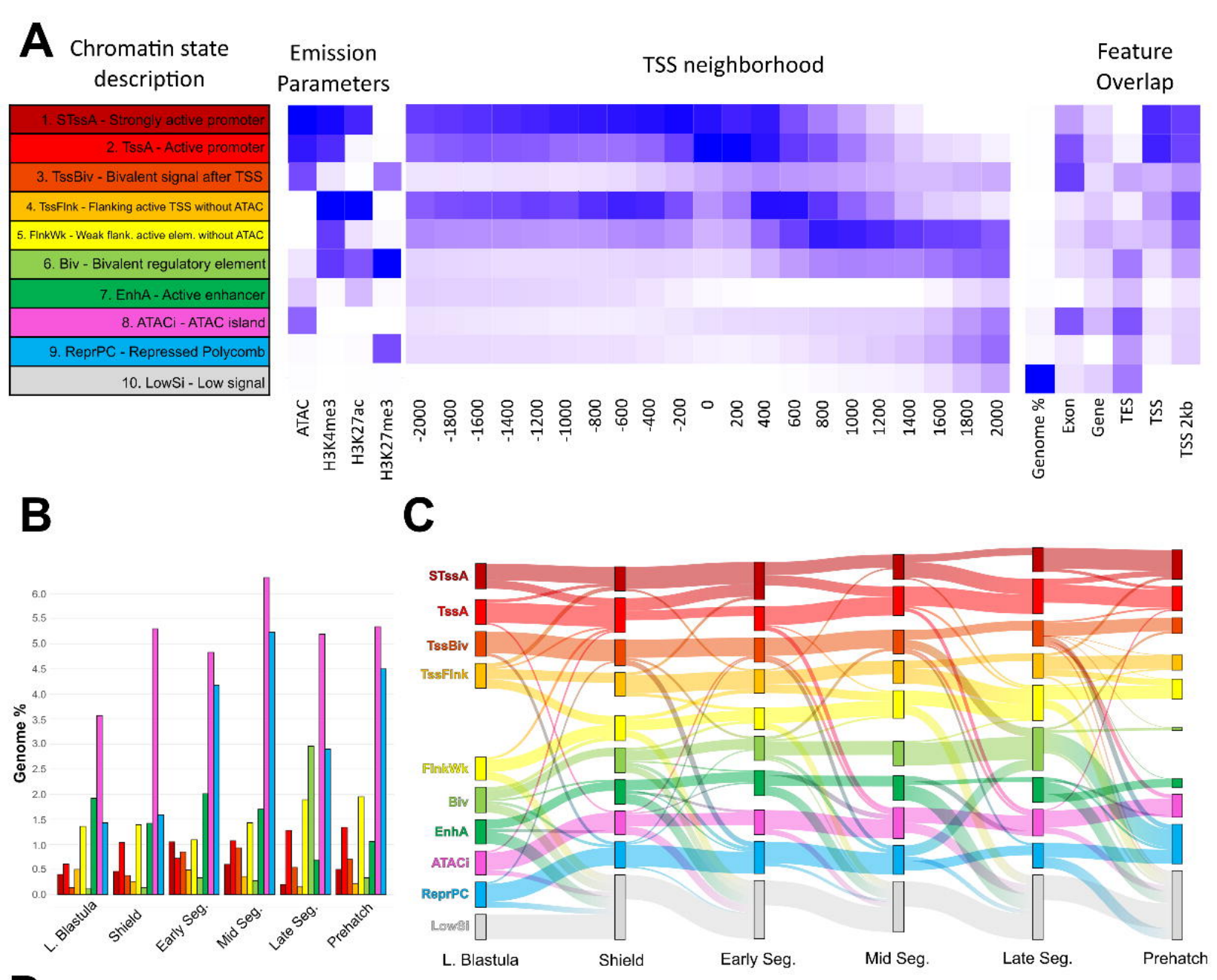


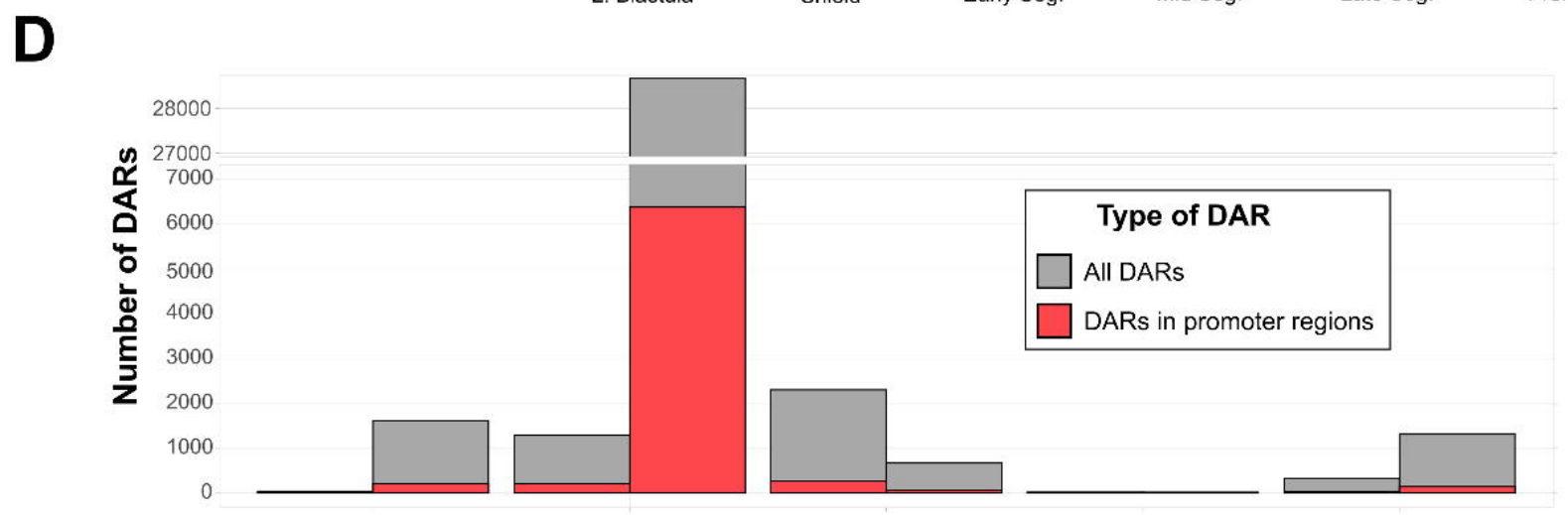


	L. Blastula vs Shield		Shield vs Early Seg		Early Seg vs Mid Seg		Mid Seg vs Late Seg		Late Seg vs Prehatch	
All DEG	5079	5080	5484	6109	3827	4113	2272	2659	1945	2085

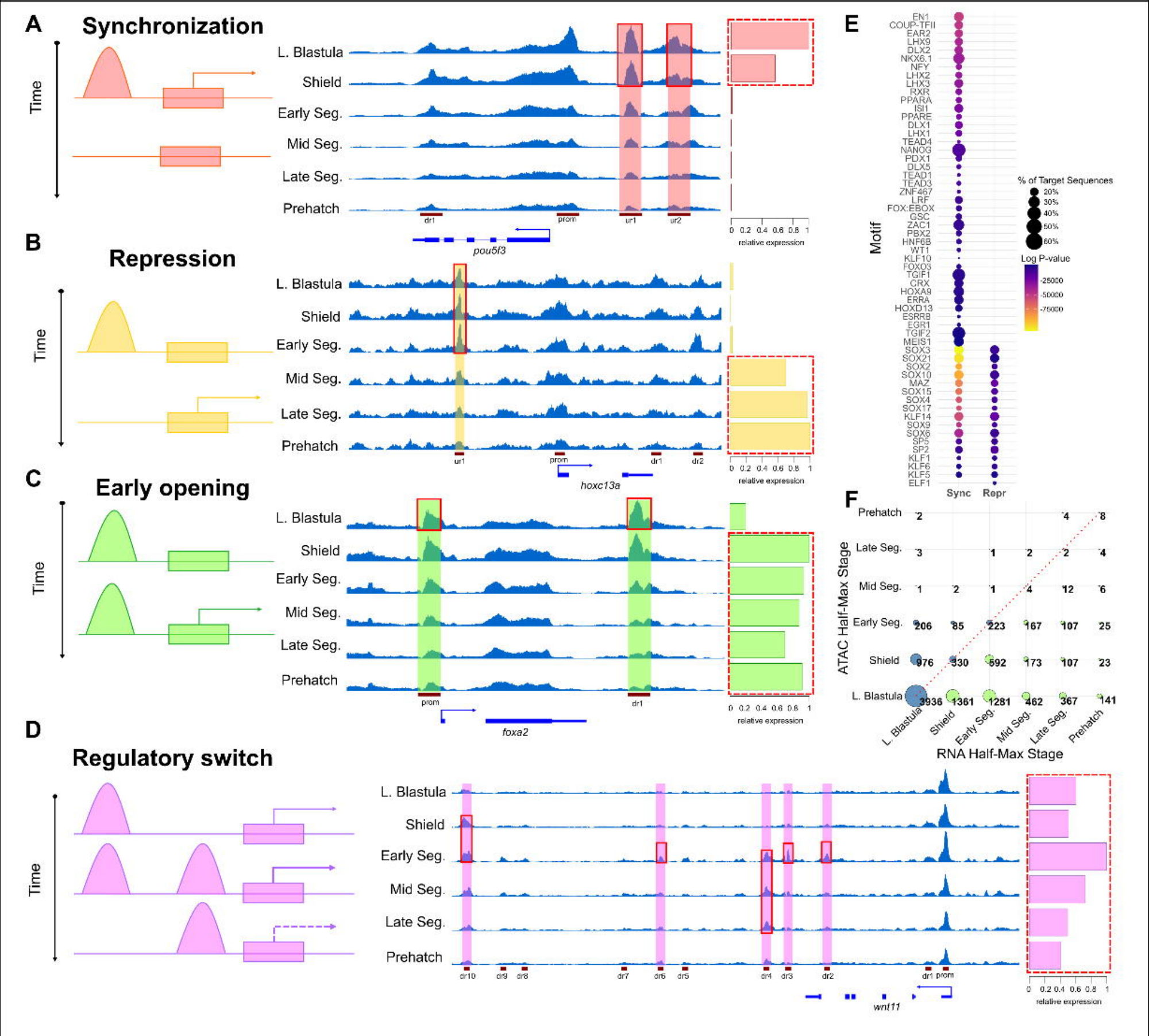

A
translation
ncRNA metabolic process
mRNA metabolic process
ncRNA processing
generation of precursor metabolites and energy
ubiquitin-dependent protein catabolic process
organonitrogen compound catabolic process
DNA metabolic process
amino acid metabolic process
organic cyclic compound catabolic process
small molecule biosynthetic process
carbohydrate derivative biosynthetic process
translational initiation
macromolecule catabolic process
ATP metabolic process
electron transport chain
cellular nitrogen compound catabolic process
heterocycle catabolic process
DNA replication
RNA modification
protein modification by small protein conjugation or removal
RNA catabolic process
aromatic compound catabolic process
amino acid activation
organophosphate biosynthetic process
protein folding
sulfur compound metabolic process
phospholipid biosynthetic process
macromolecule methylation
pattern specification process
chordate embryonic development
morphogenesis of an epithelium
regionalization
heart development
central nervous system development
sensory system development
head development
eye development
developmental growth involved in morphogenesis
cell morphogenesis
mesoderm development
negative regulation of metabolic process
regulation of macromolecule biosynthetic process
positive regulation of cellular metabolic process
regulation of cellular component organization
regulation of molecular function
regulation of translation
regulation of organelle organization
regulation of protein metabolic process
regulation of catalytic activity
regulation of amide metabolic process
regulation of cell cycle
post-transcriptional regulation of gene expression
regulation of developmental process
ribonucleoprotein complex biogenesis
actin cytoskeleton organization
chromosome organization
supramolecular fiber organization
ribosome biogenesis
protein-RNA complex organization
protein localization
Golgi vesicle transport
nuclear transport
cell cycle
nucleocytoplasmic transport
endoplasmic reticulum to Golgi vesicle-mediated transport
protein import
localization within membrane
cellular response to stress
Wnt signaling pathway
mitotic cell cycle process
methylation
actin filament-based process
growth
cell division
B
multicellular organism development
developmental process
anatomical structure morphogenesis
embryonic morphogenesis
tissue development
embryo development
tissue morphogenesis
pattern specification process
striated muscle cell development
skeletal muscle tissue development
synaptic signaling
neuron projection morphogenesis
axon development
cell migration
cell-cell adhesion
regionalization
neuron projection guidance
circulatory system development
locomotion
skeletal muscle fiber development
myotube cell development
neuron projection development
axon guidance
sarcomere organization
regulation of neurotransmitter levels
response to wounding
response to external stimulus
reproduction
reproductive process
gastrulation
convergent extension
cranial nerve development
BMP signaling pathway
cardiac muscle tissue development
striated muscle tissue development
somite development
dorsal/ventral pattern formation
anterior/posterior pattern specification
response to growth factor
regulation of sprouting angiogenesis
hemostasis
regulation of coagulation
mesenchymal cell differentiation
heart development
endodermal cell differentiation
head development
brain development
neurotransmitter secretion
regulation of locomotion
locomotory behavior
digestive system development
gland development
neuron differentiation
Intersection size
110
100
90
80
70
60
50
40
30
20
10
0
1
2
1
6
1
2
1
11
1
45
2
3
7
4
33
13
15
12
14
34
96
9
L. Blastula vs Shield
Shield vs Early Seg
Early Seg vs Mid Seg
Mid Seg vs Late Seg
Late Seg vs Prehatch
220
200
180
160
140
120
100
80
60
40
20
0
Set size

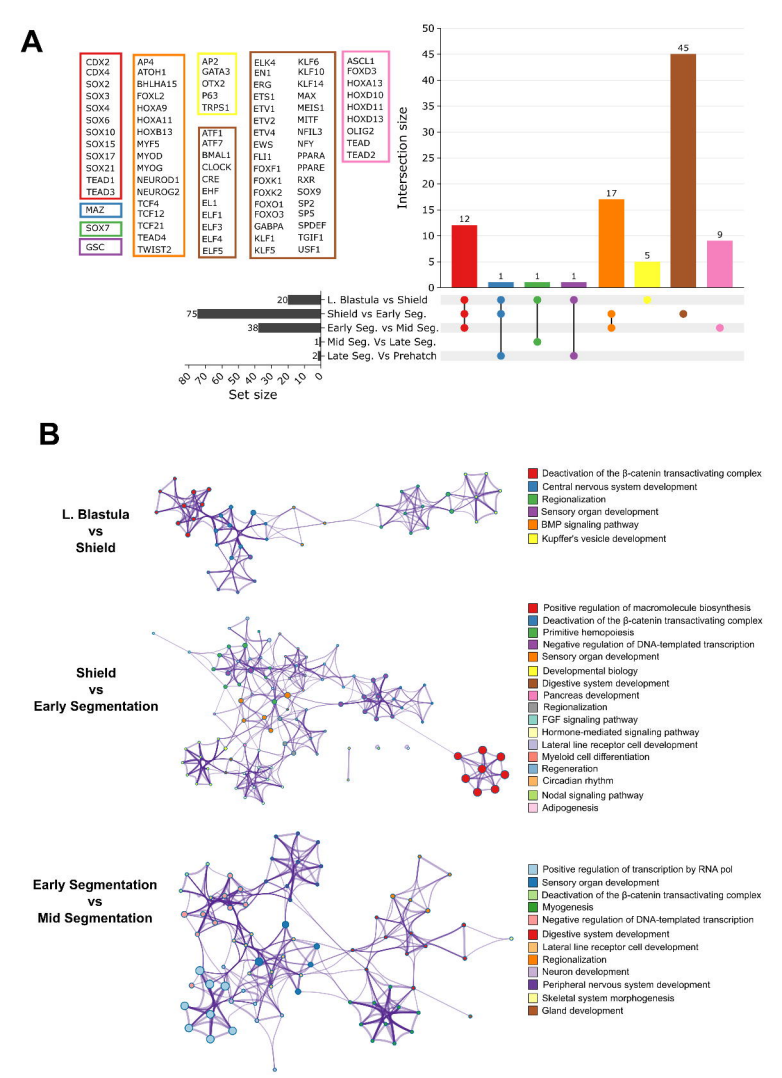

A
Peaks
60000
40000
20000
0
ATAC
H3K4me3
H3K27ac
H3K27me3
Stage
L. Blastula
Shield
Early Segm.
Mid Segm.
Late Segm.
Prehatch
B
Principal Component #2 [20%]
Principal Component #1 [35%]
ATAC
H3K4me3
H3K27ac
H3K27me3
Principal Component #2 [18%]
Principal Component #1 [51%]
Principal Component #2 [16%]
Principal Component #1 [38%]
Principal Component #2 [17%]
Principal Component #1 [22%]
Principal Component #2 [19%]
Principal Component #1 [41%]
Mark
ATAC
H3K4me3
H3K27ac
H3K27me3
Stage
L. Blastula
Shield
Early Segm.
Mid Segm.
Late Segm.
Prehatch
C
ATAC
H3K4me3
H3K27ac
H3K27me3
Stage
L. Blastula
Shield
Early Segm.
Mid Segm.
Late Segm.
Prehatch
Correlation values
0
0.25
0.5
0.75
1

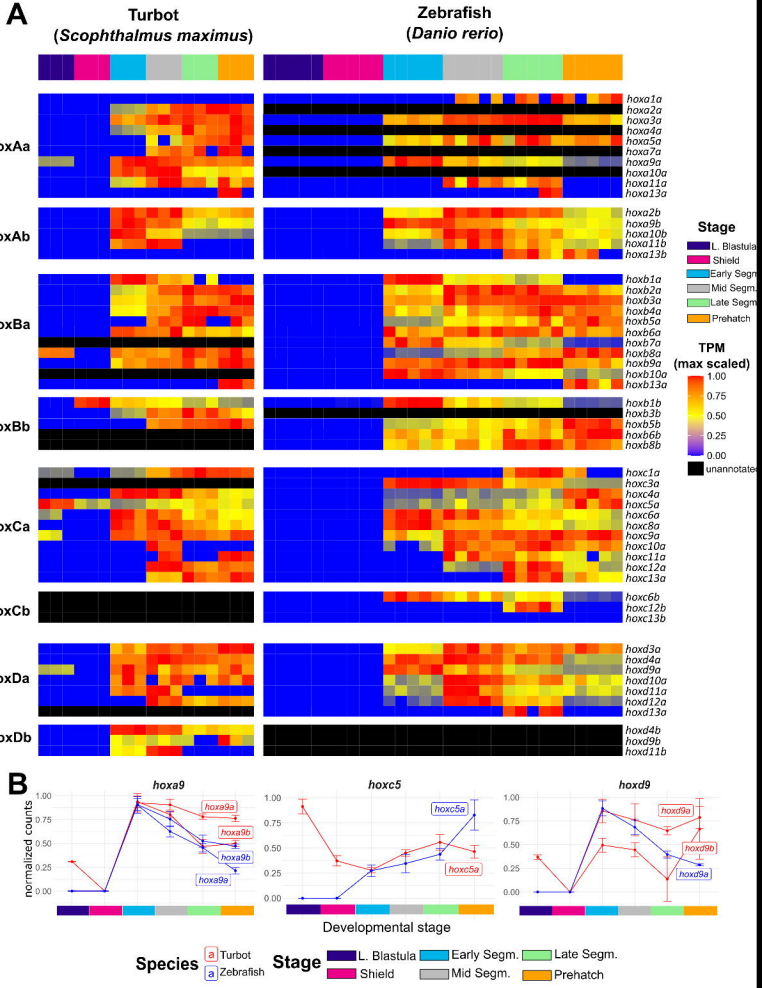




	L. Blastula vs Shield		Shield vs Early Seg		Early Seg vs Mid Seg		Mid Seg vs Late Seg		Late Seg vs Prehatch	
All DARs	30	1602	1285	28670	2299	659	8	7	315	1305
DARs in promoter	2	192	194	6385	259	52	3	1	23	130


A
Synchronization
Time
L. Blastula
Shield
Early Seg.
Mid Seg.
Late Seg.
Prehatch
dr1
prom
ur1
ur2
0 0.2 0.4 0.6 0.8 1
relative expression
pou5f3
B
Repression
Time
L. Blastula
Shield
Early Seg.
Mid Seg.
Late Seg.
Prehatch
ur1
prom
dr1
dr2
0 0.2 0.4 0.6 0.8 1
relative expression
hoxc13a
C
Early opening
Time
L. Blastula
Shield
Early Seg.
Mid Seg.
Late Seg.
Prehatch
prom
dr1
0 0.2 0.4 0.6 0.8 1
relative expression
foxa2
D
Regulatory switch
Time
L. Blastula
Shield
Early Seg.
Mid Seg.
Late Seg.
Prehatch
dr10
dr9
dr8
dr7
dr6
dr5
dr4
dr3
dr2
dr1
prom
0 0.2 0.4 0.6 0.8 1
relative expression
wnt11
E
Motif
EN1
COUP-TFII
EAR2
LHX9
DLX2
NKX6.1
NFY
LHX2
LHX3
RXR
PPARA
ISL1
PPARE
DLX1
LHX1
TEAD4
NANOG
PDX1
DLX5
TEAD1
TEAD3
ZNF467
LRF
FOX:EBOX
GSC
ZAC1
PBX2
HNF6B
WT1
KLF10
FOXO3
TGIF1
CRX
HOXA9
ERRA
HOXD13
ESRRB
EGR1
TGIF2
MEIS1
SOX3
SOX21
SOX2
SOX10
MAZ
SOX15
SOX4
SOX17
KLF14
SOX9
SOX6
SP5
SP2
KLF1
KLF6
KLF5
ELF1
Sync
Repr
% of Target Sequences
20%
30%
40%
50%
60%
Log P-value
-25000
-50000
-75000
F
ATAC Half-Max Stage
RNA Half-Max Stage
Prehatch
Late Seg.
Mid Seg.
Early Seg.
Shield
L. Blastula
2 4 8
3 1 2 2 4
1 2 1 4 12 6
206 85 223 167 107 25
976 330 592 173 107 23
3936 1361 1281 462 367 141


A
CDX2
CDX4
SOX2
SOX3
SOX4
SOX6
SOX10
SOX15
SOX17
SOX21
TEAD1
TEAD3
MAZ
SOX7
GSC
AP4
ATOH1
BHLHA15
FOXL2
HOXA9
HOXA11
HOXB13
MYF5
MYOD
MYOG
NEUROD1
NEUROG2
TCF4
TCF12
TCF21
TEAD4
TWIST2
AP2
GATA3
OTX2
P63
TRPS1
ATF1
ATF7
BMAL1
CLOCK
CRE
EHF
EL1
ELF1
ELF3
ELF4
ELF5
ELK4
EN1
ERG
ETS1
ETV1
ETV2
ETV4
EWS
FLI1
FOXF1
FOXK1
FOXK2
FOXO1
FOXO3
GABPA
KLF1
KLF5
KLF6
KLF10
KLF14
MAX
MEIS1
MITF
NFIL3
NFY
PPARA
PPARE
RXR
SOX9
SP2
SP5
SPDEF
TGIF1
USF1
ASCL1
FOXD3
HOXA13
HOXD10
HOXD11
HOXD13
OLIG2
TEAD
TEAD2
Intersection size
50
45
40
35
30
25
20
15
10
5
0
12
1
1
1
17
5
45
9
20
75
38
1
2
L. Blastula vs Shield
Shield vs Early Seg.
Early Seg. vs Mid Seg.
Mid Seg. Vs Late Seg.
Late Seg. Vs Prehatch
80
70
60
50
40
30
20
10
0
Set size
B
L. Blastula vs Shield
Deactivation of the β-catenin transactivating complex
Central nervous system development
Regionalization
Sensory organ development
BMP signaling pathway
Kupffer's vesicle development
Shield vs Early Segmentation
Positive regulation of macromolecule biosynthesis
Deactivation of the β-catenin transactivating complex
Primitive hemopoiesis
Negative regulation of DNA-templated transcription
Sensory organ development
Developmental biology
Digestive system development
Pancreas development
Regionalization
FGF signaling pathway
Hormone-mediated signaling pathway
Lateral line receptor cell development
Myeloid cell differentiation
Regeneration
Circadian rhythm
Nodal signaling pathway
Adipogenesis
Early Segmentation vs Mid Segmentation
Positive regulation of transcription by RNA pol
Sensory organ development
Deactivation of the β-catenin transactivating complex
Myogenesis
Negative regulation of DNA-templated transcription
Digestive system development
Lateral line receptor cell development
Regionalization
Neuron development
Peripheral nervous system development
Skeletal system morphogenesis
Gland development


A
Turbot
(Scophthalmus maximus)
Zebrafish
(Danio rerio)
HoxAa
HoxAb
HoxBa
HoxBb
HoxCa
HoxCb
HoxDa
HoxDb
hoxa1a
hoxa2a
hoxa3a
hoxa4a
hoxa5a
hoxa7a
hoxa9a
hoxa10a
hoxa11a
hoxa13a
hoxa2b
hoxa9b
hoxa10b
hoxa11b
hoxa13b
hoxb1a
hoxb2a
hoxb3a
hoxb4a
hoxb5a
hoxb6a
hoxb7a
hoxb8a
hoxb9a
hoxb10a
hoxb13a
hoxb1b
hoxb3b
hoxb5b
hoxb6b
hoxb8b
hoxc1a
hoxc3a
hoxc4a
hoxc5a
hoxc6a
hoxc8a
hoxc9a
hoxc10a
hoxc11a
hoxc12a
hoxc13a
hoxc6b
hoxc12b
hoxc13b
hoxd3a
hoxd4a
hoxd9a
hoxd10a
hoxd11a
hoxd12a
hoxd13a
hoxd4b
hoxd9b
hoxd11b
Stage
L. Blastula
Shield
Early Segm
Mid Segm.
Late Segm
Prehatch
TPM
(max scaled)
1.00
0.75
0.50
0.25
0.00
unannotate
B
hoxa9
hoxc5
hoxd9
normalized counts
1.00
0.75
0.50
0.25
0.00
hoxa9a
hoxa9b
hoxa9b
hoxa9a
hoxc5a
hoxc5a
hoxd9a
hoxa9b
hoxa9a
Developmental stage
Species
a Turbot
a Zebrafish
Stage
L. Blastula
Shield
Early Segm.
Mid Segm.
Late Segm.
Prehatch