

Article

Unveiling Gene Expression Dynamics during Early Embryogenesis in *Cynoglossus semilaevis*: A Transcriptomic Perspective

Xinyi Cheng^{1,2}, Wei Jiang³, Qian Wang^{4,5} , Kaiqiang Liu^{4,5}, Wei Dai² , Yuyan Liu^{4,5}, Changwei Shao^{4,5,*} and Qiye Li^{2,3,6,*} 

- ¹ School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China; bixinyicheng@outlook.com
 - ² BGI Research, Wuhan 430074, China; daiwei5@genomics.cn
 - ³ BGI Research, Shenzhen 518083, China; jiangwei2@genomics.cn
 - ⁴ State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China; wangqian2014@ysfri.ac.cn (Q.W.); liukq@ysfri.ac.cn (K.L.); liuyy@ysfri.ac.cn (Y.L.)
 - ⁵ Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China
 - ⁶ College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
- * Correspondence: shaochangwei303@163.com (C.S.); liqiye@genomics.cn (Q.L.)

Abstract: Commencing with sperm–egg fusion, the early stages of metazoan development include the cleavage and formation of blastula and gastrula. These early embryonic events play a crucial role in ontogeny and are accompanied by a dramatic remodeling of the gene network, particularly encompassing the maternal-to-zygotic transition. Nonetheless, the gene expression dynamics governing early embryogenesis remain unclear in most metazoan lineages. We conducted transcriptomic profiling on two types of gametes (oocytes and sperms) and early embryos (ranging from the four-cell to the gastrula stage) of an economically valuable flatfish—the Chinese tongue sole *Cynoglossus semilaevis* (Pleuronectiformes: Cynoglossidae). Comparative transcriptome analysis revealed that large-scale zygotic genome activation (ZGA) occurs in the blastula stage, aligning with previous findings in zebrafish. Through the comparison of the most abundant transcripts identified in each sample and the functional analysis of co-expression modules, we unveiled distinct functional enrichments across different gametes/developmental stages: actin- and immune-related functions in sperms; mitosis, transcription inhibition, and mitochondrial function in oocytes and in pre-ZGA embryos (four- to 1000-cell stage); and organ development in post-ZGA embryos (blastula and gastrula). These results provide insights into the intricate transcriptional regulation of early embryonic development in Cynoglossidae fish and expand our knowledge of developmental constraints in vertebrates.

Keywords: *Cynoglossus semilaevis*; Cynoglossidae; gametes; early embryonic development; transcriptomics



Citation: Cheng, X.; Jiang, W.; Wang, Q.; Liu, K.; Dai, W.; Liu, Y.; Shao, C.; Li, Q. Unveiling Gene Expression Dynamics during Early Embryogenesis in *Cynoglossus semilaevis*: A Transcriptomic Perspective. *Life* **2024**, *14*, 505. <https://doi.org/10.3390/life14040505>

Academic Editor: Einar Ringø

Received: 24 March 2024

Revised: 10 April 2024

Accepted: 11 April 2024

Published: 15 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Following fertilization, the early developmental stages of an embryo play a decisive role in the correct patterning of the body plan and the fitness of a multicellular organism [1–3]. Employing maternal transcripts and materials stored in the cytoplasm, embryonic development generally begins with several rounds of cleavage and the critical maternal-to-zygotic transition (MZT), which occurs through zygotic genome activation (ZGA) [4,5]. The ZGA triggers the large-scale transcription of zygotic genes, enabling the zygotic genome to replace residual maternal transcripts and dominate embryogenesis (including cell differentiation and organogenesis) [4,5]. The pace of embryogenesis varies greatly amongst metazoan lineages. Zebrafish (*Danio rerio*) and frogs (*Xenopus laevis*) exhibit rapid cleavage followed

by large-scale ZGA in the mid-blastula stage (three and seven hours after fertilization, respectively), while for mammals, initial cleavage cycles last several days, accompanied by ZGA at an early embryonic stage [4,5]. However, the embryogenesis of most species remains to be studied, and the regulatory mechanisms underlying the early embryonic development of many lineages—especially in non-mammals—are unclear.

Flatfish (Pleuronectiformes) are a large group of marine fish that are widely distributed along coastlines and perfectly adapted to a benthic lifestyle. Besides their economic and culinary value, these fish have attracted the scientific community's attention due to their unique biological features: an asymmetrical body shape, a "fin-feet" walking style adapting to benthic movement [6–8], and rapidly evolved sex determination systems [9]. The Chinese tongue sole (*Cynoglossus semilaevis*), widely cultivated in Asia, is a representative species of the Cynoglossidae family of flatfish. Over the past two decades, plentiful genetic resources have been obtained for *C. semilaevis*, such as reliable molecular-marker genetic maps [10–14], high-quality genome assembly and annotation data [9], and accumulating epigenetic and transcriptomic data for different tissues and developmental stages [15–23]. In addition, cell lines have been established from different tissues of this flatfish [24–28]. This progress highlights the potential of *C. semilaevis* to serve as a model species for flatfish. However, little is known about the early stages of *C. semilaevis* embryogenesis thus far.

This study aims to fill the gap in our knowledge about the transcriptional regulation of early embryonic development in *C. semilaevis* through transcriptomic profiling. We collected two types of gametes and early embryos from the four-cell to gastrula stages of *C. semilaevis* development for strand-specific RNA sequencing. By comparing the transcriptomic landscape of gametes and early embryos, we revealed the gene regulatory mechanisms underlying the early embryogenesis of Cynoglossidae. We also compared our results with other teleosts to investigate the conservation and diversification of early embryogenesis in fish.

2. Materials and Methods

2.1. Sampling

Experiments were carried out according to ethical guidelines from the animal care and use committee of BGI. Adult fish, including females and males, were housed in the Haiyang High-Tech Experimental Base. Unfertilized oocytes and sperms were collected and stored in liquid nitrogen. Zygotes were produced by artificial fertilization and reared in aerated water at a temperature of 21 to 23 °C. The developing embryos (including 4-cell, 32-cell, 64-cell, 128-cell, 1000-cell, blastula, and gastrula) were collected and stored in liquid nitrogen.

2.2. RNA Extraction and RNA Sequencing

RNA extraction was performed on two types of gametes (oocytes and sperms) and on early embryos at different stages (4-cell through to gastrula). Except for oocytes ($n = 2$) and sperms ($n = 4$), three biological replicates ($n = 3$) were obtained for each sample category (Figure 1), while for each replicate, numerous gametic cells or embryos were pooled to obtain sufficient RNA material for library construction. We added TRIzol lysis buffer to samples and extracted RNA by thoroughly mixing 300 μ L of chloroform–isoamyl alcohol (24:1) and centrifugation ($12,000 \times g$ for 8 min at 4 °C). The supernatant was transferred to a 1.5 mL centrifuge tube, gently mixed with 2/3 volume of isopropyl alcohol, and kept at -20 °C for over two hours to precipitate RNA. Finally, the precipitation was washed with 0.9 mL of 75% ethanol, air-dried, and dissolved with 20 to 200 μ L of RNase-free water to obtain purified RNA.

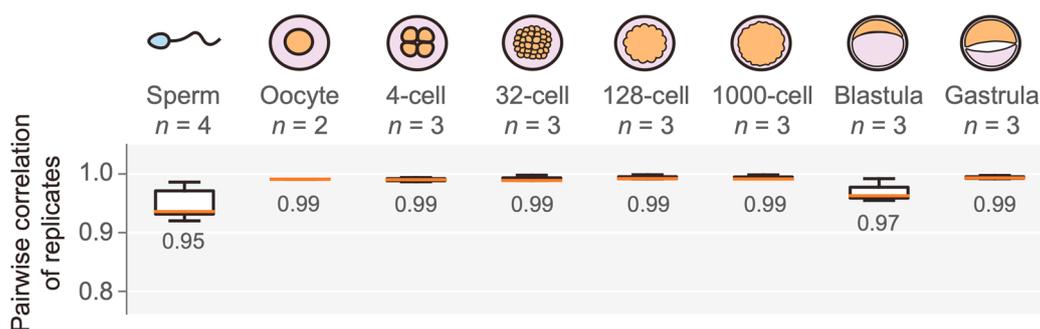


Figure 1. A schematic for sample collection with the number of replicates indicated (**top**) and the high consistency of gene expression between replicates (**bottom**). The degree of consistency was measured by Pearson's correlation of gene expression between two replicates within the same sample group. The mean correlation coefficient for each sample group is presented under the box.

The integrity of purified RNA samples was assessed with an Agilent (Santa Clara, CA, USA) 2100 Bioanalyzer. Qualified RNA samples ($n = 24$) were subjected to strand-specific RNA-seq library construction in two batches. For the first batch ($n = 9$), RNA-seq libraries were prepared using the TruSeq Stranded mRNA Sample Prep Kit (RS-122-2101, Illumina, San Diego, CA, USA) and sequenced on the Illumina HiSeq 4000 platform according to the manufacturer's instructions (Illumina, San Diego, CA, USA). For the second batch ($n = 15$), RNA-seq libraries were prepared with the MGIEasy RNA Directional Library Prep Kit (1000005272, MGI, Shenzhen, China) according to the manufacturer's instructions (MGI Tech Co., Shenzhen, China) and sequenced on the DIPSEQ-T1 platform at the China National GeneBank (Shenzhen, China). All sequencing was performed using PE100 chemistry. The library preparing kit, sequencing platform, sequencing data, and alignment metrics for each sample are presented in Table S1.

2.3. RNA-Seq Reads Filtering, Mapping, and Gene Expression Quantification

The genome assembly and gene annotations used in this study were sourced from the Chinese tongue sole genome project [9]. The RNA-seq reads were subjected to quality control by SOAPnuke v1.5.6 software [29], maintaining the reads of at least 60 bp with a proportion of Q10 bases of at least 80% and an N rate of no more than 5%. Filtered reads were mapped to the reference genome by HISAT2 v2.1.0 software [30–32] with an allowed intron length between 20 bp and 20 kb, and secondary alignments were removed by Samtools v1.11 with the *view* function (parameters: -F 256). Qualified alignments were then used to quantify the gene expression level of each gene in each sample, which produced a read count matrix and a transcripts per million (TPM) matrix for downstream analyses.

2.4. Differential Gene Expression Analysis

We performed principal component analysis (PCA) based on the TPM matrix using the *prcomp* function in R (4.3.2). Differentially expressed genes (DEGs) between sample categories were identified using the DESeq2 v1.34.0 software [33] package in R (parameters: *fitType* = "mean"). Considering the low number of replicates for each sample category, we applied a relatively stringent threshold (fold change > 2 and adjusted *p*-value < 0.05) to ensure the robustness of DEG detection following previous suggestions [34].

2.5. Gene Co-Expression Module Identification, Hub Genes Identification, and Functional Analysis

The read count matrix was first subjected to variance stabilizing transformation (VST) using DESeq2. Dynamically expressed genes were defined as genes with a square deviation of VST read count greater than the median across all samples ($n = 24$) and a TPM > 1 in at least one of the 24 samples. Then, WGCNA v1.72.1 software [35] was used to cluster all dynamically expressed genes into co-expression modules (parameters: *power* = 20, *corOptions* = list (*use* = "p", *method* = "spearman"), *networkType* = "signed",

TOMtype = “signed”, deepSplit = 4, mergeCutHeight = 0.1, reassignThreshold = 0, numericLabels = TRUE, pamRespectsDendro = FALSE). Modules with similar expression dynamics (represented by TPM Z-score) were manually merged. Hub genes of each module were identified as genes with top 30 intra-modular connectivity, and their co-expression networks were visualized using Cytoscape v1.1 [36]. To investigate the potential function of each co-expression module, Gene Ontology (GO) enrichment analysis was conducted on genes from each module using the EnrichPipeline software [37–39] (parameters: --P_Adjust_Methodfdr --TestMethodFisherChiSquare --pc 0.05).

2.6. Homology Analysis between the Embryonic Transcriptome of Tongue Sole and Zebrafish

Orthologs between *C. semilaevis* and *D. rerio* genomes were identified using Proteinortho [40], following the reciprocal best hit (RBH) rule. The distribution bias of orthologs in different co-expression modules was checked using Fisher’s exact test. To analyze the similarity of ortholog expression dynamics in zebrafish, transcriptome data of zebrafish embryogenesis were collected from Levin et al. [41] and compared with the expression profile in *C. semilaevis* embryos.

3. Results

3.1. RNA-Seq Data Exhibits High Sequencing Quality

We extracted total RNA from two types of gametes (sperms and oocytes) and early embryos in different developmental stages (including 4-cell, 32-cell, 128-cell, 1000-cell, blastula, and gastrula). Two to four (mostly three) biological replicates were prepared for each sample category, accumulating 24 RNA samples for strand-specific RNA-seq library construction. The RNA integrity numbers were all greater than eight, as assessed by the Agilent 2100 Bioanalyzer. For each sample, 43.8 to 204.6 million 100 bp pair-end reads were generated either by the HiSeq 4000 platform or the DIPSEQ-T1 platform. Adapter sequences were trimmed, and low-quality reads were filtered, resulting in 43.4 to 121.2 million clean reads with high sequencing quality (Q30, 92.37–95.81%; Table S1). More than 90.7% of these clean reads could be mapped to the *C. semilaevis* reference genome, and over 84.5% of clean reads were properly paired with their counterparts (Table S1). The properly paired reads were used for gene expression quantification by calculating the TPM value of each gene in each sample. Of note, although two different platforms collected the RNA-seq data of different replicates for each sample category, the gene expression quantification showed a high level of consistency between each pair of replicates regardless of their sequencing platforms (Figure 1, Figures S1 and S2). Therefore, our RNA-seq data were of a sufficient quality to quantify gene expression and conduct comparative analyses across samples.

3.2. Gene Expression Profile and Inferred Zygotic Genome Activation

Of the 22,152 annotated genes in the *C. semilaevis* genome, 21,920 were covered by at least one RNA-seq read across the 24 samples. Principal component analysis (PCA) based on the expression level of these 21,920 genes revealed that the sperm exhibited a substantially different transcriptomic landscape to the oocytes and all of the embryonic samples. Additionally, the blastula and gastrula were clearly separated from the cleavage stages of four cells to one thousand cells (Figure 2A). We then identified the DEGs between gametes and early embryos and between consecutive developmental stages with DESeq2. This showed a greater difference between sperm and early embryos than between oocyte and early embryos and also revealed a major transition of the transcriptomic landscape as embryos developed from the one thousand-cell stage to blastulas (Figure 2B and Figure S3). Considering the similarity of transcriptomes in oocyte and four-cell to one thousand-cell embryos and the notable increase in the number of DEGs at the blastula stage, we inferred that large-scale ZGA—when the transcription of the zygotic genome dramatically alters the expression profile [4]—occurs in the blastula.

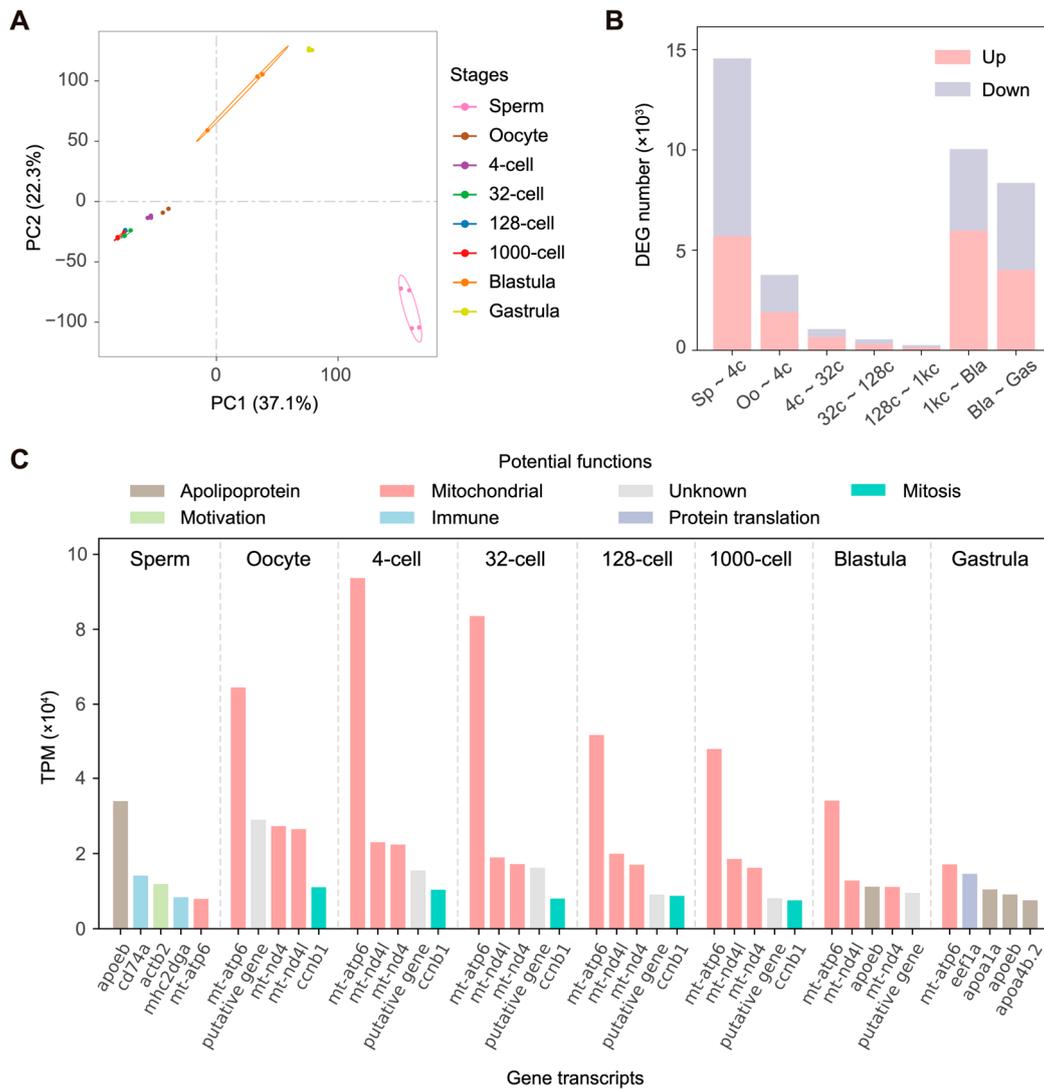


Figure 2. (A) Principal component analysis (PCA) based on 21,920 *Cynoglossus semilaevis* genes covered by at least one read. (B) Numbers of up-regulated (pink) and down-regulated (grey) differentially expressed genes (DEGs) between pairs of consecutive embryogenesis stages in *C. semilaevis*. Sp: sperm; Oo: oocyte; 4c: 4 cells; 32c: 32 cells; 128c: 128 cells; 1kc: 1000 cells; Bla: blastula; Gas: gastrula. (C) The five most abundant gene transcripts in different gametes and embryonic developmental stages in *C. semilaevis*. The potential functions of genes are indicated by color. The putative gene (gene ID: Cse_R021186, located in scaffold7208: 12,125–16,516) was structurally predicted but did not correspond to orthologs in the current NCBI-nr database; its biological function is unclear. TPM, transcripts per million.

Our inference of ZGA timing was also supported by comparing the high-abundance transcripts across samples. From the five most abundant gene transcripts in each gamete and developmental stage (Figure 2C), we found that highly expressed genes were identical for oocytes and embryos in the four-cell to the one thousand-cell stages: coding mitochondrial ATP synthase (*mt-atp6*), mitochondrial NADH dehydrogenase (*mt-nd4l*, *mt-nd4*), and mitosis regulatory protein (*ccnb1*). The relative abundance of these genes decreased as the early embryo developed, and genes encoding three apolipoproteins (*apoeb*, *apoa1a*, *apoa4b.2*) and eukaryotic elongation factor 1 alpha (*ef1a*, which is vital for protein translation [42]) became more dominant in the blastula and gastrula stages. Distinct from gene expression patterns in oocytes and early embryos, the dominant transcripts in sperm rep-

resented different functions, including the regulation of motivation (*actb2*) and immunity (*cd74a*, *mhc2dga*).

3.3. Gene Co-Expression Modules and Corresponding Biological Functions

To gain insight into the gene regulatory networks that drive early embryogenesis in *C. semilaevis*, we next defined gene co-expression modules along the time series. Among all the expressed genes, we identified 9549 genes that were dynamically expressed in development—defined as genes with a square deviation of VST read count higher than the median across all samples and a TPM of one in at least one sample. Hierarchical clustering of all samples based on these dynamic genes presented a clustering pattern identical to that generated by PCA for all genes (Figures 2A and 3A). Then, we divided the dynamically expressed genes into 33 small co-expression modules by WGCNA [35]. After excluding two modules that resulted from batch effects, we manually merged the remaining modules with similar expression dynamics. This resulted in five representative co-expression modules that accounted for 9448 (98.9%) of all of the dynamically expressed genes. Genes in these five modules exhibited their maximum expression levels in different sample stages. We named these modules Sperm, Maternal, Pre-ZGA, Post-ZGA, and Sperm–post-ZGA, respectively (Figure 3B,C).

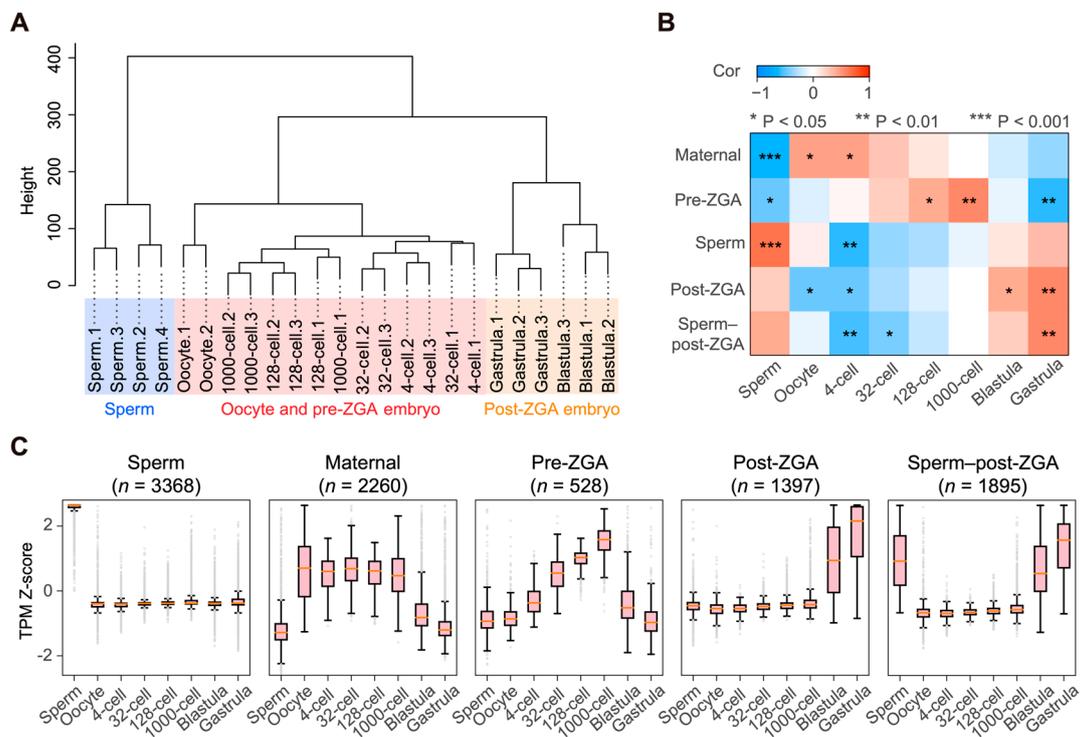


Figure 3. (A) Hierarchical clustering of the 24 samples based on 9549 dynamically expressed genes. Three distinct clusters of samples are highlighted by different colors and named after the developmental stages, namely “Sperm” (blue), “Oocyte and pre-ZGA embryo” (red), and “Post-ZGA embryo” (yellow). (B) The correlation (Cor) between the eight sample categories and five co-expression modules. The significance (P) of Spearman’s correlation analysis is indicated by asterisks. (C) Relative expression levels of genes in different modules, represented by Z-score of transcripts per million (TPM). The title of each plot indicates the highly expressed stage(s) of gene members in the relevant module. The number of genes is shown in parentheses. Large-scale zygotic genome activation (ZGA) was inferred to occur in the blastula stage.

Next, we calculated each gene’s intra-modular connectivity and defined hub genes as those with the top 30 connectivities. Gene co-expression networks of hub genes in different

modules are shown in Figure 4. The Pre-ZGA module, as the smallest module, presented a weaker intra-modular connectivity than the other four modules.

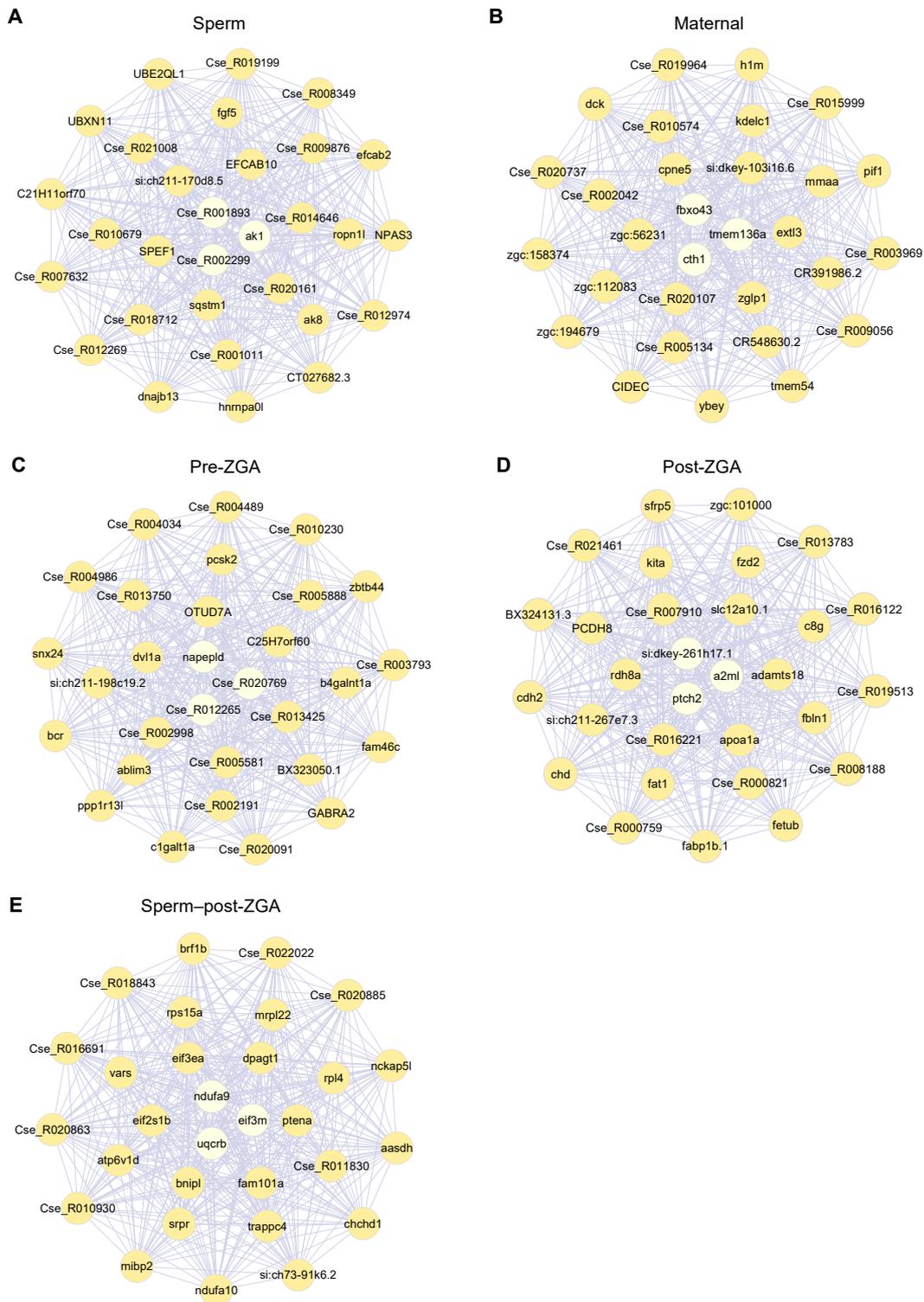


Figure 4. Hub genes with the top 30 intra-modular connectivity in five co-expression modules, respectively, namely Sperm (A), Maternal (B), Pre-ZGA (C), Post-ZGA (D), and Sperm-post-ZGA (E). Purple lines represent the correlation (0.95~1) between two linked genes, with line width representing value size. Genes closer to the center of the network (indicated by light yellow) have higher intra-modular connectivity.

The GO enrichment analysis on these modules suggested diverse biological functions specific to different growth or developmental tasks (Figure 5; Tables S2–S6). For example, genes highly expressed in the Sperm module mainly conferred immune response and cell activation; genes from the Maternal module—whose transcripts were relatively abundant in oocytes and four-cell to one thousand-cell embryos—mainly contributed to mitosis, transcription inhibition, oogenesis, and sperm–egg fusion; genes from the Post-ZGA module (abundant in the transition from blastula to gastrula) were enriched for cell differentiation, organ development, transcription regulation, and several biosynthetic processes. In addition, genes from the Sperm–post-ZGA module contributed to protein synthesis or disassembly and cellular respiration. Nevertheless, the Pre-ZGA module—which might be composed of relatively slowly degraded maternal transcripts—was not significantly enriched for any GO terms.

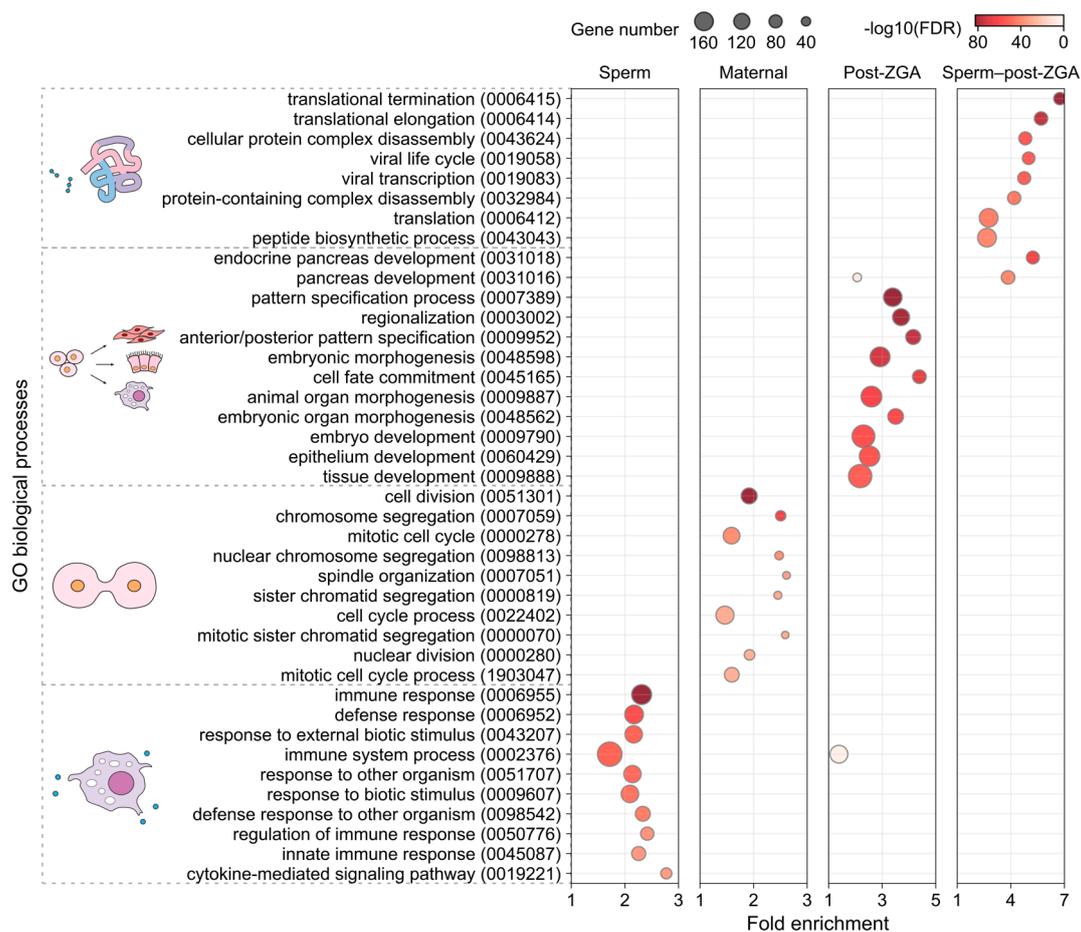


Figure 5. The top most significant Gene Ontology (GO) biological processes which are enriched in different gene co-expression modules, respectively. GO ID of each term is indicated in parentheses. Biological processes with similar functions are clustered and represented by a single icon. FDR: false discovery rate.

3.4. Comparative Analysis of Co-Expression Genes between Tongue Sole and Zebrafish

After exploring the gene regulatory networks governing early embryogenesis in *C. semilaevis*, we next questioned whether these networks may also drive embryonic development in other teleost fish. Therefore, we compared the *C. semilaevis* gene expression data with that of *D. rerio* [41] to measure the interspecific conservation based on 14,087 orthologs identified between *C. semilaevis* and *D. rerio*. We first checked for the distribution of orthologs in different co-expression modules. We observed that the orthologs are enriched in the Maternal, Post-ZGA, and Sperm–post-ZGA modules of *C. semilaevis* (Figure 6A).

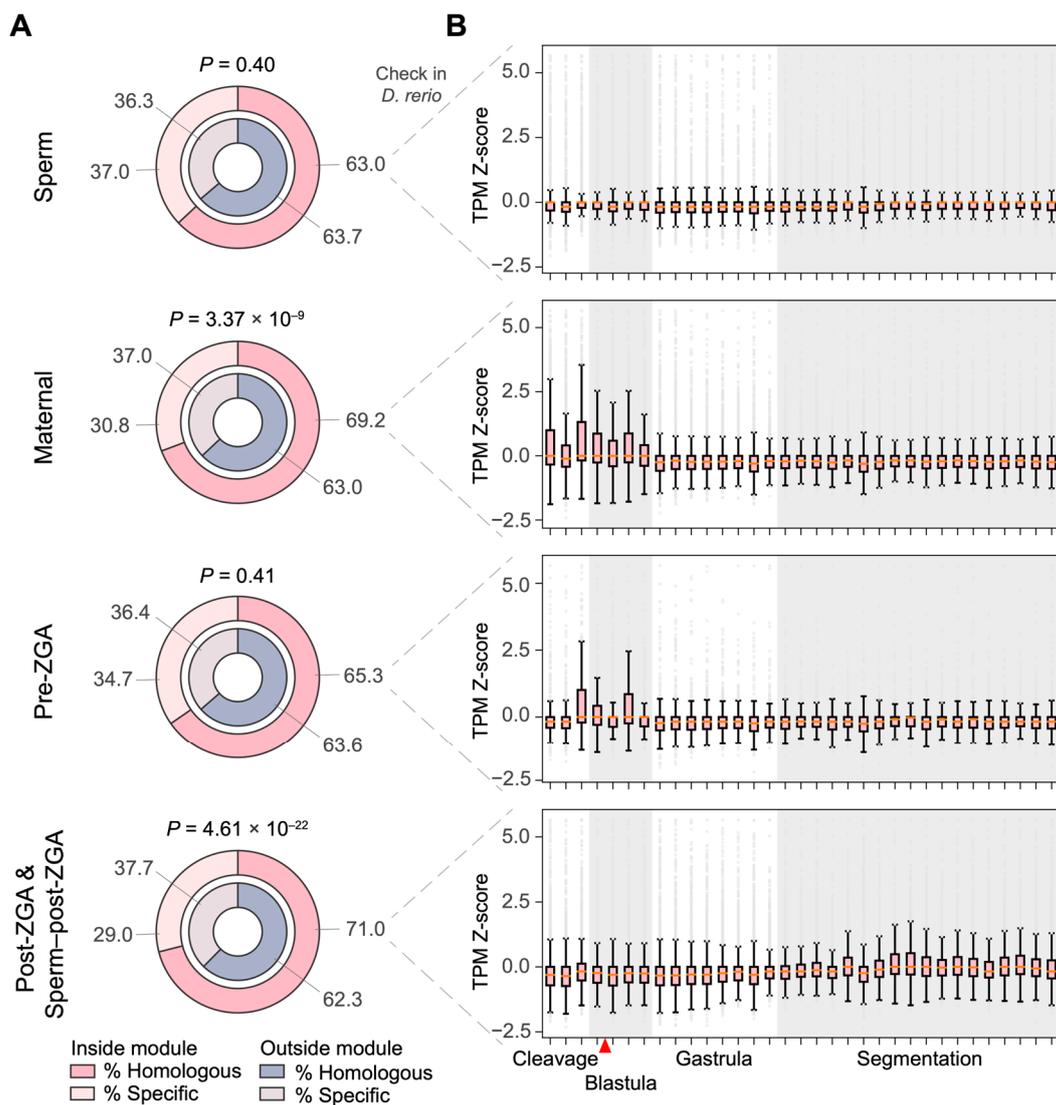


Figure 6. (A) Interspecific conservation of genes in co-expression modules, compared with *Danio rerio*. The distribution bias of orthologs was tested by Fisher's exact test. (B) Relative expression level of the orthologs in *D. rerio* during embryogenesis. The horizontal scale indicates samples taken every 40 min from fertilization. The red arrow indicates the putative ZGA of the *D. rerio* embryo. The *D. rerio* gene expression data were obtained from Levin et al. [41].

We next examined the expression dynamics of these orthologs in zebrafish during embryogenesis (Figure 6B). Specifically, the zebrafish orthologs from the Sperm module of *C. semilaevis* showed a very stable expression level throughout embryonic development; orthologs from the Maternal module displayed a higher expression level in early embryogenesis, i.e., from zygote to the end of blastula (0.66–4.66 h post-fertilization, or, hpf), than in mid and later embryonic development; orthologs from the Pre-ZGA module showed the highest expression level from 64-cell (2 hpf) to sphere (4 hpf); orthologs from the Post-ZGA and Sperm-post-ZGA modules remained unchanged from zygote to early segmentation, but slightly rose in the mid and late segmentation stages. Considering that ZGA of zebrafish embryos occurs in the mid-blastula stage [4,43], these results imply that the gene regulatory networks underlying pre-ZGA cleavages are mostly conserved in teleosts, while the networks driving post-ZGA development have been re-wired after the divergence of these two species.

4. Discussion

This study generated high-quality RNA-seq data for *C. semilaevis* gametes (sperm and oocyte) and early embryos at different developmental stages (from the four-cell to gastrula stages) to track the global gene expression changes during early embryogenesis. For the first time, we confirmed that large-scale ZGA occurs at the blastula stage in *C. semilaevis*, consistent with the findings for Japanese flounders [44], zebrafish, and frogs [4]. Accordingly, gene transcripts that were abundant in *C. semilaevis* oocytes were also abundant in four-cell to one thousand-cell embryos, which conferred functions that are vital for pre-ZGA cleavage (e.g., the enrichment of mitosis-related genes in the Maternal module). Maternal transcripts may gradually degrade during early embryogenesis, among which the slower ones occupied an increasing proportion of transcripts and presented higher relative abundance before large-scale ZGA. Yet, no significant biological functions were enriched in these slowly degraded transcripts. After large-scale ZGA, the newly generated zygotic transcripts replaced the dominant role of maternal transcripts, leading embryonic development by promoting organ differentiation and the biosynthesis process. Indeed, some genes identified in the Post-ZGA module have been previously verified to play a role in embryonic developmental regulation in *C. semilaevis*, such as *rspo2* (involved in muscle development during embryogenesis) [45,46], *gata5* (increased at the blastula stage and peaked at the heart-beating period) [47], and *sox10* (starting to be expressed in blastula and highly expressed in neurula) [48].

An interspecific comparison revealed similar embryonic regulatory patterns which have been reported in previous gene expression analyses in multiple teleost species; the functional enrichment of cell division or DNA repair has been discovered in two-cell embryos of Atlantic halibut (where cyclin B1 is abundant) [49], zygotes of zebrafish at 3.5 hpf [50], and the morula of both European seabass [51] and three-spined sticklebacks [52]. All these studies observe signals for organ development among genes highly expressed in the later stages—in line with our findings. The consistency of ZGA timing and embryonic transcriptome functions in different teleost species suggests a homology of early embryogenesis in fish. Notably, our transcriptome comparisons between *C. semilaevis* and zebrafish revealed conserved gene regulatory networks underlying pre-ZGA cleavages but not post-ZGA development in fish (Figure 6). The interspecific conservation and divergence of gene expression is also noteworthy for aquaculture researchers, as it may necessitate a precise adjustment in the target growth stage during the development of breeding techniques across distinct species.

Interestingly, genes highly expressed in the sperm of *C. semilaevis* conferred actin- and immune-related functions. The abundance of actin was expected, as actin is crucial for spermatogenesis in animals as it forms the filaments that support capacitation and motility [53,54]. However, the strong correlation between sperm transcripts and the immune system in fish has rarely been reported. Previous studies have primarily limited the functions of the sperm transcriptome to sperm capacity, fertilization, and embryonic development [55–57]. Immune regulatory functions in mammalian sperm and semen transcripts are assumed to protect the maternal genital tract from infection or inhibit maternal immune responses to sperm [56–58]. Transcriptome analysis of the ovaries of rockfish—a viviparous teleost fish that stores sperm inside the ovary cavity for months—have likewise revealed changes in immune function before and after mating [59]. Together, these studies emphasize the effect of semen on maternal immune regulation during *in vivo* fertilization, while the immunological effect of semen among *in vitro* fertilized species remains unclear. Our functional analysis of sperm-specific transcripts in *C. semilaevis* in this study suggests that sperm cells may promote immune activation and, therefore, be critical for safeguarding *in vitro* fertilization in fish, which is worthy of further research.

In conclusion, our study comprehensively analyzes gametes and early embryonic gene expression profiles in *C. semilaevis*. The high-depth RNA-seq dataset will serve as a valuable resource for studying gene regulatory mechanisms underlying the early life of Cynoglossidae. It may assist in searching for candidate genes (e.g., related to growth

and development) that benefit aquaculture practices. The transcriptomic comparison between species also contributes to knowledge about the conservation and divergence of gene regulation in fish early embryogenesis. Future studies are expected to explore the evolutionary implications of these regulatory strategies for speciation and adaptation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life14040505/s1>, Table S1: Sequencing methods, quality control and mapping statistics of all samples in *Cynoglossus semilaevis*; Table S2: GO enrichment results of gene members in Sperm module; Table S3: GO enrichment results of gene members in Maternal module; Table S4: GO enrichment results of gene members in Post-ZGA module; Table S5: GO enrichment results of gene members in Sperm–post-ZGA module; Table S6: Symbol descriptions of the GO enrichment results. Figure S1: Pairwise comparison and Pearson’s correlation of gene expression between biological replicates within the same sample category. The sequencing platform of each replicate is shown in parentheses. Figure S2: Pearson’s correlation coefficient obtained from intra-platform comparisons versus those from inter-platform comparisons. NS: not significant assessed by Student’s *t*-test. The correlation coefficient data for boxplot are from Figure S1. Figure S3: Number of differentially expressed genes (DEGs) in pairwise comparison.

Author Contributions: Conceptualization, Q.L. and C.S.; Methodology, X.C., Q.W., K.L., W.J., W.D. and Y.L.; Software, X.C.; Validation, X.C.; Formal analysis, X.C.; Investigation, Q.W., K.L. and Y.L.; Resources, Q.W., K.L., Y.L. and C.S.; Data curation, X.C.; Writing—original draft preparation, X.C.; Writing—Review And Editing, Q.L.; Visualization, X.C.; Supervision, Q.L. and C.S.; Project administration, C.S. and Q.L.; Funding acquisition, C.S. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 31501057.

Institutional Review Board Statement: The animal study protocol was approved by the Institutional Review Board of BGI (protocol code BGI-IRB A20012; date of approval: 28 June 2020).

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw RNA-seq data supporting this study’s findings have been deposited into the CNGB Sequence Archive (CNSA) (<https://db.cngb.org/>) with accession number CNP0001602, accessed on 10 April 2024.

Acknowledgments: We sincerely thank the China National GeneBank (CNGB) for computing resources.

Conflicts of Interest: Xinyi Cheng, Wei Jiang, Wei Dai, and Qiye Li are hired by BGI research.

References

1. Zhai, J.; Xiao, Z.; Wang, Y.; Wang, H. Human embryonic development: From peri-implantation to gastrulation. *Trends Cell Biol.* **2022**, *32*, 18–29. [[CrossRef](#)] [[PubMed](#)]
2. Rivera-Pérez, J.A.; Hadjantonakis, A.-K. The Dynamics of Morphogenesis in the Early Mouse Embryo. *Cold Spring Harb. Perspect. Biol.* **2014**, *7*, a015867. [[CrossRef](#)] [[PubMed](#)]
3. Arnold, S.J.; Robertson, E.J. Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 91–103. [[CrossRef](#)] [[PubMed](#)]
4. Jukam, D.; Shariati, S.A.M.; Skotheim, J.M. Zygotic Genome Activation in Vertebrates. *Dev. Cell* **2017**, *42*, 316–332. [[CrossRef](#)] [[PubMed](#)]
5. Lee, M.T.; Bonneau, A.R.; Giraldez, A.J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 581–613. [[CrossRef](#)] [[PubMed](#)]
6. Fox, C.H.; Gibb, A.C.; Summers, A.P.; Bemis, W.E. Benthic walking, bounding, and maneuvering in flatfishes (Pleuronectiformes: Pleuronectidae): New vertebrate gaits. *Zoology* **2018**, *130*, 19–29. [[CrossRef](#)] [[PubMed](#)]
7. Holmes, R.A.; Gibson, R.N. A comparison of predatory behaviour in flatfish. *Anim. Behav.* **1983**, *31*, 1244–1255. [[CrossRef](#)]
8. Janvier, P. Palaeontology: Squint of the fossil flatfish. *Nature* **2008**, *454*, 169–170. [[CrossRef](#)] [[PubMed](#)]
9. Chen, S.; Zhang, G.; Shao, C.; Huang, Q.; Liu, G.; Zhang, P.; Song, W.; An, N.; Chalopin, D.; Volff, J.-N.; et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* **2014**, *46*, 253–260. [[CrossRef](#)] [[PubMed](#)]

10. Chen, S.-L.; Li, J.; Deng, S.-P.; Tian, Y.-S.; Wang, Q.-Y.; Zhuang, Z.-M.; Sha, Z.-X.; Xu, J.-Y. Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). *Mar. Biotechnol.* **2007**, *9*, 273–280. [[CrossRef](#)]
11. Shao, C.-W.; Chen, S.-L.; Scheuring, C.F.; Xu, J.-Y.; Sha, Z.-X.; Dong, X.-L.; Zhang, H.-B. Construction of two BAC libraries from half-smooth tongue sole *Cynoglossus semilaevis* and identification of clones containing candidate sex-determination genes. *Mar. Biotechnol.* **2010**, *12*, 558–568. [[CrossRef](#)]
12. Sha, Z.; Wang, S.; Zhuang, Z.; Wang, Q.; Wang, Q.; Li, P.; Ding, H.; Wang, N.; Liu, Z.; Chen, S. Generation and analysis of 10,000 ESTs from the half-smooth tongue sole *Cynoglossus semilaevis* and identification of microsatellite and SNP markers. *J. Fish Biol.* **2010**, *76*, 1190–1204. [[CrossRef](#)] [[PubMed](#)]
13. Song, W.; Li, Y.; Zhao, Y.; Liu, Y.; Niu, Y.; Pang, R.; Miao, G.; Liao, X.; Shao, C.; Gao, F.; et al. Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). *PLoS ONE* **2012**, *7*, e52097. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, J.; Shao, C.; Zhang, L.; Liu, K.; Gao, F.; Dong, Z.; Xu, P.; Chen, S. A first generation BAC-based physical map of the half-smooth tongue sole (*Cynoglossus semilaevis*) genome. *BMC Genom.* **2014**, *15*, 215. [[CrossRef](#)] [[PubMed](#)]
15. Wang, W.; Yi, Q.; Ma, L.; Zhou, X.; Zhao, H.; Wang, X.; Qi, J.; Yu, H.; Wang, Z.; Zhang, Q. Sequencing and characterization of the transcriptome of half-smooth tongue sole (*Cynoglossus semilaevis*). *BMC Genom.* **2014**, *15*, 470. [[CrossRef](#)] [[PubMed](#)]
16. Dong, Y.; Lyu, L.; Wen, H.; Shi, B. Brain and Pituitary Transcriptome Analyses Reveal the Differential Regulation of Reproduction-Related lncRNAs and mRNAs in *Cynoglossus semilaevis*. *Front. Genet.* **2021**, *12*, 802953. [[CrossRef](#)] [[PubMed](#)]
17. Dong, Y.; Lyu, L.; Zhang, D.; Li, J.; Wen, H.; Shi, B. Integrated lncRNA and mRNA Transcriptome Analyses in the Ovary of *Cynoglossus semilaevis* Reveal Genes and Pathways Potentially Involved in Reproduction. *Front. Genet.* **2021**, *12*, 671729. [[CrossRef](#)] [[PubMed](#)]
18. Wang, P.; Zheng, M.; Liu, J.; Liu, Y.; Lu, J.; Sun, X. Sexually Dimorphic Gene Expression Associated with Growth and Reproduction of Tongue Sole (*Cynoglossus semilaevis*) Revealed by Brain Transcriptome Analysis. *Int. J. Mol. Sci.* **2016**, *17*, 1402. [[CrossRef](#)] [[PubMed](#)]
19. Lu, Y.-F.; Liu, Q.; Liu, K.-Q.; Wang, H.-Y.; Li, C.-H.; Wang, Q.; Shao, C.-W. Identification of global alternative splicing and sex-specific splicing via comparative transcriptome analysis of gonads of Chinese tongue sole (*Cynoglossus semilaevis*). *Zool Res.* **2022**, *43*, 319–330. [[CrossRef](#)] [[PubMed](#)]
20. Lin, G.; Gao, D.; Lu, J.; Sun, X. Transcriptome Profiling Reveals the Sexual Dimorphism of Gene Expression Patterns during Gonad Differentiation in the Half-Smooth Tongue Sole (*Cynoglossus semilaevis*). *Mar. Biotechnol.* **2021**, *23*, 18–30. [[CrossRef](#)] [[PubMed](#)]
21. Ye, Z.; Wang, W.; Zhang, Y.; Wang, L.; Cui, Y.; Li, H. Integrative analysis reveals pathways associated with sex reversal in *Cynoglossus semilaevis*. *PeerJ* **2020**, *8*, e8801. [[CrossRef](#)] [[PubMed](#)]
22. Xu, W.; Cui, Z.; Wang, N.; Zhang, M.; Wang, J.; Xu, X.; Liu, Y.; Chen, S. Transcriptomic analysis revealed gene expression profiles during the sex differentiation of Chinese tongue sole (*Cynoglossus semilaevis*). *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2021**, *40*, 100919. [[CrossRef](#)] [[PubMed](#)]
23. Shao, C.; Li, Q.; Chen, S.; Zhang, P.; Lian, J.; Hu, Q.; Sun, B.; Jin, L.; Liu, S.; Wang, Z.; et al. Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res.* **2014**, *24*, 604–615. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, B.; Wang, X.; Sha, Z.; Yang, C.; Liu, S.; Wang, N.; Chen, S.-L. Establishment and characterization of a testicular cell line from the half-smooth tongue sole, *Cynoglossus semilaevis*. *Int. J. Biol. Sci.* **2011**, *7*, 452–459. [[CrossRef](#)] [[PubMed](#)]
25. Sun, A.; Wang, T.Z.; Wang, N.; Liu, X.F.; Sha, Z.X.; Chen, S.L. Establishment and characterization of an ovarian cell line from half-smooth tongue sole *Cynoglossus semilaevis*. *J. Fish Biol.* **2015**, *86*, 46–59. [[CrossRef](#)] [[PubMed](#)]
26. Wang, X.L.; Wang, N.; Sha, Z.X.; Chen, S.L. Establishment, characterization of a new cell line from heart of half smooth tongue sole (*Cynoglossus semilaevis*). *Fish Physiol. Biochem.* **2010**, *36*, 1181–1189. [[CrossRef](#)] [[PubMed](#)]
27. Sun, A.; Chen, S.-L.; Gao, F.-T.; Li, H.-L.; Liu, X.-F.; Wang, N.; Sha, Z.-X. Establishment and characterization of a gonad cell line from half-smooth tongue sole *Cynoglossus semilaevis* pseudomale. *Fish Physiol. Biochem.* **2015**, *41*, 673–683. [[CrossRef](#)]
28. Zheng, Y.; Wang, N.; Xie, M.-S.; Sha, Z.-X.; Chen, S.-L. Establishment and characterization of a new fish cell line from head kidney of half-smooth tongue sole (*Cynoglossus semilaevis*). *Fish Physiol. Biochem.* **2012**, *38*, 1635–1643. [[CrossRef](#)] [[PubMed](#)]
29. Chen, Y.; Chen, Y.; Shi, C.; Huang, Z.; Zhang, Y.; Li, S.; Li, Y.; Ye, J.; Yu, C.; Li, Z.; et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **2018**, *7*, 1–6. [[CrossRef](#)] [[PubMed](#)]
30. Sirén, J.; Välimäki, N.; Mäkinen, V. Indexing Graphs for Path Queries with Applications in Genome Research. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2014**, *11*, 375–388. [[CrossRef](#)]
31. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)] [[PubMed](#)]
32. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
33. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
34. Schurch, N.J.; Schofield, P.; Gierliński, M.; Cole, C.; Sherstnev, A.; Singh, V.; Wrobel, N.; Gharbi, K.; Simpson, G.G.; Owen-Hughes, T.; et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **2016**, *22*, 839–851. [[CrossRef](#)] [[PubMed](#)]

35. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)] [[PubMed](#)]
36. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
37. Chen, S.; Yang, P.; Jiang, F.; Wei, Y.; Ma, Z.; Kang, L. De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS ONE* **2010**, *5*, e15633. [[CrossRef](#)] [[PubMed](#)]
38. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13. [[CrossRef](#)] [[PubMed](#)]
39. Beissbarth, T.; Speed, T.P. Gostat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **2004**, *20*, 1464–1465. [[CrossRef](#)] [[PubMed](#)]
40. Lechner, M.; Findeiss, S.; Steiner, L.; Marz, M.; Stadler, P.F.; Prohaska, S.J. Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Bioinform.* **2011**, *12*, 124. [[CrossRef](#)] [[PubMed](#)]
41. Levin, M.; Anavy, L.; Cole, A.G.; Winter, E.; Mostov, N.; Khair, S.; Senderovich, N.; Kovalev, E.; Silver, D.H.; Feder, M.; et al. The mid-developmental transition and the evolution of animal body plans. *Nature* **2016**, *531*, 637–641. [[CrossRef](#)] [[PubMed](#)]
42. Mateyak, M.K.; Kinzy, T.G. eEF1A: Thinking outside the ribosome. *J. Biol. Chem.* **2010**, *285*, 21209–21213. [[CrossRef](#)] [[PubMed](#)]
43. Kimmel, C.B.; Ballard, W.W.; Kimmel, S.R.; Ullmann, B.; Schilling, T.F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **1995**, *203*, 253–310. [[CrossRef](#)] [[PubMed](#)]
44. Hao, X.; Wang, Q.; Hou, J.; Liu, K.; Feng, B.; Shao, C. Temporal Transcriptome Analysis Reveals Dynamic Expression Profiles of Gametes and Embryonic Development in Japanese Flounder (*Paralichthys olivaceus*). *Genes* **2021**, *12*, 1561. [[CrossRef](#)] [[PubMed](#)]
45. Niu, J.; Guan, J.; Li, R.; Li, X.; Zhai, J.; Qi, J.; He, Y. *Cynoglossus semilaevis* Rspo3 Regulates Embryo Development by Inhibiting the Wnt/ β -Catenin Signaling Pathway. *Int. J. Mol. Sci.* **2018**, *19*, 1915. [[CrossRef](#)] [[PubMed](#)]
46. Wang, B.; Yang, F.; Li, R.; Li, X.; Wu, X.; Sun, Z.; Zhai, J.; He, Y.; Qi, J. Functional characterization of *Cynoglossus semilaevis* R-spondin2 and its role in muscle development during embryogenesis. *Genes Genet. Syst.* **2018**, *93*, 181–190. [[CrossRef](#)] [[PubMed](#)]
47. Liu, Y.; Zhu, H.; Liu, Y.; Qu, J.; Han, M.; Jin, C.; Zhang, Q.; Liu, J. Molecular characterization and expression profiles provide new insights into GATA5 functions in tongue sole (*Cynoglossus semilaevis*). *Gene* **2019**, *708*, 21–29. [[CrossRef](#)] [[PubMed](#)]
48. Deng, S.P.; Chen, S.L. cDNA cloning, tissues, embryos and larvae expression analysis of Sox10 in half-smooth tongue-sole, *Cynoglossus semilaevis*. *Mar. Genom.* **2008**, *1*, 109–114. [[CrossRef](#)]
49. Bai, J.; Solberg, C.; Fernandes, J.M.O.; Johnston, I.A. Profiling of maternal and developmental-stage specific mRNA transcripts in Atlantic halibut *Hippoglossus hippoglossus*. *Gene* **2007**, *386*, 202–210. [[CrossRef](#)] [[PubMed](#)]
50. Aanes, H.; Winata, C.L.; Lin, C.H.; Chen, J.P.; Srinivasan, K.G.; Lee, S.G.P.; Lim, A.Y.M.; Hajan, H.S.; Collas, P.; Bourque, G.; et al. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **2011**, *21*, 1328–1338. [[CrossRef](#)] [[PubMed](#)]
51. Kaitetzidou, E.; Xiang, J.; Antonopoulou, E.; Tsigenopoulos, C.S.; Sarropoulou, E. Dynamics of gene expression patterns during early development of the European seabass (*Dicentrarchus labrax*). *Physiol. Genom.* **2015**, *47*, 158–169. [[CrossRef](#)] [[PubMed](#)]
52. Kaitetzidou, E.; Katsiadaki, I.; Lagnel, J.; Antonopoulou, E.; Sarropoulou, E. Unravelling paralogous gene expression dynamics during three-spined stickleback embryogenesis. *Sci. Rep.* **2019**, *9*, 3752. [[CrossRef](#)] [[PubMed](#)]
53. Soda, T.; Miyagawa, Y.; Fukuhara, S.; Tanaka, H. Physiological role of actin regulation in male fertility: Insight into actin capping proteins in spermatogenic cells. *Reprod. Med. Biol.* **2020**, *19*, 120–127. [[CrossRef](#)] [[PubMed](#)]
54. Breitbart, H.; Finkelstein, M. Actin cytoskeleton and sperm function. *Biochem. Biophys. Res. Commun.* **2018**, *506*, 372–377. [[CrossRef](#)] [[PubMed](#)]
55. Selvaraju, S.; Ramya, L.; Parthipan, S.; Swathi, D.; Binsila, B.K.; Kolte, A.P. Deciphering the complexity of sperm transcriptome reveals genes governing functional membrane and acrosome integrities potentially influence fertility. *Cell Tissue Res.* **2021**, *385*, 207–222. [[CrossRef](#)] [[PubMed](#)]
56. Jodar, M.; Sendler, E.; Krawetz, S.A. The protein and transcript profiles of human semen. *Cell Tissue Res.* **2016**, *363*, 85–96. [[CrossRef](#)] [[PubMed](#)]
57. Sahoo, B.; Choudhary, R.K.; Sharma, P.; Choudhary, S.; Gupta, M.K. Significance and Relevance of Spermatozoal RNAs to Male Fertility in Livestock. *Front. Genet.* **2021**, *12*, 768196. [[CrossRef](#)] [[PubMed](#)]
58. Schjenken, J.E.; Sharkey, D.J.; Green, E.S.; Chan, H.Y.; Matias, R.A.; Moldenhauer, L.M.; Robertson, S.A. Sperm modulate uterine immune parameters relevant to embryo implantation and reproductive success in mice. *Commun. Biol.* **2021**, *4*, 572. [[CrossRef](#)] [[PubMed](#)]
59. Li, R.; Song, W.; Qu, J.; Liu, H.; Qi, J.; He, Y.; Niu, J. Transcriptome sequencing reveals ovarian immune response and development during female sperm storage in viviparous black rockfish (*Sebastes schlegelii*). *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2023**, *45*, 101050. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.