Marine Biotechnology
https://doi.org/10.1007/s10126-018-9798-4

ORIGINAL ARTICLE

Gonad Transcriptome Analysis of the Pacific Oyster *Crassostrea gigas* Identifies Potential Genes Regulating the Sex Determination and Differentiation Process

Chenyang Yue[1] · Qi Li[1,2] · Hong Yu[1]

Received: 10 September 2017 / Accepted: 12 February 2018

Abstract
The Pacific oyster *Crassostrea gigas* is a commercially important bivalve in aquaculture worldwide. *C. gigas* has a fascinating sexual reproduction system consisting of dioecism, sex change, and occasional hermaphroditism, while knowledge of the molecular mechanisms of sex determination and differentiation is still limited. In this study, the transcriptomes of male and female gonads at different gametogenesis stages were characterized by RNA-seq. Hierarchical clustering based on genes differentially expressed revealed that 1269 genes were expressed specifically in female gonads and 817 genes were expressed increasingly over the course of spermatogenesis. Besides, we identified two and one gene modules related to female and male gonad development, respectively, using weighted gene correlation network analysis (WGCNA). Interestingly, GO and KEGG enrichment analysis showed that neurotransmitter-related terms were significantly enriched in genes related to ovary development, suggesting that the neurotransmitters were likely to regulate female sex differentiation. In addition, two hub genes related to testis development, lncRNA LOC105321313 and *Cg-Sh3kbp1*, and one hub gene related to ovary development, *Cg-Malrd1-like*, were firstly investigated. This study points out the role of neurotransmitter and non-coding RNA regulation during gonad development and produces lists of novel relevant candidate genes for further studies. All of these provided valuable information to understand the molecular mechanisms of *C. gigas* sex determination and differentiation.

Keywords Gonad transcriptome · Sex determination and differentiation · Reproductive regulation · Weighted gene correlation network analysis · *Crassostrea gigas*

Introduction

The Pacific oyster, *Crassostrea gigas*, a member of marine bivalves originating from Northwestern Pacific, has been introduced and translocated worldwide for aquaculture purposes. Nowadays, *C. gigas* is one of the most important commercial oyster species cultivated in the world. For decades, the complex and fascinating modes of sexual reproduction of *C. gigas* caught attention from many biologists who regarded the Pacific oyster as a typical organism to understand the mollusk sex determination. It is generally accepted that *C. gigas* is dioecious without secondary sex characteristics; interestingly, individuals of this species can undergo sex transformation (Coe 1943; Guo et al. 1998). This may occur several times or not in an individual lifetime. Moreover, the occasional hermaphroditism can also be observed (Coe 1943).

Sexual fate is determined by activating the testis or ovarian pathway and repressing the alternative pathway with many genes being expressed in a sexually dimorphic manner (Piferrer 2013; Munger and Capel 2012). Two major types of sex-determining mechanisms have been recognized in organisms that reproduced sexually: genotypic sex determination (GSD), where sex is determined at conception and genetic differences are expected between the sexes, and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10126-018-9798-4) contains supplementary material, which is available to authorized users.

✉ Qi Li
qili66@ouc.edu.cn

[1] Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China

[2] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Published online: 08 March 2018

environmental sex determination (ESD), where there are no consistent genetic differences between sexes and sex is determined after fertilization in response to an environmental cue (Piferrer 2013). Studies employing controlled crosses suggested that genetic factors were important in *C. gigas* (Hedrick and Hedgecock 2010; Guo et al. 1998). Through the analysis of sex ratios in both full- and half-sib families of *C. gigas*, two genetic models of two genotypes and three genotypes were proposed for sex determination (Teaniniuraitemoana et al. 2014; Guo et al. 1998; Hedrick and Hedgecock 2010).

Both of the models proposed above were justified under the assumption that sex in *C. gigas* would be single-gene inheritance. In fact, whether the sex would be controlled by a dominant gene or multiple genes is still questioned. Up to now, some potential downstream genes of the molecular cascade of sex determination and differentiation of *C. gigas* have been identified, including female-determining genes *Cg-β-catenin* and *CgFoxl2* and its natural antisense transcript *Cg-Foxl2os*, as well as the male-determining genes, *Cg-SoxE*, *CgDsx*, and *CgSoxH* (Naimi et al. 2009a, 2009b; Santerre et al. 2012; Zhang et al. 2014; Santerre et al. 2014). In addition, a proposed model was presented where a novel testis-determining *CgSoxH* might serve as a primary regulator, directly or indirectly interacting with a testis-promoting *CgDsx* and an ovary-promoting *CgFoxL2* (Zhang et al. 2014). Similarly, the vertebrate sex-determining genes were also identified using RNA-seq in some other bivalves, such as the pearl oyster *Pinctada margaritifera*, scallop *Chlamys farreri*, and the Yesso scallop *Patinopecten yessoensis* (Teaniniuraitemoana et al. 2014; Li et al. 2016; Liu et al. 2012), which suggested that some key sex-determining genes that were thought to be vertebrate-specific may have participated in sex determination before the divergence of vertebrates and invertebrates (Li et al. 2016). To date, most studies focused on a small number of genes homologous to sex-determining pathway genes in model species, and this might limit the view on the molecular mechanisms of *C. gigas* sex determination and differentiation. To provide a better understanding of this topic, more potential genes should be identified.

The development of gonad involves a wide variety of biological phenomena, such as sex determination, oogenesis, and spermatogenesis. In other words, sex determination and differentiation of *C. gigas* is mainly the process by which an undifferentiated gonad is transformed into ovary or testis (Piferrer 2013). Thus, it is significant to select samples at multiple stages of gonad development for more information about the sexual reproduction of *C. gigas*. A microarray-based analysis to identify sex and stage-specific genes had been carried out in the *C. gigas* (Dheilly et al. 2012), but the disadvantage of sequencing technique and the absence of whole genome dataset limited this research for novel genes. The availability of the *C. gigas* genome (Zhang et al. 2012) and the reduced cost of next-generation sequencing provide the opportunity for a new global view on the complex regulation mechanisms of the whole reproduction cycle. In the present study, we investigated the transcriptional changes of male and female individuals at four gonad development stages including the resting, early gametogenesis, maturation, and spawning stage.

Materials and Methods

Animal Material and Tissue Sampling

Two-year-old adult Pacific oysters were collected monthly between December 2015 and July 2016 from Rushan, Shandong, China. For each oyster, five tissues including gill, mantle, adductor muscle, gonad, and visceral mass were sampled for RNA extraction, and gonad was also fixed for histological analysis. For total RNA extraction, individual samples were preserved in RNAwait (Solarbio) and stored at − 80 °C.

Firstly, gonad development stage and sex of the samples were determined by histological methods. Microscopic examination revealed that there was no trace of sexuality as follicles consist of undifferentiated germinal epithelium at the resting stage (stage 0) of gonad development. At the early gametogenesis stage (stage I), numerous spermatogonia or oogonia appears and can be distinguished according to cellular morphology and size. Follicles are actively developing with primary gametes and some secondary oocytes and spermatozoa (Enríquez-Díaz et al. 2008; Li et al. 2006) at the advanced gametogenesis stage (stage II). Gonads completely filled with mature gametes are classified into the maturation stage (stage III), and the sex of an individual at this stage can be determined easily without histological observations. The next stage of oyster gonad development is known as spawning stage (stage IV), during which numerous primary and mature gametes still remain at the same time (Enríquez-Díaz et al. 2008). In this study, 21 samples were selected according to their representativeness of a gonadic category, with 3 individuals per sex and stage except for stage II. For convenience, s0, s1_F, s1_M, s3_F, s3_ M, s4_F, and s4_M stood for samples assigned to stage 0, stage I, stage III, and stage IV with suffix by "_F" or "_M" to distinguish female or male, respectively.

RNA Extraction

Total RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. RNA concentration was measured on a Nanodrop (Thermo Fisher Scientific). Meanwhile, RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies), and RNA degradation and contamination was monitored on 1% agarose gels. These RNA samples

conformed to the required purity criteria (A260/A230 and A260/A280 > 1.8) and quality levels (RIN > 8) for cDNA library preparation.

cDNA Library Construction and Sequencing

A total of 21 cDNA libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. At first, mRNA was purified from 3 μg total RNA per sample using poly-T oligo-attached magnetic beads. Then, the NEBNext First Strand Synthesis Reaction Buffer (5×) was added to break the mRNA into short fragments, from which first-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Subsequently, second-strand cDNA synthesis was carried out using DNA Polymerase I and RNase H. After purification, end repair, and adenylation of 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. A range of final cDNA fragments of 420 ± 20 bp was selected using AMPure XP beads and then enriched by PCR. Finally, PCR products were purified (AMPure XP system), and library quality was evaluated on the Agilent Bioanalyzer 2100 system. After cluster generation on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions, the library preparations were sequenced on an Illumina Hiseq X Ten platform and 150 bp paired-end reads were generated.

Data Analysis

Clean data were obtained through in-house perl scripts, which removed reads containing adapter, reads containing poly-N, and low-quality reads from raw data generated from Illumina sequencing. Meanwhile, Q20, Q30, and GC content of the clean data were calculated. *C. gigas* genome and gene model annotation files were downloaded from the NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/*Crassostrea_gigas*/). Index of the reference genome was built by Bowtie v2.2.3 (Langmead et al. 2009), and paired-end clean reads were then aligned to the reference genome using TopHat v2.0.12 (Trapnell et al. 2009). The reads numbers mapped to each gene were counted using HTSeq v0.6.1 (Anders et al. 2015). Then, FPKM, expected number of fragments per kilobase of transcript sequence per million base pairs sequenced (Trapnell et al. 2010), was calculated based on the length of the gene and reads count mapped to this gene. Differential expression analysis of groups between the seven different gonadic categories was performed by the DESeq R package (1.18.0) (Anders and Huber 2010) using a model based on the negative binomial distribution. To control the false discovery rate less than 0.05, the resulting *P* values were adjusted using the Benjamini and Hochberg's approach. Genes with $|\log_2(\text{FoldChange})| > 1$ and an adjusted *P* value < 0.05 found by DESeq were assigned as differentially expressed. And hierarchical cluster analysis of DEG union was performed to assess the transcriptional pattern variations among different gonadic categories using in-house R scripts. Meanwhile, heat map of genes differentially expressed was drawn by R scripts.

GO and KEGG Enrichment Analysis of Differentially Expressed Genes

Gene ontology (GO) enrichment analysis of differentially expressed genes was implemented by the GOseq R package (Young et al. 2010), in which gene length bias was corrected and *C. gigas* genome was used as the reference set. GO terms with adjusted *P* value less than 0.05 were considered significantly enriched compared with the entire genome. KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism, and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (Kanehisa et al. 2008). The statistical enrichment of differential expressed genes in KEGG pathways was tested using KOBAS software under the entire genome background (Mao et al. 2005), and KEGG pathways with corrected *P* value less than 0.05 were regarded as significantly enriched compared with genome-wide dataset.

Protein-Protein Interaction Analysis of Differentially Expressed Genes

To find these networks of protein-protein interactions and discover the hub proteins that mediate interactions with other proteins during gonad development, protein-protein interaction (PPI) analysis of differentially expressed genes was carried out based on the known interaction of proteins in the STRING database. For *C. gigas* not existing in the database, Blastx (v2.2.28) was used to align the target gene sequences to the selected reference protein sequences of *Lottia gigantean*, a marine gastropod, and then, the networks were built according to the experimental and predicted interactions with reliability string score more than 700.

Weighted Gene Correlation Network Analysis

Weighted gene correlation network analysis (WGCNA), a correlation-based method that describes and visualizes networks of data points, was implemented by the WGCNA R package (Langfelder and Horvath 2008), for identifying specific modules of co-expressed genes associated with the stage of gonad development. In the consideration of the two

mutually exclusive processes by which an undifferentiated gonad is transformed into an ovary or a testis, we applied WGCNA to investigate specific gene modules associated with the male gonad development and the female gonad development separately. The expression data within each calculated module were used to determine the module eigengenes (ME, the first principle component of the respective module) (Xue et al. 2014), which were correlated to the sample trait. By means of the softConnectivity function in WGCNA package, we selected hub genes of co-expression networks related to gonad development as they usually served as key regulators connected to a significantly large number of nodes. For real-time PCR validation, potential hub genes with sex difference in the transcript abundance took precedence.

Quantitative Real-Time PCR Validation

Forty-two gonad samples, six biological replicates at per same gonadic category with samples for RNA sequencing, were selected for quantitative real-time PCR (qRT-PCR) analysis to validate the RNA-seq. In addition, total RNA of gill, mantle, adductor muscle, and visceral mass was also extracted with six replicates to analyze the gene expression pattern in tissue. These RNA samples conformed to the required purity criteria (A260/A230 > 2.0 and A260/A280 of 1.8–2.0) for cDNA preparation, and the integrity of the RNA samples was also assessed by agarose gel electrophoresis. Then, cDNA was synthesized for qRT-PCR quantification by Prime Script TM RT Reagent Kit with gDNA Eraser (TaKaRa, Dalian, China). The amplification was performed on the LightCycler 480 real-time PCR instrument (Roche Diagnostics, Burgess Hill, UK) using SYBR® Premix Ex Taq™ (TaKaRa). Cycling parameters were 95 °C for 5 min, then 40 cycles of 95 °C for 5 s and 60 °C for 20 s. Each microplate included negative control (total RNA treated with DNase) and blank control (sterile water). Elongation factor I (EF I) gene expression was used to normalize gene expression by real-time PCR (Renault et al. 2011). GeNorm v3.5 (Vandesompele et al. 2002) was used to confirm that EF I gene was stably transcribed in all of the templates (M = 0.281). PCR efficiency and proper dilution of cDNA was determined for each primer pair by constructing a five-point standard curve from fivefold serial dilutions of a s3_F or s3_M individual template. The melting curves of PCR products (60 to 95 °C) were performed to ensure the detection of a single specific product. Relative gene expression levels were calculated by the 2^-(ΔΔCt) method (Schmittgen and Livak 2008). After being examined for homogeneity of variances, data were analyzed by t test using software SPSS 18.0 and levels were accepted as significant at P value < 0.05.

Results

Data Preprocessing and Normalization

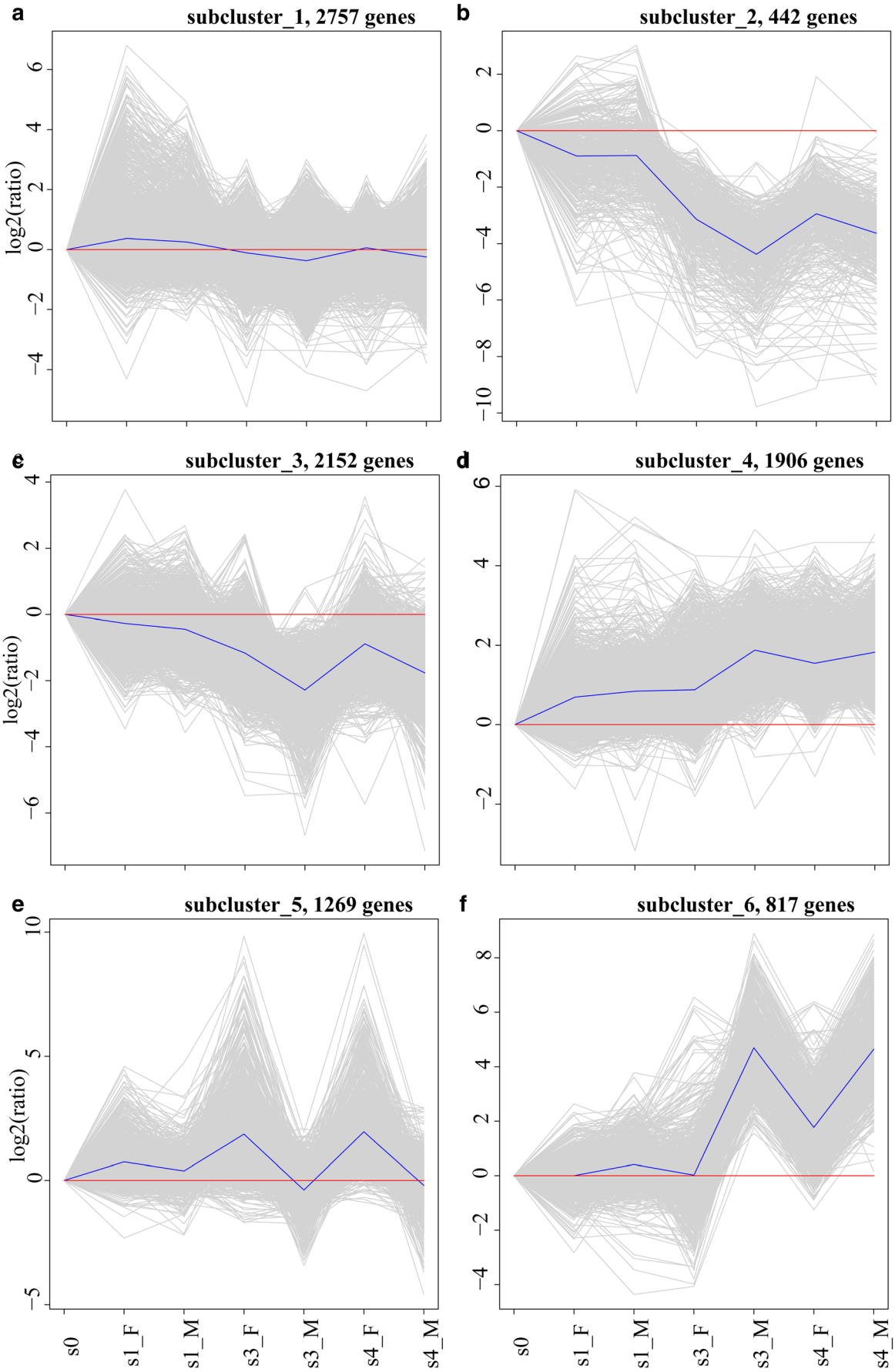
A total of 484,282,552 clean reads were filtered from 556,507,460 pair-end raw reads of 150 bp generated by Illumina sequencing, with Q20 (%) varying from 94.74 to 98.14%. Thus, the total size of the clean data generated from each sample exceeded 3.0 Gb, about six times the genome size (Supplementary Table 1). For the 21 samples, 57.07–68.37% of the total reads were aligned to the genome (Supplementary Table 2). All raw data were submitted to the Short Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/) at the National Center for Biotechnology Information (NCBI), in Bioproject PRJNA394546 under the accession number SRP112367 (Supplementary Table 3).

The abundance of all the genes (32,220) was normalized and calculated by FPKM method using HTSeq v0.6.1 software with parameter of union module. About half of the genes were considered not to be expressed or to be presented at very low levels (0 < FPKMs < 1), whereas a small proportion, less than 5%, were highly expressed (FPKM > 60) (Supplementary Table 4).

Differential Expression and Cluster Analysis

In accordance with adjusted P value < 0.05, the DESeq method identified 9343 genes differentially expressed among the seven gonadic categories of *C. gigas*. Hierarchical clustering was performed on the 21 samples for the 9343 genes based on their relative expression level $\log_2$ (ratio), and six subclusters were identified (Fig. 1). Biological interpretation of the data led us to five expression profiles: (1) 2757 genes expressed more in gonads at stage I (Fig. 1a); (2) 4594 genes expressed decreasingly over the course of oogenesis and spermatogenesis (Fig. 1b, c); (3) 1906 genes expressed increasingly over the course of oogenesis and spermatogenesis (Fig. 1d); (4) 1269 genes expressed specifically in female gonads (Fig. 1e); and (5) 817 genes expressed more in male gonads with an increasing expression over the course of spermatogenesis (Fig. 1f). Interestingly, three reported key genes, *CgDsx*, *CgFoxL2*, and

Fig. 1 Subclusters generated by hierarchical clustering based on genes differentially expressed. Every subgraph exhibits the expression level of genes in a subcluster, in which the *X*-axis represents seven gonadic categories and the *Y*-axis represents relative expression of genes. Every gray line in every subgraph is the line chart of relative expression of a specific gene during different gonadic categories, while the blue line stands for the mean relative expression of all genes in the subcluster and the red line represent the baseline. (1) Two thousand fifty-seven genes expressed more in gonads at stage I (**a**); (2) 4594 genes expressed decreasingly over the course of oogenesis and spermatogenesis (**b**, **c**); (3) 1906 genes expressed increasingly over the course of oogenesis and spermatogenesis (**d**); (4) 1269 genes expressed specifically in female gonads (**e**); (5) 817 genes expressed more in male gonads with an increasing expression over the course of spermatogenesis (**f**)


a
subcluster_1, 2757 genes
b
subcluster_2, 442 genes
c
subcluster_3, 2152 genes
d
subcluster_4, 1906 genes
e
subcluster_5, 1269 genes
f
subcluster_6, 817 genes
log2(ratio)
s0
s1_F
s1_M
s3_F
s3_M
s4_F
s4_M

CgSoxH, were assigned to subcluster_1 (Fig. 1a), subcluster_5 (Fig. 1e), and sucluster_6 (Fig. 1f), respectively. Heat map based on $\log_{10}$(FPKM + 1) of DEGs revealed that hierarchical clustering grouped the samples into three clusters: gonads at early developmental stages (s0, s1_F, and s1_M), male gonads at maturation and spawning stages (s3_M and s4_M), and female gonads at maturation and spawning stages (s3_F and s4_F) (Fig. 2).

Enrichment Analysis of DEGs and Subclusters Generated by Hierarchical Clustering

GO enrichment analysis was performed to identify the significantly enriched GO terms based on DEGs in each pair-wise comparison among seven gonadic categories, but there was no significantly enriched GO term showing consistent. KEGG enrichment analysis was implemented to discover the metabolic processes and signal transduction pathways, which were significantly enriched compared with the entire genome during gonad development. Besides "DNA replication," "Mismatch repair," and "Ribosome," a number of pathways significantly enriched were shared by some pair-wise comparisons (Supplementary Table 5). In most comparisons, "FoxO signaling pathway" was significantly enriched (Supplementary Table 5), which takes part in cellular physiological events including apoptosis, cell cycle control, glucose metabolism, oxidative stress resistance, and longevity. In addition, when s0 compared with others, s1_M compared with s3_M, s1_F compared with s3_F, and s3_F compared with s3_M, "ECM-receptor interaction pathway" showed

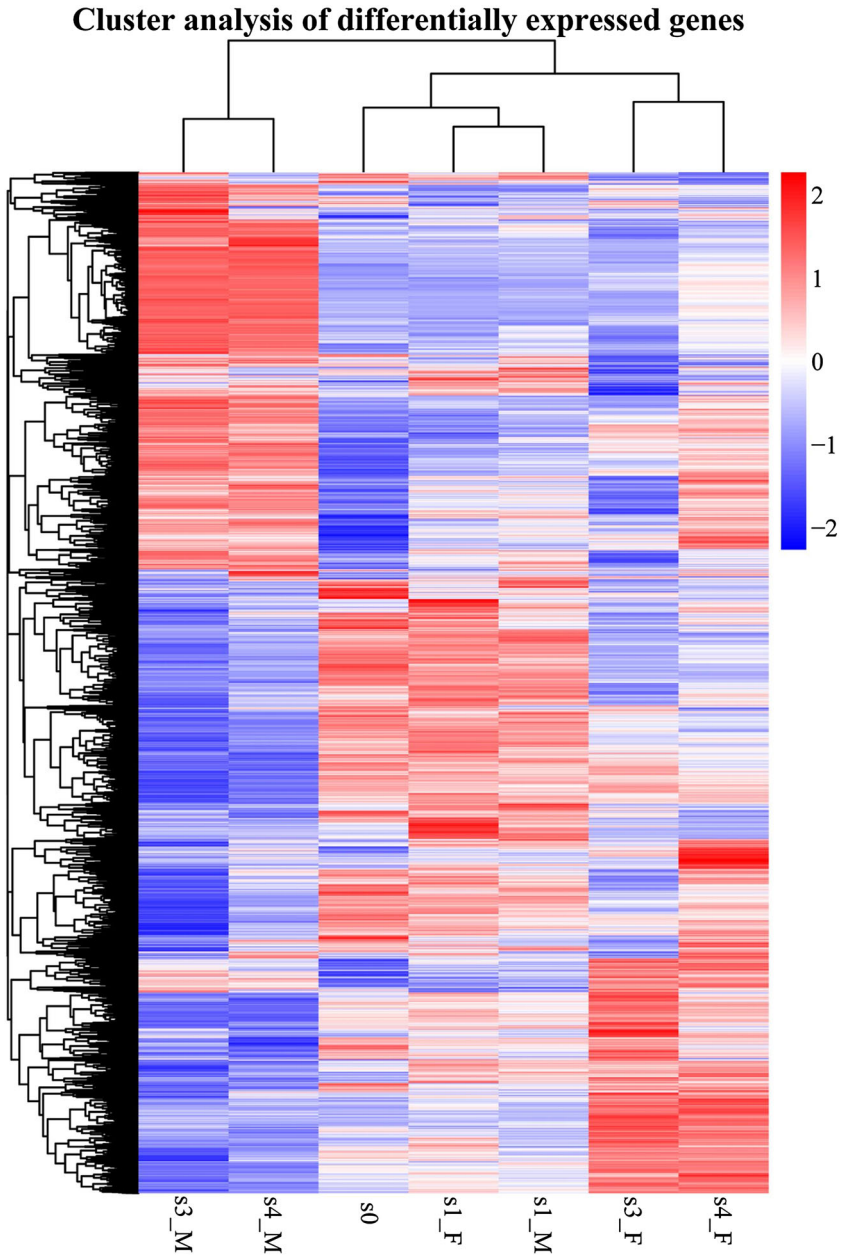


Fig. 2 Heat map of genes differentially expressed between the seven different gonadic categories of *C. gigas*. In the heat map based on $\log_{10}$(FPKM + 1) of genes differentially expressed, genes showing similar expression profiles on samples clustered together. Expression levels are depicted with a color scale, in which shades of red represent higher expression and shades of blue represent lower expression. Three major clusters of samples were determined: early developmental stages (s0, s1_F, and s1_M), male gonads in maturation stage and spawning and reabsorbing stage (s3_Mand s4_M), and female gonads at maturation stage and spawning and reabsorbing stage (s3_F and s4_F)

significantly enriched (Supplementary Table 5), of which interactions lead to a direct or indirect control of cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis. What drew our attention is that "neuroactive ligand-receptor interaction pathway" was significantly enriched in several comparisons (Supplementary Table 5), and this indicates the regulation of nervous system in sex determination and differentiation of *C. gigas*.

To investigate the biological roles of the genes clustered into specific subclusters generated by hierarchical clustering, GO and KEGG enrichments analyses were also implemented under the entire genome background. According to the GO annotations (Fig. 3), the subcluster_1 was significantly enriched with genes functioning in "oxidoreductase activity," "monovalent inorganic cation transmembrane transporter activity," "lipid metabolic process," and so on. These biological activities were closely associated with metabolic process of *C. gigas*. The GO enrichment for subcluster_5 revealed that this subcluster was enriched with genes functioning in "neurotransmitter transporter activity," "neurotransmitter/sodium symporter activity," and "neurotransmitter transport." The top three GO terms significantly enriched compared with the complete genome in subcluster_6 were "microtubule motor activity," "microtubule-based movement," and "movement of cell or subcellular component." KEGG enrichment analysis showed that "Oxidative phosphorylation," "Citrate cycle (TCA cycle)," and "Carbon metabolism and Glycerophospholipid metabolism" were top enriched terms in subcluser_1. "Arachidonic acid metabolism," "Nucleotide excision repair," "Taurine and hypotaurine metabolism," and "DNA replication" were significantly enriched in subcluster_5, while "Ubiquitin mediated proteolysis," "Purine metabolism," and "FoxO signaling pathway" were enriched in subcluster_6 (Fig. 4).

Protein-Protein Interaction Analysis of Differentially Expressed Genes

With the setting that 700 as the minimum of string score to select the reliable interactions, some networks of PPIs were built in nine pair-wise comparisons, including s0 vs s3_M, s0 vs s3_F, s0 vs s4_M, s0 vs s4_F, s1_F vs s3_F, s1_M vs s3_M, s1_M vs s4_M, s3_F vs s3_M, and s4_F vs s4_M. Among these networks, most of the hub proteins were related to DNA replicate and cell division, such as DNA replication licensing factor mcm4-like, DNA polymerase delta catalytic subunit like, and mitotic spindle assembly checkpoint protein MAD2A-like (Supplementary Table 6).

Weighted Gene Correlation Network Analysis

Based on the gene expression data of 12 samples assigned to s0, s1_M, s3_M, and s4_M, we identified 16 distinct co-expression modules (Fig. 5a). Notably, the turquoise module showed highest statistical significance between ME and the male gonad development stage. Based on the module sizes, we selected the top 2% genes with the highest connectivity in the module as hub genes, which may be more biologically significant than others. Four hub genes in the turquoise module, lncRNA LOC105321313, uncharacterized LOC105332100, SH3 domain-containing kinase-binding protein 1 (*Cg-Sh3kbp1*), and uncharacterized LOC105345697, were chosen for real-time qPCR validation as they also belong to the subcluster_6 generated by hierarchical clustering. Similarly, 19 modules were identified using the gene expression data of 12 samples assigned to s0, s1_F, s3_F, and s4_F (Fig. 5b). Two modules, black module and dark red module, were detected that may represent pathways positively associated with the female gonad development, and the hub genes

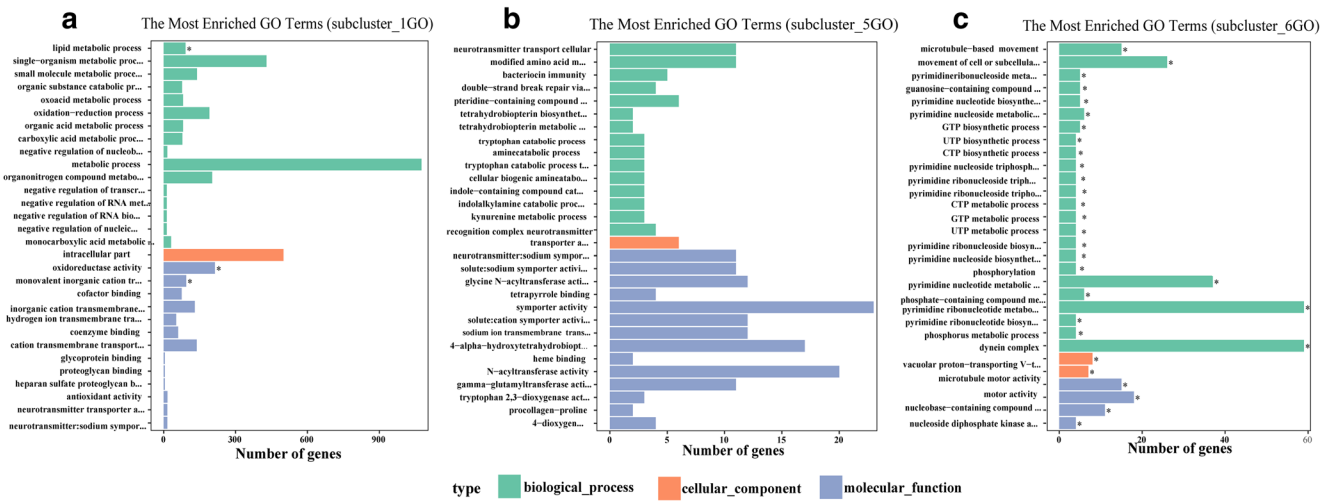


Fig. 3 Gene ontology (GO) annotation of genes in the subcluster_1, subcluster_5, and subcluster_6 generated by hierarchical clustering. The GO terms in each panel were organized by decreasing *P* value within each GO category. Bar chart of genes enriched in GO term, it can directly reflect the number of genes distributing into different GO terms

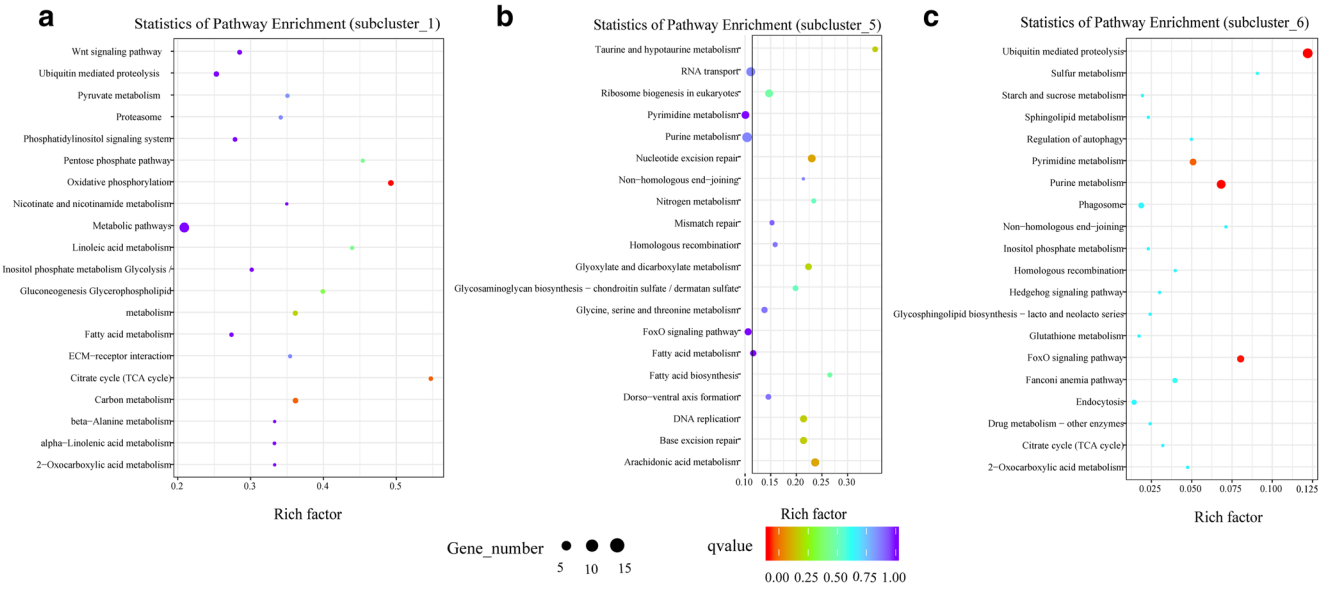


Fig. 4 KEGG annotation of genes in the subcluster_1, subcluster_5, and subcluster_6 generated by hierarchical clustering. "Rich factor" means that the ratio of the number of the genes in the specific subcluster and the number of genes annotated in this pathway. The greater the rich factor, the greater the degree of enrichment

were filtered in a similar way. Four hub genes also assigned to the subcluster_5, MAM and LDL-receptor class A domain-containing protein 1-like (*Cg-Malrd1-like*), uncharacterized LOC105336037, *Cg-caveolin-1*, and uncharacterized LOC105340741, were selected for real-time qPCR validation.

Functional Annotation of Target Modules

To further understand the biological roles of modules positively related to the gonad development of *C. gigas*, GO and KEGG enrichments were performed under the entire genome background. GO enrichment (Table 1) showed that the turquoise module related to male gonad development was significantly enriched with genes functioning in "intracellular part," "microtubule motor activity," and "organelle and microtubule-based process," which showed similarity with subcluster_6. "Organelle," "intracellular organelle," and "intracellular part" were the top 3 GO terms enriched in the black module that associated with the female gonad development. In the dark red module, "single-organism metabolic process," "threonine-type endopeptidase activity," and "threonine-type peptidase activity" were significantly enriched. According to KEGG pathway enrichment (Table 2), "DNA replication," "Ubiquitin mediated proteolysis," "Fatty acid degradation," and "Mismatch repair" were enriched in the turquoise module. In the black module that associated with the female gonad development, "Base excision repair," "DNA replication," "Protein export," and "Spliceosome" were significantly enriched. And pathway enrichment also revealed that

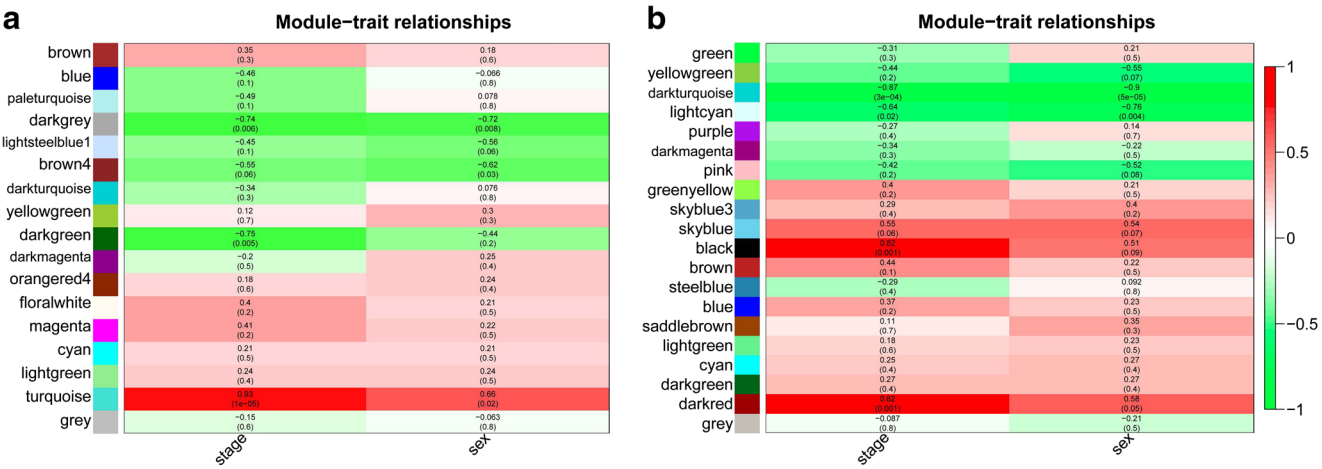


Fig. 5 Module-trait associations. **a**, **b** The module-stage associations of male and female gonadic samples, respectively. Each row corresponds to a module eigengene, and column to a trait. Each cell contains the corresponding correlation and *P* value. The table is color-coded by correlation according to the color legend

Table 1 Top-rank functional annotations enriched in gonad development-related modules

Module	GO accession	Description	Term type	*P* value
Turquoise	GO:0044424	Intracellular part	Cellular component	4.39E−09
	GO:0003777	Microtubule motor activity	Molecular function	3.89E−08
	GO:0043226	Organelle	Cellular component	9.93E−08
Black	GO:0043226	Organelle	Cellular component	3.91E−10
	GO:0043229	Intracellular organelle	Cellular component	3.91E−10
	GO:0044424	Intracellular part	Cellular component	5.33E−09
Dark red	GO:0044710	Single-organism metabolic process	Biological process	1.64E−05
	GO:0004298	Threonine-type endopeptidase activity	Molecular function	1.64E−05
	GO:0070003	Threonine-type peptidase activity	Molecular function	1.64E−05

"DNA replication," "Proteasome," and "Fatty acid metabolism" were enriched in the dark red module.

Quantitative Real-Time PCR Validation

For real-time qPCR validation, eight sex-specific genes described above were selected to explore their temporal and spatial expression (Table 3). The eight genes exhibited opposite expression levels depending on sex: increasing expression over the course of gametogenesis in males (four genes from the turquoise module) and in females (four genes from the dark red module). The patterns of transcript abundance detected for these genes in the RNA-seq and in real-time qPCR showed extremely similar profiles, and only four of the eight newly discovered genes express in gonad specifically (Fig. 6). Long non-coding RNA LOC105321313, uncharacterized LOC105345697, and *Cg-Sh3kbp1* (LOC105340479) appeared significantly more expressed in testis, and *Cg-Malrd1-like* (LOC105320617) expression was the highest in ovaries.

Functional Annotation of Genes Co-Expressed with Hub Genes

To understand the biological role of the four above hub genes, GO and KEGG enrichment were implemented based on co-expressed genes under the entire genome background. For genes co-expressed with lncRNA LOC105321313, uncharacterized LOC105345697, and *Cg-Sh3kbp1*, GO enrichment analysis revealed that "microtubule motor activity," "coenzyme A biosynthetic process," and "nucleoside bisphosphate biosynthetic process" were the top 3 terms significantly enriched in these gene sets compared with the entire genome (Supplementary Table 7). KEGG enrichment analysis showed overrepresentation of the glycolysis and gluconeogenesis pathway (Supplementary Table 8). For genes co-expressed with *Cg-Malrd1-like*, GO enrichment analysis revealed that "orotidine-5′-phosphate decarboxylase activity" and "*de novo* pyrimidine nucleobase biosynthetic process" significantly enriched (Supplementary Table 7), and KEGG enrichment analysis showed that the DNA replication and vitamin B6 metabolism were the most enriched pathways (Supplementary Table 8).

Discussion

Differential expressed genes among seven gonadic categories were clustered into six subclusters by hierarchical clustering, of which subcluster_5 and subcluster_6 (Fig. 1e, f) showed sex difference at transcriptional level. GO enrichment analysis revealed that many genes in the female-specific subcluster_5 participated in neurotransmitter transporter activity, suggesting that neurotransmitters were likely to play important roles in *C. gigas*

Table 2 Top enriched pathways in gonad development-related modules

Module	Term	ID	Overlap	Total entities	*P* value
Turquoise	DNA replication	crg03030	32	37	5.89E−07
	Ubiquitin-mediated proteolysis	crg04120	59	122	3.33E−05
	Fatty acid degradation	crg00071	21	38	0.003573204
Black	Base excision repair	crg03410	14	37	0.001217305
	DNA replication	crg03030	13	37	0.003059519
	Protein export	crg03060	9	21	0.004715896
Dark red	DNA replication	crg03030	25	37	2.30E−06
	Proteasome	crg03050	19	38	0.00069242
	Fatty acid metabolism	crg01212	22	51	0.001268431

Table 3 Primers of genes selected for real-time qPCR

Gene symbols	Gene description	Primer (5′ to 3′)	Amplicon lengths (bp)	Amplification efficiencies
LOC105321313	lncRNA LOC105321313	F: TCCATGTACGCAAACAAGAAGAC R: AACTGTGATGAGGACGAGAACG	165	1.956
LOC105340479	SH3 domain-containing kinase-binding protein 1	F: GCTGGAGGTGGATGATAGGG R: GGAACTCACAAAGGGAGACAACT	136	1.947
LOC105332100	Uncharacterized LOC105332100	F: TGCGATTTAATGACCATGACGAG R: ACGGGTTGATGTTCTGATTCCAG	197	1.920
LOC105345697	Uncharacterized LOC105345697	F: GTTTCCCACAATTCGTATTTCAAG R: GTCTGTCAGTTCTTGGCGTCTC	152	1.949
LOC105320617	MAM and LDL-receptor class A domain-containing protein 1-like	F: GAGAAGCCAGAATCGTCCTAAAC R: CGCTCTAACTCCCGGTCTATCAT	130	2.076
LOC105336037	Uncharacterized LOC105336037	F: CCACCTATCCTCACTCGTTTAGC R: GACGTACAACCATACCACCACAT	139	1.996
LOC105339682	*caveolin-1*	F: TATGCTACAAACTGATGACGCTGT R: CAGAACCTTGAACATCGGGG	119	2.018
LOC105340741	Uncharacterized LOC105340741	F: TACACTGTTGCCTAACGCCTTCT R: TCCGTGCATGTCTTATCATCTTTC	144	2.012

ovary development. Three neurotransmitters hydroxy tryptamine (5-HT), dopamine (DA), and norepinephrine (NE) distributed in the reproductive coat, cytoplasm, and nuclear membrane of the oyster oocyte at different stages (Fang et al. 2003). This suggested that nervous system might regulate the development and maturation of oocyte in *C. gigas*, which was consistent with *P. yessoensis* (Li et al. 2016). KEGG enrichment analysis displayed that taurine and hypotaurine metabolism pathway was significantly enriched in subcluster_5, corresponding with general knowledge that taurine was essential for function of the nervous system and hypotaurine served as the precursor of taurine. All of these supported the opinion that reproduction of bivalves was a genetically controlled response to the external environment, regulated by endogenous hormones and neurotransmitters (Morishita et al. 2010; Li et al. 2016). GO enrichment of genes assigned to the male-specific subcluster_6 displayed overrepresentation of microtubule motor activity, and this was expected to be relevant to the development and function of flagella through which sperm cells move. In accordance with GO enrichment analysis, KEGG enrichment indicated that male-specific gene sets were primarily involved protein metabolism, such as ubiquitin-mediated proteolysis, which might provide materials for sperm flagella formation. It seemed that there was no clue to the participation of neurotransmitters in testis development.

CgFoxL2 and *CgSoxH*, related to sex determination and differentiation of *C. gigas*, were assigned to subcluster_5 and subcluster_6, respectively. Genes clustered in the same cluster with the two genes might play similar roles. Even so, it is still difficult to discover new candidate genes for a large gene base and unclear relationship. To take this further, we identified modules of co-expressed genes using WGCNA in the setting that gonad development stage as sample trait. Enrichment analysis of the three modules revealed similar gene function with corresponding subclusters generated by hierarchical clustering, which supported the involvement of neurotransmitters in *C. gigas* reproductive regulation.

To narrow the range of potential key genes in these WGCNA modules, gene connectivity and the transcriptional difference between the genders, especially at early stage development, were both considered. In fact, *CgDsx* and *CgFoxL2* expressed more in male and female, respectively, at stage I by the comparison of s1_F vs s1_M without a withdrawn sample. This confirmed that genes participating in the pathway of sex determination showed sex difference at early gonad development. After qRT-PCR validation, two hub genes in the black module and one in the dark red module caught our attention and their possible functions were discussed preliminarily.

Long Non-Coding RNA LOC105321313

Non-coding RNAs (ncRNAs) are functional RNA molecules that are not translated into protein and have been implicated in some of the most studied complex epigenetic phenomena (Piferrer 2013; Zhou et al. 2010; Costa 2008). ncRNAs belong to many groups and play their biological roles through complex mechanism in cellular process. Generally, ncRNAs regulate gene transcription by the directed recruitment of epigenetic silencing complexes to homology-containing loci in the genome (Piferrer 2013; Chuang and Jones 2007). Long non-coding RNAs (lncRNAs) with nucleotide length > 200 bp have been implicated in sex determination and gonadogenesis in animal kingdom, such as roX and XIST participating in dosage compensation in *Drosophila* and *Mus musculus*, respectively. In *C. gigas*, the existence of *Cg-Foxl2os*, a natural

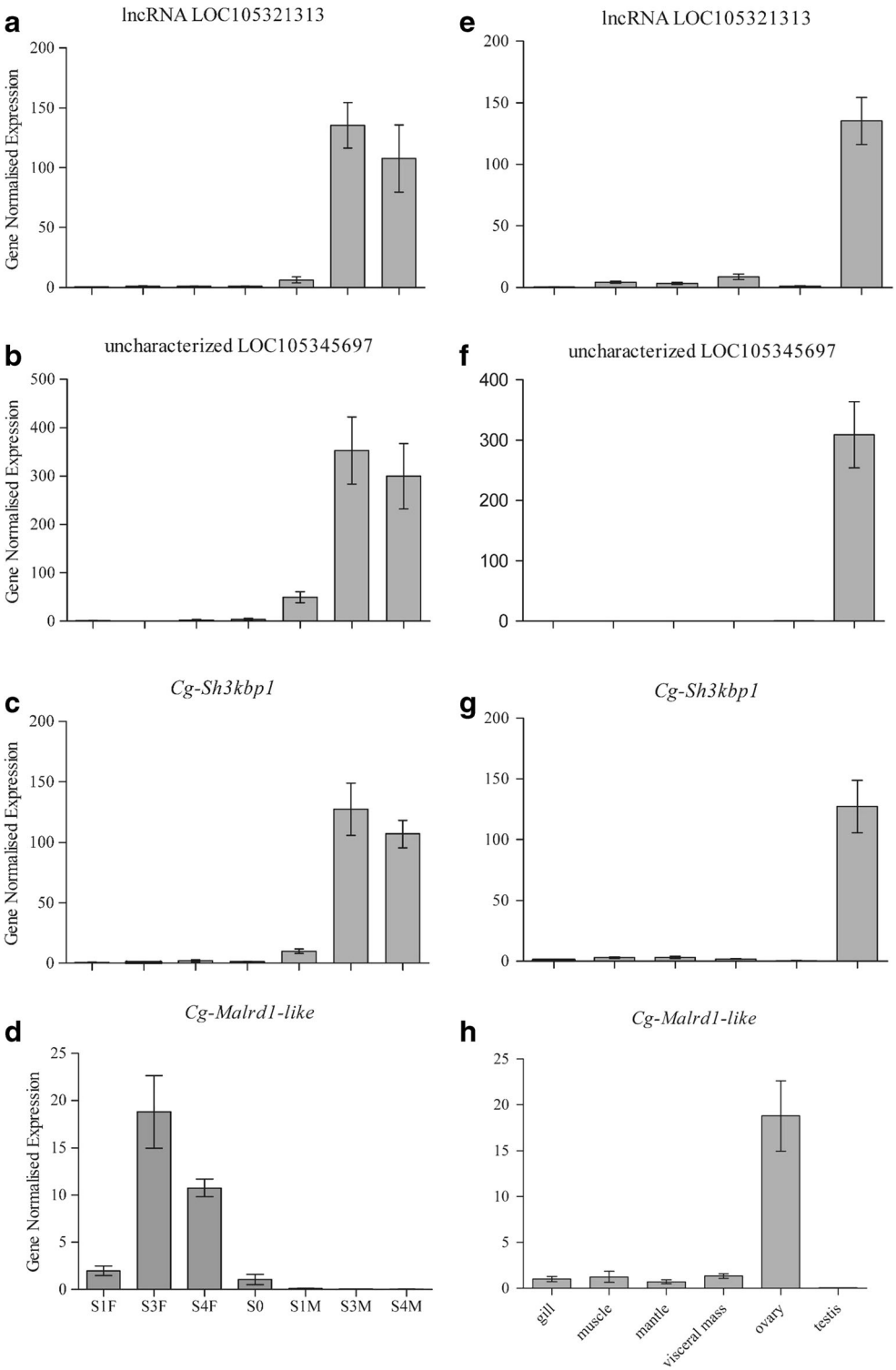


Fig. 6 Expression profiles of the four gonad specific genes. The temporal and spatial expression of four genes was measured by real-time qPCR. These histograms **a**, **b**, **c**, and **d** represent mean transcript expression during different gonadic categories, and **e**, **f**, **g** and **h** represent mean transcript expression in different tissues. In the development stage, sex and tissue are indicated at the bottom of corresponding figure. Bars represent standard error. Statistical analyses were carried out by *t* test using software SPSS 18.0, and levels were accepted as significant at P value < 0.05. Real-time PCR indicated significant differences of four gene expressions between sexes for stage III and stage IV. **a**–**c** lncRNA LOC105321313, uncharacterized LOC105345697, and *Cg-Sh3kbp1* increasing transcript expression along the spermatogenesis with expression becoming significantly higher at stage III and stage IV compared with stage 0 and stage I. **d** *Cg-Malrd1-like* increasing transcript expression along the oogenesis with expression becoming significantly higher at stage III and stage IV compared with stage 0 and stage I. **e**, **g** Significantly higher expression of lncRNA LOC105321313, uncharacterized LOC105345697, and *Cg-Sh3kbp1* in testis compared with other tissues. **h** Significantly higher expression of *Cg-Malrd1-like* in ovary compared with other tissues

antisense transcript (NAT) of *Cg-Foxl2* belonging to the lncRNA category, has been demonstrated in the testis. It may regulate *Cg-Foxl2* expression through formation of cytoplasmic RNA-RNA duplexes during the oyster gonadic differentiation (Santerre et al. 2012).

In this study, the testis-specific lncRNA LOC105321313 displayed sexually dimorphic expression during early gonad differentiation of *C. gigas*, which was verified by qRT-PCR. As there is no information about the lncRNA LOC105321313 in ncRNA database, such as CHIPBase and lncRNABase, it is challenging to investigate its function. To initially understand

the biological role of this gene, GO and KEGG enrichment were implemented based on genes co-expressed with lncRNA LOC105321313 in the black module positively related to testis development. The top enriched GO terms compared with the genome-wide dataset were necessary for spermatic flagellum and movement, and the glycolysis and gluconeogenesis pathway was significantly enriched, which was deemed to produce ATP for sperm movement. This indicated that co-expressed genes involved in spermatogenesis and the lncRNA LOC105321313 might serve as regulator lying in the center of co-expression network. Based on the above analysis, the testis-specific lncRNA LOC105321313 was regarded as a potential regulator during spermatogenesis through epigenetic modification in gene expression and the concrete mechanism should be pursued. In view of the irregular sex change of *C. gigas*, epigenetic research may make good starting points for further understanding on sex determinism as epigenetic mechanisms provide organisms with the ability to integrate genomic and environmental information to modify the activity of their genes for generating a particular phenotype (Piferrer 2013).

SH3 Domain-Containing Kinase-Binding Protein 1

The *Cg-Sh3kbp1* (SH3 domain-containing kinase-binding protein 1) gene encodes an adapter protein that contains three Src homology domains, a proline rich region, and a C-terminal coiled-coil domain. SH3 domains enable protein to play versatile and diverse roles in the cell including the regulation of enzymes, changing the subcellular localization of signaling pathway components, and mediating the formation of multiprotein complex assemblies (Havrylov et al. 2010; Yao et al. 2007; Philippe et al. 2011). Furthermore, the *Cg-Sh3kbp1* protein contains an additional domain matched with cell division factor ZapB that is required for proper Z-ring formation (Ebersbach et al. 2010), which indicates that it may take part in cell proliferation.

Though some genes of Src Homology 3 domain superfamily locate in the X chromosome in many mammals, such as human and house mouse, there is no sex difference in transcription abundance. Interestingly, the testis-specific *Cg-Sh3kbp1* gene of *C. gigas* showed an increasing expression over the course of spermatogenesis. The conserved domains of *Cg-Sh3kbp1* revealed potential roles in diverse signaling pathways, mediating the formation of multiprotein complex assemblies, and cell proliferation. In addition, the result of function enrichment analysis based on the co-expressed genes was similar with genes co-expressed with the lncRNA LOC105321313, which also revealed participation in spermatogenesis. It seemed to support an assumption that this adapter protein involved in regulating uncharted signal transduction pathways during spermatogenesis or mediating the formation of sperm cell microtubule complex assemblies.

MAM and LDL-Receptor Class A Domain-Containing Protein 1-Like

The *Cg-Malrd1-like* (MAM and LDL-receptor class A domain-containing protein 1-like) gene of *C. gigas* encodes a membrane protein that contains three MAM domains and one CUB domain. MAM is an extracellular domain which mediates protein-protein interactions and is found in a diverse set of proteins, many of which are known to function in cell adhesion (Bixby 2003; Aricescu et al. 2014). CUB domain, an extracellular domain, presents in proteins mostly known to be involved in development, such as spermadhesins (Bork and Beckmann 1993; Töpfer-Petersen et al. 2009). These developmentally regulated proteins are involved in a diverse range of functions, including complement activation, developmental patterning, tissue repair, axon guidance and angiogenesis, cell signaling, fertilization, hemostasis, inflammation, neurotransmission, receptor-mediated endocytosis, and tumor suppression.

These conserved domains of *Cg-Malrd1-like* clued us in an assumption that the encoded protein might serve as oogonia or oocyte membrane protein mediating diverse signal transduction to regulate oocyte development. Furthermore, the two enriched GO terms in gene sets co-expressed with *Cg-Malrd1-like* seemed to be absolutely necessary for nucleic acid biosynthesis. Consistent with the GO enrichment result, KEGG enrichment analysis also showed overrepresentation of the DNA replication. All this indicated that the *Cg-Malrd1-like* gene might mediate signal-cell interactions to activate the oogonia or oocyte DNA replication process. This prediction is tentative, and we need more experimental evidence to investigate the role of *Cg-Malrd1-like* gene.

Conclusions

The most significant outcome of this study is the identification of gene modules that related to gonad development of *C. gigas*, which produce lists of genes for further studies. In addition, the function analysis of gene sets provides clues for the role of amino acid neurotransmitters in oyster ovary development. Three new hub genes have been predicted as participators in the gonad development process. Among these candidate genes, the lncRNA LOC105321313 aroused us the interest about the non-coding RNA regulation during sex determination and differentiation process of *C. gigas*. These results would be important resources for future research.

Acknowledgements This work was supported by the grants from National Natural Science Foundation of China (31772843, 31672649), Shandong Province (2016ZDJS06A06), Fundamental Research Funds for the Central Universities (201762014), and Major Project for Tianjin Seed Technology (15ZXZYNC00050).

Author Contributions CY carried out the molecular genetic studies, participated in the data analysis, and drafted the manuscript. QL conceived of the study, participated in experimental design and coordination, and contributed to the manuscript preparation. HY participated in the data analysis. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Anders S, Pyl PT, Huber W (2015) HTSeq—a python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169

Aricescu AR, Hon WC, Siebold C, Lu W, Pa VDM, Jones EY (2014) Molecular analysis of receptor protein tyrosine phosphatase mu-mediated cell adhesion. EMBO J 25:701–712

Bixby JL (2003) Receptor protein tyrosine phosphatases as mediators of cellular adhesion. Front Biosci 8:d87–d99

Bork P, Beckmann G (1993) The CUB domain. A widespread module in developmentally regulated proteins. J Mol Biol 231(2):539–545

Chuang JC, Jones PA (2007) Epigenetics and microRNAs. Pediatr Res 61:24R–29R

Coe WR (1943) Sexual differentiation in mollusks. I. Pelecypods. Q Rev Biol 18:154–164

Costa FF (2008) Non-coding RNAs, epigenetics and complexity. Gene 410:9–17

Dheilly NM, Lelong C, Huvet A, Kellner K, Dubos MP, Riviere G, Boudry P, Favrel P (2012) Gametogenesis in the Pacific oyster *Crassostrea gigas*: a microarrays-based analysis identifies sex and stage specific genes. PLoS One 7:e36353

Ebersbach G, Galli E, Møllerjensen J, Löwe J, Gerdes K (2010) Novel coiled-coil cell division factor ZapB stimulates Z ring assembly and cell division. Mol Microbiol 68:720–735

Enríquez-Díaz M, Pouvreau S, Chávez-Villalba J, Le Pennec M (2008) Gametogenesis, reproductive investment, and spawning behavior of the Pacific giant oyster *Crassostrea gigas*: evidence of an environment-dependent strategy. Aquac Int 17:491–506

Fang Q, Fang Y, Weng Y, Dai Y (2003) Distribution of neurotransmitters and regulatory peptide immunoreactive substances in ovary of *Crassostrea gigas*. J Oceanogr Taiwan Strait 22:299–302

Guo X, Hedgecock D, Hershberger WK, Cooper K, Allen SK (1998) Genetic determinants of protandric sex in the Pacific oyster, *Crassostrea gigas* Thunberg. Evolution 52:394–402

Havrylov S, Redowicz MJ, Buchman VL (2010) Emerging roles of Ruk/CIN85 in vesicle-mediated transport, adhesion, migration and malignancy. Traffic 11:721–731

Hedrick PW, Hedgecock D (2010) Sex determination: genetic models for oysters. J Hered 101:602–611

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480–D484

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Li Y, Zhang L, Sun Y, Ma X, Wang J, Li R, Zhang M, Wang S, Hu X, Bao Z (2016) Transcriptome sequencing and comparative analysis of ovary and testis identifies potential key sex-related genes and pathways in scallop *Patinopecten yessoensis*. Mar Biotechnol 18:453–465

Li Q, Liu W, Shirasu K, Chen W, Jiang S (2006) Reproductive cycle and biochemical composition of the Zhe oyster *Crassostrea plicatula* Gmelin in an eastern coastal bay of China. Aquaculture 261:752–759

Liu XL, Zhang ZF, Shao MY, Liu JG, Muhammad F (2012) Sexually dimorphic expression of foxl2 during gametogenesis in scallop *Chlamys farreri*, conserved with vertebrates. Dev Genes Evol 222:279–286

Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21:3787–3793

Morishita F, Furukawa Y, Matsushima O, Minakata H (2010) Regulatory actions of neuropeptides and peptide hormones on the reproduction of molluscs. Can J Zool 88:825–845

Munger SC, Capel B (2012) Sex and the circuitry: progress toward a systems-level understanding of vertebrate sex determination. Wiley Interdiscip Rev Syst Biol Med 4:401–412

Naimi A, Martinez AS, Specq ML, Diss B, Mathieu M, Sourdaine P (2009a) Molecular cloning and gene expression of Cg-Foxl2 during the development and the adult gametogenetic cycle in the oyster *Crassostrea gigas*. Comp Biochem Physiol B Biochem Mol Biol 154:134–142

Naimi A, Martinez AS, Specq ML, Mrac A, Diss B, Mathieu M, Sourdaine P (2009b) Identification and expression of a factor of the DM family in the oyster *Crassostrea gigas*. Comp Biochem Physiol A Mol Integr Physiol 152:189–196

Philippe D, Ababou A, Yang X, Ghosh R, Daviter T, Ladbury JE, Pfuhl M (2011) Making ends meet: the importance of the N- and C-termini for the structure, stability, and function of the third SH3 domain of CIN85. Biochemistry 50:3649–3659

Piferrer F (2013) Epigenetics of sex determination and gonadogenesis. Dev Dyn 242:360–370

Renault T, Faury N, Barbosa-Solomieu V, Moreau K (2011) Suppression substractive hybridisation (SSH) and real time PCR reveal differential gene expression in the Pacific cupped oyster, *Crassostrea gigas*, challenged with *Ostreid herpesvirus 1*. Dev Comp Immunol 35:725–735

Santerre C, Sourdaine P, Martinez AS (2012) Expression of a natural antisense transcript of Cg-Foxl2 during the gonadic differentiation of the oyster *Crassostrea gigas*: first demonstration in the gonads of a lophotrochozoa species. Sex Dev 6:210–221

Santerre C, Sourdaine P, Adeline B, Martinez AS (2014) Cg-SoxE and Cg-β-catenin, two new potential actors of the sex-determining pathway in a hermaphrodite lophotrochozoan, the Pacific oyster *Crassostrea gigas*. Comp Biochem Physiol A Mol Integr Physiol 167:68–76

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3:1101–1108

Teaniniuraitemoana V, Huvet A, Levy P, Klopp C, Lhuillier E, Gaertner-Mazouni N, Gueguen Y, Le Moullac G (2014) Gonad transcriptome analysis of pearl oyster *Pinctada margaritifera*: identification of potential sex differentiation and sex determining genes. BMC Genomics 15:491

Töpfer-Petersen E, Romero A, Varela PF, Ekhlasi-Hundrieser M, Dostàlovà Z, Sanz L, Calvete JJ (2009) Spermadhesins: a new protein family. Facts, hypoteses, and perpectives. Andrologia 30:217–224

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515

Vandesompele J, Preter KD, Pattyn F, Poppe B, Roy NV, Paepe AD, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:research0034

Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, De Nardo D, Gohel TD, Emde M, Schmidleithner L, Ganesan H (2014) Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity 40(2):274–288

Yao B, Zhang J, Dai H, Sun J, Jiao Y, Tang Y, Wu J, Shi Y (2007) Solution structure of the second SH3 domain of human CMS and a newly identified binding site at the C-terminus of c-Cbl. Biochim Biophys Acta 1774:35–43

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11:R14

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z (2012) The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490:49–54

Zhang N, Xu F, Guo X (2014) Genomic analysis of the Pacific oyster (*Crassostrea gigas*) reveals possible conservation of vertebrate sex determination in a mollusc. G3 (Bethesda) 4:2207–2217

Zhou H, Hu H, Lai M (2010) Non-coding RNAs and their epigenetic regulatory mechanisms. Biol Cell 102:645–655