Functional & Integrative Genomics
https://doi.org/10.1007/s10142-018-0595-y

ORIGINAL ARTICLE

Transcriptomics analysis revealing candidate networks and genes for the body size sexual dimorphism of Chinese tongue sole (*Cynoglossus semilaevis*)

Na Wang [1,2] · Renkai Wang [1,3] · Ruoqing Wang [1,3] · Songlin Chen [1,2]

Received: 11 August 2017 / Revised: 14 February 2018 / Accepted: 16 February 2018

Abstract

The Chinese tongue sole (*Cynoglossus semilaevis*) is a typical female heterogamete species that exhibits female-biased sexual size dimorphism, which has severely hindered the sustainable development of the species in aquaculture. In the present study, four important somatotropic and reproductive tissues including brain, pituitary, liver, and gonad from 15 females and 15 males were used for transcriptome analysis via RNA-seq. A mean of 37,533,991 high-quality clean reads was obtained from each library and 806, 1482, 818, and 14,695 differentially expressed genes in female and male were identified from the brain, pituitary, liver, and gonad, respectively (fold change ≥ 2 and $q < 0.05$). Enrichment analyses of GO terms and KEGG pathways showed that nucleic acid-binding transcription factor activity, G-protein-coupled receptor activity, MAPK signaling pathway, steroid biosynthesis, and neuroactive ligand-receptor interaction may be involved in the sexual growth differences. Furthermore, via weighted gene co-expression network analyses, two modules (yellowgreen and salmon4) were identified to be significantly positive-correlated with female-biased sexual size dimorphism. An illustrated network map drawn by these two modules enabled the identification of a series of hub genes, including nipped-B-like protein A (*nipbla*), transcriptional activator protein Pur-beta-like (*purb*), and BDNF/NT-3 growth factors receptor (*ntrk2*). Detailed functional investigation of these networks and hub genes will further improve our understanding of the underlying molecular mechanism of sexual size dimorphism in fish.

Keywords Chinese tongue sole (*Cynoglossus semilaevis*) · Sexual size dimorphism · Transcriptome · Weighted gene co-expression network analysis · Steroid biosynthesis

Introduction

As a common phenomenon in metazoans, sexual dimorphism is characterized by differences in size, shape, color, physiology, and behavior between male and female individuals. This complex phenomenon captivated many biologists and subsequently led to a great amount of theories about a hypothesized link between sex chromosomes and sexual dimorphism (Kirkpatrick and Hall 2004; Albert and Otta 2005). Although several species-specific reports provided supporting evidence for these theories (Iyengar et al. 2002; Lindholm and Breden 2002; Saether et al. 2007), other studies did not conform to these theoretical predictions (Fairbairn and Roff 2006; Fitzpatrick 2004; Dean and Mank 2014; Mank et al. 2006a, b).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10142-018-0595-y) contains supplementary material, which is available to authorized users.

✉ Na Wang
wangna@ysfri.ac.cn; chensl@ysfri.ac.cn

[1] Key Laboratory for Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, 106 Nanjing Road, Qingdao 266071, China

[2] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266235, China

[3] College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China

Published online: 12 March 2018

Furthermore, a growing body of evidences suggests autosomes to be a major player, especially in those species without sex chromosomes (Mank et al. 2006a, b, 2007; Possiant et al. 2008; Ducrest et al. 2008). These conflicting results imply that the underlying mechanism of sexual dimorphism is far more complex than that represented by current models.

Sexual dimorphism in size has been described in more than 20 fish species (Parker 1992; Mei and Gui 2015). Among these, the Chinese tongue sole (*Cynoglossus semilaevis*) is a typical female heterogamete species (ZW/ZZ) and the body size of females is bigger than that of males by two- to four-fold (Chen et al. 2007). The significant difference in growth speed between female and male *C. semilaevis* has dramatically increased the feeding cost and severely hindered the sustainable development of aquaculture.

As a complex polygenic trait, fish growth is mainly regulated by the environment, nutrients, energy metabolism, and reproduction activity. Increasing evidences point to the importance of interaction between the somatotropic endocrine axis (brain-pituitary-liver) and the reproductive axis (brain-pituitary-gonad) in the regulation of growth (Holloway and Leatherland 1998; Melamed et al. 1998; Gomez et al. 1999).

To date, in *C. semilaevis*, sexual expression patterns have only been reported in several growth-related genes including growth hormone (*gh*), growth hormone receptor (*ghr*), insulin growth factor (*igf*), growth hormone-releasing hormone (*ghrh*), and pituitary adenylate cyclase-activating polypeptide (*pacap*). Also, limited investigations by cDNA array or transcriptome analysis have been performed in one single tissue such as the pituitary gland or the brain (Ji et al. 2011a, b; Ma et al. 2011a, b, 2012; Wang et al. 2016). No report on tripartite interactions among the brain, pituitary gland, and target organs (gonad, liver, muscle, and others) is available for *C. semilaevis*.

In the present study, four important tissues (brain, pituitary, gonads, and liver) of the somatotropic and reproductive axes from 15 females and 15 males were used for transcriptome analysis. As a result, a large number of differentially expressed genes between female and male tissues were obtained and analyzed via GO term and KEGG pathway enrichment. Subsequently, weighted gene co-expression network analysis (WGCNA) was employed for growth trait-related module identification and hub genes screening.

Materials and methods

Ethical statement

The collection and handling of all animals used in this study were approved by the Animal Care and Use Committee of the Chinese Academy of Fishery Sciences. Furthermore, all experimental procedures were performed in accordance with the guidelines for the Care and Use of Laboratory Animals of the Chinese Academy of Fishery Sciences.

Tissue preparation and total RNA extraction

Two-year-old Chinese tongue sole individuals were obtained from the Haiyang Yellow Sea Fisheries Limited Company: 15 females and 30 males. Due to the existence of pseudomales among male individuals, a sex determination experiment was firstly conducted to exclude pseudomales, following a previously described method (Liu et al. 2014) using the primers sex-F and sex-R (listed in Table S1), thus amplifying two products of 169 and 134 bp in females and one product of 169 bp in males. Subsequently, brain, pituitary, gonad, and liver tissues from 15 females and 15 males were sampled and immediately stored in liquid nitrogen.

To increase biological reproducibility, every five tissues from males or females were firstly mixed into one sample, and three samples for each tissue were used for the following experiment. Thus, 24 samples named CSE_MB1-3, CSE_MP1-3, CSE_MG1-3, CSE_ML1-3, CSE_FB1-3, CSE_FP1-3, CSE_FG1-3, and CSE_FL1-3 were acquired and their total RNAs were extracted via Trizol reagent (Invitrogen, USA) and measured with the Agilent 2100 RNA 6000 Nano kit (Agilent, USA) for quality control.

RNA-seq sequencing, gene annotation, and novel gene identification

Twenty-four cDNA libraries were constructed using 3 μg RNA (RIN > 7.0) from each sample via the conventional protocol, and the subsequent sequencing was conducted with an Illumina HiseqTM 4000 by the Gene Denovo Biotechonolgy Co. (Guangzhou, China).

After the adapters were removed from raw reads, those reads with low quality including reads with more than 10% unknown nucleotides and more than 50% low-quality bases (Q value < 20) were removed from the data set. The data were subsequently aligned with Bowtie2 and mapped to the *C. semilaevis* genome via TopHat2 (version 2.0.3.12) for rRNA removal and gene annotation. Then, transcript assemblies were conducted with Cuffinks and the reference annotation-based transcripts (RABT) program to obtain a comprehensive set of transcripts.

To identify novel transcripts, the assembled transcripts were aligned to the *C. semilaevis* genome and were divided into 12 categories using Cuffcompare. Genes with classcode "u," lengths ≥ 200 bp, and exon numbers ≥ 2 were defined as novel genes, and were further annotated via alignment with the Nr and KEGG database.

Analysis of differentially expressed genes

Gene abundances of known and new transcripts were quantified with the software RNA-Seq by Expectation-Maximization (RSEM). The gene expression level was further normalized by using the fragments per kilobase of transcript per million (FPKM) mapped reads method to eliminate the influence of different gene lengths and amount of sequencing data on the calculation of gene expression. The edgeR package (http://www.r-project.org/) was used to identify differentially expressed genes (DEGs) across samples with fold changes ≥ 2 and a false discovery rate-adjusted P (q value) < 0.05.

DEGs were then subjected to an enrichment analysis of GO function and KEGG pathways, and q values were corrected using < 0.05 as threshold.

Analyses of replicas' correlation and principle component

To evaluate reproducibility between samples, the correlation coefficient among three replicas was calculated. Values closer to one indicated better reproducibility. Principle component analysis (PCA) was performed with the R package gmodels (http://www.r-project.org) to reveal the relationship between samples.

Gene co-expression network construction

To reveal the relationship of the numerous DEGs and the growth traits in samples (body weight, body length, body width, and body height; shown in Table S2), WGCNA was conducted with the R package WGCNA, following published procedure (Langfelder and Horvath 2008). Prior to WGCNA, low-quality genes (which were not expressed in more than half of all samples) and low-quality samples (of which, more than half of the genes were not expressed) were filtered to improve the accuracy of the resulting network. Expression correlation coefficients of remaining genes were then calculated to search a suitable soft threshold for building gene networks using a scale-free topology model. Subsequently, gene modules with similar expression patterns were identified based on the resulting gene cluster dendrogram and using the dynamic tree cut method (minModuleSize = 50 and mergeCutHeight = 0.25). The top 1 or 5% of genes with the highest connectivity in the module network were defined as hub genes (Yang et al. 2014). GO term and KEGG pathway enrichment analyses of the annotated genes were performed for each module and q value < 0.05 was defined as threshold for significance.

Module-trait relationship and hub gene identification

To identify modules that were significantly associated with the trait of samples, the module eigengene was firstly calculated, using all genes in each module and then correlated with weight, length, width, and height of samples (Table S2). Modules with high correlation values and $p < 0.05$ were considered as significantly trait-related modules, and 1 or 5% of which were defined as hub genes. To provide a visualized network for hub genes associated with the trait of samples, co-expression patterns and interactions of hub genes were constructed via cytoscape (Shannon et al. 2003).

Validation of RNA-seq data via quantitative real-time PCR assay

To further validate the confidence of high-throughput transcriptome sequencing, 20 differentially expressed genes were selected and analyzed via qPCR. Primers (listed in Table S1) were designed based on sequences from the NCBI database. The *C. semilaevis β-actin* gene was used as an internal reference (Li et al. 2010). One microgram total RNAs for high-throughput transcriptome sequencing was reverse transcribed into cDNA with the PrimeScript™ RT reagent Kit with gDNA Eraser (Takara, Japan). Then, qPCR was conducted using SYBR Premix Ex Taq (Takara, Japan) in 20-μl reactions, containing 10 μl SYBR Premix Ex Taq (2X), 0.4 μl forward primer (10 μM), 0.4 μl reverse primer, 0.4 μl ROX reference dye, 1 μl cDNA, and 8.2 μl ddH_2O. The PCR amplification procedure was carried out at 95 °C for 10 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 34 s; this was followed by a disassociation curve analysis in an ABI 7500 fast real-time PCR system (Applied Biosystems, USA). Amplification reaction without template was used as no template control, and all reactions were performed in triplicate. The relative expression fold changes of 20 genes in female versus male tissues were analyzed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). The fold changes of 20 genes in female versus male tissues obtained via RNA-seq were calculated via FPKM. These genes' log2 fold change values of qPCR and RNA-seq were used for graphical presentation.

Results

High-throughput sequencing data

To better understand *C. semilaevis* growth difference mechanisms, we conducted a comparative transcriptomic analysis among 15 female and 15 male individuals. The exclusion of pseudomales was carried out following a previously described method (Liu et al. 2014) and the primers are listed in Table S1. A total of 24 libraries, named CSE_MB1-3, CSE_MP1-3,

CSE_MG1-3, CSE_ML1-3, CSE_FB1-3, CSE_FP1-3, CSE_FG1-3, and CSE_FL1-3, were constructed using total RNAs from brain, pituitary, gonad, and liver tissues. Each library was sequenced with an Illumina Hiseq™ 4000. As a result, a mean of 37,533,991 filtered clean reads with a Q20 above 97.86% was obtained from each library (Table S3). Subsequently, 1.01–3.30% of reads mapped with rRNA were removed, and 82.80–86.49% of the remaining reads were mapped with the *C. semilaevis* genome (NCBI CSEe_v1.0) (Table S4).

Subsequently, a total of 20,373 known genes and 2400 novel genes were identified via further transcript assembly, based on the *C. semilaevis* genome. The RNA-seq data from the present study have been uploaded to the NCBI GEO database (accession number GSE104688).

Analyses of differentially expressed genes and sample clustering

Based on the expression levels of each gene within samples, a correlation coefficient was calculated to evaluate repeatability. The Pearson correlation coefficients of three replicas were between 92.95 and 99.86% (Figure S1), which was further confirmed via PCA (Figure S2). Furthermore, the hierarchical clustering analysis revealed correct relationships among 24 samples (Fig. 1a).

Via differential expression analysis, 806 genes in brain (Table S5), 1482 genes in pituitary gland (Table S6), 818 genes in liver (Table S7), and 14,695 genes in gonad (Table S8) were found to be differentially expressed between female and male individuals (Fig. 1b) (fold change ≥ 2 and $q < 0.05$).

The validation of RNA-seq

To validate the high-throughput sequencing data, 20 genes including serine/threonine-protein kinase A-Raf-like isoform X1 (*araf*), early growth response protein 1 (*egr1*), *egr4*, growth/differentiation factor 8-like (*gdf8*), insulin-like growth factor-binding protein 5 (*igfbp5*), MAP kinase-activated protein kinase 5 (*mapk5*), BDNF/NT-3 growth factors receptor (*ntrk2*), leptin receptor (*lepr*), low-density lipoprotein receptor-related protein 3 (*lrp3*), transcriptional activator protein Pur-beta (*purb*), proto-oncogene c-Fos-like (*fos*), G-protein subunit alpha 12 (*gna12*), Jun proto-oncogene, AP-1 transcription factor subunit (*jun*), *lrp1b*, *lrp2*, LDL receptor-related protein-associated protein 1 (*lrpap1*), platelet-derived growth factor receptor alpha (*pdgfra*), E3 SUMO-protein ligase PIAS2-like isoform X2 (*pias2*), son of sevenless homolog 1 (*sos1*), and the uncharacterized LOC103392547 (*un*) were selected for qPCR analysis. The relative fold change of expression for these genes in females versus males was firstly calculated by the $2^{-\Delta\Delta Ct}$ method and the log2 fold change was then used for a comparison with the log2 fold change of RNA-seq. Similar up- or downregulation patterns of these genes were observed in qPCR and RNA-seq results (Fig. 2), although few genes showed differentiation between the values. These differentiations may be due to the different calculated methods used between RNA-seq and qPCR.

The GO and KEGG enrichment of differentially expressed genes

To better understand the function of these differentially expressed genes, GO term enrichment was conducted and a 50, 86, 173, and 123 GO terms with $p < 0.05$ were significantly enriched in brain, pituitary, liver, and gonad tissues (Table S9), respectively. Among these, with the exception of the brain, 3, 22, and 15 GO terms with $q < 0.05$ were enriched in pituitary, liver, and gonad tissues, respectively (Fig. 3a). Nucleic acid-binding transcription factor activity was enriched in both pituitary and liver tissues. Pituitary DEGs were also enriched in G-protein-coupled receptor activity and the G-protein-coupled receptor signaling pathway. In a gonad molecular function analysis, the enriched GO terms focused on several kinase activity terms such as kinase, protein kinase, phosphotransferase, and transferase activities; many binding terms included carbohydrate derivative, purine nucleoside, and ribonucleoside binding. In addition, catalytic activity and hydrolase activity contained the highest number of genes. In a liver molecular functional analysis, binding terms such as transition metal ion, growth factor, cation, and ion binding were also included. Further significantly enriched activity in the liver included nucleic acid-binding transcription factor activity, triglyceride lipase activity, and oxidoreductase activity.

Subsequently, KEGG pathway enrichment revealed that 3, 5, 16, and 21 pathways were significantly enriched in brain, pituitary, liver, and gonad tissues, respectively ($p < 0.05$) (see Fig. 3b and Table S10). Three pathways (the MAPK signaling pathway, the vascular smooth muscle contraction, and the FoxO signaling pathway) were enriched in both liver and gonad tissues. In addition to these, neuroactive ligand-receptor interaction was significantly enriched in the pituitary ($q < 0.05$). In the liver, steroid biosynthesis, p53 signaling pathway, and nicotinate and nicotinamide metabolism were also enriched. In the gonads, adherens junction, Wnt signaling pathway, endocytosis, focal adhesion, adipocytokine signaling pathway, and progesterone-mediated oocyte maturation were also enriched.

Construction of the gene co-expressed network

One thousand fifty-six low-quality genes, not expressed in half of the samples, were removed and the remaining 21,717 genes were used for the subsequent WGCNA. Based on the correlation coefficients of these genes, a gene cluster dendrogram was

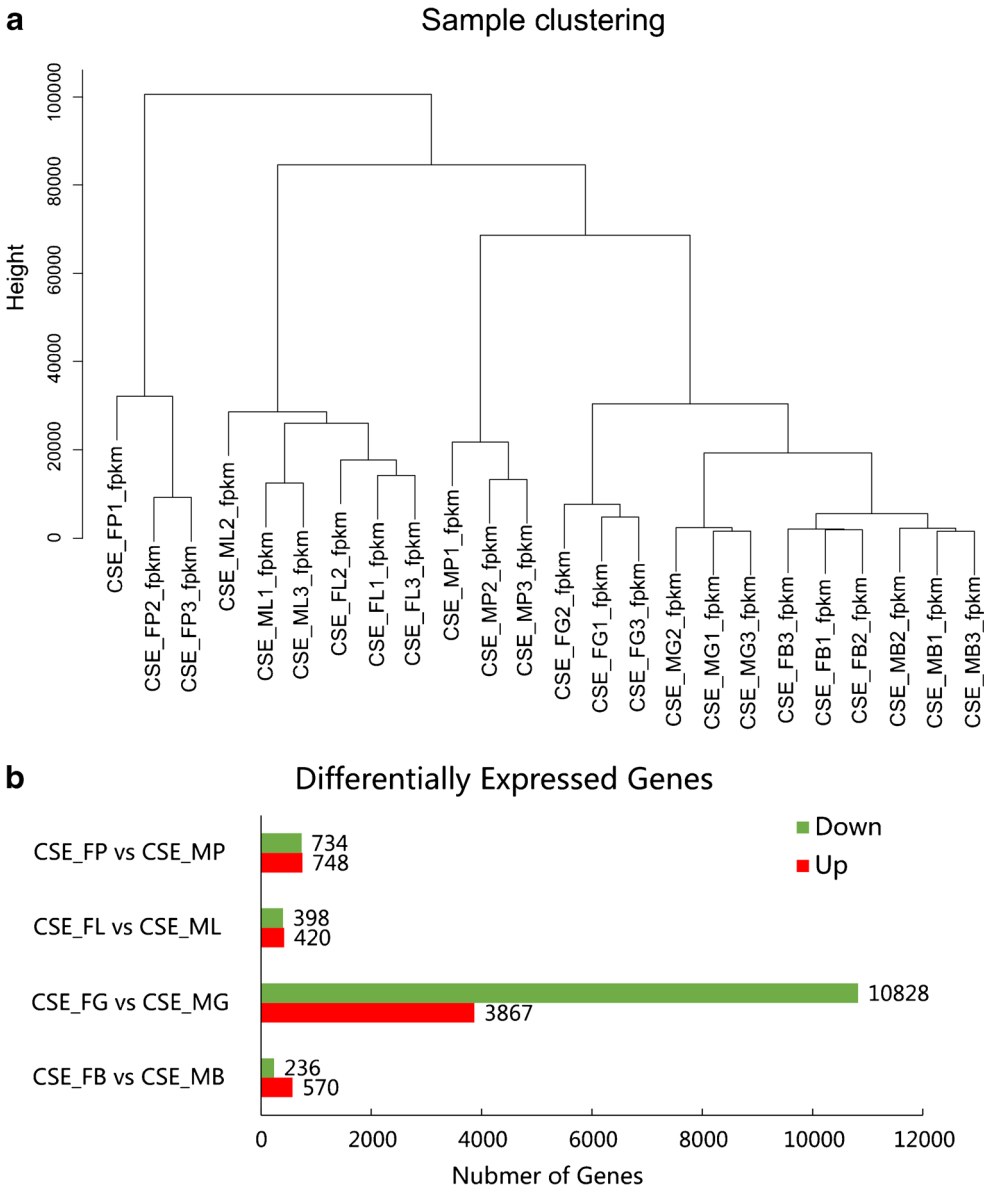

Fig. 1 The sample clustering and DEGs in RNA-seq. **a** Hierarchical clustering result based on all expressed genes in 24 samples. In the brain and liver, the samples were firstly clustered by different sexes and then clustered by different tissues. In the pituitary and gonad, samples were just clustered by different sexes. **b** DEGs identified from four groups of tissues. Red and green indicate upregulated genes and downregulated genes in female vs male tissues, respectively

constructed with a power value = 15 (Fig. 4). The gene modules were further classified and clustered by similarity = 0.8 and minModuleSize = 50. Finally, 13 modules were identified, with module sizes ranging from 83 to 5094 (Fig. 5, Table S11).

Identification and functional annotation of growth-related modules

To identify trait-related modules, the correlation coefficient between modules and sample growth traits (weight, length, width, and height; see Table S2) was calculated. As a result, both yellowgreen and salmon4 were significantly positive-correlated with growth traits ($p < 0.05$) (Fig. 6). The eigengenes of the yellowgreen module were highly expressed in the female brain and the female pituitary gland (Figure S3), while the eigengenes of the salmon4 module were highly expressed in the female brain (Figure S4). Based on a chromosome location analysis, more than 83% of the yellowgreen and 53% of the salmon4 eigengenes were found to be located on the W chromosome, which is one of the sex chromosomes (Tables S12 and S13). The modules magenta, darkorange2, and mediumpurple3 also exhibited a positive correlation with growth traits, although no significant difference was observed. GO term enrichment analyses for two significantly positively correlated modules with $p < 0.05$ are provided in Table S14.

According to the cellular component, yellowgreen was mainly enriched in the chromosomal part and the chromosome. According to molecular function, yellowgreen had been

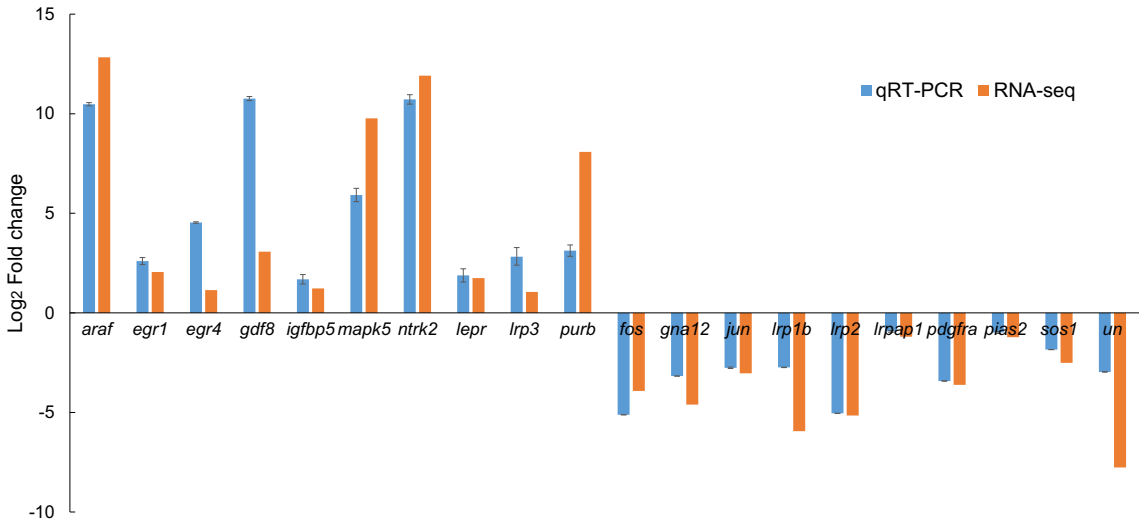

Fig. 2 The relative expression levels of 20 selected DEGs by qPCR and RNA-seq. Twenty DEGs including ten upregulated genes (*araf*, *egr1*, *egr4*, *gdf8*, *igfbp5*, *mapk5*, *ntrk2*, *lepr*, *lrp3*, *purb*) and ten downregulated genes (*fos*, *gna12*, *jun*, *lrp1b*, *lrp2*, *lrpap1*, *pdgfra*, *pias2*, *sos1*, *un*) were selected for qPCR. *β-Actin* was used for the internal reference. The expression fold changes of 20 genes in female versus male tissues detected by qPCR and RNA-seq were calculated by $2^{-\Delta\Delta Ct}$ and FPKM, respectively. And, these genes' log2 fold change values of qPCR and RNA-seq are shown in blue and red, respectively. In qPCR, values are indicated as means ± standard error (SE) derived from triplicate experiments

enriched in activity and binding, such as in phosphoric ester hydrolase activity, *N*-acetyltransferase activity, transition metal ion binding, and heterocyclic compound binding. Salmon4 was also enriched in activity and binding, including sterol transporter activity, protein binding, and sterol binding.

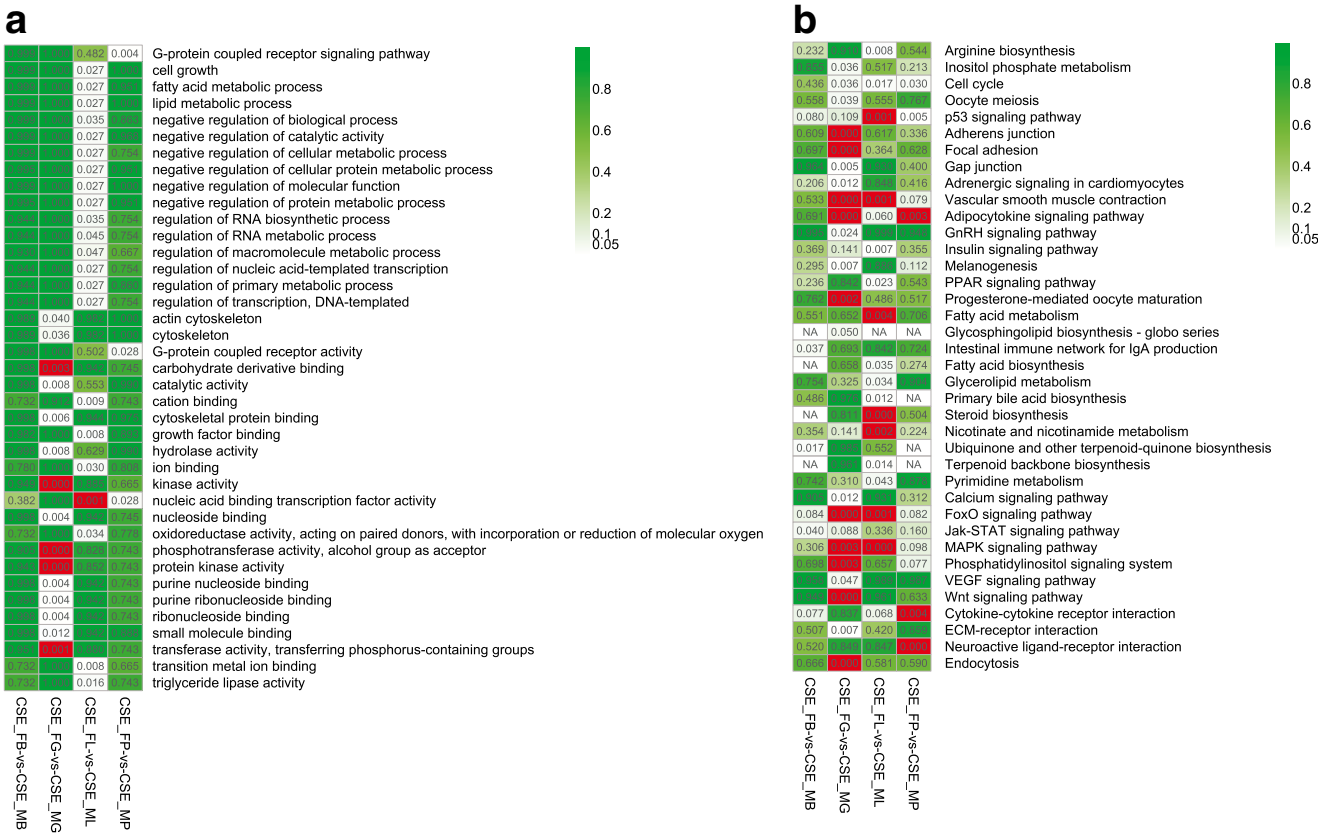

Fig. 3 GO term and KEGG pathway enrichment in four groups of RNA-seq. **a** Enriched GO terms in the brain, gonad, liver, and pituitary ($q < 0.05$). **b** Enriched KEGG pathways in the brain, gonad, liver, and pituitary ($p < 0.05$). The color bar indicates q or p value from low (red) to high (green)

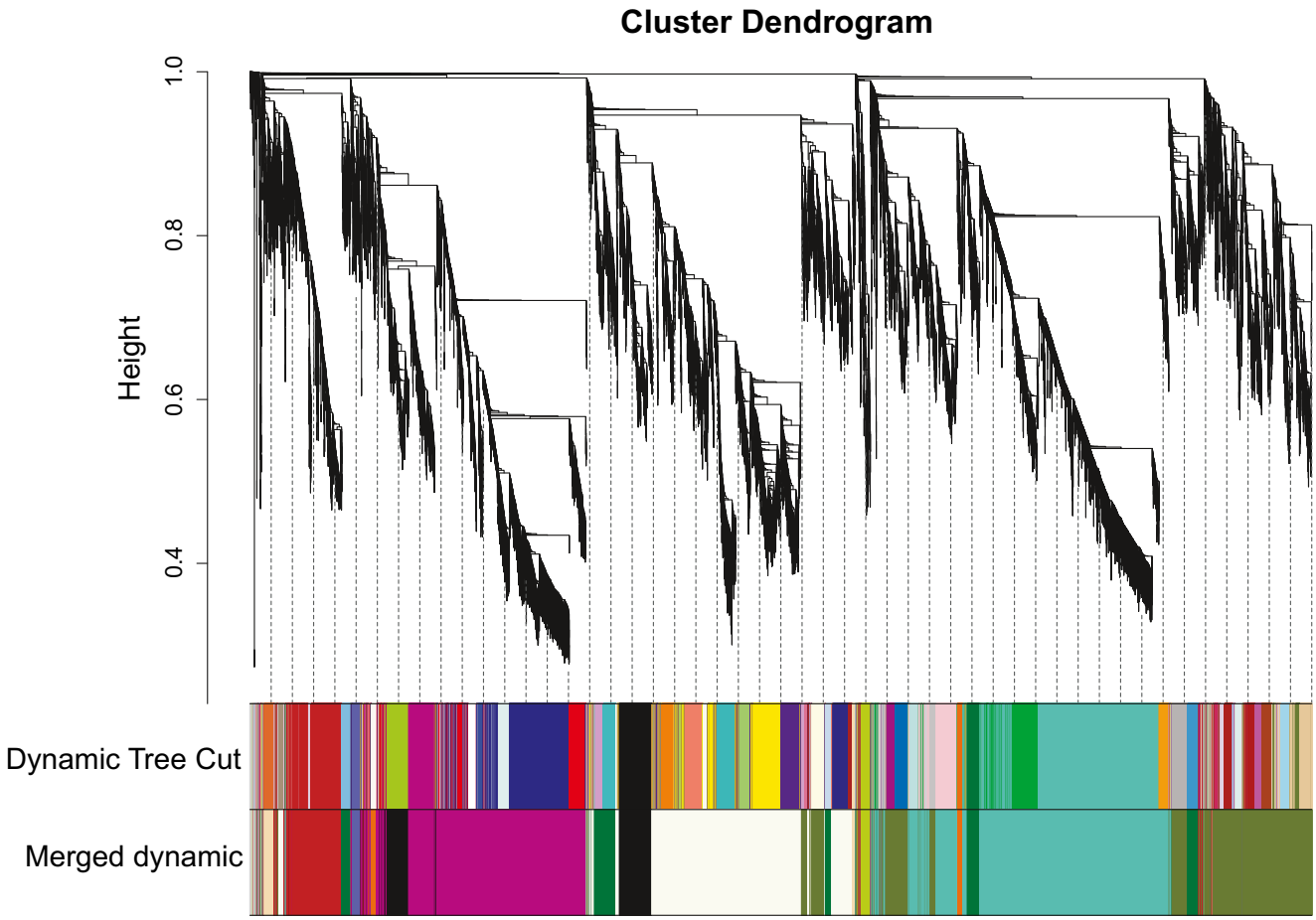

Fig. 4 The gene cluster dendrogram constructed by genes' correlation coefficients. The vertical distance shows the distance between two separate genes. Dynamic tree cut indicates the divided modules based on genes' clustering result. Merged dynamic indicates the divided modules by combining modules with similar expression patterns

In biological processes, yellowgreen significantly enriched the acute inflammatory response, the acute-phase response, direct ossification, and nucleosome organization. Salmon4 was enriched in the cell wall macromolecule metabolic process, cell wall organization or biogenesis, peptidoglycan metabolic process, sterol transport, and sulfate transport.

In KEGG pathway enrichment analysis (Table S15), yellowgreen was significantly enriched in lysine degradation, oocyte meiosis, ubiquitin-mediated proteolysis, progesterone-mediated oocyte maturation, insulin signaling pathway, vascular smooth muscle contraction, and NOD-like receptor signaling pathway ($p < 0.05$). Salmon4 was significantly enriched in melanogenesis, cell adhesion molecules (CAMs), Notch signaling pathway, phenylalanine metabolism, ubiquinone, and in other terpenoid-quinone biosynthesis ($p < 0.05$).

Hub gene identification and network construction

In the yellowgreen module, the top 150 pairs with the highest weight (54 genes) were selected for network construction (Fig. 7a). Among these, seven genes exhibited a close relationship with other genes, including the transcriptional activator protein Pur-beta-like (LOC103396891), the nipped-B-like protein A (LOC103397102), the activated RNA polymerase II transcriptional coactivator p15-like (LOC103397063), the AN1-type zinc finger protein 5-like (LOC103397127), the myotubularin-related protein 3-like (LOC103397032), the phosphatidate cytidylyltransferase 1-like isoform X1 (LOC103397022), and the histone-lysine *N*-methyltransferase NSD3-like isoform X4 (LOC103397008). In the salmon4 module, the top 200 pairs with the highest weights (31 genes) were selected for network construction (Fig. 7b).

Discussion

As a central trait relating to aquaculture yield, growth has received much attention in fish. *C. semilaevis* is a fish species that exhibits apparent sexual size dimorphism and is an excellent model for the study of growth difference mechanisms. Previous studies showed that significant differences in body weight between the sexes of *C. semilaevis* were found

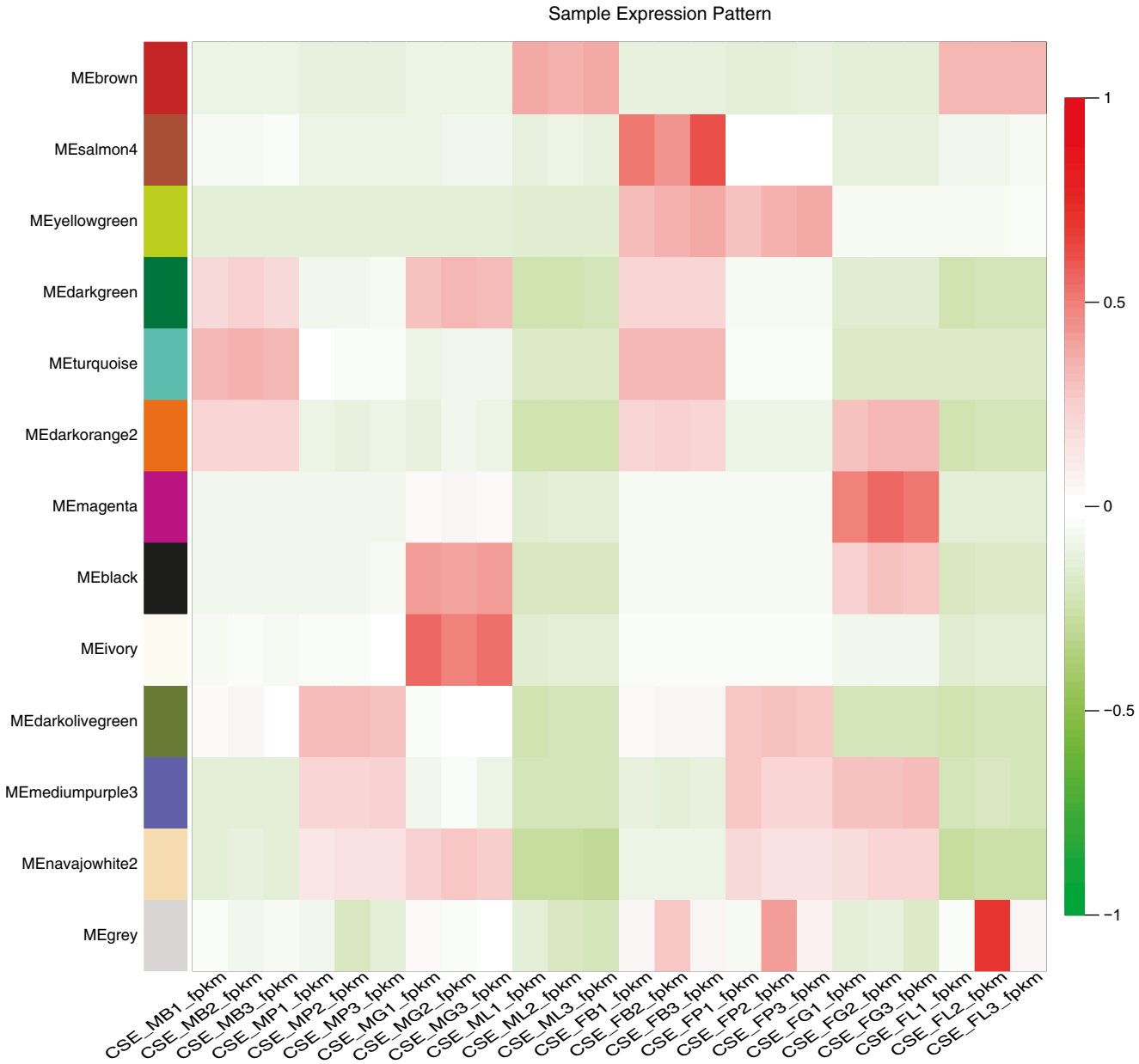

Fig. 5 The heatmap of samples expression pattern. The expression pattern of 13 modules including brown, salmon4, yellowgreen, darkgreen, turquoise, darkorange2, magenta, black, ivory, darkolivegreen, mediumpurple3, navajowhite2, and grey were shown by a heatmap. The color bar indicates expression levels from low (green) to high (red)

following 9 months of growth and the body weight of 21-month-old females was 3.28 times higher than that of males (Ji et al. 2011b). In the present study, sexual size dimorphisms of 2-year-old individuals were also observed, with the females exhibiting 6.07 times in body weight and 1.78 times in body length than the males (Table S2). Thus, brain, pituitary, liver, and gonad tissues of 2-year-old individuals, four main tissues from the somatotropic and the reproductive axes, were selected as the material for RNA-seq and WGCNA.

As numerous studies have underscored the importance of biological and technical replicates within RNA-seq (Bullard et al. 2010; Auer and Doerge 2010; McIntyre et al. 2011), we selected 15 biological and three technical replicates for each tissue to achieve more reliable results. High Q20 values, high Pearson correlation coefficients, PCA, and clustering analyses all proved that we obtained high-quality sequencing data with good reproducibility.

The brain plays an important upstream regulating function in the somatotropic and reproductive axes by secreting numerous neuroendocrine factors. In the present study, 570 genes exhibited upregulated patterns in the female brain, which is more than twice as many as downregulated genes. Among these upregulated genes, some primary response genes such as *egr1* and *egr4* interested us because they are required for

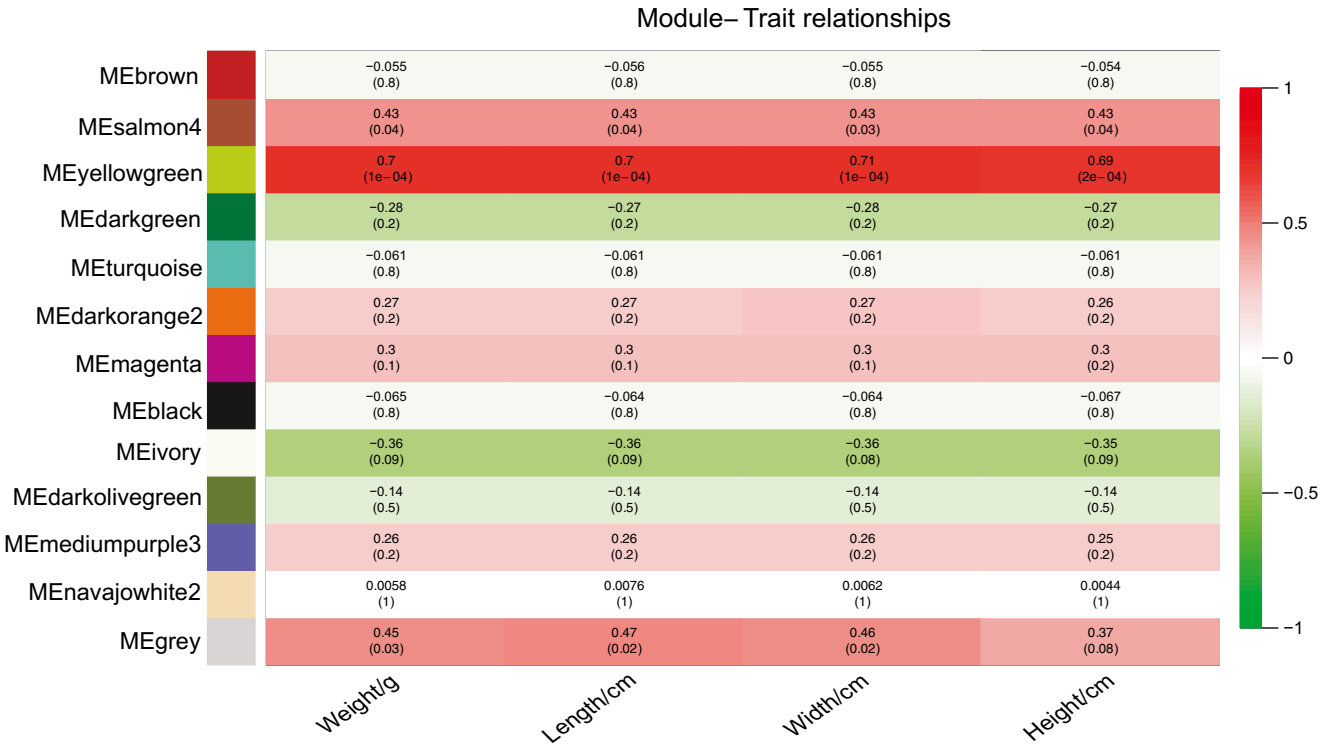

Fig. 6 The relationship between 13 modules and growth traits in samples. The eigengene in each module was calculated and shown in each row. The columns represent the sample traits of weight, length, width, and height. The modules with high correlation value and $p < 0.05$ were considered as significantly trait-related modules. The color bar indicates correlation value from low (green) to high (red)

both differentiation and mitogenesis in response to growth factors (Herschman 1991; Tarcic et al. 2012). *Araf*, which is involved in the MAPK signaling pathway and leads to both cell cycle progression and cell proliferation (Mercer et al.

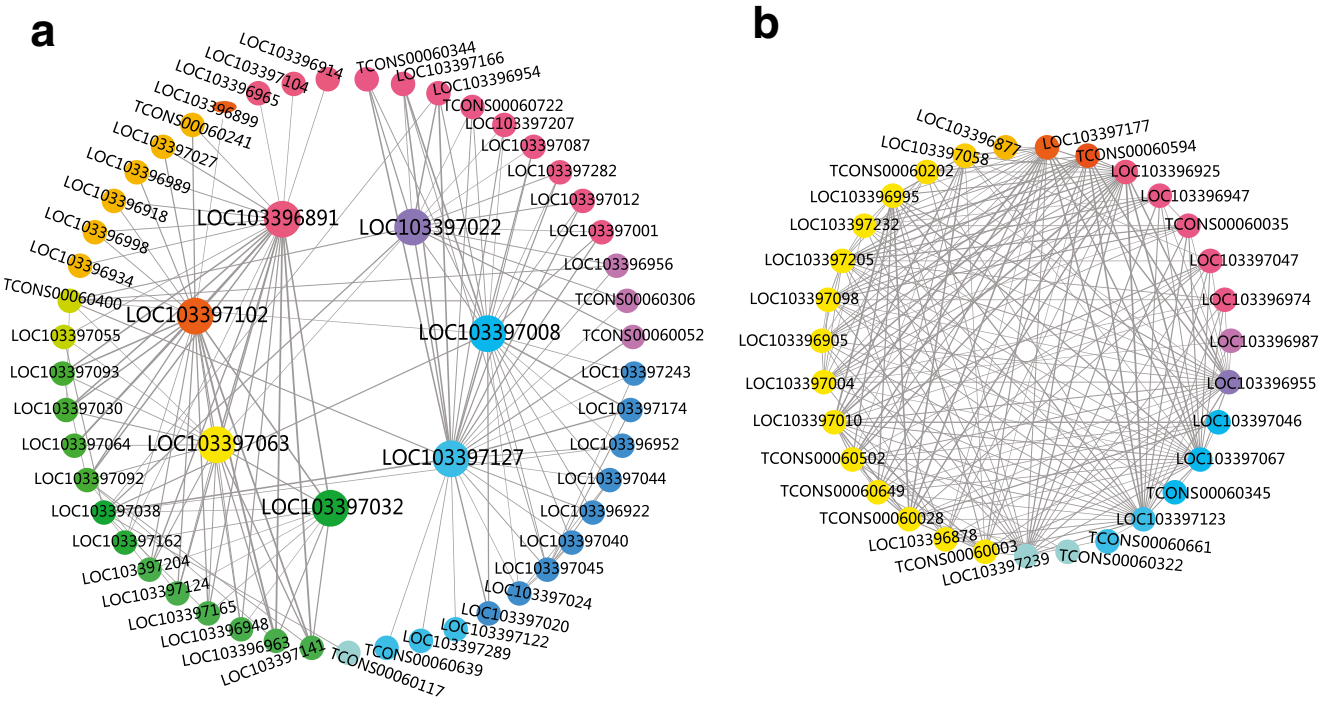

Fig. 7 The network relationship in yellowgreen and salmon4 modules. Fifty-four genes with the highest weight in yellowgreen and 31 genes with the highest weight in salmon4 were separately used for the construction of network illustration by cytoscape

2005), also exhibits an upregulated expression pattern in the female brain by 12.83 log2FC.

As a shared tissue located in the somatotropic and reproductive axes, the pituitary gland is composed of adenohypophysis and neurohypophysis. The neurohypophysis mainly releases vasopressin and oxytocin, while the adenohypophysis mainly releases adrenocorticotropic hormone (encoded by *acth*), thyrotropin (encoded by *tsh*), gonadotropic hormone (encoded by *gth*), growth hormone (encoded by *gh*), prolactin (encoded by *prl*), and melanophore-stimulating hormone (encoded by *msh*) in different cell types. In the present study, 1482 genes exerted differential expression patterns in female and male pituitaries, which were significantly enriched in the G-protein-coupled receptor signaling pathway, nucleic acid-binding transcription factor activity, and the neuroactive ligand-receptor interaction pathway. Among the neuroactive ligand-receptor interaction pathways, *tsh* and *gh* exhibited upregulated expression patterns in female versus male pituitary tissues. The expression pattern of *gh* was similar to the previously reported pattern in *C. semilaevis* (Holloway and Leatherland 1998; Ji et al. 2011b). Interestingly, another teleost-specific *prl* family member somatolactin (*sl*) (Rand-Weaver et al. 1991; Power 2005) was downregulated in the female pituitary gland, suggesting that these two somatotropic genes may regulate fish growth in a complementary manner (Company et al. 2001). Leptin (Denver et al. 2011), a member of the neuroactive ligand-receptor interaction pathway, is a fat cell-specific hormone (which influences development, growth, metabolism, and reproduction by binding to the leptin receptor (*lepr*) (Tartaglia et al. 1996)) and showed no significant difference in our samples. However, *Lepr* exerts an upregulated expression pattern in the female pituitary, which may provide a clue about its prospective role in regulating growth sexual differences.

The gonads are important tissues with obvious sexual differences, and 14,695 DEGs were enriched in protein kinase activity, phosphotransferase activity, Wnt signaling pathway, and MAPK signaling pathway. The MAPK signaling pathway existed in most cells and participated in the transmission of extracellular signals into their intracellular targets via activated MAPK, leading to a variety of cellular responses containing growth, differentiation, inflammation, and apoptosis (Seger and Krebs 1995). A study in brown trout (*Salmo trutta*) revealed that this pathway might be involved in the inhibitory effect of insulin and insulin-like growth factor-I (*igf-i*) on luteinizing hormone (*lh*)-stimulated steroid production (Méndez et al. 2005). For the carp (*Cyprinus carpio*), the interaction of the cAMP/PKA pathway with MAPK cascades may trigger *gh* transcription by activated *lh* receptor (Sun et al. 2014). In our study, 36 and 297 genes derived from this pathway exerted different expression patterns in livers and gonads, respectively.

Several GO terms and KEGG pathways including growth factor binding, cell growth, steroid biosynthesis, FoxO signaling pathway, and MAPK signaling pathway were significantly enriched in the liver. In the GO term of growth factor binding, WNT1-inducible-signaling pathway protein 3 (*wisp3*) interested us due to its essential role in human skeletal growth and cartilage homeostasis (Hurvitz et al. 1999). Its expression level increased by 2.11 log2FC in the female liver of *C. semilaevis*. IGFBP5 served as a carrier protein for IGFs (Hwa et al. 1999), and its mRNA expression was upregulated by 1.23 log2FC in female liver.

Steroid biosynthesis is an anabolic pathway, which produces steroids from simple precursors. Furthermore, the steroid cholesterol produces sex steroid hormones including estrogens, progestins, and androgens, which have been reported to affect GH production and to be released in a sex-dependent manner, contributing to skeletal sexual dimorphisms in mammals (Kerrigan and Rogol 1992; Veldhuis et al. 2000; Lindberg et al. 2001; Moverare et al. 2003). Similar to mammals, interactive regulation effects between these sexual steroid hormones and growth-related genes such as *gh* and *igf* also exist in several fish species (Manzoor and Rao 1989; Trudeau et al. 1992; Weber et al. 2007). The main enzymes involved in steroid biosynthesis, including sterol-C5-desaturase (*sc5d*), delta24-sterol reductase (*dhcr24*), 7-dehydrocholesterol reductase (*dhcr7*), emopamil-binding protein (sterol isomerase) (*ebp*), and lanosterol 14-alpha demethylase (*cyp51*) were all upregulated in the female liver of *C. semilaevis*, suggesting that cholesterol-dependent sex steroid hormones may be involved in the sexual size dimorphism of *C. semilaevis*.

WGCNA has been used as a powerful tool in systematic biology for the identification of key genetic networks involved in human diseases and crop quality (Zhang et al. 2016; Bai et al. 2015). Preliminary information about WGCNA is currently available in the zebrafish (*Danio rerio*) and the lake whitefish (*Coregonus clupeaformis*) (Filteau et al. 2013; Wong et al. 2014). In the present study, WGCNA revealed the two modules yellowgreen and salmon4 to be significantly positive-correlated with growth traits of *C. semilaevis*. Among both modules, genes were mainly upregulated in the female brain and the pituitary, suggesting that these two upstream tissues may play important roles in the growth of the different sexes. The eigengene expression patterns and chromosome locations of these two modules led to identification of 204 W chromosome-located genes, which occupied 64.35% (204/317) of the predicted functional protein-coding genes on the W chromosome (Chen et al. 2014). These data suggest that the W chromosome may be critical for interpreting sexual size dimorphism in *C. semilaevis*.

In the yellowgreen module, expression of the key gene nipped-B-like protein A (*nipb1a*) was upregulated by 6.06–10.69 log2FC in four tested female tissues. In humans, *nipb1a* mutation results in a disorder characterized by growth delay and limb reduction defects (Schoumans et al. 2007). The

transcriptional activator protein Purb-like is an important protein implicated in the control of DNA replication and transcription (Bergemann et al. 1992). Expression of the *purb-like* gene was upregulated by 7.73–9.82 log2FC in four tested female tissues.

In the salmon4 module, *ntrk2* (LOC103397098), which is also known as tyrosine receptor kinase B (*trkb*), is a receptor for the brain-derived neurotrophic factor (*bdnf*) and related diseases include obesity and developmental delay (Gupta et al. 2013). Its expression level was upregulated by 11.91 log2FC in the female brain.

The present study provides important insights into the sexual size dimorphism of *C. semilaevis*. The multiple tissues utilized in this study and the application of WGCNA enabled us to uncover novel networks and hub genes in both the somatotropic and reproductive axes. In future, a detailed functional survey of these networks and hub genes will further improve our understanding for the sexual growth differences in fish.

Author contributions NW and SLC conceived and designed the experiments. NW, RKW, and RQW performed fish tissue sampling and pseudomales identification. RKW conducted qPCR validation experiment. NW and RKW analyzed the data and wrote the paper. All authors read and approved the final manuscript.

Funding information This work was supported by grants from the Central Public-interest Scientific Institution Basal Research Fund CAFS (No. 2016GH03), the National Natural Science Foundation of China (31130057), the AoShan Talents Cultivation Program Supported by Qingdao National Laboratory for Marine Science and Technology (No. 2017ASTCP-OS15), and the Taishan Scholar Project of Shandong Province.

Compliance with ethical standards

The collection and handling of all animals used in this study were approved by the Animal Care and Use Committee of the Chinese Academy of Fishery Sciences. Furthermore, all experimental procedures were performed in accordance with the guidelines for the Care and Use of Laboratory Animals of the Chinese Academy of Fishery Sciences.

Competing interest The authors declare that they have no competing interests.

References

Albert AYK, Otta SP (2005) Sexual selection can resolve sex-linked sexual antagonism. Science 310:119–221

Auer PL, Doerge RW (2010) Statistical design and analysis of RNA sequencing data. Genetics 185(2):405–416

Bai Y, Dougherty L, Cheng L, Zhong GY, Xu K (2015) Uncovering co-expression gene network modules regulating fruit acidity in diverse apples. BMC Genomics 16:612

Bergemann AD, Ma ZW, Johnson EM (1992) Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. Mol Cell Biol 12(12):5673–5682

Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11:94

Chen SL, Li J, Deng SP, Tian YS, Wang QY, Zhuang ZM, Sha ZX, Xu JY (2007) Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol 9:273–280

Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, Song W, An N, Chalopin D, Volff JN, Hong Y, Li Q, Sha Z, Zhou H, Xie M, Yu Q, Liu Y, Xiang H, Wang N, Wu K, Yang C, Zhou Q, Liao X, Yang L, Hu Q, Zhang J, Meng L, Jin L, Tian Y, Lian J, Yang J, Miao G, Liu S, Liang Z, Yan F, Li Y, Sun B, Zhang H, Zhang J, Zhu Y, du M, Zhao Y, Schartl M, Tang Q, Wang J (2014) Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat Genet 46(3):253–260

Company R, Astola A, Pendón C, Valdivia MM, Pérez-Sánchez J (2001) Somatotropic regulation of fish growth and adiposity: growth hormone (GH) and somatolactin (SL) relationship. Comp Biochem Physiol C Toxicol Pharmacol 130(4):435–445

Dean R, Mank JE (2014) The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. J Evol Biol 27:1443–1453

Denver RJ, Bonett RM, Boorse GC (2011) Evolution of leptin structure and function. Neuroendocrinology 94(1):21–38

Ducrest AL, Keller L, Roulin A (2008) Pleiotropy in the melanocortin system, coloration and behavioral syndromes. Trends Ecol Evol 23:502–510

Fairbairn DJ, Roff DA (2006) The quantitative genetics of sexual dimorphism: assessing the importance of sex-linage. Heredity 97:319–328

Filteau M, Pavey SA, St-Cyr J, Bernatchez L (2013) Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. Mol Biol Evol 30(6):1384–1396

Fitzpatrick MJ (2004) Pleiotropy and the genomic location of sexually selected genes. Am Nat 163:800–808

Gomez JM, Weil C, Ollitrault M, Le Bail PY, Breton B, Le Gac F (1999) Growth hormone (GH) and gonadotropin subunit gene expression and pituitary and plasma changes during spermatogenesis and oogenesis in rainbow trout (*Oncorhynchus mykiss*). Gen Comp Endocrinol 113(3):413–428

Gupta VK, You Y, Gupta VB, Klistorner A, Graham SL (2013) Trk B receptor signaling: implications in neurodegenerative, psychiatric and proliferative disorders. Int J Mol Sci 14(5):10122–10142

Herschman HR (1991) Primary response genes induced by growth factors and tumor promoters. Annu Rev Biochem 60:281–319

Holloway AC, Leatherland JF (1998) Neuroendocrine regulation of growth hormone secretion in teleost fishes with emphasis on the involvement of gonadal sex steroids. Rev Fish Biol Fish 8(4):409–429

Hurvitz JR, Suwairi WM, Van Hul W, El-Shanti H, Superti-Furga A, Roudier J et al (1999) Mutations in the CCN gene family member WISP3 cause progressive pseudorheumatoid dysplasia. Nat Genet 23(1):94–98

Hwa V, Oh Y, Rosenfeld RG (1999) The insulin-like growth factor-binding protein (IGFBP) superfamily. Endocr Rev 20(6):761–787

Iyengar VK, Reeve HK, Eisner T (2002) Paternal inheritance of a female moth's mating preference. Nature 419:830–832

Ji XS, Chen SL, Jiang YL, Xu TJ, Yang JF, Tian YS (2011a) Growth differences and differential expression analysis of pituitary adenylate cyclase activating polypeptide (PACAP) and growth hormone-releasing hormone (GHRH) between the sexes in half-smooth tongue sole *Cynoglossus semilaevis*. Gen Comp Endocrinol 170(1):99–109

Ji XS, Liu HW, Chen SL, Jiang YL, Tian YS (2011b) Growth differences and dimorphic expression of growth hormone (GH) in female and male *Cynoglossus semilaevis* after male sexual maturation. Mar Genomics 4(1):9–16

Kerrigan JR, Rogol AD (1992) The impact of gonadal steroid hormone action on growth hormone secretion during childhood and adolescence. Endocr Rev 13:281–298

Kirkpatrick M, Hall DW (2004) Sexual selection and sex linkage. Evolution 58:683–691

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9(1):559

Li Z, Yang L, Wang J, Shi W, Pawar RA, Liu Y, Xu C, Cong W, Hu Q, Lu T, Xia F, Guo W, Zhao M, Zhang Y (2010) beta-Actin is a useful internal control for tissue-specific gene expression studies using quantitative real-time PCR in the half-smooth tongue sole *Cynoglossus semilaevis* challenged with LPS or *Vibrio anguillarum*. Fish Shellfish Immunol 29(1):89–93

Lindberg MK, Alatalo SL, Halleen JM, Mohan S, Gustafsson JA, Ohlsson C (2001) Estrogen receptor specificity in the regulation of the skeleton in female mice. J Endocrinol 171:229–236

Lindholm A, Breden F (2002) Sex chromosomes and sexual selection in poeciliid fishes. Am Nat 160(suppl):S214–S224

Liu Y, Chen SL, Gao FT, Meng L, Hu QM, Song WT et al (2014) SCAR-transformation of sex-specific SSR marker and its application in half-smooth tongue sole. J Agric Biotechnol 22(6):787–792

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. Methods 25(4):402–408

Ma Q, Liu SF, Zhuang ZM, Lin L, Sun ZZ, Liu CL, Su YQ, Tang QS (2011a) Genomic structure, polymorphism and expression analysis of growth hormone-releasing hormone and pituitary adenylate cyclase activating polypeptide genes in the half-smooth tongue sole (*Cynoglossus semilaevis*). Genet Mol Res 10(4):3828–3846

Ma Q, Liu SF, Zhuang ZM, Sun ZZ, Liu CL, Su YQ, Tang QS (2011b) Molecular cloning, expression analysis of insulin-like growth factor I (IGF-I) gene and IGF-I serum concentration in female and male tongue sole (*Cynoglossus semilaevis*). Comp Biochem Physiol B Biochem Mol Biol 160(4):208–214

Ma Q, Liu S, Zhuang Z, Lin L, Sun Z, Liu C, Ma H, Su YQ, Tang QS (2012) Genomic structure, polymorphism and expression analysis of the growth hormone (GH) gene in female and male half-smooth tongue sole (*Cynoglossus semilaevis*). Gene 493(1):92–104

Mank JE, Hall DW, Kirkpatrick M, Avise JC (2006a) Sex chromosome and male ornaments: a comparative evaluation in ray-finned fishes. Proc R Soc B Biol Sci 273:223–236

Mank JE, Promislow DEL, Avise JC (2006b) Evolution of alternative sex determining mechanisms in teleost fishes. Biol J Linn Soc 87:83–93

Mank JE, Hultin-Rosenberg L, Axelsson E, Ellegren H (2007) Rapid evolution of female-biased, but not male-biased, genes expressed in avian brain. Mol Biol Evol 24:2698–2706

Manzoor PKM, Rao GPS (1989) Growth improvement in carp (*Cyprinus carpio*) sterilized with 17α-methyltestosterone. Aquaculture 76: 157–164

McIntyre LM, Lopiano KK, Morse AM, Amin V, Oberg AL, Young LJ et al (2011) RNA-seq: technical variability and sampling. BMC Genomics 12:293

Mei J, Gui JF (2015) Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. Sci China Life Sci 58:124–136

Melamed P, Rosenfeld H, Elizur A, Yaron Z (1998) Endocrine regulation of gonadotropin and growth hormone gene transcription in fish. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 119(3): 325–338

Méndez E, Montserrat N, Planas JV (2005) Modulation of the steroidogenic activity of luteinizing hormone by insulin and insulin-like growth factor-I through interaction with the cAMP-dependent protein kinase signaling pathway in the trout ovary. Mol Cell Endocrinol 229(1–2):49–56

Mercer K, Giblett S, Oakden A, Brown J, Marais R, Pritchard C (2005) A-Raf and Raf-1 work together to influence transient ERK phosphorylation and Gl/S cell cycle progression. Oncogene 24(33): 5207–5217

Moverare S, Venken K, Eriksson AL, Andersson N, Skrtic S, Wergedal J, Mohan S, Salmon P, Bouillon R, Gustafsson JA, Vanderschueren D, Ohlsson C (2003) Differential effects on bone of estrogen receptor alpha and androgen receptor activation in orchidectomized adult male mice. Proc Natl Acad Sci U S A 100:13573–13578

Parker GA (1992) The evolution of sexual size dimorphism in fish. J Fish Biol 41(supplement B):1–20

Possiant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW (2008) Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. Proc R Soc B Biol Sci 275:623–628

Power DM (2005) Developmental ontogeny of prolactin and its receptor in fish. Gen Comp Endocrinol 142(1–2):25–33

Rand-Weaver M, Noso T, Muramoto K, Kawauchi H (1991) Isolation and characterization of somatolactin, a new protein related to growth hormone and prolactin from Atlantic cod (*Gadus morhua*) pituitary glands. Biochemistry 30(6):1509–1515

Saether SA, Saetre GP, Borge T, Wiley C, Svedin N, Andersson G, Veen T, Haavie J, Servedio MR, Bures S, Kral M, Hjernquist MB, Gustafsson L, Traff J, Qvarnstrom A (2007) Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science 318:95–97

Schoumans J, Wincent J, Barbaro M, Djureinovic T, Maguire P, Forsberg L, Staaf J, Thuresson AC, Borg Å, Nordgren A, Malm G, Anderlid BM (2007) Comprehensive mutational analysis of a cohort of Swedish Cornelia de Lange syndrome patients. Eur J Hum Genet 15(2):143–149

Seger R, Krebs EG (1995) The MAPK signaling cascade. FASEB J 9: 726–735

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Sun C, He M, Ko WK, Wong AO (2014) Mechanisms for luteinizing hormone induction of growth hormone gene transcription in fish model: crosstalk of the cAMP/PKA pathway with MAPK-and PI3K-dependent cascades. Mol Cell Endocrinol 382(2):835–850

Tarcic G, Avraham R, Pines G, Amit I, Shay T, Lu Y, Zwang Y, Katz M, Ben-Chetrit N, Jacob-Hirsch J, Virgilio L, Rechavi G, Mavrothalassitis G, Mills GB, Domany E, Yarden Y (2012) EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. FASEB J 26(4):1582–1592

Tartaglia LA, Dembski M, Weng X, Deng N, Culpepper J, Devos R et al (1996) Identification and expression cloning of a leptin receptor, OB-R. Cell 83(7):1263–1271

Trudeau VL, Somoza GM, Nahorniak CS, Peter RE (1992) Interactions of estradiol with gonadotropin-releasing hormone and thyrotropin-releasing hormone in the control of growth hormone secretion in the goldfish. Neuroendocrinology 56:483–490

Veldhuis JD, Roemmich JN, Rogol AD (2000) Gender and sexual maturation-dependent contrasts in the neuroregulation of growth hormone secretion in prepubertal and late adolescent males and females—a general clinical research center-based study. J Clin Endocrinol Metab 85(7):2385–2394

Wang P, Zheng M, Liu J, Liu Y, Lu J, Sun X (2016) Sexually dimorphic gene expression associated with growth and reproduction of tongue sole (*Cynoglossus semilaevis*) revealed by brain transcriptome analysis. Int J Mol Sci 17(9):1402

Weber GM, Moore AB, Sullivan CV (2007) In vitro actions of insulin-like growth factor-I on ovarian follicle maturation in white perch (*Morone americana*). Gen Comp Endocrinol 151(2):180–187

Wong RY, McLeod MM, Godwin J (2014) Limited sex-biased neural gene expression patterns across strains in zebrafish (*Danio rerio*). BMC Genomics 15:905

Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H (2014) Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun 5:3231

Zhang B, Tran L, Emilsson V, Zhu J (2016) Characterization of genetic networks associated with Alzheimer's disease. Methods Mol Biol 1303:459–477