

OPEN

# Transcriptomic analysis of female and male gonads in juvenile snakeskin gourami (*Trichopodus pectoralis*)

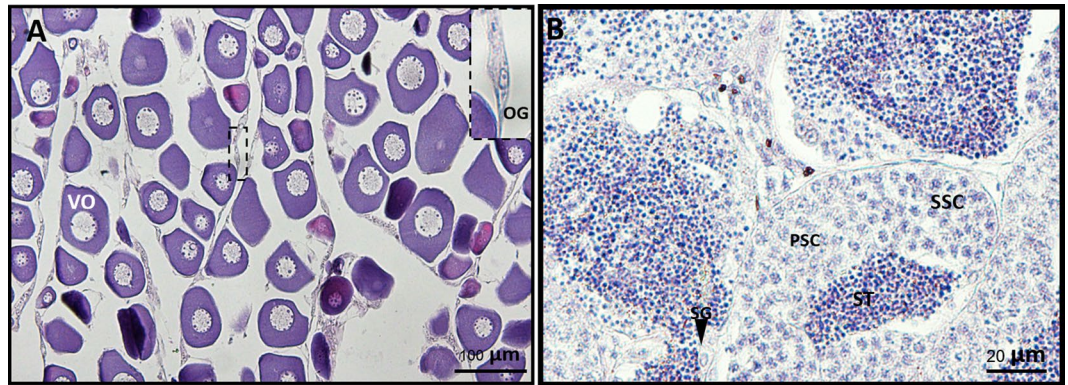
Surintorn Boonanuntasarn<sup>1\*</sup>, Araya Jangprai<sup>1</sup> & Uthairat Na-Nakorn<sup>2\*</sup>

The snakeskin gourami (*Trichopodus pectoralis*) exhibits sexual dimorphism, particularly in body size. Since the snakeskin gourami is usually marketed during sexual maturation, the sexual size dimorphism has become an economically important trait. Sex-biased gene expression plays a key role in phenotypic sexual dimorphism. Therefore, using high-throughput RNA sequencing (RNA-seq) technology, we aimed to explore the differentially expressed genes (DEGs) in ovary and testis during sex differentiation in juvenile snakeskin gourami. Our results revealed a number of DEGs were demonstrated to be overexpressed in ovary (11,625 unigenes) and testis (16,120 unigenes), and the top 10 female-biased (*rdh7*, *dnajc25*, *ap1s3*, *zp4*, *polb*, *parp12*, *trim39*, *gucy2g*, *rtbs*, and *fdxr*) and male-biased (*vamp3*, *nbl1*, *dnah2*, *ccdc11*, *nr2e3*, *spats1*, *pih1d2*, *tekt3*, *fbxo36*, and *mybl2*) DEGs were suggested to be mainly associated with ovary and testis differentiation, respectively. Additionally, using real-time reverse transcription polymerase chain reaction (qRT-PCR), validation of the differential expression of 21 genes that were previously shown to be related to gonad development was performed (*ar*, *bHLH*, *cyp19a1*, *daz*, *dead-end*, *esrb*, *esrrg*, *gnrhr*, *gpa*, *gsg1l*, *hsd17B*, *mospd1*, *nanos-1*, *nanos-2*, *p53*, *piwi-1*, *piwi-2*, *regr*, *rps6ka*, *tgf-beta*, and *VgR*). The results showed a significantly positive correlation (0.84;  $P < 0.001$ ) between the results of RNA-seq and qRT-PCR. Therefore, RNA-seq analysis in our study identified global genes that were associated with ovary and testis differentiation in the juvenile phase of the snakeskin gourami. Our findings provide valuable transcriptomic bioinformation for further investigation of reproductive biology and applications of sex manipulation.

Sexual states in fish are diverse and have been widely classified as gonochorism or hermaphroditism. Thus, fish are a suitable model to study reproductive strategies among vertebrates, particularly for sex development. Additionally, sexual plasticity in fish diversifies sexual development systems, including genetic and environmental sex-determination factors<sup>1</sup>. Several mammals possess sex chromosomes that allow the identification of genetic markers for sex determination and gene expression associated with sex differentiation and maturation, providing various approaches for reproductive manipulation<sup>2,3</sup>. Although fish exhibit the largest biodiversity among vertebrates, bioinformation regarding their sex chromosomes is limited. For example, single-nucleotide polymorphisms, linkage genes, and quantitative trait loci in a few genes or loci have been demonstrated to be related to sex determination in several fish species<sup>4–6</sup>. Furthermore, different sex-determining genes have been shown to exhibit diversity across different fish lineages<sup>7</sup>, and it has been demonstrated and suggested that young Y chromosome drives the sex-biased expression of genes<sup>8</sup>. Currently, to provide a holistic view of the evolution of reproduction biology, genetic studies on sexual system in fish have focused on sex-specific genetic differences and/or sex-biased gene expression, particularly in hermaphroditic and gonochoristic fish, which exhibit strong sexual dimorphism<sup>9–18</sup>.

Occasionally, phenotypic sexual dimorphism in fish has been observed in aquaculture-related and ornamental fish<sup>8</sup>. The snakeskin gourami (*Trichopodus pectoralis*), which is native to Southeast Asia, is an economically important food source as well as an aquarium fish. It is a type of labyrinth fish that is able to breathe air directly, in

<sup>1</sup>School of Animal Technology and Innovation, Institute of Agricultural Technology, Suranaree University of Technology, 111 University Avenue, Muang, Nakhon Ratchasima, 30000, Thailand. <sup>2</sup>Department of Aquaculture, Faculty of Fisheries, Kasetsart University, 50 Ngamvongwan Road, Chatujak, Bangkok, 10900, Thailand. \*email: [surinton@sut.ac.th](mailto:surinton@sut.ac.th); [ffisurn@ku.ac.th](mailto:ffisurn@ku.ac.th)



**Figure 1.** Histological characterization of ovary (A) and testis (B) of juvenile snakeskin gourami. Various stages of ovarian cells were found, including vitellogenic oocyte (VO) and oogonia (OG). Different stages of testicular cells were observed, including spermatogonia (SG), primary spermatocytes (PSC), secondary spermatocytes (SSC), and spermatid (ST). Scale bars represent 100  $\mu\text{m}$  (A) and 20  $\mu\text{m}$  (B).

Sample	Number of reads	Number of clean reads	Number of clean bases	Total mapped (%)	Error (%)	Q20 (%)	Q30 (%)	GC content (%)
JO1	51,743,352	48,749,338	7.3 Gb	39,541,700 (81.11)	0.02	95.87	89.94	46.56
JO2	52,352,342	50,173,912	7.5 Gb	40,528,680 (80.78)	0.02	96.05	90.50	47.11
JO3	49,695,566	46,330,550	6.9 Gb	37,638,188 (81.24)	0.02	96.32	90.79	46.46
JO4	47,515,470	45,539,184	6.8 Gb	37,123,730 (81.52)	0.02	95.89	90.08	47.58
JT1	43,092,642	49,018,366	7.4 Gb	37,674,564 (76.86)	0.02	95.18	88.54	47.85
JT2	45,252,560	48,342,728	7.3 Gb	37,553,264 (77.68)	0.02	95.86	89.99	47.43
JT3	47,702,228	42,090,162	6.3 Gb	32,983,988 (78.37)	0.02	96.12	90.52	47.53
JT4	51,291,726	39,872,726	6.0 Gb	31,520,992 (79.05)	0.02	95.85	89.79	47.61

**Table 1.** Summary of the sequencing results.

addition to absorbing oxygen through its gills, enabling it to inhabit oxygen-poor water. Snakeskin gourami naturally occurs in shallow ponds, swamps, flooded forests, and rice paddies. Morphologically, the snakeskin gourami is elongated and compressed. Compared with female fish, male fish have a slimmer and more compressed body. Snakeskin gourami mature when they are 8 months old, having a body length > 12 cm, and this is also the harvesting size of farmed fish. Therefore, the sexual size dimorphism has become an economically important trait, and female populations are more productive and preferable. Therefore, the establishment of all-female stock is attractive. Although the snakeskin gourami is one of the most important aquaculture-related species in Thailand, its genomic resources are limited. The identification of sex-biased gene expression would provide a genetic approach to understanding not only sexual dimorphism but also comparative bioinformatics in fish reproduction.

Recently, high-throughput RNA sequencing (RNA-seq) technology has been used as a novel tool for studying the global networks of gene expression. RNA-seq has enabled a rapid and cost-effective approach for preliminarily examining differences in gene expression without the requirement of prior sequence information. A genetic approach to understanding sexual dimorphism during sex differentiation would contribute to application of biotechnology for future monosex culture. In our study, therefore, comparative transcriptomic analysis during sex differentiation in juvenile snakeskin gourami was conducted for preliminarily screening sex-biased gene expression in gonads, which are the primary organs responsible for sexual development. The *de novo* transcriptome assembly served as a reference for read mapping and comparing gene expression. Functional annotation and enrichment analysis of the genes demonstrating sex-biased expression were performed. Additionally, the level of differential gene expression of 21 genes that had previously been associated with gonadal development were validated using real-time reverse transcription polymerase chain reaction (qRT-PCR).

## Results

**Histological study.** Although gender dimorphism of external morphology was not observed in snakeskin gourami during the juvenile phases, ovary and testis were distinctly observed (Fig. 1). The ovaries of sexually immature juvenile female fish contained primary oocytes and oogonia. The testis of sexually immature juvenile male fish contained various stages of testicular germ cells, including spermatids, spermatocytes, and spermatogonia.

**Sequencing assembly and functional annotation of assembled unigenes.** Approximately 56 Gb of raw reads were generated, including 388,645,886 raw reads with 370,116,966 clean reads (Table 1). Moreover, Phred quality scores of the clean reads at Q20 and Q30 ranged 95.18%–96.32% and 88.54%–90.79%, respectively.

Database	Number of unigenes (%)
Total unigenes	54,353 (100)
Annotated in nr	46,052 (84.7)
Annotated in nt	46,413 (85.4)
Annotated in KO	14,915 (27.4)
Annotated in Swiss-Prot	42,698 (78.6)
Annotated in Pfam	39,268 (72.2)
Annotated in GO	39,290 (72.3)
Annotated in KOG	27,647 (50.9)
Annotated in all databases	5,300 (9.8)
Annotated in at least one database	50,517 (92.9)

**Table 2.** Annotation of RNA-seq results. NCBI non-redundant protein database (nr), NCBI nucleotide sequences database (nt), KEGG ortholog (KO), a protein sequence database (Swiss-Prot), Protein Family (Pfam), Gene Ontology (GO), and Eukaryotic Ortholog Groups (KOG) databases were used for alignment of all the assembled unigenes.

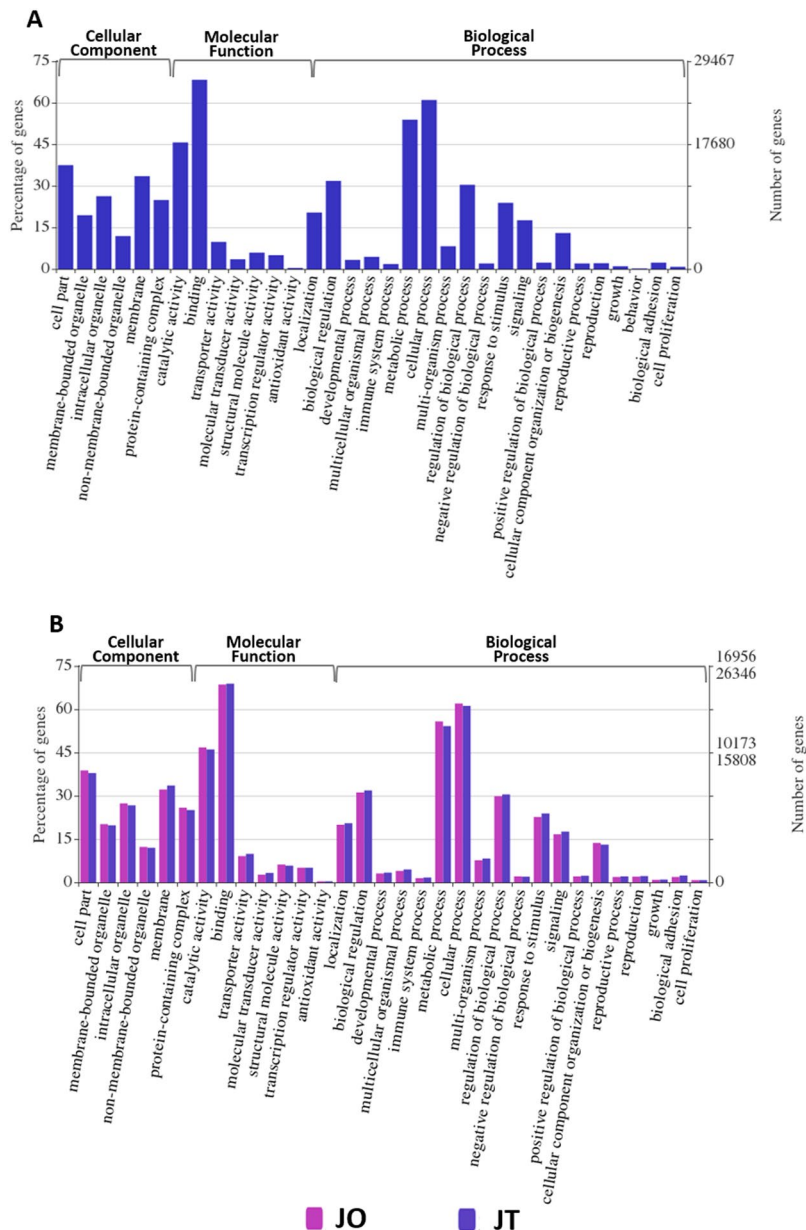
The *de novo* assembly of each clean read resulted in >80% of each JO and 76% of each JT mapping to the reference transcriptome. Among the 54,353 assembled unigenes, 50,517 (92.9%) were able to be annotated in at least 1/7 databases, (nr, nt, Swiss-Prot, KO, GO, KOG, and Pfam) (Table 2). There were 5,300 unigenes (9.8%) annotated in all seven databases. Among teleost databases, the experimental transcripts had high level of sequence identity with *Larimichthys crocea* (31.5%), *Stegastes partitus* (29.1%), *Oreochromis niloticus* (9.1%), *Maylandia zebra* (5.2%), and *Haplochromis burtoni* (4.2%) (Fig. S1). A Venn diagram was constructed to indicate the distribution of the expressed genes between testis and ovary of the juvenile fish (Fig. S2). There were 47,626 and 30,805 unigenes expressed in testis and ovary of juvenile fish, respectively, and among them, 24,078 were co-expressed between testis and ovary.

**GO classification.** To interpret global gene function, GO classification revealed 39,290 unigenes annotated into 76 functional groups of three ontologies (Fig. 2A). GO enrichment of unigenes that were expressed in either ovary or testis were provided in Table S1. In addition, Tables S2 and S3 displayed GO enrichment of unigenes that were found in ovary and testis, respectively. Our results revealed that the genes most expressed during sex differentiation were those involved in biological processes (cellular process, metabolic process and regulation of biological process, etc.). Enriched GO terms identified in ovary and testis demonstrated similar patterns (Fig. 2B). The highest enriched GO terms were matched to “binding,” “cellular process,” and “metabolic process.” Fig. 3 presents a scatterplot of the top 20 enriched KEGG pathways of DEGs. KEGG analysis of over-expressed genes in testis (JOvsJT\_down) showed a number of genes in various pathways such as amino acid metabolism, endocytosis, cell signaling and hormone metabolism (Fig. 3A, Table S4). KEGG analysis of over-expressed genes in ovary (JOvsJT\_up) displayed a number of genes in various pathways, including amino metabolism and synthesis, RNA transport and degradation, ribosome and its biogenesis, fatty acid metabolism and elongation, glycomolecule synthesis, cell cycle and apoptosis (Fig. 3B, Table S5).

**DEG and real-time RT-PCR analysis.** The volcano plotting of differential expression analysis of the assembled transcripts between ovary and testis was depicted (Fig. 4). Our study focused on the DEGs that had a  $\log_2$  fold change >1 and  $q$ -value <0.01. The results showed that 11,625 unigenes were overexpressed in ovary when compared with that in testis (Table S6), and 16,120 unigenes were overexpressed in testis when compared with that in ovary (Table S7). The top 50  $\log_2$  fold changes of DEGs were chosen to construct a hierarchical clustering, and two clusters of DEGs were delineated (Fig. 5). Table 3 presents the top 10 female-biased and male-biased genes, respectively. Additionally, the 21 expression genes that had been reported as involved in reproductive processes were validated using real-time RT-PCR analysis (Fig. 6). The  $\log_2$  fold change (JO vs JT) results obtained from qRT-PCR demonstrated a significant positive correlation (0.84;  $P < 0.001$ ) with that of the RNA-seq results. Furthermore, the coefficient of determination of a linear regression analysis ( $y = 0.796x - 2.066$ ) of expression levels between RNA-seq ( $x$ ) and real-time RT-PCR ( $y$ ) was 0.706 ( $P < 0.001$ ).

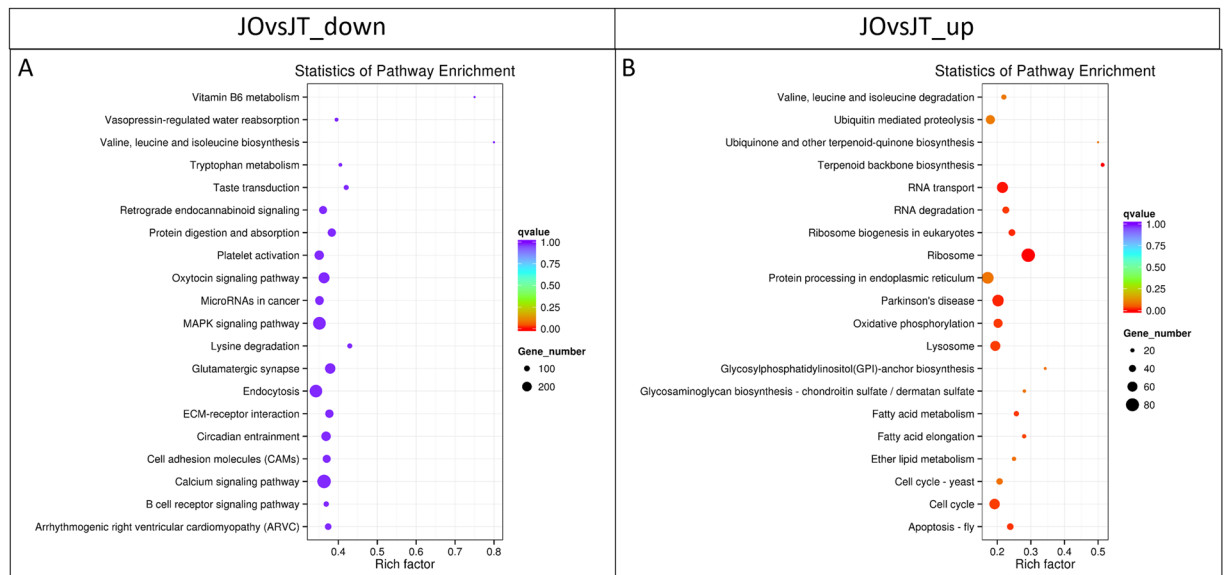
## Discussion

In our study, transcriptomic analysis between testis and ovary demonstrated global differential expression of transcripts during sex differentiation in juvenile snakeskin gourami. Our results showed that 9.8% (5,300) of the assembled unigenes had a significant annotation to all databases, which was considerably high when compared with previous reports of ovarian transcriptomic analysis of ricefield eel (*Monopterus albus*), showing only 3.36% of assembled matching to all databases<sup>12</sup>. Therefore, our transcriptome sequencing and annotating were considered acceptable for further analysis. The expressed transcripts were involved in various functional roles, indicating that a number of biological and physiological functions are involved in testis and ovary development. Similarly, high-throughput transcriptomic analysis has been applied to investigate global DEGs between sexes in gonads, brain, liver, muscle, and caudal fin in several teleosts, including gonochoristic and hermaphroditic fish<sup>9–18</sup>. These high-throughput transcriptome analyses could provide a holistic approach that will be a useful tool for further investigation of various aspects of reproductive biology.

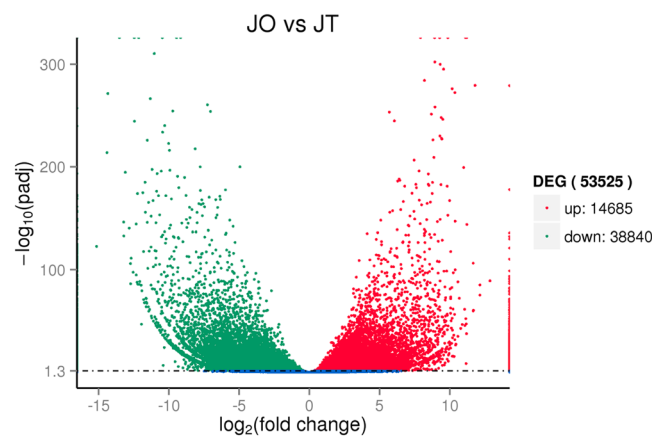


**Figure 2.** Histogram of GO classification (assigned by WEGO). **(A)** All genes that were expressed in ovary and testis. **(B)** Comparative view of GO annotations that were expressed in ovary (pink) and testis (blue). The results are presented in three main categories: biological process, cellular component, and molecular function. The left y-axis indicates the percentage of a specific category of genes in the main category. The right y-axis indicates the number of genes in each category.

The snakeskin gourami strongly exhibits sexual dimorphism in growth and external morphology. Our study focused on genes that were differentially expressed between sexes in juvenile fish. The histological study of ovary and testis revealed that they contained all the developmental stages of differentiated ovarian and testicular cells, respectively, demonstrating the sex differentiation phase of gonad development. By means of transcriptome analysis, global functional profiling of gene expression during sex differentiation in males and females revealed that a number of genes involved in biological processes, cellular components, and molecular functions. Our results implied that these genes might be associated with biological and physiological processes during gonad differentiation. Similar findings were demonstrated in bluehead wrasses (*Thalassoma bifasciatum*)<sup>9</sup>, southern bluefin tuna (*Thunnus maccoyii*)<sup>10</sup>, red porgy (*Pagrus pagrus*)<sup>13</sup>, common pandora (*Pagellus erythrinus*)<sup>13</sup>, yellow perch (*Perca flavescens*)<sup>15</sup>, and ricefield eel (*Monopterus albus*)<sup>12</sup>. Our results identified similar enriched GO terms between ovary and testis, with the highest enriched GO terms being for binding, cellular processes, and metabolic processes. In contrast, enriched GO terms for metabolic processes were observed in ovary and for signal transduction and receptor activity in testis of bluehead wrasses<sup>9</sup>.



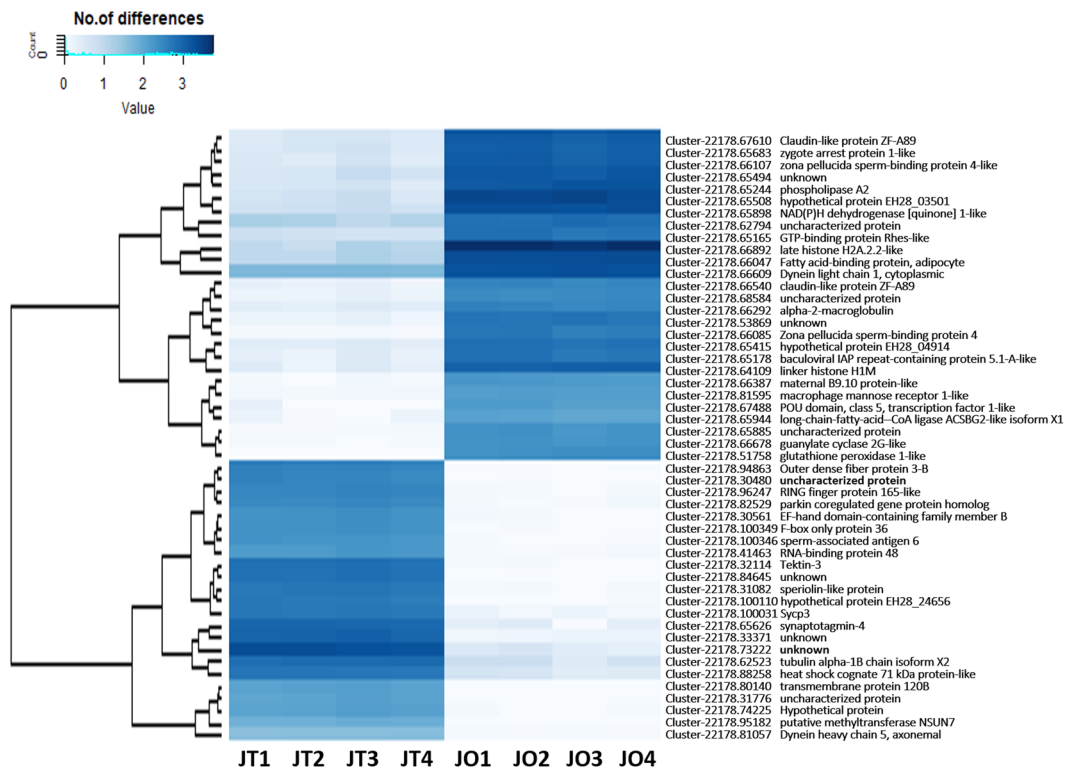
**Figure 3.** Scatter plot of the top 20 enriched KEGG pathways of DEGs. DEGs in JOvsJT\_down (A) refer to male-biased expressed genes, and those in JOvsJT\_up (B) refer to female-biased expressed genes. The x-axis represents the rich factor, which refers to the ratio of DEG numbers annotated in the pathway term to all gene numbers annotated in the pathway term. The greater the rich factor, the greater the degree of enrichment. The y-axis shows each KEGG pathway name. Each round point represents a specific KEGG pathway. The circle size indicates the number of DEGs that are associated with each significant pathway. The circle color indicates the significance level as a q-value. A q-value  $\leq 0.05$  was considered significantly enriched. Light purple shows least significant, and orange represents most significant.



**Figure 4.** Volcano plot displaying DEGs identified between ovary (JO) and testis (JT). Significantly upregulated ((female-biased genes) and downregulated (male-biased genes) expressed genes ( $\text{padj} < 0.05$ ) are denoted as red and green dots, respectively.

Our results demonstrated that the divergence of transcriptomic profiles was observed in both the number of expressed genes and their expression levels. Since the early stages of ovary development comprise vitellogenic oocytes that contain a high amount of maternal 5 S rRNA<sup>19</sup>, the high proportion of 5 S rRNA might be the reason for the relatively low number of expressed genes in ovary. Low RIN values have been reported in several transcriptomic analyses of ovary<sup>15</sup>. Therefore, during the preparation of RNA samples in our study, although low RIN values were found in some ovary samples, all RIN values  $> 5.0$  were used for further library preparation to ensure a similar amount of analyzed RNA between the testis and ovary samples. Our study found a higher number of transcripts detected in testis compared with ovary. Additionally, the number of male-biased overexpressed transcripts was higher than that of female-biased transcripts. These findings were similar to those of transcriptomic analysis of gonads of southern bluefin tuna<sup>10</sup> and bluehead wrasses<sup>9</sup>.

DEGs that were overexpressed in ovary compared with testis were considered female-biased. From the 11,625 unigenes that were overexpressed in ovary, our study focused on the top 10 unigenes that were upregulated when



**Figure 5.** Heatmap plots for differentially expressed transcripts between male and female snakeskin gourami. The intensity of the blue color in the matrix indicates the  $\log_2$  fold change. Dark blue, high  $\log_2$  fold change; light blue, low  $\log_2$  fold change. Only the top 50 DEGs were used to construct the heatmap.

compared with testis. These DEGs were identified as involved in several physiological processes in the reproductive system. For example, retinol dehydrogenase 7 (*rdh7*), also known as cis-retinol/androgen dehydrogenase type 2 (*crad2*), was demonstrated as involved in the regulation of retinol and androgen metabolism in mice (*Mus musculus*)<sup>20</sup>. DnaJ homolog, subfamily C, member 25 (*dnajc25*) encodes a protein belong to heat shock protein (HSP) 40 subfamily C that functions as co-chaperone by binding to HSP 70 to stimulate adenosine triphosphate hydrolysis. In the Pacific white shrimp (*Litopenaeus vannamei*), high levels of HSP 40 transcripts were detected in ovary, and it was suggested to play important role in ovarian development<sup>21,22</sup>. Zona pellucida sperm-binding protein (*zp*) is one of the major proteins of egg envelopment and is essential for sperm penetration and fusion of the plasma membrane to prevent polyspermy<sup>23</sup>. The expression of DNA polymerase beta (*polb*), which performs base excision repair of DNA, was demonstrated in ovarian tissue, and its overexpression increased the potential for mutagenesis<sup>24</sup>. Poly (ADP-ribose) polymerases (*parp12*) and E3 ubiquitin-protein ligase TRIM39 (*trim39*) were reported as important proteins of DNA damage repair<sup>25</sup>. Additionally, inhibition of the expression of *polb*, *parp12*, and *trim39* has been investigated for the potential to manage ovarian cancer in mammals<sup>25–27</sup>. Guanylate cyclases (*gucy*) catalyze the conversion of guanosine triphosphate to pyrophosphate and cyclic guanosine monophosphate, which is a second messenger and involved in a number of physiological processes including oocyte maturation<sup>28</sup>. Transposons in germlines were recently demonstrated as a factor driving genome evolution<sup>29</sup>. Our study identified high expression levels of RNA-directed DNA polymerase from transposon BS (*rtbs*), which is involved in retrotransposon processes, in ovary. Adrenodoxin oxidoreductase type I (*fdxr*), which obtains electrons from nicotinamide-adenine dinucleotide phosphate (NADPH) via mediation by ferredoxin, was demonstrated to localize on mitochondria and was expressed in steroidogenic tissues including ovary<sup>30</sup>. Our study provided some indication of the top 10 female-biased genes expressed, which included several related to the reproductive development of the ovary in the juvenile stage. For a holistic view of ovary development in fish, a number DEGs that were upregulated in ovary still require elucidation.

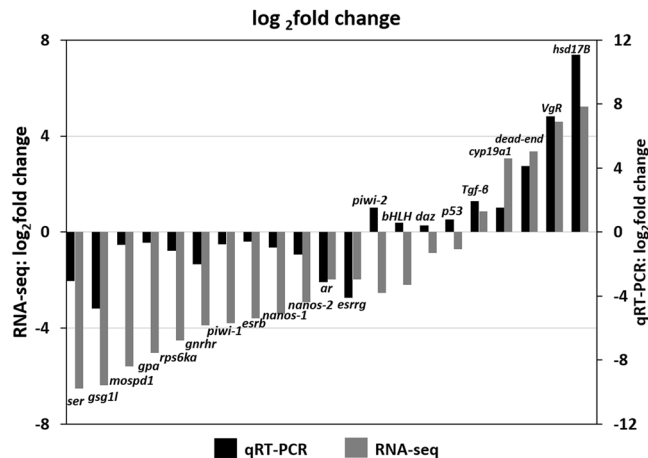
The top DEGs that were upregulated in testis when compared with ovary were identified, and they were involved in several biological mechanisms related to testis development. This included vesicle-associated membrane protein 3 (*vamp3*), a component of ternary trans-soluble N-ethylmaleimide-sensitive factor attachment protein receptor complexes of membrane structure and whole capacitated sperm<sup>31</sup>. Using transcriptome analysis, upregulation of the expression of neuroblastoma suppressor of tumorigenicity 1 (*nbl1*) was detected when zebrafish (*Danio rerio*) were exposed to 17 $\alpha$ -ethinylestradiol, and its fertility decreased<sup>32</sup>. Additionally, *nbl1* was proposed to be a candidate tumor marker for prostate cancer<sup>33</sup>. The axonemal dynein family is one of the cytoskeletal motor protein families, and its heavy chain (*dnah2*) is attributed to motor activity<sup>34</sup>. The coiled-coil domain-containing (*ccdc*) protein family was demonstrated as essential for sperm development and male fertility in mammals<sup>35,36</sup>. The PIH1 domain-containing proteins (*pih1*) have been described in the preassembly of axonemal dyneins and are required for cilia and flagella motion<sup>37</sup>. Additionally, tektin-3 (*tekt3*) is an axonemal structural

Gene name	Gene description	Gene ID	Sex demonstrating upregulation	Log <sub>2</sub> fold change
<i>rdh7</i>	retinol dehydrogenase 7	Cluster-22178.61468	F	12.86
<i>dnajc25</i>	DnaJ homolog, subfamily C, member 25	Cluster-22178.55749	F	12.14
<i>ap1s3</i>	AP-1 complex subunit sigma-3-like isoform X2	Cluster-22178.79126	F	12.14
<i>zp4</i>	Zona pellucida sperm-binding protein 4	Cluster-22178.66085	F	11.80
<i>polb</i>	DNA polymerase beta	Cluster-22178.45476	F	11.66
<i>parp12</i>	Poly (ADP-ribose) polymerase 12 isoform X2	Cluster-22178.72728	F	11.66
<i>trim39</i>	E3 ubiquitin-protein ligase TRIM39-like	Cluster-22178.64478	F	11.21
<i>gucy2g</i>	Guanylate cyclase 2G-like	Cluster-22178.66678	F	11.20
<i>rtbs</i>	RNA-directed DNA polymerase from transposon BS	Cluster-22178.58026	F	11.19
<i>fdxr</i>	NADPH:adrenodoxin oxidoreductase, mitochondrial-like isoform X1	Cluster-22178.65491	F	11.06
<i>vamp3</i>	Vesicle-associated membrane protein 3	Cluster-22178.99880	M	15.14
<i>nbl1</i>	Neuroblastoma suppressor of tumorigenicity 1	Cluster-22178.31453	M	13.21
<i>dnah2</i>	Dynein heavy chain 2, axonemal	Cluster-22178.100689	M	13.21
<i>ccdc11</i>	Coiled-coil domain-containing protein 11	Cluster-22178.100133	M	13.10
<i>nr2e3</i>	Photoreceptor-specific nuclear receptor-like	Cluster-22178.30265	M	12.72
<i>spats1</i>	Spermatogenesis-associated serine-rich protein 1 isoform X2	Cluster-22178.28827	M	12.68
<i>pih1d2</i>	PIH1 domain-containing protein 2	Cluster-22178.89036	M	12.54
<i>tekt3</i>	Tektin-3	Cluster-22178.32114	M	12.47
<i>fbxo36</i>	F-box only protein 36	Cluster-22178.100349	M	12.45
<i>mybl2</i>	Myb-related protein B-like	Cluster-22178.100072	M	12.33

**Table 3.** Top 10 DEGs between female and male of juvenile snakeskin gourami. F, female; M, male.

protein essential to sperm motility<sup>38</sup>. It was described in mammals that disruption of the expression of *dnah*, *ccdc*, or *tekt3* led to asthenozoospermia<sup>36,38–41</sup>. Furthermore, in zebrafish, genome editing to generate mutants of genes encoding PIH proteins resulted in abnormal sperm motility<sup>37</sup>. The photoreceptor-specific nuclear receptor (*nr2e3*) is an orphan nuclear receptor essential for the development and function of photoreceptor cells<sup>42</sup>. The function of *nr2e3* was related to testis development<sup>43</sup>. The expression of spermatogenesis-associated serine-rich protein 1 (*spats1*) was demonstrated as testis-specific and expressed during spermatogenesis, and a particularly high expression level was observed during meiosis of the first spermatogenic wave, mainly in pachytene spermatocytes<sup>44,45</sup>. F-box proteins contain an F-box domain that mediates protein-protein interaction<sup>46</sup>. Although information regarding the function of the F-box only protein 36 (*fbxo36*) relating to the reproductive system is limited, a high level of *fbxo36* transcripts was observed in our study. Similarly, by means of proteomic analysis, upregulation of the FBXO36 protein was detectable during sperm maturation and capacitation in the bivalve *Pecten maximus*<sup>47</sup>. Myb-related protein B (*mybl2*) was reported to activate the transcription of mitotic genes<sup>48</sup>. Taken together, these testis-upregulated genes appear to be involved in spermatogenesis and sperm motility function. Further investigation of specific roles of these and the other testis-upregulated genes are necessary to determine the entire physiological process of testis development.

qRT-PCR is a highly sensitive and specific technique used to quantitatively analyze gene expression. qRT-PCR should rather be complementary to than replaced by RNA-seq<sup>49</sup> and used to validate and confirm DEGs. Our study validated DEGs between testis and ovary using qRT-PCR for 21 genes that had previously been demonstrated as involved in gonadal development in teleosts and were specifically expressed in germ cells in various animals<sup>50–58</sup>. Our findings showed a significantly positive correlation of DEG analysis results between RNA-seq and qRT-PCR, particularly for DEGs that had a high log<sub>2</sub> fold change. Similarly, several investigations of transcriptomic analysis have described a positive correlation between RNA-seq and qRT-PCR data<sup>11,59,60</sup>. With both methods of gene expression analysis, the differential expression of several genes that play an important role in regulating reproductive functions were similar to previous transcriptomic analysis findings. For example, *hsd17b* and *cyp19a1* are involved in regulating the level of biologically active estrogens and androgens and in steroid hormone biosynthesis<sup>61</sup>. A similar overexpression of *hsd17b* and *cyp19a1* were observed in ovarian cells of southern bluefin tuna, bluehead wrasses, common pandora, and red porgy<sup>9,10,13</sup>. Additionally, the expression of *gnrhr* and *gpa*, which are related to endocrine regulation-related genes, were overexpressed in testis. Also, the testicular overexpression of *nanos-1* and *nanos-2*, which are classified as germ cell-specific genes, has been previously detected<sup>10</sup>. Our study results identified a high expression level of *piwi-1* in testis, which was similar to the findings demonstrated in bluehead wrasses<sup>9</sup>. Our study also revealed the high expression of *esrb* and *esrrg* in testis, correlating with those reported in testis of bluehead wrasses (*esr1*, *esr2a*, and *esr2b*), common pandora (*esr1a*), and red porgy (*esr1* and *esr1a*)<sup>9,13</sup>. Nevertheless, no significant differences in the expression levels of *esr1* and *esrr2b* were observed in common pandora and red porgy, respectively<sup>13</sup>. Moreover, several estrogen receptors were demonstrated to be female-biased genes in common pandora (*esrrb* and *esrr2b*) and red porgy (*esrb*)<sup>13</sup>. Furthermore, contradictory results of sex-biased gene expression were observed in *dead-end* and *ar*. For instance, while overexpressed *dead-end* was observed in ovary in our study, no significant expression differences were detected for *dead-end* in southern bluefin tuna<sup>10</sup>. Additionally, our results showed that *ar* was overexpressed in



**Figure 6.** Validation of DEGs (JO vs JT) generated from RNA-seq by qRT-PCR ( $n = 4$ ). By qRT-PCR, the expression levels of the 21 selected genes were each normalized to that of elongation factor 1 alpha (*ef-1 $\alpha$* ). The functions of the selected genes were related to gonad development and included androgen receptor (*ar*), basic helix-loop-helix domain-containing protein (*bHLH*), cytochrome P450 aromatase (*cyp19a1*), deleted in azoospermia-like protein (*daz*), dead end (*dead-end*), estrogen receptor beta (*esrb*), estrogen-related receptor gamma (*esrrg*), gonadotropin-releasing hormone II receptor (*gnrhr*), gonadotropin common alpha subunit (*gpa*), germ cell-specific gene 1-like protein (*gsg11*), estradiol 17-beta-dehydrogenase 12 (*hsd17B*), motile sperm domain-containing protein 1 (*mospd1*), nanos homolog 1 (*nanos-1*), nanos homolog 2 (*nanos-2*), cell death-inducing p53-target protein 1 (*p53*), piwi-like protein 1 (*piwi-1*), piwi-like protein 2 (*piwi-2*), ras-related and estrogen-regulated growth inhibitor-like protein (*rerg*), ribosomal protein S6 kinase alpha-3 (*rps6ka*), transforming growth factor-beta receptor-associated protein 1 (*tgf-beta*), and vitellogenin receptor isoform 2 (*VgR*).

testis. However, it exhibited no sex-biased expression in bluehead wrasses, and androgen receptor alpha and beta (*ara* and *arb*) were demonstrated as female-biased genes in common pandora and red porgy<sup>9,13</sup>.

## Conclusions

In conclusion, our study compared transcriptomic analysis between testis and ovary, demonstrating sex-biased gene expression in juvenile snakeskin gourami. Most expressed genes were annotated and categorized as involved biological processes. A number of DEGs in ovary and testis were demonstrated, and their expression was proposed as female- or male-biased, with their functions suggested to be related to ovary and testis differentiation, respectively. A positive correlation of DEGs obtained from RNA-seq and qRT-PCR were observed. Our findings demonstrated sex-associated gene expression database during sex differentiation in juvenile snakeskin gourami. This bioinformatics would offer various application for reproductive biotechnology such as sex-biased RNA-based molecular markers.

## Methods

**Ethics statement.** The study protocols of fish care and experiments were reviewed and approved by the KASETSART UNIVERSITY Institutional Animal Care and Use Committee (ACKU 61-FIS-004) and conducted in accordance with relevant guidelines.

**Fish sampling and experimental design.** The snakeskin gourami (*T. pectoralis*) used in our study were obtained from four commercial farms in central and north-east Thailand. Fifty male (body weight,  $19.22 \pm 0.23$  g; gonad weight,  $0.01 \pm 0.001$  g; gonadosomatic index [GSI],  $0.05 \pm 0.003\%$ ) and 25 female (body weight,  $18.68 \pm 0.39$  g; gonad weight,  $0.14 \pm 0.008$  g; GSI,  $0.73 \pm 0.03\%$ ) snakeskin gourami were collected for gonad sampling from each farm. The fish were euthanized using 150 mg/L MS222 before dissection. The sampled gonads were divided into two portions. One portion was stored at  $-80^\circ\text{C}$ , and the other portion was used for histological study to confirm the sex microscopically. For each replication (pond), pools of ten testis or five ovaries were collected in a sterile tube, and five tubes of each pond were kept at  $-80^\circ\text{C}$  for RNA extraction.

To perform transcriptomic analysis for comparison of differentially expressed genes (DEGs) between testis and ovary, the testis and ovary samples, each containing four replications, were used. Each replication of testicular RNA was obtained from pooled equal amounts of RNA from five tubes each containing 10 testes. Therefore, each replication of testicular RNA was obtained from 50 fish. Each replication of ovarian RNA was obtained from pooled equal amounts of RNA from five tubes each containing five ovaries. Therefore, each replication of ovarian RNA was obtained from 25 fish.

**Histological study of gonad samples.** The presence of testes or ovaries in the juvenile fish were confirmed histologically. Briefly, a portion of either ovary or testis tissue was fixed in Bouin's solution for 16 h at  $4^\circ\text{C}$ . Then, the solution was replaced with 80% ethanol, and the tissues were stored at  $4^\circ\text{C}$  until use. The fixed gonad tissues were dehydrated using a standard xylene-ethanol series and embedded in a paraffin block using a standard



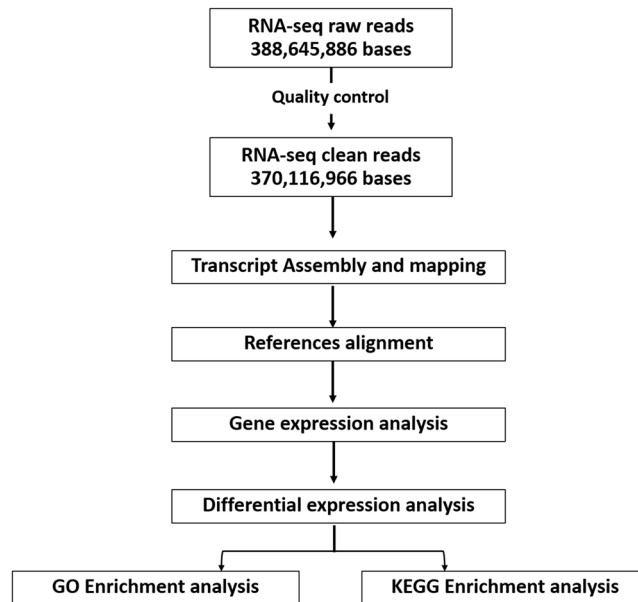
method. The paraffin tissue block was cut into 5- $\mu$ m transverse sections. The sections were then dewaxed, rehydrated, and stained with hematoxylin and eosin.

**RNA extraction and sequencing.** Total RNA was extracted from the sampled tissues (approximately 100 mg) using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and digested with RNase-free DNase I (Promega, Madison, WI, USA) according to the manufacturer's instructions. The quantity of the isolated RNA was determined using a NanoDrop<sup>®</sup> ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Agarose gel electrophoresis was also performed to assess RNA degradation. Additionally, the quality and integrity of the isolated RNA was analyzed using an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The RNA obtained from the testes samples that had an RNA integrity number (RIN) value >7.0 were used for further analysis. Since the RNA isolated from ovary samples, particularly during the juvenile stage, contained a high amount of low molecular weight 5S RNA, the RNA obtained from ovary samples with a RIN value >4.5 were used for RNA processing. Four replicates of testicular and ovarian RNA samples (2  $\mu$ g) were used to perform mRNA paired-end library construction with a TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. mRNA enrichment was carried out via poly-A mRNA isolation with oligo-dT beads. Subsequently, mRNA fragmentation was performed. First-strand cDNA synthesis was then carried out on all samples using random hexamer primers and reverse transcriptase. To generate the second strand by nick-translation, a custom second-strand synthesis buffer (Illumina) was added with dNTPs, RNase H, and *Escherichia coli* polymerase I. Agencourt AMPure XP beads were used to purify the cDNA (Beckman Coulter, Brea, CA, USA). Terminal repair, A-tailing, ligation of sequencing adapters, size selection, and PCR enrichment were applied to produce the cDNA libraries. Four libraries each were prepared for ovaries (JO1, JO2, JO3, and JO4) or testis (JT1, JT2, JT3, and JT4). Subsequently, the cDNA libraries were quantified using a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA) and then diluted to 1 ng/ $\mu$ L. The insert size (150–200 bp) was checked using an Agilent 2100 Bioanalyzer, and the concentration was determined by quantitative PCR (library activity >2 nM). Cluster generation was performed on a cBot cluster Generation System using a TruSeq PE Cluster Kit v3-cBot-HS (both Illumina). All libraries were loaded onto a HiSeq. 2500 Sequencing System (Illumina) as per the manufacturer's protocol at Novogene Bioinformatics Institute, Beijing, China.

**Preprocessing, *de novo* assembly, and read preprocessing.** Figure 7 displayed the workflow of RNA-seq and differential expression analysis. The original raw data from sequencing were converted to sequenced reads by base calling. In total, 388.65 million bases (Gb) of raw reads were generated by the Illumina HiSeq platform. (Table 1). The raw sequence files generated from 8 files (fastq) has been deposited to NCBI's Sequence Read Archive (SRA) database with the accession number PRJNA597181 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA597181>). Then, to obtain clean reads, the raw reads were filtered to remove those with adapter sequences, containing poly-N, or of low quality (low quality nucleotides constitute more than 50% of the read). The base calling and quality assignment were evaluated using the Phred score, calculated as  $Q_{\text{phred}} = -10 \log_{10}(e)$  (Table 1). Eight sets of clean reads were assembled *de novo* using Trinity 2.0.6 software with the default parameters (k-mer, 25; minimum length, 200 nucleotides)<sup>62</sup>, and then Corset software was used to perform hierarchical clustering to remove redundancy<sup>63</sup>. Finally, the longest transcript of each cluster was selected as the unigene.

**Transcriptome functional annotation.** For transcriptome functional annotation, seven databases were searched. A similarity search of nr (National Center for Biotechnology Information [NCBI] non-redundant protein sequences), nt (NCBI BLAST 2.2.28 +), and Swiss-Prot databases was performed with the E-value threshold at  $10^{-5}$  to obtain the top 10 alignment results for annotation of the assembled transcripts. The unigenes were also searched against the Eukaryotic Orthologous Groups (KOG) database, with the significant E-value threshold set at  $10^{-3}$ . Additionally, the unigene annotations were searched against the Protein Family (Pfam) database using the HMMER 3.0 software package with the E-value threshold at 0.01. For gene ontology (GO) mapping, based on the protein annotation results of the BLAST search and Pfam, Blast2GO 2.5<sup>64</sup> was used to assign the GO annotation associated with the hits obtained by the BLAST search and Pfam, with a significant E-value threshold of  $10^{-6}$  for describing biological processes, molecular functions, and cellular components. The sequences were also annotated and characterized by Kyoto Encyclopedia of Genes and Genomes (KEGG), with a significant E-value threshold of  $10^{-10}$ . The species distribution of the top BLASTX results matched to the nt database was also analyzed.

**Analysis of gene expression levels and differential expression.** The *de novo* transcriptome assembly served as a reference for read mapping. The clean reads were mapped back onto the assembled transcriptome with Bowtie 2 software, and the read count of each gene from each sample was estimated. The gene expression levels of each sample were quantified using RSEM software<sup>65</sup>. The read count for each gene of each sample was normalized to fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM)<sup>66</sup>. Contigs with a low expression (FPKM < 0.3) were excluded. The read count and FPKM for each gene of each sample were demonstrated in Tables S8 and S9. A Venn diagram was constructed to present the number of unigenes that were uniquely expressed and co-expressed in ovary and testis. Differential expression analysis was estimated at unigene level by pairwise comparisons between males and females using the R package DEGseq<sup>67</sup>. The false discovery rate (FDR) *P*-value was adjusted using the *q*-value. A *q*-value < 0.01 and  $\log_2$  fold change > 1 were set as the thresholds for significant DEGs. GO enrichment analysis of the DEGs was carried out for biological processes, cellular components, and molecular function by WEGO 2.0 software (<http://wego.genomics.org.cn/>). Enrichment analysis of DEGs in KEGG pathways was carried out for KEGG enrichment scattered plots to determine the top 20 significant DEG enriched pathways. A volcano plot for inferring the overall distribution of DEGs was constructed using



**Figure 7.** Workflow of RNA sequencing, data processing, and analysis.

a padj (FDR) threshold of  $<0.05$ . Heatmaps of DEGs were constructed using the R package gplots to display the top 50 significant DEGs.

**qRT-PCR analysis.** qRT-PCR analysis was used to quantitatively validate 21 genes that had been demonstrated as involved in the reproductive system, including androgen receptor (*ar*), basic helix-loop-helix domain-containing protein (*bHLH*), cytochrome P450 aromatase (*cyp19a1*), deleted in azoospermia-like protein (*daz*), dead end (*dead-end*), estrogen receptor beta (*esrb*), estrogen-related receptor gamma (*esrrg*), gonadotropin-releasing hormone II receptor (*gnrhr*), gonadotropin common alpha subunit (*gpa*), germ cell-specific gene 1-like protein (*gsg1l*), estradiol 17-beta-dehydrogenase 12 (*hsd17B*), motile sperm domain-containing protein 1 (*mospd*), nanos homolog 1 (*nanos-1*), nanos homolog 2 (*nanos-2*), cell death-inducing p53-target protein 1 (*p53*), piwi-like protein 1 (*piwi-1*), piwi-like protein 2 (*piwi-2*), ras-related and estrogen-regulated growth inhibitor-like protein (*reng*), ribosomal protein S6 kinase alpha-3 (*rps6ka*), transforming growth factor-beta receptor-associated protein 1 (*tgf-beta*), and vitellogenin receptor isoform 2 (*VgR*). Briefly, using the same RNA samples (2  $\mu$ g) as for the RNA-seq library preparation, the first-strand cDNA was synthesized using the ImProm-II<sup>TM</sup> Reverse Transcription System kit (Promega). qRT-PCR amplification (in duplicate) was conducted using LightCycler<sup>®</sup> 480 SYBR Green I Master (Roche Applied Science, Indianapolis, IN, USA). Elongation factor 1 alpha (*ef-1 $\alpha$* ) was used as an internal reference for data normalization. Table S10 lists the primers and annealing temperatures used in our study and the expected size of the amplicons. PCR samples were prepared in a final volume of 15  $\mu$ l consisting of 7.5  $\mu$ l of LightCycler<sup>®</sup> 480 SYBR Green I Master, 500 nM of each primer, and 1.5  $\mu$ l of cDNA template. PCR was performed at 95  $^{\circ}$ C for 10 min followed by 50 reaction cycles consisting of 15 s at 95  $^{\circ}$ C, 15 s at the annealing temperature (Table S1), and 20 s at 72  $^{\circ}$ C. Upon completion of amplification, all samples were subjected to melting curve analysis to distinguish between the PCR products. Each PCR assay included triplicate PCR amplification and the negative controls (reverse-transcriptase- and cDNA-template-free samples). To analyze the mRNA levels, relative quantification of target gene expression was carried out using the Roche Applied Science E-method. The data were analyzed using the comparative cycle threshold method. PCR efficiency was measured by the slope of a standard curve constructed using serial dilutions of cDNA. In all cases, PCR efficiency values ranged between 1.8 and 2.2. The expression of *ef-1 $\alpha$*  was used as an internal reference gene prior to calculating DEGs using  $\log_2$  fold change. In addition, Pearson's correlation coefficient was used to calculate the relationship of the results of RNA-seq and qRT-PCR. Moreover, regression analysis of the results of RNA-seq and qRT-PCR and goodness of fit ( $R^2$ ) were performed.

**Ethics approval.** This study followed the Kasetsart University Institutional Animal Care and Use Committee. Fish care and experiments were approved by the Kasetsart University Institutional Animal Care and Use Committee (ACKU-61-FIS-004).

### Data availability

All data from this work are public and included in this article in the main manuscript or as additional files. Additional datasets and information are available from the corresponding author under reasonable request.

Received: 30 October 2019; Accepted: 19 February 2020;

Published online: 23 March 2020

## References

- Kobayashi, Y., Nagahama, Y. & Nakamura, M. Diversity and plasticity of sex determination and differentiation in fishes. *Sexl Dev* 7(1–3), 115–125 (2013).
- Gilbert, S. F. *Developmental Biology*. 6th edition. Sunderland (MA): Sinauer Associates; Chromosomal Sex Determination in Mammals. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK9967/> (2000).
- Angelopoulou, R., Lavranos, G. & Manolakou, P. Sex determination strategies in 2012: towards a common regulatory model? *Reprod. Biol. Endocrin.* 10(1), 13 (2012).
- Loukovitis, D. *et al.* Quantitative trait loci involved in sex determination and body growth in the gilthead sea bream (*Sparus aurata* L.) through targeted genome scan. *PLoS One* 6(1), e16599 (2011).
- Bradley, K. M. *et al.* An SNP-Based Linkage Map for Zebrafish Reveals Sex Determination Loci. *G3: Genes, Genom. Genet.* 1(1), 3–9 (2011).
- Kamiya, T. *et al.* A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genet.* 8(7), e1002798 (2012).
- Kikuchi, K. & Hamaguchi, S. Novel sex determining genes in fish and sex chromosome evolution. *Dev. Dyn.* 242(4), 339–353 (2013).
- Dixon, G., Kitano, J. & Kirkpatrick, M. The origin of a new sex chromosome by introgression between two stickleback fishes. *Mol. Boil. Evol.* 36(1), 28–38 (2018).
- Liu, H. *et al.* Large-scale transcriptome sequencing reveals novel expression patterns for key sex-related genes in a sex-changing fish. *Biol. sex differ.* 6(1), 26 (2015).
- Bar, I., Cummins, S. & Elizur, A. Transcriptome analysis reveals differentially expressed genes associated with germ cell and gonad development in the Southern bluefin tuna (*Thunnus maccoyii*). *BMC genomics* 17(1), 217 (2016).
- Wang, P. *et al.* 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 (2016).
- Cai, J. *et al.* Transcriptomic analysis of the differentiating ovary of the protogynous ricefield eel *Monopterus albus*. *BMC Genomics* 18(1), 573 (2017).
- Tsakogiannis, A. *et al.* The transcriptomic signature of different sexes in two protogynous hermaphrodites: Insights into the molecular network underlying sex phenotype in fish. *Sci. Rep.* 8(1), 3564 (2018).
- Cribbin, K. M., Quackenbush, C. R., Taylor, K., Arias-Rodriguez, L. & Kelley, J. L. Sex-specific differences in transcriptome profiles of brain and muscle tissue of the tropical gar. *BMC Genomics* 18(1), 283 (2017).
- Li, Y.-H. *et al.* De novo transcriptome sequencing and analysis of male, pseudo-male and female yellow perch, *Perca flavescens*. *PLoS One* 12(2), e0171187 (2017).
- Sutherland, B. J., Prokkola, J. M., Audet, C. & Bernatchez, L. Sex-specific co-expression networks and sex-biased gene expression in the salmonid Brook Charr *Salvelinus fontinalis*. *G3: Genes, Genom. Genet.* 9(3), 955–968 (2019).
- Dick, C., Reznick, D. N. & Hayashi, C. Y. Sex-biased expression between guppies varying in the presence of ornamental coloration. *Peer J* 6, e5782 (2018).
- Martinez, P. *et al.* Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front. in Genet.* 5, 340 (2014).
- Diaz de Cerio, O., Rojo-Bartolomé, I., Bizarro, C., Ortiz-Zarragoitia, M. & Cancio, I. 5S rRNA and accompanying proteins in gonads: powerful markers to identify sex and reproductive endocrine disruption in fish. *Environ. Sci. Technol.* 46(14), 7763–7771 (2012).
- Tomita, K. *et al.* Gene structure and promoter for *Crad2* encoding mouse cis-retinol/3 $\alpha$ -hydroxysterol short-chain dehydrogenase isozyme. *Gene* 251(2), 175–186 (2000).
- Young, J. C. Mechanisms of the Hsp70 chaperone system. *Biochem. Cell Biol.* 88(2), 291–300 (2010).
- Chen, T. *et al.* Heat shock protein 40 (HSP40) in Pacific white shrimp (*Litopenaeus vannamei*): Molecular cloning, tissue distribution and ontogeny, response to temperature, acidity/alkalinity and salinity stresses, and potential role in ovarian development. *Front. Physiol.* 9, 1784 (2018).
- Arukwe, A. & Goksoyr, A. Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. *Comp. Hepatol.* 2(1), 4 (2003).
- Chan, K. *et al.* Overexpression of DNA polymerase  $\beta$  results in an increased rate of frameshift mutations during base excision repair. *Mutagenesis* 22(3), 183–188 (2007).
- Ohmoto, A. & Yachida, S. Current status of poly (ADP-ribose) polymerase inhibitors and future directions. *Oncotargets Ther.* 10, 5195 (2017).
- Katchman, B. A. *et al.* Autoantibody biomarkers for the detection of serous ovarian cancer. *Gynecol. Oncol.* 146(1), 129–136 (2017).
- Markman, M. Poly (ADP-ribose) polymerase inhibitors in the management of ovarian cancer. *Women's Health* 14:1745505717750694 (2018).
- Potter, L. R. Guanylyl cyclase structure, function and regulation. *Cell. Signal.* 23(12), 1921–1926 (2011).
- Durdevic, Z., Pillai, R. S. & Ephrussi, A. Transposon silencing in the *Drosophila* female germline is essential for genome stability in progeny embryos. *Life Sci. Alliance* 1(5), e201800179 (2018).
- Yazawa, T. *et al.* Transcriptional Regulation of Ovarian Steroidogenic Genes: Recent Findings Obtained from Stem Cell-Derived Steroidogenic Cells. *BioMed Res. Int.* (2019).
- Tsai, P.-S., Garcia-Gil, N., Van Haeften, T. & Gadella, B. M. How pig sperm prepares to fertilize: stable acrosome docking to the plasma membrane. *PLoS One* 5(6), e11204 (2010).
- Porseryd, T. *et al.* Testis transcriptome alterations in zebrafish (*Danio rerio*) with reduced fertility due to developmental exposure to 17 $\alpha$ -ethinyl estradiol. *Gen. Comp. Endocrin.* 262, 44–58 (2018).
- Hayashi, T. *et al.* The search for secreted proteins in prostate cancer by the *Escherichia coli* ampicillin secretion trap: expression of NBL1 is highly restricted to the prostate and is related to cancer progression. *Pathobiol.* 80(2), 60–69 (2013).
- Roberts, A. J., Kon, T., Knight, P. J., Sutoh, K. & Burgess, S. A. Functions and mechanics of dynein motor proteins. *Nature Rev. Mol. Cell Biol.* 14(11), 713 (2013).
- Pasek, R. C. *et al.* Coiled-coil domain containing 42 (*Ccdc42*) is necessary for proper sperm development and male fertility in the mouse. *Dev. Biol.* 412(2), 208–218 (2016).
- Lehti, M. S. & Sironen, A. Formation and function of sperm tail structures in association with sperm motility defects. *Biol. Reprod.* 97(4), 522–536 (2017).
- Yamaguchi, H., Oda, T., Kikkawa, M. & Takeda, H. Systematic studies of all PIH proteins in zebrafish reveal their distinct roles in axonemal dynein assembly. *Elife* 7, e36979 (2018).
- Roy, A., Lin, Y. N., Agno, J. E., DeMayo, F. J. & Matzuk, M. M. Tektin 3 is required for progressive sperm motility in mice. *Mol. Reprod. Dev.: Incorporat. Gamete Res.* 76(5), 453–459 (2009).
- Neesen, J. *et al.* Disruption of an inner arm dynein heavy chain gene results in asthenozoospermia and reduced ciliary beat frequency. *Human Mol. Genetics* 10(11), 1117–1128 (2001).
- Roy, A., Yan, W., Burns, K. H. & Matzuk, M. M. Tektin3 encodes an evolutionarily conserved putative testicular microtubules-related protein expressed preferentially in male germ cells. *Mol. Reprod. Dev.: Incorporat. Gamete Res.* 67(3), 295–302 (2004).
- Becker-Heck, A. *et al.* The coiled-coil domain containing protein *CCDC40* is essential for motile cilia function and left-right axis formation. *Nature Genet.* 43(1), 79 (2011).

42. Achermann, J. C., Schwabe, J., Fairall, L. & Chatterjee, K. Genetic disorders of nuclear receptors. *J. Clin. Invest.* **127**(4), 1181–1192 (2017).
43. Bodofsy, S., Koitz, F. & Wightman, B. Conserved and exapted functions of nuclear receptors in animal development. *Nucl. Recept. Res.* **4** (2017).
44. Geisinger, A., Wettstein, R. & Benavente, R. Stage-specific gene expression during rat spermatogenesis: application of the mRNA differential display method. *Int. J. Dev. Biol.* **40**(1), 385–388 (2003).
45. Capovano, C. A., Wettstein, R., Kun, A. & Geisinger, A. Spats 1 (Srsp1) is differentially expressed during testis development of the rat. *Gene Express. Patterns* **10**(1), 1–8 (2010).
46. Kipreos, E. T. & Pagano, M. The F-box protein family. *Genome Biol.* **1**(5): **3001**, reviews3002 (2000).
47. Boonmee, A. *et al.* Differential protein expression during sperm maturation and capacitation in an hermaphroditic bivalve, *Pecten maximus* (Linnaeus, 1758). *J. Molluscan Stud.* **82**(4), 575–584 (2016).
48. Guiley, K. Z. *et al.* Structural mechanism of Myb–MuvB assembly. *Proc. Natl. Acad. Sci.* **115**(40), 10016–10021 (2018).
49. Costa, C., Giménez-Capitán, A., Karachaliou, N. & Rosell, R. Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl. Lung Cancer Res.* **2**(2), 87 (2013).
50. Raz, E. Primordial germ-cell development: the zebrafish perspective. *Nature Rev. Genet.* **4**(9), 690 (2003).
51. Siep, M. *et al.* Basic helix–loop–helix transcription factor Tcf15 interacts with the Calmegin gene promoter in mouse spermatogenesis. *Nucleic Acids Res.* **32**(21), 6425–6436 (2004).
52. Sugiura, K., Naito, K., Endo, T. & Tojo, H. Study of germinal vesicle requirement for the normal kinetics of maturation/M-phase-promoting factor activity during porcine oocyte maturation. *Biol. Reprod.* **74**(3), 593–600 (2006).
53. Yasuda, T. *et al.* Gamma-ray irradiation promotes premature meiosis of spontaneously differentiating testis–ova in the testis of p53-deficient medaka (*Oryzias latipes*). *Cell Death & Disease* **3**(10), e395 (2012).
54. Clark, J. P. & Lau, N. C. Piwi proteins and piRNAs step onto the systems biology stage. In: *Systems Biology of RNA Binding Proteins*. Springer 159–197 (2014).
55. Zhu, F., Schlupp, I. & Tiedemann, R. Sequence evolution and expression of the androgen receptor and other pathway-related genes in a unisexual fish, the Amazon molly, *Poecilia formosa*, and its bisexual ancestors. *PLoS One* **11**(6), e0156209 (2016).
56. Cahoreau, C., Klett, D. & Combarnous, Y. Structure–function relationships of glycoprotein hormones and their subunits' ancestors. *Front. Endocrinol.* **6**, 26 (2015).
57. Lu, J. *et al.* Transcriptomic analyses reveal novel genes with sexually dimorphic expression in yellow catfish (*Pelteobagrus fulvidraco*) brain. *Marine Biotechnol.* **17**(5), 613–623 (2015).
58. Zheng, S., Long, J., Liu, Z., Tao, W. & Wang, D. Identification and Evolution of TGF- $\beta$  Signaling Pathway Members in Twenty-Four Animal Species and Expression in Tilapia. *Int. J. Mol. Sci.* **19**(4), 1154 (2018).
59. Hu, F. *et al.* Different expression patterns of sperm motility-related genes in testis of diploid and tetraploid cyprinid fish. *Biol. Reprod.* **96**(4), 907–920 (2017).
60. Chen, Y. *et al.* Gonadal transcriptome sequencing of the critically endangered *Acipenser dabryanus* to discover candidate sex-related genes. *PeerJ* **6**, e5389 (2018).
61. Mindnich, R. & Adamski, J. Zebrafish 17 $\beta$ -hydroxysteroid dehydrogenases: an evolutionary perspective. *Mol. Cell Endocrinol.* **301**(1–2), 20–26 (2009).
62. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.* **29**(7), 644 (2011).
63. Davidson, N. M. & Oshlack, A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology* **15**(7), (2014).
64. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids Res.* **36**(10), 3420–3435 (2008).
65. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**(1), 323 (2011).
66. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**(5), 511 (2010).
67. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**(10), R106 (2010).

## Acknowledgements

The authors express their gratitude for funding support from Thailand Science Research and Innovation (former Thailand Research Fund-TRF and Betagro Science Center Co. Ltd. (Contract No. DPG5980003; Distinguished Research Professor TRF FY 2016) awarded to UN. This work was also supported by Suranaree University of Technology (SUT) and by the Office of the Higher Education Commission under NRU Project of Thailand.

## Author contributions

S. Boonanuntanasarn conceived research idea and directed the experimental design. S. Boonanuntanasarn and A. Jangprai participated in sample preparation, carried out experiments and data analysis. In addition, U. Nakanorn provided conceptual of the whole project. S. Boonanuntanasarn prepared manuscript. All authors have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-61738-0>.

**Correspondence** and requests for materials should be addressed to S.B. or U.N.-N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020