

RESEARCH ARTICLE

Open Access



Transcriptome profiling analysis of sex-based differentially expressed mRNAs and lncRNAs in the brains of mature zebrafish (*Danio rerio*)

Wenliang Yuan^{1,2,3,4,5†}, Shouwen Jiang^{1,2,3†}, Dan Sun^{1,2,3}, Zhichao Wu^{1,2,3}, Cai Wei^{1,2,3}, Chaoxu Dai, Linhua Jiang^{4*} and Sihua Peng^{1,2,3*}

Abstract

Background: Similar to humans, the zebrafish brain plays a central role in regulating sexual reproduction, maturation and sexual behavior. However, systematic studies of the dimorphic patterns of gene expression in the brain of male and female zebrafish are lacking.

Results: In this study, the mRNA and lncRNA expression profiles were obtained from the brain tissue samples of the three male and three female zebrafish by high-throughput transcriptome sequencing. We identified a total of 108 mRNAs and 50 lncRNAs with sex-based differential expression. We randomly selected four differentially expressed genes for RT-qPCR verification and the results certified that the expression pattern showed a similar trend between RNA-seq and RT-qPCR results. Protein-protein interaction network analysis, Gene Ontology (GO) analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed to obtain the biological significance of differentially expressed mRNA in the brain dimorphism of zebrafish. Finally, a Pearson correlation analysis was performed to construct the co-expression network of the mRNAs and lncRNAs.

Conclusions: We found that 12 new lncRNAs not only have significant gender specificity in the brain of zebrafish, and this finding may provide a clue to further study of the functional difference between male and female zebrafish brain.

Keywords: Zebrafish, Brain, The dimorphic patterns, mRNA and lncRNAs, High-throughput transcriptome sequencing

Background

The zebrafish (*Danio rerio*) is a very useful model animal for the comparative study of neuroscience [1], because its brain has behavioral and morphological sexual dimorphisms. Although the sex-related chromosomal regions of zebrafish are not fixed [2, 3] and their genome of different sexes can be very similar; we can observe its sexual dimorphism by differences in gene expression [4]. By

using microarray technology, studies have found 24 gender differential gene expression in mature zebrafish brain [5]. The zebrafish exposed to the sex hormone showed a regional (forebrain, midbrain and hindbrain) and gender-related differences in gene expression [6]. The expression of 61 genes in the four zebrafish showed significant gender differences by RNA-seq sequencing data analysis [7]. Arslan-Ergul et al. also found that there are expression differences of various genes between genders and for different ages, which are associated with multiple pathways in zebrafish brain [8].

Long non-coding RNAs (long ncRNAs, lncRNAs), which may appear anywhere in the genome, are defined as transcripts longer than 200 nucleotides that are not

* Correspondence: lhjiang@usst.edu.cn; shpeng@shou.edu.cn

[†]Wenliang Yuan and Shouwen Jiang contributed equally to this work.

⁴School of Optical-Electric and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

¹Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources (Shanghai Ocean University), Ministry of Education, Shanghai 201306, China
Full list of author information is available at the end of the article



translated into protein [9]. Non-functional lncRNAs are likely to be the result of transcriptional noise, whereas functional lncRNAs act in *cis* and/or in *trans* [10]. The over-expression or deficiency of lncRNAs is involved in numerous human diseases [11]. Guttman M et al. predicted that lncRNA plays an important role in cell pluripotency and cancer [12]. It was also found that some regulatory genes and lncRNAs may play a key role in development and hematopoiesis through processing functional coupling network using deep RNA-seq sequencing [13]. Durga, a novel non-coding RNA, arising from the first exon of Kalirin, is a key player in axonal development in zebrafish and maintains dendritic length and density by regulating kalrna expression [14]. Similarly, many lncRNAs with biological functions have been found in zebrafish [15–18]; however the function of the sex-based lncRNAs in the brain is still unknown.

In this study, we obtained the RNA expression data of the brains of the adult male and female zebrafish by using transcriptome sequencing (RNA-seq). Differential expression mRNAs and lncRNAs were screened using computational methods. Some of the differential expression mRNAs and lncRNAs have been verified by RT-qPCR. Additionally, we found the mRNAs significantly enriched in many pathways according to the GO and KEGG functional enrichment analyses. Finally, the lncRNA-mRNA interaction network was constructed using the Pearson correlation coefficients, based on the FPKM values of the lncRNAs and mRNAs. This study expanded the zebrafish brain sex-based lncRNA catalogue and constructed a regulatory network of the zebrafish brains at the transcriptional level, providing clues for more in-depth depiction of gender differences in zebrafish brain neurons.

Results

Sequencing data, raw data filtering, and mapping of RNA sequencing reads onto the zebrafish genome

High-throughput sequencing generated 113.71 G bp of raw data (Additional file 1), and after filtering, the clean data of 98.52 G bp were extracted. Then, these high-quality reads were mapped to the reference zebrafish genome by 89.2%. The uniquely mapped reads ranged from 78.3 to 83.4%.

Differentially expressed genes

We analyzed the transcriptome data of the six zebrafish brain tissue samples to obtain the sex-based gene expression of zebrafish in brain. Using the TopHat2 and Cufflinks packages, 14,315 annotated genes were obtained, accounting for 93.4% of the total genes assembled in the danRer10 zebrafish genome. Volcano and hierarchical clustering showed that the sex-based gene expression levels were distinguishable and statistically significant (Fig. 1a, b). We

identified seven female-based genes and 101 male-based genes (fold-change > 2 and *P*-value < 0.05) (Additional file 2). When taking into account the direction of expression, approximately 93% (101/108) of the differentially expressed genes showed male-based expression (Additional file 2). Through literature search, we found that the detailed interpretation for majority of the differentially expressed male-based genes were not available. The expression levels of *f13a1a.1*, *zgc:114181* and *hbaa2* were up-regulated in the female zebrafish by 4.3, 2.8 and 2.6-folds, respectively, while *apoa2*, *leg1.1*, and *c3a.1* were down-regulated by 10.37, 9.38 and 8.54-folds, respectively. We randomly selected four differentially expressed genes (*zgc: 114181*, *f13a1a.1*, *vtna* and *rbp2a*, 2 up-regulated and 2 down-regulated) for RT-qPCR verification, with the results indicating that the expression pattern showed a similar trend between RNA-seq and RT-qPCR results (Fig. 1c).

To verify the interrelationship between the differentially expressed genes, we constructed a protein-protein interaction network of these regulated genes with 44 nodes and 54 interactions (Fig. 1d). In this network, kininogen 1 (*kng1*, degree = 8, the degree of a node is the number of edges connecting to other nodes), serpin peptidase inhibitor clade E 1 (*serpine1* degree = 7) and coagulation factor XIII A1 polypeptide a tandem duplicate1 (*f13a1a.1*, degree = 7) indicated higher degrees.

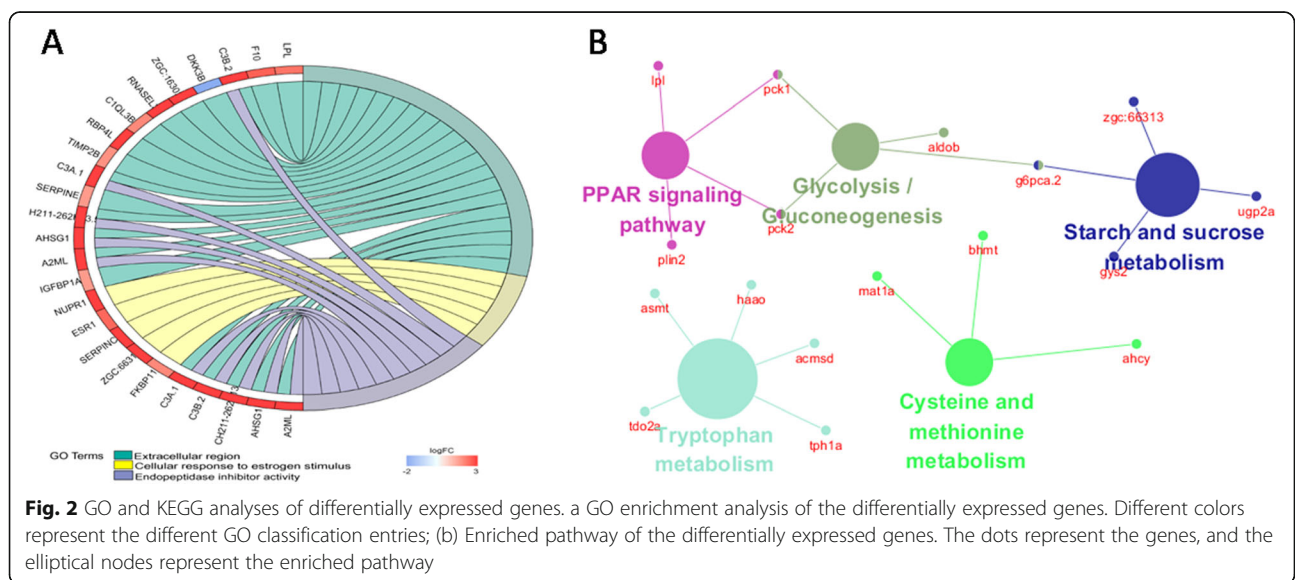
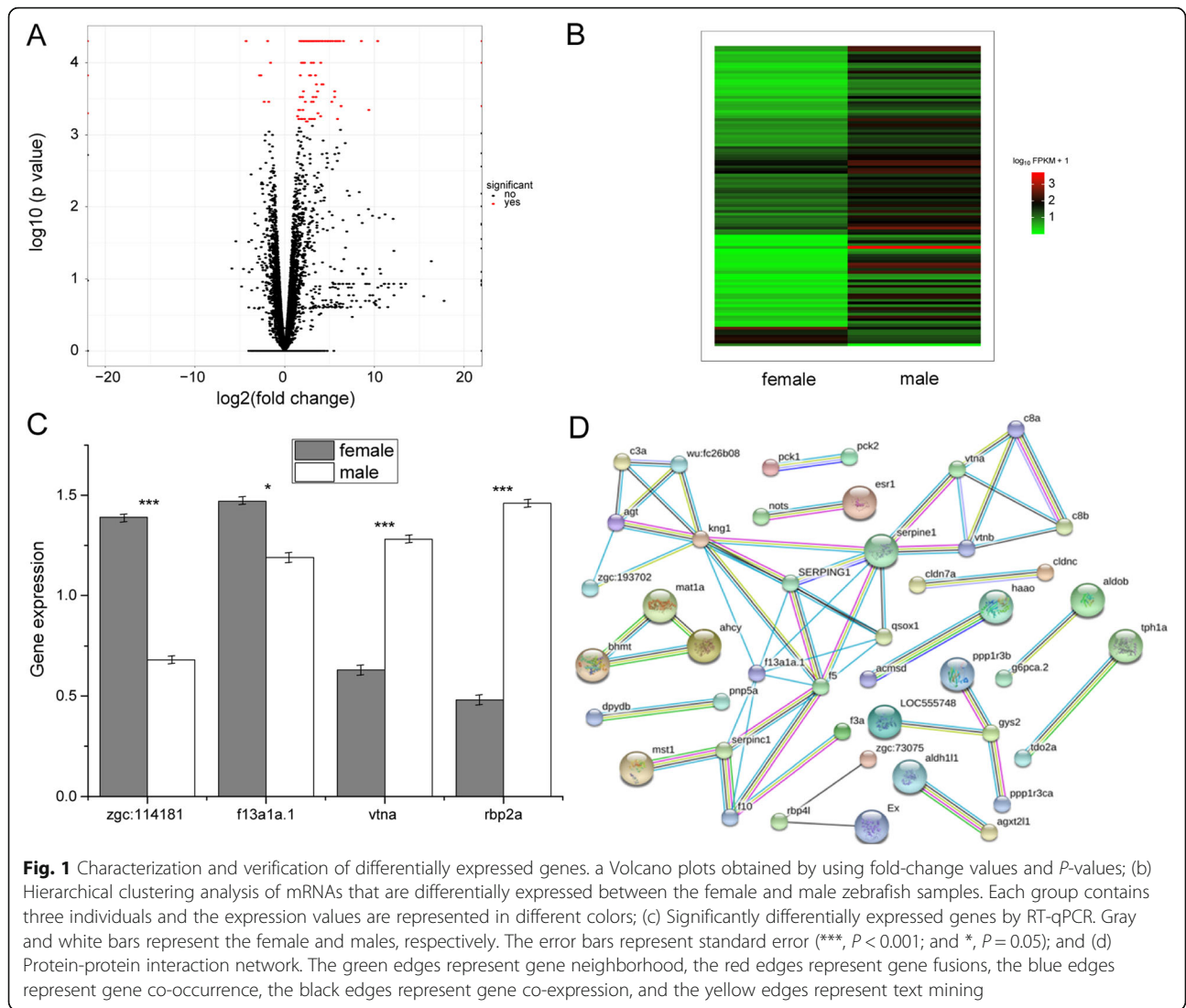
Gene ontology and KEGG analyses

To investigate the function of the differentially expressed sex-based genes, GO enrichment analysis and KEGG pathway annotation were performed. GO analysis showed that these differentially expressed sex-based genes were generally associated with the extracellular region, cellular response to estrogen stimulus and endopeptidase inhibitor activity (fold-change > 2 and FDR < 0.05, Fig. 2a). We found that the enrichment degree of the male-bias genes in gene ontology was significantly higher than that of the female-biased genes. These gene ontology terms may be associated with the observed behavioral differences between genders [19].

The KEGG analysis showed that the differentially expressed mRNAs remarkably enriched in PPAR signaling pathway, glycolysis/gluconeogenesis, starch and sucrose metabolism, tryptophan metabolism, and cysteine and methionine metabolism pathways (Fig. 2b). Previous studies found that PPAR signaling pathway plays important roles in mammalian reproductive system during the processes of the ovarian cycle, luteal formation, embryo implantation, placentation and male reproduction [20].

Identification and characterization of long non-coding RNA

To investigate the biological function of the sex-related lncRNAs in zebrafish brain, the lncRNAs were



identified in this study. Using coding prediction and ORF (open reading frame) identification software, we found that 3709 potential lncRNAs were expressed in the six zebrafish samples. Then the cuffdiff software was employed to analyze the potential lncRNAs, with a result of 28 female-based and 22 male-based differentially expressed lncRNAs ($|\log_2(\text{fold-change})| > 1.5$, $P\text{-value} < 0.05$) (Additional file 3). The scatter plot showed a high degree of positive correlation ($P = 0.913$) between lncRNA expression in the female and male zebrafish samples (Fig. 3a). The volcano plots showed the differential expression levels of the lncRNA in the female and male samples (Fig. 3b).

lncRNAs-mRNAs co-expression

To examine the collaboration between the lncRNAs and mRNAs in the gender-based zebrafish brain tissues, a co-expression analysis of the differentially expressed lncRNAs and the corresponding differentially expressed mRNAs was performed based on their FPKM values. The Pearson correlation coefficient between lncRNAs and mRNAs was calculated with the thresholds of absolute correlation coefficient $|R| > 0.8$ and $P < 0.05$, and significantly correlated lncRNA-mRNA couplings were obtained in 12 lncRNAs and 19 mRNAs (Fig. 4a). lncRNAXLOC_038516 (degree = 12) and XLOC_023087 (degree = 9) indicated higher degrees, and *fntb* (farnesyltransferase, CAAX box, beta) and *hbaa2* (hemoglobin alpha adult 2) showed similar results. The lncRNA-mRNA coupling suggested that the regulation of *cldn7a* by multiple lncRNAs was likely to occur in the brain. Further GO analysis showed that

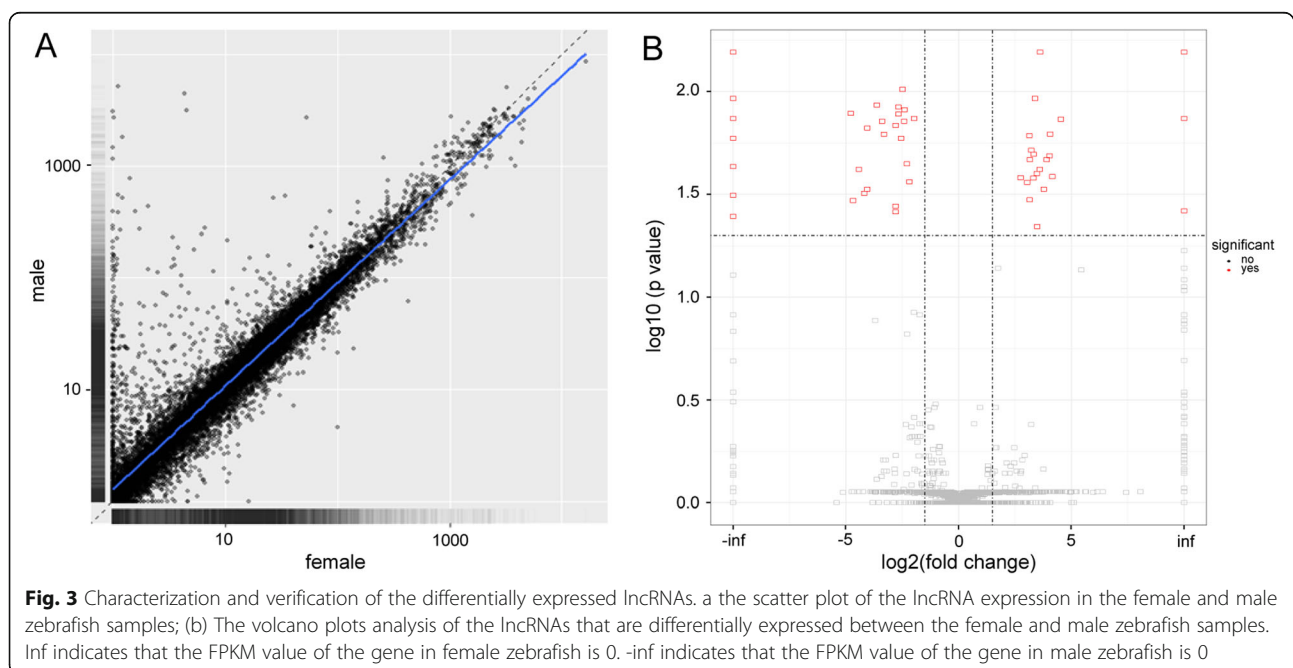
mRNAs in the co-expression network were enriched on “response to hormone” (Fig. 4b).

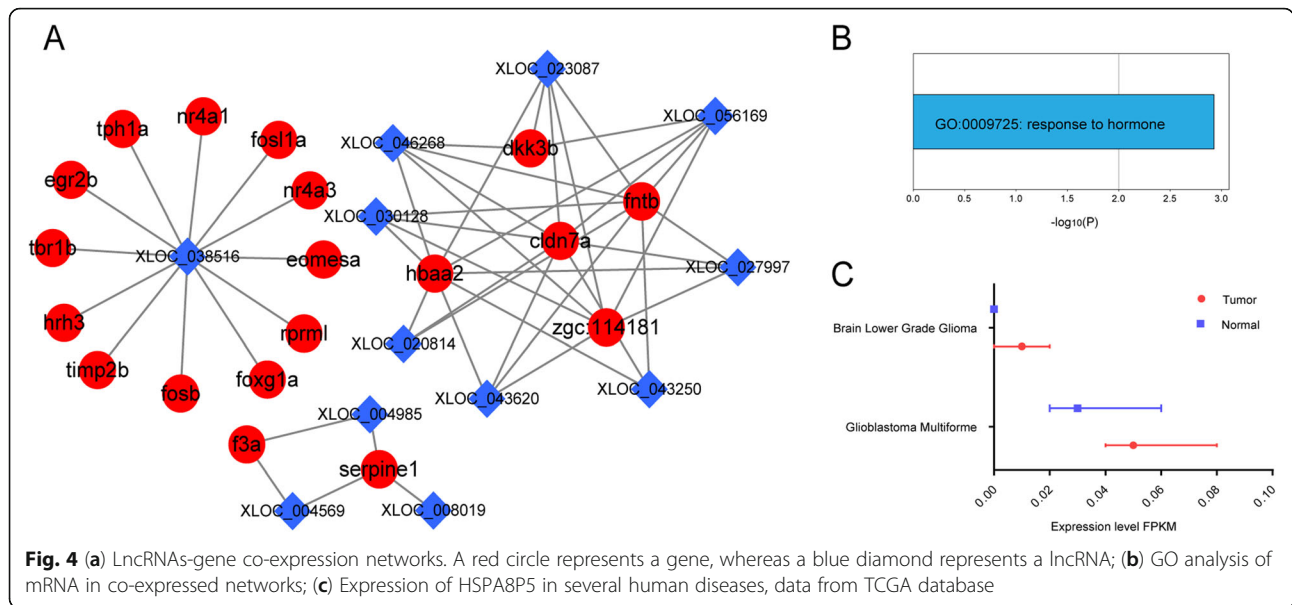
The information of the 12 lncRNAs in the lncRNA-gene networks was shown in Table 1. For each of these lncRNAs, we found that the mRNA co-expressed with them was not within 300 kb from the same chromosome, indicating that they did not have *cis* regulatory functions and were not directly involved in the regulation of gene transcription or post-transcriptional levels. The ZFLNC database was used to perform a conservative analysis of the lncRNAs in the co-expression network. We found that XLOC_038516 and human pseudogene HSPA8P5 were considered as orthologs. Further analysis showed that HSPA8P5 was differentially expressed in multiple human brain neurological diseases (Fig. 4c).

Discussion

The zebrafish, as a model animal, has a nearly 70% similarities in genes between its genome and human genome [21], so exploring its brain-related dimorphism not only expands our understanding of the interaction between its reproductive processes and environmental stressors, but also has a positive effect on the analysis of human brain diseases [22].

In this study, we investigated the differential gene expression between 8-month-old male and female zebrafish brain tissues by using RNA-Seq sequencing technique, with a result of differentially expressed seven female-based genes and 101 male-based genes (fold-change > 2 , and $P\text{-value} < 0.05$). These differentially expressed genes account for less than 1% of all genes identified from the brains of the male





and female zebrafish. By RNA-seq data analysis, we found that only 61 genes showed significant gender effects in all four strains [7]. Sreenivasan et al. generated a gonad-derived zebrafish cDNA microarray and only observed 24 candidate genes showing a sexual dimorphic pattern [23]. These studies indicated that most gene expression levels are not significantly different between male and female zebrafish. Therefore, we speculated that they should also be similar in both male and female brains. Santos et al. found that 7478 expressed genes does not show a clear separation of the individual transcriptomes according to gender, suggesting that gender is not the main determinant of the variation between individual brain gene expression profiles [5]. Later, same conclusion was reached by Wong RY et al., suggesting that behavioral and physiological gender

differences may be more easily facilitated by other factors such as the hormonal, ecological, or social environment [7].

Available evidence showed that many genes in the list of 108 sex-based genes we obtained in this study involves in neural circuit or brain development, e.g., *egr2b* (early growth response 2b, also known as KROX20) [24], *mych* (myelocytomatosis oncogene homolog) [25], and *hrh3* (histamine receptor H3, [26]). In addition, we found that some genes play a similar role in the brain of zebrafish and humans. Zhang T et al. reported that *apoa2* (apolipoprotein A-II) is abundant in the zebrafish brain and may perform a function during embryonic development [27], whereas in human, mutations in the gene may lead to adult glioma [28] and meningioma [29]. We have found that some genes have not been

Table 1 Information of 12 lncRNAs in lncRNA-gene networks

lncRNA	Chromosome	Regulation ^a	log ₂ (FC)	pvalue
XLOC_004569	chr11:14098307–14,098,777	Up	2.76891	0.0421
XLOC_004985	chr11:38600939–38,601,149	Up	2.05469	0.0476
XLOC_008019	chr12:48960587–48,964,182	Up	5.43661	0.0012
XLOC_020814	chr19:17458973–17,459,464	Down	-3.31062	0.04015
XLOC_023087	chr2:31617390–31,623,697	Down	-2.7921	0.0258
XLOC_027997	chr21:32311853–32,312,662	Down	-1.72466	0.00235
XLOC_030128	chr22:24451902–24,454,348	Down	-1.98589	0.00215
XLOC_038516	chr3:25989342–25,989,841	Up	2.46855	0.02155
XLOC_043250	chr5:16082283–16,083,083	Down	-1.77434	0.0274
XLOC_043620	chr5:31375274–31,382,222	Down	-1.81491	0.01585
XLOC_046268	chr6:9650891–9,651,332	Down	-2.83877	0.0378
XLOC_056169	chrUn_KN150103v1:1249–2053	Down	-2.08877	0.01365

^aUp- or down-regulated expression compared to expression levels in the female zebrafish FC fold change

thoroughly studied in zebrafish, but have been identified to be important in human brain neural networks and diseases. For example, *kng1* (kininogen 1) demonstrated tumor suppression and anti-angiogenic properties in glioblastoma [30]. *Aldh1l1* (aldehyde dehydrogenase 1 family member L1) was also identified as a new astrocyte-specific marker in the brain [31]. In addition, *dkk3b* (dickkopf WNT signaling pathway inhibitor 3b), belonging to the family of secreted wnt-inhibitors with conserved cysteine-rich domains, displays a specific role during neuronal differentiation [32] and encodes a vital intracellular regulator of cell proliferation [33]. However, the function of most sex-based genes we found in this study is still unknown.

When performing GO enrichment analyses, an interesting result was obtained in the differentially expressed genes. The result showed that five genes (*NUPR1*, *ESR1*, *SERPINC1*, *ZGC: 66313*, and *FKBP11*) enriched in cellular response to estrogen stimulus, but these five genes were expressed in the brain of the male zebrafish. There is evidence that male behavior requires estrogen signaling, and adult gonads of any sex can support male behavior [34]. Furthermore, previous work had also reported that the sexually dimorphic gene expression in the zebrafish does not correspond to specific pathways, from which we can ascertain that commonalities in their regulatory mechanisms have the sex determining pathways in mammals [8].

For each lncRNA locus, the 300 kb upstream and downstream protein-coding genes were identified as *cis*-acting target genes; however these genes were not differentially expressed in our study. In the lncRNAs and mRNAs co-expression networks, some of the differentially expressed genes were previously reported to be involved in neural circuit such as *egr2b* and *hrh3*, indicating that lncRNA *XLOC_038516* and *XLOC_038516* may also have the same function. Further analysis revealed that *HSPA8P5*, one of the orthologs of *XLOC_038516*, was differentially expressed in multiple human brain neurological diseases. However, these conclusions have some limitations because all results are obtained in silicon. Further *in vivo* studies will help to fully understand the role of these lncRNAs in the brain of zebrafish.

Conclusion

In this study, the mRNAs and lncRNAs with the sex-based differential expression were screened by transcriptome sequencing (RNA-seq) in the zebrafish brains. Based on the various biological analyses, we found that 12 new lncRNAs have significant gender specificity in the brain of zebrafish by analyzing the biological functions of the co-expressing mRNA. Our finding may

provide a clue to further study of the functional difference between male and female zebrafish brain.

Methods

Acquisition of the Zebrafish Specimens

The wild-type AB zebrafish were purchased from the Wuhan Institute of Hydrobiology, Chinese Academy of Sciences, China, and the adult zebrafish of the same size (3 male and 3 female) were selected as experimental specimens.

Animal euthanasia of zebrafish

We followed the NIH guidelines for zebrafish euthanasia (<https://oacu.oir.nih.gov/sites/default/files/uploads/ara-guidelines/zebrafish.pdf>). Immobilization by submersion in ice water (5 parts ice/1 part water, 0–4 °C) for 25 min following cessation of opercular movement. The fish were confirmed death by hypoxia after cessation of all movement. In this process, MS-222 solution (tricaine methane sulfonate, 168 mg/l) was used as anesthetic, which was buffered with sodium bicarbonate to pH = 7 before immersing the fish. The fish were left in the solution for 15 min following cessation of opercular movement. After anesthesia with MS222, the fish were frozen quickly in liquid nitrogen.

Total RNA extraction and quality testing

Total RNA was extracted using Invitrogen Ambion RNA Extraction Kit according to the manufacturer's protocol (ThermoFisher Scientific, MA, USA). RNA degradation and contamination were monitored on a 1% agarose gel. The RNA integrity number (RIN) was measured using an Agilent Bioanalyzer 2100 (Agilent Technologies, CA, USA) (Agilent, CA, USA) to assess the RNA quality. Sequencing was performed if the samples RIN values were greater than eight. The total RNA concentration was determined using a Qubit 2.0 fluorometer (Life Technologies, CA, USA).

Sequencing library preparation, RNA-seq sequencing, and raw data preprocessing

Library Preparation was created using VAHTS Stranded mRNA-seq Library Prep Kit according to the manufacturer's protocol (Vazyme, Nanjing, China). After the RNA samples passed the quality test, 2 µg total RNA was enriched by magnetic beads with Oligo (dT). Subsequently, the Frag/Prime Buffer was used to break the mRNA into short fragments at 85 °C for 6 min. RNA fragments were converted to cDNA using random primers, followed by second-strand cDNA synthesis and end repair. The double-stranded cDNA was subsequently purified using AMPure XP beads (Beckman Coulter, CA, USA). The purified double-stranded cDNA was added an A and ligated to the sequencing linker,

and AMPure XP beads were used for size selection of adapter-ligated DNA. Finally, PCR amplification was performed and the PCR product was purified using AMPure XP beads to obtain the final library. After the library was constructed, preliminary quantification was performed using Qubit 2.0 (Thermo Fisher Scientific, MA, USA), and then the size of the library was detected using the DNA High Sensitivity DNA Kit (Bioanalyzer 2100, Agilent, CA, USA) to ensure the proper insert size of 350–450 bp. The concentration of the library was accurately quantified using KAPA Library Quantification Kits according to the manufacturer's protocol (KAPA Biosystems, MA, USA). Subsequently, the library was sequenced using an Illumina HiSeq X Ten sequencing platform (Illumina, CA, USA). By using Trimmomatic software [35], Low-quality reads were filtered according to the following criteria: quality scores are less than 20, and reads with average quality scores of each read less than 20. FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) [36] was employed to assess the quality of the raw reads.

Mapping, annotation, and differential expression analysis for mRNA-seq data

The zebrafish reference genome (GRCz10/danRer10, Sep.2014) and the reference Index (the GTF file) were downloaded from UCSC (<http://genome.ucsc.edu/>). Firstly, bowtie2-build was used to index the reference genome, and then TopHat (version 2.1.0) was used [37] to map the reads to the reference genome. TopHat initially removed a small portion of the reads based on the quality information of each reads and then mapped the qualified reads to the reference genome [37]. In addition, the parameter of “--library-type” was set to “fr-firststr”, and the other parameters were set to default values. Then the result file of TopHat was input into Cufflinks software for further analyses, including transcript assembly, abundance estimation, and differential expression of RNA-Seq samples [38]. The confidence intervals for estimation of fragments per kilobase of transcript per million mapped reads FPKM were calculated using a Bayesian inference method [39]. Differentially expressed genes were characterized according to the criterion of a fold change > 1.5 and q-value < 0.05.

Gene ontology analyses and KEGG analysis

DAVID online tool (<https://david.ncifcrf.gov>) was used for identifying enriched biological themes [40]. Enriched GO terms with Gene-Count > 5 and P-value < 0.05 were selected as the thresholds for the subsequent analyses. Cytoscape software (two tools: ClueGO and CluePedia) was used for the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis [41], showing only pathways with P-value < 0.05.

ClueGO network diagram was created based on Kappa statistics. Every node in the diagram represented a term that reflected the relationships between nodes, and the color of the nodes reflected the enrichment of the node classification.

Validation by RT-qPCR

For the characterization of the zebrafish brain, RT-qPCR of mRNAs was performed using SYBR Green PCR Master Mix (Fermentas, Guangzhou, China) following the manufacturer's instructions. The experiments were repeated at least in triplicate. The gene-specific primers were as follows:

zgc:114181(forward: 5'-TCACCGCCTTCCTCAGAAAT-3';

reverse: 5'-ACTGAGCCGTGACCACTTTA-3');

f13a1a.1(forward: 5'-GCGTGTTCATTCCAAAACCCT-3';

reverse: 5'-CAACTTGACAGCCAGGATT-3');

vtna(forward: 5'-GACATTCCGCCGGCTTGTATT-3'.

reverse: 5'-CAAGCGGACACTAAGGATGC-3');

rbp2a(forward: 5'-TGACTAAACAAAAGGGCGCC-3';

reverse: 5'-CGCCTCTGTGCATCTTCTTC-3').

β-actin(forward primer:5'-TCACCACCACAGCCGAAAG-3';

reverse primer:5'-GGTCAGCAATGCCAGGGTA-3').

β-actin was used as an internal control. The efficiencies of all sets of primers were between 91.6–97.3%. We used 96-hole RT-qPCR plates including three negative controls to discard false positive amplification signals. In each PCR plate, synthesized cDNAs without adding reverse transcriptase were used to confirm no genomic contamination. PCR reactions were performed at 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. The fold changes were calculated by $\Delta\Delta\text{CT}$ method [42].

lncRNA identification and characterization

Cufflinks script was used to determine whether the detected transcripts were annotated by Refseq genes of zebrafish genome (build GRCz10/danRer10, Sep., 2014). Transcripts with length < 200 nt or > 10,000 nt were discarded, and only transcripts with exon number > 1 were retained. Coding Potential Calculator software (CPC, <http://cpc.cbi.pku.edu.cn/>) [43] and Coding Potential Assessment Tool (CPAT, <http://lilab.research.bcm.edu/cpat/>) [44] were used for the coding potential prediction analysis. Only transcripts that were considered “non-coding” by both of these tools were considered potential lncRNAs and software cuffdiff [38] was employed for subsequent analysis.

Protein-protein interaction network and lncRNA-gene co-expression network construction

The STRING online software [45] was employed to construct the interaction network of the proteins encoded

by the differentially expressed genes. A combined score > 0.7 was adopted as the cut-off criterion. According to the FPKM values of the different expression of the lncRNAs and the mRNAs in the six samples, the Pearson correlation coefficient between the lncRNAs and the mRNAs were calculated according to their FPKM values. The threshold for positive correlation was set to $PCC > 0.8$ and $P < 0.05$. Finally, the protein-protein interaction network and the regulatory network of lncRNAs-mRNA were established via Cytoscape software (<http://www.cytoscape.org/>) [41]. We used the ZFLNC database to conduct a conservative analysis of lncRNA [46].

Statistical analysis

We used Pearson's correlation coefficient to assess relationships between gene expression and stationary behavior and determined significance with two-tailed P -values, and a P -value of less than 0.05 was considered to be statistically significant.

In RT-qPCR validation experiment, a two-tailed t test was used to assess the statistical deference between male and female zebrafish, and a P -value of less than 0.05 was considered to be statistically significant.

All the statistical analyses were performed using R (<https://www.r-project.org/>, version 3.5.1).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6197-9>.

Additional file 1. The raw data files.

Additional file 2. Differentially expressed mRNAs in the brains of male and female zebrafish.

Additional file 3. Differentially expressed lncRNAs in the brains of male and female zebrafish.

Abbreviations

FPKM: Reads Per Kilobase of exon model per Million mapped reads; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

Acknowledgements

Not Applicable.

Authors' contributions

WLY performed the bioinformatics analyses. SWJ constructed the library and participated in data analysis and RT-qPCR validation experiment. DS performed RT-qPCR validation experiment. ZCW prepared experimental samples and designed an experimental technical solution. CW and CXD performed quality control and preprocessed the raw sequencing data. WLY and SHP wrote the manuscript. SHP and LHJ designed the research and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Shanghai Natural Science Foundation (15ZR1420800 to Sihua Peng), and the National Natural Science Foundation of China (61775139 to Linhua Jiang), and Shanghai Ocean University Youth Fund Project (A2-0203-00-100237 to Shouwen Jiang). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The data set generated and analyzed during this study is was submitted to Sequence Read Archive (SRA), with Project ID: PRJNA521037.

Ethics approval and consent to participate

The animal experiments were approved by Animal Ethics Committee of Shanghai Ocean University. All experiments were performed in strict accordance with the requirements of the Animal Ethics Procedures and Guidelines of the People's Republic of China.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources (Shanghai Ocean University), Ministry of Education, Shanghai 201306, China. ²National Pathogen Collection Center for Aquatic Animals, Ministry of Agriculture, Shanghai 201306, China. ³International Research Center for Marine Biosciences at Shanghai Ocean University, Ministry of Science and Technology, Shanghai 201306, China. ⁴School of Optical-Electric and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China. ⁵College of Mathematics and Information Engineering, Jiaying University, Jiaying 314001, China.

Received: 6 July 2019 Accepted: 16 October 2019

Published online: 08 November 2019

References

- Rupp B, Reichert H, Wullmann MF. The zebrafish brain: a neuroanatomical comparison with the goldfish. *Anat Embryol.* 1996;194(2):187–203.
- Anderson JL, Mari AR, Braasch I, Amores A, Hohenlohe P, Batzel P, Postlethwait JH. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One.* 2012;7(7):e40701.
- Liew WC, Bartfai R, Lim Z, Sreenivasan R, Siegfried KR, Orban L. Polygenic sex determination system in zebrafish. *PLoS One.* 2012;7(4):e34397.
- Griffin RM, Dean R, Grace JL, Rydén P, Friberg U. The shared genome is a pervasive constraint on the evolution of sex-biased gene expression. *Mol Biol Evol.* 2013;30(9):2168–76.
- Santos EM, Kille P, Workman VL, Paull GC, Tyler CR. Sexually dimorphic gene expression in the brains of mature zebrafish. *Comp Biochem Physiol A Mol Integr Physiol.* 2008;149(3):314–24.
- Pradhan A, Olsson P-E. Zebrafish sexual behavior: role of sex steroid hormones and prostaglandins. *Behav Brain Funct.* 2015; 11:23.
- Wong RY, McLeod MM, Godwin J. Limited sex-biased neural gene expression patterns across strains in Zebrafish (*Danio rerio*). *BMC Genomics.* 2014;15:905.
- Arslan-Ergul A, Adams MM. Gene expression changes in aging zebrafish (*Danio rerio*) brains are sexually dimorphic. *BMC Neurosci.* 2014;15:29.
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009;458(7235):223–227.
- Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell.* 2013;154(1):26–46.
- Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet.* 2011;12(12):861–874.
- Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol.* 2010;28(5):503–510.
- Wan Y, Zhang Q, Zhang Z, Song B, Wang X, Zhang Y, Jia Q, Cheng T, Zhu X, Leung AY-H. Transcriptome analysis reveals a ribosome constituents disorder involved in the RPL5 downregulated zebrafish model of diamond-Blackfan anemia. *BMC Med Genet.* 2016;9:13.
- Sarangdhar MA, Chaubey D, Bhatt A, KM M, Kumar M, Ranjan S, Pillai B: A novel long non-coding RNA, durga modulates dendrite density and expression of kalirin in Zebrafish. *Front Mol Neurosci* 2017;10:95.

15. Ray MK, Wiskow O, King MJ, Ismail N, Ergun A, Wang Y, Plys AJ, Davis CP, Kathrein K, Sadreyev R: CAT7 and cat7l long non-coding RNAs tune polycomb repressive complex 1 function during human and zebrafish development. *J Biol Chem* 2016;291(37):19558-19572.
16. Kaushik K, Leonard VE, Shamsudheen K, Lalwani MK, Jalali S, Patowary A, Joshi A, Scaria V, Sivasubbu S. Dynamic expression of long non-coding RNAs (lncRNAs) in adult zebrafish. *PLoS One*. 2013;8(12):e83616.
17. Lee K-T, Nam J-W. Post-transcriptional and translational regulation of mRNA-like long non-coding RNAs by microRNAs in early developmental stages of zebrafish embryos. *BMB Rep*. 2017;50(4):226-231.
18. Haque S, Kaushik K, Leonard VE, Kapoor S, Sivasubbu S, Joshi A, Scaria V, Sivasubbu S. Short stories on zebrafish long noncoding RNAs. *Zebrafish*. 2014;11(6):499-508.
19. Yang CF, Shah NM. Representing sex in the brain, one module at a time. *Neuron*. 2014;82(2):261-78.
20. Zhao Y, Yang Z. PPARs signaling pathway in mammalian reproduction. *Chin J Cell Biol*. 2005;27(1):14-8.
21. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013; 496(7446):498-503.
22. Bao W, Volgin AD, Alpyshov ET, Friend AJ, Strekalova TV, de Abreu MS, Collins C, Amstislavskaya TG, Demin KA, Kalueff AV. Opioid neurobiology, neurogenetics and neuropharmacology in zebrafish. *Neuroscience*. 2019;404:218-32.
23. Sreenivasan R, Cai M, Bartfai R, Wang X, Christoffels A, Orban L. Transcriptomic analyses reveal novel genes with sexually dimorphic expression in the zebrafish gonad and brain. *PLoS One*. 2008;3(3):e1791.
24. Wang Y, Zhang Y, Li X, Sun M, Wei Z, Wang Y, Gao A, Chen D, Zhao X, Feng X. Exploring the effects of different types of surfactants on zebrafish embryos and larvae. *Sci Rep*. 2015;5:10107.
25. Hong S-K, Tsang M, Dawid IB. The mych gene is required for neural crest survival during zebrafish development. *PLoS One*. 2008;3(4):e2029.
26. Sundvik M, Kudo H, Toivonen P, Rozov S, Chen Y-C, Panula P. The histaminergic system regulates wakefulness and orexin/hypocretin neuron development via histamine receptor H1 in zebrafish. *FASEB J*. 2011;25(12): 4338-47.
27. Zhang T, Yao S, Wang P, Yin C, Xiao C, Qian M, Liu D, Zheng L, Meng W, Zhu H. ApoA-II directs morphogenetic movements of zebrafish embryo by preventing chromosome fusion during nuclear division in yolk syncytial layer. *J Biol Chem*. 2011;286(11):9514-25.
28. Rajaraman P, Brenner AV, Butler MA, Wang SS, Pfeiffer RM, Ruder AM, Linet MS, Yeager M, Wang Z, Orr N. Common variation in genes related to innate immunity and risk of adult glioma. *Cancer Epidemiol Prevention Biomarkers*. 2009;18(5):1651-8.
29. Rajaraman P, Brenner AV, Neta G, Pfeiffer R, Wang SS, Yeager M, Thomas G, Fine HA, Linet MS, Rothman N. Risk of meningioma and common variation in genes related to innate immunity. *Cancer Epidemiol Prevention Biomarkers*. 2010;19(5):1356-61.
30. Ren Y, Ji N, Kang X, Wang R, Ma W, Hu Z, Liu X, Wang Y. Aberrant ceRNA-mediated regulation of KNG1 contributes to glioblastoma-induced angiogenesis. *Oncotarget*. 2016;5.
31. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci*. 2008;28(1):264-78.
32. Untergasser G, Martowicz A, Hermann M, Töchterle S, Meyer D. Distinct expression patterns of dickkopf genes during late embryonic development of *Danio rerio*. *Gene Expr Patterns*. 2011;11(8):491-500.
33. Leonard JL, Leonard DM, Wolfe SA, Liu J, Rivera J, Yang M, Leonard RT, Johnson JP, Kumar P, Liebmann KL. The Dkk3 gene encodes a vital intracellular regulator of cell proliferation. *PLoS One*. 2017;12(7):e0181724.
34. Xu X, Coats JK, Yang CF, Wang A, Ahmed OM, Alvarado M, Izumi T, Shah NM. Modular genetic control of sexually dimorphic behaviors. *Cell*. 2012;148(3):596-607.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.
36. Andrews S: FastQC: a quality control tool for high throughput sequence data. 2010; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
37. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511-U174.
38. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc*. 2012;7(3):562-578.
39. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol*. 2011; 12(3):R22.
40. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2008;4(1): 44-57.
41. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2010; 27(3):431-2.
42. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29(9):e45.
43. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G: CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 2007, 35(suppl_2):W345-W349.
44. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41(6):e74.
45. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Von Mering C. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2012; 41(D1):D808-15.
46. Hu X, Chen W, Li J, Huang S, Xu X, Zhang X, Xiang S, Liu C. ZFLNC: a comprehensive and well-annotated database for zebrafish lncRNA. *Database*. 2018;2018.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

