Cloning and Expression Analysis of GATA1 gene in Carassius auratus red var.

Yusu Tian  
Hunan University of Science and Technology  
https://orcid.org/0000-0002-4678-1290

Yuandong Sun (✉ syd@hnust.edu.cn )  
Hunan University of Science and Technology  
https://orcid.org/0000-0001-7437-9058

Mi Ou  
Chinese Academy of Fishery Sciences Pearl River Fisheries Research Institute

Xiaojuan Cui  
Hunan University of Science and Technology

Dingguang Zhou  
Hunan University of Science and Technology

Wen’an Che  
Hunan University of Science and Technology

Research article

Keywords: Carassius auratus red var., GATA1, cloning, methylation, expression characteristics

Posted Date: June 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-29269/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published on March 18th, 2021. See the published version at https://doi.org/10.1186/s12863-021-00966-3.
Abstract

Background

GATA1 is a key transcription factor in the GATA family, and promotes the differentiation and maturation of red blood cell, which is essential for normal hematopoiesis.

Results

Our results showed that the cDNA sequence of GATA1 was 2730 bp long encoding 443 amino acids. qRT-PCR analysis demonstrated that GATA1 had the highest expression in testis(T), followed by pituitary(P) and spleen(S). The expression of GATA1 gene in C auratus red var. embryo from the neuroblast stage (N) to the embryo hatching(H); and the gene expression levels of NP-treated and control embryos were significantly different. Methylation results in NP-treated and control embryos indicated that NP affected the methylation level of GATA1. NP increases the methylation level of GATA1 gene in embryos.

Conclusions

Our study provides important information for further studying the function of GATA1 gene in fish development and the molecular mechanism of NP leading to abnormal development of fish embryos.

Background

Nonylphenol (NP) is an environmental hormone that mimics estrogen and binds to its receptors in the cell, interferes with endocrine metabolism and has toxic effects on animals [1]. There are many studies on the toxicity of NP to reproduction. NP can give rise to male reproductive dysfunction, affects the development of testis, and leads to the decline of male fertility and sperm counts [2]. Comparing with the ovariectomized adult rats, the lower dose of 4-NP can induce the uterine nutrition response on the prepuberty rats [3]. Research conducted by Tanaka et al. showed that Rivulus marmoratus had abnormal gonadal development and testis insufficiency when exposed to NP [4]. The percentage of motile spermatozoa after sperm exposure to NP in Oryzias latipes was significantly lower than that in the control group [5]. But 4-NP only affected the sperm production of Oncorhynchus mykiss, while the sperm density, sperm motility and sperm fertility were not affected [6]. NP not only affects adult fish, but also greatly affects the embryonic development of fish. When Puntius conchonius embryo was exposed to NP, the embryo showed abnormal development, such as egg coagulation, spinal deformity, and delayed development [7]. NP has significantly developmental toxicity to goldfish embryos, and goldfish embryos are more sensitive to low concentrations of NP [8]. 4-NP also affects the development of embryos and larvae of Oncorhynchus mykiss at the end of the yolk sac, reducing the survival rate of embryos and larvae [6]. In NP-exposure zebrafish embryos, the distribution of PGCs along the anterior–posterior axis in
24-h-old embryos changed, and these changes may influence the juvenile and adult gonadal structures [9].

There have been some studies on the effects of NP on gene expression in vivo. Xia et al. reported that the expression of cy5 and cy3 in the rat was down-regulated under the influence of NP [10]. When Chironomus riparius larvae were exposed to NP, the expression level of CrEcR was significantly up-regulated, suggesting that nonylphenol could modulate the ecdysone nuclear receptor and may have significant implications in different developmental stages of C. riparius [11]. P353-NP produced the embryonic dysplasia in zebrafish (Danio rerio), the expression of ntl and spt was unchanged in p353-NP stressed embryos, and the expression of tbx6 significantly increased [12]. Nonylphenol exposure reduced Na+/K+-ATPase activity, plasma cortisol and triiodothyronine levels in Salmo salar gills [13]. In addition, another study suggests NP could regulate the hepatic enzyme activities of Salmo salar that mediated by Cyp3a and Cyp1a1 through Pxr and Ahr. Furthermore, it might have impacts on the metabolism of endogenous and exogenous substrates respectively, and [14]. Paolo et al. found that significantly higher mRNA level of PPARα was found in Solea solea treated with 4-NP 3 d after exposure comparing to the controls. Anh the highest dose of 4-NP also caused the up-regulation of retinoid X receptor α (RXRα) transcript levels [15].

GATA1 is a key transcription factor for erythropoiesis, and it contains three conserved functional domains: C-zinc finger, N-zinc finger, and N-terminal activation domains [16]. The two zinc finger domains are responsible for DNA binding and protein-protein interactions, which allow them to recognize typical GATA binding sites with a consensus sequence WGATAR [17]. GATA1 is indispensable in the differentiation of erythroid cells and megakaryocytes. In the development of erythroid cells, GATA1 functions early in megakaryocytes. GATA1 controls terminal maturation and its deficiency induces proliferation [18]. Galloway established a transcriptional hierarchy dependent on GATA in the process of hematopoiesis and demonstrated that GATA1 play an integral role in the fate determination of myeloid-erythroid lineage during embryogenesis [19]. Chan et al. found that the reduced hematopoiesis in Choonodraco hamatus was regulated by the miR-152-mediated down-regulation of GATA1 [16]. More importantly, studies have found that the abnormal localization of P-selectin induced by the GATA1 (low) mutation, and increased pathological interactions with leucocytes, are responsible for the increase of thrombosis in mice [20].

Carassius auratus red var. meets the basic requirements of experimental animals. It is convenient for artificial breeding, easy to observe and eliminate mutant individuals, and has high sensitivity to NP [21.22]. C auratus red var. embryos developed malformations under NP stress, including spine curvature, tail deformity, pericardial abnormalities and thrombosis [23]. In our previous study, the results of transcriptome showed that GATA1 was differentially expressed in the C auratus red var. embryos between control and NP-treated groups, which may be one of the causes for embryonic malformation of C auratus red var. [23]. In this study, the full-length cDNA sequence of GATA1 gene in C auratus red var. was obtained, and bioinformatics analysis was conducted. Realtime fluorescence quantitative PCR (qRT-PCR) was used to study the expression patterns of GATA1 in different tissues of C auratus red var. and the
expression variation of GATA1 after the treatment with different concentrations of NP. At the same time, DNA methylation levels of the C. auratus red var. embryos in the NP treatment group and the control group at each developmental stage were measured, and the effect of NP on GATA1 methylation was analyzed. This experiment investigated the expression of GATA1 gene in abnormal development of C. auratus red var. embryos under NP stress, and explored the relationship between thrombosis and GATA1 gene in malformed embryos. Our study will provide preliminary data for further research on the molecular mechanism of embryo development deformity of C. auratus red var. embryos caused by NP.

Results

Sequences analyses of GATA1 from C. auratus red var.

The cDNA sequences of GATA1 from C. auratus red var. (GenBank Accession no. MT322308) is 2730 bp in length with an ORF of 1332 bp encoding 443 amino acids (aa), 541 bp 5′-UTR and 857 bp 3′-UTR with three poly(A) signal sequences (AATAA), three RNA instability motifs (ATTTA), and a poly (A) tail. Two ZnF domains (aa 225–275, aa 279–329) were also predicted in GATA1 protein (Fig. 1).

The genomic sequences of GATA1 from C. auratus red var. was 14759 bp in length, which contained 5 exons and 4 introns following the consensus rule of GT/AG (Fig. 2). Comparison of GATA1 genomic structures from Carassius auratus (Gene ID: 113081347), Cyprinus carpio (Gene ID: 109098530), Sinocyclocheilus rhinocerous (Gene ID: 107749468), Sinocyclocheilus grahami (Gene ID: 107581944), Danio rerio (Gene ID: 564960), Mastacembelus armatus (Gene ID: 113130813) and Monopterus albus (Gene ID: 109968602) demonstrated that the genomic structure of GATA1 from C. auratus red var. is identical to the GATA1 from other teleost fish, consisting of 5 exons and 4 introns.

Multiple alignments and phylogenetic analysis

BLASTP analysis (Fig. 3) showed that GATA1 had highest similarity to CaGATA1 (99.10%) and CcGATA1 (83.97%), and lowest similarities to MaGATA1 (20.77%) and MoGATA1 (20.77%). Besides, GATA1 had middle similarities with SrGATA1 (81.07%), SgGATA1 (80.36%), DrGATA1 (59.78%), and ChGATA1 (39.6%).

Phylogenetic analysis further supported gene homology among those species (Fig. 4). Homologous amino acid sequences of GATA1 from other teleost fish and non-fish animals were gained from NBCI to construct a phylogenetic tree. According to the phylogenetic tree, these homolog proteins could be divided into five groups, consisting of mammals, birds, amphibians, fishes and invertebrates. It can be seen from the phylogenetic tree that the GATA1 protein of C. auratus red var. is closest to the GATA1 protein of Carassius auratus, and their bootstrap values reached 99%. All the fish GATA1 proteins clustered together, and diverged from their counterparts in species of other groups. GATA1 proteins in invertebrates were far separated from those in vertebrates. The phylogenetic tree reflected a genetic consistency among those species in evolution.

Tissue distribution of GATA1 in C. auratus red var.
qRT-PCR was performed to analyze the tissue distribution of GATA1 mRNA expression. As shown in Fig. 5, GATA1 expression was detected in all organs tested, and the data were calibrated against expression level in heart (H). GATA1 had the highest expression level in testis (T) (100.44 folds, \( P < 0.05 \)); intermediate levels in pituitarium (P) (7.91 folds, \( P < 0.05 \)), spleen (S) (5.70 folds, \( P < 0.05 \)), gills (G) (3.90 folds, \( P < 0.05 \)), brain (B) (3.43 folds, \( P < 0.05 \)); and low levels in muscle (M) (0.68 folds), liver (L) (0.35 folds), and ovary (O) (0.33 folds).

**Gene expression in different developmental stages after NP treatment**

To determine the effect of NP on GATA1 gene expression, the levels of GATA1 mRNA in different developmental stages were examined (Fig. 6). During the normal development of the embryo, the GATA1 gene had been detected at the N stage, and the expression level increased at the S5 stage, decreased at S14 and S21, then increases again at the P5 stage, and the expression of GATA1 reached to the highest at the P25 stage, and decreased again after the embryo hatching. Figure 6 indicated that the expression of GATA1 gene of 3 \( \mu \text{mol} / \text{L} \) NP-exposed group was the most different from the control group at S14 stage. Similarly, the expression of GATA1 gene of 5 \( \mu \text{mol} / \text{L} \) NP-exposed group was the most different from the control group at S14 stage. When embryos developed to the 21 somite stage, the expression of GATA1 gene had the biggest differences between 7 \( \mu \text{mol} / \text{L} \) NP-exposure group and the control group. Compared with other two experimental groups, GATA1 mRNA levels at neuroblast stage had the biggest difference with the control group under 3 \( \mu \text{mol} / \text{L} \) NP treatment. When embryos developed to the 5 somite stage, 7 \( \mu \text{mol} / \text{L} \) NP-exposure group had the greatest effect on the expression of GATA1 gene. The expression of GATA1 gene changed the most in the 14 somite stage embryo at 3 \( \mu \text{mol} / \text{L} \) NP treatment. Under the treatment of 7 \( \mu \text{mol} / \text{L} \) NP, the expression level of GATA1 gene in embryos at the 21 somite stage was the biggest difference from the control group. Among pharyngeal stage-primordium-5, pharyngeal stage-primordium-25 and hatching stage embryos, the expression of GATA1 gene was most affected by NP with the concentration of 3 \( \mu \text{mol} / \text{L} \). It was determined that NP affected the expression of GATA1 gene during the development of C auratus red var. embryos. NP has the greatest effect on the expression of GATA1 gene in somatic embryos.

**The methylation level of GATA1 in NP treatment and control group**

The PCR target fragment was 277 bp in size and had 10 CpG sites. In the control groups, the methylation rates of GATA1 gene in N, S5, S14, S21, P5, P25 and H stages were: 85.88%, 94.33%, 92.86%, 89.61%, 92.67%, 98.00%, and 89.33%, respectively. In the NP treatment groups, the methylation rates of GATA1 gene in N, S5, S14, S21, P5, P25 and H stages were: 93.52%, 96.67%, 98.00%, 97.06%, 98.00%, 98.67%, and 97.00%, respectively. The results showed that the methylation level of GATA1 gene in the embryos of the control group was lower than that of the NP-treated group. Figure 7 showed the results of correlation analysis between GATA1 mRNA expression and methylation of GATA1 gene in the control group and NP.
treatment group. The expression of GATA1 mRNA in the control group was significantly positively correlated with the methylation level of the GATA1 gene \((r = 0.771, P < 0.05)\), while the expression of GATA1 mRNA in the NP treatment group was not correlated with methylation level of the GATA1 gene \((r = 0.533, P > 0.05)\).

**Discussion**

In this experiment, *C. auratus* red var. was used as laboratory animals. The full-length cDNA sequence of GATA1 was obtained by homologous cloning and RACE technology from *C. auratus* red var. GATA1 is 2730 bp in length, 1332 bp in ORF, encodes 443 amino acids (aa), with 541 bp 5'-UTR and 857 bp 3'-UTR. *C. auratus* red var. is a variant of *Carassius auratus*. Alignment analysis revealed that the similarity between the *C. auratus* red var. GATA1 and *C. auratus* GATA1 protein was as high as 99.1%. Phylogenetic analysis showed that the GATA1 protein of *C. auratus* red var. was closest to that of *C. auratus*, with bootstrap values reaching 99%. The high similarity between the *C. auratus* red var. and *C. auratus* amino acid sequences is in line with our expectations. In vertebrates, members of the GATA family generally consist of five or six exons and include two conserved type IV zinc finger domains: the amino terminal zinc finger (N) and the carboxyl terminal zinc finger (C) [24]. SMART was used to analyze the conserved domains of the GATA1 amino acid sequence, and it was determined that the *C. auratus* red var. GATA1 protein has two ZnF domains (aa 225–275, aa 279–329); The *C. auratus* red var. GATA1 gene consists of five exons, which is consistent with the results of GATA1 in other species.

The tissue distribution of GATA1 mRNA expression was analyzed by qRT-PCR. GATA1 expression was detected in all tested tissues of *C. auratus* red var. GATA1 had the highest expression level in testicle (T); intermediate levels in pituitarium (P), spleen (S), gills (G), brain (B); and low levels in muscle (M), liver (L), and ovary (O). GATA1 is abundantly transcribed in mouse testis and regulates genes during the earliest stages of spermatogenesis [25]. Studies have shown that spermatogenesis is induced by the expression of GATA-1 in Sertoli cells, and transcription factor GATA-1 is a developmental stage- and spermatogenic cycle-specific regulator of gene expression in Sertoli cells [26]. The *C. auratus* red var. used in the experiment has been sexually mature, so the GATA1 gene is most expressed in the testes in all tissues and organs. The GATA transcription factor family is essential for pituitary cell differentiation and gonadotropin subunit expression [27]. GATA1 inhibits the formation of rat cortical neurons, and overexpression of GATA1 in the hippocampus can cause depressive behavior in rats [28]. Spleen is an important hematopoietic organ in the body, the gills are the respiratory organs of fish, and there are a large number of capillaries, and the GATA1 gene is a key regulator of red blood cell production [18]; Therefore, the GATA1 gene is expressed in high levels in P, S, G, and B of *C. auratus* red var. The expression of GATA1 in L and O of Nile tilapia is very low [29], this result is similar to the expression of GATA1 in *C. auratus* red var. Studying the expression pattern of GATA1 in various tissues and organs of *C. auratus* red var. is helpful to understand the function of GATA1 in *C. auratus* red var. adult.

In *C. auratus* red var. embryos, the GATA1 gene has been detected during the neural embryo stage, and has been continuously expressed during embryonic development, and the expression level is relatively
stable. It is indicated that GATA1 gene is involved in the entire embryonic development of C auratus red var. The GATA1 gene is also involved in early embryonic development in other fish. Early blood islands have emerged in the yolk sac endoderm and splanchnic mesoderm during early embryonic development. After 9 h of fertilization, the expressed GATA can be detected in zebrafish ectoderm [30]. In Branchiostoma belcheri, the GATA1 also detected expression signals in the mesendoderm of gastrula stage [31]. NP affects the expression of GATA1 during the development of C auratus red var. embryos. The expression of GATA1 in C auratus red var. embryos at somatic stage were affected the most at the administration of NP, and 3 µmol / L NP treatment had the greatest effect on the expression of GATA1 in C auratus red var. embryos. During vertebrate early embryogenesis, the ventral development is directed by the ventral-to-dorsal activity gradient of the bone morphogenetic protein (BMP) signaling [32]. Abnormal BMP signaling pathways can cause strong dorsalization phenotypes in embryos [33]. GATA1 gene is a downstream target gene of the BMP signaling pathway [34] and is shown to exert repressive effects on spine formation in cortical neurons [35]. Under NP stress, the abnormal expression of GATA1 gene in C auratus red var. embryos may be the reason for the dorsalization after NP treatment. The transcriptional activity of GATA1 is related to the transcription level of vitellogenin (Vg) [36]. Vg measurement has been used as a biomarker of exposure to endocrine-disrupting chemicals [37]. Up-regulation of GATA1 gene expression in C auratus red var. embryos under NP exposure may increase Vg gene expression. It may prove that NP is an environmental endocrine disruptor. Yokomizo et al.’s experiments in mouse embryos provided evidence of the presence of GATA-1(+) hemangioblastic cells in the extra-embryonic region, demonstrating that the GATA1 is involved in definitive hematopoiesis at embryonic stage in close association with endothelial development [38]. GATA1 or GATA2 is required to initiate blood formation in the embryo, GATA1 and GATA2 double deficient mice exhibit no visible blood cells [39]. GATA1 (low) mutations lead to increased thrombosis in mice [20]. The occurrence of thrombosis in C auratus red var. embryos under NP stress may be caused by the differentially down-regulation of GATA1 gene expression.

DNA methylation usually means heritable changes in gene expression without changes in DNA sequences. This change is crucial to embryonic development. If the methylation level is too high or too low, it will affect the normal growth and development of the embryo [40]. Reduced methylation of H3-K4 in Lsd1 mutant fruit flies results in tissue-specific defects during development [41]. Compared with normal embryos, H19 gene methylation is abnormal in abnormally developing embryos [42]. In the offspring of vitamin-deficient rats, the embryos showed a higher incidence of heart defects, possibly due to the high methylation level of the GATA4 gene [43]. In this study, we found that the methylation degree of the GATA1 gene in the control group was lower than that in the NP-treated group during the developmental stages of C auratus red var. embryos. This shows that NP stress increase the methylation level of GATA1 in C auratus red var. during embryonic development. The expression of GATA1 mRNA in the control group was significantly positively correlated with methylation level of the GATA1 gene; the expression of GATA1 mRNA in the NP treatment group was not correlated with methylation level in the GATA1 gene. In addition, high levels of GATA1 expression during the same developmental period are not necessarily low in methylation level. It is speculated that the expression of GATA1 gene may not be directly related to the degree of DNA methylation during the development of C auratus red var. embryos.
This result is similar to that of Okada et al. reported. In 3T3-L1 preadipocytes, demethylation did not increase leptin gene expression, and the diet-induced up-regulation of leptin, Mest/Peg1, and sFRP5 gene expression in white adipose tissue (WAT) during the development of obesity in mice is not mediated directly by changes in DNA methylation [44]. In addition, when studying the effect of monomeric and oligomeric flavanols (MOF) consumption on the gene expression profile of leukocytes, it was found that 8 week daily supplementation with 200 mg MOF modulates the expression of genes associated with cardiovascular disease pathways without major changes of their DNA methylation state [45].

**Conclusions**

In this study, the full-length cDNA sequence of GATA1 gene of C auratus red var. was cloned, and the expression pattern of GATA1 gene in various tissues and embryonic developmental stages of adult C auratus red var. was analyzed. The expression of GATA1 during NP-stressed embryonic development was revealed initially, and its expression was closely related to NP-stress. It provides important information for further studying the function of GATA1 gene in fish development and the molecular mechanism of NP leading to abnormal development of fish embryos.

**Methods**

**Fish and sampling**

Two-year-old healthy C auratus red var., weighting about 200 ± 10 g with an average length of 15 ± 3 cm, were obtained from the Engineering Research Center of Polyploid Fish Breeding and Reproduction of the State Education Ministry at Hunan Normal University. All experiments performed were approved by the Animal Care Committee of Hunan Normal University. Before experiments, the fish were acclimatized in an indoor freshwater tank at 25 ± 1 °C and fed twice daily with a commercial diet for one week. After no abnormal symptoms were observed, the C. auratus red var. were subjected to further study.

Three healthy fish were sacrificed as one group, and samples from the gills (G), liver (L), spleen (S), intestines (I), middle kidney (MK), muscle (M), head kidney (HK), heart (H), brain (B), pituitarium (P), and gonads (testis (T) or ovary (O)) were collected, respectively. All samples were immediately homogenized in TRIzol reagent (Invitrogen, USA) and stored at −80 °C until RNA extraction. At the same time, fin tissues were isolated and fixed in 95% ethanol. To minimize suffering, 100 mg/L MS-222 (Sigma-Aldrich, St Louis, MO, USA) was used to anaesthetize fish before dissection.

**NP treatment**

NP was used for challenge experiments. All the embryos were exposed to NP with the concentrations of 0 µmol/L (blank control, 0.01% ethanol), 3 µmol/L, 5 µmol/L and 7 µmol/L, respectively. Each group was employed for 5 parallel repetitions. Embryo incubation and NP exposure were carried out in 25 cm glass at 25 ± 1 °C.
Embryos were collected at 7 stages: neuroblast stage (N), 5 somite stage (S5), 14 somite stage (S14), 21 somite stage (S21), pharyngeal stage-primordium-5 (P5), pharyngeal stage-primordium-25 (P25) and hatching stage (H) after NP exposure. Six groups we collected at each time points and used liquid nitrogen to stop embryo development. These samples were divided into two part. One part is used for DNA extraction and the other part is used for RNA isolation.

**RNA extraction and cDNA synthesis**

The total RNAs were extracted according to the manufacturer's instruction for TRIzol reagent. Later, the RNA samples were incubated in RNase-free DNase I (Promega, USA) to eliminate any contaminating genomic DNA. Random primers and a ReverTra Ace kit (Toyobo, Japan) were used for reverse transcription to generate cDNA. Samples that need to be extracted total RNAs include: various tissues of healthy adult fish, embryos of the treatment group and the control group at different developmental stages. SMART™ RACE cDNA Amplification Kit (Takara, Japan) was used to obtain 5′-RACE Ready cDNA and 3′-RACE Ready cDNA.

**Full-length cDNA cloning and analysis**

To identify the cDNA sequence of *GATA1* from *C. auratus* red var., primers GATA1-F1/R1 (Table 1) were designed and synthesized based on the highly conserved regions of known fish *GATA1* sequences, including *Carassius auratus* GATA1 (*CaGATA1*, Accession no. XM_026253445.1) and *Sinocyclocheilus rhinocerous* GATA1 (*SrGATA1*, Accession no. XM_016537268.1). The 5′ and 3′ untranslated regions (UTRs) were obtained according to the manufacturer's instruction for SMART™ RACE cDNA Amplification Kit. The full-length cDNA sequences were amplified by PCR using GATA1-F2/R2 primers (Table 1) within the 5′and 3′UTRs, respectively.

Sequence Manipulation Suite (STS) ([http://www.bio-soft.net/sms/](http://www.bio-soft.net/sms/)) was used to analyse the sequences of *GATA1* from *C. auratus* red var.. The BLASTP program ([https://blast.ncbi.nlm.nih.gov/Blast.cgi](https://blast.ncbi.nlm.nih.gov/Blast.cgi)) was used to search for GATA1 protein sequence from other species in the NCBI ([http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)). Multiple sequence alignments were performed by the ClustalX 2.1 program ([http://www.ebi.ac.uk/tools/ clustalx2.1](http://www.ebi.ac.uk/tools/ clustalx2.1)). Simple Modular Architecture Research Tool (SMART) ([http://smart.embl-heidelberg.de/](http://smart.embl-heidelberg.de/)) was used to predict the protein domain features. A phylogenetic tree was constructed by the neighbor-joining (NJ) algorithm embedded in Mega 5.0 software ([http://www.megasoftware.net/ index.html](http://www.megasoftware.net/ index.html)) with a minimum of 1000 bootstraps.

**Genomic sequence cloning**

Genomic DNA (gDNA) was extracted from the tail fin using the Universal Genomic DNA Kit (CWBio, China) according to the manufacturer's instructions. Based on the cDNA sequences of *GATA1*, primers (Table 2) were designed to amplify the genomic sequences gradually. Five overlapping fragments were amplified from gDNA and sequenced.

The 5′unknown sequence of the *GATA1* gene was obtained from the existing gDNA sequence using the Genome Walking Kit (Takara, Japan) according to the manufacturer's instructions. The gDNA sequence
was confirmed by sequencing the PCR product amplified by primers (Table 3) within the 5′ unknown sequences.

**Quantification of gene expression**

qRT-PCR was carried out in StepOnePlus Real-Time PCR System (ABI, USA) to quantify the mRNA expression of *GATA1* in different tissues, including intestine (I), liver (L), spleen (S), gills (G), middle kidney (MK), muscle (M), head kidney (HK), heart (H), brain (B), pituitary (P), testicle (T), and ovary (O). Specific primers (Table 1) were designed for qRT-PCR. The housekeeping gene β-actin [46] (Table 4) was utilized as an internal control for cDNA normalization, and the expression level in the heart (H) was used as the baseline (1.0) for qRT-PCR.

To determine the effects of NP stress on *GATA1* mRNA expression, the expression levels of *GATA1* in different developmental stages of *C auratus* red var. embryos (neuroblast stage (N), 5 somite stage (S5), 14 somite stage (S14), 21 somite stage (S21), pharyngeal stage-primordium-5 (P5), pharyngeal stage-primordium-25 (P25) and hatching stage (H)) treated with different concentrations of NP (0 µmol/L, 3 µmol/L, 5 µmol/L and 7 µmol/L) were analyzed. The housekeeping gene β-actin was used as the reference gene, and the *GATA1* expression level in neuroblast stage under 0 µmol/L NP stress was used as the baseline for qRT-PCR (1.0).

Three replicates were performed per sample. Expression levels of corresponding genes were calculated using the $2^{-\Delta\Delta CT}$ method [47]. The *GATA1* expression levels were measured by one-way analysis of variance, followed by Dunnett's tests for multiple comparisons using SPSS Statistics 20 software. $P < 0.05$ was considered statistically significant.

**Methylation of the GATA1 from C auratus red var.**

The genomic DNAs in different developmental stages from the 5 µmol/L NP stress group and control group were extracted, respectively. The DNA was subjected to sulfite modification using the EZ DNA Methylation-Gold™ Kit (Zymo Research, China) according to the manufacturer's instructions. The software Methyl Primer Express v1.0 was used to design specific primers GATA-F4 / R4 (Table 4) in the 5'UTR region of the *GATA1* gene. The PCR products were purified by a Gel Extraction Kit (Omega, USA), and the purification products were ligated into pMD19-T vectors (Takara, Japan). The ligation products were then transformed into competent Escherichia coli DH5α cells (TransGen, China) and cultured at 37 °C. Positive colonies were selected and sequenced by a commercial company (TIANYI HUIYUAN, China). The sequencing results were sorted and methylation status was analyzed. The degree of methylation was expressed as the percentage of the methylation number of the measured CpG sites to the total number of the methylation sites measured. Correlation analysis was performed on the expression of *GATA1* mRNA and the degree of methylation in the 5'UTR region of *GATA1* gene using SPSS Statistics 20 software. The correlation between the two variables was showed by the correlation coefficient (r).
Abbreviations

NP: Nonylphenol; qRT-PCR: Realtime fluorescence quantitative PCR; G: gills; L: liver; S: spleen; T: testis; P: pituitary; I: intestines; MK: middle kidney; M: muscle; HK: head kidney; H: heart; B: brain; T: testis; O: ovary; N: neuroblast stage; S5: 5 somite stage; S14: 14 somite stage; S21: 21 somite stage; P5: pharyngeal stage-primordium-5; P25: pharyngeal stage-primordium-25; H: hatching stage; STS: Sequence Manipulation Suite; SMART: Simple Modular Architecture Research Tool; NJ: neighbor-joining; gDNA: Genomic DNA

Declarations

Ethics approval and consent to participate

This study was approved by the Animal Ethical Review Committee (AERC) of Hunan Normal University and followed the guidelines statement of the Administration of Affairs Concerning Animal Experimentation of China. This manuscript does not involve the use of any human data or tissue. The animals used in the study came from Hunan Normal University, and we have obtained written consent from Hunan Normal University to use these animals in our research.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Scientific Research Project of National Natural Science Foundation of China (31873038), Hunan Education Department (17A072) and State Key Laboratory of Developmental Biology of Freshwater Fish (2018KF008). Funding bodies played no role in the design of the study or analysis or interpretation of data or in writing the manuscript.

Authors' contributions

YDS, DGZ and MO: initial conceptual and experimental design of the study. YST, XJC and WAC: performed the experiment, interpretation of data, key discussions on principle findings. YST and MO: wrote and edited the manuscript. All authors read and approved the final version of the manuscript.

Availability of data and materials

Data and materials are available from the authors on reasonable request. The GATA1 cDNA sequence is available in the GenBank (Accession number MT322308)
Acknowledgements

We thank the laboratory members for their technical assistance and constructive comments. Work in our laboratories is supported by the Hunan University of Science and Technology, State Key Laboratory of Developmental Biology of Freshwater Fish and Pearl River Fisheries Research Institute.

Supporting Information

Checklist S1.

Completed “The ARRIVE Guidelines Checklist” for reporting animal data in this manuscript.

References

1. Yu J, Wu J, Zhang B, Xu J. Toxic effects of nonylphenol on the organisms and its mechanism. Journal of Environmental Hygiene 2013(3):268–272.

2. Lukac N, Lukacova J, Pinto B, Knazicka Z, Tvrdá E, Massanyi P. The effect of nonylphenol on the motility and viability of bovine spermatozoa in vitro. Journal of Environmental Science. 2013;48(8):973–9.

3. Laws SC, Carey SA, Ferrell JM, Bodman GJ, Cooper RL. Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. Toxicol Sci. 2000;54(1):154–67.

4. Tanaka JN, Grizzle JM. Effects of nonylphenol on the gonadal differentiation of the hermaphroditic fish, Rivulus marmoratus. Aquat Toxicol. 2002;57(3):117–25.

5. Kawana R, Strüssmann CA, Hashimoto S. Effect of p-Nonylphenol on sperm motility in Japanese medaka (Oryzias latipes). Fish Physiol Biochem. 2003;28(1):213–4.

6. Lahnsteiner F, Berger B, Grubinger F, Weismann T. The effect of 4-nonylphenol on semen quality, viability of gametes, fertilization success, and embryo and larvae survival in rainbow trout (Oncorhynchus mykiss). Aquat Toxicol. 2005;71(4):297–306.

7. Xiao Q, Xu YY. On the effects of nonylphenol on the embryonic development of Puntius conchonius. Journal of Safety Environment. 2010;10(6):9–12.

8. Zhang QY, Sun YD, Wang ZJ, Hu XJ, Kui X. Toxic Effects of Nonylphenol on the Embryonic Development of Goldfish (Carassius auratus). Progress in Modern Biomedicine 2016(16):3040–3043.

9. Willey JB, Krone PH. Effects of endosulfan and nonylphenol on the primordial germ cell population in pre-larval zebrafish embryos. Aquat Toxicol. 2001;54(1–2):113–23.

10. Lü B, Zhan P. Effects of nonylphenol on brain gene expression profiles in F1 generation rats. Toxicol Environ Chem. 2009;91(3):559–65.

11. Nair PMG, Choi J. Modulation in the mRNA expression of ecdysone receptor gene in aquatic midge, Chironomus riparius upon exposure to nonylphenol and silver nanoparticles. Environ Toxicol
12. Klempt M, Vobach M, Wiegand H, Preuss TG, Schäffer AKU. 353-nonylphenol induces expression of the t-box6 gene in zebrafish embryos-linking transcriptional information with deformities. Journal of Fisheries Sciencescom. 2013;7(1):30–42.

13. Robertson LS, McCormick SD. The effect of nonylphenol on gene expression in Atlantic salmon smolts. Aquat Toxicol. 2012;122–123:36–43.

14. Meucci V, Arukwe A. The xenoestrogen 4-nonylphenol modulates hepatic gene expression of pregnane X receptor, aryl hydrocarbon receptor, $\text{CYP3A}$ and $\text{CYP1A1}$ in juvenile Atlantic salmon ($\text{Salmo salar}$). Comp Biochem Physiol C: Toxicol Pharmacol. 2006;142(1–2):142–50.

15. Cocci P, Mosconi G, Palermo FA. Effects of 4-nonylphenol on hepatic gene expression of peroxisome proliferator-activated receptors and cytochrome P450 isoforms ($\text{CYP1A1}$ and $\text{CYP3A4}$) in juvenile sole (Solea solea). Chemosphere. 2013;93(6):1176–81.

16. Chan JL, Hu XX, Wang CC, Xu QH. miRNA-152 targets GATA1 to regulate erythropoiesis in Chionodraco hamatus. Biochem Biophys Res Commun. 2018;501(3):711–7.

17. Bábara FZ, Pavón L, Calés C. CDC6 expression is regulated by lineage-specific transcription factor GATA1. Cell Cycle. 2012;11(16):3055–66.

18. Ling T, Crispino JD, Zingariello M, Martelli F, Migliaccio AR. GATA1 insufficiencies in primary myelofibrosis and other hematopoietic disorders: consequences for therapy. Expert Review of Hematology. 2018;11(3):169–84.

19. Galloway JL, Wingert RA, Thisse C, Thisse B, Zon LI. Loss of Gata1 but Not Gata2 Converts Erythropoiesis to Myelopoiesis in Zebrafish Embryos. Dev Cell. 2005;8(1):109–16.

20. Zetterberg E, Verrucci M, Martelli F, Zingariello M, Sancillo L, D'Amore E, Rana RA, Migliaccio AR. Abnormal P-selectin localization during megakaryocyte development determines thrombosis in the gata1 low model of myelofibrosis. Platelets. 2014;25(7):539–47.

21. Wu DS. Current Situation and Prospect of the Standardization Research and Application of Laboratory Red Crucian Carp. Lab Anim Sci. 2016;33(3):56–60.

22. Lü XH, Gu Y, Song Y. Toxicity and tissue accumulation of nonylphenol in Carassius auratus red variety. Grass Carp and Sliver Carp. Journal of Hygiene Research. 2012;41(5):785–9.

23. Tian YS, Sun YD, Ou M, Liu YF, Cui XJ, Zhou DG, Che WA. Preliminary studies on the mechanism of nonylphenol-induced malformation of Carassius auratus red var. Journal of Fisheries of China. 2020;44(5):658–70.

24. Lowry JA, Atchley WR. Molecular Evolution of the GATA Family of Transcription Factors: Conservation Within the DNA-Binding Domain. J Mol Evol. 2000;50(2):103–15.

25. Ito E, Toki T, Ishihara H, Ohtani H, Gu L, Yokoyama M, Engel JD, Yamamoto M. Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. Nature. 1993;362(6419):466–8.

26. Yomogida K, Ohtani H, Harigae H, Ito E, Nishimune Y, Engel JD, Yamamoto M. Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells.
Development. 1994;120(7):1759–66.

27. Thomas RL, Crawford NM, Grafer CM, Zheng W, Halvorson LM. **GATA augments GNRH-mediated increases in Adcyap1 gene expression in pituitary gonadotrope cells.** J Mol Endocrinol. 2013;51(3):313–24.

28. Miyeon C, Wang SE, Ko SY, Kang HJ, Chae SY, Lee SH, Kim YS, Duman RS, Hyeon S. **Overexpression of Human GATA-1 and GATA-2 Interferes with Spine Formation and Produces Depressive Behavior in Rats.** PloS one. 2014;9(10):e109253.

29. Ye K. Preliminary studies on expression patterns and functional analysis of GATA factors in Nile tilapia. Chongqing: Southwest University; 2012.

30. Kwan W, North TE. Netting Novel Regulators of Hematopoiesis and Hematologic Malignancies in Zebrafish. Curr Top Dev Biol. 2017;124:125–60.

31. Zhang YJ, Mao BY. **Developmental Expression of an Amphioxus (Branchiostoma belcheri) Gene Encoding a GATA Transcription Factor.** Zoological Research. 2009;30(2):137–43.

32. Martyn U, Merker SS. The ventralized ogon mutant phenotype is caused by a mutation in the zebrafish homologue of Sizzled, a secreted Frizzled-related protein. Dev Biol. 2003;260(1):58–67.

33. Bauer H, Lele Z, Rauch GJ, Geisler R, Hammerschmidt M. The type I serine/threonine kinase receptor Alk8/Lost-a-fin is required for Bmp2b/7 signal transduction during dorsoventral patterning of the zebrafish embryo. Development. 2001;128(6):849–58.

34. Kim JH, Park JB, Lee JY, Kim J. **PV.1 Suppresses the Expression of FoxD5b during Neural Induction in Xenopus Embryos.** Mol Cells. 2014;37(3):220–5.

35. Shibata K, Ishimura A, Maéno M. **GATA-1 Inhibits the Formation of Notochord and Neural Tissue in Xenopus Embryo.** Biochem Biophys Res Commun. 1998;252(1):241–8.

36. Sun ZX, Kang K, Cai YJ, Zhang JQ, Zhai YF, Zeng RS, Zhang WQ. **Transcriptional regulation of the vitellogenin gene through a fecundity-related single nucleotide polymorphism within a GATA-1 binding motif in the brown planthopper, Nilaparvata lugens.** Insect Mol Biol. 2018;27(3):365–72.

37. Lee KW, Hwang DS, Rhee JS, Ki JS, Park HG, Ryu JC, Raisuddin S, Lee JS. **Molecular cloning, phylogenetic analysis and developmental expression of a vitellogenin (Vg) gene from the intertidal copepod Tigriopus japonicus.** Comp Biochem Physiol B: Biochem Mol Biol. 2008;150(4):395–402.

38. Tomomasa Y, Satoru T, Naomi M, Takashi K, Masatsugu E, Asami W, Ritsuko S, Osamu O, Motomih O, Hitoshi O. **Characterization of GATA-1+ hemangioblastic cells in the mouse embryo.** EMBO J. 2007;26(1):184–96.

39. Fujiwara Y, Chang AN, Williams AM, Orkin SH. **Functional overlap of GATA-1 and GATA-2 in primitive hematopoietic development.** Blood. 2003;103(2):583–5.

40. Salvaing J, Peynot N, Bedhane MN, Veniel S, Pellier E, Boulesteix C, Beaujean N, Daniel N, Duranthon V. Assessment of ‘one-step’ versus ‘sequential’ embryo culture conditions through embryonic genome methylation and hydroxymethylation changes. Hum Reprod. 2016;31(11):2471–83.
41. Stefano LD, Ji JY, Moon NS, Herr A, Dyson N. *Mutation of Drosophila Lsd1 Disrupts H3-K4 Methylation, Resulting in Tissue-Specific Defects during Development*. Curr Biol. 2007;17(9):808–12.

42. Samira IR, Mohamed AK, Rita K, Thierry B, Guérin JF, Annick L. *Analysis of H19 methylation in control and abnormal human embryos, sperm and oocytes*. Eur J Hum Genet. 2011;19(11):1138–43.

43. Feng Y, Zhao LZ, Hong L, Shan C, Shi W, Cai W. *Alteration in methylation pattern of GATA-4 promoter region in vitamin A-deficient offspring’s heart*. The Journal of Nutritional Biochemistry. 2013;24(7):1373–80.

44. Okada Y, Sakaue H, Nagare T, Kasuga M. Diet-induced up-regulation of gene expression in adipocytes without changes in DNA methylation. Kobe J Med Sci. 2009;54(5):241–9.

45. Milenkovic D, Berghe WV, Boby C, Leroux C, Declerck K, Szic KS, Heyninck K, Laukens K, Bizet M, Defrance M, et al. Dietary Flavanols Modulate the Transcription of Genes Associated with Cardiovascular Pathology without Changes in Their DNA Methylation State. PloS one. 2014;9(4):e95527.

46. Wang JS, Wei YH, Li XM, Cao H, Xu MQ, Dai JY. *The identification of heat shock protein genes in goldfish (Carassius auratus) and their expression in a complex environment in Gaobeidian Lake, Beijing, China*. Comp Biochem Physiol C: Toxicol Pharmacol. 2007;145(3):350–62.

47. Livak KJ, Schmittgen TD. *Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT Method*. Methods. 2001;25(4):402–8.

**Tables**

### Table 1 Primers for full-length cDNA cloning and qRT-PCR

| Primer name | Sequence (5′→3′) | Application |
|-------------|------------------|-------------|
| GATA1-F1    | GCTCCACAAAAAGAAAGTCAT | partial sequence obtaining |
| GATA1-R1    | ACGAGGCTGTTAGAGAAGT | |
| GATA1-F2    | CCTCAATACCCACATATGCC | ORF qualifying |
| GATA1-R2    | GTGGATTTGAGATCCGACAT | |
| GATA1-R-out | GCTCTGGCATAGTGGGTGATTGAGGTTA | 5′-Race PCR amplification |
| GATA1-R-in  | ATAATCGAAACACATCACCTCACACCCA | |
| GATA1-F-out | GGCGTCTACAGCCACCATCCCATCCATCCATCAG | 3′-Race PCR amplification |
| GATA1-F-in  | GGTAGCTTTAGTCACAGATGTCCGGAAT | |
| GATA1-qF    | CCTCCCCCTCTCTTATCCAG | qRT-PCR amplification |
| GATA1-qR    | GGTAGTGTCCCCGTCCATC | |

### Table 2 Primers for genomic DNA sequences.

| Primer name     | Sequence (5′→3′) | Product Length (bp) |
|-----------------|------------------|---------------------|
| GATA1-gDNA-F1   | CAATCACCCACATATGCCAGAGC | 914 bp |
| GATA1-gDNA-R1   | GCTGAAATAAGAGGCCAGGCGT | |
| GATA1-gDNA-F2   | TGGTCCACACTGAGGGAGGTTC | 1238 bp |
| GATA1-gDNA-R2   | GGAACACTGTACAGGGAGGCG | |
| GATA1-gDNA-F3   | CTGAGCCACACTGAGGGCTATG | 1172 bp |
| GATA1-gDNA-R3   | AGGGGTCTGTGTTCTTACATAG | |
| GATA1-gDNA-F4   | GATGGACCGCCAGCATCCACTCT | 664 bp |
| GATA1-gDNA-R4   | TAGAGTCGGCAAGACATTACA | |
| GATA1-gDNA-F5   | GGAACCTCATGTCGACACTGTC | 528 bp |
| GATA1-gDNA-R5   | CTGTTCTTGCTGACATCTTAC | |
### Table 3 Primers for 5′unknown sequences

| Primer name | Sequence (5′→3′) | Application                  |
|-------------|------------------|------------------------------|
| GATA1-SP1   | CAGAGCAAGGCTGTTAGGAAGTCATTT | 5′- Walking PCR amplification |
| GATA1-SP2   | GTCCCTGTGTTTGAGAGGTTGTTGCC   |                              |
| GATA1-SP3   | GCTTCCACCTTTGATAGGCTGTA     |                              |
| GATA1-F3    | ATGGGCTGTAGTGCTCATTTCATCGCT | verification                 |
| GATA1-R3    | CAAGAGATTACAAACTATGACTGCG   |                              |

### Table 4 Primers for others.

| Primer name | Sequence (5′→3′) | Application                |
|-------------|------------------|----------------------------|
| β-actin-F   | GCCTCCCTGTCTATCTCC | qRT-PCR                    |
| β-actin-R   | TTGAGAGGTTTGGTGTC | qRT-PCR                    |
| GATA1-F4    | TTTATTTCTGTTGGAGGATC | methylation sequence obtaining |
| GATA1-R4    | CGCTATCTAAAATCTTCCACG |                              |
Figure 1

Nucleotide and putative amino acid sequences of GATA1 and its product. The sequences numbers of nucleotide (lower row) and putative amino acid (upper row) are shown on the left. The translation initiation codon (ATG), stop codons (TGA) are in bold and yellow background. The motif associated mRNA instability (ATTTA) is doubly underscored, and poly-adenylation signal sequence (AATAA) is emphasized by wavy line. The ZnF domains are marked with gray background.
Figure 2

Genomic structure of GATA1 genes. The lengths of the elements are shown in base pairs (bp), and the numbers above and below each schematics represent the lengths of exons and introns of corresponding gene, respectively.
Figure 3

Multiple alignments of GATA1 with GATA1 proteins from various species. The amino acid sequences of GATA1 from typical organisms were aligned using the ClustalW 2.1 program. The black shade represents 100% identity, dark gray represented 80% identity. CaGATA1 stands for GATA1 protein in Carassius auratus (Protein ID. XM_026253445.1), CcGATA1 stands for GATA1 protein in Cyprinus carpio (Protein ID. XM_019103428.1), SrGATA1 stands for GATA1 protein in Sinocyclocheilus rhinoceros (Protein ID. XM_024145060.1).
XM_016537268.1), SgGATA1 stands for GATA1 protein in Sinocyclocheilus grahami (Protein ID. XM_016271639.1), DrGATA1 stands for GATA1 protein in Danio rerio (Protein ID. XP_021334219.1), ChGATA1 stands for GATA1 protein in Chionodraco hamatus (Protein ID. KP221299.1), MaGATA1 stands for GATA1 protein in Mastacembelus armatus (Protein ID. XP_026189425.1), MoGATA1 stands for GATA1 protein in Monopterus albus (Protein ID. XM_020614979.1).

Figure 4

Phylogenetic tree of the GATA1 proteins in different species. A neighbor-joining phylogenetic tree was constructed using MEGA 5.0 software. The bootstrap values of the branches were obtained by testing the tree 1000 times and values were over 50% percent marked. The GenBank accession numbers of GATA1 proteins are given after the species names in the tree.
qRT-PCR analysis of the distribution of GATA1 in different tissues. Expression of β-actin was used as an internal control for qRT-PCR. The relative expression was the ratio of gene expression in different tissues relative to that in the heart (H). Detailed values are listed at the bottom of the figure. The assay was performed three times, and data were analyzed by the unpaired t-test. *P < 0.05, compared with control.
Figure 6

Expression levels of GATA1 in the treatment and control groups at various developmental stages CK: 0 μmol / LNP-treated embryos; NP3: 3 μmol / LNP-treated embryos; NP5: 5 μmol / LNP-treated embryos; NP7: 7 μmol / LNP-treated embryos

Figure 7

Correlation analysis between GATA1 mRNA expression and GATA1 gene methylation a: Scatter plot of correlation between GATA1 mRNA expression and methylation degree of 5'UTR region of GATA1 gene in
control group embryos; b: Scatter plot of correlation between GATA1 mRNA expression and methylation degree of GATA1 gene 5'UTR region of NP treatment group.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement11.docx](#)