https://doi.org/10.1186/s12863-021-00966-3

(2021) 22:12

Tian et al. BMC Genomic Data

# **Open Access**

# Cloning and expression analysis of GATA1 gene in Carassius auratus red var



Yusu Tian<sup>1</sup>, Yuandong Sun<sup>1\*</sup>, Mi Ou<sup>2</sup>, Xiaojuan Cui<sup>1</sup>, Dinggang Zhou<sup>1</sup> and Wen'an Che<sup>1</sup>

# Abstract

**Background:** *GATA*1 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 *GATA*1 was 2730 bp long encoding 443 amino acids. qRT-PCR analysis demonstrated that *GATA*1 had the highest expression in testis (T), followed by pituitary (P) and spleen (S). *GATA*1 gene expression in *C. auratus* red var. embryo from the neuroblast stage (N) to the embryo hatching (H) changes continuously; and the gene expression levels of nonylphenol (NP)-treated and those of control embryos were significantly different. Moreover, Methylation levels of *GATA*1 gene in NP-treated embryos were higher than those in control embryos, indicating that NP affected *GATA*1 methylation.

**Conclusions:** Our study provides cues for further studying the roles of *GATA*1 gene in fish development, and suggested a potential molecular mechanism by which NP leads to abnormal development of fish embryos.

Keywords: Carassius auratus red var.,, GATA1, Cloning, Methylation, Expression characteristics

# Background

Nonylphenol (NP), 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]. Studies demonstrated that NP was detrimental to reproduction. For example, NP causes male reproductive dysfunction, damages the development of testis, and leads to the decline of male fertility and sperm counts [2]. A low 4-NP (a typical isomer of para-NP) dosage induced uterine nutrition response in prepuberty rats, but not in ovariectomized adult rats [3]. Tanaka et al. showed that *Rivulus marmoratus* had abnormal gonadal development and testis insufficiency when exposed to NP [4]. In *Oryzias latipes*, the percentage of motile spermatozoa after sperm exposure to NP decreased dramatically [5]. However, although 4-NP did

<sup>1</sup>Hunan Key Laboratory of Economic Crops Genetic Improvement and Integrated Utilization, School of Life Sciences, Hunan University of Science and Technology, Xiangtan 411201, Hunan, People's Republic of China Full list of author information is available at the end of the article



affect sperm production in Oncorhynchus mykiss, it showed no effect on sperm density, motility and fertility [6]. NP not only affected adult fish, but also interfered with fish embryonic development. When Puntius conchonius embryos were exposed to NP, they showed developmental abnormalties such as egg coagulation, spinal deformity, and delayed development [7]. NP had drastic toxicity to development of goldfish embryos, which showed higher sensitivity to low concentrations of NP than adult fish [8]. 4-NP also affected the development of embryos and larvae of Oncorhynchus mykiss at the end of the yolk sac stage, reducing their survival rate [6]. In NP-exposured zebrafish embryos, the distribution of PGCs along the anterior-posterior axis in 24-h-old embryos changed, which 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 NP exposure [10]. When *Chironomus riparius* larvae were treated with NP, the expression level of CrEcR

© The Author(s). 2021 **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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

<sup>\*</sup> Correspondence: syd@hnust.edu.cn

was significantly up-regulated, through which nonylphenol might have significant implications in various developmental stages of C. riparius [11]. P353-NP (a typical isomer of para-NP) caused embryonic dysplasia in zebrafish (Danio rerio), with the expression of ntl and spt unchanged but that of *tbx6* significantly increased [12]. Nonylphenol exposure reduced Na<sup>+</sup>/K<sup>+</sup>-ATPase activity, plasma cortisol and triiodothyronine levels in Salmo salar gills [13]. In addition, another study in Salmo salar suggests NP could regulate the hepatic enzyme activities that were mediated by Cyp3a and Cyp1a1 through Pxr and Ahr. Furthermore, NP might have impacts on metabolism of both endogenous and exogenous substrates [14]. Paolo et al. found that significantly higher PPAR $\alpha$ mRNA levels in Solea solea were associated with 4-NP treatment for 3 days while the highest dose of 4-NP in their study also led to up-regulation of retinoid X receptor  $\alpha$  (RXR $\alpha$ ) transcription [15].

GATA1 is a key transcription factor for erythropoiesis, and it contains three conserved functional domains: Czinc finger, N-zinc finger, and N-terminal activation [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 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 played an integral role in the fate determination of myeloerythroid lineage during embryogenesis [19]. Chan et al. found that reduced hematopoiesis in Choonodraco hamatus was associated with miR-152-mediated downregulation of GATA1 [16]. More importantly, studies have found abnormal localization of P-selectin induced by GATA1 (low) mutations, and increased pathological interactions with leucocytes as well, which were responsible for increased thrombosis in mice [20].

*Carassius auratus* red var. fulfills our basic requirements of experimental animals. It is convenient for artificial breeding, easy to discover and eliminate mutant individuals, and highly sensitive to NP [21, 22]. *C. auratus* red var. embryos developed malformations under NP stress, such as spine curvature, tail deformity, pericardial abnormalities and thrombosis [23]. Our previous transcriptome study revealed that *GATA1* expression in *C. auratus* red var. embryos was affected by NP-treatment, which may be one of the causes for embryonic malformation *C. auratus* [23]. In this study, we cloned and sequenced the full-length *GATA1* cDNA in *C. auratus* red var., and conducted bioinformatics analysis. In addition,

we used realtime fluorescence quantitative PCR (qRT-PCR) to explore expression patterns of GATA1 in different tissues of C. auratus red var. and transcriptional changes of GATA1 after exposure to different concentrations of NP. Moreover, we measured differences in DNA methylation levels of the C. auratus red var. embryos between the NP treatment groups and the control ones at various developmental stages, and measured NP treatment effects on GATA1 methylation. This study 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 provides cues for further research on the molecular mechanism of embryo development deformity in C. auratus red var. caused by NP.

# Results

#### Analysis of GATA1 sequences from C. auratus red var

The cDNA sequence of *GATA*1 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, 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 predicted in GATA1 protein (Fig. 1).

The genomic sequence of *GATA*1 from *C. auratus* red var. is 14,759 bp in length, which contains 5 exons and 4 introns following the consensus rule of GT/AG (Fig. 2). Comparison of *GATA*1 genomic structures among *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: 109968602) demonstrated that the genomic structure of *GATA*1 from *C. auratus* red var. is identical to the *GATA*1 from other teleost fish, all consisting of 5 exons and 4 introns.

## Multiple alignments and phylogenetic analysis

BLASTP analysis (Fig. 3) showed that GATA1 in *C. auratus* red var.shared highest similarity to *Ca*GATA1 (99.10%) and *Cc*GATA1 (83.97%), medium similarities to *Sr*GATA1 (81.07%), *Sg*GATA1 (80.36%), *Dr*GATA1 (59.78%), and *Ch*GATA1 (39.6%), and lowest similarities to *Ma*GATA1 (20.77%) and *Mo*GATA1 (20.77%).

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 collected from NBCI to construct a phylogenetic tree, which indicated that these homolog proteins could be divided into five groups, representing

	ctccttt	tgtgct	ccacaa	aaagaa	agtca	ataca	ggtt	ggaa	aggca	attgg	ggtga	agtga	tgtg	tttcg	atta	tttt	cata	atttt	caaq	tt	
91	ttttaa	attgag	tttta	tatgco	cttat	acttt	aagt	ataat	tgta	aaata	gttt	aaagc	ctaat	atttg	agaa	attt	<u>a</u> ctc	aaat	gtt	ctt	
181	gatgaat	taggtt	caact	tccaca	ataga	aacct	ttct	gtgtt	ttag	cagat	ctaa	aaatt	agat	tttgg	ftctt	tat	actg	tatg	ttg	gca	
271	ttgcata	itttgta	attct	cattte	gcatg	tggca	atatt	tccag	actt	ttttg	gtgca	tttga	actt	ttgag	Jacad	cacc	attt	gctg	aag	ctg	
361	acctgtt	tgtggc	tccgc	ccact	gtccc	cacat	gtct	aacct	caat	cacco	cacta	tgcca	agagc	acaga	aato	cttc	acat	cttc	ata	aag	
451	ttctgad	catcag	cggtg	ctcat	tttcg	tcaca	agaca	cctgt	cctga	aagac	cgaa	ctgat	aagc	tcagt	acaa	aag	gagt	tcct	gac	aga	
1	ΜE	T S	T E	Q A	R	w v	S	s s	м	V S	S	E V	М	P N	Y	Ρ	P D	S	S	Y	
541	c <mark>atg</mark> gao	ACCTCC	ACTGA	ACAGG	CTCGC	TGGGI	TTTCC	ICCTC	CATG	GTATC	CATCA	GAGGT	GATG	CCCAA	TTAC	CCA	CCTG	ACTC	CAG	CTA	
31	M V	н т	ΕE	G S	V	Y P	Y	ГD	A	E H	S	S L	Ρ	S L	F	s	S P	V	N	G	
631	TATGGT	CACACT	GAGGA	GGGTT	CAGTG	TACCO	CTAC	ACCGA	TGCC	GAACA	CAGC	AGCCI	rgccc	TCTTI	TATTO	CAGC.	AGCC	CTGT	CAA	rgg	
61	R P	S G	A F	Q T	S	s v	Y	P V	Y	s s	Ρ	F L	G	N L	S	W	L E	S	S	N	
721	TCGTCCA	TCTGGA	GCTTT	CCAAA	CCAGC	TCAGI	TATAT	CCCGI	TTAC	ICTTC	CTCCA	TTCCT	GGGA	AACCI	GTCC	TGG	CTGG	AAAG	TTC	GAA	
91	G P	S L	T N	L F	Ρ	s s	Р	s s	W	H S	S	V F	S	S S	F	Н	G S	т	Ρ	Н	
811	CGGCCCZ	TCTCTA	ACCAA	CCTCT	TCCCA	TCCTO	CCCCA	ICATC	CTGG	CACAG	GCAGT	GTGTT	TTCC	TCATC	CTTC	CAC	GGCT	CCAC	GCC'	TCA	
121	S S	A R	P P	R S	А	L P	S	LI	Q	DQ	K	DT	L	I V	Q	Е	S M	K	G	Q	
901	CTCCTC	GCCAGA	.ccgcc	TCGTT	CTGCC	CTTCC	CCTCC	CTTAI	CCAG	GACCA	GAAG	GATAC	CCTT	ATTGI	GCAG	GAG	AGCA	TGAA	AGG	ACA	
151	R L	S P	P G	G G	Е	A F	G	G V	F	S P	S	L S	S	V Y	А	Q	т н	S	S	K	
991	GAGGCTO	AGTCCT	CCTGG	AGGAG	GGGAG	GCGTI	TGGT	GGTGI	GTTT	rcccc	CTCA	TTGAG	GCAGT	GTGTA	TGCC	GCAG	ACAC	ACTC	CTC	AAA	
181	Т Н	SQ	S L	S H	Y	S P	Y	G S	F	Г Е	Ν	Y N	S	S L	L	Y	T P	S	S	F	
1081	AACACAC	TCGCAG	TCACT	GAGCC	ACTAC	AGTCO	CCTAT	GGGAG	CTTC	ACGGA	GAAC	TACAA	ACAGT	TCACI	TCTC	CTAC.	ACAC	CCTC	GTC	CTT	
211	P P	K L	C S	K M	K	F S	Ρ	L V	A	г е	P	R E	С	V N	С	G.	A T	А	S	P	
1171	CCCACCO	AAATTA	TGCAG	CAAGA	IGAAA	TTCTO	CTCCT	ITAGI	GGCG	ACAGA	GCCG	CGTGA	GTGT	GTGAA	CTGI	GGG	GCCA	CTGC	ATC	CCC	
241	L W	RR	D G	T G	Н	Y L	С	N A	C	G L	Y	H K	М	N R	Q	N	R P	L	I	R	
1261	TCTGTGC	CGGCGT	GATGG	AACGG	GACAC	TACCI	CTGC	AATGC	CTGC	GGTCI	GTAC	CACAA	GATG	AATAG	GACAG	GAAC	AGAC	CCCT	CAT	CCG	
271	ΡK	KR	L V	I S	K	R T	G	ΓQ	C '	V N	С	Q T	S	т т	Т	L	WR	R	N	А	
1051					0	CONAC			a mom			01010								CGC	
1351	ACCCAA	AAGAGA	CTGGT	CATCA	GTAAG	CGAAC	CAGGA	ACTCA	ATGT	3'I'GAA	ACTGT	CAGAC	CAGC	ACCAC	CACG	GCTG	TGGA	GACG	AAA		
301	ACCCAAA S G	AAGAGA	V C	N A	C	G L	Y Y	ACTCA F K	L I	H N	V	N R	P	ACCAC	M	K	TGGA K E	GACG	AAA I	Q	
301 1441	ACCCAAA S G CAGTGGA	AAGAGA E P AGAGCCC	CTGGT V C GTGTG	N A	C C CTTGC	G L GGACI	Y Y CTAT	ACTCA F K ITCAA	L :	H N	V V ATGTG	N R AACAG	P BGCCG	ACCAC L A CTCGC	M TATO	SCTG' K GAAG	IGGA K E AAGG	GACG G AAGG	AAA I CAT	Q	
301 1441 331	ACCCAAA S G CAGTGGA T R	AAGAGA E P AGAGCCC N R	V C GTGTG K M	N A TAACGO S S	C C CTTGC K	G L GGACI N R	Y Y ICTAT K	ACTCA F K ITCAA G K	L i GCTT( K	H N CACAA F S	V V ATGTG P	N R AACAG T E	P GGCCG E	ACCAC L A CTCGC N L	M TATO Y	GCTG K GAAG F	IGGA K E AAGG S K	GACG G AAGG N	AAA I CAT P	Q CCA G	
1331 301 1441 331 1531	ACCCAAA S G CAGTGGA T R GACGCGG	LAAGAGA E P LGAGCCC N R CAACCGT	V C GTGTG K M AAGAT	N A TAACGO S S GTCCAO	C C CTTGC K GCAAG	G L GGACI N R AACAC	Y TCTAT K GGAAA	ACTCA F K ITCAA G K GGGAA	L I GCTTO K I GAAG	GTGAA H N CACAA F S ITTAG	V V ATGTG P GCCCC	N R AACAG T E ACAGA	CAGC P GCCCG E AGGAG	ACCAC L A CTCGC N L AATCI	M TATO Y CATAT	GCTG K GAAG F TTTT	IGGA K E AAGG S K TCAA	GACG G AAGG N AGAA	AAA I CAT P TCC	Q CCA G TGG	
1331 301 1441 331 1531 361	ACCCAAA S G CAGTGGA T R GACGCGG	AAGAGA E P AGAGCCC N R CAACCGT Q H	V C GTGTG K M AAGAT F D	N A TAACGO S S GTCCAO L Y	C C CTTGC K GCAAG S	G L GGACI N R AACAC Q S	Y TCTAT K GGAAA P	ACTCA F K ITCAA G K GGGAA G A	L i GCTT K i GAAG	H N CACAA F S FTTAG G V	V ATGTG P GCCCC Y	N R AACAG T E ACAGA S H	P GCCCG E AGGAG S	ACCAC L A CTCGC N L AATCI S H	M TATO Y S S	GCTG K GAAG F TTTT L	IGGA K E AAGG S K TCAA P P	GACG G AAGG N AGAA T	AAA I CAT P TCC A	Q CCA G TGG A	
1331 301 1441 331 1531 361 1621	ACCCAAA S G CAGTGGA T R GACGCGGG S D ATCTGAT	AAAGAGA E P GAGCCCC N R CAACCGT Q H CAGCAT	V C GTGTG K M AAGAT F D TTCGA	N A TAACGO S S GTCCAO L Y CTTGTZ	C CTTGC K GCAAG S ATTCT	G L GGACI N R AACAC Q S CAGAC	Y Y TCTAT K GGAAA P STCCA	ACTCA F K ITCAA G K GGGAA G A GGAGC	L : GCTT( K : GAAG' L ( TCTG(	H N CACAA F S ITTAG G V GGCGI	V ATGTG P GCCCCC Y CCTAC	N R AACAG T E ACAGA S H AGCCA	CAGC P GGCCG E AGGAG S ACTCA	ACCAC L A CTCGC N L AATCI S H TCCCA	M TATO Y S ATTCA	GCTG K GAAG F TTTT L ACTG	IGGA K E AAGG S K TCAA P P CCGC	GACG G AAGG N AGAA T CCAC	AAA I CAT P TCC A CGC	Q CCA G TGG A TGC	
1331 301 1441 331 1531 361 1621 391	ACCCAAA S G CAGTGGZ T R GACGCGG S D ATCTGAT F H	AAAGAGA E P AGAGCCCC N R CAACCGT Q H CAGCAT A Q	V C GTGTGTG K M AAGAT F D TTCGA P A	N A TAACGO S S GTCCAO L Y CTTGTZ L P	C CTTGC K SCAAG S ATTCT M	G L GGACI N R AACAC Q S CAGAC P P	Y TCTAT K GGAAA P STCCA T	ACTCA F K ITTCAA G K GGGAA G A GGAGC S C	L CONTRACTOR	H N CACAP F S ITTTAG G V GGCGI L T	V ATGTG P SCCCCC Y CCTAC	N R AACAG T E ACAGA S H AGCCA H G	CAGC P GGCCG E AGGAG S ACTCA V	ACCAC L A CTCGC N L AATCI S H TCCCA T G	M TATO Y SATAT S ATTCA	K GAAG F TTTT L ACTG M	IGGA K E AAGG S K ICAA P P CCGC L Y	GACG G AAGG N AGAA T CCAC K	AAAA I CAT P TCC A CGC T	Q CCA G TGG A TGC L	
1331 301 1441 331 1531 361 1621 391 1711	ACCCAAA S G CAGTGGZ T R GACGCGC S D ATCTGAT F H CTTCCAC	AAAGAGA E P GAGGCCC N R CAACCGT Q H CCAGCAT A Q CGCCCAG	V C GTGTGTG K M AAGAT F D TTCGA P A CCTGC	N A TAACGO S S GTCCAO L Y CTTGTZ L P TCTACO	C CTTGC K GCAAG S ATTCT M CCATG	G L GGACI N R AACAC Q S CAGAC P P CCACC	Y TCTAT K GGAAA P STCCA T CCACC	ACTCA F K ITTCAA G K GGGAAA G A GGGAGC S C AGCTC	L CCATC	H N CACAA F S F TTTAG G V GGCGI L T CTTAC	V ATGTG P GCCCCC Y CCTAC Q CCCAG	N R AACAG T E ACAGA S H AGCCA H G CATGG	CAGC P GGCCG E AGGAG S ACTCA V STGTG	ACCAC L A CTCGC N L AATCI S H TCCCA T G ACGGG	M TATO Y SATAT S ATTCA I SAATC	SCTG K SAAG F TTTT L ACTG M CATG	IGGA K E AAGG S K ICAA P P CCGC L Y ITGT.	GACG G AAGG N AGAA T CCAC K K ACAA	AAAA I CAT P TCC A CGC T T AAAC	Q CCA G TGG A TGC L ACT	
1331 301 1441 331 1531 361 1621 391 1711 421	ACCCAAM S G CAGTGGZ T R GACGCGC S D ATCTGAD F H CTTCCAC	AAGAGA E P GAGCCCC N R CAACCGT Q H CAGCAT A Q CGCCCAG A Q	V C GTGTGG K M AAGAT F D TTCGA P A CCTGC P C	N A TAACGO S S GTCCAO L Y CTTGTZ L P TCTACO V N	C CTTGC K GCAAG S ATTCT M CCATG I	G L GGACI N R AACAC Q S CAGAC P P CCACC L S	Y TCTAT K GGAAA P GTCCA T CCACC K	ACTCA F K ITTCAA G K GGGAA G A GGAGC S C AGCTG N M	L COTTO	H N CACAA F S FTTAG G V GGCGI L T CTTAC E N	V ATGTG P GCCCCC Y CCTAC Q CCCAG L	N R AACAG T E ACAGA S H AGCCA H G CATGG T D	CAGC P GGCCG E AGGAG S ACTCA V GTGTG V	ACCAC L A CTCGC N L AATCI S H TCCCA T G ACGGG D A	M CTATO Y CATAT S ATTCA I SAATC	K GAAG. F TTTT L L ACTG M M CATG	rgga k E AAGG S K TCAA P P P CCGC CCGC L Y ITGT.	GACG G AAGG N AGAA T CCAC K ACAA	AAAA I CAT P TCC A CGC T T AAAC	Q CCA G CGG A CGC L L CCT	
1331 301 1441 331 1531 361 1621 391 1711 421 1801	ACCCAAA S G CAGTGGA T R GACGCGG S D ATCTGAT F H CTTCCAG P A CCCAGCI	AAGAGA E P GAGGCCC N R CAACCGT Q H CAGCAT A Q CGCCCAG A Q CGCCCAG	CTGGT V C GTGTGG K M AAGAT F D TTCGA P A CCTGC P C CCATG	N A TAACGO S S GTCCAO L Y CTTGTI L P TCTACO V N TGTTAI	C CTTGC K SCAAG S ATTCT M CCATG I ACATA	G L GGACI N R AACAC Q S CAGAC P P CCACC L S CTCAC	Y TCTAT K GGAAA P STCCA T CCACC K STAAA	ACTCA F K ITTCAA G K GGGAA GGAGC S C AGCTC N M AACAT	L CCATC	H N CACAP F S TTTAG G V GGCGI L T CTTAC E N GAAAP	V ATGTG P SCCCCC Y CCTAC Q CCCAG L ACTTA	N R AACAGA T E ACAGA S H AGCCA H G CATGG T D ACAGA	CAGC P GCCCG E AGGAG S ACTCA V STGTG V V ATGTG	ACCAC L A CTCGC N L AATCI S H TCCCA T G ACGGG D A GATGC	M TTATO Y SATAT S TTTCA I SAATC *	K GAAG F TTTTT L ACTG M M CATG	rgga K E AAGG S K TCAA P P P CCCGC L Y TTGT. aaac	GACGJ G N AAAGGG T T CCCACC K K AACAAJ atgt	AAAA I CAT P TCC A CGC T AAAC	Q CCA G G G G G G G G G G C C C C C C C C	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891	ACCCAAA S G CAGTGGA T R GACGCGC S D ATCTGAT F H CTTCCAC P A CCCCAGCT aaaata	AAGAGA E P GAGCCCC N R CAACCGT Q H CAGCAT A Q GCCCAG A Q GCCCAG	V C GTGTG K M AAGAT F D TTCGA P A CCTGC P C CCATG tgctt	N A TAACGO S S GTCCAO L Y CTTGTA L P TCTACO V N TGTTAA tagtca	C C C C C T T C C C A C C A C A C A C A	G L GGACT N R AACAC Q S CAGAC P P P CCCACC L S CTCAC CTCAC CTCAC	Y Y CCTAT K GGAAA P FTCCA T T CCCACC K K STAAA	ACTCA F K TTCAA G K GGGGAA G A GGAGCA S C AGCTC N M AAACAT	L :: GCTTC K :: GAAG L : CCATC GCTTC atcc: atcc:	H N F S G V GGCGI L T CTTAC E N GGAAAA A CTTAC	V ATGTG P GCCCCC Q CCCAG L L CCCAG L CCCAG	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaac	P GGCCG E S GGCGGG S S ACTCA V V V TGTGG GTGTG G ACTCA	ACCAC L A CTCGC N L AATCI S H TCCCA T G ACGGG D A GATGC aatca	M TATC Y S TATAT S ATTCZ I SAATC * *	K GAAG. F TTTT L ACTG M Aaaa aggt	IGGA K E AAAGG S K TCAA P P P P CCGC CCGC L Y ITTGT. aaaac ttat	GACGJ G N AAGGG T T CCCACC K ACAAJ ACAAJ	AAAA I CAT P TCC A CGC T T AAAC tgc aag	Q CCA G G G G G G G G G G C C C C C C C C	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981	ACCCAAA SG CAGTGGA TR GACGCGC SD ATCTGAA FH CTTCCAAC PA CCCAGCI aaaaata gacagaa	AAGAGA E P GAGCCC N R CACCGT Q H CCAGCAT A Q CCCCCAG A Q CCCCCAG Ltttgga	V C GTGTG K M AAGAT F D TTCGA P A CCTGC P C CCATG tgctt cactt	N A TAACGO S S GTCCAO L Y CTTGTZ L P TCTACO V N TGTTAZ tagto	C CTTGC K GCAAG S ATTCT M CCCATG I ACATA ACATA	G L GGACT N R AACAC Q S CAGAC CAGAC L S CTCAC atgtc ctagz	Y TCTAT K GGAAA P FTCCA T T CCACC K STAAA Cggaa atttg	ACTCA F K TTCAA G K GGGAA GGAGCA S C AGCTC N M AACAT tctca aaaaq	L : GCTTC GGAAG L : TCTGG H :: CCATC GCTTC atcc: cttg;	H N CACAAA F S G V GGCGTI L T CTTAC E N GGAAAA actgt aaaga	V ATGTG P GCCCCC V V CCTAC C CCCAG L CCCAG L CCCAG L CCCAG L	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaac aacta	P GGCCGG E GGCCG S GGCCG S GGCCG S GGCCG V V U GTGTG G C CCCAC C CCACC C CCACC C CCACC C CCACC C CCACC C C C C C C C C C C C C C C C C C C C	ACCACC L A CTCGCC N L AAATCT S H TCCCA T G ACGGG D A GATGC aatca gaagc	M TTATO Y S TTTCA I SAATO * *	K GAAG. F TTTTT L ACTGO M CATGO Aaaaa gcgt .ggt.	IGGA K E AAAGG S K TCAA P P P P P P P CCCGC CCGC L Y ITGT. aaaac ttat	G G N AAAGGAA T T CCCACC K AACAA A ACAA C Atta	AAAA I CAT P TCC A CGC T T AAAC tgc aagt	Q CCA G G G G G G G C A C G C C C C C C A C G C C A C G C C A C G C C A C G C G	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161	ACCCAAA SG CAGTGGZ TR GACGCGC SD ATCTGAT FH CCTCCAC PA CCCAGCI aaaaata gacagaa	AAGAGA E P GAGCCCC N R CACCGT Q H CCAGCAT A Q CCCCAG A Q CCCCAG A Q CCCCAG A Q CCCCAG CCCCAG CCCCAG	V C GTGTG K M AAAGAT F D TTCGA P A CCCTGC P C CCATG tgctt tataa	N A TAACGO S S GTCCAC L Y CTTGTA L P TCTACC V N TGTTAA tagtca tacaac agctac	C CTTGC K GCAAG S ATTCT M I ACATA ACATA Acata	G L GGGACT N R AAACAC Q S CAGAC CCACC L S CTCAC atgtc ctaga acatt	Y TCTAT K GGAAA P T CCCACC K STAAA CGGGAA STAAA	ACTCA F K TTCAA G K GGGGAA G A GGGGGAA G A GGGGCC S C S C S C S C S C S C S C S C S C S	L : GGTT( K : GAAG' L : TCTGG H : CCAT( atcc: cttg; gtttd	H N CACAA F S TTTTAG G V GGCGI L T CTTAC E N A CTTAC E N A CTTAC E N A CTTAC E N A CTTAC E N A CTTAC E N E N CTTAC E N E S C C C C E S C E S E S E S E S E S E S E S E S E S E S	V V TGTGTG P GCCCCC V V CCTAC Q CCCAG L CCCAG L CCTAC L CCTAC C CCAG L CCTAC	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaaac aaaaac aaaaa	P P GGCCCG E AGGAG S S ACTCA V V TGTGTG Catca ACTCA V V TGTGTG Catca	ACCAC L A CTCGC N L AATCI S H TCCCA T G ACGGC D A GATGC aatca gaagc tttaa	M TATO Y SATAT S TTTCA I SAATC * * CATG2 catat	K GAAG F TTTT L ACTG M Xaaaa Yegt aatt.	TGGA K E AAAGG S K TCAA P P P CCCGC L Y TTGT. aaaac ttat	G G N AAAGGA T T CCCACC K AACAAA Attta cgtta cgtta	AAAA I CAT P TCC A CGC T T AAC tgc aagu agt	Q CCA G G G G G G G G G G C C A C C C C A C G C G	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161	ACCCAAA SG CAGTGGA TR GACGCGC SD ATCTGAT FH CTTCCAC PA CCCCAGCT aaaata gacagaa acatgtc gcaagga	AAGAGA E P GAGCCCC N R CAACCGT Q H CAGCAT A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A C GCCCAG A C GCCCAG A C C GCCCAG C A C C GCCCAG C A C C C C C C C C C C C C C C C C C C	V C GTGTG K M AAAGAT F D TTCGA P A CCTGC P C CCATG tgctt cactt tataa agatt	N A TAACGG S S S GTCCAG L Y CTTGTA L P TCTACG V N TGTTACG tacaca agctac agctac	C CTTGC K GCAAG S ATTCT M ACATG ACATG ACATG ACATG ACATG ACATG ACATG ACATG ACATG ACATG ACATG ACATG C C C C C C C C C C C C C C C C C C C	G L GGGACI N R AAACAC Q S CAGAC CAGAC P P P CCACC L S CTCAC atgtc ctaga acatt gactc	Y TCTAT K K GGAAA P F GTCCA T CCACC K K GTAAA K CGTAAA A tttgg atttgg	ACTCA F K TTTCAA G K GGGGAA G A GGGAGC S C AGCTC S C N M AAACAT tctca aaaaag caaaag gacat	L : GGCTTC K : GGAAG L : TCTGG H : CCATC CCCATC CCATCC	H N CACAA F S TTTAG G V GGGCGI L T CTTAC E N GAAAA A actgt aaaga ggggt aaaga	V ATGTG P GCCCCC Q CCCAG L CCCCAG L CCCAG CCCAG L CCCCAG L CCCAG L CCCAG L CCCAG L CCCAG L CCCAG L CCC	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaac aacta gaaag aaaat	CAGC P GGCCG E AGGAG S S ACTCA V V STGTG G STGTG V V U TGTG G CACCA S ACTCA	ACCAC L A CTCGC N L AATCT S H TCCCA T G ACGGG D A GATGC aatca gaagc tttaa ttctt	M TATATO Y SATATO S ATTOP I SAATO X X TGA TG2 CATG2 Catat aagaaa	K GAAG. F TTTT L ACTGO M CATGO Aaaaa gcgt 	IGGA K E AAAGG S K TCAA P P P CCCGC L Y ITGT. aaaac ttat aaaag aata tttt	GACGA G AAAGGO N AGAAA T T CCCACC K ACAAA Attta atgtt ttta cgtta cgtta	AAAA I CATO P TCCC A CGCC T T AAAC tgc aagto tat: agc	Q CCA G G G G G G G G C L C C T C C C T G C C A C G G G G G G G G G G G G G G G	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161 2251 2341	ACCCAAA SG CAGTGGZ TR GACGCGC SD ATCTGAD FH CTTCCAC PA CCCCAGC aaaata gacagaa acatgtc gcaagaa aaaatt	AAGAGA E P GAGCCCC N R CACCGT Q H CCAGCAT A Q GCCCCAG A Q GCCCCAG A Q GCCCCAG A Q GCCCCAG A Q GCCCCAG A C C GCCCCAG C C C C C C C C C C C C C C	V C GTGTG GTGTG K M AAGAT F D TTCGA P A CCTGC P C CCATG CCATG CCATG tgctt tataa agatt atgta	N A TAACGO S S S GTCCAO L Y CTTGTZ L P TCTACO V N TGTTAZ tagtcao agctao gatcao aagcao gata	C CTTGC K GCAAG S ATTCT M CCATG I ACATA ACATA AcATA acacg ytttg catga	G L GGGACI N R AAACAC Q S CAGAC CAGAC L S CTCAC atgtc ctaga acatt gactc	Y TCTAT K GGAAA P STCCA T CCCACC K STAAA STAAA CGGaa atttgg ytaaa	ACTCA F K TTCAA G K GGGAA GGAGC GGAGC GGAGC S C AGCTC N M AAACAI tctca aaaaag caaaag gac <u>at</u> acatg	L : GCTTC K : GAAG' L : GCATC GCTTC GCTTC GCTTC GCTTC GCTTC GTTCC GCTTC G	H N CACAP F S TTTAG G V GGCGT L T CTTAC E N GGAAP A actgt aaaga ggggt aaaga	V ATGTG P SCCCCC Q CCCCAG L ACTTA Cgtag attaa cttee ctaca	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaac aaacta gaaag gaaag aaaat	CCAGC P GGCCCG E AGGAG S ACTCA V V ATGTG G Catca AGGAG S V V ATGTG G Catca	ACCAC L A CTCGC N L AATCI S H TCCCA T G ACCGC D A GATGC aatca gaagc tttaa ttett	M TTATO Y S ATTO S ATTO S ATTO S ATTO A S ATTO A S ATTO A S A TO S A TO S A TO S A TO S A TO S A TO S S A TO S S A TO S S A TO S S A TO S S A TO S S A TO S S A TO S S S A TO S S S S S S S S S S S S S S S S S S	GCTG <sup>7</sup> K F TTTT <sup>7</sup> L CCTG <sup>6</sup> M CCTG <sup>6</sup> CCTG <sup>6</sup> CCTG <sup>6</sup> Actor CCTG <sup>6</sup> CCTG <sup>6</sup> CCTG <sup>6</sup> CCTG <sup>6</sup> CCTG <sup>6</sup> CCTG <sup>7</sup> CCTG <sup>6</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>6</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>7</sup> CCTG <sup>6</sup> CCTG <sup>6</sup> CCTG <sup>7</sup> CCTG <sup>6</sup> CCTG <sup>6</sup> C	IGGA K E AAGG S K TCAA P P P CCGC L Y ITGT. Aaaac ttat taaaag aata tttt	GACGA G AAGGG N AGAAA T T CCCACC K ACAAA ACAAA atgt tttaa cgtta ccttc catcc	AAAA I CATO P TCCC A CGCC T AAACC tgc aagt tat: agc	Q CCA G G G G G G G G G G G C C C C C C C	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161 2251 2341 2431	ACCCAAA SG CAGTGGZ TR GACGCGGC SD ATCTGAD FH CTTCCAC PA CCCAGCT aaaata gacagaa acatgtc gcaagaa acatgtc	AAGAGA E P GAGCCCC N R CAACCGT Q H CCAGCAT A Q GCCCCAG A Q GCCCCAG A Q GCCCCAG A Q GCCCCAG A Q GCCCCAG A Q CGCCCAG CACCCAG CCCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCCAG CCCCCAG CCCCAG CCCCCAG CCCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCCAG CCCCCCAG CCCCCCAG CCCCCCCAG CCCCCCCC	V C GTGTG K M AAGAT F D TTCGA P A TTCGA P A CCCTGC P C CCCATG tgctt tataa agatt tataa agatt atgta	N A TAACGO S S S GTCCA( L Y CTTGT) L P TCTACO V N TGTTA: tagtc: tacaac agcta: aggta: aggta:	C CTTGC K S GCAAG S ATTCT M CCATG I ACATA A acacgg gtttg gtttg catga aaagt ttgat	G L GGGACI N R AACAC Q S CAGAC P P CCACCC CCACCC CCACCC CCACCC CCACCC CCACCCC CCACCCC CCACCCCCC	Y TCTAT K GGAAA P GTCCA T T CCACC K K GTAAA A CCACC K GTAAA A CCTAAA CCACC K CCACC K CCACC C K CCACC C K CCACT T CCACT T CCACT T C CCACT T C CCACT K S GGAAA C CCACT T C CCACT C C C CCACT C C C C	ACTCA FK TTCAA GGGAA GGGAAGC SA GGGAGC SC AGCTCC NM AAACAT tctca aaaaag caaaag gac <u>at</u> caaaag tcaacag	L : GCTTC K : GAAG L : TCTGG H : CCCATC L : GCTTC atcc: cttg: gtttc ttac: gata	H N CACAA F S TTTAC G V GGCGT L T CTTAC E N GAAAA A actgt aaaggggt aaaggggt aaagt	V ATGTG P SCCCCC Q CCCAG L CCCCAG L CC	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaac aacta gaaag gaaag gaaag tatat	CAGC P GGCCG E AGGAG S S ACTCA V V TGTGTG V V TGTGTG C Attat actg ggttt tcaa	ACCAC L A CTCGC N L AATCT S H TCCCA T G ACCGC D A GATGC D A GATGC datca gaagc tttaa ttett tgaca ca <u>att</u>	M CTATO Y S TATAT S NTTCZ I SAATO * * * * * * * * * *	K SAAG. F TTTT L L CATG Aaaaa gcgt cggt aaata aata agcc	IGGA K E AAGG S K TCAA P P P CCCGC L Y ITGT. aaaac ttat taaaag aata ttttt tatt	GACG G AAAGG N AGAAA T CCCACC K ACAAA ACAAA atgt tttaa cgtt cttc: atcc. caaa	AAAA I I CCATO P TTCCC A CGCC T T AAACC tgc aagto tat agc aagc	Q CCA G G G G G G G C G C C C C C C C C A C C C C	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161 2251 2341 2431	ACCCAAA SG CAGTGGA TR GACGCGCG SD ATCTGAT FH CTTCCAA PA CCCCAGCT aaaaata gaaaata gaaaagt aaaagt atgtgat aagatt#	AAGAGA E P GAGCCCC N R CACCGT Q H CCAGCAT A Q CCCCAG A Q CCCCAG A Q CCCCAG A Q CCCCAG A Q CCCCAG A Q CCCCAG A Q CCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCAG CCCCCCCAG CCCCCAG CCCCCCAG CCCCCCCC	V C GTGTG K M AAGAT F D TTCGA P A CCTGC P C CCATG CCATG CCATG tgctt tataa agatt agatt agacta gtagt	N A TAACGG S S S GTCCAG L Y CTTGTAG L P TCTACG V N TGTTAG tagata agataa agataa agataa	C CTTGC K S GCAAG S ATTCT M CCATG I I ACATA A acaccg gtttg catga aaagt tgcaaa	G L GGGACI N R AACAC Q S CAGAC P P P CCACC CCACC CL S CTCAC atgtc ctaga acatt gacts gacts gacts	Y TCTAT K SGAAA P STCCA T T CCACC K STAAA CGGaa atttgg gtaaa atttgg gtaaa tttca	ACTCA FK TTCAA GGGAA GGGAA GGGAGC SC AGCTC NM AACAI tctca aaaag gac <u>at</u> acatg tcagc acagt	L : GCTTY K : GAAG' CCAT( H : CCAT( C	H N CACAA F S TTTAC G V GGCGT L T CTTAC E N GGCGT E N GGAAA A actgt aaaga ggggt aaatgt aaatgt aaatgt catca	V ATGTG P SCCCCC Q CCCAG CCCAG L CCCCAG L CCCAG L CCCCAG L CCCAG L CCCAG L CCCAG L CCCAG L CCCAG L CCC	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaact agaaag aaaat aqtaa tctga	CAGC P GGCCG E AGGAG S CTCA V V GTGTG G CACTCA V V GTGTG G CACTCA V V CTCA CACTCA S CACTCA S CACTCA S CACTCA S CACTCA S CACTCA S CACGAG S S CACGAG S S CACGAG S S CACGAG S S CACGAG S S CACGAG S S CACGAG S S CACGAG S S CACGAG S S CACTCA V V CACGAG S CACTCA S CACTCA S CACTCA S CACGAG S S CACTCA S CACGAG S S CACTCA S C CACGAG S S CACTCA S CACGAG S S CACTCA S CACGAG S S CACTCA S CACGAG S S CACTCA S CACGAG S S CACTCA S CACTCA S CACTCA S CACTCA S CACGAG S S CACTCA S CACTCA S CACTCA S CACTCA S CACTCA S CACTCA S S CACTCACTCA S CACTCA S S CACTCA S CACTCA S CACTCACTCA S CACTCAC S CACTCA S CAC	ACCAC L A CTCGC N L AATCT S H TCCCA T G ACCGC D A GATCC D A GATGC aatca gaagc tttaa ttett tgaca ca <u>att</u> attaa a	M TATATO Y SATATO S ATTCZ I SAATCO * * CATG2 aaagaa attgaa aaagaa tttgaa	K SGAG F TTTT L ACTG M CATG Aaaaa agcgt aatt aatt aata agcc ttt	IGGA K E AAGG S K TCAA P P P CCGC L Y ITGT. Aaaac ttat taaaag aaata ttttt tatt tatt	GACGJ G AAAGGG N AGAAA T CCCACC K ACCAAA ACCAAA atgtt ttttaa ccttc: atccc. tctgg ccaaa tcag	AAAA I CATC P TCCC A CGCC T AACC tgc aagt tagc aagt tagc aagt tagc	Q CCA G G G G G G G C G C C C C C C C C A C C C C	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161 2251 2341 2431 2521 2521	ACCCAAA SG CAGTGGZ TR GACGCGG SD ATCTGAT FH CTTCCAC PA CCCCAGCT aaaaata gacagaa acatgtc gaaagga aaaagt aaaagt aaaagta aaaatta	AAGAGA E P GAGCCCC N R CACCGT Q H CAGCAT A Q GCCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q C C C C A Q C C C C A Q C C C C A Q C C C C C C C C C C C C C C C C C C C	V C GTGTG K M AAGAT F D TTCGA P A CCTGC P C CCATG CCATG tgett tataa agatt aggatt aggacta gtagt aaacc	N A TAACGG S S S GTCCAG L Y CTTGTT L P TCTACG V N TGTTAG tagtca aggtad aaggad a	C CTTGC K GCAAG S ATTCT M CCATG I I ACATA Aacacg ytttg catga aaagt gcaaa aaagt ttgat tattt	G L GGACI N R AACAC Q S CAGAC P P CCACC L S CTCAC atgtc ctaga actat gacta gactat gactat tttta	Y TCTAT K GGAAA P GTCCA T T CCCACC K GTAAA CCCACC K GTAAA Atttgg gtaaa atttgg gtaaa	ACTCA F K TTCAA G K GGGGAA G A GGGGGAG S C AGCTG N M AACAI tctca aaaaag gac <u>at</u> acaaag tcaaaag tcaaaag	L : GCTT( K : GAAG' L : TCTGG H : CCATC L : GCTT( atcc: cttg; gtta tttg; tttt; gta; ttt; ttt;	H N CACAPA F S TTTAG G V GGGCGI L T CTTAC E N GAAAPA actgt aaaga ggggt aatgt aaaga ggggt catca aactct	V ATGTG P GCCCCC V CCTAC Q CCCAG L CCCAG L CCCAG C CCCAG C L CCCAG C L CCCAG C L CCCAG C L CCCAG C L CCCAG C CCCAG C CCCAC C Q CCCACCAG C CCCAC C Q CCCACCAG C Q CCCACCAG C C CCCACCAG C Q CCCACCAG C C CCCACCAG Q CCCACCAG C C CCCACCAG Q CCCACCAG C C CCCACCAG Q CCCACCAG C C CCCACCAG Q CCCACCAG C C CCCACCAG C C CCCACCAG C C CCCACCAG Q C CCCACCAG C C CCCACCAG C C CCCACCAG C C CCCACCAG C C CCCACCAG C C CCCACCAG C C CCCACCAG C C CCCAG C C CCCAG C C C C	N R AACAG T E ACAGA S H AGCCA G AGCCA G CATGG T D ACAGA aaaact aaaact aaaact aaaact aatat aqtaa tctga gtctt	CAGC P GGCCG E AGGAG S S ACTCA V V GTGTG G Catca Attat tattat tattat cagtt ttat	ACCACA L A CTCGC N L AATCI S H TCCCA T G ACCGG D A GATGC aatca gaagc tttaa ttett ttgaca ca <u>att</u> atata gtttt	M TATC Y CATAT S TTCZ I SATCZ * CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATGZ CATATO S SATTCZ S S S S S S S S S S S S S S S S S S S	SCTG' K F F TTTT' L CTGG M CATG' CAT	IGGA K E AAGG S K TCAA P P P CCCGC L Y TTGT aaaac ttat taaaag aata ttttt tatt tatt	GACG G AAAGG N AGAA T T CCCAC K ACAA C ACAA C C C C C C C C C C	AAAA I CATC P TCC' A CCC' T AACC T AACC tgc aagu tat tag tag tag tag tag tag	Q CCA G G G G G G G G G G G G G G G G G G	
1331 301 1441 331 1531 361 1621 391 1711 421 1801 1891 1981 2071 2161 2251 2341 2431 2521 2611 2701	ACCCAAA S G CAGTGGZ T R GACGCGGC S D ATCTGAD F H CTTCCAC P A CCCAGCI aaaata gacagaa acatgto gcaagaa acatgto gcaagaa acatgto tgcaagaa acatgto tgcaagaa acatgto	AAGAGA E P GAGCCCC N R CAACCGT Q H CAGCAT A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q GCCCAG A Q CGCCCAG A Q GCCCAG A Q CGCCCAG A Q CGCCAG A Q CGCCAG CAG CAG CAG CAG CAG CAG CAG CAG	V C GTGTG K M AAGAT F D TTCGA P A TTCGA P A CCCTGC P C CCCATG tgctt tataa agatt tataa agatt agacta gtagt, aaaac aaaaa	N A TAACGO S S GTCCAC L Y CTTGTZ L P TCTACC V N TGTTAC U N TGTTAC tacaca agctaca agctaca agctaca agctaca agctaca tacaca control tacaca control tacaca gagta aataa aacacc tttctt	C CTTGC K S GCAAG S ATTCT M CCATG I ACATA A acacg gtttg gcatga aaagt ttgat tattt tattt tattt	G L GGGACI N R AACAC Q S CAGAC P P CCACCCC CCACCCC CCACCCC CCACCC CCACCC CCACCC CCACCC CCACCC CCACCC	Y TCTAT K GGAAA P GTCCA T T CCACC K K GTAAA A CCACC K GTAAA A CCACC K GTAAA A CCACC K CCACC C K CCACC C K CCACC C K CCACT T	ACTCA F K TTCAA GGGAA GGGAGC G A GGGAGC S C AGCTC N M AACAI tctca aaaaag gac <u>at</u> tcaaca tcaacag tcaacagt tcagc	L : GCTTT K : GAAG' L C TCTGG C CCATC C C C	H N CACAA F S TTTAC G V GGCGT L T CTTAC E N GAAAA A CTTAC E N GAAAA GGAAAA ggggt aaaga ggggt aaaga aaaat	V ATGTG P SCCCCC Q CCCAG CCCCAG CCCC	N R AACAG T E ACAGA S H AGCCA H G CATGG T D ACAGA aaaac aacta gaaag gaaag gaaag gaaag tatat tatga	CAGC P GGCCG E AGGAG S CTCA V V TGTGTG Gatca Catca attat tattat ggttt cactg catca attat gaaaa	ACCACA L A CTCGC N L AATCT S H TCCCA T G ACCGC D A GATCC D A GATCC D A GATCC Caatca tttaa ttott tgaca caatt gattt gaatg	M TATC Y SATAT S ATTCZ I SAATC * * CATGZ C	K GAAG F TTTT L CATG CATG CATG CATG CATG CATG CATG CATG	IGGA K E AAAGG S K TCAA P P P CCGGC L Y ITTGT. aaaac ttat ttat taaaag aata ttttt tcgg gtaa ca <u>aa</u>	GACGI G AAAGGG N AGAAA T CCCACG K ACAAA atgt tttaa ccgtt ccaca ttcga ccaaa tgag tcaca	AAAA I CATO P TCCC' A CGC' T AAAC AAAC AAAC AAAC AAAC AAAC AAAC	Q CCA G G G G G G G L L CCT CCT CCT CCT CCT CCT CCT CCT CCT	

#### (See figure on previous page.)

Fig. 1 Nucleotide and putative amino acid sequences of *GATA*1 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

mammals, birds, amphibians, fishes and invertebrates, respectively. The phylogenetic tree revealed that the GATA1 protein in *C. auratus* red var. is closest to of its ortholog in *C. auratus*, with a high bootstrap value of 99%. All the fish GATA1 proteins clustered together, and diverged from their counterparts in other groups. GATA1 proteins in invertebrates were far separated from those in vertebrates. Thus, the phylogenetic tree reflected a genetic consistency among those species in evolution.

#### Distribution of GATA1 in C. auratus red var. tissue

qRT-PCR was performed to analyze the tissue distribution of *GATA*1 mRNA expression. As shown in Fig. 5, *GATA*1 expression was detected in all organs tested, and the values were calibrated against the expression level in heart (H). *GATA*1 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 the lowest levels in muscle (M) (0.68 folds), liver (L) (0.35 folds), and ovary (O) (0.33 folds).

# GATA1 expression in different developmental stages after NP treatment

To determine the effect of NP exposure on GATA1 gene expression, the levels of GATA1 mRNA in different developmental stages were examined (Fig. 6). During the normal embryonic development, the GATA1 gene expression can be detected from the N stage, and the expression level increased at the S5 stage, decreased continuously at S14 and S21 stages, and then increased again at the P5 stage. Afer that, it reached to the highest at the P25 stage, and dropped again after embryo hatching. GATA1 expression levels in both the 3 µmol / L and 5 µmol / L NP-treated groups showed the biggest difference from that in the control group at the S14 stage, while the biggest difference in GATA1 expression between the 7 µmol / L NP-treated and control groups happened at S21 stage. (Fig. 6). On the other hand The biggest difference in GATA1 mRNA levels at the neuroblast stage was found between the control and the 3 µmol / L NP treated groups when compared with other group pairs. When embryos developed to the 5 somite stage, the 7 µmol / L NP-exposure group had greater effect on the expression of GATA1 gene than other treated groups with lower NP dosages (Fig. 6).



Page	5	of	13
------	---	----	----

GATA1 CaGATA1 SrGATA1 SrGATA1 DrGATA1 ChGATA1 MaGATA1 MaGATA1	1 1 1 1 1 1 1 1	METSTECARWVSSSMVSSEVMPNYPEDSSYMVHTEEGSVYPYTD METSTECARWVSSSMVSSEVMPNYPEDSSYMVHTEEGSVYPYTD METSTECARWVSSSMVSSEVMPTYPDSSYMVHTEEGSVYPYTD METSSECARWVSSSMVSSEVMPTYPEDSSYMVHTEEGS METSSECARWVSSSMVSSEVMPTYPEDSSYMVHTEEGSVYPYTD MESSVOQVRWGSSEVMPTYTSDSSFLLHGEEASIFPCTDADHSG MEDSEQSHWVPPALISSDPTAGFSSEPGLLPPGDEAEAFFSGOP MTAQSEASRINIKVGSAPSQTLFQSENPAHDACHRVTSPCLWLD MTAQSEASRINIKVGSAPSQTLFQSENPAHDACHRVTSPCLWLD	APHSSLPSLFSSPVNGRPSGAFOTSSVYPVYSSPFL APHSSLPSLFSSPVNGRPSGAFOTSSVYPVYSSPFL APHSSLPSLFSSPVNGRPSGAFOTSSVYPVYSSPFL VYPYTDADHSSLPSLFSSPVNGRPSGAFOTSSVYPV APHSSLPSLFSSPVNGRPSGAFOTSSVYPVYSSPIL LPSLFSSPVHSRPTAAFQHSPVYPIYSSFILGNLSW PYSGLHSFFSTPSHSRAPPTYRHSSVRQVFSSPSL DSGHQSLSSDSIPPPPSPSLYGNPALTPPCSCILSS DSGHQSLSSDSIPPPPSPSLYGNPALTPPCSCILSS
GATA1 CaGATA1 SrGATA1 SrGATA1 DrGATA1 ChGATA1 MaGATA1 MoGATA1	81 81 81 81 81 81 81 81	GNLSWLESSNGPSLINLFPSSPSSWHSSVFSSSFHGSTPHSSAR GNLSWLESSNGPSLINLFPSSPSSWHSSVFSSSFHGSTPHSSAR GNLSWLESSNGPSLISLFSSSPSSWHSSAFSSSFHGSTPLSSAR YSSPFLGNLSWLESSSGPSLINLFPSSPSSWHSSAFSSFHGSTSLSSAR GNLSWLESLSSPSLINLFPSSPSSWHSSAFSSSFHGSTSLSSAR LEGSGSSLINNFPSSPSSWHSSAFSKASSSFSSSARPEQKE LGNLQLLDGACSSHSPYTPPGSSWSSSPLSKMHSHTPTSLYTPC TAGWDSYYGNAYPSWVPGKPVPEQRECVCCGTSSASLWRRDATG TAGWDSYYGNAYPSWVPGKPVPEQRECVCCGTSSASLWRRDATG	PPRSALPSLIQDQKDTLIVQESMKGQRLSPPGGGEA PPRSALPSLIQDQKDTLIVQESMKGQRLSPPGGEA PPRSALPSIIQDQKDTLIVQESMKGERLSPPGGEA SLSSARPHSALPSLIQDQKDTIIVQDSMKGERPSP PPRSALPSLIQDQKDTLIVQDSMKGEMPSPPGGES SVKGERSSPPGGGLFPLNSSLSGVYPHSHSTKTHYS LTSSFTSSRDGYGSPGRESPRLQEALKAERLSPLGG RYLCHTCSLQHKPNNRPVLRPKRRATVTRRAGTQCV RYLCHTCSLQHKPNNRPVLRPKRRATVTRRAGTQCV
GATA1 CaGATA1 SrGATA1 SgGATA1 DrGATA1 ChGATA1 MaGATA1 MoGATA1	161 161 161 161 161 161 161 161	FGGVFSPSLSSVYAQTHSSKTHSQSLSHYSPYGSFTENYNSSLL FGGVFSPSLSSVYAQTHSSKTHSQSLSHYSPYGSFTENYNSSLL FGGVFSPSVSSVYAQTHSSKTHSQSLSHYSPYGSFTQDYNSSLL PGGGSFGGVFSPSVSSVYAQTHSSKTHSQSLSHYSPYGSFTQD FSVFSPSVSSVYAQTHSSKTHLQSLSHYSPYGSTQDYSSLL PYGGHTQDYSSLLYAPATFTPKLCSKMNFSPLDARECVNCGAT SCVSSSFMNLTPAGSVYTSSSHPHMLSPYSPYMTGPQDYNYPG NCETVTTTLWRNNAGQPVCNACGLYYKLHRVNRPLTMKREETQ NCETVTTTLWRNNAGQPVCNACGLYYKLHRVNRPLTMKREETQ	YTPSSFPPKLCSKMKFSPLVATEPRECVNCGATASP YTPSSFSPKLCSKMKFSPLVATEPRECVNCGATASP YTPSSFSPKLCSKMKFSPLEPRECVNCGATATPLWR YSSSLYTPSSFSPNLCSKMKFSPLEPRECVNCGAT APSSFSPNLCCKMKFSPLEPRECVNCGATATPLWR ATPLWRRDGTGHYLCNACGLYHKMNGQNRPLIRPKK PGAWINPSFSPKLRNKMRISTPEARECVNCGATATP TRNRKVINKMNKKRISGAKSETEWCWLASPTDEAILH TRNRKVINKNKKRISGAKSETEWCWLASPTDEAILH
GATA1 CaGATA1 CcGATA1 SrGATA1 SgGATA1 DrGATA1 ChGATA1 MaGATA1 MoGATA1	241 241 241 241 241 241 241 241 241 241	LWRRDGTGHYLCNAGGLYHKMNRQNRELIRPKKRLVISKRTGTO LWRRDGTGHYLCNAGGLYHKMNRQNRELIRPKKRLVISKRTGTO RDGTGHYLCNAGGLYHKMNGQNRELIRPKKRLVISKRTGTOCAN ATPLWRRDGTGHYLCNAGGLYOKMNRQNRELIRPKKRLVISKRTGTQCAN RDGTGHYLCNAGGLYHKMNGQNRELIRPKKRLVISKRTGTQCAN RPVVSKRIGTQCANGGLYHKMNGQNRELIRPKKRLVISKRTGTQCAN SFTHLPPTELCSWERSAGHDQSHWESEAGTISCERAESRATLRE SFTHLPPTELCSWERSAGHDQSHWESEAGTISCERAESRATLRE	CVNCQTSTTTTWRRNASGEPVCNACGLYFK HNVNR CVNCQTSTTTWRRNASGEPVCNACGLYFK HNVNR CQTSTTTWRRNASGEPVCNACGLYFK HNVNRPLA GTQCANCQTSTTTWRRNAGGEPVCNACGLYFKLHN QTSTT LWRRNAGGEPVCNACGLYFK HNVNRPLA VNRPLAMKKEGIQTRNRKVSSKNRKGKKFSTMEENL CANCHTSTTTWRRNASGEPVCNACGLYFK HNVNR ERSAEVRGGGUHGC
GATA1 CaGATA1 CcGATA1 SrGATA1 SgGATA1 DrGATA1 ChGATA1 MaGATA1 MoGATA1	321 321 321 321 321 321 321 321	PLAMKKEGIOTANRKMSSKNRKGKKFSPTEENLYFSKNPGSDQH PLAMKKEGIOTANRKMSSKNRKGKKFSPTEENLYFSKNPGSDQH MKKEGIQTRNRKVSSKNRKGKKFSATEENLYFSKNPASDQHFDL VNRPLAMKKKGIQTRNRKVSSKNRKGKKFSATEENLYFSNNPSS MKKEGIQTRNRKVSSKNRKGKKFSATEENLYFSKYPSSDQNFDL YCDFPKTPACDQHFDMYSQSPAALGVYSHSGQSTAYLPYPYHSS PLAMKKEGIOTANRKVSNKNKKSKKVAMFESYSDGPQPMDDCGP	FDLYSQSPGALGVYSHSSHSLPPTAAFHAQPALPMP FDLYSQSPGALGVYSHSSHSLPPTAAFHAQPALPMP YSQSPGALGIYSHSSHSLPPTAAFHSHASLPYPYHP DQHFDLYSQSPGALGIYSHSSHSLPPTAAFHSHASL YSQSPGALGVYSHSCHSLPPTAAFHSHASLPYPYHP PAVLPSTV
GATA1 CaGATA1 CcGATA1 SrGATA1 SgGATA1 DrGATA1 ChGATA1 MaGATA1 MoGATA1	401 401 401 401 401 401	PTSCHLTQHGVTGIMLYKTLPAAQPCVNILSKNMLENLTDVDA PTSCHLTQHGVTGIMLYKTLPAAQPCVNILSKNILENLTDVDA PAAILPGLV PYPYHPPAAILPSMV PAAILPSTV	Identity 99.10 % 83.97 % 81.07 % 80.36 % 59.78 % 39.60 % 20.77 % 20.77 %
Fig. 3 (See leg	end on	next page.)	

#### (See figure on previous page.)

**Fig. 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. *Ca*GATA1 stands for GATA1 protein in *Carassius auratus* (Protein ID. XM\_026253445.1), *Cc*GATA1 stands for GATA1 protein in *Cyprinus carpio* (Protein ID. XM\_019103428.1), *Sr*GATA1 stands for GATA1 protein in *Sinocyclocheilus rhinocerous* (Protein ID. XM\_016537268.1), *Sg*GATA1 stands for GATA1 protein in *Sinocyclocheilus grahami* (Protein ID. XM\_016271639.1), *Dr*GATA1 stands for GATA1 protein in *Danio rerio* (Protein ID. XP\_021334219.1), *Ch*GATA1 stands for GATA1 protein in *Chionodraco hamatus* (Protein ID. KP221299.1), *Ma*GATA1 stands for GATA1 protein in *Mastacembelus armatus* (Protein ID. XP\_026189425.1), *Mo*GATA1 stands for GATA1 protein in *Monopterus albus* (Protein ID. XM\_020614979.1)

Apparently, NP affected *GATA1* expression during the development of *C. auratus* red var. embryos, and the greatest effect took place in somatic embryos.

# The methylation levels of GATA1 changed significantly in NP -treated groups

Our selected *GATA1*PCR target fragment was 277 bp in size and had 10 CpG sites. Table 1 shows the methylation status of 10 CpG sites in the control and NP5 treatment groups. In the control groups, the methylation rates of *GATA*1 gene at 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. While in the NP-treated groups, the methylation rates of *GATA*1 gene at 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 (Fig. 7). Obviously, methylation levels of *GATA*1 gene in the NP-treated embryo groups were mostly higher than that in the control group. We analyzed correlation between *GATA*1 mRNA expression and methylation of *GATA*1 gene in both the control



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



and NP-treated groups, and found *GATA1* expression was significantly positively correlated with its methylation level in the control group (r = 0.771, P < 0.05), but not in the NP-treated groups (r = 0.533, P > 0.05).

# Discussion

In this study, *C. auratust*he full-length *GATA*1 cDNA sequence was obtained from *C. auratus* red var. by

homologous cloning and RACE Technology*c. auratus. GATA*1 cDNA is 2730 bp in length with a 1332 bplong ORF encodeing 443 amino acids (aa) a 541 bplong 5'-UTR and an 857-long 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%. Also, phylogenetic analysis



CpG sites	Methylation level of the CK/%	Methylation level of NP5/%
- 3681	100.00	100.00
- 3668	100.00	100.00
- 3657	97.94	100.00
- 3528	86.60	96.08
- 3518	79.38	97.06
- 3490	60.82	69.61
- 3478	89.69	97.06
- 3466	98.79	100.00
- 3434	100.00	100.00
- 3413	100.00	100.00

Table 1 Methylation status of the 5'UTR region of the GATA1 gene

The CpG sites were located between - 3413 ~ - 3681 upstream to the start codon

CK: 0 µmol / LNP-treated embryos; NP5: 5 µmol / LNP-treated embryos

Degree of methylation = methylation number of the measured CpG sites/total number of the CpG sites measured

showed that the GATA1 protein in *C. auratus* red var. was closest to that in *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: an amino terminal zinc finger (N) and a carboxyl terminal one (C) [24]. We analyzed the conserved domains in the predicted GATA1 protein with SMART, and found that the *C. auratus* red var. GATA1 contains two ZnF domains (aa 225–275, aa 279–329); The *C. auratus* red var. *GATA*1 gene consists of five exons, which is consistent with its paralogs in other species.

The tissue distribution of *GATA*1 mRNA was analyzed through qRT-PCR. *GATA*1 expression was detected in all tested tissues of *C. auratus* red var. *GATA*1 had the highest expression level in testicle (T); intermediate levels in pituitarium (P), spleen (S), gills (G), brain (B); and lower levels in muscle (M), liver (L), and ovary (O). *GATA*1 is abundantly transcribed in mouse testis and regulates genes involved in the earliest stages of spermatogenesis [25]. Studies have shown that spermatogenesis is induced by GATA-1 expression in Sertoli cells. As atranscription factor, GATA-1 is a developmental stage and spermatogenic cycle-specific regulator of gene expression in Sertoli cells [26]. In sexually mature *C. auratus* red var. individuals used in this study, the *GATA*1 expression is most enriched in the testes among all



tissues and organs. The GATA transcription factor family is essential for pituitary cell differentiation and gonadotropin subunit expression [27]. GATA1 inhibits formation of rat cortical neurons, and GATA1 overexpression ofin the hippocampus can cause depressive behavior in rats [28]. Spleen is an important hematopoietic organ in animal bodies, and the gills are the respiratory fish organs with a large number of capillaries. Accordingly as a key regulator of red blood cell production [18], the GATA1 gene is expressed at high levels in P, S, G, and B of C. auratus red var. We also found very low GATA1 expression in L and O of C. auratus red var. which was consistent with a report in Nile tilapia (Oreochromis niloticus) [29]. C. auratus. Thus, our results with GATA1 expression pattern in various tissues and organs C. auratus provided essential cues to understand GATA1 functions in C. auratus red var. adults.

In C. auratus red var. embryos, the GATA1 gene starts to be detectable from the neural embryo stage, and stays continuously expressed during embryonic development with relatively stable expression levels, which indicates that GATA1 is involved in the entire embryonic development of C. auratus red var. GATA1is also involved in early embryonic development in other fish. In general, early blood islands emerged in the yolk sac endoderm and splanchnic mesoderm during early embryonic development. GATA expression became detectable in zebraectoderm 9 h after fertilization fish [**30**]. In Branchiostoma belcheri, GATA1 expression signal could also be detected at the mesendoderm of gastrula stage [31]. NP affects GATA1 expression during the development of C. auratus red var. embryos, C. auratus with a strongest effect at the somatic stage, and with a most effective NP concentration at 3 µmol / LC. auratus. Durvertebrate early embryogenesis, the ventral ing development is directed by the ventral-to-dorsal activity gradient of the bone morphogenetic protein (BMP) signaling [32]. Abnormalties in the BMP signaling pathway may cause strong dorsalization phenotypes in embryos [33]. GATA1 gene is a downstream target 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 count for dorsalization after NP treatment. The transcriptional activity of GATA1 is related to the expression level of vitellogenin (Vg) [36]. Vg expression measurement has been used as a biomarker of exposure to endocrine-disrupting chemicals [37]. Up-regulation of GATA1 expression in C. auratus red var. embryos under NP exposure may increase Vg expression, further proving NP is an environmental endocrine disruptor. Yokomizo et al.'s experiments in mouse embryos provided evidence showing 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 production in the embryo, so *GATA1* and *GATA2* double deficient mice exhibit no visible blood cells [39]. In addition, *GATA1* mutations lead to increased thrombosis in mice [20]. Therefore, the occurrence of thrombosis in *C. auratus* red var. embryos under NP stress may be related to down-regulation of *GATA1* gene expression.

DNA methylation is a heritable modification that affects gene expression without changing DNA sequences. This modification is crucial to embryonic development. Either abnormally higher or lower methylation levels could be detrimental to the normal growth and embryonic development [40]. Reduced methylation on H3-K4 in Lsd1 mutant fruit flies results in tissue-specific developmental defects [41]. Compared with that in normal embryos, H19 gene methylation is severely altered in abnormally developing embryos [42]. In 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 level of GATA1 in the control group was lower than that in the NP-treated C. auratus red var. embryos, which demonstrates that NP stress increases methylation level of GATA1 in C. auratus red var. during embryonic development. GATA1 expression is significantly positively correlated with its methylation level in the control group, but not in the NP-treated group. In addition, high levels of GATA1 expression during the same developmental period are not necessarily low in methylation level. Since GATA1 expression levels at different developmental stages in the control group were apparently correlated with GATA1 methylation levels, we suspected that NP exposure affected GATA1 expression through changing its methylation status. However, due to the lack of correlation between GATA1 expression and its methylation in the NP-treated groups, other mechanisms must have been involved. C. auratus This is similar to what Okada et al. reported in mice. In 3 T3-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, Milenkovic et al. found that daily supplementation with 200 mg MOF for 8 weeks modulates the expression of genes associated with cardiovascular disease pathways without major changes of their DNA methylation status [45].

# Conclusions

In this study, the full-length cDNA sequence of *GATA1* gene in *C. auratus* red var. was cloned, and the special and temporal expression patterns of *GATA1* gene in various tissues/organs and embryonic developmental stages of *C. auratus* red var. were analyzed. Changes in *GATA1* expression during NP-stressed embryonic development was measured, which revealed a role of NP-stress in regulation of *GATA1* expression. It provides important cues for unravel *GATA1* functions in fish development and molecular mechanisms through which NP leads to abnormal development in fish embryos.

# Methods

#### Fish and sampling

Two-year-old healthy *C. auratus* red var., weighting about  $200 \pm 10$  g with an average length of  $15 \pm 3$  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  $25 \pm 1$  °C and fed twice daily with a commercial diet for 1 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 2 min after fertilization were exposed to NP with the concentrations of 0  $\mu$ mol/L (blank control, 0.01% ethanol), 3  $\mu$ mol/L, 5  $\mu$ mol/L and 7  $\mu$ mol/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.

Intact 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 with 30 embryos in each group were collected at each time points and we used liquid nitrogen to stop embryo development.

Samples within each group were homogenized, and aliquots of homogenized tissues were taken for DNA and RNA isolation, respectively.

# 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<sup>™</sup> 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. aur*atus red var., primers GATA1-F1/R1 (Table 2) were designed and synthesized based on the highly conserved regions of known fish *GATA1* sequences, including *Car*assius 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<sup>TM</sup> RACE cDNA Amplification Kit. The full-length cDNA sequences were amplified by PCR using GATA1-F2/R2 primers (Table 2) within the 5' and 3'UTRs, respectively.

Sequence Manipulation Suite (STS) (http://www.biosoft.net/sms/) was used to analyse the sequences of *GATA*1 from *C. auratus* red var.. The BLASTP program (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/). Multiple sequence alignments were performed by the ClustalX 2.1 program (http://www.ebi.ac.uk/tools/ clustalx2.1). Simple Modular Architecture Research Tool (SMART) (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) 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 *GATA*1, primers (Table 3) were designed to amplify the genomic sequences gradually. Five overlapping fragments were amplified from gDNA and sequenced.

 Table 2 Primers for full-length cDNA cloning and qRT-PCR

Primer name	Sequence $(5' \rightarrow 3')$	Application
GATA1-F1	GCTCCACAAAAGAAAGTCAT	partial sequence obtaining
GATA1-R1	ACGAGGGTGTGTAGAGAAGT	
GATA1-F2	CCTCAATCACCCACTATGCC	ORF qualifying
GATA1-R2	GTGGATTGAGATTCCGACAT	
GATA1-R- out	GCTCTGGCATAGTGGGTGATTGAG GTTA	5'-Race PCR amplification
GATA1-R-in	ATAATCGAAACACATCACTCACCC CA	
GATA1-F- out	GGCGTCTACAGCCACTCATCCCAT TCAC	3'-Race PCR amplification
GATA1-F-in	GGATGCTTTAGTCACACGATGT CGGAAT	
GATA1-qF	CCTTCCCTCCCTTATCCAG	qRT-PCR amplification
GATA1-qR	GGTAGTGTCCCGTTCCATC	

The 5 'unknown sequence of the *GATA*1 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 4) 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

Table 3 Primers for genomic DNA sequences

Primer name	Sequence $(5' \rightarrow 3')$	Product Length (bp)
GATA1-gDNA- F1	CAATCACCCACTATGCCAGAGC	914 bp
GATA1-gDNA- R1	GCTGAATAAAGAGGGCAGGC TG	
GATA1-gDNA- F2	TGGTCCACACTGAGGAGGGTTC	1238 bp
GATA1-gDNA- R2	GGAAACTGTGTACCAGGG ACGG	
GATA1-gDNA- F3	CTGAGCCACTACAGTCCCTATG	1172 bp
GATA1-gDNA- R3	AGGGGTCTGTTCTGTCTATTCA	
GATA1-gDNA- F4	GATGGAACGGGACACTACCTCT	664 bp
GATA1-gDNA- R4	TAGAGTCCGCAAGCATTACACA	
GATA1-gDNA- F5	GGAACTCAATGTGTGAACTGTC	528 bp
GATA1-gDNA- R5	CTGTTCTTGCTGGACATCTTAC	

Table 4	Primers	for	5'unknown	sequences
---------	---------	-----	-----------	-----------

Primer name	Sequence $(5' \rightarrow 3')$	Application
GATA1-SP1	CAGAGCAAGGCTGTGGAA GTCATTT	5'- Walking PCR amplification
GATA1-SP2	GTCCTGGTTTGGAGGTTGTTTG CC	
GATA1-SP3	GCTTCCACCTTTGATAGA GGCTGA	
GATA1-F3	ATGGCTGTAGTGCTCATTCATC GCT	verification
GATA1-R3	CAAGAGATTCACAACTATGACT GCG	

expression of *GATA*1 in different tissues, including intestine (I), liver (L), spleen (S), gills (G), middle kidney (MK), muscle (M), head kidney (HK), heart (H), brain (B), pituitarium (P), testicle (T), and ovary (O). Specific primers (Table 2) were designed for qRT-PCR. The housekeeping gene  $\beta$ -actin [46] (Table 5) 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  $\mu$ mol/L, 3  $\mu$ mol/L, 5  $\mu$ mol/L and 7  $\mu$ mol/L) were analyzed. The housekeeping gene  $\beta$ -actin was used as the reference gene, and the *GATA1* expression level in neuroblast stage under 0  $\mu$ mol/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  $2^{-\Delta CT}$  method [47]. The *GATA*1 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.

Table 5 Primers for others

Primer name	Sequence $(5' \rightarrow 3')$	Application
β-actin -F	GGCCTCCCTGTCTATCTTCC	qRT-PCR
β-actin -R	TTGAGAGGTTTGGGTTGGTC	
GATA1-F4	TTTATTTCGTTGGAGGAG ATC	methylation sequence obtaining
GATA1-R4	CGCTATCTAAAATACTTT CCACG	

# 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<sup>™</sup> 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 5) 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 DH5a cells (TransGen, China) and cultured at 37 °C. Positive colonies were selected and sequenced by a Bio-tech company (TIANYI HUIYUAN, China). Fifteen groups of colonies were selected for sequencing at each developmental stage. 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: Pituitarium; I: Intestines; MK: Middle kidney; M: Muscle; HK: Head kidney; H: Heart; B: Brain; 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

#### Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12863-021-00966-3.

Additional file 1 Checklist S1. Completed "The ARRIVE Guidelines Checklist" for reporting animal data in this manuscript.

#### 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.

#### 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.

#### Funding

This work was supported by the Scientific Research Project of National Natural Science Foundation of China (31873038), Hunan Provincial science

and technology Department (2019NK4218, 2019JJ70038), 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.

#### Availability of data and materials

Data and materials are available from the authors on reasonable request. The *GATA*1 cDNA sequence is available in the GenBank (Accession number MT322308)

# **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.

#### Author details

<sup>1</sup>Hunan Key Laboratory of Economic Crops Genetic Improvement and Integrated Utilization, School of Life Sciences, Hunan University of Science and Technology, Xiangtan 411201, Hunan, People's Republic of China. <sup>2</sup>Key Laboratory of Tropical and Subtropical Fishery Resources Application and Cultivation, Ministry of Agriculture, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, Guangdong, Hunan, People's Republic of China.

#### Received: 22 May 2020 Accepted: 2 March 2021 Published online: 18 March 2021

#### References

- 1. Yu J, Wu J, Zhang B, Xu J. Toxic effects of nonylphenol on the organisms and its mechanism. J Environ Hygiene. 2013;3:268–72.
- Lukac N, Lukacova J, Pinto B, Knazicka Z, Tvrda E, Massanyi P. The effect of nonylphenol on the motility and viability of bovine spermatozoa in vitro. J Environ Sci. 2013;48(8):973–9.
- 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.
- Tanaka JN, Grizzle JM. Effects of nonylphenol on the gonadal differentiation of the hermaphroditic fish, *Rivulus marmoratus*. Aquat Toxicol. 2002;57(3): 117–25.
- 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.
- Lahnsteiner F, Berger B, Grubinger F, Weismann T. The effect of 4nonylphenol 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.
- Xiao Q, Xu YY. On the effects of nonylphenol on the embryonic development of *Puntius conchonius*. J Saf Environ. 2010;10(6):9–12.
- Zhang QY, Sun YD, Wang ZJ, Hu XJ, Kui X. Toxic effects of Nonylphenol on the embryonic development of goldfish (*Carassius auratus*). Progress Modern Biomed. 2016;16:3040–3.
- 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.
- Lü B, Zhan P. Effects of nonylphenol on brain gene expression profiles in F1 generation rats. Toxicol Environ Chem. 2009;91(3):559–65.
- 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 Pharmacol. 2012;33(1):98–106.

- Klempt M, Vobach M, Wiegand H, Preuss TG, Schäffer AKU. 353nonylphenol induces expression of the t-box 6 gene in zebrafish embryos-linking transcriptional information with deformities. J Fish Sci. 2013;7(1):30–42.
- 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, *CYP3A* and *CYP1A1* in juvenile Atlantic salmon (Salmo salar). Comparative Biochem Physiol Part C. 2006;142(1–2):142–50.
- Cocci P, Mosconi G. Palermo F a: effects of 4-nonylphenol on hepatic gene expression of peroxisome proliferator-activated receptors and cytochrome P450 isoforms (*CYP1A1* and *CYP3A4*) in juvenile sole (Solea solea). Chemosphere. 2013;93(6):1176–81.
- 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.
- Bárbara FZ, Pavón L, Calés C. CDC6 expression is regulated by lineagespecific transcription factor GATA1. Cell Cycle. 2012;11(16):3055–66.
- Ling T, Crispino JD, Zingariello M, Martelli F, Migliaccio AR. GATA1 insufficiencies in primary myelofibrosis and other hematopoietic disorders: consequences for therapy. Expert Rev Hematol. 2018;11(3):169–84.
- 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.
- 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.
- 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.
- Lü XH, Gu Y, Song Y. Toxicity and tissue accumulation of nonylphenol in Carassius auratus red variety, Grass Carp and Sliver Carp. J Hygiene Res. 2012; 41(5):785–9.
- 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. J Fish China. 2020;44(10):1619–36.
- 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.
- 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.
- 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.
- 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.
- 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.
- Kwan W, North TE. Netting novel regulators of hematopoiesis and hematologic malignancies in Zebrafish. Curr Top Dev Biol. 2017;124:125–60.
- Zhang YJ, Mao BY. Developmental expression of an Amphioxus (*Branchiostoma belcheri*) gene encoding a GATA transcription factor. Zool Res. 2009;30(2):137–43.
- 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.
- 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.
- Kim JH, Park JB, Lee JY, Kim J. PV.1 suppresses the expression of *FoxD5b* during neural induction in Xenopus embryos. Mol Cell. 2014;37(3):220–5.

- Shibata K, Ishimura A, Maéno M. GATA-1 inhibits the formation of notochord and neural tissue inXenopusEmbryo. 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.
- 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.
- Tomomasa Y, Satoru T, Naomi M, Takashi K, Masatsugu E, Asami W, Ritsuko S, Osamu O, Motomih O, Hitoshi O. Characterization of *GATA-1*<sup>+</sup> hemangioblastic cells in the mouse embryo. EMBO J. 2007;26(1):184–96.
- 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.
- 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.
- 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.
- 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.
- 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. J Nutr Biochem. 2013;24(7):1373–80.
- 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. Comparative Biochem Physiol Part C. 2007;145(3):350–62.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using realtime quantitative PCR and the 2<sup>-act</sup> method. Methods. 2001;25(4):402–8.

# **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

