


Long noncoding RNAs implicated in embryonic development in *Ybx1* knockout zebrafish

Chen Huang¹, Bo Zhu², Dongliang Leng², Wei Ge² and Xiaohua Douglas Zhang² 

¹ State Key Laboratory of Quality Research in Chinese Medicine, Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, China

² Faculty of Health Sciences, University of Macau, Taipa, China

Keywords

deep RNA sequencing; long noncoding RNAs; REDOX; YBX1; zebrafish embryo development

Correspondence

X. D. Zhang and W. Ge, Faculty of Health Sciences, University of Macau, Taipa, Macau, SAR, China. E-mails: douglaszhang@um.edu.mo (XDZ); weige@um.edu.mo (WG)

Chen Huang and Bo Zhu contributed equally to this article

(Received 24 September 2020, revised 29 November 2020, accepted 3 December 2020)

doi:10.1002/2211-5463.13057

Y-box-binding protein 1 (Ybx1, YB-1), also known as Y-box transcription factor, is involved in a variety of biological processes (BPs) and pathways, including embryogenesis, reproduction and development in vertebrates. Several noncoding RNAs regulate Ybx1 signaling. However, the role of long noncoding RNAs (lncRNAs) in embryogenesis remains incompletely understood. Here, we investigated the possible involvement of lncRNAs in *Ybx1*-mediated regulation of vertebrate development by performing systematic transcriptome analysis of RNA sequencing data derived from *ybx1* homozygous mutant zebrafish on day 5 (*day5_ybx1^{-/-}*) and wild-type zebrafish on days 5 and 6 (*day5_ybx1^{+/+}* and *day6_ybx1^{+/+}*). We identified several lncRNAs affected by *ybx1* disruption that may target reduction–oxidation-related genes, such as *duox* (NADPH oxidase) and *noxo1a* (NADPH oxidase organizer). Knockdown of three selected lncRNAs led to morphological deformation of larvae, implying an involvement of these lncRNAs in zebrafish embryo development. In summary, our study provides new insights into the lncRNA-mediated mechanisms underlying development in *Ybx1*-deficient zebrafish larvae.

Y-box-binding protein 1 (Ybx1, YB-1), also known as Y-box transcription factor, is a member of a large family of proteins harboring cold shock domain, and it has been proved to have multiple functions involved in a variety of BPs and pathways, including proliferation, differentiation, regulation of apoptosis, translation, stress response, etc. This protein was found in diverse vertebrates and extensively studied in *Homo sapiens* because of its great clinical values. For instance, overexpression of Ybx1 was found in tumor cells and is associated with tumor phenotype [1]. In addition, Ybx1 is demonstrated to harbor the capacity of preventing oncogenic cell transformation via the

phosphatidylinositol 3'-kinase(PI3K)-Akt signaling pathway [2]. In as much as the crucial role of Ybx1 is implicated in tumorigenesis, this protein has been regarded as a well-established biomarker and novel therapeutic target in cancer research [3–6].

In addition to its medicinal benefits, the multifunctional Ybx1 also has been widely used in the study of vertebrate development because it has been proved to be an essential gene to most vertebrate organisms for growth and development. Typically, many studies showed that *ybx1* knockout in mice could result in severe disorders in embryonic development and even death [7,8]. On the one hand, Ybx1 was regarded as a

Abbreviations

BP, biological process; DE, differentially expressed; *duox*, NADPH oxidase; GO, Gene Ontology; lncRNA, long noncoding RNA; *noxo1a*, NADPH oxidase organizer; nt, nucleotide; PPI, protein–protein interaction; REDOX, reduction–oxidation; ROS/RNS, reactive oxygen and reactive nitrogen species; RNA-seq, RNA sequencing; TOM, Topology Overlap Matrix; Ybx1, Y-box-binding protein 1.

crucial protein in early embryogenesis because Ybx1-deficient mice were found to die at a very early stage of embryogenesis [8]. On the other hand, it was reported that *ybx1*^{-/-} embryos exhibit severe growth retardation and progressive mortality after 13.5 embryonic days, which suggests its important role in late-stage embryonic development in mice [7]. Moreover, maternal Ybx1 has been demonstrated to play a safeguard to protect oocyte maturation and maternal-to-zygotic transition in zebrafish [9]. The similar roles of Ybx1 were found in Kumari *et al.*'s study [10]. It was shown that Ybx1 is an essential protein to regulate the maternal control of the Nodal signaling pathway, which has been reported to be essential to the axis formation and germ layer specification in zebrafish [10].

In summary, Ybx1 as a key regulator in embryogenesis and reproduction, as well as development, has already been well demonstrated. However, the underlying mechanism of the processes and pathways mediated by Ybx1 still has much room for deep investigation. For example, Ybx1 exerts multiple functions depending on different subcellular localizations; however, what triggers the translocation of Ybx1 from cytoplasm into nucleus has not been fully understood. Moreover, does noncoding transcriptome participate in Ybx1-mediated regulation in embryogenesis and development?

Noncoding RNA indeed has emerged in recent years as a crucial regulatory layer to participate in and govern diverse BPs and pathways. In particular, long noncoding RNA (lncRNA) becomes a research hotspot in many fields because it has been demonstrated as a multifunctional regulator involved in a range of BPs and pathways and is associated with diverse diseases [11–13]. However, regarding whether lncRNA is important to vertebrate development is always debatable, particularly in embryogenesis. Goudarzi *et al.*'s study [14] knocked down 24 specific lncRNAs in zebrafish using CRISPR/Cas9 to investigate their functions in embryo development. The result is not so satisfying because no lncRNAs are found to have overt functions in zebrafish embryogenesis, viability and fertility. Notably, a recent review by Suresh *et al.* [15] demonstrated that a variety of noncoding RNAs mediate Ybx1 signaling, which provides a new perspective to investigate the role of lncRNA in vertebrate development. Hence Ybx1 might be a good breakthrough to uncover the potential lncRNAs and their possible roles in embryonic development. Our study also might be regarded as complementary research providing clues to support the view that some lncRNAs are important to vertebrate development.

In short, our study used Ybx1 as a springboard to investigate the relationship of lncRNAs with zebrafish

embryonic development, and successfully uncovered several relevant lncRNAs and their possible roles in it. We believe our study provides new insights into the lncRNA-mediated mechanisms underlying development in *Ybx1*-deficient zebrafish larvae.

Materials and methods

Animals

The wild-type zebrafish (*Danio rerio*) AB strain was used in this study. Zebrafish were maintained in the ZebTEC multi-linking rack system (Tecniplast, Buguggiate, Italy) under an artificial photoperiod of 14 h light/10 h dark. The temperature, pH and conductivity of the system were 28 ± 1 °C, 7.5 and 400 $\mu\text{S}\cdot\text{cm}^{-1}$, respectively. The fish were fed twice a day with Otohime fish diet (Marubeni Nisshin Feed, Tokyo, Japan) by the Tritone automatic feeding system (Tecniplast).

All experiments were performed under a license from the Government of Macau Special Administrative Region and approved by Animal Experimentation Ethics Committee of the University of Macau.

Larvae RNA preparation and RNA sequencing

Larvae from wild-type zebrafish of AB strain (day5_*ybx1*^{+/+}, day6_*ybx1*^{+/+}) and *ybx1* homozygous mutants (day5_*ybx1*^{-/-}), which were established by transcription activator-like effector nucleases, were collected for larval RNA preparation. Around 15–20 zebrafish larvae of each type were collected and pooled for a sufficient amount of RNA per sample. A total of nine samples of day5_*ybx1*^{+/+} (three duplicates), day6_*ybx1*^{+/+} (three duplicates) and day5_*ybx1*^{-/-} (three duplicates) were included in RNA sequencing (RNA-seq).

Total RNA was extracted using Tri-Reagent (Molecular Research Center, Cincinnati, OH) according to the protocol of the manufacturer and our previous report [16]. Total RNA was then treated with DNase for 10 min at 37 °C to remove genomic DNA [10 μg RNA in 100 μL reaction buffer with 2 U DNase I from NEB (Ipswich, MA)] followed by phenol-chloroform extraction and ethanol precipitation.

Basic RNA-seq data analysis

Initially, the raw data were processed using Trimmomatic v0.36 (ILLUMINACLIP: TruSeq3-PE.fa:2:30:10:8:true SLIDINGWINDOW:4:15 LEADING:3 TRAILING:3 MINLEN:50) to remove low-quality reads and adaptors. Clean reads were aligned against zebrafish genome (Ensembl release 91) [17] using STAR v020201 [18] with two-pass mode. Then, the alignment results were assembled into transcripts using StringTie v1.3.3b [19] with gene annotation

reference. Finally, the assembled transcripts from each sample were merged into a consensus of final transcripts. The quality of the assembled transcripts was evaluated using REF-EVAL from DETONATE (v1.11) [20].

lncRNAs identification

To identify lncRNAs, we applied a stringent stepwise filtering pipeline, which has been widely used in our previous studies [21,22]. At first, the assembled transcripts were annotated and excluded by aligning against known protein sequences from National Center for Biotechnology Information (NCBI) nr database, UniProt database [23], and the mRNA and protein datasets of zebrafish were derived from the Ensembl database using BLAST (v2.6.0+). This step aims at removing protein-coding sequences as much as possible. Then, the remaining unannotated transcripts with length larger than 200 nucleotides (nt) and longest ORF less than 100 residues were retained. In addition, Coding Potential Calculator [24] was used in the second trimming of protein-coding sequences. Finally, the remaining transcripts were translated (stop-to-stop codon) using an in-house perl script and then subjected to the Pfam database [25] to search for sequences probably encoding protein domains or motifs. The unmatched transcripts were retained as the final high-confidence dataset of lncRNAs.

Differential expression analysis

The assembled transcripts in each sample were quantified by featurecounts (v1.5.3) [26]. Then, the expression profile was normalized using a median of ratios method by R package DESEQ2 [27]. Differential expression analysis was also performed using DESEQ2. Differentially expressed (DE) RNAs were obtained by cutoff of adjusted $P < 0.05$ (Benjamini–Hochberg correction) [27].

Functional prediction of lncRNAs based on network analysis

To investigate the possible functions of lncRNAs implicated in Ybx1 gene knockout, we performed a weighted coexpression network analysis based on the DE lncRNAs and mRNAs using a package WGCNA in R [28]. Before WGCNA analysis, the expression profiles of lncRNAs and mRNAs were normalized using regularized log normalization in DESeq2. To construct the network and detect the modules related to traits (day5_ybx1^{-/-} and day6_ybx1^{+/+}), we used the function *softConnectivity* from WGCNA with the ‘randomly selected genes’ parameter set at 5000 and the power parameter precalculated by pickSoft-Threshold function of WGCNA. Then, the predicted modules of highly correlated RNAs were subjected to Metascape to search for relevant BPs and pathways. The intramodular connectivity

of mRNA and lncRNA in trait-related modules was assessed using TOM (Topology Overlap Matrix) similarity, which was calculated by *TOMsimilarityFromExpr* of WGCNA. TOM similarity reflects the potential interactions between transcripts.

Prediction of lncRNA–protein interactions

The lncRNA–protein interactions were predicted using the *catRAPID* server (*catRAPID* signature: identification of ribonucleoproteins and RNA-binding regions). *catRAPID* (http://s.tartagialab.com/page/catrapid_group) is an algorithm to estimate the binding propensity of protein–RNA pairs, which has been demonstrated as a program to predict protein–RNA associations with great accuracy.

Reverse transcription and real-time quantitative PCR

Reverse transcription was performed at 37 °C for 2 h in a total volume of 10 µL reaction solution containing 3 µg RNA, 0.5 µg oligo (dT), 1X moloney murine leukemia virus reverse transcriptase (MMLV RT) buffer, 0.5 mM each dNTP, 0.1 mM dithiothreitol and 100 U MMLV RT (Invitrogen, Carlsbad, CA, USA). To validate the RNA-seq data, we determined the expression levels of several selected lncRNA transcripts by real-time quantitative PCR in both day5_ybx1^{+/+} and day5_ybx1^{-/-}, including MSTRG12630.1, MSTRG24792.1, ENSDART00000171757, MSTRG.30533.1 and MSTRG.33365.1. To study the correlation between the selected lncRNA and their targets, including *duox* (NADPH oxidase), *nox1a* (NADPH oxidase organizer) and *ybx1*, we determined the mRNA expression levels of targets by real-time quantitative PCR after targeted lncRNA morpholino microinjection. The expression levels were normalized to that of the housekeeping gene *ef1a*. The standard for each gene was prepared by PCR amplification of cDNA fragments with specific primers (Table 1). The real-time quantitative PCR assay was performed on the CFX96 Real-time PCR Detection System (Bio-Rad, Hercules, CA, USA) and repeated twice. All values were expressed as the mean ± standard error of the mean ($n = 3$), and the data were analyzed by *t*-test using Prism 6 on Macintosh OS X (GraphPad Software, San Diego, CA, USA).

Morpholino knockdown and morphological analysis

To investigate the functions of three selected lncRNAs, ENSDART00000171757, MSTRG.30533.1 and MSTRG.33365.1, we synthesized 300 nmol morpholino and 100 nmol standard morpholino oligo by Gene Tools (Gene Tools, Philomath, OR, USA). The morpholino targeted

Table 1. Sequence information of primers for real-time quantitative PCR.

Gene	Primer ID	Strand	Sequence (5'–3')	Size (bp)
MSTRG.12630.1	4975	Sense	TAATATAGTCGCGCGCACAG	94
MSTRG.12630.1	4976	Antisense	TAGAGTCCCCCGTTGTTTTG	94
MSTRG.24792.1	4977	Sense	CAGCAACAAGTCACACTCAACA	143
MSTRG.24792.1	4978	Antisense	CTCTAAACGAGCAAGGTCCAA	143
ENSDART00000171757	4979	Sense	ACGAGGTGTCCTCTCACCAG	134
ENSDART00000171757	4980	Antisense	CCAGGATGGTCCCTTGATAGC	134
MSTRG.30533.1	4981	Sense	ACCAAATCAAGCAAGCGTTC	112
MSTRG.30533.1	4982	Antisense	TGTAGGACCTCCGCTTTCAC	112
MSTRG.33365.1	5085	Sense	CCTCTGTTGAACCGGAAAAG	107
MSTRG.33365.1	5086	Antisense	TTTGGAGTTACCTGAAAGGATG	107
duox	4963	Sense	CATGGATGGGGTTTATCTGG	109
duox	4964	Antisense	TCTGTTCTTGTGGGACAGCA	109
noxo1a	5573	Sense	ACACGCAAAAACATGCAGAG	131
noxo1a	5574	Antisense	CCGTCTGTCCATTAGCCATT	131
ybx1	1919	Sense	TCTGATTGCCTTTGTTTCTGTGTC	110
ybx1	1920	Antisense	CCCTTCTACCACGTGCAACT	110

sequences for synthesis regarding the earlier lncRNA were represented in Table 2. For morpholino microinjection, three different doses (5, 10 and 20 ng) of morpholino were injected into one cell stage embryos for targeted lncRNA, respectively, within 30 min just after spawning. Equal amounts (5, 10 and 20 ng) of standard morpholino oligo were injected as control. On day 5, the embryo number was counted both with normal and deformed morphology individually for different morpholinos and different dose injections. The morphology of morpholino-injected larvae were photographed by Nikon SMZ18 microscope (Nikon, Tokyo, Japan).

Immunoblotting

To further study the correlation between lncRNA ENSDART00000171757 and *ybx1* expression at protein level, we used western blotting for evaluation of Ybx1 protein expression between morpholino-injected (ENSDART00000171757) and control morpholino-injected groups. The larvae on day 5 with normal morphology were collected for protein sample preparation. The Ybx1 antibodies are from GenScript in Nanjing, China.

In brief, larvae were lysed with 100 μ L 1 \times SDS sample buffer (62.5 mM Tris-HCl, 1% w/v SDS, 10% glycerol, 5% mercaptoethanol, pH 6.8), and the lysates were heated at

95 $^{\circ}$ C for 10 min and then centrifuged 17 000 *g* for 15 min at 4 $^{\circ}$ C, followed by protein concentration measurement. The samples (total 100 μ g proteins) from ENSDART00000171757_morpholino (MO) and control MO were separated on 12% polyacrylamide gels and transferred to poly(vinylidene difluoride) membranes. The membranes were blocked with 5% nonfat milk in 1 \times TBST for 1 h at room temperature. After washing three times with 1 \times TBST, the membranes were incubated with Ybx1 primary antibody (1 : 5000) and beta-actin primary antibody (1 : 2000) in 5 mL of 1 \times TBST with 5% nonfat milk overnight at 4 $^{\circ}$ C. They were washed with 1 \times TBST three times, followed by incubation with the horseradish peroxidase-conjugated secondary antibody (1 : 2000) in 5 mL 1 \times TBST for 1 h at room temperature. After washing, the membranes were incubated with the SuperSignal West Femo Maximum Sensitivity Substrate (Thermo Scientific, Waltham, MA, USA), and signals were detected on the ChemiDoc MP imaging system (Bio-Rad, Hercules, CA, USA).

Whole-mount *in situ* hybridization

To further study the spatial expression pattern for the three target lncRNAs, ENSDART00000171757, MSTRG.30533.1 and MSTRG.33365.1, whole-mount RNA *in situ* hybridization was undertaken for investigation. All their lncRNA

Table 2. Morpholino information for three lncRNAs.

Oligo name	Quantity	Production number	Sequence (5'–3')
MSTRG.30533.1	300 nmol	04–13Aug18A	GCTTGATTTGGTTGATTCCTAACTG
MSTRG.33365.1	300 nmol	05–13Aug18A	TGGAGTCCACTGACACTGACTGATC
ENSDART00000171757	300 nmol	06–13Aug18A	ATACTGCATTCCAACCTGAATCAG

probes were transcribed *in vitro* and for storage at -80°C . Embryos were under 0.003% 1-phenyl-2-thiourea treatment to remove pigment until the larvae reached day 5. On day 5, the *ybx1^{+/+}* and *ybx1^{-/-}* larvae were fixed in fresh 4% PFA overnight at 4°C . The larvae were washed with PBST for 5 min at room temperature, and this was repeated four times. For permeabilization, the larvae were immersed with $10\ \mu\text{g}\cdot\text{mL}^{-1}$ Protease K in PBST for 15 min at room temperature. The larvae were refixed in 4% PFA for 20 min at room temperature, followed by PBST rinsing for 5 min that was repeated four times. Before mRNA probe hybridization, larvae were immersed in hybridization mix for 2 h at 65°C . We diluted 100 ng labeled RNA probe in hybridization mix. We removed prehybridization solution and added prewarmed hybridization mix plus probe to the larvae. The hybridization was kept for 40 h at 65°C . For posthybridization washes, we washed for 5 min in 66% hybridization mix, 33% $2\times$ sodium chloride/sodium citrate (SSC) at 65°C ; washed 5 min in 33% hybridization mix, 66% $2\times$ SSC at 65°C ; and washed 5 min in $2\times$ SSC at 65°C . We washed $1\times$ 20 min in $0.2\times$ SSC + 0.1% Tween 20 at 65°C and then washed $2\times$ 20 min in $0.1\times$ SSC + 0.1% Tween 20 at 65°C . At room temperature, we washed for 5 min in 66% $0.2\times$ SSC, 33% PBST; washed for 5 min in 33% $0.2\times$ SSC, 66% PBST; and washed for 5 min in PBST. After washing, we incubated larvae in blocking solution (PBST plus 2% sheep serum, $2\ \text{mg}\cdot\text{mL}^{-1}$ BSA) for 1 h at room temperature and prepared the first antibody (alkaline-phosphatase-conjugated anti-digoxigenin) by diluting it in blocking solution, 1:5000. We incubated in the antibody for shaking overnight at 4°C and washed $5\times$ 15 min in PBST. For colorization, we washed $4\times$ 5 min in coloration buffer. We mixed 200 μL Nitro Blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate stock mixture with 10 mL coloration buffer. We added 500 μL of this mix to larvae and incubated in the dark at room temperature until a blue reaction product was visible. We stopped the reaction by quickly washing larvae three times by stop solution (PBST, pH 5.5), then washed twice in stop solution for 15 min. After glycerol mounting, we photographed immediately by Nikon SMZ18 microscope (Nikon, Tokyo, Japan).

Results

Identification and characterization of lncRNAs

To detect lncRNAs in zebrafish larvae, we conducted a basic bioinformatics analysis. In brief, the RNA-seq yielded an average of 51 879 984 raw pair-end reads with length of 150 nt per sample (Table S1). After trimming adaptor sequences and low-quality reads, the remaining average of 49 379 818 clean reads per sample was aligned to zebrafish genome using STAR and assembled into 77 252 transcripts using Stringtie

(Table S2). Then all the assembled transcripts were subjected to quality assessment using DETONATE. The results showed that the majority of transcripts have high quality (Table S3). In addition, the BLAST search against known zebrafish transcripts in the Ensembl database indicated that 61 926 transcripts ($\sim 80.16\%$) have been well recorded in the Ensembl database. Notably, the remaining 16 316 transcripts for which no hits were found in the Ensembl database were probably novel transcripts representing a new subset of functional mRNAs or lncRNAs.

Based on the well-established zebrafish transcriptome, a stringent stepwise filtering pipeline was used to identify a high-confidence lncRNA dataset in this study (Fig. 1). Two core filtering criteria were applied to screen these lncRNAs according to the general definition of lncRNA: (a) the capability of protein coding, and (b) the length of transcripts. In brief, 73 018 potential protein-coding transcripts obtained by BLAST search against nr, Swiss-Prot database, as well as the known zebrafish mRNAs/proteins from the Ensembl database, were excluded. The remaining 1555 unannotated transcripts shorter than 200 nt in length and larger than 100 residues in the longest ORF were filtered out. A second protein-coding filtering was performed to further remove the transcripts with potential protein-coding capacity, including those containing functional domains/motifs predicted by Pfam scanning and those with protein-coding potential classified by Coding Potential Calculator. This step filtered out only one transcript. A total of 2678 lncRNAs were finally identified for further study.

We compared those identified lncRNAs with known zebrafish lncRNAs derived from the NONCODE database (v5.0) and ZFLNC database, using BLASTn (e-value was set to 0.1). The results indicate that a total of 1083 lncRNAs ($\sim 40.44\%$) identified in this study could be well compared with known zebrafish lncRNAs, 460 lncRNAs of which have been included in the NONCODE and ZFLNC database. The remaining mapped lncRNAs exhibit high similarity to the known zebrafish lncRNAs in some specific regions. It is not surprising because lncRNAs have been proved to be quite distinct from mRNAs as mRNAs have a more conserved mutational rate. Instead, lncRNAs have less evolutionary constraints, except the selective pressure to strictly conserve the short functional region, e.g., sequence-specific interactions, protein-binding sites, etc. The rest of the lncRNAs that could not be found in any hit in NONCODE and ZFLNC databases might represent a group of novel zebrafish lncRNAs.

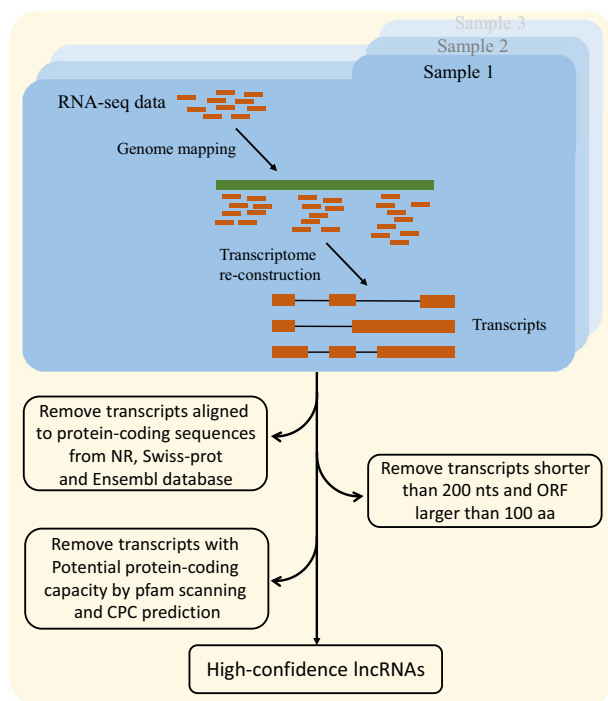


Fig. 1. The bioinformatics pipeline used for identification of zebrafish lncRNAs in this study. aa, amino acids; CPC, Coding Potential Calculator.

The expression pattern of the DE lncRNAs implicated in Ybx1 gene knockout

The expression profile for each sample was initially assessed using featurecounts. To identify the lncRNAs associated with Ybx1 gene knockout, differential expression analysis between day5_ *ybx1*^{+/+} and day5_ *ybx1*^{-/-} was conducted using DESeq2. The result shows that 44 lncRNAs (22 are up-regulated and 22 are down-regulated) and 1901 mRNAs (849 are up-regulated and 1052 are down-regulated) are detected as significantly DE (Figs S1 and S2). Despite no change of the RNA expression level of Ybx1 as observed in day6_ *ybx1*^{+/+} compared with that of day5_ *ybx1*^{+/+}, we found that the protein expression level of Ybx1 was dramatically decreased in day6_ *ybx1*^{+/+} versus day5_ *ybx1*^{+/+} (data not shown). Hence we also compared the expression level of transcripts between day5_ *ybx1*^{+/+} and day6_ *ybx1*^{+/+} to explore

the underlying mechanisms resulting in decline of Ybx1 in day6_ *ybx1*^{+/+} at protein level. DE analysis showed that a total of 542 RNAs were DE in day5_ *ybx1*^{+/+} versus day6_ *ybx1*^{+/+}, including 2 DE lncRNAs and 540 DE mRNAs (Figs S3 and S4). Then both results of DE analysis were subjected to the DAVID server to search for the possible functions of those DE RNAs. Intriguingly, both functional results indicated that the hormone synthesis-related and oxidative stress-related processes were commonly enriched with an extremely high level of statistical significance, including steroid biosynthesis process, oxidation-reduction process, etc. (Fig. 2A,B). This result disclosed a close association of Ybx1 with the processes of hormone synthesis and oxidative stress response in zebrafish.

Also, the difference of RNA changes and the corresponding roles caused by two distinct mechanisms resulting in down-regulation of Ybx1 is another focus of attention for us to investigate. Therefore, we conducted a systematic comparison of day5_ *ybx1*^{+/+} versus day5_ *ybx1*^{-/-} and day5_ *ybx1*^{+/+} versus day6_ *ybx1*^{+/+} (Fig. 2C). Concretely, 143 transcripts were found to be consistently up/down-regulated in both comparisons. Remarkably, most of them are up-regulated, and only five transcripts were observed as down-regulated. Gene Ontology (GO) functional enrichment analysis for those 143 DE transcripts showed that many BP categories involved in hormone-related and oxidative stress-related processes were significantly enriched, such as steroid biosynthesis, oxidation-reduction process, etc. (Fig. 2D). Furthermore, we found that 731 transcripts (including 22 lncRNAs) were specifically up-regulated in day5_ *ybx1*^{+/+} versus day5_ *ybx1*^{-/-} but exhibit no changes in day5_ *ybx1*^{+/+} versus day6_ *ybx1*^{+/+}. Subsequent functional analysis on DAVID indicated that the functions of these specific up-regulated RNAs involved reproduction processes, that is, egg coat formation, binding of sperm to zona pellucida, etc. Notably, the oxidation-reduction process was also significantly enriched (Fig. 2D). A total of 1062 transcripts (including 22 lncRNAs) were found to be specifically down-regulated in day5_ *ybx1*^{+/+} versus day5_ *ybx1*^{-/-}, which involve protein transport, glycolytic process, gluconeogenesis, etc.

Fig. 2. Differential expression analysis between day5_ *ybx1*^{+/+} and day5_ *ybx1*^{-/-} and day6_ *ybx1*^{+/+}. GO functional enrichment analysis of DE transcripts achieved on David server: (A) day5_ *ybx1*^{+/+} versus day5_ *ybx1*^{-/-} and (B) day5_ *ybx1*^{+/+} versus day6_ *ybx1*^{+/+}. The results are summarized only in the category of BP. The y axis indicates functional groups. The x axis indicates $-\log_{10}(P\text{-value})$. (C) Heatmap of all DE transcripts that were clustered into five subsets, including specific down-regulated in day5_ *ybx1*^{-/-}, specific up-regulated in day5_ *ybx1*^{-/-}, specific down-regulated in day6_ *ybx1*^{+/+}, specific up-regulated in day6_ *ybx1*^{+/+} and consistently dysregulated in day5_ *ybx1*^{-/-} and day6_ *ybx1*^{+/+}. The expression profile was normalized using DESeq2 and then hierarchically clustered using heatmap2. Values have been centered and scaled for each row. Each row represents a single transcript. (D) Functional analysis of subsets using DAVID (FDR < 5%).

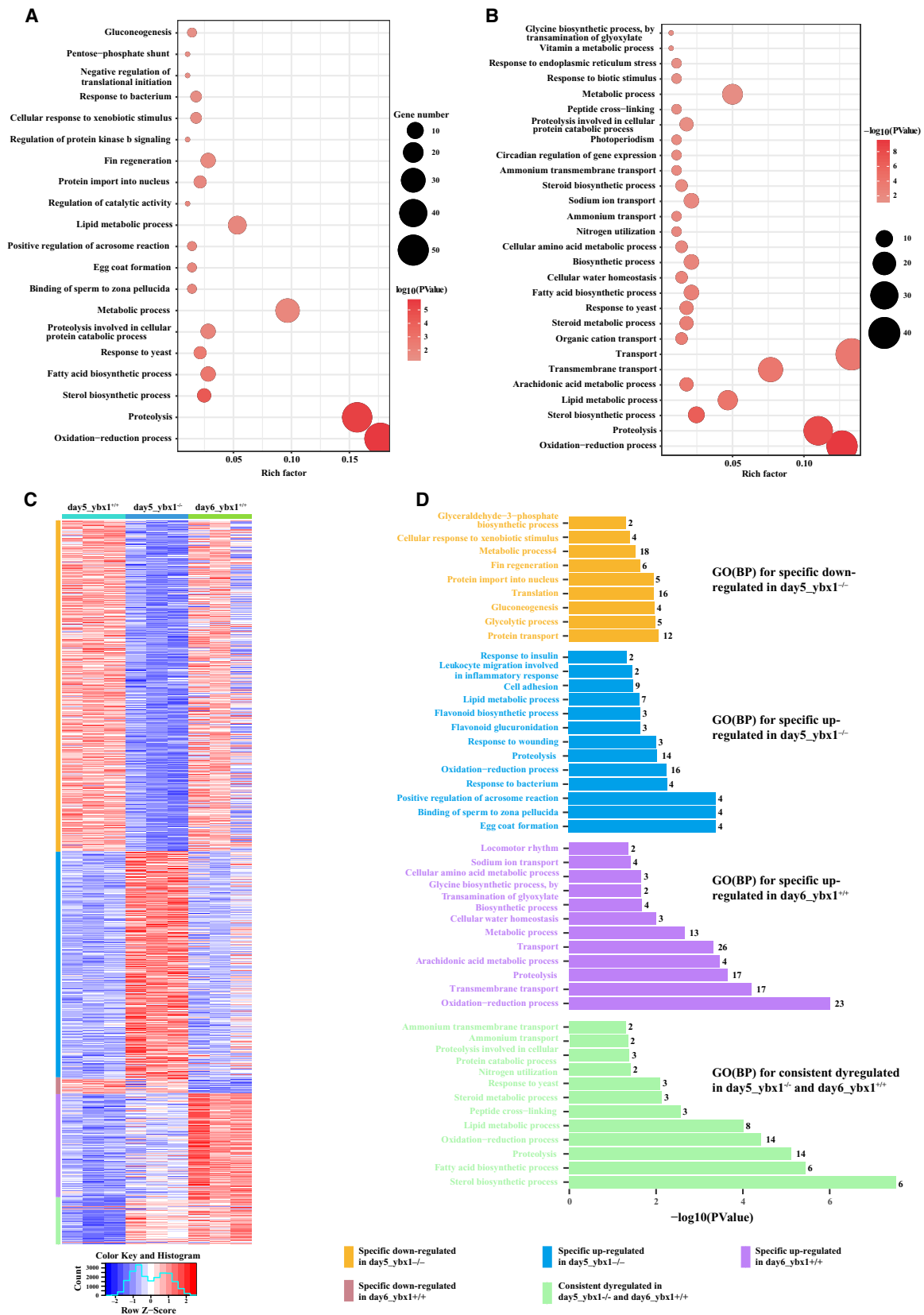


Table 3. Basic statistics of DE transcripts in two different comparisons. 5D WT, Wide Type in 5 Days; 5D MT, Mutant Type in 5 Days; 6D WT, Wide Type in 6 Days.

5D WT versus 5D MT	5D WT versus 6D WT	Number (mRNAs/ lncRNAs)
Up	Up	138/0
Down	Down	5/0
Down	Up	7/0
Up	Down	2/0
Up	Non	709/22
Down	Non	1040/22
Non	Up	340/2
Non	Down	48/0

(Fig. 2D). Similarly, we have also detected many transcripts DE in *day5_ybx1^{+/+}* versus *day6_ybx1^{+/+}*, but not present in *day5_ybx1^{+/+}* versus *day5_ybx1^{-/-}* (Table 3). Concretely, 342 transcripts (including two lncRNAs) exhibit up-regulation in *day6_ybx1^{+/+}*, and functional analysis showed that the functions of those DE RNAs might involve the oxidation–reduction process, transmembrane transport, proteolysis, etc. (Fig. 2D). Forty-eight specific down-regulated RNAs were found in *day6_ybx1^{+/+}* compared with *day5_ybx1^{+/+}*, and their corresponding functional analysis showed that no BP categories were significantly enriched. These functional analyses based on the different subsets of DE transcripts indicated that reduced expression of *ybx1* could trigger up-regulation of oxidative stress-related BPs and pathways (e.g., the oxidation–reduction process was significantly enriched by specific up-regulated transcripts in *day5_ybx1^{-/-}* and *day6_ybx1^{+/+}*). These findings suggest that reduced expression of *ybx1*, either caused by *Ybx1* knockout or by innate autoregulation, exhibits strong correlation with the reduction–oxidation (REDOX) regulation in zebrafish. Notably, many DE lncRNAs are included in those DE RNA datasets (the majority of them are found in the comparison of *day5_ybx1^{+/+}* versus *day5_ybx1^{-/-}*), implying specific lncRNAs are present in REDOX regulation after *Ybx1* gene knockout.

Network analysis reveals lncRNAs implicated in REDOX regulation

To investigate the lncRNAs, as well as their possible roles, implicated in *Ybx1* gene knockout, we

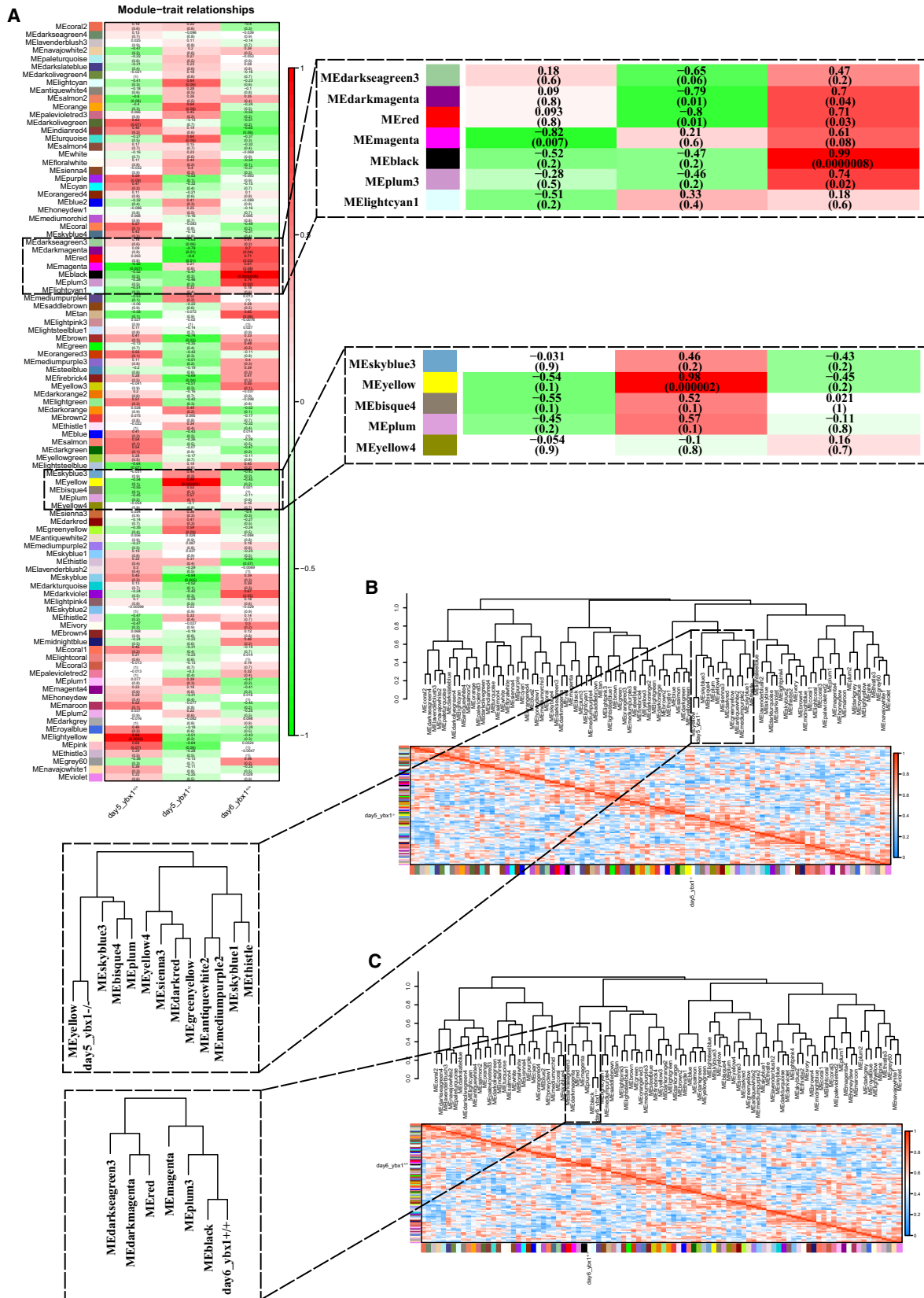
Table 4. The number of mRNAs and lncRNAs in the top 10 modules.

Module	Size	lncRNA	mRNA
Turquoise	4731	318	4413
Blue	4284	328	3956
Brown	2752	382	2370
Yellow	2746	352	2394
Green	2267	355	1912
Red	1563	176	1387
Black	1230	155	1075
Pink	1206	207	999
Magenta	1109	98	1011
Purple	970	201	769

conducted a systematic coexpression network analysis using WGCNA in R. Initially, the flashClust tools package was used to conduct cluster analysis on these samples to detect outliers; the results showed that all samples are in the clusters and passed the cuts (Fig. S5). Then a network-topology analysis was conducted for several soft-thresholding powers to have relative balance scale independence and mean connectivity of the WGCNA. The lowest power for the scale-free topology fit index was selected for construction of a hierarchical clustering tree (Fig. S6). This analysis colocalizes both correlated lncRNAs and mRNAs into 95 modules (Fig. S7). Each module contains independent datasets of transcripts (Table 4). The interactions among those modules were visualized in Fig. S8. The modules with common expression patterns that were significantly associated with specific traits (*day5_ybx1^{-/-}* and *day6_ybx1^{+/+}*) were detected based on the correlation between module eigengene and traits. Then the function plotEigengeneNetworks of WGCNA was used to identify groups of correlated eigengenes. The results indicated that two modules (yellow and black) were significantly correlated to the traits: *day5_ybx1^{-/-}* and *day6_ybx1^{+/+}*, respectively (Fig. 3). The relationship of transcript significance with module membership was visualized in Fig. S9. The yellow modules significantly related to *day5_ybx1^{-/-}* include 2394 mRNAs and 352 lncRNAs, whereas black modules correlated to *day6_ybx1^{+/+}* contain 1075 mRNAs and 155 lncRNAs (Table 4).

GO enrichment analyses based on the DAVID server were subsequently performed using the annotated transcripts (mRNAs) derived from two trait-related

Fig. 3. Detection of two trait-related modules. (A) Module–trait associations. Each row corresponds to a module eigengene, and each column to a trait. Each cell displays the corresponding correlation and *P*-value. (B) The eigengene dendrogram and heatmap identify that the yellow module is highly related to *day5_ybx1^{-/-}*. (C) The eigengene dendrogram and heatmap identify that the black module is highly related to *day6_ybx1^{+/+}*.



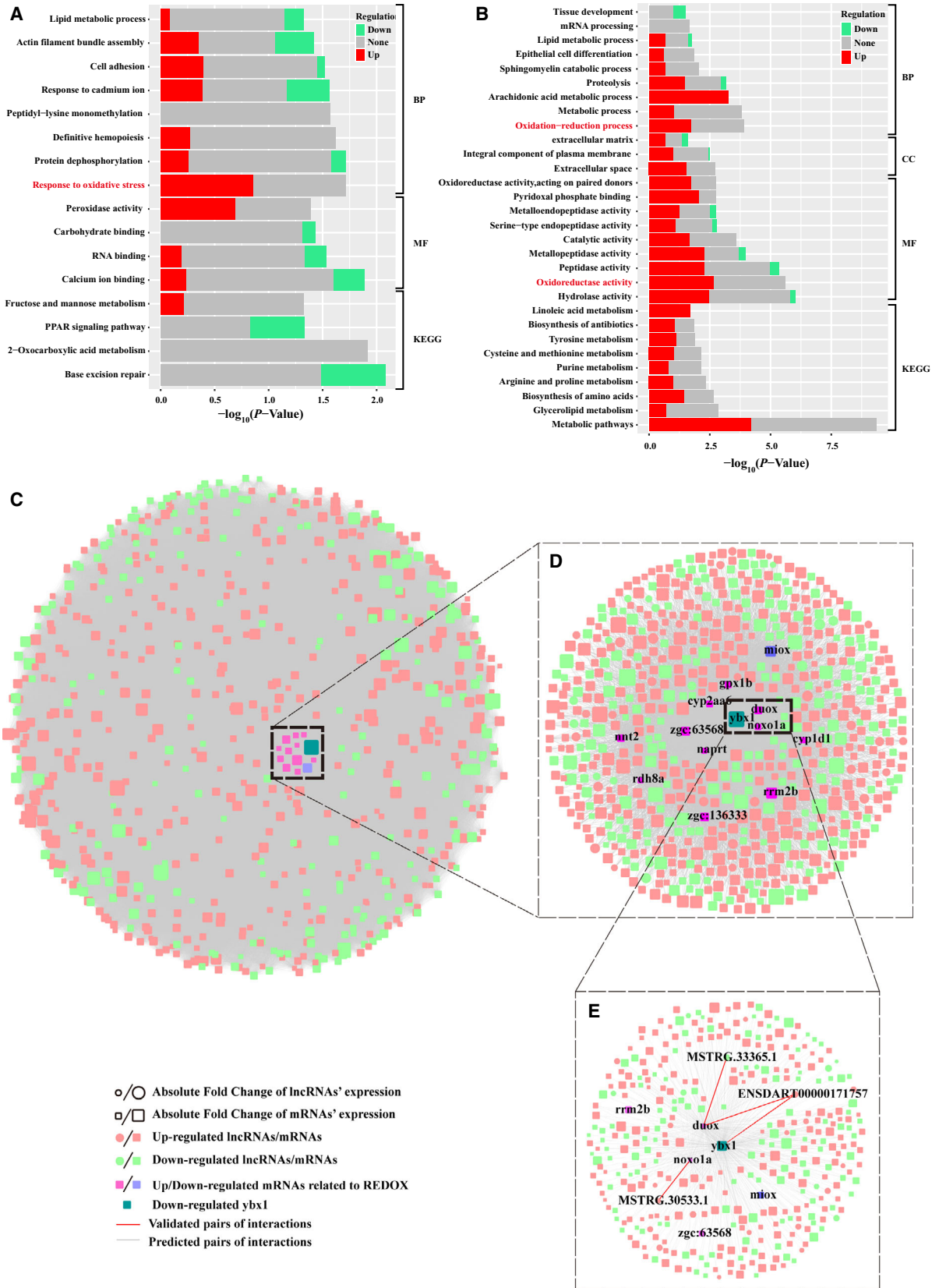


Fig. 4. Coexpression network analysis defined REDOX-related interactome in day5_ *ybx1*^{-/-}-related module. GO functional enrichment analysis of two trait-related modules: (A) day5_ *ybx1*^{-/-}-related module and (B) day6_ *ybx1*^{+/+}-related module. The colored bar reports the fraction of up-regulated and down-regulated transcripts present in that functional category. All the adjusted statistically significant values of the terms were negative 10-base log transformed. KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, Biological Process; MF, Molecular Function; CC, Cellular Component. (C) Cytoscape map of the interaction (including mRNA–mRNA, mRNA–lncRNA and lncRNA–lncRNA) network defined by DE lncRNAs and mRNAs using the function *TOMsimilarityFromExpr* of *WGCNA* package. (D) A zoomed view of a core network displaying interactions (including mRNA–mRNA and mRNA–lncRNA) involved in REDOX-related genes and *YBX1*. (E) Zoomed in subnetwork view of interactions involved in *ybx1* and two key REDOX-related genes: *duox*, *nox1a*.

modules. Unsurprisingly, the results indicated that many REDOX processes were significantly enriched by up-regulated transcripts in two trait-related modules with considerably high significance (highlighted in red text in Fig. 4A,B), including GO: 0006979 (response to oxidative stress), GO: 0016491 (oxidoreductase activity) and GO: 0055114 (oxidation–reduction process), etc. This finding is consistent with the results of functional enrichment analysis using DE RNAs described in the previous section (Fig. 2D), which further suggests that the decline of *Ybx1* can enhance up-regulation of REDOX. Indeed, many studies [29–31] have illustrated the influence of oxidative stress to embryonic development in zebrafish, e.g., oxidative stress condition could change the cellular REDOX state, which further causes oxidation of molecules, such as membrane peroxidation, loss of ions, protein cleavage and DNA strand breakages, and finally leads to cell death.

Next, to explore possible roles of lncRNAs in REDOX regulation triggered by decreased *ybx1*, we used the expression profile of DE lncRNAs and DE mRNAs included in the yellow module related to day5_ *ybx1*^{-/-} to assess the intramodular connectivity (including mRNA–mRNA and mRNA–lncRNA interactions) using TOM similarity, which was done by the function *TOMsimilarityFromExpr* of *WGCNA* package. We did not further investigate the lncRNAs implicated in REDOX in black module related to day6_ *ybx1*^{+/+} because our previous DE analysis showed that just two DE lncRNAs were found in the comparison of day5_ *ybx1*^{+/+} versus day6_ *ybx1*^{+/+}. Regarding the yellow module, we detected a great number of interactions that involved 631 DE mRNAs and 45 DE lncRNAs (Fig. 4C). Among them, several lncRNAs, as well as mRNAs, were found to interact with two REDOX-related mRNAs, that is, *duox* and *nox1a* (Fig. 4D). These two genes have been demonstrated to play important roles in response to oxidative stress [32,33]. Particularly, we observed that several lncRNAs not only exhibit strong correlations to *ybx1* but can also interact with two REDOX-related mRNAs: *Duox* and *Nox1a* (Fig. 4E).

Correlations of lncRNA ENSDART00000171757 with embryonic development

Based on the previous network analysis, several lncRNAs associated with embryonic development, as well as the REDOX process, have been figured out. However, the underlying mechanism is yet unclear. For this purpose, a large-scale screening of lncRNA–protein interactions was conducted based on the *Ybx1*-related, REDOX-related lncRNAs (displayed in Fig. 4E) via the *catRAPID* server. We observed that the lncRNA ENSDART00000171757 exhibited moderate interactions with several regulatory proteins (Table S4), including transcription factors, RNA-binding protein, etc. Subsequent functional analysis achieved by Metascape indicated that those proteins significantly enriched in diverse development processes (Fig. 5B), including regionalization, endocrine system development, central nervous system development, embryonic morphogenesis, etc. In addition, through screening their possible targets via a PPI (protein–protein interaction) database STRING, we found that those proteins cooperated with *Ybx1*, and two key REDOX-related proteins, *duox* and *nox1a*, might participate in a complex regulatory network involved in many developmental processes (Fig. 5A). Notably, *nfk* (ranked among top 3 proteins that interacted with lncRNA ENSDART00000171757 in *catRAPID*) was found to interact with many ribosomal proteins and might act as a bridge to connect *duox*, *nox1a* and *ybx1* in the PPI network (Fig. 5B). Furthermore, it has been shown to play an essential role in early zebrafish embryonic development (insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development). Our analysis indicated that knockout of *ybx1* in zebrafish might affect the expression of p53 and further disrupt the balance of the whole regulatory network and the expression of many key proteins involved in development, as well as lncRNA ENSDART00000171757, which might lead to overexpression of the REDOX-related genes, that is, *duox*. Overexpression of REDOX-related genes might further generate too much reactive oxygen and reactive

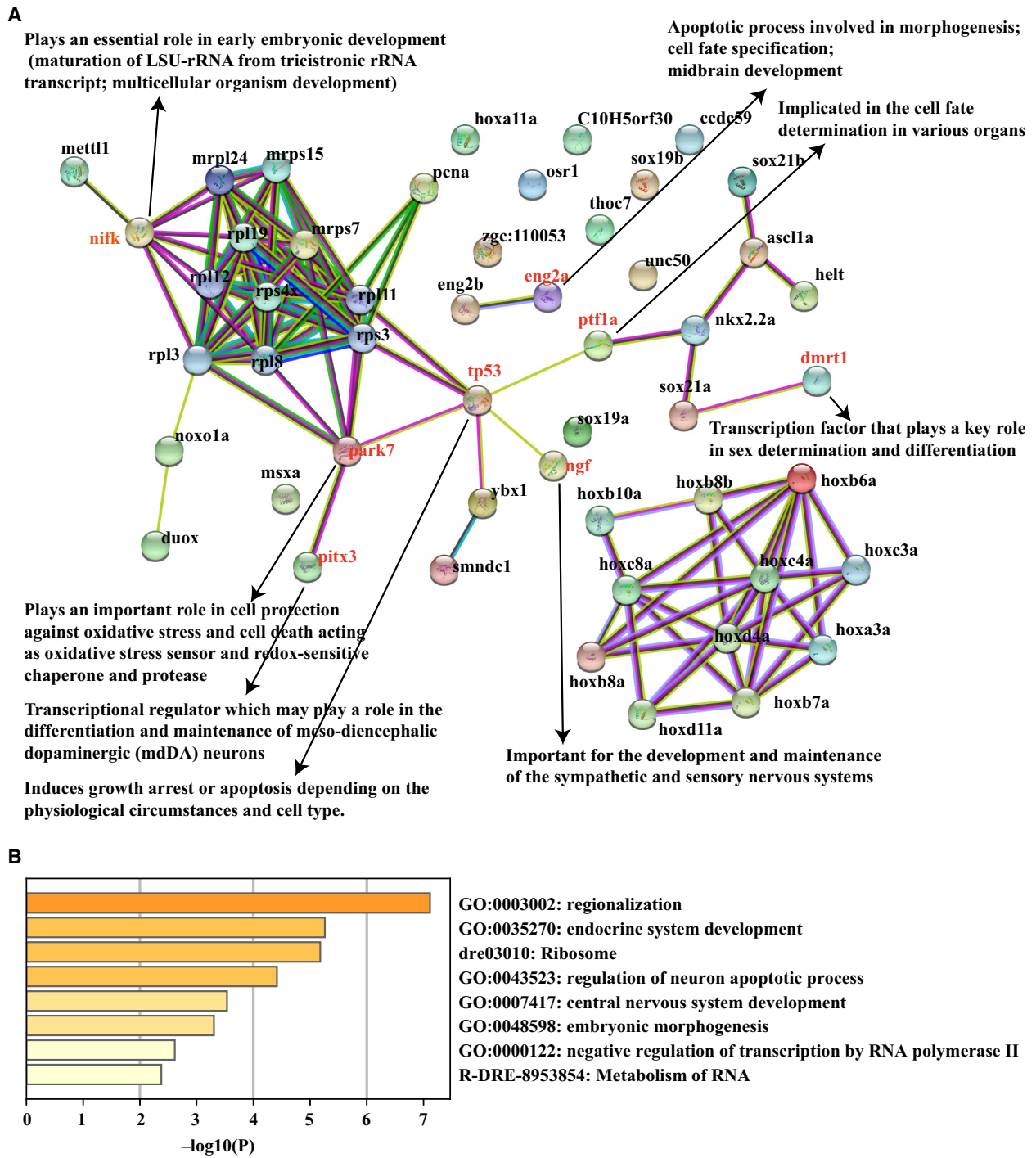


Fig. 5. Prediction of lncRNA–protein interactions reveals close association of lncRNA ENSDART00000171757 with zebrafish embryonic development. (A) The lncRNA–protein interaction network derived from the STRING database. LSU-rRNA, Large SubUnit (LSU) ribosomal RNA (rRNA). (B) Functional enrichment analysis of the possible interacted proteins in Metascape server.

nitrogen species (ROS/RNS), which result in oxidative stress and cause either cell death or senescence. These processes finally affect the zebrafish embryonic development.

Bioassay validations

To test and confirm our findings, we subjected the lncRNA ENSDART00000171757 to an experimental

validation study. Meanwhile, another two novel lncRNAs, MSTRG.30533.1 and MSTRG.33365.1, were selected for the same experiments, which also exhibit relatively significant differential expression in day5_ybx1^{+/+} compared with day5_ybx1^{-/-}, and simultaneously have strong correlations with *Duox* and *Noxola*, as well as *Ybx1*, in the network (Fig. 4).

Real-time quantitative PCR validated five DE lncRNAs

For the real-time quantitative PCR validation, we selected five lncRNAs, including three earlier mentioned lncRNAs, as well as another two lncRNAs, MSTRG.12630.1 and MSTRG.24792.1, which have relatively high fold change in differential analysis between day5_ybx1^{+/+} and day5_ybx1^{-/-} (Fig. S10). The expressions of MSTRG.12630.1, MSTRG.24792.1, ENSDART00000171757 and MSTRG.30533.1 were all apparently increased in day5_ybx1^{-/-} larvae (Fig. S10A–D), and MSTRG33365.1 presented a sharp decrease in day5_ybx1^{-/-} larvae (Fig. S10E). All the real-time quantitative PCR results were consistent with the differential expression analysis.

Three REDOX-related lncRNAs tend to be expressed in the gut

We next tested whether the expression of three lncRNAs (ENSDART00000171757, MSTRG30533.1 and MSTRG.33365.1) has tissue specificity by whole-mount *in situ* hybridization assay. Interestingly, we observed that all three lncRNAs were predominantly expressed in the gut region (Fig. 6A). Moreover, the expression level of those lncRNAs in day5_ybx1^{+/+} and day5_ybx1^{-/-} larvae was consistent with the real-time quantitative PCR results (Fig. 6B–D).

lncRNA knockdown leads to larvae morphological deformation

Our network and PPI analysis revealed a lncRNA ENSDART00000171757 might be associated to zebrafish embryonic development. To test it, we targeted the lncRNA ENSDART00000171757, as well as other novel lncRNAs MSTRG.30533.1 and MSTRG.33365.1 (also exhibit strong correlations with *Duox* and *Noxola*, as well as *Ybx1*, in the network analysis), for morpholino knockdown, with their specific target sequence for morpholino synthesis. After serial morpholino injection (5, 10 and 20 ng), partial larvae on day 5 with severe morphological deformation were photographed (Fig. 6E). As a result, a large proportion of deformed larvae was

observed for each lncRNA's morpholino, but only a few deformed larvae by control morpholino injection. In particular, the rate of larval deformation after morpholino injection was dose dependent for all three lncRNAs (Fig. 6H). Our experiments showed that the knockdown of each of the three selected lncRNAs leads to larvae morphological deformation in a dose-dependent manner.

Correlation of lncRNAs with *duox*, *noxola* and *ybx1*

Our network analysis indicated that the lncRNAs ENSDART00000171757, MSTRG.30533.1 and MSTRG.33365.1 exhibited strong correlations to the REDOX-related genes *duox* and *noxola*. To validate it, we divided larvae (day 5) injected with 20 ng morpholino for the three lncRNAs into two groups, normal morphology and deformed morphology. We observed that knockdown of ENSDART00000171757 slightly reduced the mRNA expression of *duox* in both the normal and the deformed groups (Fig. 6F) and remarkably reduced the mRNA expression of *ybx1* in the normal group (Fig. 6G). MSTRG.30533.1 knockdown apparently activated *noxola* mRNA expression in the normal group (Fig. 6I), whereas inhibition of MSTRG.33365.1 led to a sharp decrease of *duox* mRNA expression in the deformed group (Fig. 6J). Our experiments demonstrated that both lncRNAs ENSDART00000171757 and MSTRG.33365.1 might positively regulate the expression of *duox*, whereas MSTRG30533.1 probably exhibits negative regulation to *noxola*. lncRNA ENSDART00000171757 also exhibited positive regulation to *ybx1*.

Also, we tested the potential interaction between ENSDART00000171757 and *Ybx1* in protein level. Immunoblotting was performed to evaluate *Ybx1* protein expression level under ENSDART00000171757 morpholino knockdown, and beta-actin was used as an internal reference. The results showed that the expression of *Ybx1* protein was evidently reduced after ENSDART00000171757 knockdown (Fig. S11A). It was suggested that inhibition of ENSDART00000171757 would result in significant reduction (approximately 50%) of the relative intensity of *Ybx1*/Actin in day 5 larvae (Fig. S11B).

Discussion

Given that lncRNAs have been proposed to carry out diverse functions, for instance, regulation of gene expression by acting as signaling, guiding, sequestering or scaffolding molecules, an increasing number of

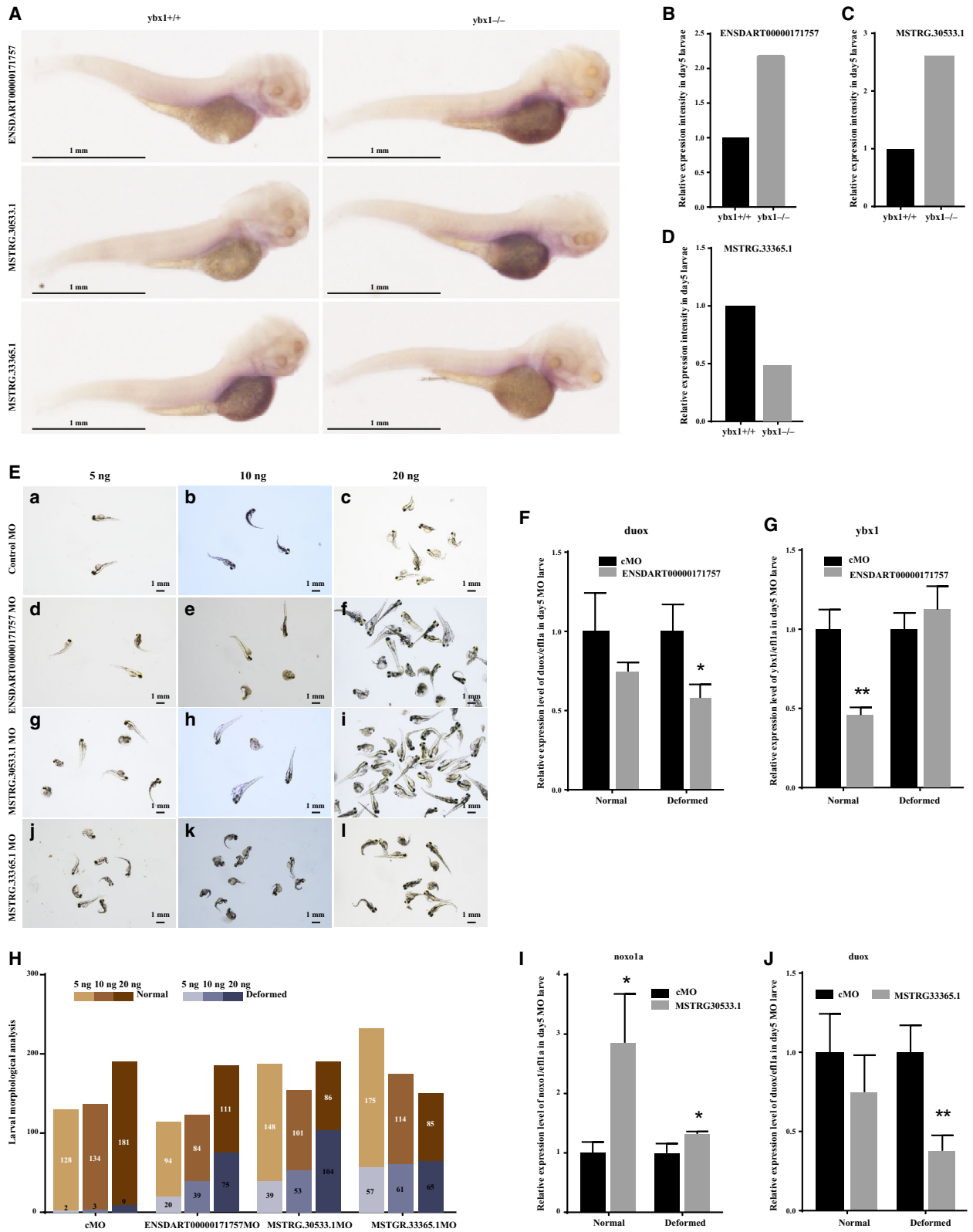


Fig. 6. lncRNA knockdown led to larvae morphological deformation. (A) Whole-mount *in situ* hybridization analysis of the three selected lncRNAs (ENSDART00000171757, MSTRG.30533.1 and MSTRG.33365.1) in day 5 *ybx1^{+/+}* and *ybx1^{-/-}* larvae. Top: the expression of ENSDART00000171757 lncRNA in day 5 *ybx1^{+/+}* and *ybx1^{-/-}* larvae. Middle: the expression of MSTRG.30533.1 lncRNA in day 5 *ybx1^{+/+}* and *ybx1^{-/-}* larvae. Bottom: the expression of lncRNA MSTRG.33365.1 in day 5 *ybx1^{+/+}* and *ybx1^{-/-}* larvae. Scale bars: 1 mm. (B–D) Relative expression intensity of ENSDART00000171757, MSTRG30533.1 and MSTRG33365.1 in day 5 *ybx1^{+/+}* and *ybx1^{-/-}* larvae. (E) Larval morphological analysis after lncRNA morpholino injection on day 5. (a–l) Photographs of deformed larvae injected with different doses (5, 10 and 20 ng) of morpholino for lncRNA (ENSDART00000171757, MSTRG.30533.1 and MSTRG.33365.1), and control morpholino injected larvae were used as reference. Scale bars: 1 mm. (F,G) Relative expression of *duox* mRNA and *ybx1* mRNA in day 5 larvae after injection of 20 ng ENSDART00000171757 morpholino. (H) Larval morphological statistics for different lncRNA morpholino larvae injected with different doses (5, 10 and 20 ng); control morpholino injection was used for reference. Column in brown represents larvae with normal morphology, and purplish blue column stands for deformed morphological larvae. (I) Relative expression of *noxo1a* mRNA in day 5 larvae after injection of 20 ng MSTRG.30533.1 morpholino. (J) Relative expression of *duox* mRNA in day 5 larvae after 20 ng MSTRG.33365.1 morpholino injection. 20 ng control morpholino injected larvae was used as reference. On day 5, morpholino-injected larvae were divided into two groups, normal and deformed morphology. Columns in black represent control morpholino (cMO), and gray columns stand for targeted lncRNA morpholino. * $P < 0.05$, ** $P < 0.01$. Error bars indicate standard deviation. Wilcoxon test was used to determine significance or lack thereof. The experiments were replicated three times.

studies focused on the disclosure of the roles of lncRNA in vertebrate development [34–36]. In some cases, however, similar works on the same lncRNAs come to contradictory conclusions. A typical example is that Lin *et al.*'s study [37] and Ulitsky *et al.*'s study [38] showed that knockdown of the lncRNA megamind led to zebrafish embryonic defects, whereas Kok *et al.*'s study [39] demonstrated that megamind had no influence on embryonic development. These inconsistent results have caused a dispute about whether lncRNA is important to vertebrate development. Notably, Goudarzi *et al.*'s study [14] used the CRISPR/Cas9 system to knock out 24 specific lncRNAs in zebrafish, which was picked out based on synteny, conservation, expression dynamics and proximity to development-related genes and concluded that individual lncRNAs have no overt functions in zebrafish embryogenesis, viability and fertility. Given these facts, it is likely that lncRNAs exert limited functions in vertebrate development, particularly in zebrafish embryonic development. *Ybx1* is an important protein to embryogenesis and development, disclosure of the correlation between *ybx1* and lncRNAs transcriptome will shed new insights for unveiling organelle-based mechanisms of lncRNAs implicated in zebrafish embryo development. We believe that our study also might be regarded as complementary research to provide some clues for the issue whether the lncRNAs are important to vertebrate development.

Worthy of a special mention is the fact that despite that the RNA library in this study is prepared using oligo-dT to capture sequences with poly-A tail, a great amount of lncRNAs could still be sequenced. The reason is that lncRNAs are structurally similar to mRNAs in some way, that is, the majority of mature lncRNAs are produced by the same RNA polymerase II transcriptional machinery and can be

polyadenylated, suggesting some of them are indistinguishable from mRNAs in structure. In this study, we reconstructed the zebrafish transcriptome and identified a high-confidence dataset of lncRNAs using a stringent filtering pipeline. Next, to figure out the lncRNAs implicated in zebrafish embryonic development, we performed a series of systematic transcriptomic analyses. These analyses concluded that decreasing *ybx1* by acquired knockout or innate autoregulation exhibits close correlations to REDOX processes that involved several lncRNA regulations.

Embryo development is a complicated process that underlies precise regulation and control. The roles of REDOX in embryonic development have been well elucidated so far. Concretely, oxygen can be regarded as a double-edged sword because it is essential to embryogenesis but also acts as a potential hazard via formation of ROS/RNS [40]. Too much ROS leads to oxidative stress and further causes either cell death or senescence by oxidation of cellular molecules [41]. REDOX balance is thus critical to embryonic development. Notably, ROS generated by NADPH oxidases is a crucial REDOX signal to establish homeostasis [32]. This knowledge directs us to narrow down the huge number of mRNA–lncRNA interactions to a relatively small subnetwork exhibiting strong correlation with the two key REDOX-related genes, *duox* and *noxo1a*. Among them, analysis of lncRNA–protein interactions next figured out a lncRNA (ENSDART00000171757) might interact with many proteins, which has already been demonstrated to be associated to various zebrafish embryonic development processes. For instance, *nifk* encodes a protein that is important to early embryonic development. The protein *ang2a* was found to be involved in the apoptotic process of morphogenesis, cell fate specification and midbrain development. *Park7* was shown to play an important role in cell

protection against oxidative stress and cell death acting as an oxidative stress sensor and REDOX sensitive. It was implied that the lncRNA might be a key molecule participating in a complex regulatory network involved in those critical proteins, that is, *ybx1*, *nox1a* and *duox*, etc. Indeed, subsequent validation bioassays support our predictions. Certainly, further studies are needed to clarify more detailed mechanisms.

In conclusion, our systematic transcriptome analysis successfully disclosed many lncRNAs that might be important to zebrafish embryonic development, particularly in lncRNA ENSDART00000171757. This study provides new insights into a molecular mechanism of zebrafish embryonic development. This could also potentially pave a new way to investigate the functions of lncRNAs for the regulatory elements through binding to their respective partners during the development process.

Acknowledgements

The authors thank Kuan Cheok Lei for the suggestions of manuscript revision. A preprint version of this manuscript has been released at bioRxiv [42]. This work was funded by University of Macau through research grants MYRG2018-00071-FHS, EF005/FHS-ZXH/2018/GSTIC and FHS-CRDA-029-002-2017, and by the Science and Technology Development Fund, Macau SAR (File no. 0004/2019/AFJ and 0011/2019/AKP).

Conflict of interest

The authors declare no conflict of interest.

Data availability

This Transcriptome Shotgun Assembly (TSA) project was deposited at Gene Expression Omnibus repository under the accession number [GSE134334](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134334).

Author contributions

CH and BZ designed the research and wrote the manuscript. CH and DL made bioinformatics analysis. BZ validated the experiments. XDZ and WG reviewed the manuscript and supervised the research. All authors read and approved the final manuscript.

References

1 Eliseeva IA, Kim ER, Guryanov SG, Ovchinnikov LP and Lyabin DN (2011) Y-box-binding protein 1 (YB-1)

and its functions. *Biochemistry Biokhimiia*. **76**, 1402–1433.

- 2 Troiano A, Lomoriello IS, di Martino O, Fusco S, Pollice A, Vivo M, La Mantia G and Calabro V (2015) Y-box binding protein-1 is part of a complex molecular network linking DeltaNp63alpha to the PI3K/akt pathway in cutaneous squamous cell carcinoma. *J Cell Physiol* **230**, 2067–2074.
- 3 Lyabin DN, Eliseeva IA and Ovchinnikov LP (2014) YB-1 protein: functions and regulation. *Wiley Interdiscip Rev RNA* **5**, 95–110.
- 4 Shibahara K, Sugio K, Osaki T, Uchiumi T, Maehara Y, Kohno K, Yasumoto K, Sugimachi K and Kuwano M (2001) Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer. *Clin Cancer Res* **7**, 3151–3155.
- 5 Shiota M, Izumi H, Tanimoto A, Takahashi M, Miyamoto N, Kashiwagi E, Kidani A, Hirano G, Masubuchi D, Fukunaka Y *et al.* (2009) Programmed cell death protein 4 down-regulates Y-box binding protein-1 expression via a direct interaction with Twist1 to suppress cancer cell growth. *Can Res* **69**, 3148–3156.
- 6 Huang J, Tan PH, Li KB, Matsumoto K, Tsujimoto M and Bay BH (2005) Y-box binding protein, YB-1, as a marker of tumor aggressiveness and response to adjuvant chemotherapy in breast cancer. *Int J Oncol* **26**, 607–613.
- 7 Lu ZH, Books JT and Ley TJ (2005) YB-1 is important for late-stage embryonic development, optimal cellular stress responses, and the prevention of premature senescence. *Mol Cell Biol* **25**, 4625–4637.
- 8 Uchiumi T, Fotovati A, Sasaguri T, Shibahara K, Shimada T, Fukuda T, Nakamura T, Izumi H, Tsuzuki T, Kuwano M *et al.* (2006) YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation. *J Biol Chem* **281**, 40440–40449.
- 9 Sun J, Yan L, Shen W and Meng A (2018) Maternal Ybx1 safeguards zebrafish oocyte maturation and maternal-to-zygotic transition by repressing global translation. *Development* **145**, dev166587.
- 10 Kumari P, Gilligan PC, Lim S, Tran LD, Winkler S, Philp R and Sampath K (2013) An essential role for maternal control of Nodal signaling. *eLife*. **2**, e00683.
- 11 Kung JT, Colognori D and Lee JT (2013) Long noncoding RNAs: past, present, and future. *Genetics* **193**, 651–669.
- 12 Rinn JL and Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145–166.
- 13 Sun M and Kraus WL (2015) From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. *Endocr Rev* **36**, 25–64.
- 14 Goudarzi M, Berg K, Pieper LM and Schier AF. (2019) Individual long non-coding RNAs have no overt

- functions in zebrafish embryogenesis, viability and fertility. *eLife* **8**, e40815.
- 15 Suresh PS, Tsutsumi R and Venkatesh T (2018) YBX1 at the crossroads of non-coding transcriptome, exosomal, and cytoplasmic granular signaling. *Eur J Cell Biol* **97**, 163–167.
 - 16 Wang Y and Ge W (2003) Gonadotropin regulation of follistatin expression in the cultured ovarian follicle cells of zebrafish. *Danio rerio, Gen Comp Endocrinol.* **134**, 308–315.
 - 17 Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C *et al.*, (2018) Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res* **46**, D802–D808.
 - 18 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
 - 19 Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT and Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295.
 - 20 Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R and Dewey CN (2014) Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biol* **15**, 553.
 - 21 Huang C, Morlighem JRL, Cai J, Liao Q, Perez CD, Gomes PB, Guo M, Radis-Baptista G and Lee SM (2017) Identification of long non-coding RNAs in two anthozoan species and their possible implications for coral bleaching. *Sci Rep* **7**, 5333.
 - 22 Huang C, Leng D, Sun S and Zhang XD (2019) Re-analysis of the coral *Acropora digitifera* transcriptome reveals a complex lncRNAs-mRNAs interaction network implicated in Symbiodinium infection. *BMC Genom* **20**, 48.
 - 23 UniProt C (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506–D515.
 - 24 Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L and Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* **35**, W345–W349.
 - 25 Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J *et al.*, (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**, D222–D230.
 - 26 Liao Y, Smyth GK and Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.
 - 27 Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.
 - 28 Langfelder P and Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
 - 29 Zhao X, Wang S, Wu Y, You H and Lv L (2013) Acute ZnO nanoparticles exposure induces developmental toxicity, oxidative stress and DNA damage in embryo-larval zebrafish. *Aquat Toxicol* **136–137**, 49–59.
 - 30 Mugoni V, Camporeale A and Santoro MM (2014) Analysis of oxidative stress in zebrafish embryos. *J Vis Exp* **17**, 51328.
 - 31 Wu M, Xu H, Shen Y, Qiu W and Yang M (2011) Oxidative stress in zebrafish embryos induced by short-term exposure to bisphenol A, nonylphenol, and their mixture. *Environ Toxicol Chem* **30**, 2335–2341.
 - 32 Ewald CY (2018) Redox signaling of NADPH oxidases regulates oxidative stress responses. *Immunity and Aging, Antioxidants* **7**.
 - 33 Sharma P, Chakraborty R, Wang L, Min B, Tremblay ML, Kawahara T, Lambeth JD and Haque SJ (2008) Redox regulation of interleukin-4 signaling. *Immunity* **29**, 551–564.
 - 34 Anderson KM, Anderson DM, McAnally JR, Shelton JM, Bassel-Duby R and Olson EN (2016) Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **539**, 433–436.
 - 35 Ip JY, Sone M, Nashiki C, Pan Q, Kitaichi K, Yanaka K, Abe T, Takao K, Miyakawa T, Blencowe BJ *et al.*, (2016) Gomafu lncRNA knockout mice exhibit mild hyperactivity with enhanced responsiveness to the psychostimulant methamphetamine. *Sci Rep* **6**, 27204.
 - 36 Bell CC, Amaral PP, Kalsbeek A, Magor GW, Gillinder KR, Tangermann P, di Lisio L, Cheatham SW, Gruhl F, Frith J *et al.*, (2016) The *Evx1/Evx1as* gene locus regulates anterior-posterior patterning during gastrulation. *Sci Rep* **6**, 26657.
 - 37 Lin N, Chang KY, Li Z, Gates K, Rana ZA, Dang J, Zhang D, Han T, Yang CS, Cunningham TJ *et al.*, (2014) An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* **53**, 1005–1019.
 - 38 Ulitsky I, Shkumatava A, Jan CH, Sive H and Bartel DP (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550.
 - 39 Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, Kirchmaier BC, Peterson-Maduro J, Kourkoulis G, Male I *et al.*, (2015) Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell* **32**, 97–108.
 - 40 Ufer C, Wang CC, Borchert A, Heydeck D and Kuhn H (2010) Redox control in mammalian embryo development. *Antioxid Redox Signal* **13**, 833–875.

- 41 Tsunoda S, Kimura N and Fujii J (2014) Oxidative stress and redox regulation of gametogenesis, fertilization, and embryonic development. *Reprod Med Biol* **13**, 71–79.
- 42 Huang C, Zhu B, Leng D, Ge W and Zhang XD (2019) The lncRNAs implicated in Redox regulation in Ybx1 deficient zebrafish larvae. *BioRxiv* **679167**.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Basic statistics of zebrafish RNA-seq data before and after quality trimming.

Table S2. Basic statistics of assembly results of transcriptome in zebrafish.

Table S3. Basic statistics of quality assessment of assembled transcripts achieved by DETONATE.

Table S4. Potential interacted proteins of lncRNA ENSDART00000171757 predicted from *catRAPID* server.

Fig. S1. Heatmap of DE transcripts detected in comparison between day5_ybx1^{+/+} and day5_ybx1^{-/-}.

Fig. S2. Volcano plot of DE transcripts detected in comparison between day5_ybx1^{+/+} and day5_ybx1^{-/-}.

Fig. S3. Heatmap of DE transcripts detected in comparison between day5_ybx1^{+/+} and day6_ybx1^{+/+}.

Fig. S4. Volcano plot of differentially expressed transcripts detected in comparison between day5_ybx1^{+/+} and day6_ybx1^{+/+}.

Fig. S5. Sample clustering to detect outliers.

Fig. S6. Analysis of network topology for various soft-thresholding powers.

Fig. S7. Clustering dendrograms of transcripts, with dissimilarity based on topological overlap, together with assigned module colors.

Fig. S8. Visualizing the gene network using a heatmap plot.

Fig. S9. Scatterplots of Gene Significance (GS) for recurrence versus Module Membership (MM) in the yellow module (A) and black module (B).

Fig. S10. Reverse transcription and real-time quantitative PCR to validation of DE lncRNA identified by RNA-seq between day 5 *ybx1*^{+/+} and *ybx1*^{-/-} larvae.

Fig. S11. Immunoblotting study for validation of the correlation between Ybx1 expression and ENSDART00000171757 lncRNA knockdown.