Golden pompano genome and embryo transcriptome shed light on regulation of minor ZGA gene on heart and vascular development

Honglin luo[1, #], Yongde Zhang[1, #], Changmian Ji [2, #], Yongzhen Zhao[1, #], Jinxia Peng [1, #], Yin Huang[1], Qingyun Liu[1], Pingping He[1], Pengfei Feng[1], Chunlin Yang[1], Xiuli Chen[1], Haiyan yu[2], Hongkun zheng[2, *], Yong Lin[1, *], Xiaohan Chen[1, *]

[1]Guangxi Key Laboratory for Aquatic Genetic Breeding and Healthy Aquaculture, Guangxi Institute of Fishery Sciences, Nanning, 530021, China

[2] Biomarker Technologies, Beijing, 101300, China.

[#] These authors contributed equally to this work.

[*]Prime correspondence to: Xiaohan Chen. Hongkun zheng and YongLin are also corresponding authors

Running title: Golden pompano genome and transcriptome shed light on heart and vascular development

Corresponding authors addresses and emails:

NO. 8, Qingshan Road, Guangxi Key Laboratory for Aquatic Genetic Breeding and Healthy Aquaculture, Guangxi Institute of Fishery Sciences, Nanning, 530021, China.

Xiaohan Chen: Chnxhn@163.com; Hongkun zheng: zhenghk@biomarker.com.cn; Yonglin: linnn2005@126.com.

Whole-genome duplication (WGD) event affects embryo development in teleosts[1-3]. Although minor zygotic genome activation (ZGA) contributes to further activation of the genome in the mouse[4], the cardio- and vascular-linked functions of the genes in the WGD shaping minor ZGA in marine fish are largely unknown. Here, we present a *de novo* chromosome level-assembly of a marine fish golden pompano and mass transcriptome analyses from 19 embryo development stages to explore how the WGD (Ts3R) contributed to minor-ZGA. Subgenome comparative analyses showed that karyotypes-retained genes are crucial for retaining embryo development stability. Remarkably, we identified one group of specific highly expressed hub-genes within minor ZGA which is essential for the major ZGA. Of which, we discovered a function unknown gene XingHuo (designated as *XH*) belonging to medium fractionated subgenome locates in key node of specific expression network. For the first time, we report that loss of its homolog *ZfXH* decreases the subsequent ZGA activation, causes pericardial oedema, body axis bending, caudal fin defects, thinner intersegmental vessels (ISVs), and disruption of the honeycomb structure in the caudal vein plexus (CVPs) in zebrafish. Furthermore, loss of *ZfXH* downregulates the heart- and vascular-associated landmark gene network (including keystone gene *Ptprb,* Hand2, Tie2, Hey2, and S1pr1, etc.) via a hitherto unreported *ZfXH*-Nucleolin-*Ptprb* pathway. This suggests that *XH* may function similarly in golden pompano development. We also found that silencing of HARBI1, the human homolog of *XH*, does suppress the angiogenesis of the cancer cells, indicating its potential as an anti-cancer therapeutic target. Our findings provide new insights into cardio- and vascular-linked functions of the genes in the WGD shaping minor ZGA of marine fish during embryo development.

Teleost fishes demonstrate strong adaptability and show a remarkable explosion of biodiversity. The well-described while genome duplication (WGD)[5,6] may have contributed to these features. It has been evidenced that teleosts underwent a third round of WGD (Ts3R, about ~320 million years ago (Mya)[7,8], which is thought to have been a driver for the radiation and diversification of teleosts[9] and to be a cause of phenotypic innovation and complexity[3,10]. After WGD, many of the duplicated genes are lost. This loss often shows a subgenome-bias pattern [11,12]. Although the correlations between the WGD driven-subgenome-bias pattern and the emergence of evolutionary novelties have been demonstrated [13], whether the genes in the biased-subgenome affect the global zygotic genome activation (ZGA) is largely unclear.

ZGA is a vital process for subsequent embryo development[14,15] (for a review of the mechanisms underlying ZGA, see ref [15]). Although the timing of ZGA and from one to another transcriptional activation vary in different vertebrate species (reviewed in refs [15,16]), studies using genomic approaches confirmed that a minor wave of ZGA (a low level, yet distinct, early activation of transcription) exists prior to the major wave of ZGA [17,18]. The latest study in mice demonstrated that transient inhibition of minor ZGA results in compromised development of major ZGA and subsequent stages [4], suggesting the minor ZGA may contribute to signaling the further activation of the genome. However, the functions of the genes in the biased-subgenome and the minor ZGA of teleosts during embryonic development are largely unknown. In this study, we tried to reveal the functions of such genes in the embryonic development stage of a sea fish *Trachinotus ovatus*.

T. ovatus (Linnaeus 1758, also called golden pompano) is a member of the family *Carangidae* (Rafinesque, 1815)[19]. It is widely distributed and cultured in the Asia-Pacific region, and has significant economic importance for offshore cage aquaculture in China and Southeast Asian countries [19]. Due to the lack of genome data and detailed gene annotations, the genetics and gene regulation in this species remain largely unexplored. Here, we report, for the first time, the *de novo* chromosome level-assembly and annotation of golden pompano genome that underwent a subgenomic pattern (least fractionated (LF), medium fractionated (MF), and complete fractionated (Other)). We identified a minor ZGA prior to the major ZGA during its embryonic development, and a fire-new gene EVM0008813 (designated as

XingHuo, *XH*), which concurrently belongs to the subgenome MF and the minor ZGA. We studied the function of *XH* in the zebrafish model by morpholino knockdown of its homolog *zgc:113227* (designated as *ZfXH*). For the first time, we reveal that loss of *ZfXH* decreases the subsequent ZGA activation, causes heart and vascular defects and reduces the angiogenesis via *ZfXH*-nucleolin-ptprb network in zebrafish. This implies that *XH* may play a similar function in golden pompano development. Our study provides a new insight on cardio-and vascular-linked functions of the genes in the minor ZGA of teleosts during embryonic development.

RESULTS

Genome sequencing, assembly and annotation

We first estimated the genome size and GC content of golden pompano at 656.98 Mb and 41.54% based on 21-Kmer (Figure S1).Then we sequenced and assembled the genome of a female individual through a combination of three technologies: paired-end sequencing with the Illumina HiSeq platform, single-molecule real time (SMRT) sequencing with the PacBio Sequel platform, and optical genome mapping with the BioNano Genomics Saphyr System (Figure S2; TableS1, S2A), and an assembly containing 1,490 scaffolds, with a scaffold N50 length of 21.02 Mb was built (Table S1C and S2A). Next, we anchored and oriented the assembly sequences onto 24 pseudo-molecules, which accounted for ~98.60% (636.61 Mb) of genome assembly (Figure S3; Table S1 and S2) according to the interaction frequency mapping of 38.7-fold high-through chromosome conformation capture (Hi-C) data (Table S1D). We finally obtained a genome assembly containing 645.62 Mb of genome data made of 1,536 scaffolds (included 24 pseudochromosomes) with scaffold N50 size of 20.22 Mb (Table S2B). The quality evaluation of the assembled genome turned out to reveal a high level of contiguity and connectivity for the golden pompano genome and facilitated further analyses (Table S2, S3; Figure S3). By combing the evidence of *de novo*, homologs and transcriptome approaches, we annotated 24,186 high confidence protein-coding genes, 152 pseudogenes and 1,274 noncoding RNAs (Table S3, S4 S5 and S6), in comparison to zebrafish (25,642 protein-coding genes). We visualized the genomic landscape of genes, repetitive sequences, genome map markers, Hi-C data, and GC content of the golden pompano genome by circus[20]

(Figure 1). As expected, the repetitive elements accumulated in low gene density regions (Figure 1A, 1B). The optical markers and Hi-C data are evenly distributed across the genome (Figure 1C, 1D), as well as GC content (Figure 1E). Detailed genome data can be found in the supplementary information (Table S7-S17).

Subgenomic patterning suggests karyotypes-retained genes crucial for retaining embryo development stability

To examine the evolutionary position of the golden pompano, we reconstructed the evolutionary history of teleost fishes. Phylogenies strongly supported the traditional classification that the golden pompano belongs to *Carangidae*. Golden pompano is a sister of *Seriola dumerilie,* and diverged at approximate 59.99 Mya (Figure 2A). Assuming a constant rate of distribution of silent substitutions (dS) [21] of 1.5e-8, we estimated the dates of Ts3R and Ss4R at 350 Mya and 96 Mya, respectively (Figure 2B). Genome collinearity comparison among spotted gar karyotypes (preserved in golden pompano genome) showed that 894 pairs of paralogous genes that were inherited from the Ts3R event (ohnologues) retained a double-conserved synteny block in the golden pompano genome (Figure 2C and Table S18), implying that teleost-ancestral karyotypes are considerably conserved in post-Ts3R rediploidization with large fissions, fusions or translocations (Figure 2D and Table S19-S20). Next, we classified the Ts3R subgenomes according to the integrity of gene as belonging to the LF, MF, and Other subgenome [22]. The component of rediploidization-driven subgenomes (LF, MF, and Other) is unequally distributed among golden pompano subgenomes (Figure 2E), suggesting an asymmetric retention of ancestral subgenomes in teleosts, which is commomly observed in plants [23-25].

Although the gene expression evolution following the WGD event has debated in fishes[26-28], knowledge on the rediploidization process and embryo development stability is lacking. We then compared the genome-wide transcriptional levels of LF, MF, and Other karyotypes from whole-embryo development stages (OSP to YAPS) (Figures 3A). Karyotypes-retained regions (LF and MF) showed comparable expression levels during the embryo development, while karyotypes-loss regions (Other) were expressed at significantly lower levels (signed-rank sum test, $P < 0.01$) (Extended data 1A). The Ks/Ks values of karyotypes-retained regions are

significantly lower than those of karyotypes-loss regions (Extended data 1B). This observation indicated that karyotypes-loss genes evolved faster than did the karyotypes-retained regions. We propose that karyotypes-retained genes are crucial for retaining embryo development stability and that karyotypes-loss genes are more prone to contribute to genetic diversity. Detail descriptions about subgenome can be found in supplementary information.

Discovery of a minor ZGA gene that concurrently belongs to the karyotypes-retained genes during embryo development

To prove this idea, we collected 19 embryo stages of the golden pompano and sequenced the mRNAs to figure out the gene expression patterns (Figure S5 and Table S21-S26). We found that all 57 of the samples were separated into two components according to the expression profile (Figures 3A, 3B, and 3C). The first 33 samples (from OSP to MGS) cluster into a clade and the residual 24 samples (from LGS to YAPS) cluster into another. The genes in the first clade were clearly "silenced" compared with those of the second clade, in which the gene levels show an explosive increase, which is reminiscent of the ZGA process. We then identified the stages after MGS, namely from LGS to YAPS stage, are the major ZGA (Figures 3A, 3B and 3C; Extended data 2 and Figure S5). This ZGA expression profile was verified by our qPCR results (Figure S6). We also noticed that before LGS, a group of genes in three stages, HBS, EGS, and MGS, are highly expressed in the first clade (Figure 3D). This unique gene expression profile might be the minor wave of ZGA before the major wave of ZGA. The existence of minor ZGA was previously documented in mammal embryo development[4,17,18], and it was suggested to contribute to signaling the massive activation of the genome[29]. Thus, we asked what the functions of the genes in the minor ZGA are during the embryo development of golden pompano.

Next, we clustered these genes in the minor ZGA using the WGCNA R package and we found that most of them clustered into the purple_module and are co-expressed in a close network, indicating regulatory roles for these genes before the major ZGA (Figure 3E and 3F). Then we randomly picked 8 genes to verify their expression trends, and we found all of them to coincide in the expression profile of the minor ZGA genes (Figure S6). We subsequently

choose the mostly coincident gene to the minor ZGA expression trend, EVM0008813 (designated as XingHuo, *XH*; Extended data 3), as a representative gene to figure out the function of minor ZGA during embryo development of golden pompano. We found that *XH* is one of the karyotypes-retained genes (MF). We then searched the Nr database in NCBI by BLASTp and only one gene *zgc:113227* (designated as *ZfXH*) shares 54.7% similarity to *XH* at the amino acid level in zebrafish. The collinear analysis shows its homolog gene in zebrafish is also *ZfXH*, and both have the same functional domain DDE_Tnp_4 as the other seven genes have (Extended data 3), suggesting they may have similar biological functions during embryo development. What could the function of *XH* be?

Knockdown of *ZfXH* suppresses ZGA activation and causes morphological defects in the heart and caudal fin in zebrafish

Next, we investigated the functions of *ZfXH*, the zebrafish homolog of *XH*, in zebrafish. Since *ZfXH* shares high (54.7%) similarity to *XH* at the amino acid level, we used the morpholino-based knockdown method in zebrafish to study the function of *ZfXH,* which may represent the function of *XH* because they share the DDE_Tnp_4 domain. We used two types of morpholinos (MOs) (splice blocking: *ZfXH* ^{e1i1} MO and translational blocking: *ZfXH* ^{ATG} MO) to silence *ZfXH* (Figure 4A). As *XH* is supposed to be a minor ZGA gene in golden pompano, we tried to figure out if knockdown of its homolog *ZfXH* inhibits the activation of ZGA in zebrafish. We found that loss of *ZfXH* decreased the expressions of five ZGA marker genes[30] at 6 hpf in zebrafish, implying that *ZfXH* significantly suppresses the activation of zebrafish ZGA (Figure 4B). This finding is consistent with previous research in the mouse[4] and highlights the important function of *XH* and *ZfXH* during embryo development.

Heart and vascular development starts at 16 hpf [31] and 24 hpf [32]in zebrafish, respectively. This means heart and vascular development occur after the major ZGA (3 hpf). Since the loss of *ZfXH* significantly suppresses the major ZGA in zebrafish, we speculated that *ZfXH* may affect the heart and vascular development. Then we checked the changes in the heart and vascular system after the silencing of *ZfXH.* We observed that injection of 4 ng of morpholino *ZfXH* ^{e1i1} MO and *ZfXH* ^{ATG} MO both resulted in nearly identical phenotypes of pericardial

oedema, body axis bending, and caudal fin defects (Figure 4C (a-k)； Figure S7A and Figure S8) at 3 dpf (days post fertilization), suggesting that the phenotype of *ZfXH* knockdown is *ZfXH*-specific (Figure 4C). Considering the vascular system, embryos injected with *ZfXH* ^{e1i1} MO present thinner ISVs (yellow arrows) and ectopic sprouts (asterisk) of dorsal aorta compared with controls, and the *ZfXH* knockdown prevents the PAV formation, the precursor to the lymphatic system. Moreover, heartbeat and circulation in the caudal vein (CV) is visible in the control fish, but is abnormal in *ZfXH*-MO-injected fish (Supplementary Movie1, Supplementary Movie2). The efficiency of *ZfXH* ^{e1i1} MO and *ZfXH* ^{ATG} MO at 6 hpf and 3 dpf demonstrated that injection of 4 ng of *ZfXH* ^{e1i1} and *ZfXH* ^{ATG} morpholino dramatically disrupted normal splicing of *ZfXH* (Figure 4C (j))， indicating high efficiency and specificity of the morpholino knockdown of *ZfXH*. We also recorded a high percentage of embryos with defects (81.55%, n=103 embryos in *ZfXH* ^{e1i1} MO and 100%, n=106 embryos in *ZfXH* ^{ATG} MO) with comparison to control (n=128 embryos) and low survival rate at 3 dpf (45.78%, n=225 embryos in *ZfXH* ^{e1i1} MO and 17.68%, n=198 embryos in *ZfXH* ^{ATG} MO) compared to controls (n=218 embryos) (Figure 4C (k) and Figure S7B). This confirmed that knockdown of *ZfXH* certainly causes morphological defects in the heart and caudal fin in zebrafish.

Loss of *ZfXH* causes vascular defects and impairs formation of the caudal vein plexus (CVP) in zebrafish

Since *ZfXH* silence caused pericardial edema and caudal fin defects, we speculated that there is a positive connection between the vascular system and these phenotypes. We then used the *Tg(fli1a:EGFP)*^{y1} zebrafish as a model to find out if the vascular development is impaired by knockdown of *ZfXH*. Embryos were injected with 4 ng control MO or *ZfXH* ^{e1i1} MO, and the results exhibited that loss of *ZfXH* caused intersegmental vessel (ISV) growth defect and disruption of the honeycomb structure in the CVP at 52 hpf (Figure 4D (a-f)). Also, a thinner ISV growth and ectopic sprouts of dorsal aorta were observed at the rear-somite with only 10% of complete ISVs (n=365 embryos) in *ZfXH* ^{e1i1} morphants compared to 98% of complete ISVs in controls (n=335 embryos). In control embryos, the parachordal vessels (PAV) were formed normally, while the *ZfXH* ^{e1i1} MO knockdown prevented the parachordal vessels (PAV) formation (Figure 4D (g)). We also observed that CVPs formed

honeycomb-like structures at the tail around 52 hpf in control embryos, but *ZfXH* ^{e1i1} MO knockdown caused specific defects in CVP formation (Figure 4E (a-f)). Quantification of loop formation and the area at CVP showed a 8.2-fold and 3-fold decrease in *ZfXH* ^{e1i1} MO morphants (n=10 embryos) at 52 hpf, respectively (Figure 4E (g and h)). Our data indicate that *ZfXH* plays a critical role in controlling vascular integrity and in regulating ISV and CVP formation during angiogenesis, which is an explanation strongly consistent with the phenotypes observed. What is the mechanism behind?

ZfXH may play role in angiogenesis via endothelial cells and its human homolog HARBI1 is a potential anti-cancer therapeutic target

As *ZfXH* plays a critical role in controlling vascular integrity, we then speculated that it may affect angiogenesis via endothelial cells. Since human HARBI1 gene shares DDE_Tnp_4 domain with *ZfXH* and *XH* (Extended data 3), it is supposed to play similar role to *ZfXH* and *XH* in vascular epithelioid cells. Next, we designed siRNAs targeting human HARBI1, transferred siRNA into human umbilical vein endothelial (HUVEC) cells and investigated their cell migration, invasion, and tube formation. We observed that silencing of HARBI1 not only significantly inhibits the cell migration and invasion abilities but also the tube formation compared with controls in HUVECs (Extended data 4A). This highlights the anti-angiogenesis function of HARBI1, indicating that *XH* and *ZfXH* are functionally related. Because targeting angiogenesis has great potential in anti-tumor or anti-cancer therapy [33,34], we next asked whether this anti-angiogenesis function of HARBI1 affects the angiogenesis of the cancer cells. We choose HepG2 (hepatocellular carcinoma, HCC), A549 (non-small cell lung cancer, NSCLC), and HCT116 (colon carcinoma, Colo) cancer cell lines, transferred with si-HARBI1, then their supernatants incubated with HUVECs, respectively. We observed the abilities of tube formation in all of the cancer cells are significantly inhibited compared with controls (Extended data 4B). This shows that silencing of HARBI1 does suppress the angiogenesis of the cancer cells, indicating its potential as an anti-cancer therapeutic target. Such function of HARBI1 was never reported [35-37] and more works should be done to validate this potential. However, our data indicate that *ZfXH* and *XH*, like their human homolog HARBI1, may play role in angiogenesis process via endothelial cells.

Loss of *ZfXH* leads to the down-regulation of the landmark gene network associated with heart and vascular development in zebrafish

Next, we asked how *ZfXH* affects angiogenesis. We first examined transcriptome sequencing (RNA-seq) data from zebrafish after injection of 4 ng *ZfXH* ^{e1i1} MO at 3 dpf. We found that loss of *ZfXH* greatly changes the transcriptome with 1955 down-regulated and 698 up-regulated (Figure S9; Table S21). We noticed that in the KEGG pathways associated with heart and vascular development are significantly enriched in the *ZfXH*-silenced group (Figure S10; Table S22-26). We speculated that the transcription of genes linked to heart and vascular development may also be significantly changed in the *ZfXH*-silenced zebrafish. We then picked and examined the expression of 18 genes previously documented to be closely related to heart defects and/ or angiogenesis [38-42] using qPCR. Consistent with the RNA-seq data, we found that 13 of these genes (Ptprb, Tie2, Nr2f1a, S1pr1, Hey2, Dot1L, Hand2, Erbb2, Klf2a, Mef2cb, Mef2aa, EphB2a and CX40.8) were significantly decreased while two genes (VEGFaa and VEGFR2) increased sharply. S1pr2, Egfl7, and Nrg2a were kept unchanged (Figure 5A-E). Normally, the increase of VEGFaa and VEGFR2 is linked to the enhancement of vascular system [43,44]. However, in our study, both genes increased while others decreased when *ZfXH* was silenced. We speculated this was a consequence of a negative feedback regulation to avoid an excessive decrease in the vascular system. We next found that at the protein level, when *ZfXH* is silenced, the *ZfXH* protein also decreases. Ptprb, the most decreased gene, is also greatly reduced at the protein level. The S1pr1, Hand2, Dot1L, and Hey2 proteins were also downregulated compared with controls (Figure 5F). As previously reported, Ptprb, Tie2, Nr2f1a, S1pr1, VEGFaa and VEGFR2 normally contribute to vascular development and deletion of each of them leads to defects on the vascular system during embryo development [43-47], while loss of Dot1L, Hand2, Erbb2, Mef2cb, Mef2aa, EphB2a or CX40.8 always results in cardiovascular system or heart development defects [40,48-54]. Hey2 and Klf2a have been implicated in the regulation of both angiogenesis and heart development [55,56]. Based on these reports, we built a schematic diagram of the network as shown in (Figure 5F). This network demonstrated that silencing of *ZfXH* did downregulate the key genes that are essential for heart and /or vascular development. To this end, our results showed that loss

of *ZfXH* greatly affects the expression of these key genes in the network, suggesting that the heart and vascular phenotypes caused by *ZfXH* silencing are greatly due to the regulation of these genes. Thus, we then asked how *ZfXH* mediates the network.

Ptprb plays similar function to *ZfXH* in zebrafish and is downstream of *ZfXH*

As described above, we noticed that *ptprb* is the most downregulated gene after knocking down of *ZfXH* and one that is closely linked to both vascular integrity and angiogenesis[57-62]. To test whether there is a positive connection between *ZfXH* and *ptprb,* we knocked down *ptprb* by injection of 4 ng *ptprb* MO morphants as previously documented [42]. We observed that embryos injected with *ptprb* MO present slight pericardial edema, shortened body axis and severe body axis bending (Extended data 5A (a-h)； Figure S11). We also recorded a high percentage of embryos with defects (87.11%, n=287 embryos in *ptprb* MO and 1.76%, n=284 embryos in control) and lower survival rate at 48 hpf compared with controls (Extended data 5A (i)). The efficiency of *ptprb* MO at 48 hpf demonstrated that injection of 4 ng of *ptprb* morpholino was highly efficient in producing a knockdown of *ptprb* (Extended data 5A (j-k)). In the vascular system, loss of *ptprb* leads to an indefinite absence or deformity of DLAVs and ISVs in the tail end (white arrowhead), and a decrease of PAV formation (Extended data 5B (a-f)). We also observed that CVPs formed honeycomb-like structures at the tail around 48 hpf in control embryos, but *ptprb* MO knockdown causes CVP sinus cavities defects (Extended data 5B (g, h)). Quantification of loop formation and area at CVP showed a 5.2-fold and 1.5-fold decrease in *ptprb* MO (n=83 embryos) at 48 hpf, respectively (Extended data 5B (i, j)). These data are to some extent consistent with previous reports [42] and strongly suggest that loss of *ptprb* results in heart and vascular phenotypes similar to those of *ZfXH*. Moreover, we also investigated the expression of the 15 genes that were examined in *ZfXH*-knockdown experiment. We found that most of the genes present an expression profile similar to that in *ZfXH*-knockdown experiment, except VEGFaa and VEGFR2 (Extended data 5C (a, b)). We then logically concluded that *ZfXH* and *ptprb* should act in the same signaling pathway. However, which one is downstream of the other is unclear. Therefore, we examined the gene expression of *ZfXH* after silencing of *ptprb*, and we found that it was kept unchanged (Extended data 5C (c)), but we observed a significant decrease of *ptprb* expression after

silencing of *ZfXH*. These observations demonstrate that *ptprb* acts at the downstream of *ZfXH*.

***ZfXH-NCL-Ptprb* signaling links heart and vascular molecular network to regulate *ZfXH*-induced phenotypes**

As mentioned above, among the 18 genes associated with heart and vascular development, 15 genes were significantly changed by the *ZfXH* and *ptprb* knockdown. We suppose these genes may be part of a regulatory network of their own. We then built a schematic diagram of the network according to previous reports (Figure 5E). This network presents connections between most of these genes, suggesting a cooperative regulation mechanism on the heart and vascular development. As we already knew that *ptprb* acts downstream of *ZfXH*, we next asked if *ZfXH* directly interacts with *ptprb* to mediate the network. Thus, we designed *ZfXH* probes and conducted a ChIRP-MS experiment in zebrafish to find out those proteins interacting with *ZfXH*. Eleven proteins with change folds above 2 were discovered (Figure 6A (a,b)). This indicates that the *ZfXH* RNA may interact with these proteins. Unexpectedly, *Ptprb* was not found in these interacting proteins (Figure 6A (a, b)). This suggests that proteins other than *ptprb* may be involved. We next focused on the proteins of the group of eleven proteins that are associated with the vascular system, and nucleolin (NCL) (Figure 6A (c, d)) aroused our interest because of its molecular conservation and important functions on angiogenesis[63,64]. To figure out whether *ZfXH* interacts with NCL, we performed RNA Immunoprecipitation (RIP) using the NCL protein as bait protein in 293T cells (Figure S12) and then detected the *ZfXH* RNA using qPCR. We found that *ZfXH* RNA is significantly higher than that in IgG control in the RNAs pulled-down by the NCL protein (Figure 6B (b, c, d) and Figure S13). The RNA pulled down was amplified and the sequencing results confirmed that it is ZfXH gene. This indicates that the NCL protein reversely interacts with *ZfXH* RNA. Therefore, these experiments prove that *ZfXH* RNA and NCL interact physically.

However, still no evidence was shown on the relationship between *ZfXH* and *ptprb*. Could it be that NCL interacts with *ptprb,* thus bridging *ZfXH* and *ptprb?* Such scenario was never proposed or documented before. However, a report showed that VEGF, which plays roles on vascular development as *ZfXH* does, does interact with NCL[65]. We then supposed that NCL might also interact with *ptprb* or its human homologue VE-PTP, the key molecule in vascular

development. To test this hypothesis, we detected VE-PTP RNA using the same RNAs pulled-down by NCL protein, and we found that VE-PTP RNA is significantly higher than that in IgG control. The RNA pulled down was amplified and the sequencing results proved that it is VE-PTP gene. This indicates that the NCL protein can also interact with VE-PTP RNA physically (Figure 6B (b, e, f)). We next verified this interaction in 293T cells using the VE-PTP RNA pulldown experiment in the reverse way, and the result of western blotting against NCL protein supports the interaction between VE-PTP and NCL (Figure 6C (a)). However, whether this interaction also occurs between NCL and *ptprb* in zebrafish is unclear. We next designed a zebrafish *ptprb* gene-specific probe to pull down the proteins that interact with *ptprb* in the juvenile zebrafish. We found that the NCL protein band is much stronger that in controls (Figure 6C (b)). These results indicate that the NCL protein not only interacts with VEP-PTP in 293T cells but also with *ptprb* in zebrafish. So far, we proved that *ZfXH* and *NCL, NCL* and *ptprb* interact physically. However, how *ZfXH* regulates NCL and *ptprb* is unclear. To address this issue, we micro-injected 4 ng *ZfXH*-e1i1-MO in one cell stage embryo, and found that resemble phenotypes were induced as that shown in Figure 4 (Extend data 6). Meanwhile, we found that loss of *ZfXH* not only causes a significant decrease of NCL mRNA and total protein level but also leads to reduction of T76 phosphorylation level and an increase of the K88 acetylation level of the NCL protein (Figure 6D (a, b)). This suggests that knockdown of *ZfXH* significantly affects the expression of NCL, which plays vital functions in angiogenesis[64], although the impact of phosphorylation and acetylation of NCL protein on the heart and vascular development have not been deeply understood yet [63]. Then we investigated the expression of the downstream gene *ptprb*, and found that loss of *ZfXH* also decreases *ptprb* at both mRNA and protein levels (Figure 6D (c, d)). These results suggest that silence of *ZfXH* leads to heart and vascular defect due to the downregulation of both NCL and *ptprb* via the decrease in the *ZfXH-NCL* and *NCL-ptprb* interaction.

Taken together, we conclude that the NCL protein can not only interact with the upstream *ZfXH* but also the downstream *ptprb*. This uncovers the existence of regulatory gene upstream of *ptprb* gene, one of the most important vascular-controllers. Certainly, this logically proves that *XH* may act in the same way in the golden pompano. Based on these data, we built a new schematic diagram of the network that shows the *ZfXH-NCL-Ptprb* signaling links of the heart

and vascular system molecular network (Figure 6E (a) vs. Figure 5E). We also made a schematic diagram to describe the possible mechanism underlying *ZfXH*-induced phenotypes (Extended data 7), highlighting the finding that loss of *ZfXH* significantly and broadly downregulates heart- and vascular- associated landmark gene network via the *ZfXH-NCL-ptprb* pathway. Given the extreme importance of the heart and vascular development, and the broad connections with landmark genes described in the network, we believe the finding of this novel signaling pathway to be of considerable relevance for the study of the heart and vascular development regulatory network.

Discussion

Taken together, we report the first high-quality chromosome-level genome for the golden pompano, which underwent the Ts3R WGD and shows biased-subgenomes retention. For the first time, we show the existence of minor and major ZGA during golden pompano embryo development. We also for the first time report that the karyotypes-retained gene *XH,* the homolog of zebrafish *ZfXH*, belongs to the minor ZGA gene, and it is likely to play key roles on the heart and vascular development through regulating a heart- and vascular-associated network of landmark genes. Moreover, we are the first to highlight the interactions between *ZfXH* and NCL, NCL and *ptprb*, and the existence of a hitherto unreported *ZfXH*/NCL/*ptprb* signaling pathway as a regulatory complex mediating both heart and vascular development. Finally, we show that the novel angiogenesis-linked function of HARBI1, the human homolog of *XH*, may be a potential anti-cancer therapeutic target. A detailed discussion can be found in the supplementary information.

Acknowledgements

We thank Biomarker Technologies Corporation for genome and transcriptome sequencing. This research was supported by the Guangxi science and technology major project (GuiKeAA18242031, GuiKeAA18242031-2, GuiKeAA17204080, GuiKe AA17204080-3), Guangxi Key Laboratory for Aquatic Genetic Breeding and Healthy Aquaculture, Guangxi Institute of Fishery Sciences (14-045-10 (14-A-01-02),15-140-23(15-A-01-01, 15-A-01-02, 15-A-01-03),16-380-38(16-A-01-01, 16-A-01-02)) and Guangxi research institutes of basic

research and public service special operations (GXIF-2016-03, GXIF-2016-09, GXIF-2016-18, GXIF-2016-19）.

Author Contributions

H.L.L and X.H.C designed the scientific objectives and oversaw the project. Y.D.Z, J.X.P., Y.L., Y.H., Q.Y.L., P.P.H., C.L.Y., X.L.C., and P.F.F. collected samples for sequencing DNA and RNA. C.M.J. and their colleagues performed genome sequencing, assembly and annotation. C.M.J. and H.Y.Y performed phylogenomic and whole genome duplication evolution analysis. C.M.J. H.Y.Y., H.L.L., and Y.D.Z performed RNA-seq analysis. H.L.L performed functional assay of zebrafish *XH* gene. C.M.J, H.Y.Y., H.L.L., and Y.D.Z prepared the supplemental data and method. C.M.J. and H.L.L prepared the draft manuscript with input from all other authors. H.L.L., X.H.C., Y.L., and H.K.Z. discussed and revised the manuscript.

Data availability

The authors declare that all data reported in this study are fully and freely available from the date of publication. The draft genome data (genome assembling and annotations) and RNA-Seq data of the embryo are available under BioProject PRJNA574895. Transcriptome (Illumina) data of ZfXH silence are available in the Sequence Read Archive (SRA) with accession number SRR10199007 and SRR10199008 under BioProject PRJNA573544 .

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental Information can be found with this article online version.

References

1 Moriyama, Y. & Koshiba-Takeuchi, K. Significance of whole-genome duplications on the emergence of evolutionary novelties. *Briefings in functional genomics* **17**, 329-338, doi:10.1093/bfgp/ely007 (2018).

2 Glasauer, S. M. & Neuhauss, S. C. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular genetics and genomics : MGG* **289**, 1045-1060, doi:10.1007/s00438-014-0889-2 (2014).

3 Arnegard, M. E., Zwickl, D. J., Lu, Y. & Zakon, H. H. Old gene duplication facilitates origin and diversification of an innovative communication system--twice. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22172-22177, doi:10.1073/pnas.1011803107 (2010).

4 Abe, K. I. *et al.* Minor zygotic gene activation is essential for mouse preimplantation development. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E6780-E6788, doi:10.1073/pnas.1804309115 (2018).

5 Adamska, M. *et al.* Five Nkx5 genes show differential expression patterns in anlagen of sensory organs in medaka: insight into the evolution of the gene family. *Dev Genes Evol* **211**, 338-349 (2001).

6 Robinson-Rechavi, M. *et al.* Euteleost fish genomes are characterized by expansion of gene families. *Genome research* **11**, 781-788, doi:10.1101/gr.165601 (2001).

7 Smith, J. J. *et al.* Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nature genetics* **45**, 415-421, 421e411-412, doi:10.1038/ng.2568 (2013).

8 Nakatani, Y., Takeda, H., Kohara, Y. & Morishita, S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome research* **17**, 1254-1265, doi:10.1101/gr.6316407 (2007).

9 Meyer, A. & Schartl, M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* **11**, 699-704 (1999).

10 Braasch, I., Brunet, F., Volff, J. N. & Schartl, M. Pigmentation pathway evolution after whole-genome duplication in fish. *Genome Biol Evol* **1**, 479-493, doi:10.1093/gbe/evp050 (2009).

11 Cheng, F. *et al.* Epigenetic regulation of subgenome dominance following whole genome triplication in Brassica rapa. *New Phytol* **211**, 288-299, doi:10.1111/nph.13884 (2016).

12 Hughes, T. E., Langdale, J. A. & Kelly, S. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome research* **24**, 1348-1355, doi:10.1101/gr.172684.114 (2014).

13 Mammoto, T., Mammoto, A. & Ingber, D. E. Mechanobiology and developmental control. *Annu Rev Cell Dev Biol* **29**, 27-61, doi:10.1146/annurev-cellbio-101512-122340 (2013).

14 Blythe, S. A. & Wieschaus, E. F. Coordinating Cell Cycle Remodeling with Transcriptional Activation at the Drosophila MBT. *Curr Top Dev Biol* **113**, 113-148, doi:10.1016/bs.ctdb.2015.06.002 (2015).

15 Harrison, M. M. & Eisen, M. B. Transcriptional Activation of the Zygotic Genome in Drosophila. *Curr Top Dev Biol* **113**, 85-112, doi:10.1016/bs.ctdb.2015.07.028 (2015).

16 Sandler, J. E. & Stathopoulos, A. Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. *Genetics* **202**, 1575-1584, doi:10.1534/genetics.116.186783 (2016).

17 Heyn, P. *et al.* The earliest transcribed zygotic genes are short, newly evolved, and different across species. *Cell Rep* **6**, 285-292, doi:10.1016/j.celrep.2013.12.030 (2014).

18 Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. & Klein, P. S. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Developmental cell* **19**, 220-231, doi:10.1016/j.devcel.2010.07.007 (2010).

19 Nelson, J. S. Fishes of the World.Third edition [M]. *New York: John Wiley&Sons, Inc.* (1994).

20 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).

21 Crete-Lafreniere, A., Weir, L. K. & Bernatchez, L. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PloS one* **7**, e46662, doi:10.1371/journal.pone.0046662 (2012).

22 Cheng, F. *et al.* Deciphering the diploid ancestral genome of the Mesohexaploid Brassica rapa. *The Plant cell* **25**, 1541-1554, doi:10.1105/tpc.113.110486 (2013).

23 Session, A. M. *et al.* Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**, 336-343, doi:10.1038/nature19840 (2016).

24 Emery, M. *et al.* Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS genetics* **14**, e1007267, doi:10.1371/journal.pgen.1007267 (2018).

25 Liang, Z. & Schnable, J. C. Functional Divergence between Subgenomes and Gene Pairs after Whole Genome Duplications. *Molecular plant* **11**, 388-397, doi:10.1016/j.molp.2017.12.010 (2018).

26 Braasch, I. *et al.* The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature genetics* **48**, 427-437, doi:10.1038/ng.3526 (2016).

27 Lien, S. *et al.* The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200-205, doi:10.1038/nature17164 (2016).

28 Berthelot, C. *et al.* The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications* **5**, 3657, doi:10.1038/ncomms4657 (2014).

29 Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol* **30**, 581-613, doi:10.1146/annurev-cellbio-100913-013027 (2014).

30 Joseph, S. R. *et al.* Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife* **6**, doi:10.7554/eLife.23326 (2017).

31 Crosier, P. S. *et al.* Pathways in blood and vessel development revealed through zebrafish genetics. *Int J Dev Biol* **46**, 493-502 (2002).

32 Glickman, N. S. & Yelon, D. Cardiac development in zebrafish: coordination of form and function. *Semin Cell Dev Biol* **13**, 507-513 (2002).

33 Castet, F. *et al.* Uveal Melanoma, Angiogenesis and Immunotherapy, Is There Any Hope? *Cancers* **11**, doi:10.3390/cancers11060834 (2019).

34 Qin, S. *et al.* Recent advances on anti-angiogenesis receptor tyrosine kinase inhibitors in cancer therapy. *Journal of hematology & oncology* **12**, 27, doi:10.1186/s13045-019-0718-5 (2019).

35 Smith, J. J., Sumiyama, K. & Amemiya, C. T. A living fossil in the genome of a living fossil: Harbinger transposons in the coelacanth genome. *Molecular biology and evolution* **29**, 985-993, doi:10.1093/molbev/msr267 (2012).

36 Sinzelle, L. *et al.* Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4715-4720, doi:10.1073/pnas.0707746105 (2008).

37 Kapitonov, V. V. & Jurka, J. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. *DNA and cell biology* **23**, 311-324, doi:10.1089/104454904323090949 (2004).

38 Pereira, C. F. *et al.* Hematopoietic Reprogramming In Vitro Informs In Vivo Identification of Hemogenic Precursors to Definitive Hematopoietic Stem Cells. *Developmental cell* **36**, 525-539, doi:10.1016/j.devcel.2016.02.011 (2016).

39 Boselli, F., Freund, J. B. & Vermot, J. Blood flow mechanics in cardiovascular development. *Cellular and molecular life sciences : CMLS* **72**, 2545-2559, doi:10.1007/s00018-015-1885-3 (2015).

40 Hinits, Y. *et al.* Zebrafish Mef2ca and Mef2cb are essential for both first and second heart field cardiomyocyte differentiation. *Developmental biology* **369**, 199-210, doi:10.1016/j.ydbio.2012.06.019 (2012).

41 Ruiz-Herguido, C. *et al.* Hematopoietic stem cell development requires transient Wnt/beta-catenin activity. *The Journal of experimental medicine* **209**, 1457-1468, doi:10.1084/jem.20120225 (2012).

42 Carra, S. *et al.* Ve-ptp modulates vascular integrity by promoting adherens junction maturation. *PloS one* **7**, e51245, doi:10.1371/journal.pone.0051245 (2012).

43 Toselli, C. M., Wilkinson, B. M., Paterson, J. & Kieffer, T. J. Vegfa/vegfr2 signaling is necessary for zebrafish islet vessel development, but is dispensable for beta-cell and alpha-cell formation. *Scientific reports* **9**, 3594, doi:10.1038/s41598-019-40136-1 (2019).

44 Koenig, A. L. *et al.* Vegfa signaling promotes zebrafish intestinal vasculature development through endothelial cell migration from the posterior cardinal vein. *Developmental biology* **411**, 115-127, doi:10.1016/j.ydbio.2016.01.002 (2016).

45 Winderlich, M. *et al.* VE-PTP controls blood vessel development by balancing Tie-2 activity. *The Journal of cell biology* **185**, 657-671, doi:10.1083/jcb.200811159 (2009).

46 Wu, B. J. *et al.* Nuclear receptor subfamily 2 group F member 1a (nr2f1a) is required for vascular development in zebrafish. *PloS one* **9**, e105939, doi:10.1371/journal.pone.0105939 (2014).

47 Mendelson, K., Zygmunt, T., Torres-Vazquez, J., Evans, T. & Hla, T. Sphingosine 1-phosphate receptor signaling regulates proper embryonic vascular patterning. *The Journal of biological chemistry* **288**, 2143-2156, doi:10.1074/jbc.M112.427344 (2013).

48 Schindler, Y. L. *et al.* Hand2 elevates cardiomyocyte production during zebrafish heart development and regeneration. *Development* **141**, 3112-3122, doi:10.1242/dev.106336 (2014).

49 Nguyen, A. T. *et al.* DOT1L regulates dystrophin expression and is critical for cardiac function. *Genes & development* **25**, 263-274, doi:10.1101/gad.2018511 (2011).

50 Medrano, J. L. & Naya, F. J. The transcription factor MEF2A fine-tunes gene expression in the atrial and ventricular chambers of the adult heart. *The Journal of biological chemistry* **292**, 20975-20988, doi:10.1074/jbc.M117.806422 (2017).

51 Lazic, S. & Scott, I. C. Mef2cb regulates late myocardial cell addition from a second heart field-like population of progenitors in zebrafish. *Developmental biology* **354**, 123-133, doi:10.1016/j.ydbio.2011.03.028 (2011).

52 Liu, J. *et al.* A dual role for ErbB2 signaling in cardiac trabeculation. *Development* **137**, 3867-3875, doi:10.1242/dev.053736 (2010).

53 Gerhart, S. V., Jefferis, R. & Iovine, M. K. Cx40.8, a Cx43-like protein, forms gap junction channels inefficiently and may require Cx43 for its association at the plasma membrane. *FEBS letters* **583**, 3419-3424, doi:10.1016/j.febslet.2009.09.054 (2009).

54 Gemel, J. *et al.* Connexin40 abnormalities and atrial fibrillation in the human heart. *Journal of molecular and cellular cardiology* **76**, 159-168, doi:10.1016/j.yjmcc.2014.08.021 (2014).

55 Gibb, N. *et al.* Hey2 regulates the size of the cardiac progenitor pool during vertebrate heart development. *Development* **145**, doi:10.1242/dev.167510 (2018).

56 Vermot, J. *et al.* Reversing blood flows act through klf2a to ensure normal valvulogenesis in the developing heart. *PLoS biology* **7**, e1000246, doi:10.1371/journal.pbio.1000246 (2009).

57 Hayashi, M. *et al.* VE-PTP regulates VEGFR2 activity in stalk cells to establish endothelial cell polarity and lumen formation. *Nature communications* **4**, 1672, doi:10.1038/ncomms2683 (2013).

58 Braun, L. J. *et al.* VE-PTP inhibition stabilizes endothelial junctions by activating FGD5. *EMBO*

reports, doi:10.15252/embr.201847046 (2019).

59 Juettner, V. V. *et al.* VE-PTP stabilizes VE-cadherin junctions and the endothelial barrier via a phosphatase-independent mechanism. *The Journal of cell biology* **218**, 1725-1742, doi:10.1083/jcb.201807210 (2019).

60 Mellberg, S. *et al.* Transcriptional profiling reveals a critical role for tyrosine phosphatase VE-PTP in regulation of VEGFR2 activity and endothelial cell morphogenesis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **23**, 1490-1502, doi:10.1096/fj.08-123810 (2009).

61 Dejana, E., Tournier-Lasserve, E. & Weinstein, B. M. The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications. *Developmental cell* **16**, 209-221, doi:10.1016/j.devcel.2009.01.004 (2009).

62 Dominguez, M. G. *et al.* Vascular endothelial tyrosine phosphatase (VE-PTP)-null mice undergo vasculogenesis but die embryonically because of defects in angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3243-3248, doi:10.1073/pnas.0611510104 (2007).

63 Jia, W., Yao, Z., Zhao, J., Guan, Q. & Gao, L. New perspectives of physiological and pathological functions of nucleolin (NCL). *Life sciences* **186**, 1-10, doi:10.1016/j.lfs.2017.07.025 (2017).

64 Huang, Y. *et al.* The angiogenic function of nucleolin is mediated by vascular endothelial growth factor and nonmuscle myosin. *Blood* **107**, 3564-3571, doi:10.1182/blood-2005-07-2961 (2006).

65 Watanabe, T. *et al.* Nucleolin as cell surface receptor for tumor necrosis factor-alpha inducing protein: a carcinogenic factor of Helicobacter pylori. *Journal of cancer research and clinical oncology* **136**, 911-921, doi:10.1007/s00432-009-0733-y (2010).

Figures and extended data:

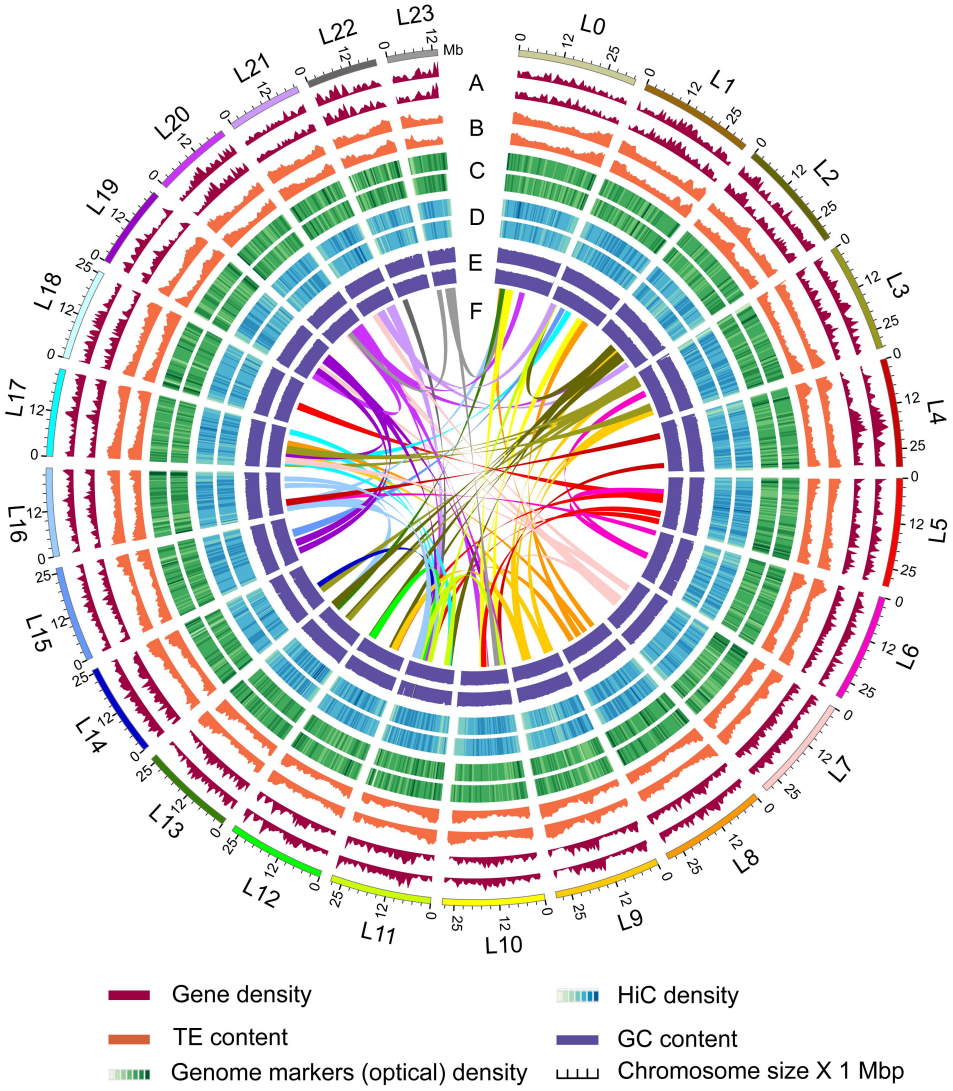


Figure 1. Overview of golden pompano. Numbers on the circumference are at the megabase scale. **A** Gene density of female *Trachinotus ovatus* (window size = 500 Kb). **B** TE content density of female *T. ovatus* (window size = 500 Kb). **C** Genome markers (optical) density of female *T. ovatus* (window size = 500 Kb). **D** Hi-C depth of female *T. ovatus* (window size = 500 Kb). **E** GC content of female *T. ovatus* (window size = 500 Kb). **F** Color bands in the middle of the Circos plot connect segmental duplication (minimum five gene pairs) from Teleost-specific whole genome duplication (Ts3R) events.

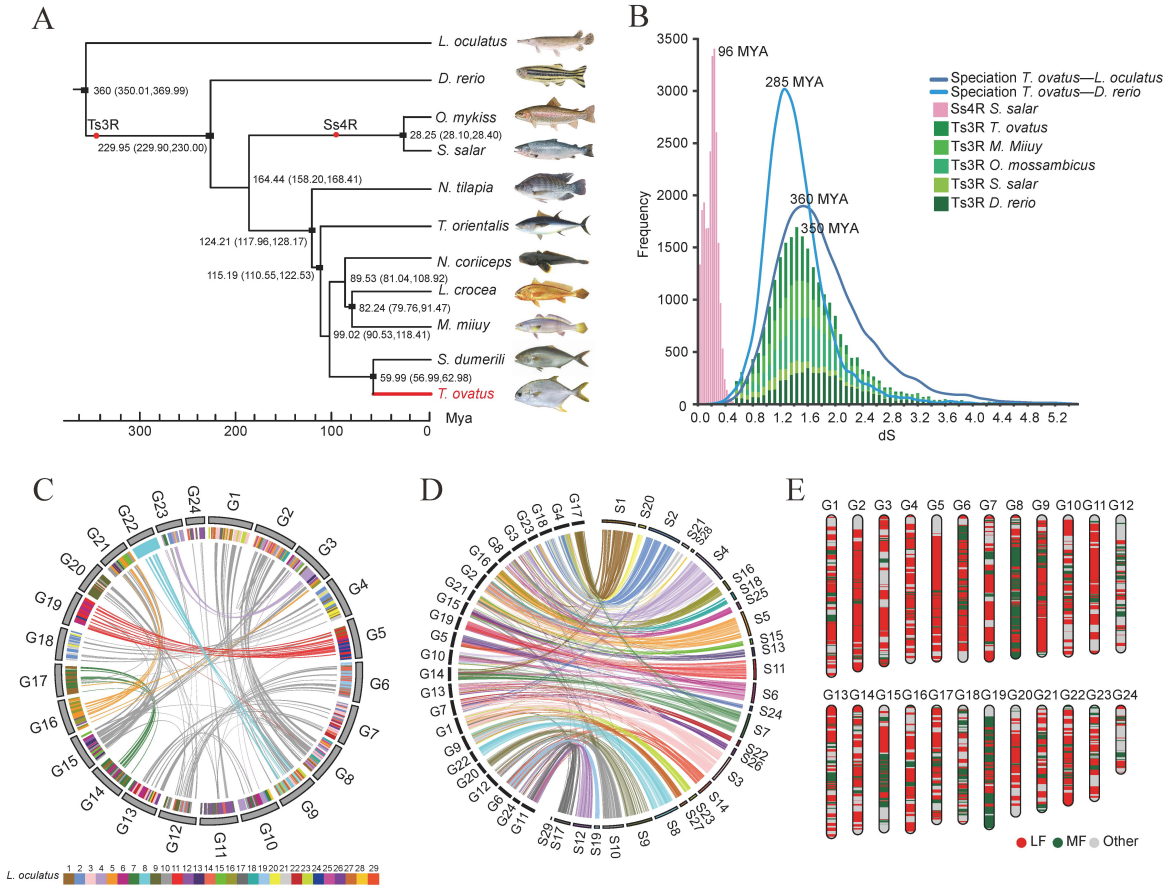


Figure 2. Genome evolution of golden pompano. A Phylogenetic relationship of Perciformes and relevant teleost lineages. The position of golden pompano is highlighted in red. Red circles represent the Teleost specific whole genome duplication (Ts3R), Salmonid-specific whole genome duplication (Ss4R), respectively. The divergence time was estimated using the nodes with calibration times derived from the Time Tree database, which were marked by a black rectangle. All estimated divergence times are shown with 95% confidence intervals in brackets. **B** Inspection of whole genome duplication events based on synonymous mutation rate (Ks) distribution. The x axis shows the synonymous distance until a Ks cut-off of 5.2. Note that in order to represent all the data on the same frequency scale, bin sizes are different for each data set. **C** Internal genome synteny of golden pompano. Double-conserved synteny between the golden pompano and spot gar genomes. Each spot gar chromosome (represented as colored blocks) is mostly syntenic with two different chromosomes in the golden pompano genome (syntenic golden pompano regions represented

by different colors according to spot gar chromosomal location), a pattern typically associated with whole-genome duplication (Ts3R). Pairs of paralogous genes in spotted gar that are inserted in a double-conserved synteny block are consistent with an origin at the Ts3R event (ohnologues), while genes that are inserted in a double-conserved synteny block but have no paralogue are singletons that have lost their duplicate copy since the Ts3R event. Only genes anchored to chromosomes are represented. **D** Macro-synteny comparison between spotted gar and golden pompano shows the overall one-to-two double-conserved synteny relationship between spotted gar to a post-Ts3R teleost genome. **E** Component of less fragment (LF) and major fragment (MF) subgenomes within golden pompano genome.

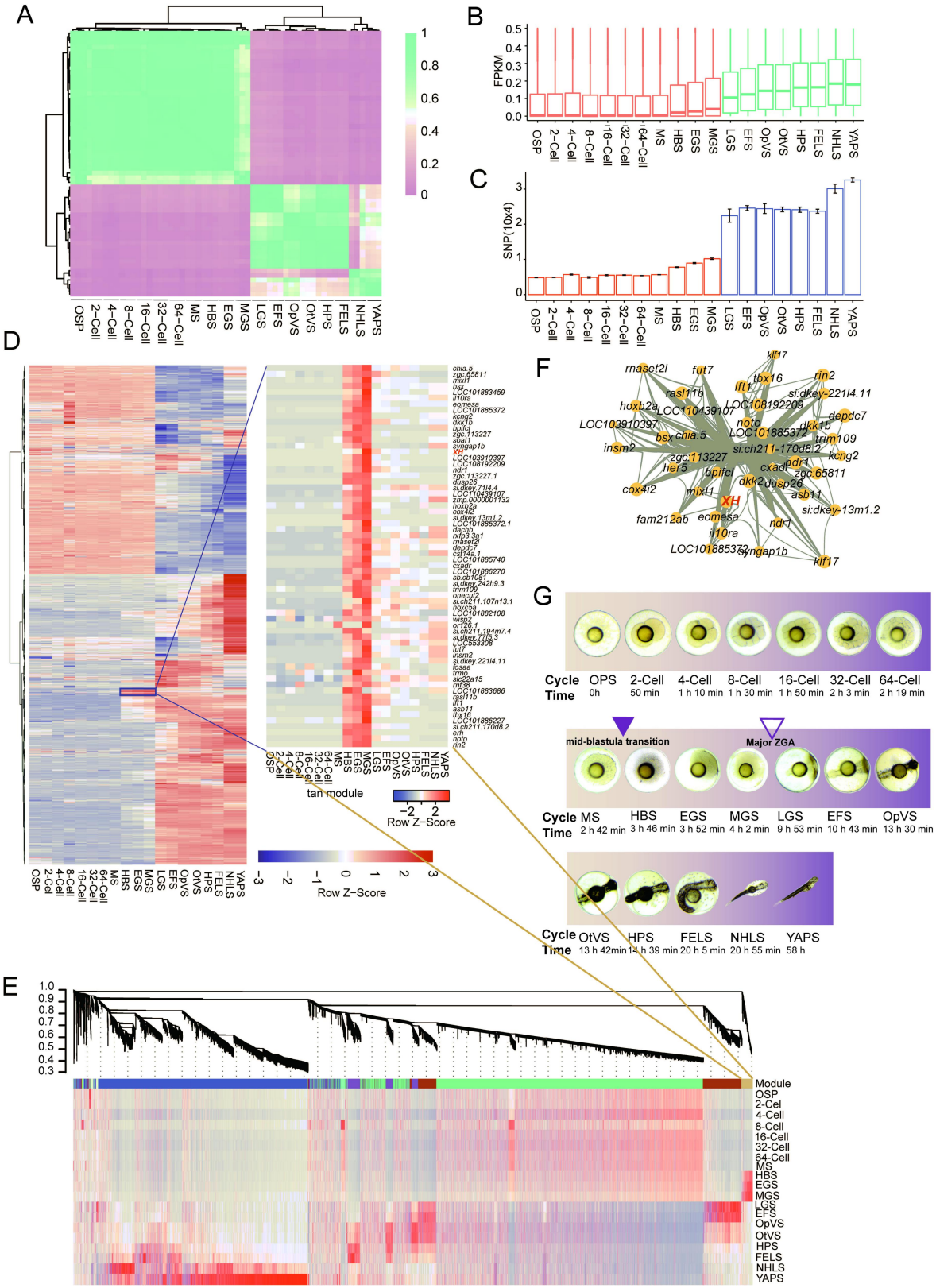


Figure 3. Characterization of Nineteen Continuous Embryonic Development stages from OSP to YAPS. A Heatmap of correlation coefficient within transcriptomics experiments of

embryonic development. Expression profile of transcriptomic experiments among embryonic development process. **B** Landscape of transcriptomic expression for 19 embryonic development stages. Cluster of stage-specifically expressed genes were highlighted in red rectangle. **C** Accumulation of transcriptome SNPs within embryonic developments. Error bar represents the standard deviation of biological replicates. **D** Expanded view of expression of all genes in purple modules. The purple model is associated with specific actively expressed genes in minor-ZGA process. **E** WGCNA analysis of Embryonic Development Stages Revealed Gene-Network Modules Enriched in D. **F** Hub-gene network of the purple module. Size of the dots represents hubness. Color of the dots represents the increasing expression level from low to high in preparation stage for ZGA (MGS). Bold text highlights the genes known for *XH* (EVM0008813). **G** Timing of minor-ZGA and major-ZGA events. The photos of embryonic development process from OSP to YAPS were observed using microscope. The activation of minor-ZGA is beginning at HBS stage and continued to MGS stage. The major-ZGA is activated during transformation stages between MGS and LGS process.

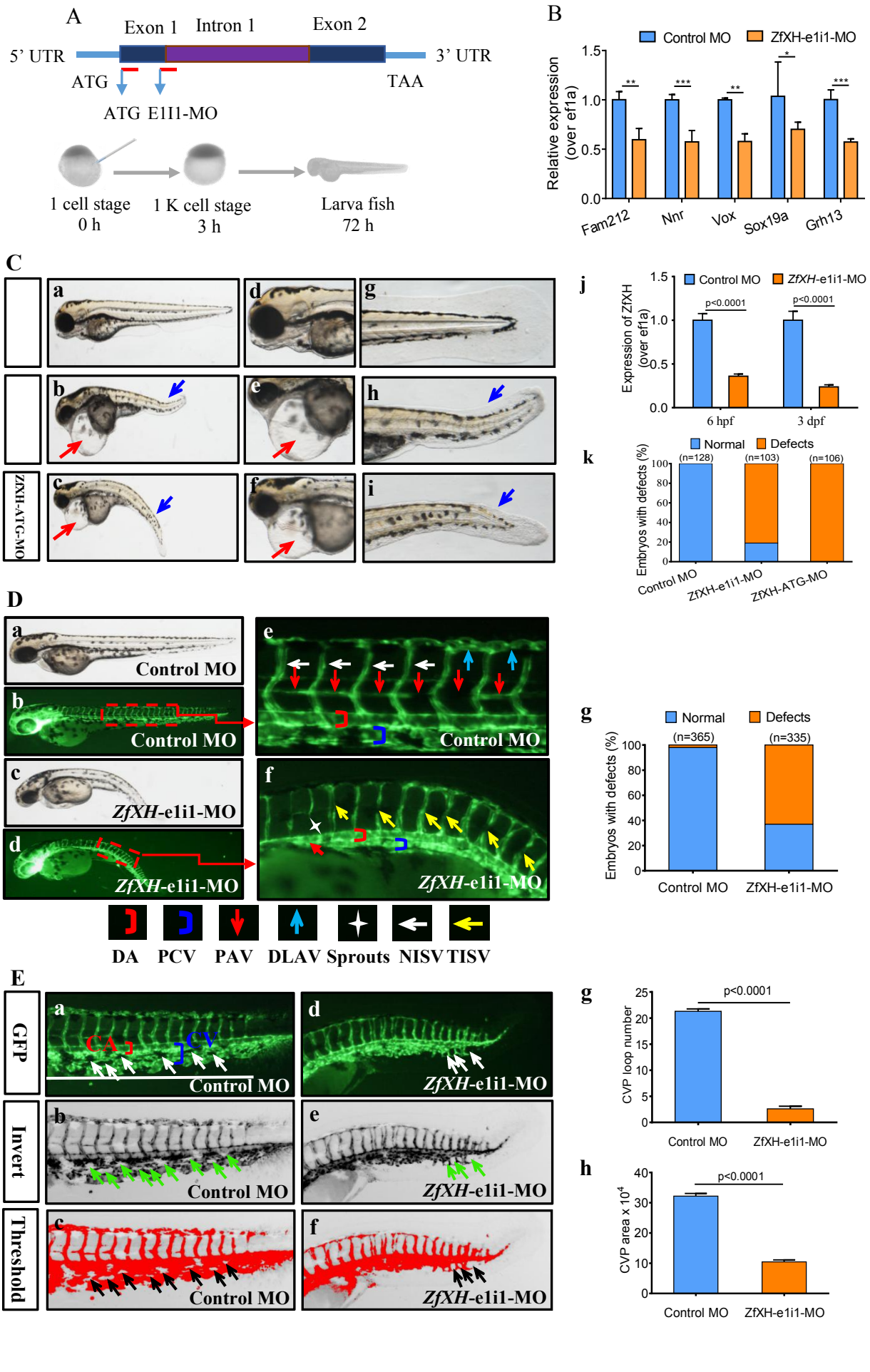


Figure 4. Loss of *ZfXH* gene affects ZGA and causes morphological and vascular defects. **A** Schematic map of *ZfXH* gene and the MO design and micro-injection strategy. **B** Knockdown of *ZfXH* significantly decreases the expression of ZGA marker genes. **C** (**a-i**) Gross morphology at 3 dpf in wild-type AB strain. Compared with control MO, knock down *ZfXH* present pericardial oedema (**b**, **e, c, f**, red arrow) and caudal fin defects (**b, c**, **h, i**, blue arrow). Heart beat and circulation in caudal vein (CV) is visible in the control fish, but is abnormal in *ZfXH*-MO injected fish (Supplementary Movie1, Supplementary Movie2). (**j**) Validation of MO against *ZfXH*. Endogenous *ZfXH* in control and *ZfXH* morphants assessed by qPCR. Samples were collected at 6 hpf after introduction of 4ng of MO at the one-cell stage (N=30). (**k**) The bar graph shows the percentage of embryos with development defects after knockdown of *ZfXH* with *ZfXH*-e1i1-MO *and* *ZfXH*-ATG-MO. **D** Morpholino knockdown of *ZfXH* causes vascular defects. *Tg(fli1a:EGFP)*^{y1} zebrafish embryos were injected with 4ng control MO or 4 ng *ZfXH*-e1i1-MO. **(a-f)** Representative bright field and fluorescent images of *Tg(fli1a:EGFP)*^{y1} embryos at 52 hpf. **(b, e)** Image of trunk regions taken at 52 hpf, with the vascular structures visualized by GFP fluorescence and labelled ISV (intersegmental vessel) and DLAV (dorsal longitudinal anastomotic vessel) showed regular development in the embryo injected with control MO. Compared with control MO, embryos injected with *ZfXH*-e1i1-MO present thinner ISVs (yellow arrows) and ectopic sprouts (asterisk) of dorsal aorta **(d, f)**. In control embryos, the parachordal vessels (PAV) form normally (**e**, red arrows). Compared with control, MO knock down *ZfXH* prevents the PAV formation, the precursor to the lymphatic system. **g**, The bar graph shows the percentage of embryos with vascular defects after knockdown of *ZfXH* with *ZfXH*-e1i1-MO. The boxed regions are shown at higher magnification in the right panels. hpf, hours post fertilization. **E** *ZfXH* knockdown impairs formation of the CVP in zebrafish. In control embryos, caudal vein plexus (CVP) were formed honeycomb-like structures at the tail around 52 hpf (**a-c**, arrowheads). In contrast, *ZfXH* knockdown resulted in specific defects in CVP formation (**d-f**). **g,** Quantification of loop formation at CVP shows a 8.2-fold decreased in *ZfXH*-e1i-MO injected embryos at 52 hpf. **h,** Quantification of area at CVP shows a 3-fold decreased in

ZfXH-e1i-MO injected embryos at 52 hpf. CA, caudal artery; CV, caudal vein. NISV, normal intersegmental vessel; TISV, thinner intersegmental vessel.

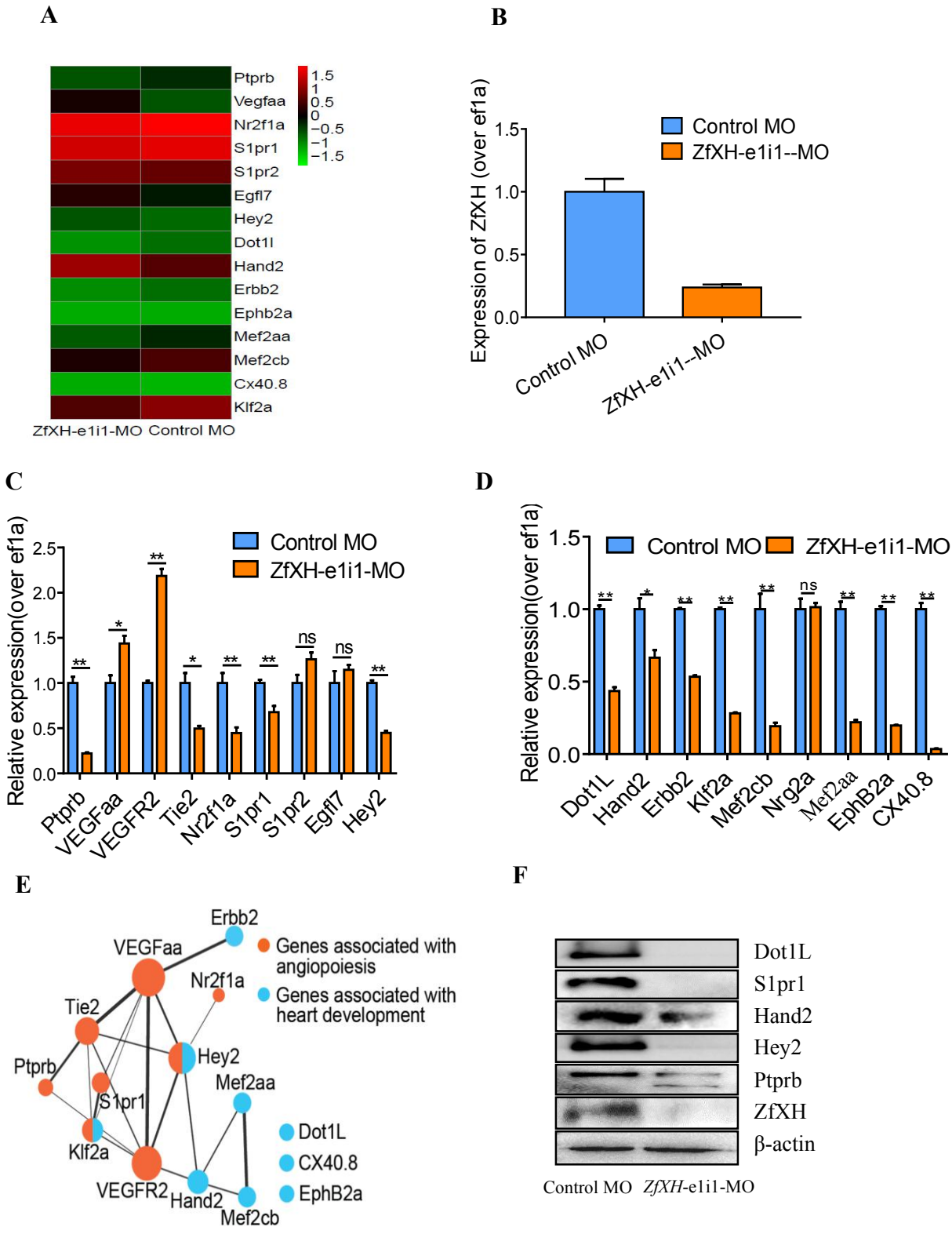


Figure 5. Loss of *ZfXH* results in the down-regulation of genes associated with vascular and heart development. A Heatmap of the 15 selected genes from zebrafishes after injection

of 4ng *ZfXH* ^{e1i1} MO at 3 dpf examined by RNA-seq. **B** Expression of *ZfXH* post injection of *ZfXH* ^{e1i1} MO 3 dpf. **C** Expression of genes associated with angiopoiesis post injection of *ZfXH* ^{e1i1} MO 3 dpf using QPCR. The data represent as mean±SEM from three independent experiments. **D** Expression of genes associated with heart development post injection of *ZfXH* ^{e1i1} MO 3 dpf using QPCR. The data represent as mean±SEM from three independent experiments. **E** Networks of the genes previously reported to be associated with angiopoiesis and heart development. Cytoscope V3.6.1 was used to build this network. **F** Protein levels of the selected genes associated with angiopoiesis and heart development post injection of *ZfXH* ^{e1i1} MO 3 dpf by using Western blotting. β-actin antibody was used as internal control.

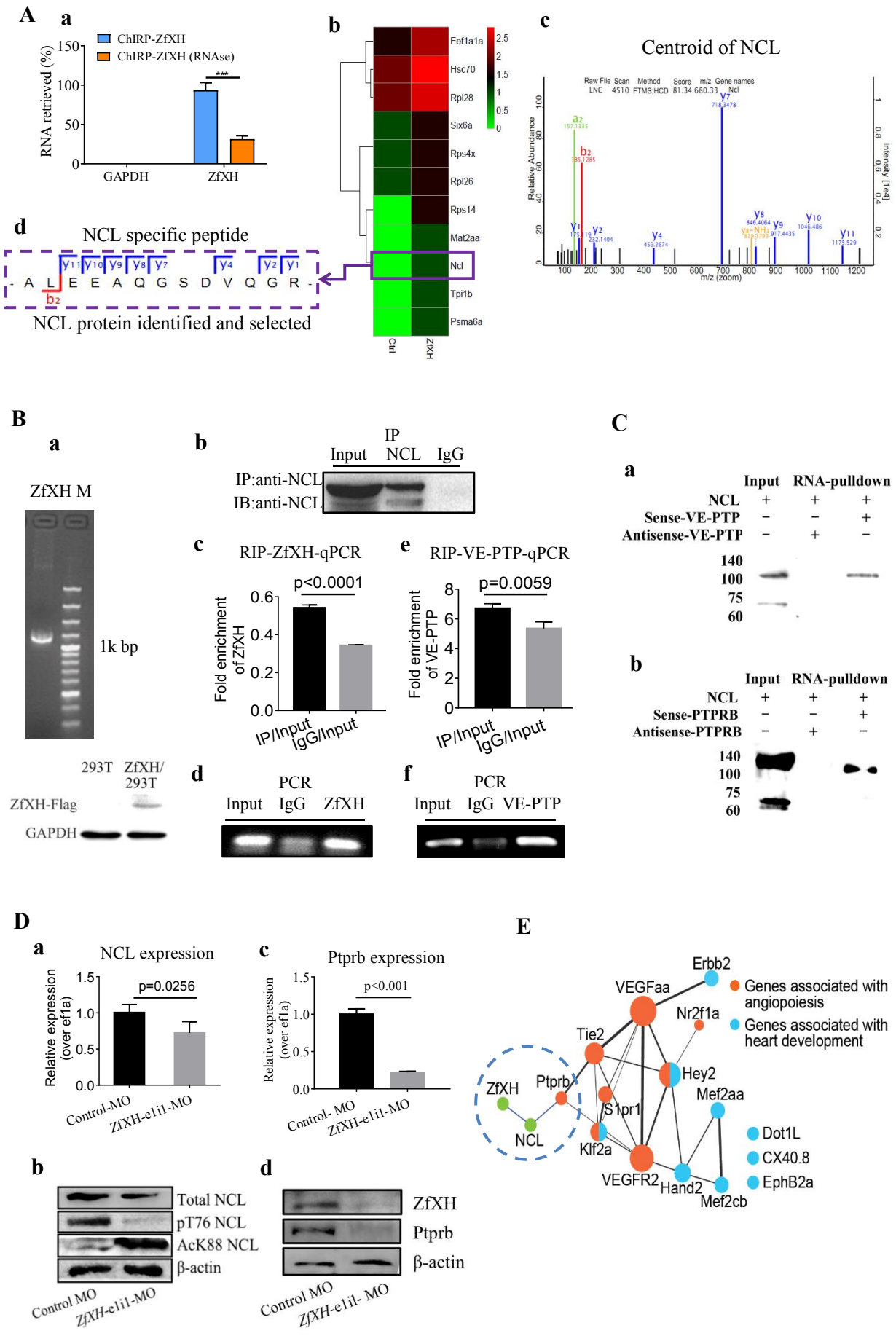
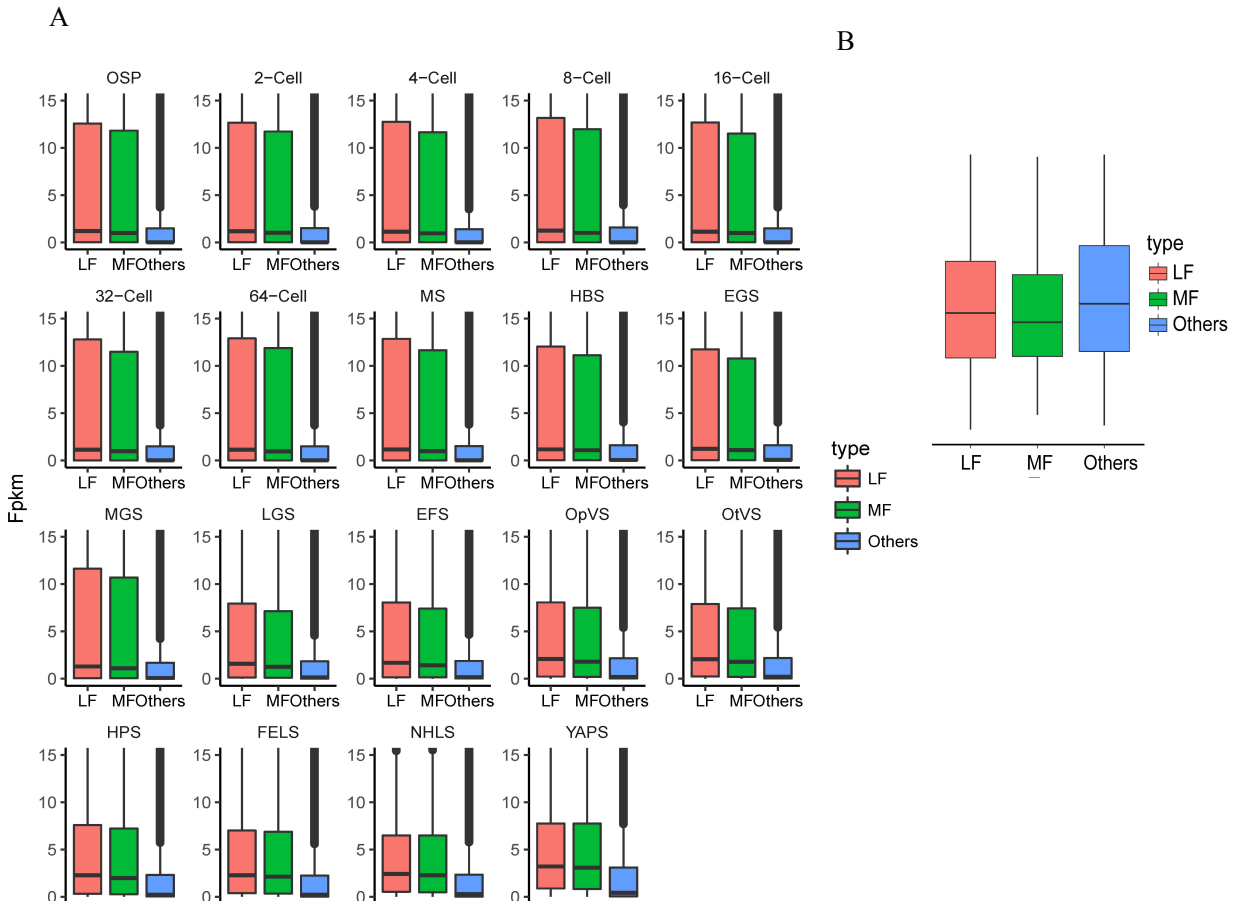
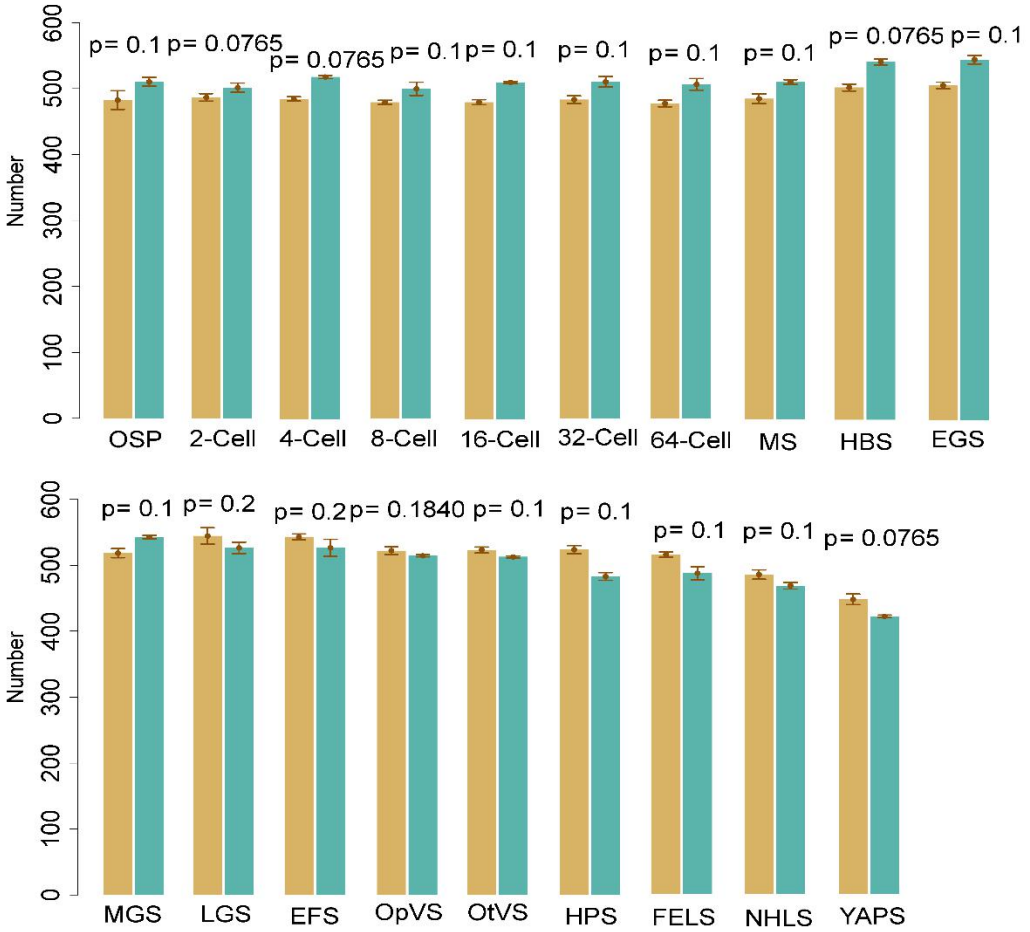


Figure 6. Mechanism of *ZfXH* on angiopoiesis development of zebrafish. A ChIRP-MS identification of *ZfXH* RNA binding proteins. (**a**) qPCR identification of *ZfXH* RNA in the eluted RNAs. Graph shows more than 90% *ZfXH* RNA was retrieved, and no GAPDH was detected. (**b**) Heat map of major proteins enriched and significantly (change fold >2 and $p<0.05$) retrieved by *ZfXH* and control probes, analyzed by LC/MS-MS. Among them, NCL protein (purple boxed) was selected as candidate for follow-up study because of its broad association with angiopoiesis. (**c**) Centroid of NCL protein. NCL protein was pull down and identified by LC/MS-MS and centroid of NCL shows its unique m/z feature. The specific peptide (**d**) identifies NCL protein. **B** Physiological interaction between *ZfXH,* VE-PTP mRNA and NCL protein, respectively. (**a**) The mRNA expression of *ZfXH* was determined by qPCR and Western blotting against Flag antibody was used to identify the successful expression of pcDNA3.1- Flag-ZfXH plasmid in 293T cells. (**b**) Immunoprecipitation (IP) against NCL antibody followed the Western blotting against NCL antibody was performed to verify whether NCL protein was pulled down successfully. (**c, d**) The interaction between ZfXH mRNA and NCL protein. qPCR and general PCR gel detection for *ZfXH* mRNA expression in the NCL-pulled down RNA. The data represent as mean±SEM from three independent experiments. (**e, f**) The interaction between VE-PTP mRNA and NCL protein. qPCR and general PCR gel detection for VE-PTP mRNA expression in the NCL-pulled down RNA. The data represent as mean±SEM from three independent experiments. **C** Physiological interaction between NCL protein and VE-PTP and *Ptprb* mRNA, respectively. (**a**) The interaction between VE-PTP mRNA and NCL protein. Western blotting was performed to detect NCL protein in the VE-PTP-biotin probe -pulled down proteins in 293T cells. (**b**) The interaction between *Ptprb* mRNA and NCL protein. Western blotting was performed to detect NCL protein in the Ptprb-biotin probe -pulled down proteins in zebrafish tissues. **D** Loss of *ZfXH* affects the expression of nucleolin (NCL) at both mRNA and protein levels. (**a**) The mRNA expression of NCL was determined by qPCR. The data represent as mean±SEM from three independent experiments. (**b**)The total NCL protein, phosphorylated NCL and acetylated NCL were detected by Western blotting using specific NCL antibodies. (**c**) The

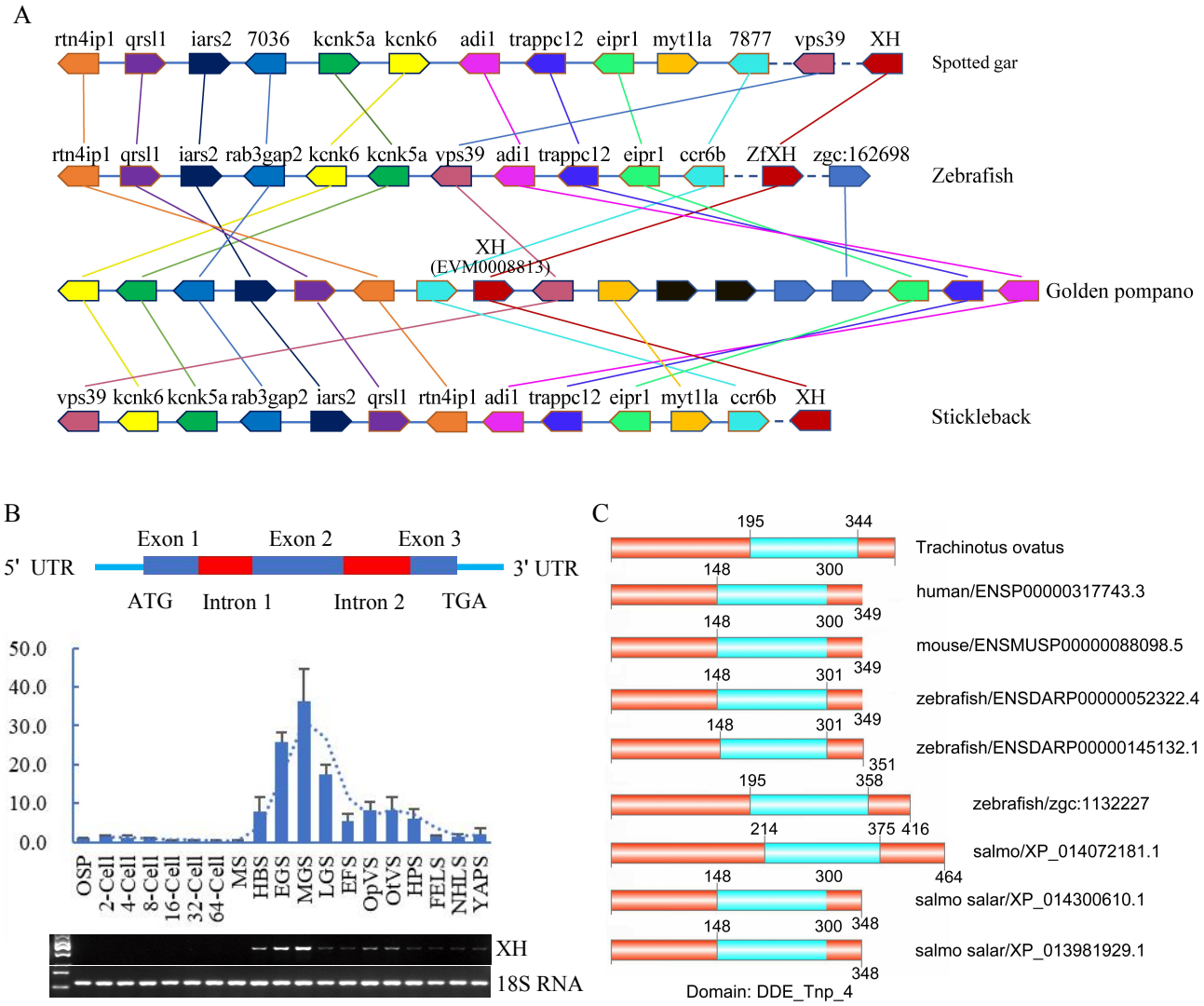
mRNA expression of *ptprb* was determined by qPCR. The data represent as mean±SEM from three independent experiments. (**d**) The total ZfXH and Ptprb protein were detected by Western blotting using specific ZfXH and Ptprb antibodies. **E** Schematic model illustrating the mechanism of *ZfXH* in zebrafish angiopoiesis and heart development. Interaction map of the proteins associated with angiopoiesis and heart development in zebrafish in this study. The interactions between *ZfXH* mRNA and NCL protein, NCL protein and ptprb mRNA are new-found interactions in this study. *ZfXH* may affect the angiopoiesis and heart development via the regulation of the network present.



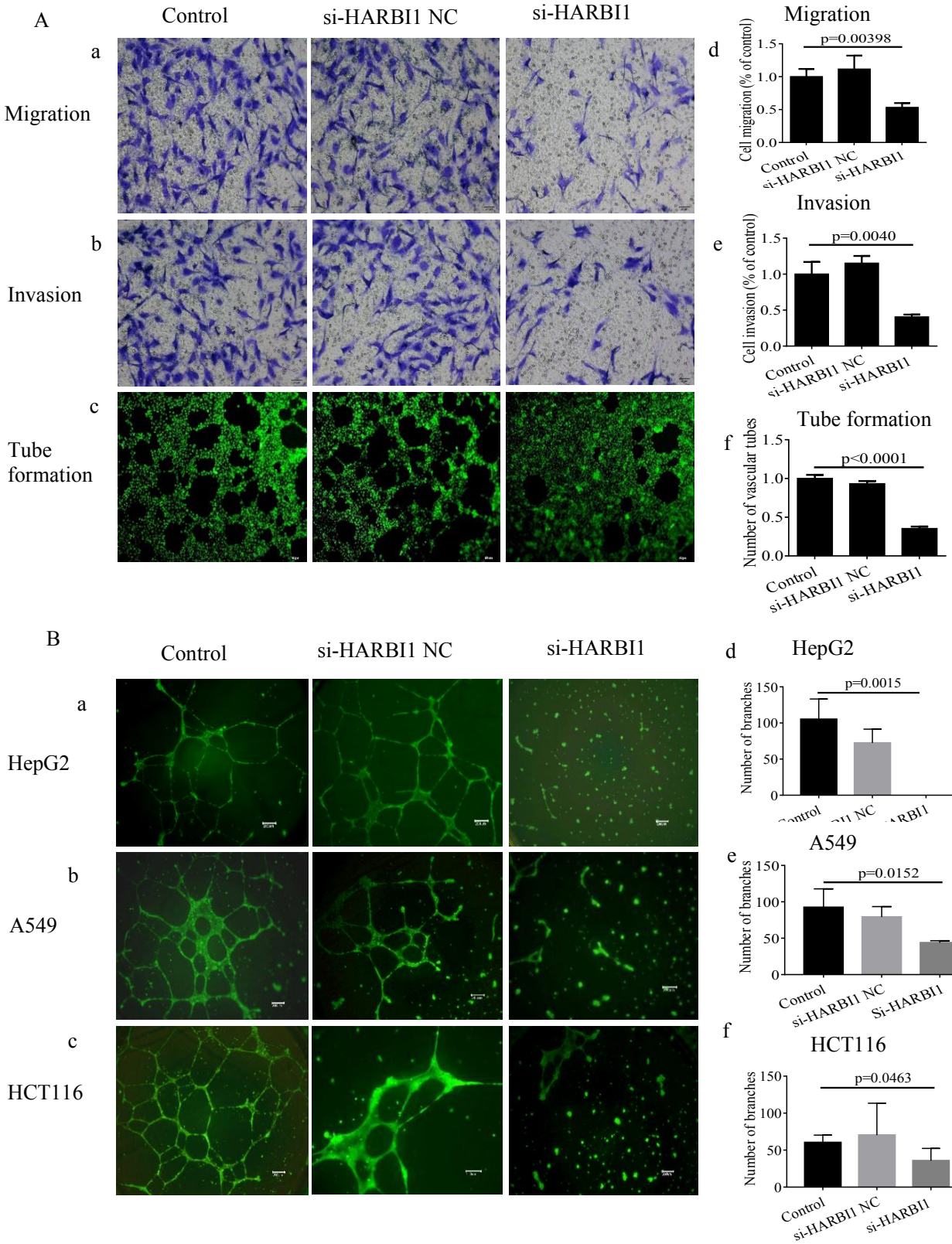
Extended data 1. Expression Divergence and Selection Bias among Ancestral Subgenomes. A Boxplot of expression level of LF, MF and Other gene sets. **B** Selection bias associated with ancestral subgenomes fragmentation. The Ka/Ks values were calculated by orthologous pairs between golden pompano and spotted gar which is outgroup species without Ts3R genome duplication events.



Extended data 2. Dynamic dominantly expressed genes within LF and MF subgenomes during embryonic development process. Genes whose expression level (FPKM) >= 0.1 are defined as expressed genes. Gene pair having at least one expressed gene is considered for expression bias analysis. Trend of dominant genes is inverse from LGS stage, which is the beginning time of major-ZGA event.

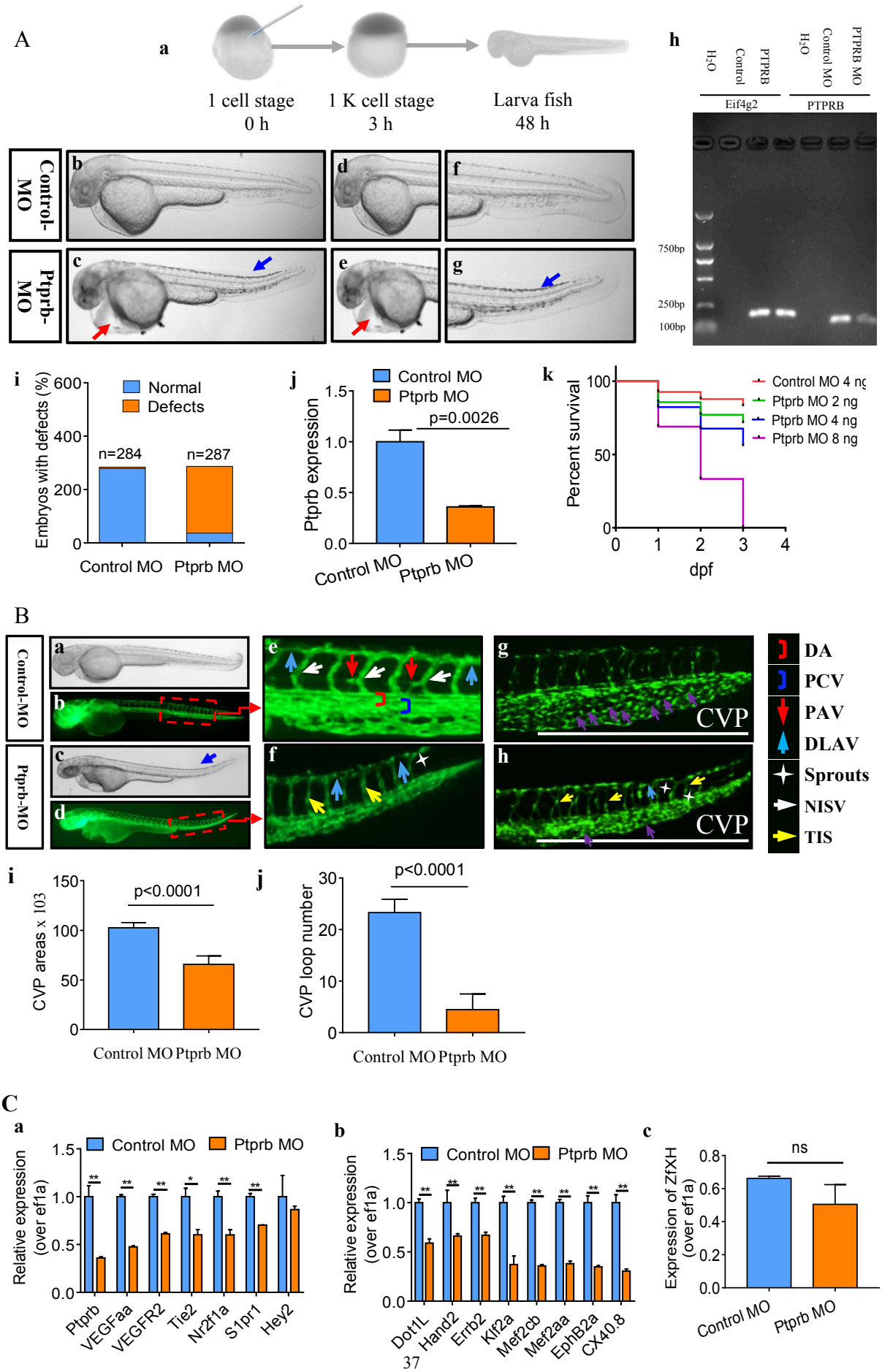


Extended data 3. Characterization of *XH* gene (EVM0008813). A Micro-synteny analysis of *XH* locus among spotted gar, zebrafish, gold pompano and stickleback. Two inversions and one insertion occurred in *XH* locus region of golden pompano genomes. **B** Validation of expression level for *XH* by QPCR technology. 18s RNA was considered as internal marker. Gene structure of *XH* was showed at upper region. **C** Domains of *XH* and other homologous protein. The domains were identified in SMART database (http://smart.embl.de/).



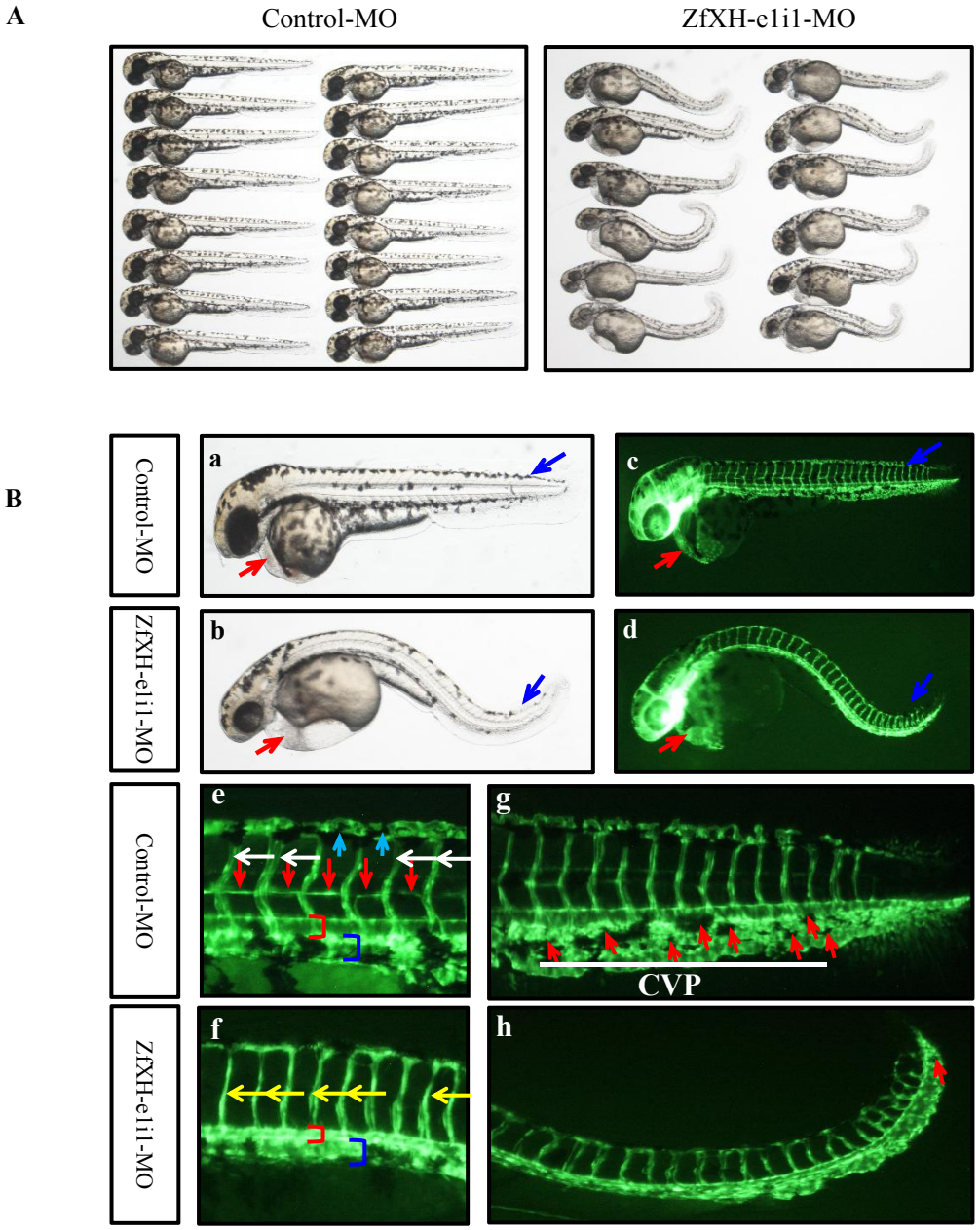
Extended data 4. Silencing of HARBI1 decreases the invasion and tube formation abilities of HUVEC cells and angiogenesis of cancer cell lines. A Silencing of HARBI1 gene suppresses HUVEC cell migration, invasion and tube formation abilities. The migratory

and invasive potential of HUVECs treated with si-HARBI1 was determined using transwell chambers as described in the "Materials and methods" section. (**a, d**) Representative images of migrated cells stained with crystal violet and inhibition of migration in si-HARBI1 treated HUVEC cells. The data represent as mean±SEM from three independent experiments. (**b, e)** Representative images of invaded cells stained with crystal violet and inhibition of invasion in si-HARBI1 treated HUVEC cells. The data represent as mean±SEM from three independent experiments. Scale bars, 20 μm. (**c, f**) Representative images of tube formation in si-HARBI1 treated HUVEC cells and inhibition of tube formation in vitro. Images were taken 6 h after addition of the supernatant. Tube networks were quantified using the Image J software and expressed as branches. The data represent as mean±SEM from three independent experiments. Scale bars, 50 μm. **B** Silencing of HARBI1 gene inhibits angiogenesis of three cancer cell lines in vitro. HUVEC cells were plated on Matrigel-coated plates at a density of 1.5 × 105 cells/well and incubated for 24 h before the cell culture supernatant of si-HARBI1 treated hepatocellular carcinoma (HepG2), non-small-cell lung carcinoma (A549) and colon carcinoma (HCT116) cell lines was added, respectively. (**a, b, c**) Representative images of tube formation in si-HARBI1 treated HUVEC cells. (**d, e, f**) shows the inhibition of angiogenesis of cancer cell lines in vitro. The data represent as mean±SEM from three independent experiments. Scale bars, 100 μm. Images were taken 24h after addition of the supernatant. Tube networks were quantified using the ImageJ software and expressed as branches.



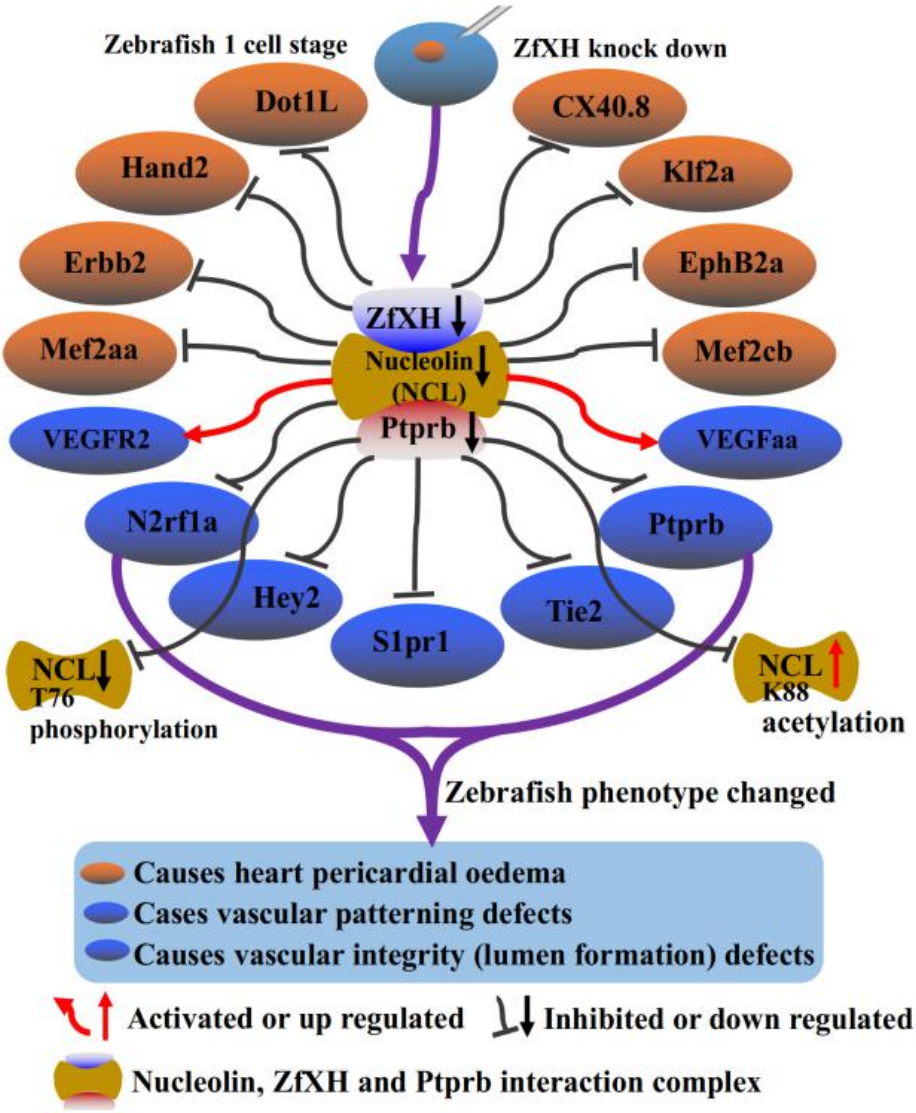
Extended data 5. Loss of *Ptprb* gene causes morphological and vascular defects. A Loss of *Ptprb* induces morphological defects. (**a**) Schematic map the micro-injection strategy. (**b-g**) Gross morphology at 48 hpf. Compared with control MO, knock down *Ptprb* present pericardial oedema (**c, e**, red arrow) and caudal fin defects (**c, e**, blue arrow). (**h**) The gel electrophoresis of PCR products of *Ptprb* gene in control and *Ptprb* MO groups. Compared with control MO, *Ptprb* MO shows quite weak expression of *Ptprb* gene. (**i**) The bar graph shows the percentage of embryos with development defects after knockdown of *Ptprb*. (**j**) Validation of MO against *Ptprb*. Endogenous *Ptprb* in control and *Ptprb* morphants assessed by qPCR. Samples were collected at 48 hpf after introduction of 4ng of MO at the one-cell stage (N=20). The data represent as mean±SEM from three independent experiments. (**k**) Percent survival in control vs. *Ptprb* morphants for 3 days. Compared with control MO, embryos with injection of 4ng *Ptprb* MO survive more than 8 ng *Ptprb* MO with similar phenotype presentation. hpf, hours post fertilization. **B** Morpholino knockdown of *Ptprb* causes vascular defects. *Tg(fli1a:EGFP)*^{y1} zebrafish embryos were injected with 4ng control MO or 4ng *Ptprb* MO. **(a-f)** Representative bright field and fluorescent images of *Tg(fli1a:EGFP)*^{y1} embryos at 48 hpf. **(b, e)** Image of trunk regions taken at 48 hpf, with the vascular structures visualized by GFP fluorescence and labelled ISV (intersegmental vessel) and DLAV (dorsal longitudinal anastomotic vessel) showed regular development in the embryo injected with control MO. Compared with control MO, embryos injected with *ptprb* MO present thinner ISVs (yellow arrows) and ectopic sprouts (asterisk) of dorsal aorta **(e, f)**. MO knock down *Ptprb* prevents the parachordal vessels (PAV) formation. In control embryos, the parachordal vessels (PAV) form normally (**e**, red arrows). The boxed regions are shown at higher magnification in the right panels. hpf, hours post fertilization. (**g, h)** *Ptprb* knockdown impairs formation of the CVP in zebrafish. In control embryos, caudal vein plexus (CVP) were formed honeycomb-like structures at the tail around 48 hpf (purple arrow). In contrast, *ptprb* knockdown resulted in specific defects in CVP formation (**h**). (**i)** Quantification of loop formation at CVP shows a 5.2-fold decreased in *Ptprb* MO injected embryos at 48 hpf. (**j)** Quantification of area at CVP shows a 1.5-fold decreased in *Ptprb* MO injected embryos at 48 hpf. The data represent as mean±SEM from three independent experiments. CVP, caudal vein

plexus; CA, caudal artery; CV, caudal vein. **C** Expression of genes associated with angiopoiesis (a) and heart development (b) post injection of *Ptprb* MO 48 hpf using QPCR. (c) Expression of *ZfXH* post injection of *Ptprb* MO 48 hpf. The data represent as mean±SEM from three independent experiments.



Extended data 6. Silence of *ZfXH* gene leads to identical hear and vascular phenotypes to that in Figure 4. A Loss of *ZfXH* induces morphological defects. Compared with control MO, knock down *ZfXH* present pericardial oedema (red arrow) and caudal fin defects (blue arrow). Samples were collected at 3 dpf after introduction of 4 ng of MO at the one-cell stage (N=20). **B** Morpholino knockdown of *ZfXH* causes vascular defects. *Tg(fli1a:EGFP)*^{y1} zebrafish embryos were injected with 4ng control MO or 4ng *ZfXH* MO. **(a, b)** representative bright field images and (c, d) representative fluorescent images of *Tg(fli1a:EGFP)*^{y1} embryos at 3 dpf. **(e, f)** Image of trunk regions taken at 3 dpf, with the vascular structures visualized by

GFP fluorescence and labelled ISV (intersegmental vessel) and DLAV (dorsal longitudinal anastomotic vessel) showed regular development in the embryo injected with control MO. Compared with control MO, embryos injected with *ZfXH* MO present thinner ISVs (yellow arrows). **(g, h)** *ZfXH* knockdown impairs formation of the CVP in zebrafish. In control embryos, caudal vein plexus (CVP) were formed honeycomb-like structures at the tail around 3 dpf (red arrow). In contrast, *ZfXH* knockdown resulted in specific defects in CVP formation. CA, caudal artery; CV, caudal vein. These phenotypes above were identical to that in Figure 4.



Extended data 7. Possible mechanism of *ZfXH* in zebrafish angiopoiesis and heart development. Knockdown of *ZfXH* may downregulate the ZfXH-NCL-ptprb complex, subsequently regulate the proteins associated with angiopoiesis and heart development, and finally result in heart pericardial oedema, vascular patterning and integrity (lumen formation)

defects.

Supplementary figure legends

Figure S1. **Investigation of genome features for Golden Pompano. A** K-mer analysis for genome size estimation (k-mer = 21). The estimated genome size is approximate 657 Mb with a heterozygosity rate of 0.19%. The genome size was estimated by using the formula: Genome size = K-mer num/Peak depth, and the heterozygosity rate causes a sub-peak at a position half of that of the main peak, whereas a certain repeat rate can cause a similar peak at the position of multiple integers of the main peak. **B** Distribution of GC contents among different sequencing depth. The x-axis represents GC content of 10 kb non-overlapping window.

Figure S2. Distribution of quality and length score of Pacbio subreads. Black line indicated accumulation. A Histogram plot of quality score of SMRT subreads. **B** Histogram plot of read length of SMRT subreads.

Figure S3. Heatmap of interaction frequency of Hi-C data. The Hi-C data was generated from muscle tissue.

Figure S4. Gene families evolution of Golden Pompano. A Venn diagram of gene family comparison in five fish species. **B** Gene family expansion and contraction of golden pompano. Gains and losses are indicated along branches and nodes.

Figure S5. Expression profile of introns in Golden Pompano embryo development. 19 embryo development stages were collected and transcriptome-sequenced. The FPKM of introns were used to draw the heatmap. A boundary lines between OSP-MGS and LGS-YAPS stages.

Figure S6. QPCR results of the 8 selected genes in five embryo development stages. 8 genes in 5 embryo stages (between MS and LGS) were randomly picked to verify the gene expression profiles perfected by sequencing. Total RNA was extracted from embryos and reverse transcribed using the the PrimeScript RT reagent Kit. Quantification of gene expression was performed in triplicates using Bio-rad iQ SYBR Green Supermix (Bio-rad) with detection on the Realplex system (Eppendorf). Relative gene expression quantification was based on the comparative threshold cycle method ($2^{-\Delta\Delta Ct}$) using 18s RNA as

endogenous control gene.

Figure S7. Confirmation of *ZfXH* gene knockdown. 4 ng morphants per fish embryo was injected. **A** The gel electrophoresis of PCR products of *ZfXH* gene in control and *ZfXH*-MO groups. Compared with control MO, *ZfXH* MO shows quite weak expression of *ZfXH* gene. **B** A time-course plot of percent survival in control vs. *ZfXH* morphants for 3 days. hpf, hours post fertilization; dpf, days post fertilization.

Figure S8. Phenotypes of *ZfXH* zebrafish morphants. 4 ng morphants per fish embryo was injected. **A** Gross morphology at 3-dpf. Compared with control MO, knock down *ZfXH* present pericardial oedema (red arrow) and caudal fin defects (blue arrow). **B** Representative bright field and fluorescent images of *Tg(fli1a:EGFP)*^{y1} embryos at 52-hpf after injection of *ZfXH* morphants.

Figure S9. Gene expression profile post knockdown of ZfXH comparing with control. (A). Heatmap of the gene clustering after knockdown of ZfXH gene comparing with control. (B). Scatter of the differential expressed genes after injection of ZfXH morphants.

Figure S10. KEGG enrichment of genes differentially expressed after knockdown of ZfXH gene comparing with control.

Figure S11. Phenotypes of *Ptprb* zebrafish morphants. 4 ng morphants per fish embryo was injected. **A** Gross morphology at 48-hpf. Compared with control MO, knock down *Ptprb* present pericardial oedema (red arrow) and caudal fin defects (blue arrow). **B** Representative bright field and fluorescent images of *Tg(fli1a:EGFP)*^{y1} embryos at 48-hpf after injection of *Ptprb* morphants.

Figure S12. Identification of pcDNA3.1-ZfXH expression plasmid. A PCR amplification of Z fXH gene (1267 bp, including enzyme cutting sites NheI and KpnI, and protective bases). **B** Four monoclones were selected by PCR and used for sequencing identification. **C** Sequencing result of Z fXH gene. The green underline is the fragment inserted into the vector, the vector sequence is on both sides. The red underline is the sequencing result of enzyme cutting site, which shows that the target sequence has been correctly constructed into the vector.

Figure S13. Identification of pcDNA3.1-ZfXH-Flag expression plasmid. A PCR amplification of ZfXH-Flag gene（1306 bp, including enzyme cutting sites NheI and KpnI,

and protective bases ） based on pcDNA3.1-ZfXH constructs. **B** Sequencing result of ZfXH-Flag fragment. The green underline is the fragment inserted into the vector, which shows that the target sequence has been correctly constructed into the vector.

Figure S14. SDS-PAGE electrophoresis of Ptprb and ZfXH protein. A Ptprb and ZfXH protein was expressed in pet-b2m Rosetta expression system with kanamycin resistance, respectively. Ptprb (33 kDa) and ZfXH (62 kDa) was successfully expressed and the purity of the recombinant protein was 85%. The purified protein was used to immunize rabbits to produce antibody, respectively. M, protein marker. 1, whole bacterial protein. 2, pet-b2m-ZfXH protein induced by IPTG. 3 and 4 pet-b2m-ptprb protein induced by IPTG. **B, C** Antibody titters of ZfXH and Ptprb after three times immunizations of ZfXH and Ptprb antigens, respectively. 2ug/ml ZfXH and Ptprb protein (100ul/well) was used to coat the plates, respectively. The final antibody concentration of ZfXH and Ptprb is 10mg/ml with titter of 1:256000, respectively.

Supplementary Tables:

Table S1a. Summary of genome sequencing strategy for golden pompano.

Table S1b. Statistics of PacBio sub-reads length distribution.

Table S1c. Summary of BioNano data collection and assembly statistics.

Table S1d. Summary of Hi-C data for golden pompanos.

Table S2a. Summary of assembly results for golden pompano.

Table S2b. Statistics of chromosomes for golden pompano based on Hi-C data.

Table S3a. PacBio sub-reads validation for golden pompano genome assembly.

Table S3b. Genome completeness assement by BUSCO.

Table S3c. Summary of pooled transcriptome data assisted for genome annotation.

Table S3d. Genome completeness evaluated by ESTs/unigenes.

Table S4. Repeat content of female and male golden pompano.

Table S5. Summary of nc-RNA of golden pompano.

Table S6a. Comparison of gene features among golden pompano, zebrafish, large yellow croaker, antarctic bullhead notothen and Pacific bluefin tuna.

Table S6b. Summary of gene models annotated by different non-redundant databases.

Table S7. Lineage Specific (LS) genes annotation for Golden pompano.

Table S8. KEEG analysis of positive selection genes for Golden pompano.

Table S9. Positive selection genes of Pacific bluefin tuna.

Table S10. Go term analysis of positive selection genes for Pacific bluefin tuna.

Table S11. KEEG analysis of positive selection genes for Pacific bluefin tuna.

Table S12. Positive selection genes of Golden pompano.

Table S13. Go term analysis of positive selection genes for Golden pompano.

Table S14. KEEG analysis of positive selection genes for Golden pompano.

Table S15. Common positive selection genes of Golden pompano and Pafici bluefin tuna.

Table S16. Pfam annotation of lineage specific genes for gold pompano.

Table S17. Summary of transcript factor (TF) among five genome sequenced fished in Perciformes and Zebrafish.

Table S18. Summary of syntenic gene pairs between Spotted gar and Golden pompano.

Table S19. The enriched molecular function of G17, G19, G20 and G22.

Table S20. The enriched KEGG pathway of G17, G19, G20 and G22.

Table S21. Annotation of the differentially expressed genes.

Table S22. Differentially expression analysis between ZfXH MO and Control MO.

Table S23. Go analysis of the down-regulated genes between ZfXH MO and Control MO.

Table S24. Go analysis of the up-regulated genes between ZfXH MO and Control MO.

Table S25. KEGG pathway analysis.

Table S26. All genes involved in KEGG pathway.

Table S27. Primers and probes used in this study.