Association analysis of intraspecific ESC resistance QTL and identification of landscape of promoters in catfish on the genome-wide level

by

Huitong Shi

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 3, 2019

Keywords: Catfish, ESC resistance, GWAS, QTL, Promoter, H3K4me3

Approved by

Zhanjiang Liu, Chair, Professor of School of Fisheries, Aquaculture, and Aquatic Sciences
Rex Dunham, Co-chair, Professor of School of Fisheries, Aquaculture, and Aquatic Sciences
Charles Y. Chen, Professor of Crop, Soil, and Environmental Sciences
Xu Wang, Assistant Professor of Pathobiology

Abstract

Catfish is the most important aquaculture species, accounting for over 50% of aquaculture production in the US. However, disease problems threaten the survival of the catfish industry. Bacterial diseases are a major issue for the catfish industry, with enteric septicemia of catfish (ESC) being the most severe, causing enormous economic losses every year.

ESC is caused by the bacterial pathogen *Edwardsiella ictaluri*, which is a rod-shape and gram-negative bacterium. Genetic architecture and detailed molecular mechanism of ESC disease resistance in catfish can be studied by genome-wide association study (GWAS), which detected the genomic markers significantly associated with the phenotypic variation and then, identified the associated genomic regions (QTL) and potentially influenced genes.

Two related, but separate studies of catfish genome analysis have been conducted in this dissertation. In the first study, GWAS was conducted to detect the quantitative trait loci (QTL) associated with ESC resistance in channel catfish. Three significant QTL, two of them located on LG1 and one on LG26, and three suggestive QTL located on LG1, LG3, and LG21, respectively, were identified. This study validated one QTL previously identified using interspecific hybrid backcross progenies and identified additional QTL among channel catfish families. There were only a few major QTL for ESC disease resistance, potentially, making marker-assisted selection an effective approach for genetic improvement of ESC resistance in channel catfish.

In the second study, applicability of ChIP-Seq technology in catfish have been confirmed. Genome-wide tri-methylated histone H3 lysine 4 (H3K4me3) profiles were generated using ChIP-Seq technology and the landscape of catfish promoter sequences determined. A total of 14,464 active promoters and 14,346 genes regulated by these promoters were identified. The catfish promoters were characteristic of low GC-content and specific structure of TATA-boxes. A total of 331 and 363 protein-coding genes were regulated by tissue-specific promoters in liver and intestine, respectively. This analysis laid a solid foundation for further analysis of histone modification epigenetic regulators in catfish, and their involvement in regulation of phenotypes such as growth, stress tolerance, and disease resistance in the future.

Acknowledgments

I would like to express my sincere appreciation to my major Professor, Dr. Zhanjiang Liu, for his guidance, support, and patience throughout my degree program. His scientific attitude and passion in education impress me deeply. He trained me to be not only a good researcher, but also a better person in the society. It would be impossible for me to finish this dissertation without his guidance, encouragement, and tremendous support.

I also would like to express my appreciation to my committee co-chair Professor, Dr. Dunham, who also provided extensive assistance to me. I thank my committee members Dr. Charles Y. Chen and Dr. Xu Wang, and my University reader Dr. Joseph Kloepper, for their suggestions and feedback, which helped me to improve the quality of my dissertation. I learned from all my committee professors how to become a good researcher.

I am grateful to Dongya Gao for her review of my papers, experimental designs and routine guidance and support. I thank Dr. Shikai Liu, Dr. Xin Geng, Dr. Lisui Bao, Dr. Qifan Zeng, Dr. Sen Gao, Dr. Tao Zhou, Dr. Ning Li, Dr. Xiaozhu Wang, Dr. Yujia Yang, Dr. Zihao Yuan, Wenwen Wang, De Xing and all the other colleagues in the laboratory for their help, collaboration, and friendship.

Finally, and as always, I want to give a special recognition and appreciation to my parents and other family members for their endless love, understanding, support, and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Problem Statement

Catfish is one of the most important aquaculture species in United States, accounting for over 50% of the total aquaculture production. The bank value of catfish production was $386 and $380 million in 2016 and 2017, respectively (USDA 2018), and the catfish industry had a multi-billion-dollar impact with the added value, including jobs and income for laborers. Mississippi, Alabama, Arkansas, and Texas the top 4 states for catfish industry, which account for over 96% of total catfish production. However, after peaking in 2003, the catfish production in US decreased due to diseases, increasing costs and competition from imports. Therefore, studies on growth, stress tolerance, and disease resistance of catfish have become more important to both catfish farmers and researchers.

Catfish production is threatened by several bacterial diseases including enteric septicemia of catfish (ESC), the columnaris disease, and the emerging *Aeromonas* disease, compromising sustainability of the catfish industry (Wagner et al. 2002). They cause a loss of approximately 40% of the catfish industry (Tucker 2012) and millions of economic. There are no efficient ways currently available to control the disease problems in catfish. Although a live attenuated vaccine against *Edwardsiella ictaluri*, the causative agent of ESC disease, was developed (Klesius and Shoemaker 1999), its effectiveness is limited to only to a fraction of field isolates. Antibiotics are effective for the control of bacterial diseases, but application of large quantities of antibiotics in the aquatic environment would have drastically adverse long-term environmental impact, human health risks, and is expensive. Hence, genetic improvement of disease resistance is an effective alternative for disease control in aquaculture (Vallejo et al. 1998) and the recently published well-

assembled and annotated catfish genome (Liu et al. 2016b) provides a solid foundation for the genetic and genomic analysis of disease resistance in catfish.

In the catfish industry in the United States, two catfish species are utilized, channel catfish (*Ictalurus punctatus*) and blue catfish (*I. furcatus*). Previously, channel catfish was the most commonly use genetic type, but in recent years, hybrid catfish generated by crossing female channel catfish and male blue catfish has become increasingly popular (Dunham 2011; Dunham and Masser 2012) due to their superiority in growth, disease resistance, and other traits, now accounting for 60-70% of catfish production. Dramatic differences exist in disease resistance between channel catfish and blue catfish, although a wide range of resistance levels exist within each species (Dunham and Smitherman 1984; Peterson et al. 2016). For example, these two species show different disease resistance when infected by different pathogens. Channel catfish is more resistant to *Flavobacterium columnare* than blue catfish while the blue catfish is much more resistant to *Edwardsiella ictaluri* than channel catfish. The channel-blue hybrid catfish has been reported to exhibit intermediate levels of ESC resistance, (Wolters et al. 1996; Dunham et al. 2008) but detailed genetic and genomic research is lacking.

Expect for the detection of function of genetic variation of DNA sequences, analysis of epigenetic regulators has also been an area of intense research in recent years. Previously, the non-coding sequences were considered as the "noise" in genome. Recently, it was realized that activation of eukaryotic gene transcription involved the orchestration and coordination of a multitude of epigenetic alternations of DNA sequences such as cis-regulatory elements and varied chromatin structure, such as chromatin accessibly and histone modifications (Lemon and Tjian 2000; Orphanides and Reinberg 2002; Nightingale et al. 2006). Although a set of genomic and genetic analysis associated with economically important traits of catfish have been conducted as

the a well-assembled and annotated reference genome of catfish has been produced (Liu et al. 2016b), the fine-structure mapping of genes, especially for the promoter sequencing and analysis of epigenetic regulatory elements, in catfish are still limited. Furthermore, the complexity of chromatin, and the incomplete understanding of dynamic processes, results in a lack of understanding of how histone modifications contribute to eukaryotic gene regulation and chromosome packaging in catfish and other species.

1.2 Review of literatures

1.2.1 Genetic studies of aquaculture species

To achieve the goal of genetic improvement, a key requirement for marker-assisted selection is the understanding of the genetic architecture of the focus trait. As the development of genotyping technology, sequencing methods, and large number of identified SNPs (single nucleotide polymorphisms) on the whole genome level, it is now possible for researchers to have a more detailed and better understanding of the genetic architecture and molecular mechanism of the trait of interest and identify the QTL (quantitative trait loci) associated with the phenotypic variation of the trait.

QTL mapping is a powerful method, which could lead to the identification of many genomic regions associated with a given trait and help to understand complex trait variation. QTL mapping has been conducted for over 20 aquaculture species. For example, detection of growth-related QTL has been programmed in barramundi (*Lates calcarifer*) (Wang et al. 2008), turbot (*Scophthalmus maximus*) (Sánchez-Molano et al. 2011), half-smooth tongue sole (*Cynoglossus semilaevis*) (Song et al. 2012), Asian seabass (Wang et al. 2015), and common carp (*Cyprinus carpio*) (Peng et al. 2016). The QTL associated with sex determination has also been detected in Atlantic halibut

(*Hippoglossus hippoglossus*) (Palaiokostas et al. 2013), half-smooth tongue sole (Song et al. 2012), common carp (Peng et al. 2016), tilapia (Liu et al. 2013a), and turbot (Viñas et al. 2012).

In the case of disease resistance, QTL detection has been conducted for many aquaculture species. Identification of QTL controlling the resistance to lymphocystis disease (Fuji et al. 2006) and *Vibrio anguillarum* disease resistance (Shao et al. 2015) have been obtained in Japanese flouder (*Paralichthys olivaceus*). A fine-mapping of the QTL for resistance to infectious pancreatic necrosis has been constructed in Atlantic salmon (*Salmo salar*) (Moen et al. 2009). QTL for resistance to *Bonamiosis* (Bonamia ostreae) in the European flat oyster (*Ostrea edulis*) were detected (Lallias et al. 2009). QTL detection of *Aeromonas salmonicida* resistance related traits was conducted in turbot (Rodríguez-Ramilo et al. 2011). QTL associated with bacterial cold water disease resistance (Liu et al. 2015) and infectious pancreatic necrosis virus (IPNV) (Ozaki et al. 2001) have been identified in rainbow trout (*Oncorhynchus mykiss*).

Over the last decade, GWAS has become the standard and one of the most powerful tools for gene discovery in genomic and genetic research. GWASs are based upon the principle of linkage disequilibrium (LD) at the population level and LD is the nonrandom association between alleles at different loci (Visscher et al. 2012). The basic approach in GWAS is to evaluate the association between each genotyped marker and a phenotype of interest that has been scored across a large number of individuals (Korte and Farlow 2013). Early in the GWAS era, costs were high and sample sizes were small, but with technological advances, prices have come down significantly and typical sample sizes are now in the thousands (Begum et al. 2012). During recent years, high-density SNP arrays and DNA re-sequencing have illuminated the development and application of GWAS, and GWAS presents a relatively effective tool to reconnect important traits back to their underlying genetics (Korte and Farlow 2013). GWAS was popular conducted in aquaculture

species for many years. GWAS for growth rate and age at sexual maturation (Gutierrez et al. 2015), fillet texture and fat content (Sodeland et al. 2013), and resistance to *Caligus regercresseyi* (Correa et al. 2017) were conducted in Atlantic salmon. A QTL associated with the resistance to viral nervous necrosis disease was detected by GWAS in Asian seabass (Wang et al. 2017a). In Pacific oysters, a GWAS was programmed to identify QTL related to host resistance to *Ostreid herpesvirus* (Gutierrez et al. 2018). Identification of QTL associated with viral nervous necrosis in Atlantic cod was achieved by GWAS using a 12K SNP array.

Genetic and genomic studies with aquaculture species suggested that in some cases, marker-assisted selection may be effective because a very limited number of major QTL were found to control the vast majority of phenotypic variance. For example, a pioneering discovery was that a single gene was responsible for the resistance to lymphocystis disease in Japanese flounder (Fuji et al. 2006), and markers linked with the locus since have been used for marker-assisted selection that are 100% effective (Fuji et al. 2007). Similarly, a major QTL was found to account for up to 83% of phenotypic variance for resistance to infectious pancreatic necrosis disease in Atlantic salmon (Moen et al. 2009; Houston et al. 2012), and a recent study further identified epithelial cadherin as the causal gene for disease resistance (Moen et al. 2015). one major QTL was identified for whirling disease of rainbow trout (*Oncorhynchus mykiss*), which could explain 50-86% of the phenotypic variance across families (Baerwald et al. 2011). A few major QTL in rainbow trout were found to explain 83-89% of phenotypic variance of bacterial cold water disease (Vallejo et al. 2010). A major QTL associated with *Vibrio anguillarum* resistance was identified in Japanese flounder (Wang et al. 2014a). A few QTL associated with columnaris resistance/susceptibility have been identified for catfish (Geng et al. 2015). In all these cases, only a limited number of major QTL accounted for a large proportion of phenotypic variations, suggesting that marker-

assisted selection may be quite effective for enhancing disease resistance with aquaculture species. However, for selection of many other traits, or simultaneous selection of several other traits along with disease resistance, genome selection apparently is more effective. Although the genome selection is used for Atlantic salmon by several international companies, its implementation is currently cost-prohibitive for most aquaculture species.

1.2.2 Genetic studies in catfish

Catfish is one of the most important aquaculture species in United States (US), accounting for over 50% of the total aquaculture production each year. There are roughly 4,100 species in the order of Siluriformes (catfish), which comprise 12% of all fish species and 6.3% of all vertebrate species (Wilson and Reeder 2005; Uetz and Hošek 2016; Gill and Wright 2006; Eschmeyer and Fong 2014; Sullivan et al. 2006). Due to the its critical position in both economics and evolution, it is necessary to have a good the understanding of genetic architecture of important traits of catfish including growth, sex determination, stress tolerance, and disease resistance.

Generally, RNA-seq and GWAS are two main methods to analyze the potential causal genes associated with these economically important traits, including growth, stress tolerance, and disease resistance. RNA-seq is an approach to transcriptome profiling that uses deep-sequencing technologies, which also provides a far more precise measurement of levels of transcripts and their isoforms than other methods (Wang et al. 2009). RNA-Seq revealed expression signatures of genes involved in oxygen transport, protein synthesis, folding and degradation in response to heat stress in F1 hybrid catfish (Liu et al. 2013b). Potential candidate genes involved in salinity tolerance in striped catfish (*Pangasianodon hypophthalmus*) was evaluated using an RNA-Seq approach (Nguyen et al. 2016). Male-biased genes (Sun et al. 2013) and genes related to testicular differentiation (Zeng et al. 2016) in channel catfish were revealed by RNA-Seq analysis of the

testis transcriptome. Transcriptomic analyses indicated novel genes with sexually dimorphic expression in yellow catfish (*Pelteobagrus fulvidraco*) brain (Lu et al. 2015). RNA-seq also uncovered the NF-κB suppression and IFN stimulation in the catfish gill following columnaris bacterial infection (Sun et al. 2012) and a number of differentially expressed influential innate immune components between resistant and susceptible catfish to columnaris disease resistance (Peatman et al. 2013). Comparative transcriptome analysis between wild and albino yellow catfish were conducted to suggest the contribution of differentially expressed genes to the phenotypic variation (Zou et al. 2015) Comparative transcriptome analysis revealed conserved branching morphogenesis related genes involved in chamber formation of catfish swimbladder (Yang et al. 2017), in response to low oxygen stress (Yang et al. 2018a) ,and identified the enrichment of genes associated with auditory system (Yang et al. 2018b).

The meta-analysis of RNA-seq datasets could be organized to recognize the function of gene family because of the enormity of RNA-seq data. Vis this method, cytochrome P450 genes were found to be involved in disease defense responses (Zhang et al. 2014). 52 Rab GTPase genes were identified for their involvement in immune responses after bacterial infections (Wang et al. 2014b). It was detected that expression of nitric oxide synthase (NOS) genes in channel catfish was highly regulated and time dependent after bacterial challenges (Yao et al. 2014). Interleukin 17 family ligand and receptor genes in channel catfish were induced by *Edwardsiella ictaluri* and *Flavobacterium columnare* infection (Wang et al. 2014c). Claudin multigene family was found to be in response to bacterial infection and hypoxia in channel catfish (Sun et al. 2015). Hsp90, Hsp60 and sHsp families of heat shock protein genes in channel catfish were recognized significantly regulated expression after bacterial infections (Xie et al. 2015). Bcl-2 genes in channel catfish potentially function by reducing cell apoptosis in response to bacterial infection and hypoxia stress

(Yuan et al. 2016). It was found that the chemokinome superfamily in channel catfish was highly involved in disease defense and hypoxia responses (Fu et al. 2017).

Lack of resources for catfish limited the development of GWAS analysis in catfish, however, GWAS has become the most powerful method to detect the QTL and potentially causal genes of mentioned traits of concerned based on a set of essential technologies for genomic analysis including the well-assembled and annotated reference genome (Liu et al. 2016b), high density SNP arrays (Liu et al. 2014; Zeng et al. 2017), high-density genetic linkage maps (Li et al. 2014; Liu et al. 2016a), as well as the well-studied intraspecific and interspecific segregation populations (Waldbieser et al. 2001; Liu et al. 2003; Kucuktas et al. 2009; Ninwichian et al. 2012).

Through GWAS analysis, the novel genes significantly affecting growth were identified (Li et al. 2018a). In this study, A genomic region of approximately 1 Mb on LG (linkage group) 5 was found to be significantly associated with body weight while four suggestively associated QTL regions on LG 1, 2, 23, and 24 were identified (Li et al. 2018a). GWAS also identified the QTL associated with catfish head size: a QTL on LG 7 was identified significantly associated with head length while a QTL on LG 9 was identified significantly associated with head width (Geng et al. 2016). The genes with functions for bone development associated with body confirmation were also detected using GWAS, LG 2 and 24 contained the significant QTL associated with deheaded body length while LG 2, 7, 18 contained the significant QTL associated with body length (Geng et al. 2017).

Climate changes and its associated pressures make genetic improvement of stress tolerance of great importance. Therefore, a set of GWAS related to stress tolerance were conducted in catfish. Three significant SNPs and a total of 14 genes were detected to be associated with heat stress (Jin et al. 2017). QTL associated with hypoxia tolerance were characterized in both channel (Wang et

8

al. 2017b) and hybrid catfish (Zhong et al. 2017). Pathway analysis indicated that the MAPK or PI3K/AKT/mTOR signaling pathways play key roles in low oxygen tolerance in both two populations (Zhong et al. 2017; Wang et al. 2017b).

Stress can lead to disease. GWAS also were conducted to identify the genetic architectures and molecular mechanisms of disease resistance in catfish. GWAS revealed the importance of QTL on LG 7 and presence of functional hubs of related gene within QTL for columnaris disease resistance in catfish (Geng et al. 2015). In this study, many candidate genes within associated QTL regions were involved in PI3K pathway, which is known to be required by many bacteria for efficient entry into the host (Geng et al. 2015). Similarly, GWAS analysis identified the QTL on LG 2, 26, and 29 to be significantly associated with Motile *Aeromonas septicemia* resistance and indicated importance of NF-κB signaling pathway in host resistance in catfish (Wang et al. 2019).

Climate change could influence sex ratios as temperature can affect sex determination in some species of fish, including channel catfish (Patiño et al. 1996), and this has important evolutionary and extinction implications. Both RNA-seq and GWAS were utilized to study the sex determination mechanisms in catfish (Bao et al. 2019). GWAS analyses placed the sex determination locus within a genetic distance less than 0.5 cM and physical distance of 8.9 Mb and the comparative RNA-Seq analysis between females and males revealed exclusive sex-specific expression of an isoform of the breast cancer anti-resistance 1 (BCAR1) gene in the male during early sex differentiation, suggesting BCAR1 as a putative sex determination gene (Bao et al. 2019).

1.2.3 Genetic studies of ESC in catfish

As other traits, two sets of genetic studies have been conducted with ESC disease in catfish: expression analysis of differentially expressed genes and genetic analysis of disease resistance loci.

Microarray analysis of gene expression after ESC infection revealed that early activation of the MHC class I pathway could be the basis for ESC disease resistance in blue catfish (Peatman et al. 2008). In addition, RNA-seq analysis of mucosal immune responses following *E. ictaluri* infection was conducted in channel catfish (Li et al. 2012). A total of 1,633 differentially expressed genes were detected after the infection of pathogen bacteria and gene pathway analysis of the differentially expressed genes in this study indicated that actin cytoskeletal polymerization/remodeling and junctional regulation might be associated with susceptibility for pathogen entry and subsequent inflammatory responses (Li et al. 2012). BSR-seq (bulk segregant RNA-seq) was used to reveal expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish in backcross catfish populations, produced by mating male channel-blue hybrid catfish and parental female channel catfish (Wang et al. 2013). Using this method, 8 LGs, LG 1, 3, 6, 9, 15, 17, 18, and 25, appeared to harbor QTL involved in ESC disease resistance; a total of 1,255 differentially expressed genes were found between resistant and susceptible fish; 56,419 SNPs residing on 4,304 unique genes were identified as significant SNPs between susceptible and resistant fish (Wang et al. 2013). Another BSR-seq analysis was conducted in channel catfish populations which revealed that the LG 1 and 16 contained the QTL significantly associated with ESC disease resistance (unpublished data). Alternative splicing analysis using previous RNA-seq data was conducted recently alternative splicing was greatly induced by *E. ictaluri* infection with a 21.9% increase in alternative splicing events (Tan et al. 2018a). Sequence analyses of splice variants of a representative alternatively spliced gene, splicing factor *srsf2*, revealed that certain spliced transcripts may undergo nonsense-mediated decay (NMD), suggesting functional significance of the induced alternative splicing (Tan et al. 2018a).

We recently conducted GWAS analysis with the third generation of interspecific backcross catfish populations and three genomic regions were identified to be associated with ESC disease resistance using the 250K catfish SNP array (Liu et al. 2014): One QTL was identified on LG 1 that was statistically significant, and two suggestive QTL were identified with one each on LG 12 and LG 16 (Zhou et al. 2017b). Detailed gene analysis within the detected associated genomic regions, revealed that the *nlrc3* duplicates were identified within all the three QTL, suggesting their importance in association with the QTL (Zhou et al. 2017b). The detection of *nlrc3*, *nlrp12*, and *nlrp3*, which were reported to be involved in NOD-like receptor signaling pathway (Castaño-Rodríguez et al. 2014), indicated that NOD-like receptor signaling plays a pivotal role in ESC resistance in interspecific backcross catfish populations (Zhou et al. 2017b). This study also verified the function of members of NLP gene family (Peatman et al. 2015) and class I major histocompatibility complex (MHC)-mediated antigen processing and presentation pathway (Peatman et al. 2008) in ESC resistance. The *nck1* gene near the significantly associated SNP (AX-85308810), which was known to function as an adaptor to facilitating the pathogenesis of enteropathogenic *Escherichia coli* (EPEC) in humans. Notably, *E. ictaluri* and EPEC pathogens belong to the same bacterial family and share many common characteristics (Zhou et al. 2017b). Later, the role of NCK and ABI genes in ESC disease response were confirmed by the analysis of RNA-seq data, which found that the majority of NCK and ABI genes were expressed at higher levels in ESC susceptible fish than ESC resistant fish (Zhou et al. 2017a).

The use of higher generations of interspecific backcross populations offered greater power for fine mapping of the disease resistance QTL but has the possibility of missing some QTL because the number of individuals used to produce higher generations of interspecific backcross populations was very small. In a parallel study, QTL analysis was conducted using first generation

of interspecific hybrid backcross populations using a recently developed 690K catfish SNP array (Zeng et al. 2017), which validated the QTL on LG 1, and discovered additional significant QTL for ESC disease resistance on LG 23 as well as a suggestively significantly QTL on LG 1 (Tan et al. 2018b). Examination of the resistance alleles indicated their origin from blue catfish, indicating that at least two major disease resistance loci exist among blue catfish populations compared to channel catfish populations (Tan et al. 2018b). In this study, a set of immune genes located within the associated genomic regions were detected. It was found that the immune-related genes observed in the significant QTL were found to be mainly involved in phagocytosis and T-cell activation (Tan et al. 2018b).

While those two studies provided important information of QTL for ESC disease resistance in the interspecific hybrid backcross populations, QTL analysis of ESC disease resistance within the channel catfish intra-specific system has not been conducted. As a result, it is unknown if the identified QTL were of interspecific in nature and, some of the resistant alleles could have been originated from channel catfish.

1.2.4 Promoter

Activation of eukaryotic gene transcription involves the orchestration and coordination of a multitude of epigenetic alternations of DNA sequences such as cis-regulatory elements and varied chromatin structure, such as chromatin accessibly and histone modifications (Lemon and Tjian 2000; Orphanides and Reinberg 2002; Nightingale et al. 2006). The cis-regulatory elements are composed of DNA (typically non-coding DNA) containing binding sites for TFs and /or other regulatory molecules that are needed to activate and sustain transcription (Ong and Corces 2011). Cis-regulatory regions are highly involved in regulation the transcription of a gene. Promoters and enhancers are the best understood types of cis-regulatory elements (Levine 2010; Bulger and

Groudine 2011). Compared to promoters, enhancers tend to be more variable among species and they are the type of cis-regulatory elements that are most often thought to be responsible for cis-regulatory divergence (Wray et al. 2003).

Promoters, an important type of cis-regulatory elements, are located at the 5′ ends of genes immediately surrounding the transcriptional start site (TSS) and serve as the point of assembly of the transcriptional machinery and initiation of transcription (Smale and Kadonaga 2003). At first, the substantial non-coding regions were considered as "junk" sequences. However, it has been discovered that this portion of the genome is full of functional elements. The promoters, in conjunction with enhancers, silencers and insulators, define the combinatorial codes that specify gene expression patterns (Tjian and Maniatis 1994). Most eukaryotic genes contain a single promoter located close to the transcription start site, although some genes contain alternative promoters that activate transcription at different positions in the genome, often under specific conditions (Wittkopp and Kalay 2012). Promoter mutations are, however, a frequent cause of human disease (Savinkova et al. 2009).

In 1983, the DNA sequence of 168 promoter regions (-50 to +10) for *Escherichia coli* RNA polymerase were first compiled and analyzed (Hawley and McClure 1983). Twenty years later, the genome-wide analysis of promoters became a popular research area. The genome-wide promoter analysis of the general stress response network in *E. coli* were conducted in 2005 (Weber et al. 2005a). Recently, many new next-generation DNA-sequencing technologies have allowed these studies in eukaryotes, especially human. A chromosome-wide and promoter-specific analyses identified sites of differential DNA methylation in normal and transformed human cells (Weber et al. 2005b). Nascent RNA sequencing revealed the widespread pausing and divergent initiation at human promoters and found that promoter-proximal polymerase reside on ∼30% of

human genes, transcription extends beyond pre-messenger RNA 3' cleavage, and antisense transcription is prevalent (Core et al. 2008). Genome-wide promoter analysis also uncovered the portions of the cancer methylome (Hoque et al. 2008) and the SOX4 transcriptional network in prostate cancer cells (Scharer et al. 2009).

Besides human, core promoter modules and their application in accurate transcription start site prediction have been identified in Drosophila (Ohler 2006). Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs provided a resource for exploring promoter Polr2a binding across pluripotent and differentiated cell types in mammals (Barrera et al. 2008). The genome-scale analysis of *in vivo* spatiotemporal promoter activity has been conducted in *Caenorhabditis elegans* (Dupuy et al. 2007). Abundance, arrangement, and function of sequence motifs have been analyzed in the chicken promoters (Abe and Gemmell 2014). Comparative promoter analyses have also been conducted to identify the species-specific promoter architecture, conserved promoters, and the promoter evolution (Xuan et al. 2005; Carninci et al. 2006; Sandelin et al. 2007; Jin et al. 2006). Until now, enormous studies of promoter sequence have been conducted and for some species, the precise promoter sequences have been produced and can be obtained from the EPD (Eukaryotic Promoter Database, https://epd.epfl.ch//index.php). However, most studies about promoters addressed important mammals or model species.

Eukaryotic gene promoters could be broadly classified as TATA-containing and TATA-less based on the presence of TATA box in their promoter sequences (Basehoar et al. 2004). TATA box was one of the most conserved, ancient, and best characterized element in most core promoters, which served as a binding site for TBP (TATA-binding protein) (Struhl et al. 1998). Usually, the promoters were regulated to be expressed by the binding of TBP to the TATA box (Levine and Tjian 2003). Although the core promoters without TATA box could ensure the recognition by the

RNA polymerase II (Pol II) transcription complex to initiate transcription by compensatory downstream promoter element, it was possible that some enhancers preferentially activate TATA-containing promoters (Ohtsuki et al. 1998). A wide variation was found in the percentage of TATA box-containing promoters reported in different studies, which is due to differences in the definition used for TATA box, the window size considered for extracting TATA-containing promoters, and to a lesser extent the datasets used (FitzGerald et al. 2004; Yang et al. 2007; Trinklein et al. 2003). Based on several criteria such as a maximum sequence length (8 bp), minimal consensus sequence, confined upstream location, and conservation across orthologous upstream regions, TATA box was defined as TATA(A/T)A(A/T)(A/G) (TATAWAWR) (Basehoar et al. 2004).

Although TATA box is the most conserved and well-known element in the promoter, the extensive existence of CpG island was found, especially in the TATA-less promoters. In vertebrates, the postreplication addition of methyl groups to the 5-position of cytosine in certain CpG dinucleotides and the maintenance of a particular genomic pattern of methylated CpGs provides an epigenetic means for differential regulation of gene expression (Saxonov et al. 2006; Jaenisch and Bird 2003; Novik et al. 2002). In humans and mice, approximately 60% of all promoters colocalize with CpG islands, which are regions devoid of methylation that have a higher G+C content than the genome average, while the rest have a methylation pattern and base composition indistinguishable from bulk DNA (Antequera 2003). The pattern of methylation often varies between cell types and different conditions, changes throughout development, and is abnormal in many disease states (Herman and Baylin 2003; Jaenisch and Bird 2003; Novik et al. 2002; Jones and Takai 2001; Geiman and Robertson 2002). It was widely accepted that the state of CpG methylation regulates and stabilizes chromatin structure, perhaps regulating accessibility of the transcription machinery to regions of DNA (Jaenisch and Bird 2003; Fahrner et al. 2002;

Geiman and Robertson 2002). Therefore, it was concluded that the methylation state of CpG island in promoters could guide the expression of the regulated genes by suppressing the expression via methylated CpGs and allowing the expression by keeping CpGs unmethylated. The study about mouse aprt gene promoter recognized that the Sp1 sites are required to prevent methylation of the CpG island and confirm the expression of the gene (Macleod et al. 1994). In humans, 72% of promoters with high CpG concentrations and 28% of promoters with low CpG concentrations based on the ratio of 25% (Saxonov et al. 2006).

1.2.5 Histone modification, histone methylation, and H3K4me3

Epigenetic alternations have been considered as a reasonable explanation of phenotypic complexity. Histone modifications were a cluster of conserved and effective epigenetic regulators. Mounting evidence suggests that histone modifications play a crucial role in diverse biological processes including cell differentiation, organismal development, and that aberrant modification of histones contributes to diseases (Füllgrabe et al. 2011). At least eleven types of post-transcription modifications (PTMs) have been reported at over 60 different amino acid residues on histones, including histone methylation, acetylation, propionylation, butyrylation, formylation, phosphorylation, ubiquitylation, sumoylation, citrullination, proline isomerization, and ADP ribosylation (Martin and Zhang 2007; Ruthenburg et al. 2007). Among these, methylation and acetylation are the two most popular and well-studied types of histone modification. Acetylation of histone tails neutralizes the positive charge of lysines and profoundly alters chromatin properties (Waterborg 2002; Shogren-Knaak et al. 2006) while methylation of particular lysines on histone tails can increase the affinity of binding modules present on a variety of proteins that are thought to act by altering chromatin packaging (Fischle et al. 2003). Histone modifications were considered to contribute to the regulation of chromatin-templated processes via two major mechanisms: (1)

directly modulate the packaging of chromatin either by altering the net charge of histone molecules or by altering inter-nucleosomal interactions, thereby regulating chromatin structure and the access of DNA-binding proteins such as transcription factors (Kouzarides 2007; Ruthenburg et al. 2007); (2) regulate chromatin structure and function by recruiting specific binding proteins, which recognize modified histones via specialized structural folds such as bromo-, chromo- and PHD domains (Wysocka et al. 2005; Wysocka et al. 2006; Zeng and Zhou 2002). As the role in epigenetic regulation of important downstream events of histone modification, the concept of a 'histone code' has been popular to determine histone modification complexity (Strahl and Allis 2000).

As mentioned, the histone methylation is one of the most popular, conserved, and well-studied cluster of histone modification. Methylation of histone could occur at arginines and lysines and methylation at lysines or arginines may be one of three different forms: mono-, di-, or trimethyl for lysines and mono- or di- (asymmetric or symmetric) for arginines (Kouzarides 2007). Methylation at arginines occurs within the tails of histone H3 (R2, R17, R26) and H4 (R3) (Kouzarides 2002). There are five known arginine methyltransferases that have a highly conserved catalytic domain: PRMT1, PRMT3 and PRMT4/CARM1 are classified as Class I enzymes as they can catalyze the formation of asymmetric dimethylated arginine; PRMT5/JBP1 is classified as a class II enzyme as it catalyzes symmetric demethylation. The PRMT2 protein has not yet been established as an enzyme (Zhang and Reinberg 2001; McBride and Silver 2001). Methylation of histones at arginine may be a stimulating event for transcription by finding CARM1/PRMT4 acted as a co-activator of nuclear receptor activity using an arginine methyltransferase domain capable of methylating specifically histone H3 (Chen et al. 1999) and this function was confirmed by using antibodies that specifically recognize the major methylation site by CARM1, Arg17 of histone H3

17

(Ma et al. 2001; Bauer et al. 2002). Therefore, arginine methylation represents a histone modification that correlates with the active state of transcription (Kouzarides 2002).

Methylation of lysine residues is mostly known to occur on histone H3 (K4, K9, K27, K36, and K79) and H4 (K20) (Kouzarides 2002) while the SUV39 protein was the first histone methyltransferase to be discovered (Rea et al. 2000). Different histone methylations are associated with different chromatin functions, and early experiments suggested that methylation of H3K4, H3K36, and H3K79 is generally associated with actively transcribed genes, whereas H3K9, H3K27, and H4K20 methylation frequently demarcate silent chromatin (Lachner and Jenuwein 2002; Berger 2007). However, the histone lysine methylation owns "many faces" since it was found that methylation of H3K4 was involved in rDNA silencing (Briggs et al. 2001) and H3K4 methylation was enriched in the silenced region (Bryk et al. 2002). To understand how H3K4 methylation could be associated with both activation and suppression of gene transcription, a study using antibodies that can distinguish between di- and tri-methylated H3K4 was conducted and revealed that tri-methylation (H3K4me3), is specific for the active state of transcription, whereas di-methylated K4 (H3K4me2) exist in both active and repressed genes (Santos-Rosa et al. 2002). In most vertebrate cells, H3K27me3 is inversely correlated with H3K4me3 (Shilatifard 2006), but they coexist in discrete or overlapping regions of individual genes, usually promoter regions, to form a bivalent control of gene transcription (Bernstein et al. 2006; Pan et al. 2007; Zhao et al. 2007; Voigt et al. 2012; Voigt et al. 2013). Methylation of H3K9 has emerged as an important modification that is associated with transcriptional silencing, heterochromatin formation (Boggs et al. 2001), and DNA methylation (Krishnan et al. 2011). However, just like the methylation of H3K4, it was verified that di- and trimethylation of K9 (H3K9me2 and H3K9me3) are enriched in the transcriptional start sites of silenced genes, whereas the monomethylated K9 (H3K9me1) is

present in the promoter regions of active genes (Barski et al. 2007). Conclusively, the number of methyl groups in a modification, and not just the particular site, appears to play an important role in the functional consequences of histone methylation (Iizuka and Smith 2003).

Again, H3K4me3 was one of the least abundant histone modifications in the genome but highly enriched in gene promoter regions. Recent genome-wide high-resolution profiling studies of H3K4me3 in the human and mouse genomes further indicated that the H3K4me3 preferentially located around the TSS, and was preferentially associated with the promoters of active genes (Barski et al. 2007; Bernstein et al. 2005). This characteristic of H3K4me3 also confirmed in other eukaryotic species, including yeast (Bernstein et al. 2002; Santos-Rosa et al. 2002; Ng et al. 2003; Pokholok et al. 2005), *Drosophila* (Schübeler et al. 2004), and chicken (Schneider et al. 2004). All these studies suggested that the histone modification H3K4me3 is a hallmark of transcription initiation at most genes and has been convincingly shown to broadly occur at the promoters of most protein-coding genes (Guenther et al. 2007). Hence, H3K4me3 was a logical choice to detect promoter regions in a non-model teleost, catfish, as an epigenetic mark (Kratochwil and Meyer 2015b). Some additional detailed studies of H3K4me3 uncovered its more precise functions. For example, it was found that binding of the MLL (mixed-lineage leukemia) PHD (plant homeodomain) 3 finger to H3K4me3 was required for MLL-dependent gene transcription (Chang et al. 2010). PHD finger protein 8 could activate transcription of rRNA genes through H3K4me3 binding (Feng et al. 2010). H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 (inhibitor of growth 1) tumor suppressor (Pena et al. 2008). It was concluded that H3K4me3 regulated preinitiation complex assembly and selective gene activation via the interaction with TAF (TBP-associated factors) 3 and H3K4me3 directed the rapid induction of specific p53 target genes (Lauberth et al. 2013). H3K4me3 breadth was identified to be linked to

cell identity and transcriptional consistency (Benayoun et al. 2014). Regulation of H3K4me3 at transcriptional enhancers characterized acquisition of virus-specific CD8+ T cell-lineage-specific function (Russ et al. 2017)

1.2.6 ChIP-seq (Chromatin Immunoprecipitation-seq)

The genome-wide mapping of epigenetic marks is essential for a full understanding of transcriptional regulation (Park 2009). It is possible to directly detect the genome-wide enrichment of histone modifications using pertinent antibodies. CHIP-seq has become the most effective and efficient tool for histone modification identification, which was first used around 2007 (Johnson et al. 2007; Barski et al. 2007; Robertson et al. 2007; Mikkelsen et al. 2007). Compared to chromatin immunoprecipitation followed by microarray (ChIP-chip), which is a similar but traditional method as ChIP-seq, ChIP-seq has higher resolution, less noise, higher genome coverage, and wider dynamic range (Park 2009).

ChIP-seq has been used in different species for identification of transcription factors. For instance, ChIP-seq identified two group of transcription factors in mouse embryonic stem cells and predicted absolute and differential gene expression (Ouyang et al. 2009). Genome-wide identification of CTCF (CCCTC-binding factor), NRSF (neuron-restrictive silencer factor) and STAT1 (signal transducer and activator of transcription protein 1) were conducted using ChIP-seq (Jothi et al. 2008). ChIP-seq of ERα and RNA polymerase II defined genes differentially responding to ligands (Welboren et al. 2009). A ChIP-seq defined genome-wide map of vitamin D receptor binding and indicated its associations with disease and evolution (Ramagopalan et al. 2010). Diverse transcription factor binding features was revealed by genome-wide ChIP-seq in *Caenorhabditis elegans* (Niu et al. 2011). A high-resolution genome-wide mapping of hypoxia-inducible factor (HIF)-binding sites was revealed by ChIP-seq (Schödel et al. 2011). Genome-

wide ChIP-seq revealed a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation (Pilon et al. 2011). Multi-read analysis of ChIP-seq data discovered transcription factor binding sites in highly repetitive regions of the genome (Chung et al. 2011). The comparative analysis of ChIP-seq of five vertebrates has been conducted to identify the evolutionary dynamics of transcription factor binding (Schmidt et al. 2010b). ChIP-seq also has been utilized to conduct the analyses of conserved regulatory regions, including reveal nucleosome architecture of human promoters (Schmid and Bucher 2007), uncovering divergent transcription over short distances from active promoters (Seila et al. 2008), identification of weakly conserved heart enhancers (Blow et al. 2010), suggesting enhancer RNA synthesis actively engaged in promoting mRNA synthesis (Kim et al. 2010), detection of general transcription factors in both enhancers and promoters (Koch et al. 2011), and detection of conserved epigenomic signals in mice and humans associated with Alzheimer's disease (Gjoneska et al. 2015).

ChIP-seq is also an effective and efficient method to conduct analysis of tissue-specific regulatory elements. For example, tissue-specific activity of enhancer prediction has identified in mouse embryonic forebrain, midbrain and limb tissue (Visel et al. 2009). The role for cohesion in tissue-specific transcription has been revealed (Schmidt et al. 2010a). Tissue-specific transcription initiation platforms and general transcription factors has been detected (Koch et al. 2011). Tissue-specific analysis of chromatin state identified temporal signatures of enhancer activity during embryonic development (Bonn et al. 2012).

ChIP-seq was also used in several aquaculture species in recent years. By using ChIP-seq, a comprehensive characterization of zebrafish *globin* loci and the regulation of *globin* switching was clarified (Ganis et al. 2012); the role of imbalance of histone methylation dynamics in X-

linked mental retardation was indicated in zebrafish (Qi et al. 2010); profiling of histone modification marks was produced during early development of zebrafish (Bogdanovic et al. 2013); cis-regulatory features in the embryonic zebrafish genome through large-scale profiling of histone modifications binding sites were identified (Aday et al. 2011); conserved cis-regulatory nodes were identified by comparing epigenomics of zebrafish and medaka (*Oryzias latipes*) (Tena et al. 2014); active promoters in in Nile tilapia (*Oreochromis niloticus*) were mapped to uncover cis-regulatory elements in ecological model teleosts (Kratochwil and Meyer 2015a); the landscape of CTCF binding landscape in jawless fish with reference to Hox cluster evolution was identified (Kadota et al. 2017); and the key role of epigenetic modification to orchestrate the thermocoupling of the immune response during behavioral fever was demonstrated in Atlanta Salmon (*Salmo salar*) (Boltana et al. 2018).

1.3 Rationale and significance

GWAS analysis is effective and efficient to identify the associated SNPs, QTL and potential causal genes for ESC resistance in catfish. Although two GWA studies were conducted using interspecific hybrid backcross progenies provided important information for QTL for ESC disease resistance in the interspecific hybrid system, QTL analysis of ESC disease resistance within channel catfish intra-specific system has not been conducted. As a result, it is unknown if the identified QTL were interspecific in nature or if the resistant alleles could have originated from channel catfish. With the objectives of determining if the previously identified QTL were interspecific and if additional major QTL are operating within channel catfish, we took the advantage of the recently developed 690K catfish SNP arrays (Zeng et al. 2017) and conducted GWAS analysis of ESC disease resistance using channel catfish families in the current study.

Comparative analysis was conducted to provide insights into the origin of the resistant alleles. Here, we report both novel QTL identified within channel catfish and overlapped QTL identified from interspecific hybrid system (Zhou et al. 2017b; Tan et al. 2018b).

Although the well-assembled and annotated catfish genome has been published (Liu et al. 2016b), the lack of fine-structure mapping of gene sequences, especially promoter sequences, limit a further understanding of catfish gene structure. Hence, we conducted genome-wide H3K4me3 profiling using ChIP-seq to identify the active promoter sequences in catfish liver and intestine. The enriched motifs, the relatively low GC-content catfish promoter sequences, and specific structure of TATA-box detected in this study indicated the characteristics of catfish promoters. Tissue-specific active promoters in catfish liver and intestine further compared and verified the specific function of each tissue. This analysis helped to construct a standard protocol for further histone modification epigenetic regulators studies in catfish for other economically important traits in the future.

Chapter 2 GWAS of intraspecific QTL from channel catfish (*Ictalurus punctatus*) associated with the resistance for ESC

2.1 Abstract

Disease resistance is one of the most important traits for aquaculture industry. For the catfish industry, enteric septicemia of catfish (ESC), caused by the bacterial pathogen *Edwardsiella ictaluri*, is the most severe disease, causing heavy economic losses every year. In this study, we used three channel catfish families with 900 individuals (300 fish per family) and the 690K catfish SNP arrays, and conducted genome-wide association study (GWAS) to detect the quantitative trait loci (QTL) associated with ESC resistance. Three significant QTL, with two of them located on LG1 and one on LG26, and three suggestive QTL located on LG1, LG3, and LG21, respectively, were identified to be associated with ESC resistance. With a well assembled and annotated reference genome sequence, genes around the involved QTL regions were identified. Among these genes, 37 genes had known functions in immunity, which may be involved in ESC resistance. Notably, *nlrc*3 and *nlrp12* identified here were also found in QTL regions of ESC resistance in the channel catfish x blue catfish interspecific hybrid system, suggesting this QTL was operating within both intra-specific channel catfish populations and interspecific hybrid backcross populations. Many of the genes of the Class I MHC pathway, for mediated antigen processing and presentation, were found in the QTL regions. The positional correlation found in this study and the expressional correlation found in previous studies indicated that Class I MHC pathway was significantly associated with ESC resistance. This study validated one QTL previously identified using the second and fourth generation of the interspecific hybrid backcross progenies and identified five additional QTL among channel catfish families. Taken together, it appears that there

are only a few major QTL for ESC disease resistance, making marker-assisted selection potentially effective approach for genetic improvement of ESC resistance.

2.2 Background

Plant and animal diseases cause billions of economic losses each year and pose a threat to global food security and agricultural sustainability. Compared to most agricultural plant and animal species, the significance of disease problems in aquaculture is paramount because of the intimate relationship of aquatic species with their environment and the intensity of aquaculture industry. Catfish is the major aquaculture species in the United States, however, several bacterial diseases including enteric septicemia of catfish (ESC), the columnaris disease, and the emerging Aeromonas disease threaten the sustainability of the catfish industry (Wagner et al. 2002), by causing millions of economical loss.

After realizing the limit of effectiveness and the environmental and human health risks of vaccine and antibodies, genetic improvement of disease resistance is an effective alternative for disease control in aquaculture (Vallejo et al. 1998). A set of genomic analysis of disease resistance could be conducted based on abundant and reliable resources, including a well-assembled and annotated reference genome sequence (Liu et al. 2016), high density SNP arrays (Liu et al. 2014; Zeng et al. 2017), high-density genetic linkage maps (Li et al. 2015; Liu et al. 2016), as well as the well-studied intraspecific and interspecific segregation populations (Waldbieser et al. 2001; Liu et al. 2003; Kucuktas et al. 2009; Ninwichian et al. 2012) to develop modern molecular genetic technology for disease resistance enhancement.

Marker-assisted selection has been suggested to be effective for genetic improvement of disease resistance because a very limited number of major QTL were found to control the vast

majority of phenotypic variance in the genetic and genomic studies with aquaculture species, including the studies of lymphocystis disease resistance (Fuji et al. 2006; Fuji et al. 2007) and *Vibrio anguillarum* resistance (Wang et al. 2014a) in Japanese flounder, infectious pancreatic necrosis disease resistance in Atlantic salmon (Moen et al. 2009; Moen et al. 2015; Houston et al. 2012), resistance to whirling disease (Baerwald et al. 2011), bacterial cold water disease (Vallejo et al. 2010; Liu et al. 2015) in rainbow trout, and columnaris disease resistance in catfish (Geng et al. 2015). All these studies identified only a limited number of major QTL accounted for a large proportion of phenotypic variations, indicating the effectiveness of marker-assisted selection genetically enhancing disease resistance with aquaculture species.

Channel catfish and blue catfish express dramatic differences to disease resistance (Dunham and Smitherman 1984). Channel catfish is much more resistant to columnaris than blue catfish, while blue catfish is generally much more resistant to ESC than channel catfish. The channel-blue hybrid catfish has been reported to exhibit intermediate levels of resistance (Wolters et al. 1996; Dunham et al. 2008).

Previous studies of ESC resistance in catfish revealed that early activation of the MHC class I pathway could be the basis for ESC disease resistance in blue catfish (Peatman et al. 2008), indicated that actin cytoskeletal polymerization/remodeling and junctional regulation might be associated with susceptibility for pathogen entry and subsequent inflammatory responses (Li et al. 2012), and detected the QTL located on LG 1, 3, 6, 9, 15, 17, 18, and 25, potentially associated with ESC resistance in hybrid catfish (Wang et al. 2013). Two GWAS directly identified the SNP and QTL associated with the phenotypic variation of ESC resistance in hybrid catfish. A total of three QTL located on LG 1, 12, and 16 were detected using third generation backcross interspecific hybrid populations (Zhou et al. 2017b) while three QTL, two located on LG1 and one on LG23

26

were detected in first generation of backcross interspecific hybrid populations (Tan et al. 2018b) . While these two studies provided important information of QTL for ESC disease resistance in the interspecific hybrid system, QTL analysis of ESC disease resistance within channel catfish intraspecific system has not been conducted. As a result, it is unknown if the identified QTL were of interspecific in nature since some of the resistant alleles could have originated from channel catfish. With the objectives of determining if the previously identified QTL were interspecific and if additional major QTL are operating within channel catfish, in this study, we took the advantage of the recently developed 690K catfish SNP arrays (Zeng et al. 2017) to conduct GWAS analysis of ESC disease resistance using channel catfish families, two families from Marion strain while one from Kansas Random strain. In addition, comparative analysis was conducted to provide insights into the origin of the resistance alleles.

2.3 Materials and methods

2.3.1 Ethics statement

All procedures involved in handling and treatment of fish for this work were approved by Auburn University Institutional Animal Care and Use Committee (AU-IACUC), prior to the start of the study. Tissue samples were collected after euthanasia. All animal procedures were carried out according to the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act in the United States.

2.3.2 Experiment fish, bacterial challenge, and sample collection

A total of 990 channel catfish from three families were used in this study. The fish were produced at the E.W. Shell Fisheries Research Center, Alabama Agricultural Experiment Station,

Auburn University, Alabama. Of the 3 families, two were Marion strain while the other was Kansas Random strain. The detailed ancestry information for these strains were previously reported (Dunham and Smitherman 1984). For each family, 300 fish were used for treatment group and 30 fish were used for control group, for bacterial challenges. The fish for treatment group were reared in three flow-through water tanks (400L) marked by "T1" "T2" and "T3" corresponding to family 1, family 2 and family 3, and they were challenged with *E. ictaluri* in the tanks. The fish for control group were all reared in one tank, with each of the 30 fish placed in a physically separated compartment. *E. ictaluri* used for challenge was freshly isolated from an obviously symptomatic fish, and the bacteria were biochemically confirmed as *E. ictaluri*. Then, the confirmed bacteria were cultured in Brain Heart Infusion (BHI) medium and incubated in a shaker incubator at 28 ℃ overnight. The bacteria were quantified by plating 10 ml of 10-fold serial dilutions onto BHI agar plates.

For bacteria challenge, the bacteria concentration of 4×10^6 CFU/ml was used for each of the three treatment group tanks. The fish were incubated with the challenging bacteria for two hours, then the flow of water was reinitiated on to prevent any water quality problems. For control groups, the same volume of BHI medium was added into the control group tank without bacteria. After the initiation of the challenge, fish were monitored real time around the clock. Blood samples were continuously collected from moribund fish, and the symptoms of ESC disease and death time were recorded. The challenge lasted for two weeks. The earliest 75 dead fish in each family were regarded as susceptible fish, while the latest dead fish and all surviving fish at the end of challenge, were regarded as resistant fish. Blood samples of all survival individuals were collected when the bacterial challenge finished.

Blood samples were collected for 288 fish and stored in a 15 mL tube containing 5 mL of cell lysis solution (10mL lysis solution, proteinase K 20 mg/mL) for DNA isolation. Time of death, body weight, and body length of the fish were all recorded. Unlike other traits, the disease resistance is a complex trait for which accurate phenotyping is difficult and can be affected by trait definitions and statistical models used in the analysis (Yáñez et al. 2014). Here, the phenotype, survival time of fish, were recorded individually during the ESC challenge. The survival time of fish varies from 0h (the first death happened) to 264h (the blood sample collection for surviving fish when the challenge ended). And it was treated as a continuous trait instead of the categorical trait in case-control analysis in this study. The resistance to ESC was considered as a continuous trait rather than a binary trait based on the following reasons: 1) the continuous data is better for small number of samples, for example, there were only 288 individuals analyzed in this study and the continuous data could avoid some bias compared to binary data; 2) the continuous data is more sensitive and powerful, which should produce more reliable results compared to binary data; and 3) the continuous data are adaptable to more methods of statistical analysis than binary data.

2.3.3 DNA extraction, genotyping, and quality control

Based on "selective genotyping" method (Darvasi and Soller 1992), 96 fish from each family were selected as the extremes of the disease resistance distribution of the 300 fish. Standard protocols were used for DNA extraction from these selected samples. First, the blood samples stored in the tube containing cell lysis solution were incubated at 55 ℃ overnight prior to DNA extraction. Then, the blood cells were broken by cell lysis solution, and protease K and protein precipitation solution were used to remove protein. Next, DNA was precipitated by isopropanol and collected by brief centrifugation, and the collected DNA was washed twice with 70% ethanol, then air-dried and resuspended in TE buffer. Nanodrop (Thermo Scientific, Wilmington, DE, USA)

was used to quantify the isolated DNA samples, and the integrity of the DNA samples was checked by electrophoresis on 1% agarose gels.

The catfish 690K SNP arrays with well-distributed markers were developed using Affymetrix Axiom genotyping technology (Zeng et al. 2017). Genotyping using the catfish 690K SNP array was performed by GeneSeek (Lincoln, NE, USA). Genotyping data were first screened to eliminate samples with low quality or low call rate ($< 95\%$), which excluded three samples from analysis. Affymetrix Genotyping Console was used to perform the SNP genotype calling. A total of 390,143 SNPs was kept for subsequent analysis after filtering out SNPs with a call rate of <95%, and a minor allele frequency (MAF) of <5%.

2.3.4 Statistical analysis

The SNP and Variation Suite (SVS, Version 8.0) was used to determine the potential associations between the SNPs and phenotypes. Linkage Disequilibrium (LD) pruning was carried out to generate independent SNPs, with a window size of 50 SNPs (the number of SNPs at each LD testing), a step of five SNPs (shift the window five SNPs forward), and r^2 threshold of 0.5. After LD pruning, 13,485 independent SNP haplotypes were kept for subsequent analysis. Identity by state (IBS) between all pairs of samples were estimated using the independent SNPs, and the population structure was assessed by the principal component analysis (PCA) with the independent SNP markers.

EMMAX (Efficient Mixed-Model Association eXpedited) (http://genetics.cs.ucla.edu/emmax/) analysis was conducted with the first two principal components and the fish body weight and body length as two covariates. We chose this statistical analysis because of its computational efficiency, based on both the effective EMMA algorithm

30

and avoiding repetitive variance component estimation procedure of this analysis. The model is as follows:

$$Y = X\beta + Z\mu + e$$

Where Y is the vector of phenotype which is the survival time in this experiment; X is the matrix of fixed effects including the two principal components, fish body weight, and fish body length; Z is the matrix of random effects; β and μ are the coefficient vector of fixed effects and random effects, respectively; Var (u) = $G\sigma_g^2$ where G is the simple IBS allele-sharing matrix and σ_g^2 is the additive genetic variance; e is the vector of random residuals.

The threshold P-value for genome-wide significance was calculated based on Bonferroni-correction with the estimated number of independent markers and LD blocks (Duggal et al. 2008). The threshold P-value of the 5% Bonferroni genome-wide significance was 0.05/13,485 [-log10(P-value) = 5.43]. The threshold P-value for the significance of "suggestive association" was set at one false positive effect in a genome-wide test, i.e., 1/13,485 [-log10(P-value) = 4.13].

To avoid any associations caused by family structure and stratification of the experimental fish, we used another statistical analysis, in addition to EMMAX. Family-based association test for quantitative traits (QFAM) was conducted using PLINK (Purcell et al. 2007) simple linear regression of phenotype on genotype. QFAM partitions the genotypes into between- and within-family components (Fulker et al. 1999; Abecasis et al. 2000). QFAM utilizes a special permutation procedure to correct for the family structure, and a within-family test with specify permutation of 104 was performed in this study. Using this method, the true association results could be identified without being affected by stratification (Purcell et al. 2007).

We also estimated variance components using fitting maximum likelihood variance component model by Linkage Disequilibrium Analysis for Quantitative and Discrete Traits (QTDT) (Abecasis et al. 2000) since these cannot be obtained from the analysis using PLINK.

2.3.5 Candidate genes and pathway analysis

Genes within ±0.5 Mb region of the significantly associated SNPs were identified using the channel catfish genome information of NCBI (National Center for Biotechnology Information). The predicted genes were annotated by BLAST. Functions of all these genes were examined to provide insights into their potential roles in response to ESC infection. GeneCards (http://www.genecards.org/), KEGG (the Kyoto Encyclopedia of Genes and Genomes) pathway database (http://www.genome.jp/kegg/pathway.html), and Reactome database (http://reactome.org/) were used for pathway analysis.

2.4 Results

2.4.1 Mortality rate and sample structure

At the end of the experiment, the cumulative mortality rate was 70.3%. The mortalities peaked between day 7-9 after the bacterial challenge. As shown in Fig. 1, the patterns of mortalities were similar overall, but with differences among families presumably due to differences of the genetic background of the experimental fish. Most mortalities occurred 6 days or later after the challenge, except for family 2, which had a significantly larger number of mortalities on day five after challenge. Overall, family 3 had a higher mortality rate than the other two families (Fig. 1).
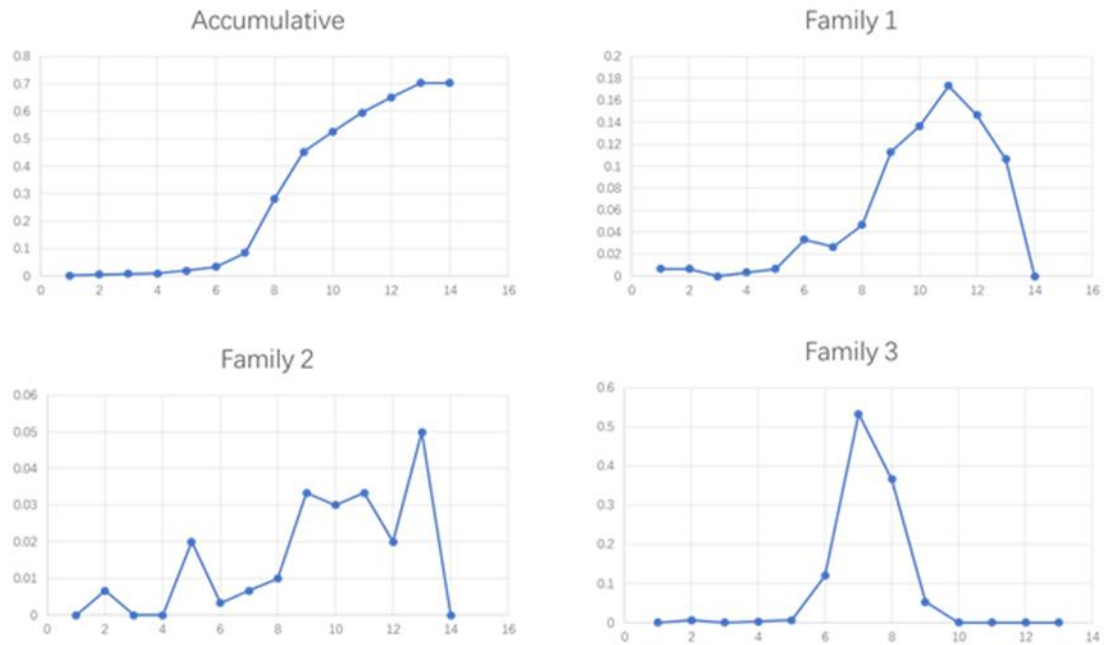
Figure 1. Mortality rate of channel catfish, *Ictalurus punctatus*, after Edwardsiella ictaluri infection utilizing an artificial tank challenge. Each tank has 300 fish in it. For all three families and the accumulative (all three families combined) graphs, the X-axis are days after infection while the Y-axis are mortality of fish. Family 1 and 2 are from Marion strain while family 3 was from Kansas Random strain.

Principal component analysis (PCA) was used to identify the sample structure with the first two principal components (Fig. 2). The three families were distantly related, and family 3 is notably more distantly related to families 1 and 2 than between themselves. The PCA analysis result was consistent with the origin of the strains (see Materials and Methods)
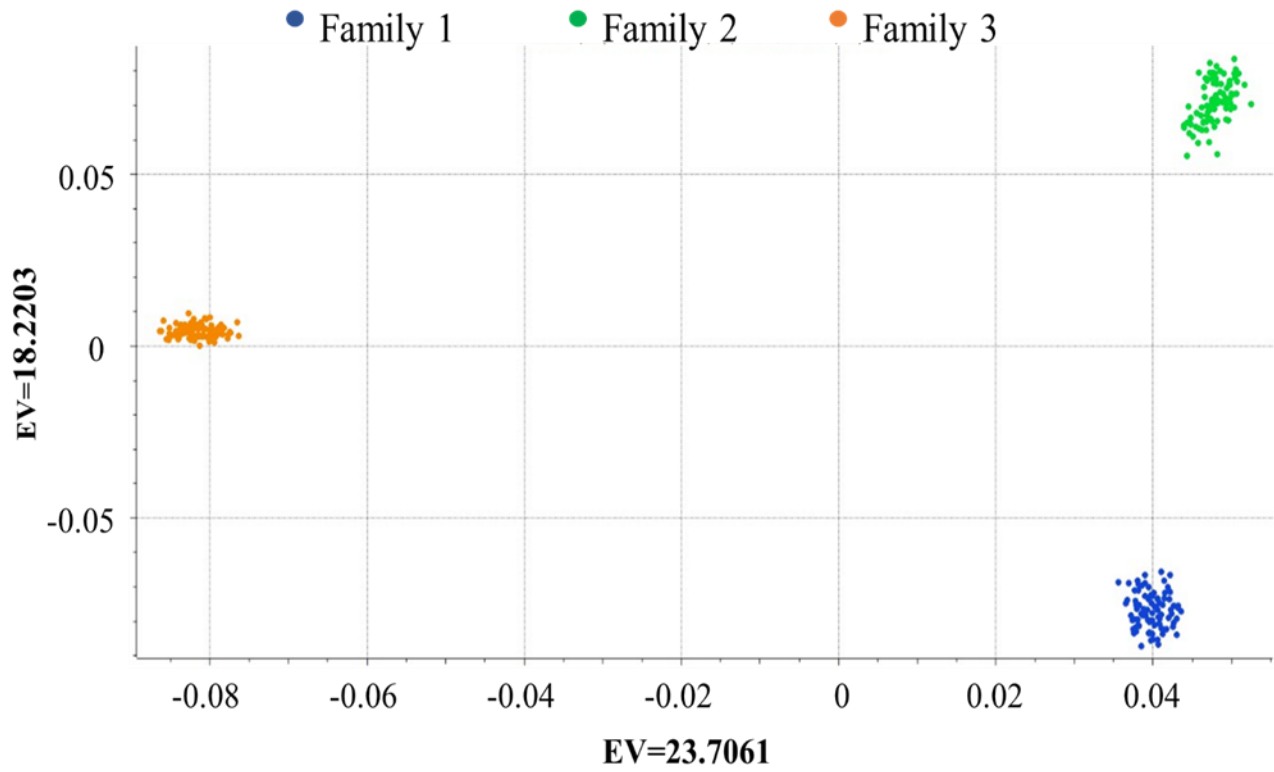
Figure 2. Principal component analysis of sample structures of 3 families of channel catfish, *Ictalurus punctatus*, using first two principal components. Each dot represents the individual fish in the experiment.

2.4.2 QTL associated with ESC resistance of channel catfish

A total of four linkage groups (LGs) were found to be associated with ESC resistance in channel catfish, with LG1 and LG26 containing significant SNP markers at the genome-wide levels (-log10(P-value) >5.43), and LG3 and LG21 containing suggestively significant SNP markers (-log10(P-value) >4.13) (Fig. 3). Ten SNPs in LG1 and three SNPs in LG26 were significantly associated with ESC disease resistance (Table 1). For LG1, the ten SNPs were located from 5,948,310 to 9,638,510 bp, a region spanning almost 3.7 Mb of physical distance. With such a broad range, detailed analysis of these associated SNPs became necessary as to if more than one QTL were involved (see below). In contrast to the situation on LG1, there was just one small region on LG26 containing significant SNPs (Fig. 3), located from 1,639,485 to 1,666,617 bp, spanning a distance of only 27 Kb. The most significant SNP in this region (AX-158328525) could explain about 8.24% of the phenotype variation, as analyzed using EMMAX. In addition to this most significant SNP, other SNPs located within this region could also explain a small part of phenotype variation, but all of them were less than 8.24%, suggesting these SNPs were all significant because of their linkage.
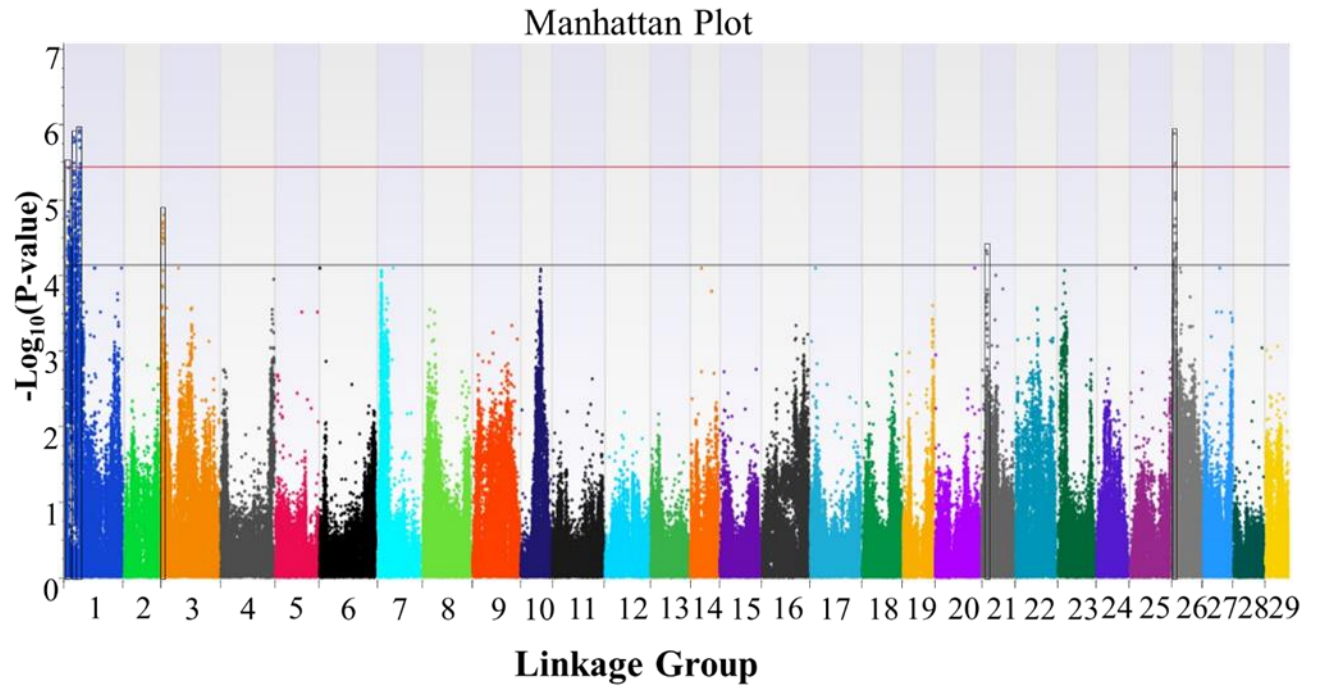
Figure 3. Manhattan plot of genome wide association analysis of channel catfish, *Ictalurus panctatus*, for enteric septicemia of catfish (ESC) resistance. The X-axis indicates chromosomes, and Y-axis indicates the P-values of the SNP markers. The red solid line indicates the threshold P-value for genome-wide significance. The black solid line indicates the threshold P-value for the significance of "suggestive association".

Table 1. Single nucleotide polymorphisms (SNPs) significantly associated with enteric septicemia of catfish (ESC) resistance in channel catfish, *Ictalurus punctatus*

Linkage group	SNP ID	Position (bp)	$-\log_{10}$(P-value)	% variance
1	AX-157892401	5,948,310	5.480	7.657
1	AX-157741156	6,019,208	5.824	8.172
1	AX-157892542	6,056,396	5.758	8.074
1	AX-157892547	6,061,469	5.776	8.101
1	AX-157894097	9,270,017	5.792	8.125
1	AX-157894099	9,276,432	5.916	8.311
1	AX-157894105	9,284,563	5.888	8.269
1	AX-157894109	9,290,743	5.782	8.109
1	AX-157894287	9,610,346	5.685	7.965
1	AX-157894300	9,638,510	5.467	7.637
26	AX-158328515	1,639,485	5.436	7.590
26	AX-158328525	1,648,547	5.870	8.242
26	AX-158328540	1,666,617	5.479	7.655

In addition to the significant QTL on LG1 and LG26, suggestive QTL were detected on LG3 and LG21, with 9 and 2 suggestively significant SNPs (4.13 < -log10(P-value) < 5.43) being detected, respectively (Table S1). On LG3, the suggestively significant SNPs were located between 802,388 to 1,215,683 bp, spanning approximately 413 Kb. The most significantly SNPs (AX-157921206) in this region could explain about 6.62% of the phenotype variation. On LG21, two suggestively significant SNPs were located at 3,278,086 bp and 3,284,725 bp, and these two SNPs (AX-158255403 and AX-158255412) could explain about 5.90% of the phenotype variation. The same QTL were detected by using both EMMAX and QFAM method in PLINK (Fig. S1). The estimation of variance component was conducted by QTDT (Abecasis et al. 2000), with which a typical model for the variance included environmental, polygenic, and additive component of variance. Because the allowed number of SNPs in QTDT is relatively small, we only used significant SNPs on LG1 and LG26 and suggestive significant SNPs on LG3 and LG21 to conduct the estimation. The following results were obtained: 1) the environmental component of variance is far smaller than polygenic and addictive components; 2) the polygenic component was necessary, which means these related QTL was highly heritable (p-value=2.33x10-19); 3) the additive component of variance was necessary in this model due to the p-values of all SNPs we selected were less than 0.05, which means there was strong evidence for the linkage of the associated SNPs within each QTL region; and 4) All selected SNPs showed the strong disequilibrium with the disease mutation, which was consistent with results from other two analysis.

2.4.3 Detailed analysis of SNPs associated with ESC resistance on LG1

Although 10 SNPs were found to be significantly associated with ESC resistance on LG1, they spanned a very large genomic region of 3.7 Mb from 5,948,310 to 9,638,510 bp. It became

necessary to further analyze the distribution of these significant SNPs along with other neighboring SNPs to determine if more than one QTL were involved in the associated region on LG1. As shown (Fig. 4), a very large number of SNPs, a total of 279, were found to be above the suggestively significant level for associations with ESC disease resistance. Detailed analysis of these SNPs and their associated P-values indicated that this region on LG1 could involve three distinct QTL: One including the SNP with the lowest P-value on LG1, AX-157894099, located between 9,270,017 to 9,638,510 bp, spanning a region of about 368 Kb. This region could explain about 8.31% of the phenotype variation. A second QTL region, located between 5,948,310 to 6,061,469 bp, spanning about 113 Kb. The most significant SNP in this region, AX-157741156, could explain about 8.17% of the phenotype variation. As shown in Figure 4A, there are many SNPs between the two significant QTL regions, but with higher P-values. In addition to these two significant QTL, a third QTL appeared to be on LG1 between 1,833,304 to 5,353,131 bp, including one SNP at 2,411,070 bp almost reaching the significant level, while many SNPs between this and the significant QTL had P-values higher than those within this suggestive QTL or those within the significant QTL, suggesting it to be a distinct QTL (Fig. 4). The most significant SNP (AX-157890714) in this region could explain about 7.56% of the phenotype variation.
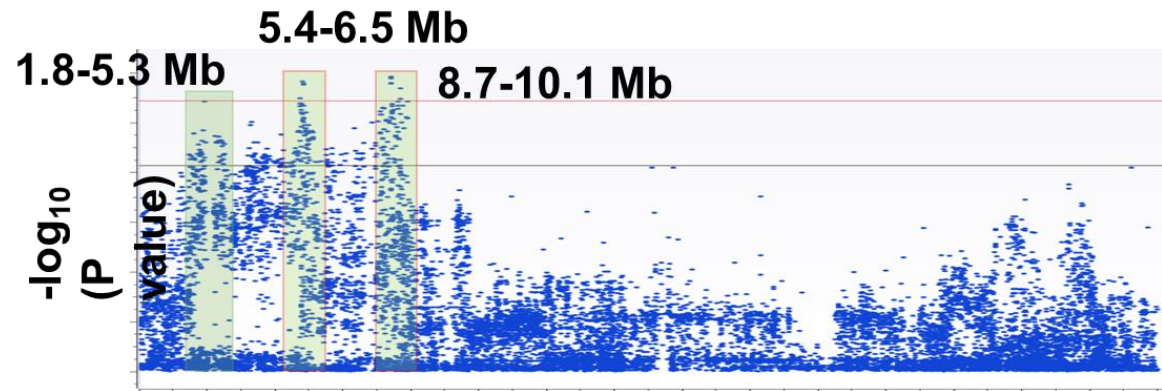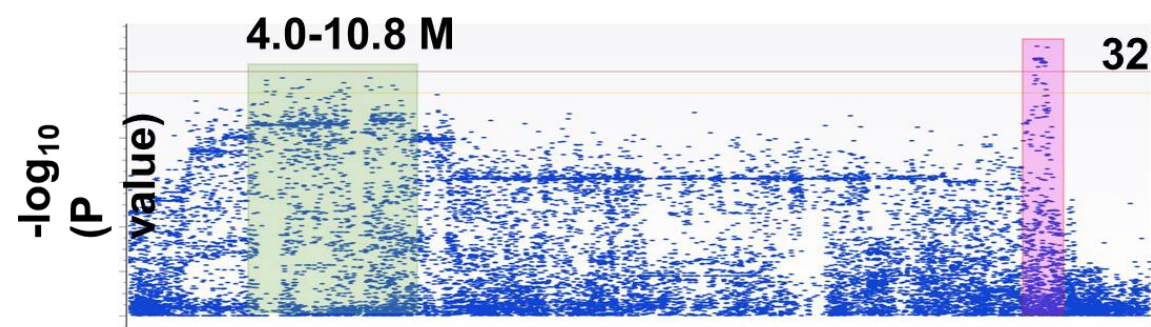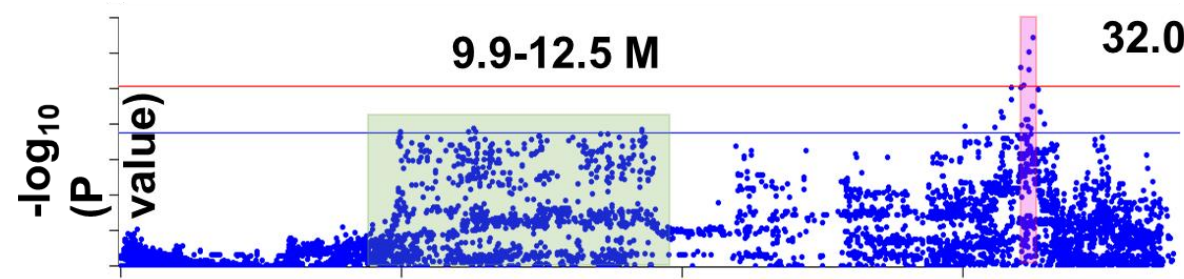
Figure 4. Detailed analysis of single nucleotide polymorphisms (SNPs) associated with enteric septicemia of catfish (ESC) resistance on linkage group (LG)1 of three populations of catfish, including first generation of hybrid backcross population, third generation of hybrid backcross population, and channel catfish (*Ictalurus punctatus*). SNPs associated with ESC resistance on LG1 in channel catfish, first generation backcross of interspecific hybrid catfish (Tan et al. 2018b), and in fourth generation backcross of interspecific hybrid catfish (Zhou et al. 2017b). The red boxes indicate the statistically significant QTL. The green boxes indicate the statistically suggestive QTL. Solid red line indicated the threshold P-value for genome-wide significance, and the other horizontal line indicated threshold P-value for the significance of "suggestive association", respectively.

2.4.4 Correlation of the SNPs associated with ESC resistance

Conditional analyses were conducted to examine the correlation of the SNPs associated with ESC resistance (Nishimura et al. 2012). In the mixed linear model, the genotype of the most significant SNP on LG1 (AX-157894099) was included as a covariate. Under this condition, the -log10(P-value) of all SNPs located on LG1 dropped below suggestive level, while the -log10(P-value) of all SNPs located on other LGs remained significant, suggesting strong correlations among these associated SNPs on LG1. The same analysis was also conducted on LG26, LG3, and LG21, and similar results were observed.

2.4.5 Genes within significant QTL regions associated with ESC resistance

In this study, the genomic region within the 0.5 Mb upstream and downstream around the significant SNPs on LG1 and LG26 were considered as the significant QTL regions. We examined the genes within these regions. There were 50, 54, and 24 genes on the three significantly associated regions, respectively (Table S2). In the subsequent analysis, we were more interested in genes related to immunity since these genes could be important in the process of disease resistance. Among the 128 genes located in the three regions, 25 genes were involved in immunity, and they included uba5, nrros, mylip, dync1i1, fbxo32, dap12, dap10, prss3, pbxip1, tlr13, s100a1, cgn, il6r, setdb1, trim39, clec4e, fbxw7, ddr1, tubb, abhd16a, nectin2, pou2f2, tia1, atp6v0a2, and nefl.

In addition to the three significantly associated regions, the genes located within the three suggestively associated regions on LG1, LG3, and LG21 were also examined. A total of 51 genes were found to be near the suggestively significant associated SNPs (Table S3), of which 12 were

involved in immunity. These included asb10, scnn1a, mylk, ccr7, crhr1, nlrc3, adcy2, ptprj, nlrp12, tmeff1, prf1, and coro1a. Along with the 25 immunity genes within the significantly associated QTL regions, a total of 37 genes involved in immunity were identified (Table 2).

Table 2. Functions and pathways of genes related to immunity located with quantitative locus (QTL) associated with enteric septicemia of catfish (ESC) in channel catfish, *Ictalurus punctatus*, identified from genome-wide association study (GWAS)

gene	function	pathway
uba5	activates UFM1 and SUMO2	Class I MHC mediated antigen processing and presentation;
nrros	host defense	
mylip	mediates ubiquitination and proteasomal degradation	Class I MHC mediated antigen processing and presentation;
dync1i1	regulate dynein function	MHC class II antigen presentation;
fbxo32	mediates ubiquitination and degradation of target proteins	Class I MHC mediated antigen processing and presentation;
dap12	neutrophil activation	DAP12 interactions; Neutrophil degranulation;
dap10	immune surveillance and cytotoxicity	Class I MHC mediated antigen processing and presentation; Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell;
prss3	definsin processing in ileum	Antimicrobial peptides; Neutrophil degranulation;
pbxip1	regulation of pre-B-cell leukemia transcription factors	
tlr13	key role in innate immune system	
s100a1	regulation of cellular processes	Toll-Like Receptors Cascades;
cgn	immune cell transmigration	Blood-Brain Barrier and Immune Cell Transmigration:
il6r	regulation of immune response	Signaling by Interleukins;
setdb1	gene silencing	
trim39	facilitate apoptosis	Class I MHC mediated antigen processing and presentation;
clec4e	pathogenesis recognition and induce inflammatory cytokines	C-type lectin receptors (CLRs);
fbxw7	mediates ubiquitination and degradation of target proteins	Class I MHC mediated antigen processing and presentation;
ddr1	role in tumor cell invasion	
tubb	component of microtubules	Neutrophil degranulation;

abhd16a	Immunity	
nectin2	modulator of T-cell signaling	
pou2f2	activating immunoglobulin gene expression	B cell receptor signaling pathway
tia1	against cytotoxic lymphocyte target cells	IL-10 Pathway
atp6v0a2	maintain golgi function	Innate Immune System;
nefl	maintenance of neuronal caliber	Signaling by interleukins; DAP12 interactions;Fc epsilon receprot (FCERI) signaling;
asb10	mediates ubiquitination and degradation of target proteins	Class I MHC mediated antigen processing and presentation;
scnn1a	control electrolyte and blood pressure hemeostasis	
mylk	inflammatory response;	Blood-Brain Barrier and Immune Cell Transmigration: VCAM-1/CD106 Signaling Pathways;
ccr7	activates B and T lymphocytes	Chemokine signaling pathway;
crhr1	regulate immnue response	
nlrc3	T-cell activation	Cytosolic sensors of pathigen-assciated DNA; NOD-Like recptor Signaling Pathways;
adcy2	increase IL6 production	DAP12 interactions;
ptprj	negativly regulate T cell receptor signaling	Class I MHC mediated antigen processing and presentation; Neutrophil degranulation;
nlrp12	regulate inflammatory response	NOD-Like receptor Signaling Pathways;
tmeff1	tumor suppressor in brain cancers	
prf1	kill non-self cells	Allograft rejection; Graft-versus-host disease; Autoimmune thyroid disease; Natural killer cell mediatd cytotoxicity;
coro1a	prevent fusion between phagosomes and lysosomes	

2.5 Discussion

In this study, a total of six genomic regions were found to be associated with ESC disease resistance within channel catfish families, of which three were significant QTL and three were suggestively significant at the genomic level. Two of the three significant QTL were on LG1 and the other one on LG26. The three suggestively significant QTL were found to be on LG1, LG3, and LG21, respectively.

In a recent GWA study of ESC resistance in the third generation of interspecific hybrid backcross populations (Zhou et al. 2017b), one significant QTL on LG1 and two suggestive QTL on LG12 and LG16 were detected. To verify if the QTL identified on LG1 in this study was overlapped with those identified from the previous study, we mapped the significant SNPs detected by Zhou et al. (2017b) to the latest genome assembly used by this study (Table S4). In the study of Zhou et al. (2017b), they believed the significant QTL region on LG1 was from about 32.0 Mb to 32.5 Mb (Fig. 4), notably, this QTL was also identified using the first generation of interspecific hybrid backcross populations (Tan et al. 2018b). At the same time, Zhou et al. (2017b) identified a suggestively significant QTL located from 9.9 Mb to 12.5 Mb of LG1 and Tan et al. (2018b) identified a broad suggestively significant QTL located from 5.0 Mb to 9.8 Mb of LG1, overlapping with the second significant QTL region detected in the present study on LG1 (Fig. 4). This overlapped genetic region detected by three GWA studies indicated that the origin of this QTL on LG1 in hybrid catfish was from channel catfish. Meanwhile, we emphasized the importance of several functional genes (discussed later) in channel catfish. These annotated genes within QTL associated with ESC resistance in channel catfish were also detected in another GWA

47

study in the interspecific hybrid catfish even if the hybrid fish carried more significant QTL and major resistant genes from blue catfish origin (Zhou et al. 2017b; Tan et al. 2018b).

In addition to the overlapped QTL, the other QTL were novel, identified from the channel catfish families, but not from the interspecific hybrid backcross families. We believed the major difference of the detected QTL was mostly due to use of different fish populations. As presented in the introduction, blue catfish is much more resistant against ESC disease than channel catfish. We believed that some significant QTL detected in interspecific hybrid catfish with stronger resistance to ESC were originated from blue catfish, such as those on LG1 (not overlapped with the QTL identified from channel catfish in this study), and those on LG 12, 16, and 23 (Zhou et al. 2017b; Tan et al. 2018b). These regions were not detectable in channel catfish due to their origins of blue catfish. For instance, NCK-related resistance was significantly important for interspecific hybrid catfish (Zhou et al. 2017a). Nevertheless, it was not functional in channel catfish. The use of channel catfish families in this study allowed detection of those QTL operating within channel catfish populations. Indirectly, this point was also supported by the results obtained from BSR-seq analysis (Wang et al. 2013). In that study, significantly differentially expressed SNPs were detected on LG1, 3, 6, 9, 15, 17, 18, and 25 using second generation interspecific hybrid backcross catfish. Additional significant SNPs were detected from several other linkage groups, perhaps due to the use of second generation interspecific of backcross families. Remarkably, QTL on LG1 were detected from all three studies, validating the importance and broad involvement of the QTL on LG1 in ESC resistance in catfish.

There were probably three distinct QTL on LG1, with two of the three reaching genome level significance while the third one reaching suggestively significant level. Conditional analysis

indicated that they may not be independent because when we treated the most significant SNP as a covariate, all the remaining significant SNPs were no longer significant. Based on the results of this conditional analysis, it was proposed that all the three peaks might be within one single QTL, however, this region might contain three QTL due to the following reasons: 1) there are clearly three peaks of significant or suggestively significant SNPs (Fig. 4). Between each of these significant SNP peaks, there are many SNPs at higher P-values. Many SNPs with lower P-values cannot be just incidental; 2) the conditional analysis cannot separate these QTL because they have close linkage. All three QTL were located within just a few centiMorgans of genetic distance; 3) if there was just one QTL, it would be difficult to explain the many SNPs with higher P-values between each SNP low P-values peaks. Given the recombination frequency, many of significantly associated regions in catfish has been just several hundred kilobases (Geng et al. 2015; Geng et al. 2016; Zhou et al. 2017b; Geng et al. 2017; Wang et al. 2017b; Zhong et al. 2017; Jin et al. 2017). The similar associated physical distances were observed in this study for QTL on LG3, LG21, or LG26 (Table S1).

With a well-assembled and well-annotated reference genome sequence (Liu et al. 2016b), it is possible to identify the genes within and surrounding the identified QTL regions. The detailed analysis of genes surrounding the most significant SNPs (Figs. 5, 6, and 7) and suggestive SNPs (Table S3) were conducted. A total of 182 genes were found within 0.5 Mb from the most significant SNPs within the associated regions, of which 37 genes have known functions in immunity.
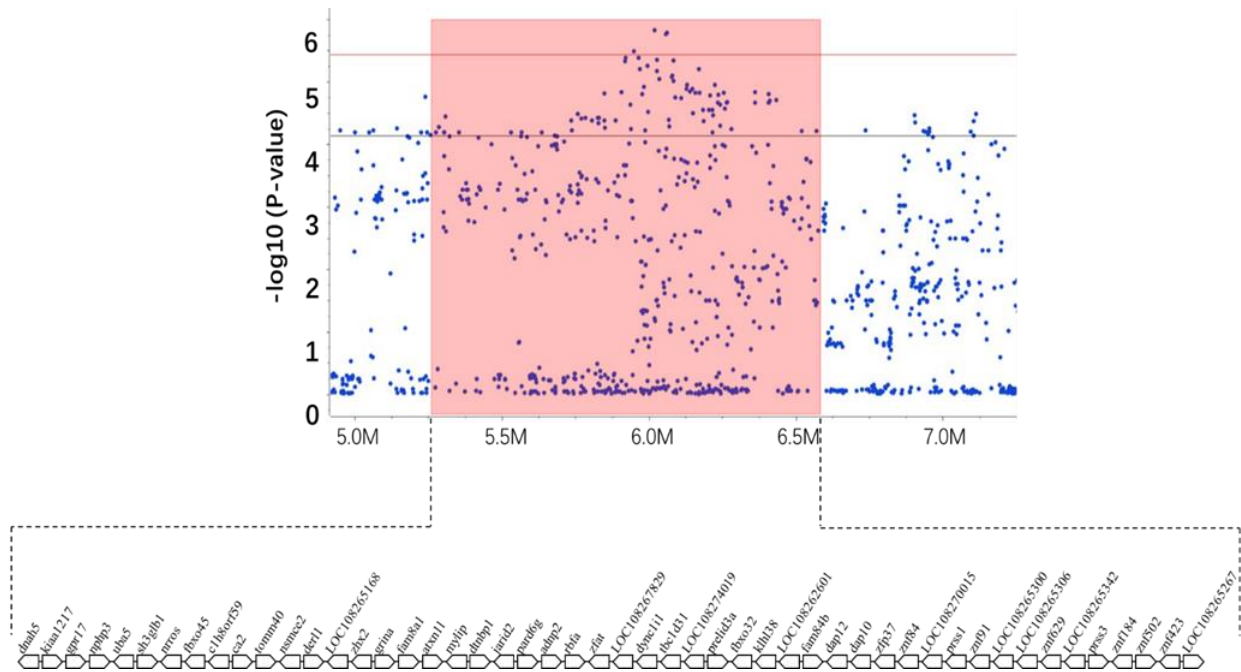
Figure 5. Genes surrounding the first set of significant SNPs (around 6Mb) associated with enteric septicemia of catfish (ESC) resistance on LG1 of channel catfish (*Ictalurus punctatus*) identified from genome-wide association study (GWAS). The red shade indicates ±0.5 Mb genomic region of the most significant SNPs. The red solid line indicates the threshold P-value for genome-wide significance. The black solid line indicates the threshold P-value for the significance of suggestive association.
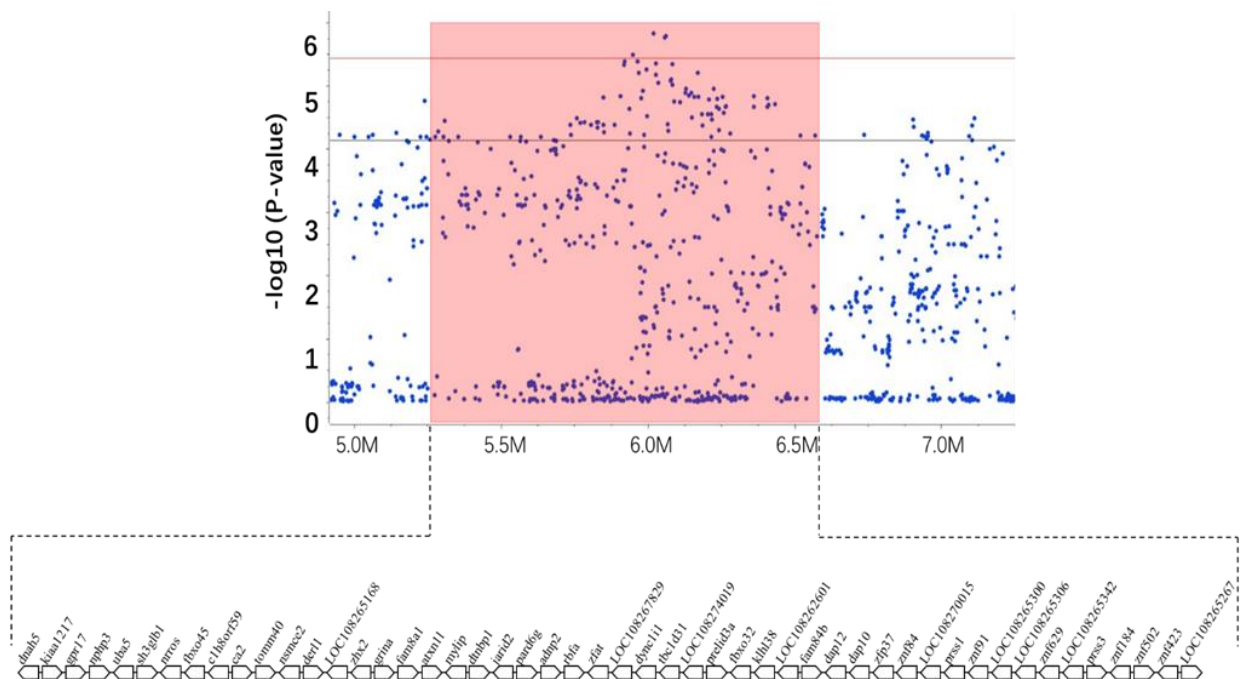
Figure 6. Genes surrounding the second set of significant SNPs (around 9Mb) associated with enteric septicemia of catfish (ESC) resistance on LG1 of channel catfish (*Ictalurus punctatus*) identified from genome-wide association study (GWAS). The red shade indicates ±0.5 Mb genomic region of the most significant SNPs. The red solid line indicates the threshold P-value for genome-wide significance. The black solid line indicates the threshold P-value for the significance of suggestive association.
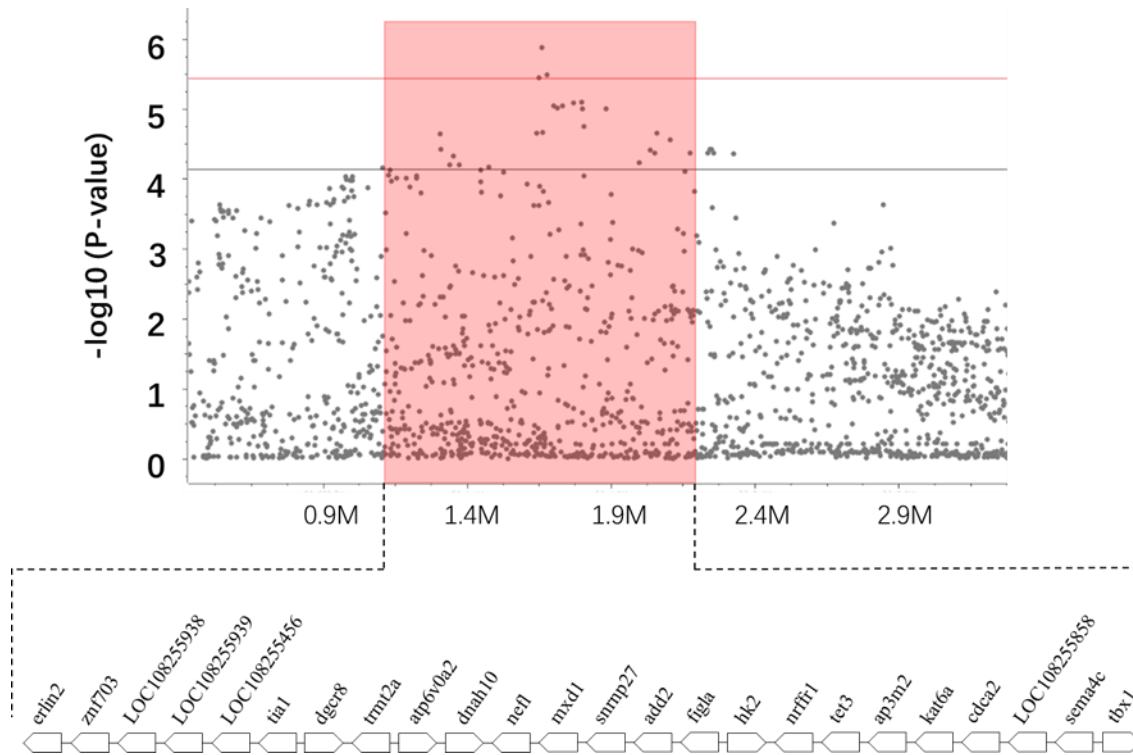
Figure 7. Genes surrounding the significant SNPs associated with enteric septicemia of catfish (ESC) resistance on LG26 of channel catfish (*Ictalurus punctatus*) identified from genome-wide association study (GWAS). The red shade indicates ±0.5 Mb genomic region of the most significant SNPs. The red solid line indicates the threshold P-value for genome-wide significance. The black solid line indicates the threshold P-value for the significance of suggestive association.

Within the first significant QTL regions on LG1, a total of eight genes including *uba5, nrros, mylip, dync1i1, fbxo32, dap12, dap10,* and *prss3* were identified to be involved in immune responses. Among them, gene *fbxo32* was up-regulated at 3h and then down-regulated at 24h and 3d in channel catfish intestine (Li et al. 2012), and gene *prss3* expressed at lower level in both channel catfish and resistant hybrid catfish compared to susceptible hybrid catfish (Li et al. 2012; Wang et al. 2013) after *E. ictaluri* infection. These two genes could be involved in ESC resistance because: 1) they were both located within the significant QTL regions identified in this study; 2) they were both differentially expressed between susceptible and resistance catfish after ESC infection. Gene *fbxo32* exerted increased expression during a proinflammatory antibacterial response but not following a viral infection, demonstrating the highly association between the transcription of this gene and bacterial infection, for instance with *E. ictaluri* (Tacchi et al. 2010). Gene *prss3* was reported its general roles in the immune defense system due to the broad distribution in the spleen, liver, kidney, and brain of normal mice (Koshikawa et al. 1998).

A total of 14 immune related genes (*pbxip1, tlr13, s100a1, cgn, il6r, setdb1, trim39, clec4e, fbxw7, ddr1, tubb, abhd16a, nectin2,* and *pou2f2,*) were identified within the second significant QTL on LG1. Notably, gene *trim39* was down-regulated at 3h and 24h, and then up-regulated at 3d in channel catfish intestine after *E. ictaluri* infection (Li et al. 2012), and gene *tubb* was induced in the liver of ESC resistant hybrid catfish (Wang et al. 2013). Gene *trim39* was reported to act as an important regulator of fish innate immune response (Wang et al. 2016). Remarkably, *btr* subfamily genes, the orthologous to *trim39*, was upregulated upon *Aeromonas hydrophila* infection and showed significant anti-microbial activity in zebrafish (Zhang et al. 2015). Gene *tubb* was also detected up-regulation after infection with *A. hydrophila* in grass carp

(*Ctenopharyngodon idellus*) (Yang et al. 2016). In contrast, only three genes were identified to be involved in immune system within the significant QTL on LG26, including *tia1, atp6v0a2,* and *nefl*. The identification of *trim39* and *tubb* in the significant QTL, and their differential expression after ESC infection (Li et al. 2012; Wang et al. 2013) suggest their roles in disease resistance.

Additionally, genes involved in immunity surrounding suggestive SNPs were also detected. A total of 10 genes located on LG1 were identified to be immunity related, including *asb10, scnn1a, mylk, ccr7, crhr1, nlrc3, adcy2, ptprj, nlrp12,* and *tmeff1*. Among them, *nlrc3* and *nlrp12*, belonging to NLR family, were also detected by the previous GWA study (Zhou et al. 2017b). Members of the NLR gene family acted as important mediators of innate mucosal response in catfish (Peatman et al. 2015). Both gene *nlrc3* and *nlrp12* were also reported down-regulation in the liver of resistant hybrid catfish compared to susceptible hybrid catfish (Wang et al. 2013) and up-regulation in the channel catfish intestine after *E. ictaluri* challenge (Li et al. 2012), all indicating these two genes were important players for ESC resistance. Meanwhile, *mylk* was also expressed at a higher level in the liver of resistant hybrid catfish after the *E. ictaluri* infection (Wang et al. 2013) and gene *adcy2* was found to be one of the key differentially expressed genes in Japanese flounder after infection of *E. tarda* (Li et al. 2018b). In contrast, only two genes (*prf1* and *coro1a*) related to immunity were detected within the QTL on LG3. Gene *prf1* was involved in immune process for infectious disease (Trapani and Smyth 2002). Gene *coro1a* was reported the differential expression in response to infection of *E. tarda* in cells using proteomic analysis (Qin et al. 2017), and down-regulation in response to PRV (*piscine orthoreovirus*) infection in Atlantic salmon (Dahle et al. 2015). No genes related to immunity were detected within the QTL on LG21. The extensively involvement of these genes in ESC as well as other infectious disease

make them promising candidates for disease resistance. Future studies are required to explore their roles in disease resistance.

Pathway analysis was conducted to provide more evidence for the association of positional candidate genes with ESC disease resistance. Eight genes identified in this study were involved in Class I MHC mediated antigen processing and presentation pathway. Similarly, the induction of several MHC Class I related pathways were observed in blue catfish liver after the infection of *E. ictalurid* (Peatman et al. 2008). Taken together, these findings suggest that Class I MHC related pathways play important roles in ESC resistance. NOD-Like Receptor (NLR) Signaling Pathway was also a well-studied pathway associated with immune system. Its involvement in ESC resistance was also observed in our previous GWA study (Zhou et al. 2017b). In addition, members of the NLR gene family seemed to be the important mediators of innate mucosal response in catfish (Peatman et al. 2015). Except the two pathways we just discussed above, the immune genes we identified were involved in many other pathways related to immune response (Table 2), indicating the relatively complex genetic architecture of resistance to ESC disease in catfish.

In summary, three significant QTL and three suggestive QTL were identified in the present study associated with ESC resistance in channel catfish using genome-wide association study. A total of 13 significant and 311 suggestively significant SNPs were identified at genome-wide level. Many genes within the QTL regions were also differentially expressed between resistant and susceptible fish, suggestion their positional and expressional candidacy for roles in ESC disease resistance. Future studies are required to pin down the candidate genes, leading to the eventual identification of the causal genes.

Chapter 3 Genome-wide H3K4me3 profiling of catfish by ChIP-seq reveals the landscape of its promoter sequence

3.1 Abstract

Orchestrating epigenetic alternations were considered to be very important for the explanation of phenotypic complexity. Histone modifications were a cluster of conserved and effective epigenetic regulators. As the development of technology, chromatin immunoprecipitation followed by sequencing (CHIP-seq) has become the most powerful method to screen histone modifications. Although the importance of these epigenetic regulators was realized for decades, the related studies mainly focused on model species. Catfish, as one of the most important aquaculture species, accounts for over half of the aquaculture production each year in United States. The reference genome sequences of catfish have been published, however the fine-structure mapping of genes especially for the promoter sequencings were lacking. In this study, we conducted the genome-wide tri-methylated histone H3 lysine 4 (H3K4me3) profiling by ChIP-seq to obtain the landscape of catfish promoter sequences. As a result, a total of 14,464 active promoters and 14,346 genes regulated by these promoters were identified. By comparing with other species, the low GC-content and specific structure of TATA-box in catfish were recognized. Analyses of two tissues allowed to obtain tissue-specific active promoters. 331 and 363 protein-coding genes were recognized to be regulated by tissue-specific active promoters in liver and intestine, respectively. The promoter of gene agtr1 and serpinf2 were active in liver but entirely inactive in intestine, nevertheless, promoter of gene ca4 was opposite. The liver specifically activated genes were highly enriched in regulation of blood pressure and smooth muscle contraction while the intestine specifically activated genes were mainly involved in regulation of

actin filament polymerization. This analysis laid a solid foundation for further histone modification epigenetic regulators studies in catfish for other economically important traits including growth, stress tolerance, and disease resistance in the future.

3.2 Background

There are roughly 4,100 species in the order of *Siluriformes* (catfish), which comprise 12% of all fish species and 6.3% of all vertebrate species. Hence, the comprehensive study of catfish genome sequence is of great significance. As the a well-assembled and annotated reference genome of catfish has been produced (Liu et al. 2016b), a set of genomic and genetic analysis associated with economically important traits of catfish have been conducted, including growth (Li et al. 2018a), head size (Geng et al. 2016), body confirmation (Geng et al. 2017), tolerance to hypoxia (Wang et al. 2017b; Zhong et al. 2017), tolerance to heat stress (Jin et al. 2017), and bacterial disease resistance (Geng et al. 2015; Zhou et al. 2017b; Tan et al. 2018b; Shi et al. 2018). However, all these studies only suggested the function of genetic variation of DNA sequences, analysis of epigenetic regulatory elements in catfish are still limited. Recently, it was realized that activation of eukaryotic gene transcription involved the orchestrate coordination of a multitude of epigenetic alternations of DNA sequences such as cis-regulatory elements and varied chromatin structure, such as chromatin accessibly and histone modifications (Lemon and Tjian 2000; Orphanides and Reinberg 2002; Nightingale et al. 2006).

Cis-regulatory regions immensely regulate the transcription of a gene. Promoters, as an important type of cis-regulatory elements, are located at the 5′ ends of genes immediately surrounding the transcriptional start site (TSS) and serve as the point of assembly of the transcriptional machinery and initiation of transcription (Smale and Kadonaga 2003). The genome-

wide analysis of promoters became an attractive point, and recently, many new next-generation DNA-sequencing technologies have allowed these studies in eukaryotes, including *C. elegans* (Dupuy et al. 2007), Drosophila (Ohler 2006), zebrafish (Aday et al. 2011), chicken (Abe and Gemmell 2014), and mammals (Barrera et al. 2008). Up to now, enormous studies of promoter sequence have been conducted and for some species, the precise promoter sequences have been produced. However, most studies about promoters only aimed at important mammals or model species till now. The fine-structure mapping of catfish promoter sequences is still lacking.

TATA box was one of the most conserved and ancient elements, which played a key role in regulation of gene transcription by provide a binding site for TBP (TATA-binding protein) (Struhl et al. 1998). Based on several criteria such as a maximum sequence length (8 bp), minimal consensus sequence, confined upstream location, and conservation across orthologous upstream regions, TATA box was defined as TATA(A/T)A(A/T)(A/G) (TATAWAWR) (Basehoar et al. 2004). However, a wide variation was found in the percentage and precise sequence structure of TATA box-containing promoters reported in different studies. Although the identification of TATA box has been conducted for a long time, the detailed TATA box structure is still flawed in catfish and the comparative analysis of TATA box structure variation is also limited. Like TATA-box, the GC-content is another significant feature of promoter sequence. Similarly, the GC-content of catfish promoter regions has not been conducted before due to the lack of fine-structure mapping of catfish promoters.

Histone modifications is a covalent post-translational modification to histone proteins. Among over 60 identified types of histone modification, histone methylation is the most popular and well-studied. H3K4me3 was demonstrated as one of the least abundant histone methylations

58

in the genome but highly enriched in gene promoter regions. Genome-wide high-resolution profiling studies of H3K4me3 indicated its characteristic that H3K4me3 located surround the TSS and was preferentially associated with the promoters of active genes in many eukaryotic species including yeast (Bernstein et al. 2002; Santos-Rosa et al. 2002; Ng et al. 2003; Pokholok et al. 2005), *Drosophila* (Schübeler et al. 2004), chicken (Schneider et al. 2004), mouse and human (Barski et al. 2007; Bernstein et al. 2005). All the studies suggested that the histone modification H3K4me3 is a hallmark of transcription initiation at most genes and has been convincingly shown to broadly occur at the promoters of most protein-coding genes (Guenther et al. 2007). Hence, H3K4me3 was a logical choice to detect promoter regions in non-model teleost, catfish, as an epigenetic mark (Kratochwil and Meyer 2015b).

Nowadays, as the development of technology, it is possible to directly detect the genome-wide enrichment of histone modifications using pertinent antibodies. CHIP-seq has become the most effective and efficient tool for histone modification identification and has been used in different species for different function including identification of transcription factors (Schmidt et al. ; Ouyang et al. 2009; Jothi et al. 2008; Ramagopalan et al. 2010; Pilon et al. 2011), detection of conserved regulatory regions (Blow et al. 2010; Gjoneska et al. 2015; Wang et al. 2011), and analysis of tissue-specific regulatory elements (Visel et al. 2009; Schmidt et al. 2010a; Bonn et al. 2012). ChIP-seq was also used in some aquaculture species including clarifying *globin* loci and the regulation of *globin* switching (Ganis et al. 2012), the role of imbalance of histone methylation dynamics in X-linked mental retardation (Qi et al. 2010), histone modification marks during early development in zebrafish (Bogdanovic et al. 2013), and cis-regulatory features in the embryonic genome in zebrafish (Aday et al. 2011), identifying conserved cis-regulatory nodes by comparative

epigenomic analysis (*Oryzias latipes*) (Tena et al. 2014); detection of active promoters in in Nile

tilapia (*Oreochromis niloticus*) (Kratochwil and Meyer 2015a), identifying the landscape of CTCF

binding landscape in jawless fish (Kadota et al. 2017), and demonstrating the role of epigenetic

modification to the immune response during behavioral fever in Atlanta Salmon (*Salmo salar*)

(Boltana et al. 2018). However, the genome-wide profiling of H3K4me3 to fine-structure map the

precise promoter regions and identify the characteristics of catfish promoter sequence have never

been conducted.

3.3 Materials and methods

3.3.1 Experimental fish, sample collection, and ethics statement

First generation of interspecific backcross catfish by mating male hybrid of channel and

blue catfish with female parent channel catfish were used in this study. Liver and intestine of

experimental fish were collected. In this study, liver and intestine from five fish were pooled

together and there were two replications for each tissue.

All procedures involved in handling and treatment of fish for this work were approved by

Auburn University Institutional Animal Care and Use Committee (AU-IACUC), prior to the start

of the study. Tissue samples were collected after euthanasia. All animal procedures were carried

out according to the Guide for the Care and Use of Laboratory Animals and the Animal Welfare

Act in the United States.

3.3.2 ChIP sequencing (ChIP-seq)

Liver and intestine of euthanatized experimental fish were sampled and frozen in the liquid

nitrogen immediately. After sampling, all collected tissue samples were stored into the -80℃

refrigerator. Then, ChIP-seq experimental step was performed based on Active Motif's HistonePath method (Active Motif) (Ramagopalan et al. 2010). To get chromatin, samples were fixed with 1% formaldehyde at room temperature for 15 min and stopped the fixation with 0.125 M glycine. Chromatin was then isolated by adding lysis buffer and disrupted with a Dounce homogenizer. Lysates were sonicated with a microtip to shear the DNA, with an average length of 300-500 bp, then cleared and stored at -80℃. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were resuspended, and the purified DNA was quantified on a NanoDrop spectrophotometer.

For chromatin immunoprecipitation reaction, 30 ug of extracted chromatin was precleared with A agarose beads (Invitrogen) and 4 ug of antibody against H3K4me3 were set up to react with the precleared chromatin. After the immunoprecipitation reaction, the immune complexes were washed, eluted from the beads with SDS buffer, and subjected to Rnase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65 ℃, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation.

Quantitative PCR (QPCR) reactions were carried out in triplicate on specific genomic regions using SYBR Green Supermix (Bio-Rad). The resulting signals were normalized for primer efficiency by carrying out QPCR for each primer pair using Input DNA.

ChIP DNA was amplified by PCR and Illumina sequencing libararies were prepared by the standard consecutive enzymatic steps. After a final PCR amplification step (18 cycles), the resulting DNA libraries were quantified and sequenced on Illumina's NextSeq 500 (75 nt reads, single end).

3.3.3 ChIP-seq reads alignment and ChIP Peak identification

Sequenced reads from ChIP-seq were aligned to published catfish genome (assembly IpCoco_1.2) (Liu et al. 2016b) using BWA algorithm (Li and Durbin 2009) with default settings . The reads passed the purity filter of Illumina, aligned with no more than 2 mismatches, and mapped uniquely to the catfish genome (mapping quality >= 25) were used in the subsequent analysis. Duplicate reads were also removed. Aligned reads were extended *in silico* at their 3'-ends to length of average genomic fragment length in sized-selected library, usually 200 bp. To identify the density of extended reads along the genome, the genome is divided into 32-nt bins and the number of reads in each bin is determined. ChIP peak location on genomic regions were determined by two main callers: MACS algorithm (v2.1.0) (Zhang et al. 2008) with a cutoff of p-value = $1e^{-7}$, and SICER (Zang et al. 2009) at a cutoff of FDR $1e^{-10}$ and a max gap parameter of 600 bp. Peaks that were on the ENCODE blacklist of known false ChIP-Seq peaks were removed. Standard normalization analysis was also conducted. When comparing all samples, the read number of all samples is reduced by random sampling to the number of tags preset in the smallest sample. This normalization method can detect site-specific as well as global differences in target enrichments between samples.

3.3.4 Peak annotation and visualization

After identification the location of ChIP peaks on the genome, their proximities to gene annotations and other genomic features could be determined and presented. The R package, ChIPseeker (Yu et al. 2015), was used to analyze characteristics of peak distribution across all catfish genomic regions and generated a coverage plot over all chromosomes a pie chart. The

genomic region was classified in the pie chart: 5'UTR and 3'UTR were classified as Exons, while downstream sequences was classified as Intergenic regions.

To compare the ChIP peaks in different samples, data visualization tool, IGB (Integrated Genome Browser; https://bioviz.org/) was used (Freese et al. 2016). The catfish genome (IpCoco_1.2) was administrated as the reference genome. DNA aggregations along the genome of sample ChIP files were expressed by blue heatmap. Peak regions of each sample could be loaded and visualized from chromosome level at the same time, which allowed the direct comparisons of each sample at the same genomic location.

3.3.5 Motif discovery by MEME Suite

The MEME Suite 5.0.0 (http://meme-suite.org/) was used to discover sequence motifs representing features such as DNA binding sites and protein interaction domains in the DNA sequences (Bailey et al. 2009). MEME (Bailey and Elkan 1994) was used in this study to identify the motifs within ChIP peak sequence, which was regarded as the potential promoter sequence. MEME has a large number of optional inputs to fine-tune its performance (Abe and Gemmell 2014). In this study, the optional inputs were set up as following: (1) zero or one occurrence per sequence model (i.e., zoops) was chosen; (2) the minimum width of motifs was 6 while the maximum width of motifs was 10 or 20 for short and long motifs, respectively; (3) both the given strand and the reverse complement strand were considered when searching for motifs in a complementable alphabet; (4) the number of discovered motifs was set to 20. In general, motifs with a longer width tend to have lower levels of E-value in MEME analysis, because the width of the motifs can affect the statistical significance, E-value, of the motifs and then affect the probability reported by MEME. Notably, to avoid losing significant motifs, the number of

63

discovered motifs was set to 20. However, some motifs with the E-value larger than 0.05 were also detected under this setting. Therefore, the "Results" section only displays the significant motifs with E-value smaller than 0.05.

TOMTOM was used to compare the discovered motifs by MEME against the database of known motifs (JASPER version 4) and report for each query a list of target motifs, ranked by p-value (Gupta et al. 2007). In TOMTOM analysis, the optional inputs were set up as following: (1) the verbosity level of the output information messages was 2; (2) the minimum positions of motif overlapped was 5; (3) Pearson's correlation coefficient was descripted; (4) only report matches with significance values no more than 0.05.

3.3.6 Tissue-specific genes identification

Merged peak regions were produced when comparing all four samples by merging the overlapping peaks from different samples into merged peak regions and defined the start coordinate of the most upstream peak and the end coordinate of the most downstream peak. In locations where only one sample has a peak, this peak was defined as the merged region. Two cutoff criteria, "Maxtag" and "log2ratio", were utilized to expose the merged peak regions with significantly different peak value between liver and intestine. Maxtag was defined as maximum of peak values of compared samples multiple length of peaks, then divided by average library fragment length. The calculation of log2ratio was log_2(peak value of liver/ peak value of intestine). Peak regions with Maxtag no less than 100 as well as log2ratio > 3 or log2ratio < -3 were recognized as the tissue-specific regions. The protein-coding genes located within ±1 Kb of these recognized regions were considered as the genes regulated by the tissue-specific promoters.

3.3.7 Gene ontology analysis

The analysis of functional gene annotations was performed using PANTHER (protein annotation through evolutionary relationship) classification system 14.0 (http://www.pantherdb.org) (Mi et al. 2013). Statistical overrepresentation test was conducted using default settings. Annotation data was set as PANTHER GO-Slim Biological Process, Fisher's Exact test was chosen, and False Discovery Rate (FDR) was calculated as the correction. Finally, the enriched biological processes with FDR less than 0.05 were reported.

3.3.8 Data access

The promoter sequences of other species were from the EPD (eukaryotic promoter database; https://epd.epfl.ch//index.php). This database allows the access to annotated non-redundant collection of eukaryotic POL II promoters, for which the TSS has been determined experimentally. Promoters of a total of 10 animal species, two plant species, two fungi and one invertebrate species were collected in this resource. And in this study, we selected promoters of some important model species to analyze including human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis familiaris*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), and fruit fly (*Drosophila melanogaster*). The genome GC-content of all species were obtained from NCBI genome databased (https://www.ncbi.nlm.nih.gov/genome/).

3.4 Results

3.4.1 Genome-wide H3K4me3 enriched regions across the catfish genome

ChIP-seq raw reads number and aligned reads number were shown in Table 3. The total number of reads produced by ChIP-seq for four samples were from about 19 million to more than

25 million. After alignment and a set of filters, there were about 10.5 million, 12 million, 14.5 million reads, what we called "tags" to clear in the Table 3, were kept for subsequent analysis for Liver-rep1, Liver-rep2, and two intestine samples, respectively. Among four samples, there were few differences of the final number of tags between two intestine samples while about 20% differences were observed between two liver samples.

Table 3. Summary of reads and alignment of Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) using liver and intestine from first generation of backcross catfish

Samples	Total number of reads	Total number of alignments	Unique alignments	Final number of tags
Liver-rep1	18,953,704	15,591,172	10,601,731	10,597,248
Liver-rep2	23,102,586	18,643,722	12,214,963	12,209,388
Intestine-rep1	24,767,084	19,546,190	14,533,306	14,529,033
Intestine-rep2	25,548,217	20,370,463	14,602,095	14,597,435

The result of peak annotation was shown in Fig. 8A-B. After annotation, it was found that most peaks (81.57%) were located within promoter regions. The other 6.33%, 8.76%, and 3.34% ChIP peaks were located within exon, intron, and intergenic regions, respectively. Comparatively, when we only focused on core promoters, 62.75% detected ChIP peaks were located within these regions while 18%, 14.03%, and 5.22% peaks were located within exon, intron, and intergenic regions, respectively.

A total of 17,266, 18,351, 17,643, and 17,991 peaks were detected in four samples from the ChIP-seq using H3K4me3 antibody, respectively (Table 4). Fig. 8C and Table 4 displayed the ChIP peaks across 29 chromosomes in catfish genome. In general, the peak number is positively relevant to the length of chromosome. However, the unit peak number (per Mb) on chromosome 28 and 18 were extremely large while on chromosome 4 and 19 were extremely small, which means the distribution of ChIP peaks are not totally even between each chromosome.

A.

B.

C.

Chromosome Size (bp)

70
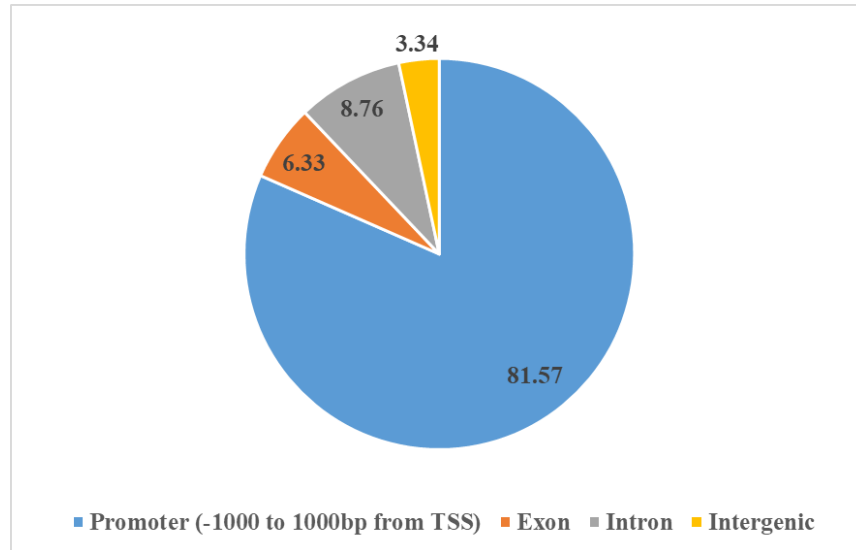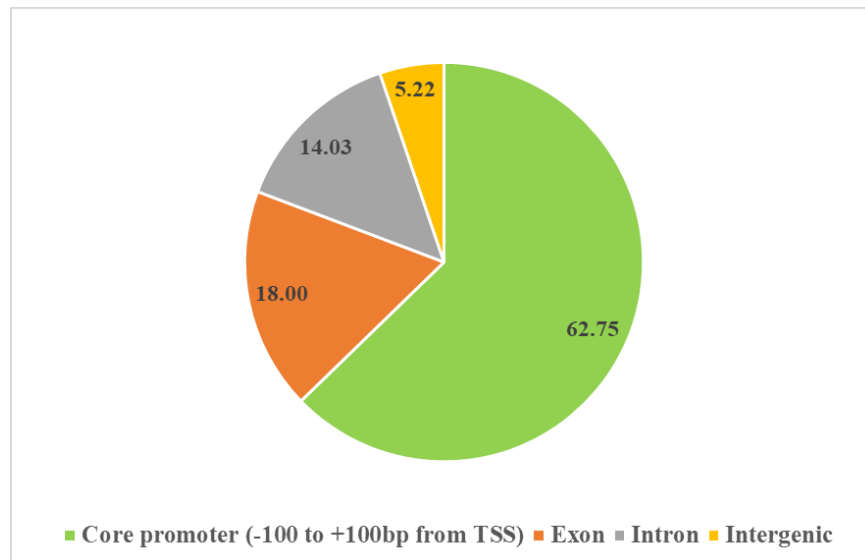
Figure 8. Distribution of Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) peaks in first generation of backcross catfish across catfish genome. (A): catfish genome was divided in four categories based on the location and they were promoters (±1 Kb around TSS), exons, introns, and intergenic regions. 5' UTR and 3' UTR were defined as part of exons. (B): similar to (A), but the core promoter was defined as ±100 bp around TSS. (C): ChIP-seq peaks distribution over 29 chromosomes of catfish.

Table 4. Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) peaks over 29 chromosomes of catfish genome

Chromosome	Length (bp)	Total Peaks	Liver-rep1	Liver-rep2	Intestine-rep1	Intestine-rep2	Peaks number/Mb
1	37,510,255	3,235	793	816	791	835	86
2	37,257,225	3,842	918	969	984	971	103
3	35,853,109	3,284	802	866	804	812	92
4	34,595,840	2,357	576	613	581	587	68
5	33,351,511	3,184	755	806	802	821	95
6	32,139,338	2,532	605	637	627	663	79
7	31,027,439	3,210	785	841	781	803	103
8	30,410,158	2,825	686	724	711	704	93
9	30,212,908	2,783	687	738	673	685	92
10	29,046,025	2,662	658	688	655	661	92
11	28,208,075	2,185	538	556	532	559	77
12	28,145,556	3,035	723	777	759	776	108
13	27,425,808	2,553	614	655	637	647	93
14	26,699,778	2,395	576	610	609	600	90
15	26,085,978	2,787	660	682	720	725	107
16	25,964,114	2,436	575	610	624	627	94
17	25,916,894	2,303	547	593	578	585	89
18	24,512,808	2,704	659	694	669	682	110
19	24,127,165	1,775	433	468	426	448	74
20	22,470,590	2,005	483	538	489	495	89
21	20,987,006	2,005	499	525	485	496	96
22	19,874,160	2,142	534	556	517	535	108
23	19,722,170	1,862	470	485	444	463	94
24	19,702,765	2,148	522	551	531	544	109
25	19,457,651	2,005	497	526	484	498	103
26	19,081,014	2,071	489	534	516	532	109
27	18,544,414	1,467	349	392	362	364	79
28	18,419,168	2,276	552	584	561	579	124
29	15,256,605	1,183	281	317	291	294	78
Total			17,266	18,351	17,643	17,991	

3.4.2 Enriched motifs of catfish promoter

  To obtain the detailed landscape of catfish promoter sequences, motif discovery was conducted in ChIP peak sequences located with promoter and core promoter regions, respectively. Two long motifs were detected in catfish promoter sequences, which were the poly-A and poly-G prevalent motifs (Fig. 9A) and 7 short conserved motifs were identified within the core promoter sequences in catfish (Fig. 9B). The detected motifs were compared to known motifs in JASPAR version 4 database. The archetype of detected conserved motifs, E-value and number of sites of each motif, and the number and known motifs they can match to were all presented (Fig. 9).

A.

Motif archetype	E-value	Sites	Matches with motif database	Matched motif
	7.50E-46	3049	2	ZNF384 ONECUT3
	1.10E-11	2201	1	ZBTB33

B.

Motif archetype	E-value	Sites	Matches with motif database	Matched motif
	2.20E-53	2048	25	ELF5 ETV2 ELF4 ELF1 EHF SPI1 ELK1 ELK3 ERF ETV3 FEV ETV1 ETS1 ELF3 FLI1 ETV4 ERG ETV6 SPIC ETV5 SPDEF ELK4 ZBTB7A Gabpa SPIB
	7.10E-50	1467	1	NRF1
	1.00E-42	1329	1	ZNF384
	6.40E-30	742	4	YY1 E2F2 YY2 E2F3
	2.20E-17	1018	13	SP1 SP2 KLF16 SP3 KLF5 SP4 ZNF263 KLF14 SP8 EGR3 ZNF740 EGR4 KLF13
	1.00E-06	502	20	ATF7 FOSB::JUN FOSB::JUNB FOSL2::JUNB FOS::JUN FOSL1::JUN JDP2 JUN::JUNB FOSL2::JUND BATF3 JUNB Creb5 FOSL2::JUN CREB1 FOSL1::JUND Crem JUND JUN Atf1 Atf3
	6.40E-07	447	3	NFYA NFYB Dux

Figure 9. Enriched motifs detected using Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) peak sequences in first generation of backcross catfish. (A): Conserved motifs detected using ChIP-seq peak sequences within ±1 Kb around TSSs and maximum of length of motifs were set as 20bp. (B): Conserved motifs detected using ChIP-seq peak sequences within ±100 bp around TSSs and maximum of length of motifs were set as 10bp.

75

3.4.3 Promoter comparison

To have a better understanding of catfish promoter sequence, the comparison between catfish and the model species including human, mouse, dog, chicken, zebrafish, and fruit fly was conducted. The discrepancy of GC-content of promoter sequences between these species was detected (Table 5). Based on the GC-content, the analyzed species could be divided into two groups with different color: (1) in promoters of human, mouse, dog, and chicken, the amount of G/C is significantly larger than A/T, even attained to or over than twice in dog and chicken; (2) in promoters of zebrafish, catfish, and fruit fly, the amount of A/T is slightly more than G/C. To be compared, the GC-content of genome of all these species were also displayed and the GC-content of zebrafish and catfish genome were found relatively lower than other species (Table 5). When calculating the genome GC-content, the mtDNA and unmapped sequences were excluded.

Table 5. Comparative analysis of GC-content of both promoters and genome in human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis familiaris*), chicken (*Gallus gallus*), fruit fly (*Caenorhabditis elegans*), zebrafish (*Danio rerio*), and catfish (*Siluriformes*)

Species	GC-content in promoter	GC-content in genome
Human	0.628	0.413
Mouse	0.608	0.419
Dog	0.678	0.411
Chicken	0.706	0.422
Zebrafish	0.49	0.367
Catfish	0.434	0.398
Fruit fly	0.44	0.421

As one of the most conserved and important elements in core promoters, the TATA-box comparison was also conducted between previously mentioned species. The TATA elements were also detected using MEME Suite 5.0.0 (Bailey et al. 2009) and the archetype of TATA elements in each species were shown (Fig. 10). To compare the TATA-box between these species directly and conveniently, we defined the TATA element based on the definition and only focused on the standard TATA sequence, which were marked in the blue table (Fig. 10). Notably, the TATA sequences were similar between zebrafish and catfish, and the number of TATA element sites were significantly less in zebrafish and catfish compared to other species.
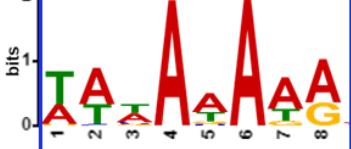
Species	Motif archetype		Sequence Count	Sites	Width
human			29598	773	8
mouse			25111	1510	8
dog			7431	1004	10
chicken			6126	687	10
fruitfly			16972	937	10
zebrafish			10726	180	8
catfish			54720	974	10

Figure 10. TATA-box comparison between human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis familiaris*), chicken (*Gallus gallus*), fruit fly (*Caenorhabditis elegans*), zebrafish (*Danio rerio*), and catfish (*Siluriformes*). Sequence count means the number of total sequences used to identify TATA-box. Sites means the number of sequences contained TATA-box in each species. Blue tables marked the core structure of TATA-box of all species based on the standard.

3.4.4 Tissue-specific active promoter analysis in catfish

After merging and annotation, the merged peak regions and the genes they were regulating were produced. The merge step is necessary because the locations and lengths of peak are rarely the same when comparing different samples. After merged, a total of 12,769 and 13,059 peak regions were identified in catfish liver and intestine, respectively (Fig. 11A). Among them, 11,364 peak regions were detected in both liver and intestine, while 1,405 peak regions were liver-specific and 1,695 peak regions were intestine-specific. Significant difference of peak value was detected in several chromosomes after visualization of the peaks (Fig .11B-E).

A.

B.

Chr1: 7,325,556 - 7,654,657

C.

Chr1: 32,231,340 - 32,560,441

D.

Chr17: 22,153,159 – 22,411,957

E.

Chr18: 22,911,0.82 – 23,160,678

82
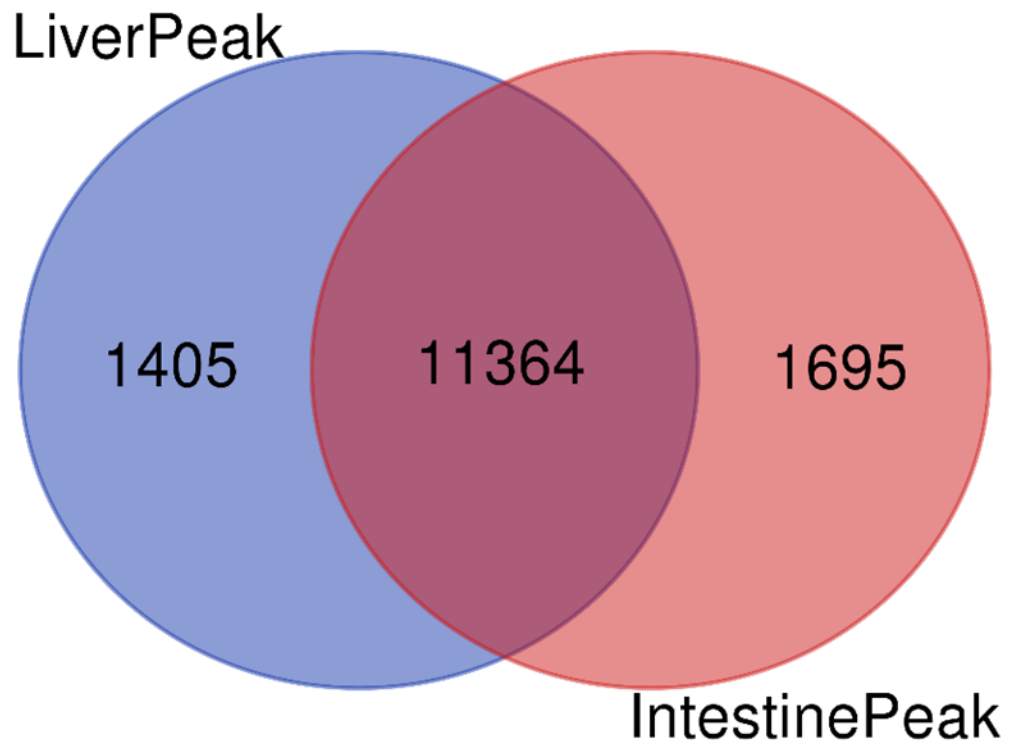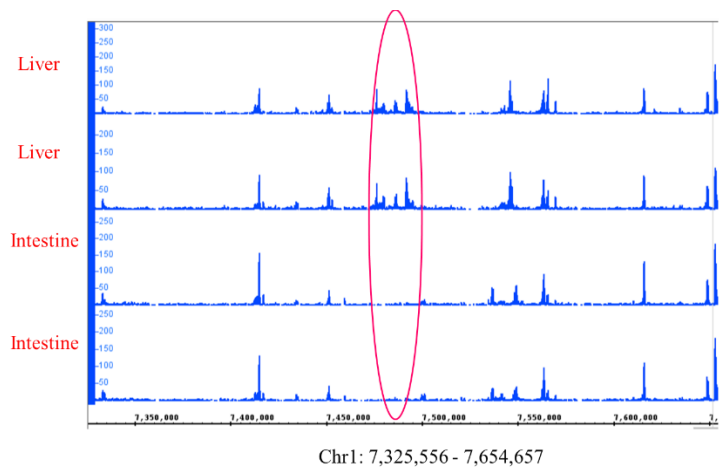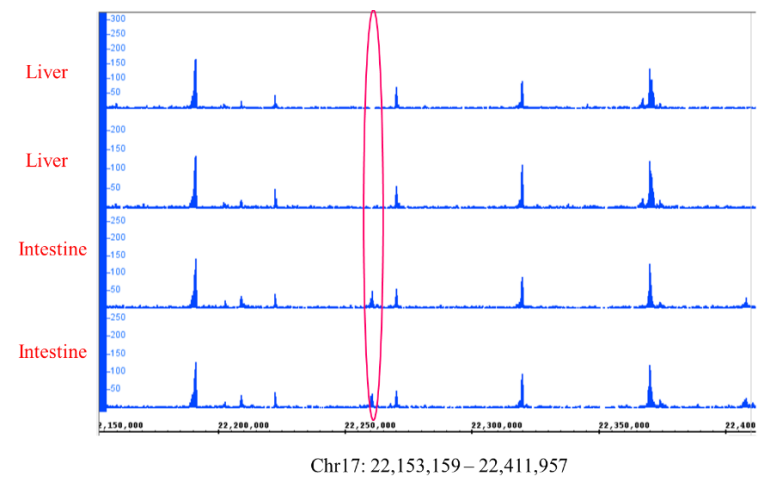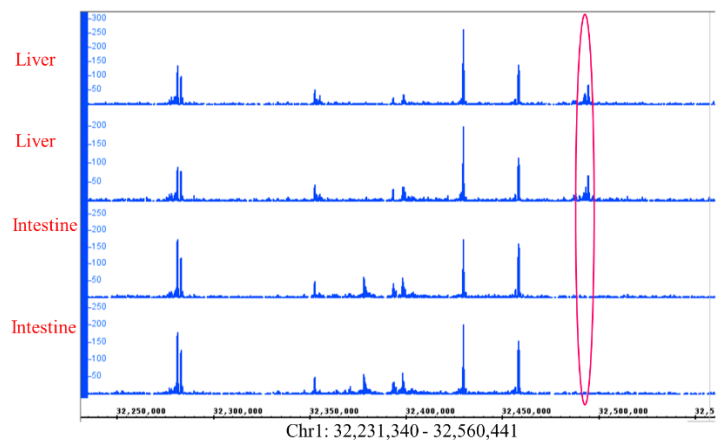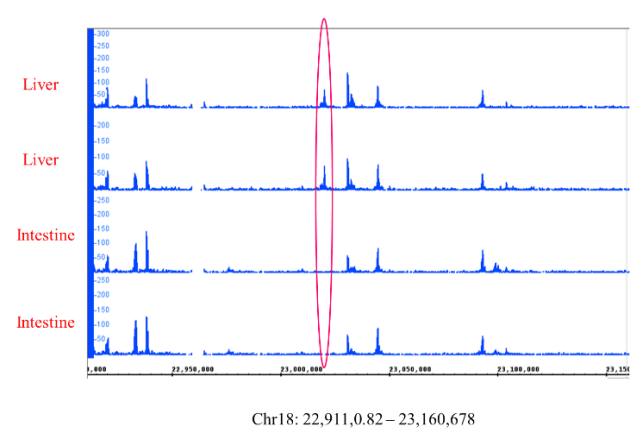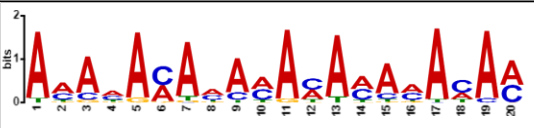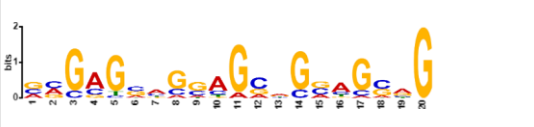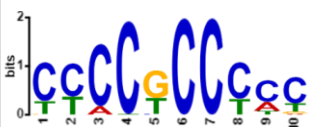
Figure 11. Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) peaks comparison in catfish liver and intestine. (A): the Venn diagram of peaks detected in catfish liver and intestine. The middle-overlapped part means the peaks detected in both liver and intestine. (B-E): visualization of peaks in four samples and the red circles indicated the significantly differences between two tissues.

To identify the landscape of tissue-specific active promoters, motif discovery was operated using liver and intestine samples, respectively. Three conserved motifs were recognized in catfish liver promoters. Except for the poly-A and poly-G conserved motif, a motif with length of 8bp and highly GC replication were detected (Fig. 12A). The conserved motifs detected in catfish intestine (Fig. 12C) were highly similar to those in liver, nevertheless, the length and count of sites were relatively distinct. The conserved motifs within core promoter regions in both liver and intestine were also characterized (Fig. 12B&D). A total of 5 conserved motifs were detected in liver core promoter regions (Fig. 12B) while 8 motifs were detected in intestine core promoters. The motifs detected in liver core promoter regions were all detected in intestine core promoter regions (Fig. 12).

A.
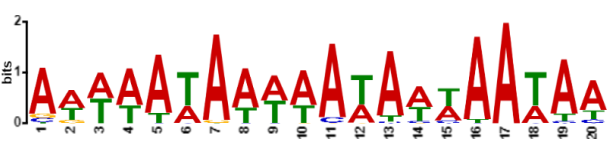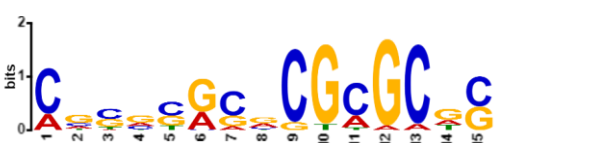
Motif archetype	E-value	Sites	Matches with motif database	Matched motif
	1.30E-77	1194	3	Foxj3 ZNF384 Foxd3
	2.90E-10	342	5	ZNF263 ASCL1 Myod1 Tcf12 Myog
	1.20E-03	572	4	Tchl5 Hes1 NRF1 HES7

B.

Motif archetype	E-value	Sites	Matches with motif database	Matched motif
	5.00E-60	367	1	NRF1
	2.50E-12	232	1	ZNF384
	4.20E-12	297	13	SP1 SP2 KLF16 SP3 SP4 KLF14 KLF5 SP8 EGR3 ZNF263 ZNF740 EGR4 KLF13
	7.00E-03	73	23	JUN::JUNB JDP2 Creb5 FOSB::JUNB FOSL1::JUN JUNB FOSL2::JUN CREB1 FOSB::JUN FOSL2::JUNB FOS::JUN FOSL1::JUND ATF7 FOSL2::JUND BATF3 Crem JUND JUN SREBF2 TFEB Srebf1 Atf3 Atf1
	7.60E-07	108	3	YY1 E2F2 E2F3

C.

Motif archetype	E-value	Sites	Matches with motif database	Matched motif
	5.70E-33	601	2	ZNF263 IRF1
	4.10E-19	317	2	ZNF384 Arid3b
	3.80E-08	643	2	NRF1 SP2

D.

Motif archetype	E-value	Sites	Matches with motif database	Matched motif
	9.30E-82	967	1	NRF1
	5.20E-64	1423	24	ELK1 ERF ETV3 FEV ETS1 ETV6 FLI1 ELK3 ERG ETV1 ETV5 ETV4 ELK4 ELF4 ELF1 ELF5 Gabpa ZBTB7A EHF ELF3 ETV2 SPDEF SPI1 SPIC
	1.50E-50	545	22	CREB1 JUN::JUNB FOSB::JUNB FOSL1::JUN JDP2 Creb5 JUNB FOSL1::JUND FOSL2::JUN FOSB::JUN FOSL2::JUNB FOSL2::JUND FOS::JUN ATF7 BATF3 Crem JUND Atf1 Atf3 JUN TFEB BHLHE41
	1.40E-47	528	1	ZNF384
	3.90E-32	609	4	YY1 E2F2 YY2 E2F3
	3.00E-32	643	3	NFYA NFYB Dux
	2.40E-28	706	14	SP1 KLF5 SP2 KLF16 SP4 SP3 SP8 KLF14 KLF13 Klf12 EGR3 Klf1 ZNF740 KLF9
	2.20E-09	281	23	USF Arntl BHLHE41 TFE3 TFEB USF1 MITF SREBF2 BHLHE40 Srebf1 TFEC MLX Arnt MLXIPL Id2 Creb3l2 HES HEY1 HES5 MYCN HEY2 MNT MYC

Figure 12. Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) Enriched Motifs detected in catfish liver and intestine. (A): Enriched motifs detected using ChIP-seq peak sequences within ±1 Kb around TSSs in liver and maximum of length of motifs were set as 20bp. (B): Enriched motifs detected using ChIP-seq peak sequences within ±100bp around TSSs in liver and maximum of length of motifs were set as 10bp. (C): Enriched motifs detected using ChIP-seq peak sequences within ±1 Kb around TSSs in intestine and maximum of length of motifs were set as 20bp. (D): Enriched motifs detected using ChIP-seq peak sequences within ±100bp around TSSs in intestine and maximum of length of motifs were set as 10bp.

As mentioned in Method part, the tissue-specific merged peak regions and regulated genes were generated. Via the cutoff criteria, a total of 331 and 363 specific protein coding genes were detected in liver and intestine, respectively (Table S6&7). Notably, the log2ratio of some regions could not be calculated since the peak value of one of the compared samples equals to zero, which means there was totally no peak detected in the region of this samples (Table 6-7, Fig. 11C-E). The top 15 tissue-specific genes, which owned the 15 highest log2ratio absolute value, and function of the genes were displayed (Table 6-7).

The biological process enriched by these genes with tissue-specific active promoters were identified by gene ontology analysis. Although the number of intestine-specific genes were more than liver, the liver-specific genes were involved in 18 biological process on statistically significant level with FDR (false discovery rate) less than 0.05, while intestine-specific genes were only involved in 10 process. Among these biological processes, three were detected in both liver and intestine, which were signal transduction, regulation of cellular process, cell surface receptor signaling pathway. Two liver-specific genes enriched biological process with highest fold enrichment were positive regulation of blood pressure and smooth muscle contraction, both of which were liver-specific process. Similarly, positive regulation of actin filament polymerization, which acquired the largest fold enrichment from intestine-specific genes, was also only identified in intestine.

88

Table 6. Top 15 liver-specific actively transcribed genes with highest log2ratio of Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) peak value between catfish liver and intestine

Peak value		Chromosome	log2ratio(L/I)	Gene	Function
Liver	Intestine				
30.567	0.000	1		agtr1	receptor of Angiotensin II
38.552	0.000	18		serpinf2	regulating the blood clotting
135.891	2.337	26	5.862	fga	the coagulation factor fibrinogen
77.400	1.515	25	5.675	cenpj	maintenance of centrosome function
77.400	1.515	25	5.675	plg	secreted blood zymogen
48.739	1.000	20	5.607	proc	vitamin K-dependent plasma glycoprotein
81.419	1.780	26	5.515	fgb	component of fibrinogen, a blood-borne glycoprotein
64.920	1.421	10	5.514	abcb11	ATP-binding cassette transporters
52.439	1.214	7	5.433	cfhr1	Involved in complement regulation
77.566	1.821	29	5.413	fgg	fibrinogen, a blood-borne glycoprotein
52.557	1.237	7	5.409	c8a	component of the complement system
100.903	2.380	9	5.406	ppp1r3c	Activates glycogen synthase, and limits glycogen breakdown.
49.319	1.174	8	5.393	etnk2	catalyzes phosphatidylethanolamine (PtdEtn) biosynthesis
41.469	1.000	11	5.374	uroc1	histidine catabolism
62.092	1.514	10	5.358	tdo2	tryptophan metabolism

Table 7. Top 15 intestine-specific actively transcribed genes with lowest log2ratio of Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) peak value between catfish liver and intestine

Peak value		Chromosome	log2ratio(L/I)	Gene	Function
Liver	Intestine				
0.000	19.725	17		ca4	catalyze the reversible hydration of carbon dioxide
1.000	46.808	13	-5.549	myo15b	
1.167	52.159	23	-5.482	il20ra	receptor for interleukin 20, a cytokine involved in epidermal function
1.000	41.679	13	-5.381	arhgap27	clathrin-mediated endocytosis
1.000	41.324	14	-5.369	ush1c	assembly of Usher protein complexes
1.355	55.620	8	-5.359	rnf128	possess E3 ubiquitin ligase activity
1.000	37.333	2	-5.222	prkg1	components of many signal transduction processes
2.115	78.393	18	-5.212	cdx1	intestine-specific gene expression and enterocyte differentiation
2.333	85.423	7	-5.194	hrh3	regulate neurotransmitter release
1.000	35.450	15	-5.148	mapk12	transduction of extracellular signals
1.000	35.286	6	-5.141	pdzk1	role in cholesterol metabolism
1.588	54.550	15	-5.102	pparg	regulator of adipocyte differentiation
1.217	40.640	5	-5.061	tmem252	
1.667	55.472	5	-5.056	lpxn	cell type-specific signaling
1.324	42.143	15	-4.992	trim35	related to Cytokine Signaling in Immune system

3.5 Discussion

This study proposed a protocol for conducting histone modification ChIP-seq on a nonmodel teleost, which initiated the identification of regulated roles of epigenetic variations in catfish. This study identified H3K4me3 enriched regions across the whole catfish genome, detected the enriched motifs of ChIP peak sequences located within both catfish promoter (±1 Kb around TSS) and core promote regions (±100bp around TSS), and characterized the tissue-specific active promoters in catfish liver and intestine, which guided the detection of tissue-specific active transcripted genes and key functions the tissues played. Altogether, this analysis provided a powerful tool and an effective workflow for genome-wide epigenomic and epigenetic regulator identification and hinted the probability of use of this protocol to identify the function of epigenetic regulators in more important and ecological traits in catfish including growth, stress tolerance, and disease resistance in the future.

This study verified the proposition that the enrichment of H3K4me3 mainly around TSSs (Barski et al. 2007; Bernstein et al. 2005) by peak annotation and visualization (Fig. 8A-B) based on the well assembled and annotated catfish genome (Liu et al. 2016b), which laid a solid foundation for successive promoter analysis. Although the detected ChIP peaks showed a positive relationship with length of chromosome and a relatively even distribution on all chromosomes, the significant fewer peaks number per Megabytes on chromosome 4 were still recognized, what may be related to the main roles of the chromosomes. For instance, the chromosome 4 was reported as the sex determination regions in catfish (Bao et al. 2019), however, the non-sex-specific experimental fish lead to the obviously fewer unit peak numbers detected on chromosome 4. This

distinctive distribution across chromosomes may clue the of great importance of biological and developmental roles of each chromosome in catfish genome.

Particularly, motifs are critical for the recognition by specific proteins affecting chromosome organization, transcription and DNA replication (Calistri et al. 2011). A total of 9 enriched motifs were identified of ChIP peak sequences located within catfish promoter and core promoter regions. The most enriched motif with a core consensus GGAA/T, was reported as the binding sites of EWS and ETS family members, which were a key oncogenic event in Ewing tumors (Guillon et al. 2009). The motif with highly GC replications were matched with the promoter nrf1, which was a transcription factor that played a destine functional role in regulating antioxidant response element-driven genes (Ohtsuji et al. 2008). This motif also indicated the existence of CpG island in catfish promoters, since the CpG islands were suggested to influence local chromatin structure and simplify regulation of gene activity (Deaton and Bird 2011). The poly-A and poly-G sequences were identified in both promoter and core promoter regions, indicating their extensive amount in catfish promoters. Previous studies also illustrated the polypurine motifs were the most and enriched in promoter regions in human genome (Anderson and Widom 2001) and chicken (Abe and Gemmell 2014). It was suggested the poly-A might tasked to alter the stability or dynamics of nucleosomes, then improved the ability of gene activator proteins to bind nearby DNA target sites in a manner (Shimizu et al. 2000). The motif with core sequence CCAT were aligned transcription factor YY. YY1 and YY2 were reported as a DNA-binding transcription factor, which could function as a repressor, activator, or transcription initiator (Shi et al. 1997) and the duplication of YY1 were uncovered in several different lineages, including flies, fish, and mammals (Kim et al. 2007), which was well supported by our observation. The

motifs detected in catfish liver and intestine were of great similarity. However, an intestine-specific motif with core consensus CACGTG was detected within core promoter regions. This highly conserved motif could affect the specificity of protein binding (Williams et al. 1992), which was found to develop the process of cellular proliferation in mouse by interacting with potential trans-activator of ornithine decarboxylase (Tobias et al. 1995).

When comparing the catfish promoters with other model species, a relatively low GC-content was detected in zebrafish, catfish, and fruit fly, comparing to chicken and three mammals (Table 5). This was consistent with the previous conclusion from the study of evolutionary trends GC/AT in promoters that the GC-rich was present in promoters of eukaryotes and GC-rich gradients increased in length from lower to higher organisms (Calistri et al. 2011). In the whole genome, GC-content was also found to be variable with different organisms. It was concluded that the GC-richer regions had a high gene density, called "genome core", while the GC-poorer regions, called "genome desert", contained a low number of genes and differed from the former in several area including chromatin structure, UTRs length, and intensity of gene expression (Mouchiroud et al. 1991). This may be the powerful interpretation of the GC-content is significantly higher in promoter regions than whole genome (Table 5). The GC-rich promoters also indicated the prevalence of CpG island, which was a clear hallmark of mammalian promoters and presented at more than two-thirds of genes (Fenouil et al. 2012). Although the relatively low GC-content in catfish promoter regions, a CG-rich motif was still detected (Fig. 12A-D), suggesting the existence of CpG island within the promoter regions of catfish. Like TATA-box, CpG islands also played a key role in transcription initiation (Carninci et al. 2006). However, genes containing different core promoter elements tended to control different biological processes: the promoters of housekeeping

genes in vertebrate organisms were often associated with CpG islands; while cell type-specific or highly regulated genes often contained TATA-boxes (Gardiner-Garden and Frommer 1987; Larsen et al. 1992; Smale and Kadonaga 2003; Yamashita et al. 2005). This was also a reasonable explanation of fewer conserved TATA motif were detected in catfish promoter regions.

Tissue-specific active promoter analysis recovered that promoters of agtr1 and serpinf2 were active in liver and entirely inactive in intestine while promoter of ca4 was only active in intestine and inactive in liver. The decreased expression of agtr1 mRNA and protein in advanced liver fibrosis and cirrhosis indicated its crucial role in prevention of cirrhosis (Töx et al. 2007). The encoded protein located primarily in the liver, kidney, adrenal gland, and lung, which was consistent with its liver-specific characteristic uncovered in this study. Gene serpinf2 encoded a major inhibitor of plasmin and was reported to be expressed specifically by the liver (Zhou et al. 2014). Similar to agtr1, decreased production of serpinf2 was also diagnosed in cirrhosis of the liver, leading to an increased risk of bleeding in liver disease (Sattar 2016). The differential expression of this gene in liver of two rainbow trout strains demonstrated serpinf2 may contribute to maintain the health status for the strain showing enhanced resistance to infections in aquaculture species (Rebl et al. 2012). Northern blot and RT-PCR analyses revealed the abundant expression of gene ca4 in distal small and large intestine in rate and human, and immunohistochemistry localized ca4 to the apical plasma membrane of the mucosal epithelium in intestine, mainly in colon (Fleming et al. 1995). Similar situation was detected in guinea-pig (Carter and Parsons 1971). ChIP peaks detected in both catfish liver and intestine, nevertheless, the peak value attained to 8-fold difference between two tissues, were also considered as the tissue-specific active promoters.

Based on the gene ontology analysis of these genes regulated by tissue-specific active promoters, it was clear concluded that liver was involved in more biological process than intestine in catfish. Apparently, catfish liver was heavily involved in regulation of blood pressure specifically. Insulin-like growth factor (IGF), derived by liver, has been proved the role of regulation of blood pressure in mice (Tivesten et al. 2002) while several IGF related genes were detected as the liver-specific genes including igfals, igf1, igf2, igfbp1, igfbp2. Targeted disruption of soluble epoxide hydrolase revealed its role in blood pressure regulation (Sinal et al. 2000), meanwhile, it was uncovered that the mammalian soluble epoxide hydrolase was identical to liver hepoxilin hydrolase (Cronin et al. 2011), which also emphasized the responsibility of liver in blood pressure regulation. Except, smooth muscle contraction was also disclosed in the enrichment analysis of genes regulated by liver-specific active promoters. Gene ontology analysis of genes regulated by intestine-specific active promoters revealed the intestine-specific function of positive regulation of actin filament polymerization. Previous studies supported this conclusion by identifying specific location of actin in the brush border of epithelial cells in chicken intestine (Tilney and Mooseker 1971), the encounter of nucleated polymerization of actin from microvillar filaments in the intestinal brush border (Mooseker et al. 1982) and that a component of intestinal brush borders could regulate the actin polymerization (Craig and Powell 1980).

This is the first-time conducting histone modification ChIP-seq to reveal the role of the epigenetic regulator H3K4me3 in catfish, which uncovered the landscape of catfish promoters and recognized the tissue-specific active promoters and specific biological process in catfish liver and intestine, respectively. Not only conserved motifs in both long (1000bp around TSSs) and core promoters (100bp around TSSs) were identified, the exclusively tissue-specific genes, agtr1 and

95

serpinf2 in liver and ca4 in intestine, were also detected. Based on the gene ontology analysis, the specific role of liver in regulation of blood pressure and of intestine in regulation of actin filament polymerization were verified. This study provided a protocol for study of epigenetic histone regulation in catfish as well as other non-model teleost species. Furthermore, mapping genome-wide occupancy of H3K4me3 in catfish could encourage further investigations (Kratochwil and Meyer 2015b) including finding novel nonannotated transcripts in the genome (Cheung et al. 2010), examining evolutionary novelties for regulatory evolution beyond promoter sequence evolution (Main et al. 2013; Nepal et al. 2013), and detecting the functional non-coding elements (Zhou et al. 2011). This study also authority the further analysis of H3K4me3 regulator in more importantly economical trait in catfish such as growth, stress tolerance, and disease resistance, meanwhile the protocol and workflow proposed could be used in studies of regulation role of other types of histone modifications in catfish. Although the importance of epigenetics regulation in gene transcription and expression has been realized for many years, few analyses have been conducted in nonmodel teleost and this is only the first step in epigenomic and epigenetic studies in catfish.

References

Abe, H., and Gemmell, N.J. 2014. "Abundance, arrangement, and function of sequence motifs in the chicken promoters." *BMC genomics* 15 (1):900.

Abecasis, G.R., Cardon, L.R., and Cookson, W. 2000. "A general test of association for quantitative traits in nuclear families." *The American Journal of Human Genetics* 66 (1):279-292.

Aday, A.W., Zhu, L.J., Lakshmanan, A., Wang, J., and Lawson, N.D. 2011. "Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites." *Developmental biology* 357 (2):450-462.

Anderson, J., and Widom, J. 2001. "Poly (dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites." *Molecular and cellular biology* 21 (11):3830-3839.

Antequera, F. 2003. "Structure, function and evolution of CpG island promoters." *Cellular and Molecular Life Sciences CMLS* 60 (8):1647-1658.

Baerwald, M., Petersen, J., Hedrick, R., Schisler, G., and May, B. 2011. "A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (Oncorhynchus mykiss)." *Heredity* 106 (6):920.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. 2009. "MEME SUITE: tools for motif discovery and searching." *Nucleic acids research* 37 (suppl_2):W202-W208.

Bailey, T.L., and Elkan, C. 1994. "Fitting a mixture model by expectation maximization to discover motifs in bipolymers."

Bao, L., Tian, C., Liu, S., Zhang, Y., Elaswad, A., Yuan, Z., Khalil, K., Sun, F., Yang, Y., and Zhou, T. 2019. "The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish." *BMC biology* 17 (1):6.

Barrera, L.O., Li, Z., Smith, A.D., Arden, K.C., Cavenee, W.K., Zhang, M.Q., Green, R.D., and Ren, B. 2008. "Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs." *Genome research* 18 (1):46-59.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. 2007. "High-resolution profiling of histone methylations in the human genome." *Cell* 129 (4):823-37. doi: 10.1016/j.cell.2007.05.009.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. 2004. "Identification and distinct regulation of yeast TATA box-containing genes." *Cell* 116 (5):699-709.

Bauer, U.M., Daujat, S., Nielsen, S.J., Nightingale, K., and Kouzarides, T. 2002. "Methylation at arginine 17 of histone H3 is linked to gene activation." *EMBO reports* 3 (1):39-44.

Begum, F., Ghosh, D., Tseng, G.C., and Feingold, E. 2012. "Comprehensive literature review and statistical considerations for GWAS meta-analysis." *Nucleic acids research* 40 (9):3777-3784.

Benayoun, B.A., Pollina, E.A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E.D., Devarajan, K., Daugherty, A.C., Kundaje, A.B., and Mancini, E. 2014. "H3K4me3 breadth is linked to cell identity and transcriptional consistency." *Cell* 158 (3):673-688.

Berger, S.L. 2007. "The complex language of chromatin regulation during transcription." *Nature* 447 (7143):407.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. 2002. "Methylation of histone H3 Lys 4 in coding regions of active genes." *Proceedings of the National Academy of Sciences* 99 (13):8695-8700.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., and Gingeras, T.R. 2005. "Genomic maps and comparative analysis of histone modifications in human and mouse." *Cell* 120 (2):169-181.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., and Plath, K. 2006. "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." *Cell* 125 (2):315-326.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., and Chen, F. 2010. "ChIP-Seq identification of weakly conserved heart enhancers." *Nature genetics* 42 (9):806.

Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de la Calle-Mustienes, E., and Gomez-Skarmeta, J.L. 2013. "The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos." *Methods* 62 (3):207-15. doi: 10.1016/j.ymeth.2013.04.011.

Boggs, B.A., Cheung, P., Heard, E., Spector, D.L., Chinault, A.C., and Allis, C.D. 2001. "Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes." *Nature genetics* 30 (1):73.

Boltana, S., Aguilar, A., Sanhueza, N., Donoso, A., Mercado, L., Imarai, M., and Mackenzie, S. 2018. "Behavioral Fever Drives Epigenetic Modulation of the Immune Response in Fish." *Frontiers in immunology* 9.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E. 2012. "Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development." *Nature genetics* 44 (2):148.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. 2001. "Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae." *Genes & development* 15 (24):3286-3295.

Bryk, M., Briggs, S.D., Strahl, B.D., Curcio, M.J., Allis, C.D., and Winston, F. 2002. "Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S. cerevisiae by a Sir2-independent mechanism." *Current Biology* 12 (2):165-170.

Bulger, M., and Groudine, M. 2011. "Functional and mechanistic diversity of distal transcription enhancers." *Cell* 144 (3):327-339.

Calistri, E., Livi, R., and Buiatti, M. 2011. "Evolutionary trends of GC/AT distribution patterns in promoters." *Molecular phylogenetics and evolution* 60 (2):228-235.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., and Frith, M.C. 2006. "Genome-wide analysis of mammalian promoter architecture and evolution." *Nature genetics* 38 (6):626.

Carter, M., and Parsons, D. 1971. "The isoenzymes of carbonic anhydrase: tissue, subcellular distribution and functional significance, with particular reference to the intestinal tract." *The Journal of physiology* 215 (1):71-94.

Castaño-Rodríguez, N., Kaakoush, N.O., Goh, K.-L., Fock, K.M., and Mitchell, H.M. 2014. "The NOD-like receptor signalling pathway in Helicobacter pylori infection and related gastric cancer: a case-control study and gene expression analyses." *PloS one* 9 (6):e98899.

Chang, P.-Y., Hom, R.A., Musselman, C.A., Zhu, L., Kuo, A., Gozani, O., Kutateladze, T.G., and Cleary, M.L. 2010. "Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription." *Journal of molecular biology* 400 (2):137-144.

Chen, D., Ma, H., Hong, H., Koh, S.S., Huang, S.-M., Schurter, B.T., Aswad, D.W., and Stallcup, M.R. 1999. "Regulation of transcription by a protein methyltransferase." *Science* 284 (5423):2174-2177.

Cheung, I., Shulha, H.P., Jiang, Y., Matevossian, A., Wang, J., Weng, Z., and Akbarian, S. 2010. "Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex." *Proceedings of the National Academy of Sciences* 107 (19):8824-8829.

Chung, D., Kuan, P.F., Li, B., Sanalkumar, R., Liang, K., Bresnick, E.H., Dewey, C., and Keleş, S. 2011. "Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data." *PLoS computational biology* 7 (7):e1002111.

Core, L.J., Waterfall, J.J., and Lis, J.T. 2008. "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." *Science* 322 (5909):1845-1848.

Correa, K., Lhorente, J.P., Bassini, L., López, M.E., Di Genova, A., Maass, A., Davidson, W.S., and Yáñez, J.M. 2017. "Genome wide association study for resistance to Caligus rogercresseyi in Atlantic salmon (Salmo salar L.) using a 50K SNP genotyping array." *Aquaculture* 472:61-65.

Craig, S.W., and Powell, L.D. 1980. "Regulation of actin polymerization by villin, a 95,000 dalton cytoskeletal component of intestinal brush borders." *Cell* 22 (3):739-746.

Cronin, A., Decker, M., and Arand, M. 2011. "Mammalian soluble epoxide hydrolase is identical to liver hepoxilin hydrolase." *Journal of lipid research* 52 (4):712-719.

Dahle, M.K., Wessel, Ø., Timmerhaus, G., Nyman, I.B., Jørgensen, S.M., Rimstad, E., and Krasnov, A. 2015. "Transcriptome analyses of Atlantic salmon (Salmo salar L.) erythrocytes infected with piscine orthoreovirus (PRV)." *Fish & shellfish immunology* 45 (2):780-790.

Darvasi, A., and Soller, M. 1992. "Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus." *Theoretical and applied Genetics* 85 (2-3):353-359.

Deaton, A.M., and Bird, A. 2011. "CpG islands and the regulation of transcription." *Genes & development* 25 (10):1010-1022.

Duggal, P., Gillanders, E.M., Holmes, T.N., and Bailey-Wilson, J.E. 2008. "Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies." *BMC genomics* 9 (1):516.

Dunham, R.A. 2011. *Aquaculture and fisheries biotechnology: genetic approaches*: Cabi.

Dunham, R.A., and Masser, M.P. 2012. *Production of hybrid catfish*. Vol. 436: Southern Regional Aquaculture Center Stoneville, Mississippi, USA.

Dunham, R.A., and Smitherman, R.O. 1984. "Ancestry and breeding of catfish in the United States." *Ancestry and breeding of catfish in the United States.* (273).

Dunham, R.A., Umali, G.M., Beam, R., Kristanto, A.H., and Trask, M. 2008. "Comparison of production traits of NWAC103 channel catfish, NWAC103 channel catfish× blue catfish hybrids, Kansas Select 21 channel catfish, and blue catfish grown at commercial densities and exposed to natural bacterial epizootics." *North American Journal of Aquaculture* 70 (1):98-106.

Dupuy, D., Bertin, N., Hidalgo, C.A., Venkatesan, K., Tu, D., Lee, D., Rosenberg, J., Svrzikapa, N., Blanc, A., and Carnec, A. 2007. "Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans." *Nature biotechnology* 25 (6):663.

Eschmeyer, W., and Fong, J. 2014. Species by family/subfamily in the Catalog of Fishes. 2016.

Fahrner, J.A., Eguchi, S., Herman, J.G., and Baylin, S.B. 2002. "Dependence of histone modifications and gene expression on DNA hypermethylation in cancer." *Cancer research* 62 (24):7213-7218.

Feng, W., Yonezawa, M., Ye, J., Jenuwein, T., and Grummt, I. 2010. "PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation." *Nature structural & molecular biology* 17 (4):445.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., and Gut, I. 2012. "CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters." *Genome research* 22 (12):2399-2408.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. 2003. "Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains." *Genes & development* 17 (15):1870-1881.

FitzGerald, P.C., Shlyakhtenko, A., Mir, A.A., and Vinson, C. 2004. "Clustering of DNA sequences in human promoters." *Genome research* 14 (8):1562-1574.

Fleming, R.E., Parkkila, S., Parkkila, A.-K., Rajaniemi, H., Waheed, A., and Sly, W.S. 1995. "Carbonic anhydrase IV expression in rat and human gastrointestinal tract regional, cellular, and subcellular localization." *The Journal of clinical investigation* 96 (6):2907-2913.

Freese, N.H., Norris, D.C., and Loraine, A.E. 2016. "Integrated genome browser: visual analytics platform for genomics." *Bioinformatics* 32 (14):2089-2095.

Fu, Q., Zeng, Q., Li, Y., Yang, Y., Li, C., Liu, S., Zhou, T., Li, N., Yao, J., and Jiang, C. 2017. "The chemokinome superfamily in channel catfish: I. CXC subfamily and their involvement in disease defense and hypoxia responses." *Fish & shellfish immunology* 60:380-390.

Fuji, K., Hasegawa, O., Honda, K., Kumasaka, K., Sakamoto, T., and Okamoto, N. 2007. "Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (Paralichthys olivaceus)." *Aquaculture* 272 (1-4):291-295.

Fuji, K., Kobayashi, K., Hasegawa, O., Coimbra, M.R.M., Sakamoto, T., and Okamoto, N. 2006. "Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (Paralichthys olivaceus)." *Aquaculture* 254 (1-4):203-210.

Fulker, D., Cherny, S., Sham, P., and Hewitt, J. 1999. "Combined linkage and association sib-pair analysis for quantitative traits." *The American Journal of Human Genetics* 64 (1):259-267.

Füllgrabe, J., Kavanagh, E., and Joseph, B. 2011. "Histone onco-modifications." *Oncogene* 30 (31):3391.

Ganis, J.J., Hsia, N., Trompouki, E., de Jong, J.L., DiBiase, A., Lambert, J.S., Jia, Z., Sabo, P.J., Weaver, M., and Sandstrom, R. 2012. "Zebrafish globin switching occurs in two developmental stages and is controlled by the LCR." *Developmental biology* 366 (2):185-194.

Gardiner-Garden, M., and Frommer, M. 1987. "CpG islands in vertebrate genomes." *Journal of molecular biology* 196 (2):261-282.

Geiman, T.M., and Robertson, K.D. 2002. "Chromatin remodeling, histone modifications, and DNA methylation—how does it all fit together?" *Journal of cellular biochemistry* 87 (2):117-125.

Geng, X., Liu, S., Yao, J., Bao, L., Zhang, J., Li, C., Wang, R., Sha, J., Zeng, P., and Zhi, D. 2016. "A genome-wide association study identifies multiple regions associated with head size in catfish." *G3: genes, genomes, genetics* 6 (10):3389-3398.

Geng, X., Liu, S., Yuan, Z., Jiang, Y., Zhi, D., and Liu, Z. 2017. "A genome-wide association study reveals that genes with functions for bone development are associated with body conformation in catfish." *Marine biotechnology* 19 (6):570-578.

Geng, X., Sha, J., Liu, S., Bao, L., Zhang, J., Wang, R., Yao, J., Li, C., Feng, J., and Sun, F. 2015. "A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance." *BMC genomics* 16 (1):196.

Gill, F.B., and Wright, M.T. 2006. *Birds of the world: recommended English names*: Princeton University Press Princeton, New Jersey.

Gjoneska, E., Pfenning, A.R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.-H., and Kellis, M. 2015. "Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease." *Nature* 518 (7539):365.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. 2007. "A chromatin landmark and transcription initiation at most promoters in human cells." *Cell* 130 (1):77-88.

Guillon, N., Tirode, F., Boeva, V., Zynovyev, A., Barillot, E., and Delattre, O. 2009. "The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function." *PloS one* 4 (3):e4932.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. 2007. "Quantifying similarity between motifs." *Genome biology* 8 (2):R24.

Gutierrez, A.P., Bean, T.P., Hooper, C., Stenton, C.A., Sanders, M.B., Paley, R.K., Rastas, P., Bryrom, M., Matika, O., and Houston, R.D. 2018. "A genome-wide association study for host resistance to Ostreid Herpesvirus in Pacific oysters (Crassostrea gigas)." *G3: Genes, Genomes, Genetics* 8 (4):1273-1280.

Gutierrez, A.P., Yáñez, J.M., Fukui, S., Swift, B., and Davidson, W.S. 2015. "Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (Salmo salar)." *PLoS One* 10 (3):e0119730.

Hawley, D.K., and McClure, W.R. 1983. "Compilation and analysis of Escherichia coli promoter DNA sequences." *Nucleic acids research* 11 (8):2237-2255.

Herman, J.G., and Baylin, S.B. 2003. "Gene silencing in cancer in association with promoter hypermethylation." *New England Journal of Medicine* 349 (21):2042-2054.

Hoque, M.O., Kim, M.S., Ostrow, K.L., Liu, J., Wisman, G.B.A., Park, H.L., Poeta, M.L., Jeronimo, C., Henrique, R., and Lendvai, Á. 2008. "Genome-wide promoter analysis uncovers portions of the cancer methylome." *Cancer research* 68 (8):2661-2670.

Houston, R.D., Davey, J.W., Bishop, S.C., Lowe, N.R., Mota-Velasco, J.C., Hamilton, A., Guy, D.R., Tinch, A.E., Thomson, M.L., and Blaxter, M.L. 2012. "Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon." *BMC genomics* 13 (1):244.

Iizuka, M., and Smith, M.M. 2003. "Functional consequences of histone modifications." *Current opinion in genetics & development* 13 (2):154-160.

Jaenisch, R., and Bird, A. 2003. "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." *Nature genetics* 33:245.

Jin, V.X., Singer, G.A., Agosto-Pérez, F.J., Liyanarachchi, S., and Davuluri, R.V. 2006. "Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs." *BMC bioinformatics* 7 (1):114.

Jin, Y., Zhou, T., Geng, X., Liu, S., Chen, A., Yao, J., Jiang, C., Tan, S., Su, B., and Liu, Z. 2017. "A genome‐wide association study of heat stress‐associated SNP s in catfish." *Animal genetics* 48 (2):233-236.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. 2007. "Genome-wide mapping of in vivo protein-DNA interactions." *Science* 316 (5830):1497-1502.

Jones, P.A., and Takai, D. 2001. "The role of DNA methylation in mammalian epigenetics." *Science* 293 (5532):1068-1070.

Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. 2008. "Genome-wide identification of in vivo protein–DNA binding sites from ChIP-Seq data." *Nucleic acids research* 36 (16):5221.

Kadota, M., Hara, Y., Tanaka, K., Takagi, W., Tanegashima, C., Nishimura, O., and Kuraku, S. 2017. "CTCF binding landscape in jawless fish with reference to Hox cluster evolution." *Scientific reports* 7 (1):4957.

Kim, J.D., Faulk, C., and Kim, J. 2007. "Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1." *Nucleic acids research* 35 (10):3442-3452.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., and Kuersten, S. 2010. "Widespread transcription at neuronal activity-regulated enhancers." *Nature* 465 (7295):182.

Klesius, P.H., and Shoemaker, C.A. 1999. "Development and use of modified live Edwardsiella ictaluri vaccine against enteric septicemia of catfish." *Advances in veterinary medicine* 41:523-537.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., De La Chapelle, A.L., Heidemann, M., and Hintermair, C. 2011. "Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters." *Nature structural & molecular biology* 18 (8):956.

Korte, A., and Farlow, A. 2013. "The advantages and limitations of trait analysis with GWAS: a review." *Plant methods* 9 (1):29.

Koshikawa, N., Hasegawa, S., Nagashima, Y., Mitsuhashi, K., Tsubota, Y., Miyata, S., Miyagi, Y., Yasumitsu, H., and Miyazaki, K. 1998. "Expression of trypsin by epithelial cells of various tissues, leukocytes, and neurons in human and mouse." *The American journal of pathology* 153 (3):937-944.

Kouzarides, T. 2002. "Histone methylation in transcriptional control." *Current opinion in genetics & development* 12 (2):198-209.

Kouzarides, T. 2007. "Chromatin modifications and their function." *Cell* 128 (4):693-705.

Kratochwil, C.F., and Meyer, A. 2015a. "Mapping active promoters by Ch IP‐seq profiling of H3K4me3 in cichlid fish – a first step to uncover cis‐regulatory elements in ecological model teleosts." *Molecular ecology resources* 15 (4):761-771.

Kratochwil, C.F., and Meyer, A. 2015b. "Mapping active promoters by ChIP-seq profiling of H3K4me3 in cichlid fish - a first step to uncover cis-regulatory elements in ecological model teleosts." *Mol Ecol Resour* 15 (4):761-71. doi: 10.1111/1755-0998.12350.

Krishnan, S., Horowitz, S., and Trievel, R.C. 2011. "Structure and function of histone H3 lysine 9 methyltransferases and demethylases." *Chembiochem* 12 (2):254-263.

Kucuktas, H., Wang, S., Li, P., He, C., Xu, P., Sha, Z., Liu, H., Jiang, Y., Baoprasertkul, P., and Somridhivej, B. 2009. "Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers." *Genetics* 181 (4):1649-1660.

Lachner, M., and Jenuwein, T. 2002. "The many faces of histone lysine methylation." *Current opinion in cell biology* 14 (3):286-298.

Lallias, D., Gomez-Raya, L., Haley, C., Arzul, I., Heurtebise, S., Beaumont, A., Boudry, P., and Lapègue, S. 2009. "Combining two-stage testing and interval mapping strategies to detect QTL for resistance to bonamiosis in the European flat oyster Ostrea edulis." *Marine Biotechnology* 11 (5):570.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. 1992. "CpG islands as gene markers in the human genome." *Genomics* 13 (4):1095-1107.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. 2013. "H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation." *Cell* 152 (5):1021-1036.

Lemon, B., and Tjian, R. 2000. "Orchestrated response: a symphony of transcription factors for gene control." *Genes & development* 14 (20):2551-2569.

Levine, M. 2010. "Transcriptional enhancers in animal development and evolution." *Current Biology* 20 (17):R754-R763.

Levine, M., and Tjian, R. 2003. "Transcription regulation and animal diversity." *Nature* 424 (6945):147.

Li, C., Zhang, Y., Wang, R., Lu, J., Nandi, S., Mohanty, S., Terhune, J., Liu, Z., and Peatman, E. 2012. "RNA-seq analysis of mucosal immune responses reveals signatures of intestinal barrier disruption and pathogen entry following Edwardsiella ictaluri infection in channel catfish, *Ictalurus punctatus*." *Fish & shellfish immunology* 32 (5):816-827.

Li, H., and Durbin, R. 2009. "Fast and accurate short read alignment with Burrows–Wheeler transform." *bioinformatics* 25 (14):1754-1760.

Li, N., Zhou, T., Geng, X., Jin, Y., Wang, X., Liu, S., Xu, X., Gao, D., Li, Q., and Liu, Z. 2018a. "Identification of novel genes significantly affecting growth in catfish through GWAS analysis." *Molecular genetics and genomics*:1-13.

Li, Y., Liu, S., Qin, Z., Waldbieser, G., Wang, R., Sun, L., Bao, L., Danzmann, R.G., Dunham, R., and Liu, Z. 2014. "Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish." *DNA Research* 22 (1):39-52.

Li, Z., Liu, X., Liu, J., Zhang, K., Yu, H., He, Y., Wang, X., Qi, J., Wang, Z., and Zhang, Q. 2018b. "Transcriptome profiling based on protein–protein interaction networks provides a core set of genes for understanding blood immune response mechanisms against Edwardsiella tarda infection in Japanese flounder (Paralichthys olivaceus)." *Developmental & Comparative Immunology* 78:100-113.

Liu, F., Sun, F., Li, J., Xia, J.H., Lin, G., Tu, R.J., and Yue, G.H. 2013a. "A microsatellite-based linkage map of salt tolerant tilapia (Oreochromis mossambicus x Oreochromis spp.) and mapping of sex-determining loci." *BMC genomics* 14 (1):58.

Liu, S., Li, Y., Qin, Z., Geng, X., Bao, L., Kaltenboeck, L., Kucuktas, H., Dunham, R., and Liu, Z. 2016a. "High‐density interspecific genetic linkage mapping provides insights into genomic incompatibility between channel catfish and blue catfish." *Animal genetics* 47 (1):81-90.

Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., Zhang, J., Feng, J., Kaltenboeck, L., and Kucuktas, H. 2014. "Development of the catfish 250K SNP array for genome-wide association studies." *BMC research notes* 7 (1):135.

Liu, S., Vallejo, R.L., Palti, Y., Gao, G., Marancik, D.P., Hernandez, A.G., and Wiens, G.D. 2015. "Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout." *Frontiers in genetics* 6:298.

Liu, S., Wang, X., Sun, F., Zhang, J., Feng, J., Liu, H., Rajendran, K., Sun, L., Zhang, Y., and Jiang, Y. 2013b. "RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding and degradation in response to heat stress in catfish." *American Journal of Physiology-Heart and Circulatory Physiology*.

Liu, Z., Karsi, A., Li, P., Cao, D., and Dunham, R. 2003. "An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family." *Genetics* 165 (2):687-694.

Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., Jiang, C., Sun, L., Wang, R., and Zhang, Y. 2016b. "The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts." *Nature communications* 7:11757.

Lu, J., Zheng, M., Zheng, J., Liu, J., Liu, Y., Peng, L., Wang, P., Zhang, X., Wang, Q., and Luan, P. 2015. "Transcriptomic analyses reveal novel genes with sexually dimorphic expression in yellow catfish (Pelteobagrus fulvidraco) brain." *Marine Biotechnology* 17 (5):613-623.

Ma, H., Baumann, C.T., Li, H., Strahl, B.D., Rice, R., Jelinek, M.A., Aswad, D.W., Allis, C.D., Hager, G.L., and Stallcup, M.R. 2001. "Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter." *Current Biology* 11 (24):1981-1985.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. 1994. "Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island." *Genes & development* 8 (19):2282-2292.

Main, B.J., Smith, A.D., Jang, H., and Nuzhdin, S.V. 2013. "Transcription start site evolution in Drosophila." *Molecular biology and evolution* 30 (8):1966-1974.

Martin, C., and Zhang, Y. 2007. "Mechanisms of epigenetic inheritance." *Current opinion in cell biology* 19 (3):266-272.

McBride, A.E., and Silver, P.A. 2001. "State of the arg: protein methylation at arginine comes of age." *Cell* 106 (1):5-8.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. 2013. "Large-scale gene function analysis with the PANTHER classification system." *Nature protocols* 8 (8):1551.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., and Koche, R.P. 2007. "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." *Nature* 448 (7153):553.

Moen, T., Baranski, M., Sonesson, A.K., and Kjøglum, S. 2009. "Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait." *BMC genomics* 10 (1):368.

Moen, T., Torgersen, J., Santi, N., Davidson, W.S., Baranski, M., Ødegård, J., Kjøglum, S., Velle, B., Kent, M., and Lubieniecki, K.P. 2015. "Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon." *Genetics* 200 (4):1313-1326.

Mooseker, M.S., Pollard, T.D., and Wharton, K.A. 1982. "Nucleated polymerization of actin from the membrane-associated ends of microvillar filaments in the intestinal brush border." *The Journal of cell biology* 95 (1):223-233.

Mouchiroud, D., D'Onofrio, G., Aissani, B., Macaya, G., Gautier, C., and Bernardi, G. 1991. "The distribution of genes in the human genome." *Gene* 100:181-187.

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M.M., Sheng, Y., Abdelhamid, R.F., and Anand, S. 2013. "Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis." *Genome research* 23 (11):1938-1950.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. 2003. "Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity." *Molecular cell* 11 (3):709-719.

Nguyen, T.V., Jung, H., Nguyen, T.M., Hurwood, D., and Mather, P. 2016. "Evaluation of potential candidate genes involved in salinity tolerance in striped catfish (Pangasianodon hypophthalmus) using an RNA-Seq approach." *Marine genomics* 25:75-88.

Nightingale, K.P., O'Neill, L.P., and Turner, B.M. 2006. "Histone modifications: signalling receptors and potential elements of a heritable epigenetic code." *Current opinion in genetics & development* 16 (2):125-136.

Ninwichian, P., Peatman, E., Liu, H., Kucuktas, H., Somridhivej, B., Liu, S., Li, P., Jiang, Y., Sha, Z., and Kaltenboeck, L. 2012. "Second-generation genetic linkage map of catfish and its integration with the BAC-based physical map." *G3: Genes, Genomes, Genetics* 2 (10):1233-1241.

Niu, W., Lu, Z.J., Zhong, M., Sarov, M., Murray, J.I., Brdlik, C.M., Janette, J., Chen, C., Alves, P., and Preston, E. 2011. "Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans." *Genome research* 21 (2):245-254.

Novik, K., Nimmrich, I., Genc, B., Maier, S., Piepenbrock, C., Olek, A., and Beck, S. 2002. "Epigenomics: genome-wide study of methylation phenomena." *Current issues in molecular biology* 4:111-128.

Ohler, U. 2006. "Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction." *Nucleic acids research* 34 (20):5943-5950.

Ohtsuji, M., Katsuoka, F., Kobayashi, A., Aburatani, H., Hayes, J.D., and Yamamoto, M. 2008. "Nrf1 and Nrf2 play distinct roles in activation of antioxidant response element-dependent genes." *Journal of Biological Chemistry* 283 (48):33554-33562.

Ohtsuki, S., Levine, M., and Cai, H.N. 1998. "Different core promoters possess distinct regulatory activities in the Drosophila embryo." *Genes & development* 12 (4):547-556.

Ong, C.-T., and Corces, V.G. 2011. "Enhancer function: new insights into the regulation of tissue-specific gene expression." *Nature Reviews Genetics* 12 (4):283.

Orphanides, G., and Reinberg, D. 2002. "A unified theory of gene expression." *Cell* 108 (4):439-451.

Ouyang, Z., Zhou, Q., and Wong, W.H. 2009. "ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells." *Proceedings of the National Academy of Sciences* 106 (51):21521-21526.

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M., Akutsu, T., and Okamoto, N. 2001. "Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (Oncorhynchus mykiss)." *Molecular Genetics and Genomics* 265 (1):23-31.

Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M.E., Oral, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., Penman, D.J., and Migaud, H. 2013. "Mapping the sex determination locus in the Atlantic halibut (Hippoglossus hippoglossus) using RAD sequencing." *BMC genomics* 14 (1):566.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. 2007. "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells." *Cell stem cell* 1 (3):299-312.

Park, P.J. 2009. "ChIP–seq: advantages and challenges of a maturing technology." *Nature reviews genetics* 10 (10):669.

Patiño, R., Davis, K.B., Schoore, J.E., Uguz, C., Strüssmann, C.A., Parker, N.C., Simco, B.A., and Goudie, C.A. 1996. "Sex differentiation of channel catfish gonads: normal development and effects of temperature." *Journal of Experimental Zoology* 276 (3):209-218.

Peatman, E., Lange, M., Zhao, H., and Beck, B.H. 2015. "Physiology and immunology of mucosal barriers in catfish (Ictalurus spp.)." *Tissue barriers* 3 (4):e1068907.

Peatman, E., Li, C., Peterson, B.C., Straus, D.L., Farmer, B.D., and Beck, B.H. 2013. "Basal polarization of the mucosal compartment in Flavobacterium columnare susceptible and resistant channel catfish (*Ictalurus punctatus*)." *Molecular immunology* 56 (4):317-327.

Peatman, E., Terhune, J., Baoprasertkul, P., Xu, P., Nandi, S., Wang, S., Somridhivej, B., Kucuktas, H., Li, P., and Dunham, R. 2008. "Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with Edwardsiella ictaluri." *Molecular immunology* 45 (2):553-566.

Pena, P., Hom, R., Hung, T., Lin, H., Kuo, A., Wong, R., Subach, O., Champagne, K., Zhao, R., and Verkhusha, V. 2008. "Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor." *Journal of molecular biology* 380 (2):303-312.

Peng, W., Xu, J., Zhang, Y., Feng, J., Dong, C., Jiang, L., Feng, J., Chen, B., Gong, Y., and Chen, L. 2016. "An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (Cyprinus carpio)." *Scientific reports* 6:26693.

Peterson, B.C., Flora, C., Wood, M., Bosworth, B.G., Quiniou, S.M., Greenway, T.E., Byars, T.S., and Wise, D.J. 2016. "Vaccination of Full‐s ib Channel Catfish Families Against Enteric Septicemia of Catfish with an Oral Live Attenuated Edwardsiella ictaluri Vaccine." *Journal of the World Aquaculture Society* 47 (2):207-211.

Pilon, A.M., Ajay, S.S., Kumar, S.A., Steiner, L.A., Cherukuri, P.F., Wincovitch, S., Anderson, S.M., Center, N.C.S., Mullikin, J.C., and Gallagher, P.G. 2011. "Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation." *Blood* 118 (17):e139-e148.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., and Herbolsheimer, E. 2005. "Genome-wide map of nucleosome acetylation and methylation in yeast." *Cell* 122 (4):517-527.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. 2007. "PLINK: a tool set for whole-genome association and population-based linkage analyses." *The American journal of human genetics* 81 (3):559-575.

Qi, H.H., Sarkissian, M., Hu, G.-Q., Wang, Z., Bhattacharjee, A., Gordon, D.B., Gonzales, M., Lan, F., Ongusaha, P.P., and Huarte, M. 2010. "Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development." *Nature* 466 (7305):503.

Qin, L., Li, F., Wang, X., Sun, Y., Bi, K., and Gao, Y. 2017. "Proteomic analysis of macrophage in response to Edwardsiella tarda-infection." *Microbial pathogenesis* 111:86-93.

Ramagopalan, S.V., Heger, A., Berlanga, A.J., Maugeri, N.J., Lincoln, M.R., Burrell, A., Handunnetthi, L., Handel, A.E., Disanto, G., and Orton, S.-M. 2010. "A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution." *Genome research*.

Rea, S., Eisenhaber, F., O'carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., and Allis, C.D. 2000. "Regulation of chromatin structure by site-specific histone H3 methyltransferases." *Nature* 406 (6796):593.

Rebl, A., Verleih, M., Korytář, T., Kühn, C., Wimmers, K., Köllner, B., and Goldammer, T. 2012. "Identification of differentially expressed protective genes in liver of two rainbow trout strains." *Veterinary immunology and immunopathology* 145 (1-2):305-315.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., and Delaney, A. 2007. "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing." *Nature methods* 4 (8):651.

Rodríguez-Ramilo, S.T., Toro, M.A., Bouza, C., Hermida, M., Pardo, B.G., Cabaleiro, S., Martínez, P., and Fernández, J. 2011. "QTL detection for Aeromonas salmonicida resistance related traits in turbot (Scophthalmus maximus)." *BMC genomics* 12 (1):541.

Russ, B.E., Olshansky, M., Li, J., Nguyen, M.L., Gearing, L.J., Nguyen, T.H., Olson, M.R., McQuilton, H.A., Nüssing, S., and Khoury, G. 2017. "Regulation of H3K4me3 at transcriptional enhancers characterizes acquisition of virus-specific CD8+ T Cell-lineage-specific function." *Cell reports* 21 (12):3624-3636.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. 2007. "Multivalent engagement of chromatin modifications by linked binding modules." *Nature reviews Molecular cell biology* 8 (12):983.

Sánchez-Molano, E., Cerna, A., Toro, M.A., Bouza, C., Hermida, M., Pardo, B.G., Cabaleiro, S., Fernández, J., and Martínez, P. 2011. "Detection of growth-related QTL in turbot (Scophthalmus maximus)." *BMC genomics* 12 (1):473.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. 2007. "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." *Nature Reviews Genetics* 8 (6):424.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. 2002. "Active genes are tri-methylated at K4 of histone H3." *Nature* 419 (6905):407.

Sattar, H.A. 2016. *Fundamentals of pathology: Medical course and step 1 review*: Pathoma. com.

Savinkova, L., Ponomarenko, M., Ponomarenko, P., Drachkova, I., Lysova, M., Arshinova, T., and Kolchanov, N. 2009. "TATA box polymorphisms in human gene promoters and associated hereditary pathologies." *Biochemistry (Moscow)* 74 (2):117-129.

Saxonov, S., Berg, P., and Brutlag, D.L. 2006. "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." *Proceedings of the National Academy of Sciences* 103 (5):1412-1417.

Scharer, C.D., McCabe, C.D., Ali-Seyed, M., Berger, M.F., Bulyk, M.L., and Moreno, C.S. 2009. "Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells." *Cancer research* 69 (2):709-717.

Schmid, C.D., and Bucher, P. 2007. "ChIP-Seq data reveal nucleosome architecture of human promoters." *Cell* 131 (5):831-832.

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. 2010a. "A CTCF-independent role for cohesin in tissue-specific transcription." *Genome research* 20 (5):578-588.

Schmidt, D., Wilson, M., Ballester, B., Schwalie, P., Brown, G., Marshall, A., Kutter, C., and Watt, S. "C. 844 P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek, and DT Odom. 2010. Five-vertebrate ChIP-845 seq reveals the evolutionary dynamics of transcription factor binding 846 1." *Science* 328:1036-1040.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., and Mackay, S. 2010b. "Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding." *Science* 328 (5981):1036-1040.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. 2004. "Histone H3 lysine 4 methylation patterns in higher eukaryotic genes." *Nature cell biology* 6 (1):73.

Schödel, J., Oikonomopoulos, S., Ragoussis, J., Pugh, C.W., Ratcliffe, P.J., and Mole, D.R. 2011. "High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq." *Blood* 117 (23):e207-e217.

Schübeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., Van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., and Delrow, J. 2004. "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote." *Genes & development* 18 (11):1263-1271.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. 2008. "Divergent transcription from active promoters." *science* 322 (5909):1849-1851.

Shao, C., Niu, Y., Rastas, P., Liu, Y., Xie, Z., Li, H., Wang, L., Jiang, Y., Tai, S., and Tian, Y. 2015. "Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (Paralichthys olivaceus): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis." *DNA research* 22 (2):161-170.

Shi, H., Zhou, T., Wang, X., Yang, Y., Wu, C., Liu, S., Bao, L., Li, N., Yuan, Z., and Jin, Y. 2018. "Genome-wide association analysis of intra-specific QTL associated with the resistance for enteric septicemia of catfish." *Molecular Genetics and Genomics*:1-14.

Shi, Y., Lee, J.-S., and Galvin, K.M. 1997. "Everything you have ever wanted to know about Yin Yang 1." *Biochimica et biophysica acta* 1332 (2):F49-66.

Shilatifard, A. 2006. "Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression." *Annu. Rev. Biochem.* 75:243-269.

Shimizu, M., Mori, T., Sakurai, T., and Shindo, H. 2000. "Destabilization of nucleosomes by an unusual DNA conformation adopted by poly (dA)· poly (dT) tracts in vivo." *The EMBO journal* 19 (13):3358-3365.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. 2006. "Histone H4-K16 acetylation controls chromatin structure and protein interactions." *Science* 311 (5762):844-847.

Sinal, C.J., Miyata, M., Tohkin, M., Nagata, K., Bend, J.R., and Gonzalez, F.J. 2000. "Targeted disruption of soluble epoxide hydrolase reveals a role in blood pressure regulation." *Journal of Biological Chemistry* 275 (51):40504-40510.

Smale, S.T., and Kadonaga, J.T. 2003. "The RNA polymerase II core promoter." *Annual review of biochemistry* 72 (1):449-479.

Sodeland, M., Gaarder, M., Moen, T., Thomassen, M., Kjøglum, S., Kent, M., and Lien, S. 2013. "Genome-wide association testing reveals quantitative trait loci for fillet texture and fat content in Atlantic salmon." *Aquaculture* 408:169-174.

Song, W., Li, Y., Zhao, Y., Liu, Y., Niu, Y., Pang, R., Miao, G., Liao, X., Shao, C., and Gao, F. 2012. "Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (Cynoglossus semilaevis)." *PLoS One* 7 (12):e52097.

Strahl, B.D., and Allis, C.D. 2000. "The language of covalent histone modifications." *Nature* 403 (6765):41.

Struhl, K., Kadosh, D., Keaveney, M., Kuras, L., and Moqtaderi, Z. 1998. "Activation and repression mechanisms in yeast." Cold Spring Harbor symposia on quantitative biology.

Sullivan, J.P., Lundberg, J.G., and Hardman, M. 2006. "A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences." *Molecular phylogenetics and evolution* 41 (3):636-662.

Sun, F., Liu, S., Gao, X., Jiang, Y., Perera, D., Wang, X., Li, C., Sun, L., Zhang, J., and Kaltenboeck, L. 2013. "Male-biased genes in catfish as revealed by RNA-Seq analysis of the testis transcriptome." *PloS one* 8 (7):e68452.

Sun, F., Peatman, E., Li, C., Liu, S., Jiang, Y., Zhou, Z., and Liu, Z. 2012. "Transcriptomic signatures of attachment, NF-κB suppression and IFN stimulation in the catfish gill following columnaris bacterial infection." *Developmental & Comparative Immunology* 38 (1):169-180.

Sun, L., Liu, S., Bao, L., Li, Y., Feng, J., and Liu, Z. 2015. "Claudin multigene family in channel catfish and their expression profiles in response to bacterial infection and hypoxia as revealed by meta-analysis of RNA-Seq datasets." *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 13:60-69.

Tacchi, L., Bickerdike, R., Secombes, C.J., Pooley, N.J., Urquhart, K.L., Collet, B., and Martin, S.A. 2010. "Ubiquitin E3 ligase atrogin-1 (Fbox-32) in Atlantic salmon (Salmo salar): sequence analysis, genomic structure and modulation of expression." *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 157 (4):364-373.

Tan, S., Wang, W., Zhong, X., Tian, C., Niu, D., Bao, L., Zhou, T., Jin, Y., Yang, Y., and Yuan, Z. 2018a. "Increased Alternative Splicing as a Host Response to Edwardsiella ictaluri Infection in Catfish." *Marine Biotechnology* 20 (6):729-738.

Tan, S., Zhou, T., Wang, W., Jin, Y., Wang, X., Geng, X., Luo, J., Yuan, Z., Yang, Y., and Shi, H. 2018b. "GWAS analysis using interspecific backcross progenies reveals superior blue catfish alleles responsible for strong resistance against enteric septicemia of catfish." *Molecular Genetics and Genomics*:1-14.

Tena, J.J., González-Aguilera, C., Fernández-Miñán, A., Vázquez-Marín, J., Parra-Acero, H., Cross, J.W., Rigby, P.W., Carvajal, J.J., Wittbrodt, J., and Gómez-Skarmeta, J.L. 2014. "Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period." *Genome research*:gr. 163915.113.

Tilney, L.G., and Mooseker, M. 1971. "Actin in the brush-border of epithelial cells of the chicken intestine." *Proceedings of the National Academy of Sciences* 68 (10):2611-2615.

Tivesten, A.s., Bollano, E., Andersson, I., Fitzgerald, S., Caidahl, K., Sjögren, K., Skøtt, O., Liu, J.-L., Mobini, R., and Isaksson, O.G. 2002. "Liver-derived insulin-like growth factor-I is involved in the regulation of blood pressure in mice." *Endocrinology* 143 (11):4235-4242.

Tjian, R., and Maniatis, T. 1994. "Transcriptional activation: a complex puzzle with few easy pieces." *Cell* 77 (1):5-8.

Tobias, K.E., Shor, J., and Kahana, C. 1995. "c-Myc and Max transregulate the mouse ornithine decarboxylase promoter through interaction with two downstream CACGTG motifs." *Oncogene* 11 (9):1721-1727.

Töx, U., Scheller, I., Kociok, N., Kern, M.A., Klanac, D., Daudi, S.M., Laue, O., Schirmacher, P., Goeser, T., and Schulte, S. 2007. "Expression of angiotensin II receptor type 1 is reduced in advanced rat liver fibrosis." *Digestive diseases and sciences* 52 (8):1995-2005.

Trapani, J.A., and Smyth, M.J. 2002. "Functional significance of the perforin/granzyme cell death pathway." *Nature Reviews Immunology* 2 (10):735.

Trinklein, N.D., Aldred, S.J.F., Saldanha, A.J., and Myers, R.M. 2003. "Identification and functional analysis of human transcriptional promoters." *Genome research* 13 (2):308-312.

Tucker, C. 2012. "Channel catfish." *Aquaculture Farming Aquatic Animals and Plants. 2nd Edition. Wiley-Blackwell, Sussex, UK*:365-383.

Uetz, P., and Hošek, J. 2016. The reptile database.

Vallejo, R.L., Bacon, L.D., Liu, H.-C., Witter, R.L., Groenen, M.A., Hillel, J., and Cheng, H.H. 1998. "Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens." *Genetics* 148 (1):349-360.

Vallejo, R.L., Wiens, G.D., Rexroad, C., Welch, T.J., Evenhuis, J.P., Leeds, T.D., Janss, L.L., and Palti, Y. 2010. "Evidence of major genes affecting resistance to bacterial cold water disease in rainbow trout using Bayesian methods of segregation analysis." *Journal of animal science* 88 (12):3814-3832.

Viñas, A., Taboada, X., Vale, L., Robledo, D., Hermida, M., Vera, M., and Martínez, P. 2012. "Mapping of DNA sex-specific markers and genes related to sex differentiation in turbot (Scophthalmus maximus)." *Marine biotechnology* 14 (5):655-663.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., and Chen, F. 2009. "ChIP-seq accurately predicts tissue-specific activity of enhancers." *Nature* 457 (7231):854.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. 2012. "Five years of GWAS discovery." *The American Journal of Human Genetics* 90 (1):7-24.

Voigt, P., LeRoy, G., Drury III, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. 2012. "Asymmetrically modified nucleosomes." *Cell* 151 (1):181-193.

Voigt, P., Tee, W.-W., and Reinberg, D. 2013. "A double take on bivalent promoters." *Genes & development* 27 (12):1318-1338.

Wagner, B.A., Wise, D.J., Khoo, L.H., and Terhune, J.S. 2002. "The epidemiology of bacterial diseases in food-size channel catfish." *Journal of Aquatic Animal Health* 14 (4):263-272.

Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J., and Wolters, W.R. 2001. "A microsatellite-based genetic linkage map for channel catfish, Ictalurus punctatus." *Genetics* 158 (2):727-734.

Wang, C., Lo, L., Feng, F., Zhu, Z., and Yue, G. 2008. "Identification and verification of QTL associated with growth traits in two genetic backgrounds of Barramundi (Lates calcarifer)." *Animal Genetics* 39 (1):34-39.

Wang, H., Zou, J., Zhao, B., Johannsen, E., Ashworth, T., Wong, H., Pear, W.S., Schug, J., Blacklow, S.C., and Arnett, K.L. 2011. "Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells." *Proceedings of the National Academy of Sciences* 108 (36):14908-14913.

Wang, L., Fan, C., Liu, Y., Zhang, Y., Liu, S., Sun, D., Deng, H., Xu, Y., Tian, Y., and Liao, X. 2014a. "A genome scan for quantitative trait loci associated with Vibrio anguillarum infection resistance in Japanese flounder (Paralichthys olivaceus) by bulked segregant analysis." *Marine biotechnology* 16 (5):513-521.

Wang, L., Liu, P., Huang, S., Ye, B., Chua, E., Wan, Z.Y., and Yue, G.H. 2017a. "Genome-wide association study identifies loci associated with resistance to viral nervous necrosis disease in Asian seabass." *Marine Biotechnology* 19 (3):255-265.

Wang, L., Wan, Z.Y., Bai, B., Huang, S.Q., Chua, E., Lee, M., Pang, H.Y., Wen, Y.F., Liu, P., and Liu, F. 2015. "Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass." *Scientific reports* 5:16358.

Wang, R., Sun, L., Bao, L., Zhang, J., Jiang, Y., Yao, J., Song, L., Feng, J., Liu, S., and Liu, Z. 2013. "Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish." *BMC genomics* 14 (1):929.

Wang, R., Zhang, Y., Liu, S., Li, C., Sun, L., Bao, L., Feng, J., and Liu, Z. 2014b. "Analysis of 52 Rab GTPases from channel catfish and their involvement in immune responses after bacterial infections." *Developmental & Comparative Immunology* 45 (1):21-34.

Wang, W., Huang, Y., Yu, Y., Yang, Y., Xu, M., Chen, X., Ni, S., Qin, Q., and Huang, X. 2016. "Fish TRIM39 regulates cell cycle progression and exerts its antiviral function against iridovirus and nodavirus." *Fish & shellfish immunology* 50:1-10.

Wang, W., Tan, S., Luo, J., Shi, H., Zhou, T., Yang, Y., Jin, Y., Wang, X., Niu, D., and Yuan, Z. 2019. "GWAS Analysis Indicated Importance of NF-κB Signaling Pathway in Host Resistance Against Motile Aeromonas Septicemia Disease in Catfish." *Marine Biotechnology*:1-13.

Wang, X., Li, C., Thongda, W., Luo, Y., Beck, B., and Peatman, E. 2014c. "Characterization and mucosal responses of interleukin 17 family ligand and receptor genes in channel catfish *Ictalurus punctatus*." *Fish & shellfish immunology* 38 (1):47-55.

Wang, X., Liu, S., Jiang, C., Geng, X., Zhou, T., Li, N., Bao, L., Li, Y., Yao, J., and Yang, Y. 2017b. "Multiple across-strain and within-strain QTLs suggest highly complex genetic architecture for hypoxia tolerance in channel catfish." *Molecular Genetics and Genomics* 292 (1):63-76.

Wang, Z., Gerstein, M., and Snyder, M. 2009. "RNA-Seq: a revolutionary tool for transcriptomics." *Nature reviews genetics* 10 (1):57.

Waterborg, J.H. 2002. "Dynamics of histone acetylation in vivo. A function for acetylation turnover?" *Biochemistry and cell biology* 80 (3):363-378.

Weber, H., Polen, T., Heuveling, J., Wendisch, V.F., and Hengge, R. 2005a. "Genome-wide analysis of the general stress response network in Escherichia coli: σS-dependent genes, promoters, and sigma factor selectivity." *Journal of bacteriology* 187 (5):1591-1603.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schuebeler, D. 2005b. "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." *Nature genetics* 37 (8):853.

Welboren, W.J., Van Driel, M.A., Janssen‑Megens, E.M., Van Heeringen, S.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. 2009. "ChIP‑Seq of ERα and RNA polymerase II defines genes differentially responding to ligands." *The EMBO journal* 28 (10):1418-1428.

Williams, M.E., Foster, R., and Chua, N.-H. 1992. "Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding." *The Plant Cell* 4 (4):485-496.

Wilson, D.E., and Reeder, D.M. 2005. *Mammal species of the world: a taxonomic and geographic reference*. Vol. 1: JHU Press.

Wittkopp, P.J., and Kalay, G. 2012. "Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence." *Nature Reviews Genetics* 13 (1):59.

Wolters, W.R., Wise, D.J., and Klesius, P.H. 1996. "Survival and antibody response of channel catfish, blue catfish, and channel catfish female× blue catfish male hybrids after exposure to Edwardsiella ictaluri." *Journal of Aquatic Animal Health* 8 (3):249-254.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. 2003. "The evolution of transcriptional regulation in eukaryotes." *Molecular biology and evolution* 20 (9):1377-1419.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. 2005. "WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development." *Cell* 121 (6):859-872.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., and Badenhorst, P. 2006. "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling." *Nature* 442 (7098):86.

Xie, Y., Song, L., Weng, Z., Liu, S., and Liu, Z. 2015. "Hsp90, Hsp60 and sHsp families of heat shock protein genes in channel catfish and their expression after bacterial infections." *Fish & shellfish immunology* 44 (2):642-651.

Xuan, Z., Zhao, F., Wang, J., Chen, G., and Zhang, M.Q. 2005. "Genome-wide promoter extraction and analysis in human, mouse, and rat." *Genome biology* 6 (8):R72.

Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K. 2005. "Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity." *Gene* 350 (2):129-136.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. 2007. "Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters." *Gene* 389 (1):52-65.

Yang, Y., Fu, Q., Liu, Y., Wang, X., Dunham, R., Liu, S., Bao, L., Zeng, Q., Zhou, T., and Li, N. 2017. "Comparative transcriptome analysis reveals conserved branching morphogenesis related genes involved in chamber formation of catfish swimbladder." *Physiological genomics*.

Yang, Y., Fu, Q., Wang, X., Liu, Y., Zeng, Q., Li, Y., Gao, S., Bao, L., Liu, S., and Gao, D. 2018a. "Comparative transcriptome analysis of the swimbladder reveals expression signatures in response to low oxygen stress in channel catfish, *Ictalurus punctatus*." *Physiological genomics*.

Yang, Y., Wang, X., Liu, Y., Fu, Q., Tian, C., Wu, C., Shi, H., Yuan, Z., Tan, S., and Liu, S. 2018b. "Transcriptome analysis reveals enrichment of genes associated with auditory system in swimbladder of channel catfish." *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 27:30-39.

Yang, Y., Yu, H., Li, H., and Wang, A. 2016. "Transcriptome profiling of grass carp (Ctenopharyngodon idellus) infected with Aeromonas hydrophila." *Fish & shellfish immunology* 51:329-336.

Yao, J., Li, C., Zhang, J., Liu, S., Feng, J., Wang, R., Li, Y., Jiang, C., Song, L., and Chen, A. 2014. "Expression of nitric oxide synthase (NOS) genes in channel catfish is highly regulated and time dependent after bacterial challenges." *Developmental & Comparative Immunology* 45 (1):74-86.

Yu, G., Wang, L.-G., and He, Q.-Y. 2015. "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization." *Bioinformatics* 31 (14):2382-2383.

Yuan, Z., Liu, S., Yao, J., Zeng, Q., Tan, S., and Liu, Z. 2016. "Expression of Bcl-2 genes in channel catfish after bacterial infection and hypoxia stress." *Developmental & Comparative Immunology* 65:79-90.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. 2009. "A clustering approach for identification of enriched domains from histone modification ChIP-Seq data." *Bioinformatics* 25 (15):1952-1958.

Zeng, L., and Zhou, M.-M. 2002. "Bromodomain: an acetyl‑lysine binding domain." *FEBS letters* 513 (1):124-128.

Zeng, Q., Fu, Q., Li, Y., Waldbieser, G., Bosworth, B., Liu, S., Yang, Y., Bao, L., Yuan, Z., and Li, N. 2017. "Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence." *Scientific reports* 7:40347.

Zeng, Q., Liu, S., Yao, J., Zhang, Y., Yuan, Z., Jiang, C., Chen, A., Fu, Q., Su, B., and Dunham, R. 2016. "Transcriptome display during testicular differentiation of channel catfish (*Ictalurus punctatus*) as revealed by RNA-Seq analysis." *Biology of reproduction* 95 (1):19, 1-17.

Zhang, J., Yao, J., Wang, R., Zhang, Y., Liu, S., Sun, L., Jiang, Y., Feng, J., Liu, N., and Nelson, D. 2014. "The cytochrome P450 genes of channel catfish: their involvement in disease defense responses as revealed by meta-analysis of RNA-Seq data sets." *Biochimica et Biophysica Acta (BBA)-General Subjects* 1840 (9):2813-2828.

Zhang, X., Zhao, H., Chen, Y., Luo, H., Yang, P., and Yao, B. 2015. "A zebrafish (Danio rerio) bloodthirsty member 20 with E3 ubiquitin ligase activity involved in immune response against bacterial infection." *Biochemical and biophysical research communications* 457 (1):83-89.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., and Li, W. 2008. "Model-based analysis of ChIP-Seq (MACS)." *Genome biology* 9 (9):R137.

Zhang, Y., and Reinberg, D. 2001. "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." *Genes & development* 15 (18):2343-2360.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.-K., Shahab, A., and Kuznetsov, V.A. 2007. "Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells." *Cell stem cell* 1 (3):286-298.

Zhong, X., Wang, X., Zhou, T., Jin, Y., Tan, S., Jiang, C., Geng, X., Li, N., Shi, H., and Zeng, Q. 2017. "Genome-wide association study reveals multiple novel QTL associated with low oxygen tolerance in hybrid catfish." *Marine Biotechnology* 19 (4):379-390.

Zhou, T., Li, N., Liu, S., Jin, Y., Fu, Q., Gao, S., Liu, Y., and Liu, Z. 2017a. "The NCK and ABI adaptor genes in catfish and their involvement in ESC disease response." *Developmental & Comparative Immunology* 73:119-123.

Zhou, T., Liu, S., Geng, X., Jin, Y., Jiang, C., Bao, L., Yao, J., Zhang, Y., Zhang, J., and Sun, L. 2017b. "GWAS analysis of QTL for enteric septicemia of catfish and their involved genes suggest evolutionary conservation of a molecular mechanism of disease resistance." *Molecular Genetics and Genomics* 292 (1):231-242.

Zhou, V.W., Goren, A., and Bernstein, B.E. 2011. "Charting histone modifications and the functional organization of mammalian genomes." *Nature Reviews Genetics* 12 (1):7.

Zhou, Y., Geng, Y., Yang, J., Zhao, C., Harrison, T.J., and Wang, Y. 2014. "Hepatitis E virus open reading frame 3 protein interacts with porcine liver‐specific plasminogen and α2‐antiplasmin." *Journal of medical virology* 86 (3):487-495.

Zou, M., Zhang, X., Shi, Z., Lin, L., Ouyang, G., Zhang, G., Zheng, H., Wei, K., and Ji, W. 2015. "A comparative transcriptome analysis between wild and albino yellow catfish (Pelteobagrus fulvidraco)." *PloS one* 10 (6):e0131504.

Supplementary

Table S1

Linkage Group	Marker	Position	-LOG10(P-Value)	% variance
1	AX-157890307	1,833,304	4.354	5.959
1	AX-157868447	1,911,441	4.259	5.815
1	AX-157890389	1,945,300	4.318	5.903
1	AX-157890433	2,010,603	4.779	6.601
1	AX-157890444	2,027,727	4.274	5.838
1	AX-157868449	2,068,488	4.249	5.800
1	AX-157717283	2,142,398	4.366	5.977
1	AX-157750774	2,193,567	4.557	6.265
1	AX-157890557	2,218,367	4.448	6.100
1	AX-158388951	2,228,169	4.324	5.913
1	AX-157890591	2,261,576	4.355	5.960
1	AX-157890594	2,264,534	4.408	6.041
1	AX-158379093	2,271,985	4.206	5.735
1	AX-157890613	2,291,803	4.355	5.960
1	AX-157890616	2,299,470	4.355	5.960
1	AX-157890617	2,299,582	4.384	6.004
1	AX-157890623	2,313,997	4.355	5.960
1	AX-158388955	2,344,434	4.631	6.378
1	AX-157890670	2,356,117	4.836	6.687
1	AX-157890672	2,358,420	4.240	5.785
1	AX-157890714	2,411,070	5.416	7.561
1	AX-157890718	2,424,488	4.293	5.866
1	AX-157890733	2,442,920	4.595	6.323
1	AX-157890843	2,806,189	4.722	6.515
1	AX-157890870	2,856,807	4.595	6.323
1	AX-157890873	2,862,213	4.595	6.323
1	AX-157716740	2,864,335	4.595	6.323
1	AX-157746862	2,868,540	4.595	6.323
1	AX-157890876	2,875,537	4.595	6.323
1	AX-157725146	2,879,104	4.595	6.323
1	AX-157890878	2,880,599	4.595	6.323
1	AX-157890879	2,881,191	4.595	6.323
1	AX-157753267	2,937,890	4.191	5.712

1	AX-157890916	2,954,436	4.713	6.501
1	AX-157890917	2,954,740	4.631	6.378
1	AX-157868452	2,955,711	4.595	6.323
1	AX-157890920	2,962,210	4.308	5.888
1	AX-157890921	2,964,251	4.631	6.378
1	AX-158404158	2,990,743	4.595	6.323
1	AX-157890943	3,007,754	4.641	6.392
1	AX-157722016	3,010,441	4.227	5.767
1	AX-157890964	3,043,053	4.525	6.217
1	AX-157890974	3,076,547	4.602	6.333
1	AX-157890982	3,082,775	4.595	6.323
1	AX-157890992	3,099,206	4.347	5.948
1	AX-157891007	3,118,750	4.164	5.671
1	AX-158404161	3,127,517	4.253	5.806
1	AX-157891042	3,160,193	4.568	6.282
1	AX-157891043	3,160,911	4.359	5.966
1	AX-157891517	4,032,463	4.328	5.919
1	AX-157891542	4,055,100	4.483	6.153
1	AX-157739789	4,117,146	4.945	6.852
1	AX-157891632	4,157,214	5.007	6.945
1	AX-157891660	4,196,020	4.323	5.912
1	AX-157891666	4,211,924	4.304	5.883
1	AX-157891673	4,235,944	4.593	6.320
1	AX-157891677	4,239,109	4.347	5.948
1	AX-157891692	4,261,955	4.375	5.990
1	AX-157891696	4,269,860	4.246	5.795
1	AX-157891769	4,370,503	4.598	6.328
1	AX-158388964	4,375,085	4.208	5.737
1	AX-157891779	4,386,069	4.216	5.750
1	AX-157740292	4,386,505	4.180	5.695
1	AX-157704920	4,452,018	4.180	5.695
1	AX-157810329	4,453,190	4.180	5.695
1	AX-157891830	4,492,170	4.180	5.695
1	AX-157891831	4,492,526	4.179	5.694
1	AX-158403712	4,501,600	4.197	5.721
1	AX-157891841	4,505,633	4.180	5.695
1	AX-158379102	4,525,525	4.180	5.695
1	AX-157874258	4,542,570	4.299	5.876
1	AX-157891911	4,596,950	4.266	5.825

114

1	AX-157772233	4,624,147	4.329	5.920
1	AX-157810348	4,626,297	4.358	5.965
1	AX-157891932	4,633,753	4.180	5.695
1	AX-157891947	4,658,024	4.180	5.695
1	AX-157891956	4,668,144	4.180	5.695
1	AX-157891960	4,678,714	4.216	5.750
1	AX-157891971	4,690,940	4.180	5.695
1	AX-157891985	4,708,007	4.299	5.875
1	AX-157891998	4,736,545	4.464	6.126
1	AX-157892017	4,773,687	4.245	5.793
1	AX-157892035	4,799,360	4.320	5.907
1	AX-157892040	4,810,434	4.261	5.818
1	AX-157892454	4,948,950	4.220	5.755
1	AX-157892475	5,000,192	4.180	5.695
1	AX-157868472	5,048,064	4.180	5.695
1	AX-157892498	5,062,120	4.216	5.750
1	AX-157810370	5,143,052	4.252	5.805
1	AX-157892103	5,225,218	4.183	5.700
1	AX-157892113	5,238,609	4.749	6.556
1	AX-157892119	5,245,866	4.180	5.695
1	AX-157892123	5,256,116	4.139	5.632
1	AX-157892127	5,275,052	4.180	5.695
1	AX-157892130	5,284,586	4.267	5.827
1	AX-157892133	5,302,390	4.180	5.695
1	AX-157892136	5,303,611	4.180	5.695
1	AX-157892137	5,307,910	4.436	6.083
1	AX-157892146	5,353,131	4.180	5.695
1	AX-157892194	5,529,957	4.180	5.695
1	AX-158388969	5,566,338	4.180	5.695
1	AX-157892247	5,634,939	4.180	5.695
1	AX-157892296	5,735,164	4.377	5.994
1	AX-157892297	5,737,401	4.218	5.753
1	AX-157810389	5,757,791	4.481	6.151
1	AX-157810393	5,770,195	4.406	6.037
1	AX-157892320	5,783,125	4.420	6.058
1	AX-157892327	5,805,972	4.377	5.994
1	AX-157892341	5,825,180	4.420	6.058
1	AX-157892345	5,827,424	4.327	5.918
1	AX-157722408	5,830,364	4.418	6.056

1	AX-157892355	5,847,123	4.377	5.994
1	AX-157892356	5,847,408	4.809	6.647
1	AX-157892359	5,849,302	4.260	5.816
1	AX-157892369	5,885,934	4.377	5.994
1	AX-157892379	5,905,531	4.826	6.672
1	AX-157750570	5,917,382	5.321	7.418
1	AX-157892387	5,920,628	5.372	7.495
1	AX-157892392	5,936,494	4.623	6.365
1	AX-157892393	5,937,445	4.234	5.777
1	AX-157892410	5,964,236	5.376	7.500
1	AX-157892412	5,968,858	5.200	7.235
1	AX-157892419	5,975,507	4.792	6.620
1	AX-157810408	5,982,540	4.917	6.810
1	AX-157892425	5,987,345	4.218	5.753
1	AX-157892429	5,994,676	5.250	7.310
1	AX-157743672	5,996,106	4.508	6.192
1	AX-157868473	6,025,913	5.164	7.181
1	AX-157892523	6,027,580	5.345	7.454
1	AX-157892530	6,035,421	5.036	6.988
1	AX-157892551	6,067,138	4.381	5.999
1	AX-157892553	6,069,479	4.281	5.847
1	AX-157892557	6,077,909	5.068	7.036
1	AX-158404231	6,079,136	5.090	7.069
1	AX-157892560	6,082,966	5.002	6.938
1	AX-157892561	6,083,925	5.336	7.440
1	AX-157892567	6,088,104	4.740	6.543
1	AX-157892580	6,108,746	4.732	6.530
1	AX-157892585	6,126,333	4.885	6.760
1	AX-157892586	6,128,189	4.944	6.850
1	AX-157892590	6,139,795	4.835	6.685
1	AX-157892594	6,146,863	4.885	6.760
1	AX-157892596	6,149,310	4.493	6.169
1	AX-157892598	6,157,138	4.819	6.661
1	AX-157759008	6,169,820	5.201	7.238
1	AX-157700592	6,171,403	4.821	6.664
1	AX-157892613	6,183,481	4.540	6.239
1	AX-157892615	6,185,934	4.363	5.973
1	AX-157892623	6,203,765	4.542	6.243
1	AX-157892630	6,208,531	4.777	6.598

116

1	AX-157892635	6,214,827	4.540	6.239
1	AX-157892636	6,215,252	4.185	5.702
1	AX-157892637	6,216,934	4.463	6.124
1	AX-157810440	6,224,545	4.940	6.845
1	AX-157892645	6,232,938	4.663	6.425
1	AX-157892648	6,236,394	4.790	6.618
1	AX-157892653	6,241,370	4.622	6.364
1	AX-158379105	6,241,835	4.293	5.866
1	AX-157892662	6,254,713	4.815	6.655
1	AX-157892664	6,257,912	4.562	6.273
1	AX-157892665	6,263,637	4.663	6.425
1	AX-157892667	6,265,452	4.667	6.432
1	AX-157892678	6,280,417	4.238	5.782
1	AX-157892718	6,360,464	4.824	6.669
1	AX-157892719	6,360,964	4.663	6.425
1	AX-157810444	6,405,994	4.663	6.425
1	AX-157892737	6,407,401	4.800	6.633
1	AX-157892738	6,410,167	4.694	6.473
1	AX-157892745	6,431,901	4.701	6.482
1	AX-157892781	6,519,219	4.206	5.735
1	AX-157810452	6,568,369	4.206	5.735
1	AX-157892875	6,736,269	4.219	5.753
1	AX-157892966	6,903,348	4.460	6.118
1	AX-157892967	6,904,382	4.344	5.943
1	AX-157892984	6,932,195	4.209	5.739
1	AX-158388975	6,938,366	4.200	5.725
1	AX-157810481	6,948,900	4.157	5.660
1	AX-157892995	6,951,619	4.200	5.725
1	AX-157766872	6,953,340	4.199	5.724
1	AX-157740702	6,953,741	4.247	5.796
1	AX-157893630	7,093,792	4.200	5.725
1	AX-157893636	7,104,050	4.360	5.968
1	AX-157893641	7,112,685	4.480	6.149
1	AX-158424012	7,396,758	4.144	5.641
1	AX-157770640	7,437,825	4.353	5.957
1	AX-157893110	7,571,812	4.225	5.763
1	AX-157893144	7,616,042	4.370	5.983
1	AX-158388979	7,920,751	4.877	6.749
1	AX-157810525	7,987,728	4.424	6.064

1	AX-157893411	8,017,568	4.269	5.830
1	AX-157893437	8,044,521	4.370	5.983
1	AX-157893485	8,130,275	4.245	5.794
1	AX-157893498	8,152,379	4.371	5.984
1	AX-157874276	8,200,474	4.371	5.984
1	AX-157893540	8,232,604	4.371	5.984
1	AX-157893543	8,241,007	4.371	5.984
1	AX-157893551	8,250,153	4.280	5.847
1	AX-158403772	8,266,229	4.586	6.309
1	AX-157893728	8,563,810	4.466	6.128
1	AX-157893730	8,577,992	4.993	6.924
1	AX-157893780	8,728,850	4.313	5.897
1	AX-157893781	8,729,943	4.586	6.309
1	AX-157893783	8,731,056	4.532	6.227
1	AX-158404278	8,743,042	4.366	5.976
1	AX-157810579	8,809,653	5.020	6.965
1	AX-157893866	8,875,179	4.780	6.602
1	AX-157893867	8,875,543	4.797	6.628
1	AX-157893873	8,884,990	5.216	7.260
1	AX-157893887	8,904,683	5.160	7.175
1	AX-157893916	8,937,759	5.158	7.172
1	AX-157893924	8,945,889	4.489	6.162
1	AX-157893933	8,956,008	4.671	6.438
1	AX-158379114	8,970,027	4.684	6.458
1	AX-157893936	8,970,241	5.295	7.379
1	AX-157893953	8,995,316	4.441	6.090
1	AX-157893955	8,998,726	4.858	6.720
1	AX-158388986	9,001,934	4.345	5.945
1	AX-157893961	9,008,219	4.507	6.191
1	AX-157893974	9,024,967	4.501	6.180
1	AX-157893981	9,033,183	5.286	7.364
1	AX-157741250	9,040,548	5.283	7.361
1	AX-157893986	9,041,957	4.366	5.977
1	AX-157810605	9,079,528	4.171	5.682
1	AX-157868487	9,112,319	5.025	6.972
1	AX-157894025	9,126,760	5.046	7.003
1	AX-157894034	9,137,472	4.342	5.940
1	AX-157894035	9,139,298	4.818	6.660
1	AX-157894045	9,155,568	4.607	6.340

118

1	AX-157810627	9,163,300	4.190	5.711
1	AX-157894061	9,172,951	5.043	7.000
1	AX-157756029	9,175,318	4.272	5.835
1	AX-157894070	9,200,806	5.015	6.957
1	AX-157894087	9,241,165	4.990	6.920
1	AX-157894095	9,263,500	5.098	7.082
1	AX-157758864	9,311,669	4.468	6.131
1	AX-157894121	9,313,743	5.321	7.418
1	AX-157894124	9,322,766	5.231	7.282
1	AX-157810643	9,326,922	5.370	7.491
1	AX-157894127	9,336,325	5.204	7.242
1	AX-157810647	9,337,369	5.173	7.196
1	AX-157744060	9,352,103	4.273	5.836
1	AX-157894143	9,372,487	5.250	7.311
1	AX-157810651	9,373,879	5.069	7.038
1	AX-157894156	9,400,003	4.196	5.719
1	AX-157894161	9,414,290	4.963	6.878
1	AX-157894177	9,437,054	4.707	6.492
1	AX-157894180	9,442,859	4.521	6.210
1	AX-157894181	9,445,249	4.248	5.799
1	AX-157894182	9,445,657	4.707	6.492
1	AX-157894195	9,462,603	4.246	5.794
1	AX-157894205	9,471,883	4.729	6.525
1	AX-157894211	9,480,424	4.905	6.791
1	AX-157810663	9,489,029	4.905	6.791
1	AX-157894240	9,530,496	4.808	6.644
1	AX-157894244	9,540,814	4.678	6.448
1	AX-157894248	9,546,290	5.271	7.343
1	AX-157894252	9,548,326	5.332	7.434
1	AX-157894257	9,554,816	5.416	7.560
1	AX-157862589	9,602,028	4.663	6.425
1	AX-157810675	9,620,337	5.167	7.186
1	AX-157894292	9,623,728	4.135	5.627
1	AX-157894303	9,643,676	4.249	5.800
1	AX-157894319	9,672,595	5.000	6.934
1	AX-157894336	9,705,852	5.329	7.430
1	AX-157894355	9,745,009	4.881	6.754
1	AX-157894417	9,838,636	5.097	7.081
1	AX-157894442	9,872,774	5.078	7.051

1	AX-157894443	9,874,583	5.078	7.051
1	AX-157894455	9,893,033	5.417	7.562
3	AX-157813250	802,388	4.619	6.359
3	AX-157920976	803,822	4.702	6.485
3	AX-157920982	813,720	4.670	6.437
3	AX-157921003	844,037	4.552	6.258
3	AX-157921043	952,424	4.419	6.056
3	AX-157921093	1,039,773	4.500	6.179
3	AX-157853794	1,181,901	4.488	6.161
3	AX-157921196	1,202,819	4.466	6.128
3	AX-157921206	1,215,683	4.794	6.623
21	AX-158255403	3,278,086	4.317	5.903
21	AX-158255412	3,284,725	4.279	5.845
26	AX-158328170	1,094,521	4.154	5.656
26	AX-158328297	1,294,513	4.633	6.381
26	AX-157848875	1,296,669	4.412	6.046
26	AX-158328319	1,328,771	4.191	5.712
26	AX-157747179	1,342,000	4.323	5.912
26	AX-158328344	1,361,682	4.191	5.712
26	AX-157806292	1,463,866	4.159	5.663
26	AX-158328507	1,630,492	4.644	6.397
26	AX-158328515	1,639,485	5.436	7.590
26	AX-158328525	1,648,547	5.870	8.242
26	AX-158328527	1,650,559	4.655	6.414
26	AX-158328540	1,666,617	5.479	7.655
26	AX-158328562	1,689,779	5.038	6.991
26	AX-158328566	1,702,093	5.009	6.947
26	AX-158328571	1,719,727	5.042	6.998
26	AX-158328594	1,759,026	5.080	7.055
26	AX-158328605	1,786,600	5.093	7.075
26	AX-158328607	1,790,225	4.993	6.924
26	AX-158328611	1,793,819	4.740	6.542
26	AX-158328651	1,871,693	4.993	6.924
26	AX-158328732	1,986,046	4.227	5.766
26	AX-158328753	2,026,741	4.407	6.039
26	AX-158328761	2,039,622	4.365	5.976
26	AX-158328767	2,048,266	4.648	6.402
26	AX-158328793	2,095,700	4.556	6.264
26	AX-158328837	2,162,642	4.365	5.976

26	AX-158328881	2,225,285	4.365	5.976
26	AX-158328883	2,232,748	4.415	6.051
26	AX-158328887	2,236,947	4.415	6.051
26	AX-158381520	2,246,812	4.365	5.976
26	AX-158328930	2,313,882	4.351	5.955

Table S2

Linkage Group	gene	start position	end position	gene symbol	gene description
1	LOC108268556	5,264,773	5,552,400	dnah5	dynein heavy chain 5, axonemal-like
1	kiaa1217	5,564,583	5,670,495	kiaa1217	kiaa1217
1	LOC108265087	5,676,993	5,681,784	gpr17	uracil nucleotide/cysteinyl leukotriene receptor-like
1	nphp3	5,686,594	5,724,617	nphp3	nephrocystin 3
1	uba5	5,728,209	5,738,277	uba5	ubiquitin like modifier activating enzyme 5
1	LOC108265118	5,746,030	5,768,441	sh3glb1	endophilin-B1-like
1	nrros	5,772,119	5,775,983	nrros	negative regulator of reactive oxygen species
1	fbxo45	5,780,418	5,789,822	fbxo45	F-box protein 45
1	c1h8orf59	5,792,038	5,793,560	c1h8orf59	chromosome 1 open reading frame
1	ca2	5,797,658	5,802,800	ca2	carbonic anhydrase 2
1	LOC108272680	5,848,956	5,858,497	tomm40	mitochondrial import receptor subunit TOM40 homolog
1	nsmce2	5,860,644	5,869,522	nsmce2	NSE2/MMS21 homolog, SMC5-SMC6 complex SUMO ligase
1	derl1	5,870,674	5,882,259	derl1	derlin 1
1	LOC108265168	5,875,776	5,878,002	pseudogene	
1	zhx2	5,883,500	5,902,236	zhx2	zinc fingers and homeoboxes 2
1	LOC108278704	5,915,808	5,926,327	grina	protein lifeguard 1-like
1	LOC108271836	5,929,449	5,935,322	fam8a1	protein FAM8A1-like
1	LOC108272177	5,951,310	5,966,079	atxn1l	ataxin-1-like
1	LOC108272890	5,971,031	5,980,818	mylip	E3 ubiquitin-protein ligase MYLIP-B-like
1	dtnbp1	5,981,111	6,005,131	dtnbp1	dystrobrevin binding protein 1
1	LOC108265176	6,009,088	6,078,846	jarid2	protein Jumonji-like
1	pard6g	6,105,366	6,108,232	pard6g	par-6 family cell polarity regulator gamma
1	LOC108265191	6,110,926	6,116,744	adnp2	ADNP homeobox protein 2-like
1	rbfa	6,123,104	6,127,640	rbfa	ribosome binding factor A (putative)
1	zfat	6,133,444	6,148,557	zfat	zinc finger and AT-hook domain containing
1	LOC108267829	6,151,786	6,153,696	ncRNA	
1	LOC108267831	6,160,706	6,179,014	dync1i1	cytoplasmic dynein 1 intermediate chain 1-like
1	tbc1d31	6,180,154	6,203,778	tbc1d31	TBC1 domain family member 31
1	LOC108274019	6,205,912	6,211,728	uncharacterized	
1	prelid3a	6,211,829	6,220,849	prelid3a	PRELI domain containing 3A

122

1	fbxo32	6,225,044	6,231,302	fbxo32	F-box protein 32
1	klhl38	6,233,511	6,238,886	klhl38	kelch like family member 38
1	LOC108262601	6,247,851	6,249,875	ncRNA	
1	LOC108270999	6,252,231	6,254,017	fam84b	protein FAM84B-like
1	LOC100305014	6,283,447	6,286,160	dap12	DAP12
1	LOC100304692	6,291,315	6,294,257	dap10	DAP10
1	LOC108269301	6,297,252	6,305,994	zfp37	zinc finger protein 37-like
1	LOC108268204	6,311,803	6,320,695	znf84	zinc finger protein 84-like
1	LOC108270015	6,323,096	6,338,113	uncharacterized	
1	LOC108265257	6,342,791	6,346,924	prss1	trypsin-like
1	LOC108265242	6,350,345	6,359,826	znf91	zinc finger protein 91-like
1	LOC108265300	6,364,840	6,371,591	uncharacterized	
1	LOC108265306	6,383,574	6,394,997	uncharacterized	
1	LOC108265278	6,408,486	6,419,193	znf629	zinc finger protein 629-like
1	LOC108265342	6,434,285	6,442,476	uncharacterized	
1	LOC108265385	6,443,805	6,445,790	prss3	trypsin-3-like
1	LOC108265295	6,447,203	6,456,786	znf184	zinc finger protein 184-like
1	LOC108265350	6,460,165	6,468,883	znf502	zinc finger protein 502-like
1	LOC108265289	6,494,474	6,515,545	znf423	zinc finger protein 423-like
1	LOC108265267	6,550,647	6,579,534	uncharacterized	
1	pbxip1	8,768,899	8,784,556	pbxip1	PBX homeobox interacting protein 1
1	LOC108265907	8,779,292	8,811,038	s100a6	protein S100-A16-like
1	LOC108265814	8,785,178	8,794,490	tlr13	toll-like receptor 13
1	LOC108265897	8,811,123	8,813,452	s100a1	protein S100-A1-like
1	chrnb2	8,816,421	8,848,069	chrnb2	cholinergic receptor nicotinic beta 2 subunit
1	she	8,871,744	8,882,966	she	Src homology 2 domain containing E
1	cgn	8,900,845	8,942,282	cgn	cingulin
1	LOC108265940	8,950,323	8,965,842	tuft1	tuftelin-like
1	c1h1orf43	8,967,835	8,978,749	c1h1orf43	chromosome 1 open reading frame
1	LOC108265956	8,982,464	9,002,225	uncharacterized	
1	LOC108265981	9,004,539	9,014,513	aqp10	aquaporin-10-like
1	atp8b2	9,016,012	9,072,202	atp8b2	ATPase phospholipid transporting 8B2
1	LOC108266000	9,072,923	9,088,727	il6r	interleukin-6 receptor subunit alpha-like
1	LOC108266011	9,102,955	9,119,917	il6r	interleukin-6 receptor subunit alpha-like
1	LOC108266079	9,149,121	9,225,288	uncharacterized	
1	LOC108266083	9,152,258	9,159,695	uncharacterized	

1	LOC108266026	9,160,301	9,188,822	uncharacterized	
1	LOC108266018	9,189,077	9,195,519	uncharacterized	
1	riiad1	9,231,026	9,242,915	riiad1	regulatory subunit of type II PKA R-subunit (RIIa) domain containing 1
1	LOC108266121	9,247,464	9,272,438	otoa	otoancorin-like
1	znf687	9,262,070	9,289,250	znf687	zinc finger protein 687
1	setdb1	9,296,100	9,316,851	setdb1	SET domain bifurcated 1
1	trim39	9,321,484	9,326,719	trim39	tripartite motif containing 39
1	LOC108266148	9,327,661	9,332,683	clec4e	C-type lectin domain family 4 member E-like
1	cops7a	9,334,160	9,352,174	cops7a	COP9 signalosome subunit 7A
1	LOC108266188	9,353,205	9,361,176	uncharacterized	
1	LOC108266172	9,361,593	9,384,552	fbxw7	F-box/WD repeat-containing protein 7-like
1	znf384	9,387,293	9,411,118	znf384	zinc finger protein 384
1	gnl1	9,412,759	9,425,419	gnl1	G protein nucleolar 1
1	LOC108268602	9,432,679	9,435,736	uncharacterized	
1	ddr1	9,438,568	9,484,797	ddr1	discoidin domain receptor tyrosine kinase 1
1	flot1	9,489,761	9,499,921	flot1	flotillin 1
1	tubb	9,508,358	9,514,068	tubb	tubulin beta class I
1	mdc1	9,517,203	9,540,224	mdc1	mediator of DNA damage checkpoint 1
1	LOC108266289	9,540,746	9,560,067	uncharacterized	
1	LOC108266309	9,564,823	9,569,529	znf33a	zinc-binding protein A33-like
1	ino80e	9,572,751	9,582,072	ino80e	INO80 complex subunit E
1	LOC108266323	9,582,962	9,587,151	uncharacterized	
1	vars	9,594,608	9,624,192	vars	valyl-tRNA synthetase
1	abhd16a	9,624,311	9,638,737	abhd16a	abhydrolase domain containing 16A
1	LOC108274489	9,644,664	9,665,090	uncharacterized	
1	LOC108266341	9,666,386	9,741,026	nectin2	nectin-2-like
1	LOC108261561	9,749,072	9,759,611	uncharacterized	
1	LOC108266346	9,781,718	9,803,377	atp1a3	sodium/potassium-transporting ATPase subunit alpha-3
1	rabac1	9,804,761	9,814,334	rabac1	Rab acceptor 1
1	dedd2	9,813,036	9,831,203	dedd2	death effector domain containing 2
1	pou2f2	9,858,168	9,906,821	pou2f2	POU class 2 homeobox 2
1	erf	9,923,076	9,970,194	erf	ETS2 repressor factor
1	LOC108266385	9,970,575	9,971,912	uncharacterized	
1	cic	9,982,904	10,039,968	cic	capicua transcriptional repressor
1	znf574	10,042,572	10,051,478	znf574	zinc finger protein 574

1	gsk3a	10,052,166	10,077,483	gsk3a	glycogen synthase kinase 3 alpha
1	LOC108266406	10,080,348	10,106,045	znf791	zinc finger protein 791-like
1	grik5	10,115,568	10,190,932	grik5	glutamate ionotropic receptor kainate type subunit 5
26	LOC108255854	1,125,804	1,141,744	erlin2	erlin-2-like
26	LOC108255937	1,157,207	1,162,254	znf703	zinc finger protein 703-like
26	LOC108255938	1,209,690	1,212,947	uncharacterized	
26	LOC108255939	1,227,221	1,231,365	uncharacterized	
26	LOC108255456	1,255,122	1,274,340	uncharacterized	
26	tia1	1,294,064	1,304,279	tia1	TIA1 cytotoxic granule associated RNA binding protein
26	dgcr8	1,306,398	1,316,196	dgcr8	DGCR8, microprocessor complex subunit
26	trmt2a	1,315,834	1,328,097	trmt2a	tRNA methyltransferase 2 homolog A
26	atp6v0a2	1,328,968	1,347,194	atp6v0a2	ATPase H+ transporting V0 subunit a2
26	dnah10	1,353,843	1,411,253	dnah10	dynein axonemal heavy chain 10
26	nefl	1,437,430	1,445,488	nefl	neurofilament light
26	mxd1	1,463,189	1,491,945	mxd1	MAX dimerization protein 1
26	snrnp27	1,492,838	1,498,644	snrnp27	small nuclear ribonucleoprotein U4/U6.U5 subunit 27
26	add2	1,500,592	1,522,465	add2	adducin 2
26	figla	1,523,093	1,528,335	figla	folliculogenesis specific bHLH transcription factor
26	LOC108255852	1,536,348	1,573,256	hk2	hexokinase-2-like
26	LOC108255451	1,640,808	1,665,581	npffr1	neuropeptide FF receptor 1-like
26	tet3	1,683,625	1,717,675	tet3	tet methylcytosine dioxygenase 3
26	LOC108255511	1,736,938	1,744,573	ap3m2	AP-3 complex subunit mu-2
26	kat6a	1,748,354	1,780,921	kat6a	lysine acetyltransferase 6A
26	cdca2	1,798,269	1,813,571	cdca2	cell division cycle associated 2
26	LOC108255858	1,821,530	1,828,052	uncharacterized	
26	LOC108255859	1,828,101	1,894,623	sema4c	semaphorin-4C-like
26	tbx1	2,179,613	2,193,660	tbx1	T-box 1

Table S3

linkage group	gene	start position	end position	gene description
1	obscn	1,777,901	1,836,336	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF
1	dtna	1,900,705	1,921,802	dystrobrevin alpha
1	spidr	1,922,786	1,948,506	scaffolding protein involved in DNA repair
1	sntg1	2,003,673	2,046,338	syntrophin gamma 1
1	st18	2,066,584	2,110,585	ST18, C2H2C-type zinc finger
1	rb1cc1	2,142,472	2,161,249	RB1 inducible coiled-coil 1
1	agap3	2,174,017	2,302,655	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3
1	gbx1	2,310,330	2,312,584	gastrulation brain homeobox 1
1	asb10	2,315,428	2,325,921	ankyrin repeat and SOCS box containing 10
1	scnn1a	2,346,717	2,416,441	sodium channel epithelial 1 alpha subunit
1	abcb8	2,425,128	2,434,711	ATP binding cassette subfamily B member 8
1	LOC108264077	2,437,826	2,446,823	myosin light chain kinase, smooth muscle-like
1	fmnl1	2,785,248	2,814,235	formin like 1
1	LOC108264371	2,847,171	2,855,141	C-C chemokine receptor type 7-like
1	LOC108264380	2,862,475	2,869,480	receptor activity-modifying protein 1-like
1	LOC108273238	2,874,212	2,879,045	endonuclease domain-containing 1 protein-like
1	LOC108273210	2,919,873	2,944,940	corticotropin-releasing factor receptor 1-like
1	LOC108269775	2,947,367	2,964,107	carnitine O-palmitoyltransferase 1, liver isoform-like
1	tsr3	2,989,497	2,991,987	TSR3, acp transferase ribosome maturation factor
1	LOC108264462	3,007,539	3,019,440	BTB/POZ domain-containing protein KCTD5-like
1	cog7	3,045,715	3,053,506	component of oligomeric golgi complex 7
1	LOC108264490	3,060,414	3,122,660	potassium voltage-gated channel subfamily C member 1-like
1	nlrc3	3,172,940	3,192,223	NLR family CARD domain containing 3
1	dnah5	4,029,552	4,135,385	dynein axonemal heavy chain 5
1	LOC108272199	4,137,980	4,333,977	adenylate cyclase type 2-like
1	LOC108278538	4,357,648	4,374,758	non-canonical poly(A) RNA polymerase PAPD7-like
1	prpf3	4,380,752	4,392,359	pre-mRNA processing factor 3
1	ice1	4,435,517	4,454,032	interactor of little elongation complex ELL subunit 1
1	taf2	4,483,326	4,499,036	TATA-box binding protein associated factor 2
1	enpp2	4,501,895	4,527,578	ectonucleotide pyrophosphatase/phosphodiesterase 2
1	ciart	4,530,352	4,544,442	circadian associated repressor of transcription

1	ipo4	4,596,047	4,609,815	importin 4
1	tm9sf1	4,625,132	4,632,513	transmembrane 9 superfamily member 1
1	fen1	4,633,282	4,643,716	flap structure-specific endonuclease 1
1	esco1	4,648,090	4,670,422	establishment of sister chromatid cohesion N-acetyltransferase 1
1	rbbp8	4,670,696	4,695,304	RB binding protein 8, endonuclease
1	tmem241	4,701,099	4,715,794	transmembrane protein 241
1	LOC108255457	4,740,905	4,761,927	receptor-type tyrosine-protein phosphatase eta-like
1	LOC108260301	4,779,287	4,790,561	probable E3 ubiquitin-protein ligase DTX2
1	c1h18orf8	4,908,611	4,971,565	chromosome 1 open reading frame
1	otud1	5,017,296	5,019,369	OTU deubiquitinase 1
1	LOC108264971	5,064,322	5,121,286	NACHT, LRR and PYD domains-containing protein 12-like
1	LOC108272748	5,141,375	5,241,305	tomoregulin-1-like
1	LOC108260814	5,244,068	5,246,974	uncharacterized
1	LOC108276803	5,254,902	5,263,540	muscle-related coiled-coil protein-like
1	LOC108268556	5,264,773	5,552,400	dynein heavy chain 5, axonemal-like
3	armc5	802,839	808,753	armadillo repeat containing 5
3	LOC108280679	855,132	861,431	perforin-1-like
3	LOC108280773	963,009	984,385	coronin-1A-like
3	shank1	1,049,690	1,162,472	SH3 and multiple ankyrin repeat domains 1
3	LOC108274951	1,174,143	1,194,440	protein kinase C and casein kinase substrate in neurons protein 3-like
3	syt3	1,228,599	1,280,652	synaptotagmin 3
21	magi1	3,159,835	3,304,759	membrane associated guanylate kinase, WW and PDZ domain containing 1

Table S4

SNP	Linkage Group	P value	-LOG10(P value)	Position
AX-85418798	1	3.29E-07	6.483	32,497,896
AX-85354203	1	8.87E-07	6.052	32,328,310
AX-85381538	1	2.44E-06	5.612	32,054,212
AX-85308810	1	2.70E-06	5.569	32,340,053
AX-85190584	1	7.49E-06	5.126	32,150,752
AX-86104274	1	8.65E-06	5.063	32,054,212
AX-86063681	1	8.67E-06	5.062	31,708,415
AX-85219421	1	1.02E-05	4.993	32,683,071
AX-85339437	1	1.89E-05	4.724	31,708,415
AX-85405156	1	2.98E-05	4.525	32,454,726
AX-85416217	1	2.98E-05	4.525	32,479,386
AX-85337921	1	4.19E-05	4.378	31,360,586
AX-85332161	1	4.31E-05	4.365	32,746,364
AX-85190829	1	5.89E-05	4.230	31,426,189
AX-85198504	1	6.06E-05	4.217	31,386,020
AX-85254477	1	7.06E-05	4.151	32,250,553
AX-85922785	1	7.06E-05	4.151	32,293,603
AX-85294217	1	9.39E-05	4.027	32,879,259
AX-85335301	1	9.39E-05	4.027	32,880,604
AX-85292899	1	0.000101893	3.992	32,091,139
AX-85351527	1	0.000108229	3.966	32,316,169
AX-85440513	1	0.000110464	3.957	30,063,971
AX-85342133	1	0.00011967	3.922	31,128,023
AX-86046706	1	0.000120462	3.919	32,433,181

AX-85338610	1	0.000122594	3.912	12,568,441
AX-85440683	1	0.00013696	3.863	18,570,884
AX-85278345	1	0.000149682	3.825	18,571,315
AX-85306395	1	0.000152663	3.816	12,384,762
AX-85319338	1	0.000153741	3.813	12,644,661
AX-86003594	1	0.000153982	3.813	9,947,945
AX-86026760	12	2.69E-05	4.570	23,872,811
AX-85301000	12	3.40E-05	4.468	23,872,811
AX-85919596	12	5.48E-05	4.261	23,858,646
AX-86042926	12	5.48E-05	4.261	23,910,486
AX-85307954	12	8.46E-05	4.073	23,996,409
AX-86008454	12	9.44E-05	4.025	23,988,850
AX-85265433	12	0.000109914	3.959	23,977,185
AX-85293337	12	0.000155091	3.809	24,067,171
AX-85374530	16	3.04E-05	4.517	1,793,012
AX-85310557	16	7.57E-05	4.121	2,527,682
AX-86035941	16	7.57E-05	4.121	2,527,682
AX-85251687	16	9.52E-05	4.021	2,545,850
AX-85256902	16	0.000103919	3.983	2,004,928
AX-85319262	16	0.000127608	3.894	2,508,863
AX-85367817	16	0.000159554	3.797	2,580,246
AX-85339272	16	0.000159857	3.796	1,819,316

Table S5

Merged Region	Chromosome	invovled gene
990	1	agtr1
14490	26	fga
13976	25	cenpj
13976	25	plg
11013	20	proc
14487	26	fgb
1804	10	abcb11
22483	8	mls1
20856	7	cfhr1
16808	29	fgg
21547	7	c8a
23392	9	ppp1r3c
21656	8	etnk2
2069	11	uroc1
1744	10	tdo2
7412	17	seprinf2
8512	18	vtn
10622	20	cld3
7849	18	vmo1
11012	20	proc
12	1	c1rl
2154	11	angptl3
2483	11	inhbb
8526	18	serpinf2
1247	10	fgf19
14746	26	wap65-2
22633	9	dio2
16879	3	glra3
16250	28	ugt2c1
4584	14	enpp2
953	1	ppp1r3g
10637	20	apoa1
18932	5	csl3
12551	23	c8g
14861	26	alg11
14861	26	cpb2
14651	26	proz

130

20376	6	erbb4
15966	28	c9
13330	24	igfals
20910	7	angptl3
1860	10	c4bpa
11086	20	tf
18297	4	ttc36
23401	9	minpp1
23158	9	serpina10
3678	13	shbg
1648	10	ahsg
4286	13	arg1
21504	7	igfbp1
6612	16	kl
9302	2	abcc1
4637	14	hal
11360	21	cda
17729	3	habp2
13029	23	hpn
502	1	c4
14906	26	gyg2
15654	27	nb0b1
3806	13	gch1
7069	17	tsku
6307	16	msmb
20609	7	cst3
4341	13	prodh2
19094	5	adra1a
22537	8	adh5
15446	27	f2
20864	7	ahsg
22057	8	f9
11207	21	itih3
7511	17	a2m
15599	27	c3
20763	7	ebi3
14655	26	f7
12146	22	c5
1474	10	serpinc1
13978	25	slc22a2
9112	19	gys2

7144	17	agxt
1197	10	hp
15653	27	nr0b1
9094	19	pah
16088	28	hpd
23242	9	smoc1
14949	26	rnaseh2a
18848	5	ntrk2
20347	6	mbl2
15066	26	c3
3786	13	itln2
7970	18	dmgdh
12102	22	ptgds
6309	16	msmb
551	1	hfe2
1735	10	mcoln2
11957	22	spink2
19799	6	upp2
19671	5	serping1
21509	7	cpn2
7516	17	a2m
15600	27	c3
8008	18	agxt2
2844	12	tmprss6
311	1	apoa4
18422	4	tat
8694	19	lgr5
15837	28	cyp2d3
4850	14	cpne8
8800	19	itih2
2289	11	agxt
3949	13	habp2
18423	4	tat
7026	17	cyp7a1
17124	3	pygl
17362	3	comtd1
8153	18	cfb
7218	17	cfb
7218	17	mfrp
18887	5	fgfr1
15095	26	hspg2

4297	13	fam20a
4267	13	sstr2
20666	7	ackr4
18905	5	serpind1
23124	9	plcb1
13860	25	serpina1
364	1	il6r
20475	6	wap65-1
5542	15	cfi
7520	17	a2m
21990	8	mtmr7
22718	9	prox1
7310	17	apom
20839	7	cbln1
12928	23	fabp1
20629	7	cys
2292	11	aln
10908	20	lpl
15093	26	hspg2
4992	14	pcdhac2
9670	2	hs3st6
9100	19	igf1
1643	10	cp
6661	16	sec14l2
1743	10	tdo2
19128	5	gramd3
10503	20	sepp1b
1661	10	uric
13518	24	tmem54
304	1	apoc1
11956	22	spink2
17970	4	raly
12879	23	oxr1
643	1	c1rb
1805	10	abcb11
17323	3	pnlip
8857	19	mfap4
11082	20	tf2
8187	18	adra1b
13622	24	hmcn2
1647	10	fetub

20021	6	kcnj3
645	1	masp2
15752	28	crim1
7232	17	mid1ip1
18553	4	onecut1
7991	18	onecut2
17243	3	ogdhl
23122	9	hao1
8390	18	c7
20075	6	f5
8051	18	slc22a6
19214	5	cfp
3141	12	nr0b2
5955	15	masp2
21989	8	slc7a2
8821	19	slco1c1
15246	27	irx3
14858	26	spp2
18853	5	adamts13
11270	21	ntn4
9413	2	coq10a
21052	7	etnppl
13513	24	klkb1
17286	3	sfrp5
7971	18	bhmt
12738	23	hgd
21232	7	znf638
14090	25	lrg1
21546	7	c8b
7956	18	daf36
12742	23	slc51a
19479	5	nr1h4
15109	26	mfap4
9021	19	tymp
7338	17	f10
7338	17	gmnc
12897	23	adam22
6185	16	ambp
7345	17	cldn1
18452	4	shank3
9637	2	hs3st3b1

22382	8	met
19826	6	tns1
15201	27	igdcc4
2554	11	mst1
7522	17	a2m
14654	26	fa7
2884	12	abcc1
20122	6	lrp1b
18193	4	cetp
21638	7	pseudo
14268	25	esr2a
4240	13	g6pc
10572	20	stx1a
4586	14	deptor
3958	13	adrb1
16989	3	hmx2
21841	8	citi1
21841	8	uvssa
9701	2	ldlr
10760	20	ephb1
15071	26	gcgr
15888	28	lrwd1
5318	15	styk1
11447	21	cyp27b1
7876	18	whrn
22585	9	clmn
3730	13	hs3st3b1
15377	27	lipc
11016	20	col10a1
11016	20	proc
1374	10	cers3
15470	27	tmem246
22300	8	syt9
14091	25	slc10a4
14209	25	ttr
11014	20	proc
8077	18	abhd15
18115	4	cdh13
19835	6	igfbp2
10520	20	inhba
6662	16	sec14l2

2791	12	apoh
15507	27	npb
18179	4	ces5a
1812	10	arrdc2
18911	5	crb1
7288	17	vtn
6456	16	psd3
5502	15	ppm1k
22127	8	glra1
8822	19	slc2a13
1437	10	kng1
4543	14	gpx3
12455	22	pseudo
4878	14	gse1
22341	8	cyp2d15
19282	5	gls2
23003	9	foxa2
19179	5	iqsec1
1918	10	nr5a2
6875	16	dhrs13
22232	8	spbc1861.05
4259	13	rbp4
17032	3	agt
20022	6	kcnj3
9709	2	nme1
11056	20	st3gal3
4096	13	bhlha15
23239	9	alk
23381	9	htra1a
2792	12	apoh
19666	5	qsox1
7911	18	rhobtb3
9710	2	nme1
12035	22	npffr2
14012	25	hey2
12695	23	liph
10694	20	gpr148
20628	7	cst3
5044	14	adra2a
11211	21	itih3
12994	23	plc

18366	4	nrn1l
18665	5	clc4f
22598	9	gch1
15335	27	igf2
14168	25	fmo5
10278	2	prox1
7687	17	fat3
17913	4	mob2
10142	2	kcnf1
48	1	masp2
3427	12	cyr61
7514	17	a2m
9388	2	svil
8858	19	mfap4
718	1	foxo6
6467	16	robo2
13636	24	hebp2
29	1	mal2
5555	15	tfr2
749	1	eva1b
19879	6	ptprn
15483	27	ftcd
1671	10	adgrl2
17797	3	fbln5
17810	3	c3h2orf50
6105	15	klf15
3785	13	gprc5c
2567	11	camkv
22343	8	cyp2d3
12635	23	urad
20132	6	tfcp2l1
8569	18	fgf1
9338	2	mylk
20245	6	dnajc3
22884	9	htr1b
18745	5	adra2b
3141	12	nr0b2
851	1	pnp
312	1	apoa4
11077	20	thpo
7541	17	apoa4

15653	27	nb0b1
10186	2	slc22a16
22483	8	glc
18905	5	serpind1
7412	17	serpinf2
6308	16	msmb

Table S6

Merged Region	Chromosome	invovled gene
7638	17	ca4
15197	27	calml4
11821	22	arhgap25
20508	6	gpr183
3867	13	slc16a3
14293	25	fermt1
1265	10	bco1
15228	27	nkd1
6310	16	dok3
21391	7	tchh
3686	13	trim39
6500	16	vasp
22811	9	inafm2
15068	26	dennd1b
21874	8	nox1
14254	25	prkch
11559	21	sraP
15159	27	plekhf1
19744	6	klhl9
21496	7	rnf138
5724	15	srgap3
12827	23	esrp1
10363	2	il20ra
5347	15	hoxc10
5752	15	arhgef3
5529	15	fam46a
16263	28	thap1
12193	22	gcnt1
20768	7	rgs21
13988	25	pln
4472	14	tgfbr2
22158	8	ccdc88a
2788	12	myadm
6351	16	rhof
11108	20	gpr4
17180	3	rnf122
3432	12	pkn2

1139	10	adgrg5
8550	18	gpr34
13364	24	laptm4a
21235	7	ovol1
508	1	rapgef5
6336	16	rassf2
5549	15	txk
12998	23	tmem176a
12998	23	rpz
6373	16	cntfr
18376	4	carmil2
2112	11	il12rb2
19760	6	aff3
6656	16	tpst2
14102	25	lrrc66
5450	15	sel1l3
19600	5	lpxn
19600	5	znf692
9580	2	hoxb7
8160	18	rasgrp2
9859	2	cyth2
9859	2	rac2
6776	16	mogat2
1958	11	krt5
10024	2	myadm
18728	5	arid5a
21489	7	foxf2
22138	8	htr4
18996	5	susd3
10389	2	themis
1809	10	il12rb2
22938	9	gpr132
1299	10	tmc6
1299	10	tmc8
18990	5	il17re
13220	24	fam83g
19982	6	hoxd11
7160	17	fxyd5
1166	10	drd4
22190	8	ckmt1b
11157	21	elf3

605	1	tnfaip8l2
916	1	slc12a7
10656	20	rin2
11947	22	entpd2
4320	13	arhgap27
14980	26	ifngr1
21693	8	spire2
13870	25	spint1
10896	20	nmrk2
5373	15	agap1
14945	26	apcs
20403	6	stat4
344	1	chrnb2
9372	2	ppl
6288	16	smtn
13275	24	card11
19612	5	crim1
4216	13	ank3
5523	15	reps2
5523	15	tmem27
15317	27	znf536
20770	7	rgs21
20306	6	cavin2
3007	12	ptpn1
5447	15	slc34a2
3096	12	s10aa
18943	5	plch2
1591	10	fgf22
9582	2	hoxb9
150	1	nlrc3
4822	14	h2
19636	5	mpnd
19203	5	was
8617	19	prr5
2226	11	dpt
8479	18	klf5
39	1	cd40
21912	8	xpnpep2
14563	26	radil
21913	8	sash3
5349	15	hoxc9

9863	2	grap2
12954	23	satb1
2858	12	cg1b
17040	3	rgs17
21782	8	p2ry10
3870	13	enpp7
20506	6	gpr183
6862	16	lgals9
14689	26	sh2d1a
9968	2	klb
8510	18	slc13a2
2536	11	pkp1
5747	15	tnfrsf1a
9606	2	anks4b
16689	29	fabp2
16689	29	klhl10
4985	14	dock2
9998	2	clnk
14600	26	atp1a1
530	1	hoxa10
15105	26	vav1
5228	15	chit
12119	22	emb
17424	3	fgfr3
21900	8	itk
19682	5	jak3
21417	7	prkd3
20029	6	dgkb
14709	26	hs3st3a1
14022	25	bcl11b
17033	3	phactr2
1272	10	rassf10
11425	21	chia
14660	26	cd247
22196	8	sh3gl3
23282	9	kcnk18
7223	17	gpr171
21824	8	rgcc
5833	15	arhgef19
256	1	dap10
21243	7	cldn15

23194	9	slc7a9
14445	25	ileu
19603	5	foxd2
5650	15	sla2
20220	6	slc15a2
16616	29	chs2
21771	8	sh21a
7706	17	cd3e
11544	21	efcc1
17078	3	grin2
5817	15	plekhn1
10774	20	gpr55
1192	10	b3galt1
14292	25	fermt1
10420	2	batf
5299	15	p2ry1
5279	15	erbb3
23360	9	lck
19979	6	hoxd9
532	1	hoxa9b
3715	13	st6galnac1
6510	16	slc43a2
2887	12	ccr7
1686	10	fyb
15493	27	krt19
3595	12	gimap8
18564	4	c4h15orf48
14535	26	cpo
20342	6	lcp1
23177	9	flvcr2
20484	6	satb2
11017	20	cld15
4621	14	cdhr5
9242	2	prf1
3637	13	trim16
18400	4	mgam
4035	13	gucy2c
1654	10	b3g5a
690	1	acp5
12585	23	xkr9
1751	10	nkpd1

7150	17	myo7b
22791	9	cdc42ep2
17438	3	enpep
2526	11	nfatc2
6936	16	g2
21329	7	slc1a1
6327	16	slk
7723	18	tcf7
9581	2	hoxb8a
20710	7	plxdc2
2527	11	nfatc2
18286	4	slc28a2
9935	2	cdhr2
2248	11	myo7b
4077	13	aqp8
6937	16	g2
12822	23	cdh17
19726	6	cd101
14984	26	fat4
5827	15	rnf223
20161	6	slc10a2
9533	2	lat
6904	16	gdnf
4853	14	hsd17b2
7704	17	cd3g
19584	5	adamts10
20527	6	gpa33
16981	3	scocb
1416	10	alp
4620	14	cdhr5
10799	20	ccl25
12669	23	xirp1
6359	16	casp3
4127	13	muc13
21220	7	slc7a7
6497	16	#REF!
11554	21	c21h1orf116
16814	29	fabp2
19301	5	dspp
1625	10	klc4
4244	13	#REF!

144

20620	7	ccr9
13406	24	scin
9138	19	cd36
7459	17	fgfr1
507	1	slc6a18
18863	5	sorbs3
5346	15	hoxc11a
506	1	slc6a19
4333	13	pax6
4333	13	prss27
8640	19	arx
21490	7	foxq1
12914	23	trim35
6160	16	plac8
1898	10	anx2a
22612	9	bcl11b
15262	27	galr1
9025	19	tspan19
1959	11	krt8
9552	2	baiap2
17707	3	runx3
6006	15	slc39a5
3152	12	apoeb
22805	9	ppp1r14b
17918	4	b4galnt3
12501	23	ccr9
5343	15	hoxc13a
18259	4	anpep
4451	14	slc6a19
19853	6	vil1
18453	4	trim35
6341	16	ggt1
15642	27	ace2
3758	13	clrn3
2443	11	slc5a8
307	1	apoa1
5490	15	trim35
19428	5	tmem252
6127	15	pparg
20200	6	pdzk1
5616	15	mapk12

21465	7	hrh3
8043	18	cdx1
10272	2	prkg1
21766	8	rnf128
4913	14	ush1c
12930	23	il20ra
3814	13	myo15b
3170	12	sh2d2a
15074	26	gimap8
10779	20	sema4e
5380	15	lpar6
4244	13	ace
20942	7	klhl6
15079	26	cd300lf
13660	24	gmip
20742	7	muc2
20103	6	p2ry8
20524	6	cd247
729	1	smpdl3b
11454	21	ssr1
2489	11	arhgap9
19463	5	traf1
6832	16	igsf5
17742	3	lbh
635	1	myh2
6381	16	epha5
6808	16	parp14
19636	5	mpnd
16505	29	crip2
12274	22	selplg
7079	17	samsn1
1533	10	matk
20265	6	arhgap15
2464	11	tespa1
13749	24	ncald
15500	27	dpp4
1571	10	arhgap45
10764	20	cd86
752	1	pl8l1
12305	22	cd53
20434	6	cxcr4

21086	7	dgkd
5045	14	rgs14
19311	5	sypl2
9858	2	rac2
17348	3	ikzf1
19586	5	myo1f
5185	14	nrip2
19988	6	wipf1
6756	16	il2rg
7239	17	lpar6
11440	21	nckap1
14840	26	tap2
21848	8	dusp1
21848	8	ergic1
3399	12	gucy2c
6497	16	dmpk
19980	6	hoxd10
9893	2	gdir2
22592	9	fut9
19296	5	was
3726	13	unc13d
10854	20	arhgap30
15116	26	lpar2
21158	7	olfr473
15226	27	adcy7
16054	28	megf9
21780	8	gpr174
8267	18	rcsd1
13371	24	arpp21
13323	24	il21r
14888	26	kcnq5
6834	16	tmprss2
3044	12	dab2ip
6570	16	ets2
22050	8	cd40lg
5640	15	sulf2
7322	17	fam129a
4045	13	sh3bp1

Figure S1

148