

Article

# Transcriptome Analysis of Epithelioma Papulosum Cyprini Cells Infected by Reovirus Isolated from Allogynogenetic Silver Crucian Carp

Cui-Yu Ba <sup>1,2</sup>, Xiao-Yan Du <sup>3</sup>, Pei-Jun Zhang <sup>4</sup>, Ping Chen <sup>5</sup>, Ya-Nan Cai <sup>1</sup> and Yue-Hong Li <sup>1,2,\*</sup>

<sup>1</sup> Animal Science and Technology, Jilin Agriculture University, Changchun 130118, China; bacuiyu@163.com (C.-Y.B.); Caiyanan0925@163.com (Y.-N.C.)

<sup>2</sup> College of Life Sciences, Jilin Agriculture University, Changchun 130118, China

<sup>3</sup> Freshwater Fisheries Research Institute of Jilin Province, Changchun 130000, China; dxypa@tom.com

<sup>4</sup> Health Monitoring and Inspection Center of Jilin Province, Changchun 130062, China; zhangpeijun63@sina.com

<sup>5</sup> College of Food Science and Engineering, Jilin Agriculture University, Changchun 130118, China; ccchenping@sina.com

\* Correspondence: liyhong@sina.com

Received: 8 January 2018; Accepted: 16 March 2018; Published: 18 March 2018

**Abstract:** The present study aimed to identify differentially expressed genes (DEGs) and major signal transduction pathways that were related to the immune response of epithelioma papulosum cyprinid (EPC) cells to reoviruses isolated from allogynogenetic silver crucian carp. The study also lays a theoretical foundation for the pathogenesis and immunity of the reovirus, which is helpful to the breeding of cyprinids fish. Reovirus infected and uninfected EPC cells were analyzed by using a new-generation high-throughput sequencing technology. DEGs were identified, annotated, and classified, and the signal pathways involved in the response to reovirus infection were identified by using bioinformatics tool. The data were assembled into 92,101 contigs with an average length of 835.24 bp and an N50 value of 1432 nt. Differential expression analysis of all the genes identified 3316 DEGs at a false discovery rate (FDR) of <0.01 and a fold-change of  $\geq 3$ , of which 1691 were upregulated genes, 1625 were downregulated, and about 305 were immune-related genes. Gene Ontology (GO) enrichment analysis resulted in the annotation of 3941 GO terms, including 2719 biological processes (37,810 unigenes), 376 cell components (7943 unigenes), and 846 molecular functions (11,750 unigenes). KEGG metabolic pathway analysis matched the DEGs from pre-and post-infection EPC cells to 193 pathways, of which 35 were immune-related, including the Toll-like receptor, cytokine-cytokine receptor interaction, and the JAK-STAT signaling pathways.

**Keywords:** transcriptome; EPC; reovirus; allogynogenetic silver crucian carp

## 1. Introduction

Aquareoviruses are characterized to have a double capsid, icosahedral symmetry, and no capsule. The average diameter of aquareoviruses is 60–70 nm and its genome consists of 11 segments of double-stranded RNA [1,2]. Aquatic reoviruses have infected various economically important economic aquaculture fishes such as grass carp, black carp, barley fish, Atlantic salmon, and rare crucian carp, thereby resulting in higher mortality rates. Besides aquatic animals, reoviruses have been isolated from ostriches, bats, and ducks. However, reports on the isolation of reoviruses that infect allogynogenetic crucian carp are limited.

The transcriptome is the sum of all the transcripts of a given organism in a given state, including not only mRNAs but also non-coding RNAs [3,4]. The transcriptome is an essential link between genetic information and biological function (proteome). New genes and transcripts are often identified

by using the latest high-throughput sequencing technology and correlations are calculated via mRNA expression analysis [5–7]. Transcriptome analysis is mainly used in studies of differential gene expression, allele-specific expression, alternative splicing, fusion genes and other related studies [8–10]. Transcriptome sequencing (RNA-seq) pertains to the direct sequencing of a cDNA sequence using a large-scale sequencing technology, thereby resulting in tens of millions of reads that could determine its transcription level by direct comparison of the number of reads in the genomic region [11]. In recent years, RNA sequencing has become a widely used method for fish-virus interaction studies. Due to our limited understanding of the molecular mechanism underlying reovirus infection in Cyprinidae, most research studies on the resistance to reoviruses have mainly focused on immune-related genes such as IL-10, HSP70, and HSP90, as well as other areas such as antimicrobial peptides (AMPs), proteomics analysis, and physical and chemical control strategies.

To understand the molecular immune mechanism of reoviruses in cyprinid fish, the present study analyzed the transcriptome of epithelioma papulosum cyprinid (EPC) cells that were infected by reovirus using the Illumina/Solexa high-throughput sequencing technique. We investigated the mRNA expression levels, immune-related genes, and pathways, and identified several simple repeat sequences (SSRs) and single nucleotide polymorphisms (SNPs). The findings of the present study may be utilized in the elucidation of the molecular mechanism underlying the potential interaction of reoviruses isolated from allogynogenetic crucian carp with EPCs.

## 2. Materials and Methods

### 2.1. Isolation, Inoculation, and Sampling of Reoviruses from Allogynogenetic Silver Carassius auratus

The allogynogenetic reoviruses used in the present study were isolated from diseased gibel carp by our laboratory in a fishing ground in Jilin Province in 2015. The EPC cells and crucian carp reoviruses were stored at the Jilin Agricultural University aquatic laboratory. The EPC cells were cultured in 75-cm<sup>2</sup> flasks (BD Biosciences, Franklin Lakes, NJ, USA) with M199 cell culture medium containing 10% fetal bovine serum (Life Technologies, Carlsbad, CA, USA) and 1% double antibody (Sigma-Aldrich, St. Louis, MO, USA) at 27 °C. Upon reaching 80% confluency, the reovirus was added to the cell culture bottle for propagation. Upon reaching a confluency of >80%, the EPCs were harvested and stored at –80 °C as the virus mother liquor. The titer of the virus was measured on a 96-well plate covered with EPC cells using the Karber method.

Approximately  $1.4 \times 10^7$  cells were seeded onto 75-cm<sup>2</sup> culture flasks (BD Biosciences) and 90% confluency was achieved after incubating for 24 h in a 27 °C incubator. Reoviruses with a multiplicity of infection (MOI) of 10 were added to the cell cultures. The same volume of M199 medium was added to a control flask (Figure 1A), and the supernatant was discarded when the proportion of cytopathic lesions (CPE) reached about 80% (about 12 h) (Figure 1B), and the cells were collected, frozen in liquid nitrogen, and kept in a –80 °C ultra-low freezer. Each group consisted of two biological replicates.



**Figure 1.** Cytopathy of epithelioma papulosum cyprinid (EPC) cells after virus infection (100×).

## 2.2. Library Construction and Sequencing

Total RNA was extracted from the samples, and the purity, concentration, and integrity of the RNA samples were determined by using a Nanodrop, Qubit 2.0, and Agilent 2100 methods to ensure that the samples to be used for transcriptional sequencing were of good quality. mRNA from the extracted total RNA was enriched and purified by using poly(T)-rich low-adsorption magnetic beads. A fragmentation buffer was added to randomly break the mRNA. Using mRNA as template, the first cDNA strand was synthesized using random hexamers. Then, the second cDNA strand was synthesized by adding a buffer, dNTPs, RNase H, and DNA polymerase I. The cDNA was purified by using AMPURE XP beads. The purified double-stranded cDNA was further subjected to terminal repair with A-tails and ligated into sequencing adapters. The AMPure XP beads were then used to select the fragment size. Finally, the cDNA library was obtained by PCR amplification. The established library was sequenced on an Illumina HiSeq<sup>TM</sup> 4000 system (San Diego, CA, USA).

## 2.3. Transcriptome Analysis

The raw data obtained by sequencing was initially processed. The data with a ratio of reads containing connectors and N (undetermined bases) >10%, along with the low-quality reads were discarded, and the corresponding clean reads database was obtained. The clean data of each sample were aligned and compared to the assembled unigene library. The reads that matched Unigenes were designated as mapped reads, which were then used in the subsequent analysis. The expression level of each gene in the sample was calculated using the reads per kb per million reads (RPKM) method [5] using only the number of reads and the total number of reads that could be matched to the reference sequence.

## 2.4. Data Assembly and Annotation

The clean data obtained after quality control was assembled using Trinity [12] (<https://github.com/trinityrnaseq/trinityrnaseq/wiki>) to produce unigenes [13]. Unigene sequences were aligned and compared to the NR [14], Swiss-Prot [15], GO [16], COG [17], KOG [18], eggNOG4.5 [19] and KEGG [20] databases by using BLAST [21]. KOBAS2.0 [22] was employed for KEGG analysis of Unigenes. The amino acid sequence of each unigene was predicted and then annotated based on the results of alignment and comparison using HMMER [23] and the Pfam [24] database.

## 2.5. Screening of Differentially Expressed Genes (DEGs)

The relative expression level of unigenes was determined by using the number of reads per kilobase of the map to the number of reads per 1 million base of the exon (FPKM). DEGs between the experimental and control groups were calculated using the DESeq [25] technique [26]. For differential expression analysis, the well-known Benjamin-Hochberg method was used to correct the p-value from the original test. Using the corrected p-value, the false discovery rate (FDR) was selected as the key index for screening DEGs. In this study, the FDR was <0.01, and the fold-change was  $\geq 3$ .

## 2.6. Functional Annotation and Enrichment Analysis of DEGs

Gene Ontology (GO) is a functional annotation of the DEGs. Using the Goatools software (<https://github.com/tanghaibao/GOatool>) for feature enrichment, significant enrichment was observed when the FDR was <0.05. COG classification of DEGs from various groups of samples was also performed. KEGG pathway analysis using the KEGG Orthology-Based Annotation System (KOBAS; <http://kobas.cbi.pku.edu.cn/home.do>) was also conducted, and the signaling pathways were verified using the Fisher's test and an FDR of  $\leq 0.05$ .

## 2.7. SSR and SNP Identification

The analysis of the unigene sequences using MicroSAteLLite (MISA) software (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) identified six types of SSRs: mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats. Single nucleotide polymorphisms (SNP) were analyzed using STAR [27] (<https://github.com/alexdobin/STAR>).

## 2.8. Real-time Polymerase Chain Reaction Analysis

To validate the data generated from Illumina HiSeq<sup>TM</sup> 4000 sequencing (San Diego, CA, USA), 10 DEGs were randomly selected for real-time quantitative PCR (qPCR) analysis. Primers were designed using a primer software (Table 1). The total RNA used for RT-PCR was the same as that for the transcriptome. A Bole (CFX96) PCR instrument was used for PCR detection with  $\beta$ -actin as internal reference. The reaction system (total volume: 25  $\mu$ L) consisted of the following: 12.5  $\mu$ L SYBR Premix Ex Taq II, 1  $\mu$ L of the upstream and downstream primers, 2.5  $\mu$ L of the cDNA template, and supplemented with dH<sub>2</sub>O to a final reaction volume of 25  $\mu$ L. To confirm that only one PCR product was amplified, the PCR product was subjected to dissociation curve analysis at the end of each PCR reaction. The data were analyzed by using the CFX (Computational Fluid Dynamics X) software (version 3.1, Hercules, CA, USA) package using Ct values ( $2^{-\Delta\Delta Ct}$  values) to analyze the expression levels of different genes.

**Table 1.** Primers, sequences, target genes, product sizes, and applications used in the present study.

Name	Sequence (5' → 3')	Product Size (bp)
IL-11F	TAGAGCACTGCTAGGCCTGA	124
IL-11R	TACCACGATGCCTGTTGACC	
MYD88F	GAGGACAGTCGCCGAAATGA	123
MYD88R	TTTGCTACAGTGGCCTCTGG	
P38F	CTTTCTCCCACTCGCTGGT	146
P38R	AATCCCTCTGCTTGCTGCC	
HSP90F	ATGGTGGAGTTGTCTCGCAG	127
HSP90R	TTGAGAAGGTCACCGTGTCC	
HERC4F	ACAGGTGGAGGGCTTAGGTT	133
HERC4R	CGGTGTGTTGGAAAACCTCT	
IRAK1F	GGCATCCCAACATAATGGAC	119
IRAK1R	GTTGCTGTCTCACAGCGTA	
CXCR4F	ACCGTCATCCTCATCCTTTG	86
CXCR4R	ACGTTTCAGCTCCACCAGAGT	
RasF	ACGACCCAACCATTGAGGAC	131
RasR	TCCGGTCCTCATGACTGGT	
PI3KF	GATGGAGCCTTCCTCATCCG	135
PI3KR	CAAGTTGTAGGGCTCTGCCA	
AMPKF	TACCGTGCCATGAAACAGCT	130
AMPKR	GGTTGTCCACCTGGTAGAGC	
$\beta$ -actinF	TGCAGTCAACATCTGGAATC	191
$\beta$ -actinR	ATTTTGGCGCTTGACTCAG	

## 3. Results and Analysis

### 3.1. RNA Sequencing and Transcriptome Sequence Assembly

In the present study, high-throughput sequencing of EPC cells (T03 and T04) infected with reovirus and uninfected EPC cells (T01 and T02) was performed via high-throughput transcriptional analysis and short sequence assembly analysis. The purity, concentration, and integrity of the RNA of the four samples were assessed by using Nanodrop, Qubit 2.0, and the Agilent 2100 system. The results showed that the concentration and total amount of the four samples were in accordance with the sequencing requirements. Four samples were sequenced using a HiSeq<sup>TM</sup> 4000 (San Diego, CA, USA)

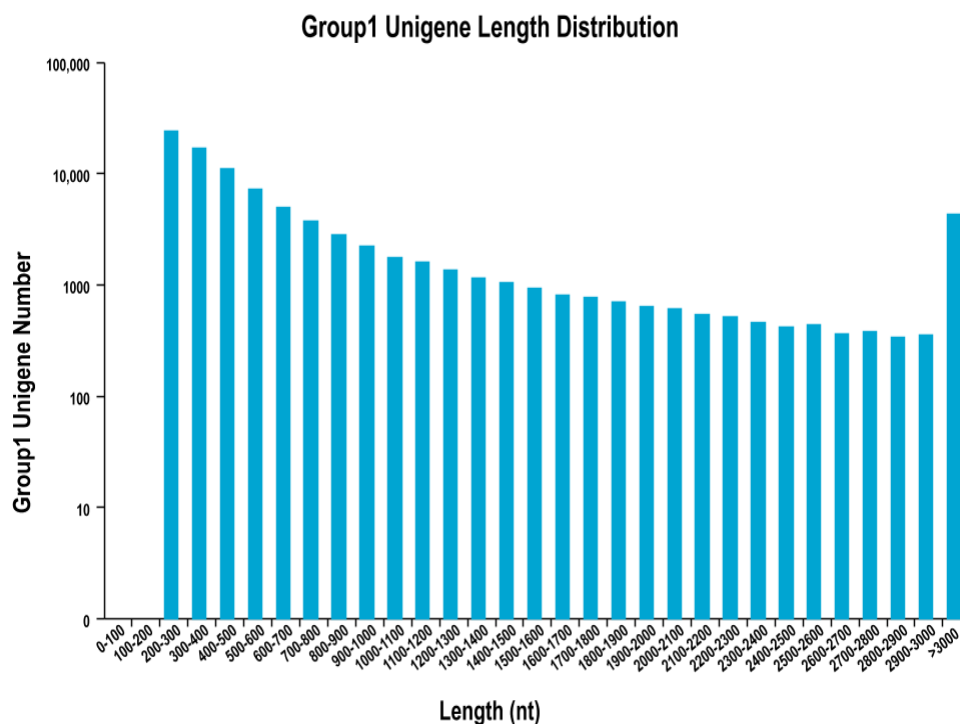
high-throughput sequencing platform. After removing the linker and primer sequences, filtering low-quality data, and performing sequencing quality control, read number and base number of each sample were obtained (Table 2). The analysis found that the percentages of Q30 bases were not >89.44% and GC Content were >47.19%, which indicated that the sequencing results were good and could generate good original data for the subsequent data assembly.

**Table 2.** Statistical data for sample sequencing.

Sample	Read Number	Base Number	GC Content	% $\geq$ Q30
T01	36,262,497	10,633,672,216	47.49%	89.44%
T02	28,349,442	8,398,124,460	47.15%	89.57%
T03	28,251,699	8,299,768,760	48.54%	89.91%
T04	29,010,341	8,549,124,730	47.87%	90.09%

Note: Samples: Sample name on the sample information sheet; Read number: Total number of paired-end reads in the clean data; Base number: Total number of bases in the clean data; GC content: GC content in the clean data, which is the percentage of G and C bases in the total bases in the clean data; %  $\geq$ Q30: Percentage of bases with a mass value  $\geq$  30 in the clean data.

Using Trinity (<http://trinityrnaseq.sourceforge.net/>) to assemble and splice the clean reads, the data was assembled into 92,101 contigs with an average length of 835.24 bp and a N50 value of 1432 nt. The length distribution of the reads is shown in Figure 2, with Unigene length 300–500 being the largest number, accounting for 30.23%, followed by Unigene length 200–300. The clean reads were mapped to Unigene, and the results are shown in Table 3, mapped ratio were >73.37%, and EPC cells (T03 and T04) infected with reovirus and uninfected EPC cells (T01 and T02) were similar, indicating better reproducibility. The assembly showed that the pattern of length distribution and average length of contigs were similar to the results of previous transcriptome studies using Illumina sequencing [28,29]. The large amount of sequence information obtained not only filled the gene expression profile of reovirus-immunized cyprinids, but also facilitated the mining of genetic information resources [30].



**Figure 2.** Unigene length distribution. The abscissa indicates the length of the unigene, and the ordinate indicates the number of unigene of a certain length.



**Table 3.** Comparison of sequencing data and assembly results.

Sample	Number of Clean Reads	Number of Mapped Reads	Mapped Ratio
T01	36,262,497	26,659,090	73.51%
T02	28,349,442	20,801,516	73.37%
T03	28,251,699	21,018,859	74.39%
T04	29,010,341	21,523,297	74.19%

Note: Samples: Sample name on the sample information sheet; Clean reads: Number of clean reads, paired end; Mapped reads: Number of mapped reads, paired end; Mapped ratio: Percentage ratio of mapped reads in the clean reads.

### 3.2. Functional Annotation and Classification of Transcriptional Group Data

The 92,101 assembled unigenes as describe above were compared to those in the NR, SWISS-PROT, eggNOG, COG, Pfam, KOG, GO, and KEGG databases respectively. By selecting the BLAST parameter E-value of  $< 1 \times 10^{-5}$  and the HMMER parameter E-value of  $< 1 \times 10^{-10}$ , 42,522 unigene annotations were obtained. The results of the comparative analysis are shown in Table 4, compared to COG, CO, KEGG, KOG, Pfam, Swissprot, eggNOG and NR in the database Unigene, respectively, 10,157 (23.9%), 20,555 (48.3%), 21,915 (51.5%), 27,599 (64.9%), 28,364 (66.7%), 22,535 (53%), 39,539 (92.9%), 41,260 (97%) Unigene. A total of 10,157 (23.9%) Unigene are annotated in at least one database.

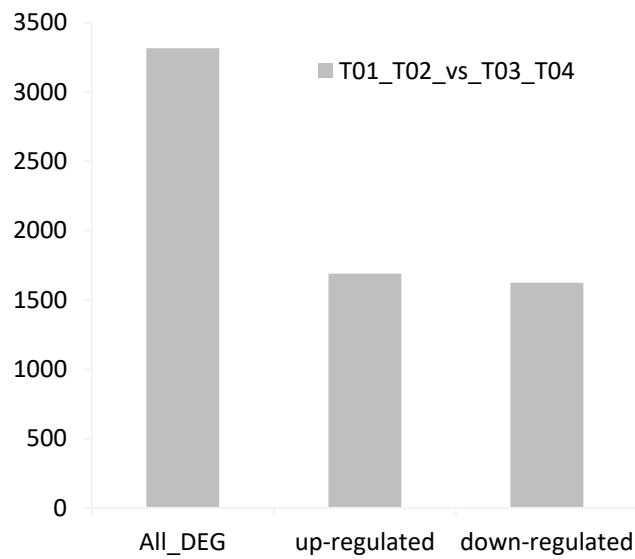
**Table 4.** Distribution of annotated unigenes.

Anno_Database	Annotated_Number	$300 \leq \text{Length} < 1000$	$\text{Length} \geq 1000$
COG_Annotation	10,157	2887	6538
GO_Annotation	20,555	6433	12,164
KEGG_Annotation	21,915	7875	11,248
KOG_Annotation	27,599	9471	14,856
Pfam_Annotation	28,364	8584	17,552
Swissprot_Annotation	22,535	7054	13,358
eggNOG_Annotation	39,539	14,403	19,799
nr_Annotation	41,260	15,294	20,454
All_Annotated	42,522	15,898	20,572

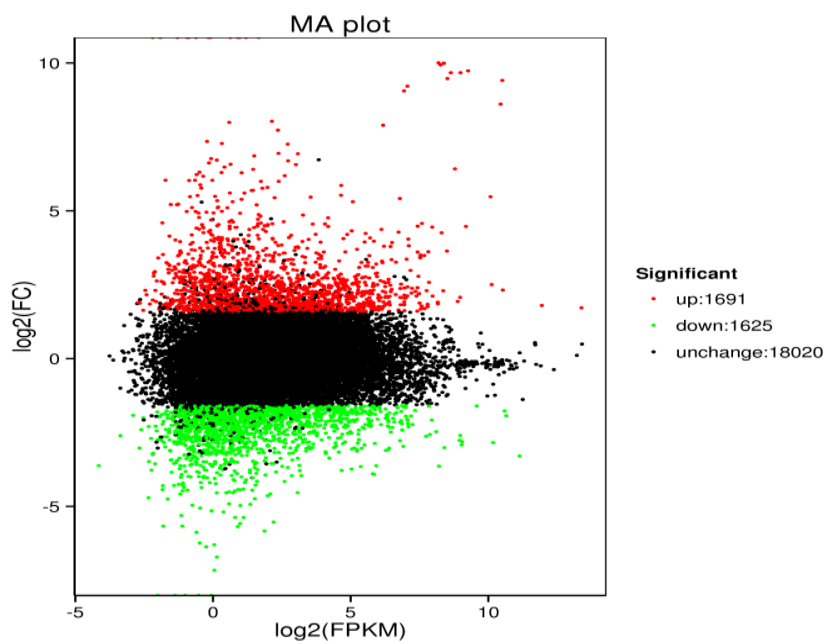
Note: Annotated databases: the function of the database; Annotated\_Number: Number of comments annotated to the unigene database;  $300 \leq \text{Length} < 1000$ : Number of unigenes annotated to the database that is  $\leq 300$  or less and  $< 1000$  bases in length;  $\text{Length} \geq 1000$ : Number of unigenes annotated to the database that are  $> 1000$  bases in length.

### 3.3. Differential Gene Expression of EPC Cells Pre-infected and Post-infected by Reovirus from Allogynogenetic Crucian Carp

To identify DEGs that may play an important role in the defense response to reovirus infection in cyprinids, the data of the test groups (T03 and T04) infected with reoviruses from allogynogenetic silver crucian carp were compared to that of the uninfected control groups (T01 and T02). A total of 3316 DEGs were identified by using the standard FDR of  $< 0.01$  and fold-change of  $\geq 3$ , of which 1691 were upregulated and 1625 were downregulated (Figure 3). The distribution of DEGs is illustrated using an MA plot (Figure 4) [31]. To visualize the gene expression profiles, we generated a heatmap using the Hcluster algorithm, where the color changes from red to green with decreasing expression levels [32].



**Figure 3.** The number of differentially expressed genes. DEG Set: Name of differentially expressed gene set; All DEG: Number of differentially expressed genes; upregulated: TNumber of upregulated genes; downregulated: Number of downregulated genes.

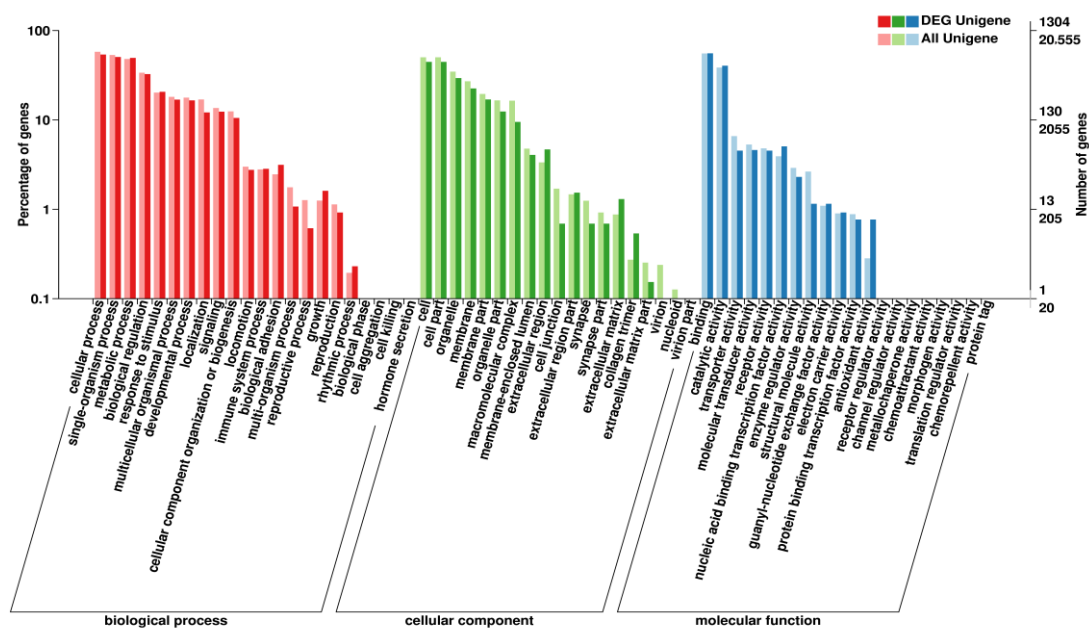


**Figure 4.** MA map of differentially expressed genes. Each point represents a gene in the MA map of differentially expressed genes. The abscissa is the A value:  $\log_2(\text{FPKM})$ , the logarithm of the mean value of the expression in the two samples. The ordinate is the M value:  $\log_2(\text{FC})$ , the logarithm of the difference in gene expression between the two samples, which was used to measure the difference in the level of expression. The green and red dots represent genes with significant differences in expression levels, green represents downregulated gene expression, red represents upregulated gene expression, and black spots represent genes with no significant differences in expression.

#### 3.4. DEGs GO Functional Enrichment Analysis

The GO database is a structured standard biology annotation system constructed by the GO organization, which aims to establish a standard vocabulary system for genes and its products and is applicable to all species. To further reveal the function of DEGs, 3316 DEGs from the test group (T03

and T04) that were infected with allogynogenetic reovirus and the uninfected control group (T01 and T02) were subjected to GO Note (Figure 5). The results showed that 3941 GO terms were annotated, including 2719 biological processes (37,810 unigenes). Among the 376 cell components (7943 unigenes) and 846 molecular functions (11,750 unigenes), it is indicated that most of the differentially expressed genes are mainly involved in the biological process of cells, but there are few genes involved in cell components and molecular functions. The DEGs were mainly involved in cellular processes (GO: 0009987), single biological processes (GO: 0044699) (GO: 0044464), organelle (GO: 0043226), and catalytic activity (GO: 0003824). These results indicated that various physiological and biochemical changes occurred in the EPC cells during reovirus infection, and these played an important role in virus defense response, it is possible to recognize the functional distribution of differentially expressed genes from the macroscopic angle of infection of the infected reovirus.



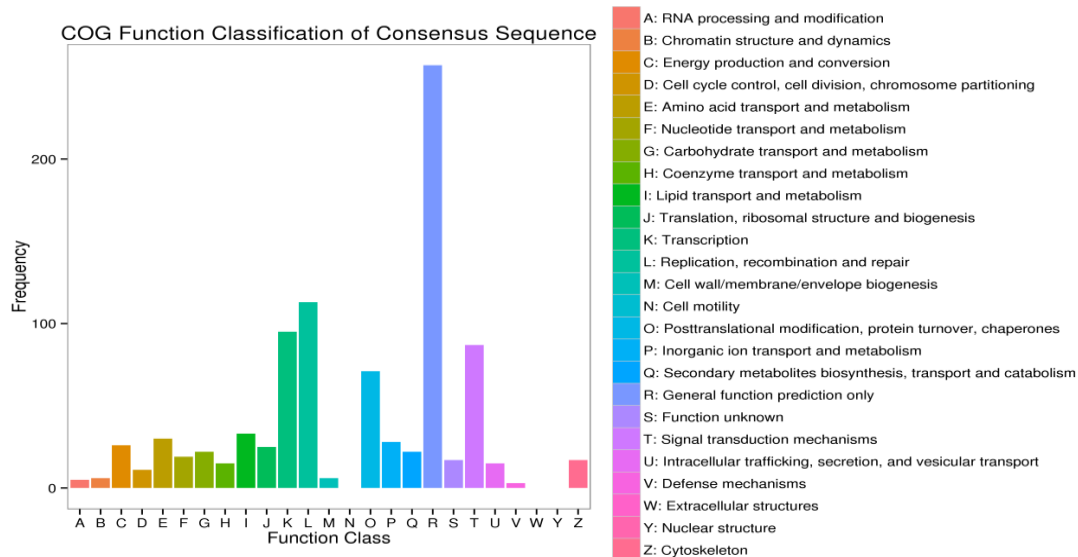
**Figure 5.** Clustering map of differentially expressed genes Gene Ontology (GO) in test group and control group. The abscissa is the GO classification, on the left of the ordinate is the percentage of the number of genes, and on the right is the number of genes. This figure shows the secondary function of each GO gene enrichment status in the background of the differentially expressed genes and the background of all the genes, thus reflecting the status of each secondary function in two backgrounds. The secondary function with significant difference indicates that the differentially expressed genes are different from the enrichment trend of all genes, and can be focused on whether this function is related to the difference.

### 3.5. COG Annotation of DEGs

COG (Clusters of Orthologous Groups) is a database of orthologous classification of gene products. Each COG protein is presumed to be derived from an ancestral protein. The COG database was constructed based on bacterial, algal, eukaryotic encoded proteins with complete genomes and phylogenetic relationships. A total of 3166 differentially expressed genes from the comparative analysis of the test group (TO3 and T04) that was infected with allogynogenetic reovirus to the uninfected control groups (T01 and T02) were directly orthologous and classified with gene products in the COG database. The resulting classification is shown in Figure 6. A total of 25 different functions were involved. According to the number of genes arranged from high to low in the functional classification, the number of functional units in the first six positions were 257 in general function prediction (R), 113 in replication, recombination and repair (L), 71 in translational modification, protein turnover,



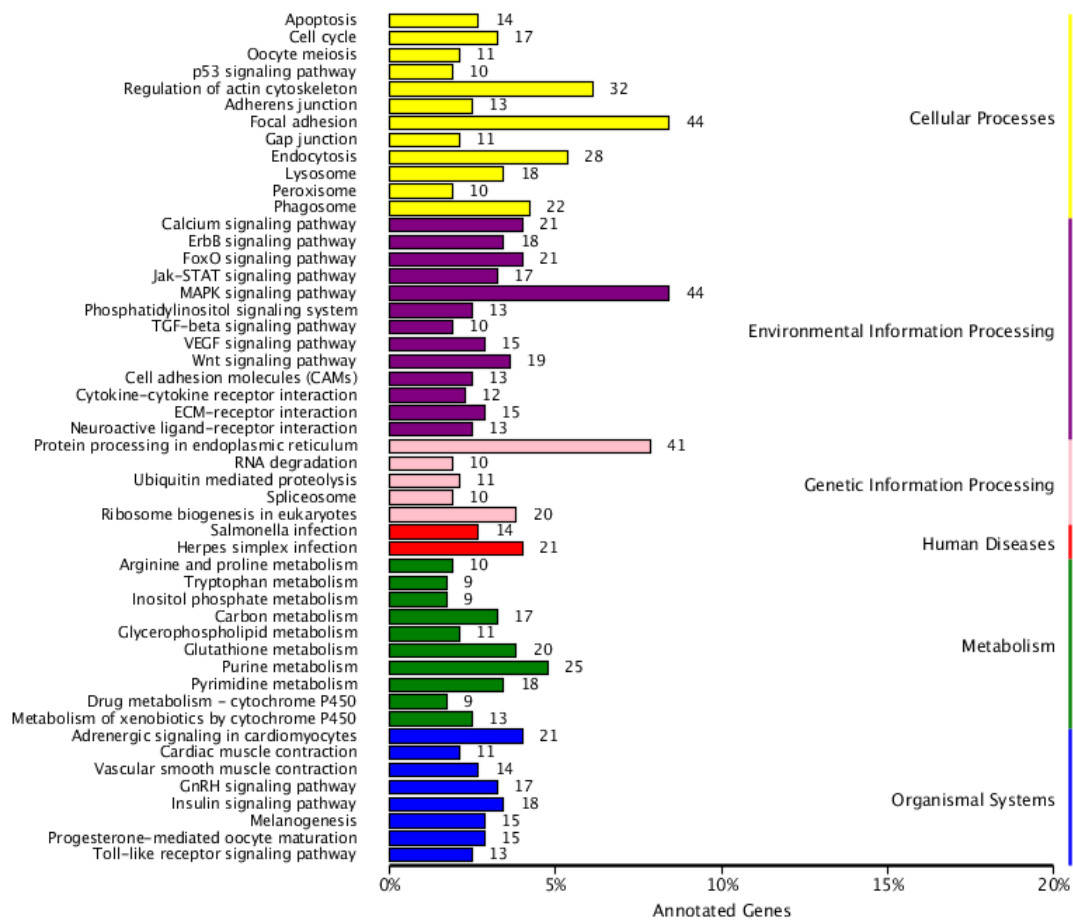
chaperone (O), 33 in lipid transport and metabolism (I), 87 in signal transduction mechanism (T), 95 in transcription (K), 36 in each of secondary metabolite synthesis, transport and decomposition (Q), and amino acid transport and metabolism (E). It shows that EPC cells can resist the injury by adjusting their physical metabolism in the face of the reovirus infection.



**Figure 6.** Orthologous classification of the differentially expressed genes in the COG (Clusters of Orthologous Groups) database between the experimental group and the control group. The abscissa is the COG (Clusters of Orthologous Groups) content of each classification, and the ordinate is the number of genes. In different functional categories, the number of genes reflects the corresponding period and the environment under the metabolic or physiological biases and other content, and the object can be combined with the distribution of various functional classes to generate a scientific explanation.

### 3.6. Pathway Enrichment Analysis of DEGs

The execution of biological functions relies on the coordination among genes. Pathway enrichment analysis of DEGs could provide better insights into the biological function of a gene and its interaction with other genes, so as to further study their complex behavior. The differential genes in the test group (T03 and T04) infected with reovirus and the uninfected control group (T01 and T02) were found to be involved in 50 subclasses of metabolic pathways in six broad categories (cellular processes, environmental information processing, genetic information processing, human diseases, metabolism, and organic systems) after these were analyzed by using the metabolic pathway database KEGG (Figure 7). Among these, several signaling pathways with a relatively high number of genes were identified, which included actin cytoskeleton regulation (ko04810), adhesion (ko04510), MAPK signaling pathway (ko04010), and protein processing in the endoplasmic reticulum (ko04141). These metabolic pathways are involved in the signal transduction of environmental information processing. There were 35 pathways associated with immune and disease resistance, including cytokine-cytokine-receptor interaction (12 genes, ko04060), apoptotic signaling pathway (21 genes, ko04068), phagocytosis (22 genes, ko04145), endocytosis (28 genes, ko04144), and Toll receptor signaling pathway (13 genes, ko04620). These pathways are basically involved in all aspects of fish immunization. This demonstrates that the experimental KEGG annotation effect is good and shows the reovirus and immune-induced regulatory pathways. Genes and pathways associated with the immune system, signal transduction and disease processes were similar to those previously reported in zebrafish [33], grouper [34], and bream [35]. Studies on immune-related genes and pathways can improve our understanding of the molecular immune mechanisms of reovirus-stimulated EPCs.



**Figure 7.** KEGG classification of differentially expressed genes. The ordinate is the name of the KEGG metabolic pathway, and the abscissa is the ratio of the number of genes annotated to the pathway and the number of genes in the annotated genes.

### 3.7. SSR Tags

Among various molecular markers, SSRs have many recognized functions and are widely used in paternity testing, genetic diversity, some aspects of genetic linkage maps, and marker-assisted breeding [36]. SSRs are tandem repeats of single nucleotides to hexanucleotides found in all prokaryotic and eukaryotic genomes. As a molecular marker, SSRs have a variety of putative functions and high polymorphisms that play an important role in certain aspects of genetic and breeding research. In addition, gene transcription and mRNA splicing or export to the cytoplasm may be affected by intron SSRs. Therefore, it is essential to understand the spectrum of SSRs. In the present study, we determined that all 21,985 unigenes in the transcriptome group contained SSRs (Table 5).

SNPs are the most abundant type of DNA sequence polymorphisms and have been increasingly used in quantitative trait loci (QTLs) for molecular marker linkage map construction, mapping, and association studies [37,38]. According to the number of alleles (Allele) on the SNP locus, i.e., the number of different bases supported by the sequencing reads, SNP sites can be divided into homozygous SNP sites (only one allele) and heterozygous SNP sites (two or more alleles). The percentage of heterozygous SNPs in different species varies. The number of SNP sites in each sample in this study is shown in Table 6. Both the Homo SNP and HeteSNP crucian carp were less in the test group (T03 and T04) infected with reovirus and the uninfected control group (T01 and T02). We can further study genetic markers related to immune-related genes and provide reference for the pathogenesis of reovirus-infected cyprinid fish and the search for genetic markers of immune genes.

**Table 5.** Statistical table of SSR analysis results.

#Type	Number
c	1556
c *	68
p1	11,943
p2	6079
p3	2078
p4	249
p5	10
p6	2
Total	21,985

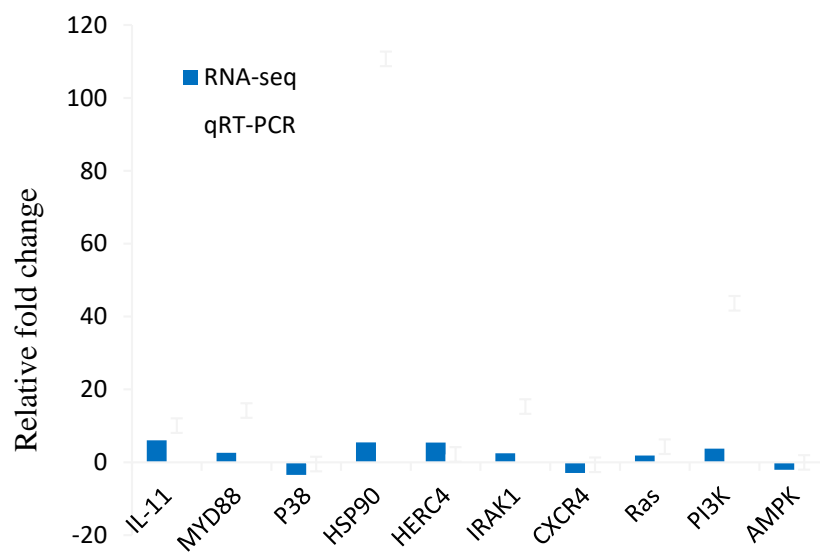
Note: Total number of identified sequences: Total number of sequences evaluated: Total size of examined sequences (bp): Number of total bases evaluated; Total number of identified SSRs: Total number of SSRs identified; Number of SSR containing sequences: number of sequences containing SSRs; Number of sequences containing >1 SSR: Number of sequences containing >1 SSR and SSR overlaps each other; Number of SSRs present in compound formation: Number of SSRs in complex form; Mononucleotide: Single-base repeat SSR; Dinucleotide: Double-base repeat SSR; Trinucleotide: Three-base repeat SSR; Tetranucleotide: Four-base repeat SSR; Pentanucleotide: Five-base repeat SSR; Hexanucleotide: Six-base repeat SSR.

**Table 6.** SNP distribution.

Sample	HomoSNP	HeteSNP	AllSNP
T01	37,230	66,496	103,726
T02	35,852	60,150	96,002
T03	35,445	57,150	92,595
T04	34,355	53,879	88,234

### 3.8. Real-Time PCR Verification of DEGs

The selected genes were associated with cytokine receptors, Toll-like receptor signaling conduction (TLR), chemokine signaling pathways, and MAPK signaling pathways. Compared to the control group, the expression level of these genes in the EPC cells infected with reoviruses was consistent with that generated from transcriptome analysis (Figure 8).

**Figure 8.** Comparison of relative fold changes between RNA-seq and qRT-PCR results.

#### 4. Discussion

The molecular response mechanisms of animals following viral infection include changes in transcriptional and translational levels, which in turn regulate different metabolic pathways and signaling pathways. Therefore, transcriptome research has become an important direction of functional genomics research. The second-generation transcriptome sequencing (RNA-Seq) technology is a high-throughput sequencing approach involving the cDNA from fragmentally processed mRNA, sequence splicing assembly, and statistical correlation of number of sequences to obtain different transcripts (mRNA). This technique has been widely applied to the determination and analysis of host-virus interaction transcriptome. So far, studies on the mechanisms underlying the molecular immune response of cyprinids to reoviruses on cyprinids are limited. To fully understand the molecular mechanism underlying reovirus response, we first analyzed the transcriptome data of EPC cells before and after reovirus infection as well as obtained gene data. We then analyzed various gene expression profiles to identify genes that are involved in the immune response and signal transduction (Table 7).

The immune system of fish provides protection from the pathogens, thereby allowing its survival and evolution. The Toll-like receptor family that mediates immune congenital immunity is the first line of defense of fish and plays a role in connecting innate immunity and acquired immune responses. Thirteen representative immune-related genes were significantly differentially expressed in the Toll-like receptor signaling pathway (Table 7, Figure 9). The MyD88 gene was significantly upregulated in the metabolic pathways of EPC cells after infection with reovirus, suggesting that it might play a key role in the generation of anti-reoviruses in carp and thus conferring pathogen resistance, which is similar to the findings of previous research studies [39–42]. In addition, the expression of the PIK3C and TAK1 genes were significantly downregulated after reovirus stimulation, which may be a negative regulatory mechanism of excessive inflammatory response. The results of this study differed from those of other viral infections, suggesting that certain genes in the Toll-like receptor signaling pathway play a key role in the response to reoviruses. Up-regulated expression of AP-1 and STAT after reovirus infection can promote the expression of proinflammatory cytokines and inflammatory cytokines, such as TNF $\alpha$ , IL-12/IL-1 $\beta$ . The release of these cytokines extracellularly can cause cells, macrophages chemotactic aggregation, increased capillary permeability, lymphocyte infiltration, and other inflammatory reactions.

The interaction between cytokines and membrane receptors is an important link in various biological activities. Cancer, autoimmune diseases, metabolic disorders, and other diseases are closely related to cytokine imbalance, and studying the interactions between cytokines and receptors becomes the cornerstone for the development of treatment of these diseases. Cytokines are a class of high activity, multi-functional protein peptide molecules or glycoproteins that are produced by activated immune cells (e.g., lymphocytes and mononuclear macrophages) and related cells (e.g., fibroblasts and endothelial cells) and thereby regulate cell function. These interact with specific receptors with high affinity, and are an indispensable component of the immune system. Approximately 12 representative immune-related genes were significantly differentially expressed in cytokine-cytokine receptor interaction signaling pathways after being infected with reovirus (Table 7, Figure 10). CXCL14 is a novel chemokine that decreases the inflammatory response. The CXCL14 gene was significantly downregulated in the metabolic pathways of EPC cells after being infected with reoviruses, thereby indicating that the inflammatory reaction may play a role in preventing the development of reoviruses. CXCR4 plays an important role in immunity, organogenesis and repair, anti-cancer, embryonic development, and hematopoiesis. CXCR4 can prevent the occurrence of excessive inflammatory reaction by negatively regulating the production of immune factors. In the present study, the CXCR4 gene was significantly downregulated in the metabolic pathways of EPC cells after infection with reovirus, indicating that it plays an important role in resistance against reovirus infection. The study of cytokine gene expression regulation may identify antagonists or activators that actively adjust the expression of specific genes. In addition, these studies may facilitate the establishment of an efficient eukaryotic expression system that generated a high yield of recombinant cytokines for the treatment of specific diseases or vaccine adjuvants.

The JAK-STAT signaling pathway mediates cytokine signal transduction. It is involved in multiple life processes such as disease infection, immunity, inflammation, and development. In this study, ten representative immune-related genes were significantly differentially expressed in the JAK-STAT signaling pathway (Table 6, Figure 11). The three main members of the JAK-STAT signaling pathway are cytokine receptors, JAK, and STAT. STAT is a unique protein family that binds to DNA. So far, seven members of the STAT family have been identified, which include STAT1, STAT2, STAT3, STAT4, STAT5a, STAT5b, and STAT6. STAT1 is mainly involved in the response to IFNs, STAT1 deficiency can cause increased susceptibility to viral infection and disorders of intracellular pathogen infection regulation. In the present study, cytokines, growth factors, and other factors activated JAK, and then activated STAT1 after binding to their associated receptors. The STAT1 and CISH (cytokine inducible SH2-containing protein) were significantly upregulated in EPC cells that were infected with reoviruses, indicating that their susceptibility to virus infection decreased, the body was stimulated to activate the immune system, and the disorders of intracellular pathogen infection regulation were reduced. The role of IL-6 and receptor genes in the JAK-STAT signaling pathway may be further investigated. The PIK3R gene was upregulated in EPC cells after reovirus infection, possibly promoting apoptosis. This study laid the foundation for further investigating the role of the fish JAK-STAT gene in fish immune regulation and disease control. By performing series of analyses and research studies on the JAK-STAT signaling pathway, schemes for disease control and targets for the development of new drugs may be established.

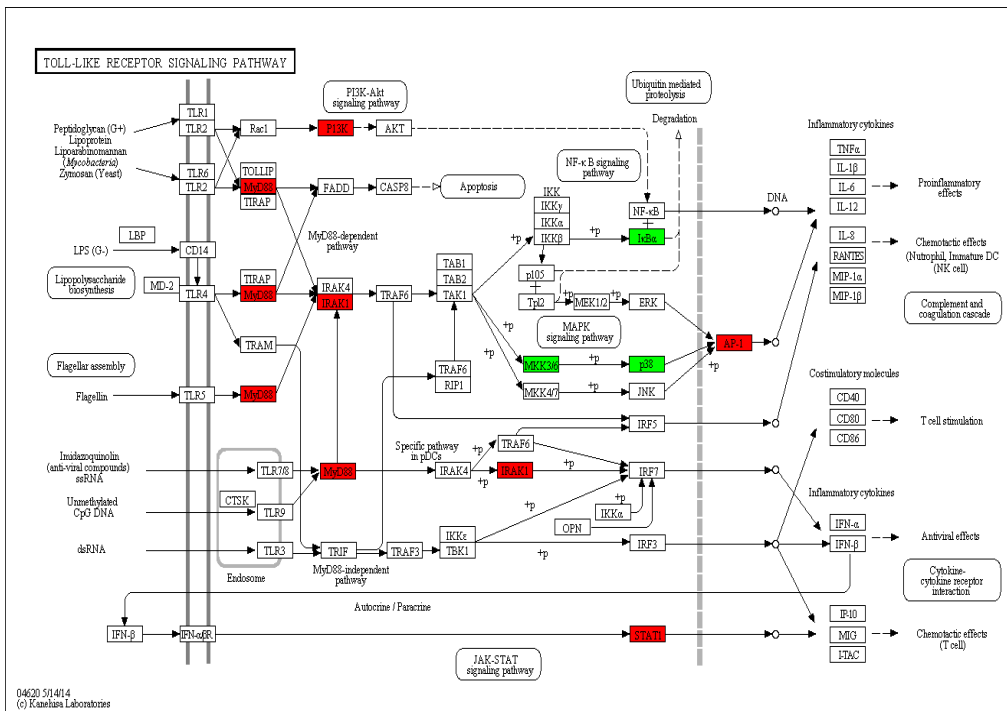
NLRs were found recently, they are cytoplasmic pattern recognition receptors (PRRS). NLRs play an important role in defending against pathogen infection by recognizing pathogen-associated molecular patterns (PAMPs) and initiating the innate immune response, leading to an adaptive immune response through signal transduction. Approximately 3 representative immune-related genes were significantly differentially expressed in NOD-like receptor signaling pathway after being infected with reovirus (Figure 12). Among them, the significant down-regulation of I $\kappa$ B and P38 inhibits NOD1 (Nucleotide-binding oligomerization domain-containing protein 1) directly recruiting RIPK2 through CARD-CARD (Caspase activation and recruitment domains) interaction during viral recognition, further activating NF- $\kappa$ B and MAPK (mitogen-activated protein kinase) signaling and stimulating the body to produce immunity. In this study, there was no difference in RIPK2, which may be due to the early stage of Infection increased and then decreased in the early stage of Infection increased and then decreased, Or the amount of expression increased but the difference was not significant compared with the control group, which can be further studied in the follow-up. It has been demonstrated that the HSP90 and SGT1 interact with the LRR (leucine-rich repeat) domain of NLRP3 (leucine-rich repeat and pyrin domain domains-containing protein 3) to maintain a non-activated but stable state, activating NLRP3 inflammation through changes in HSP90 The body causes secretion of IL-1 $\beta$  or IL-18 as well as programmed cell death. In addition HSP90 involved in a variety of physiological processes in vivo, as a molecular chaperone to improve the heat resistance of organisms and protect organisms and other functions. The infection of reovirus HSP90 in EPC cells significantly increased, indicating that not only by regulating its receptor protein activity and maintaining the stability of the receptor protein structure and indirectly involved in the regulation of multiple intracellular signaling pathways, but also to deal with resistance pathogen invasion and regulation of the body's immune system also plays an important role, suggesting that it might play a key role in the generation of anti-reoviruses in carp and thus conferring pathogen resistance, which is similar to the findings of previous research studies [43,44].

In addition to the Toll-like and JAK-STAT signaling pathways that interact with cytokines and cytokine receptors, various representative immune system-related genes are significantly expressed in phagocytosis and chemokine- and apoptosis-related pathways. Some genes play an important role in other pathways such as Ras; for example, it is involved in the regulation of the actin cytoskeleton, the MAPK signal transduction pathway, and the chemokine pathway; Furthermore, MyD88 is involved in apoptosis and the Toll pathway. The identification of these common genes in different pathways may improve our understanding of the molecular immune mechanisms of reovirus-infected cyprinids.

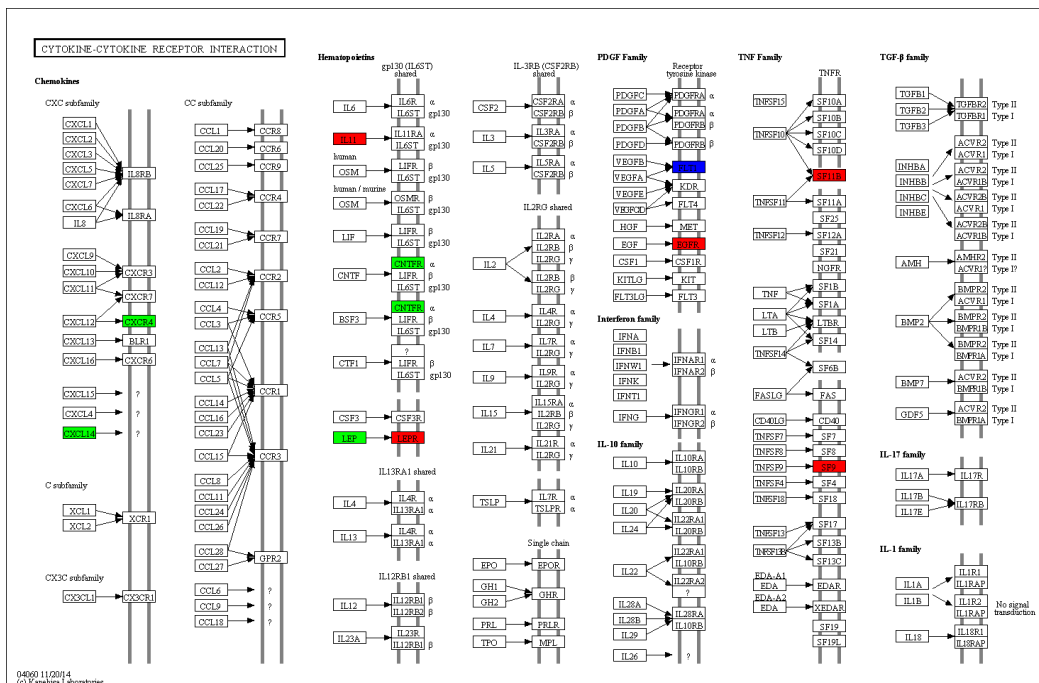


**Table 7.** Representative immune-related genes that are differentially expressed after reovirus infection.

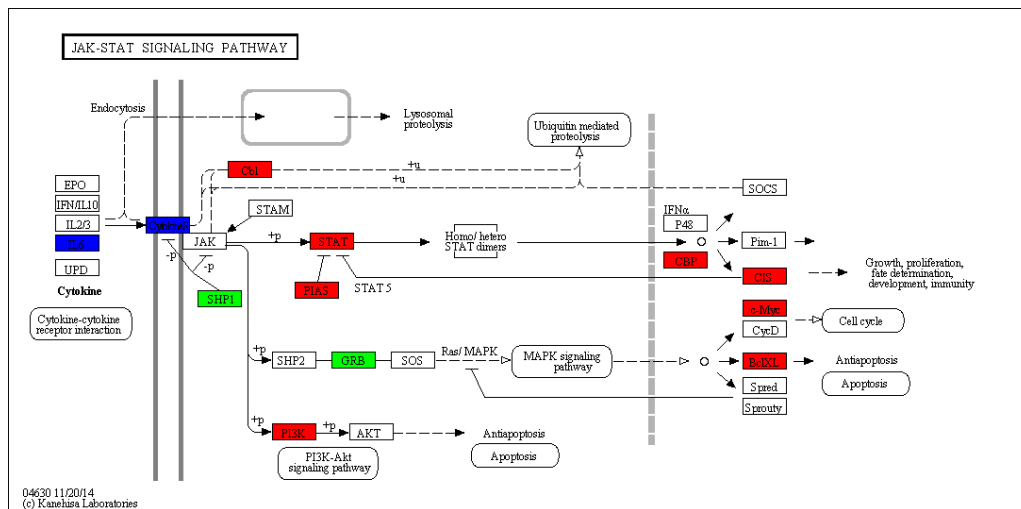
Gene Name	Description	Change	log2FC	FDR
T-cell receptor signaling pathway				
Ras	GTPase HRas	Up	1.84	0.000154
SNARE interactions in vesicular transport				
VAMP4	Vesicle-associated membrane protein 4	Up	1.68	0.005729
Nicotinate and nicotinamide metabolism				
PNP	Purine-nucleoside phosphorylase	Up	1.92	$4.45 \times 10^{-17}$
NRK1_2	Nicotinamide/nicotinate riboside kinase	Up	1.61	$4.37 \times 10^{-6}$
Phagosome				
ATPeV0A	V-type H+-transporting ATPase subunit a	Up	1.67	$6.01 \times 10^{-9}$
TUBB	Tubulin beta	Up	2.17	$1.48 \times 10^{-6}$
C1R	Complement component 1, r subcomponent	Down	-2.39	$5.62 \times 10^{-6}$
ITGB2	Integrin beta 2	Down	-1.73	$5.78 \times 10^{-5}$
MRC	Mannose receptor, C type	Up	2.91	$2.73 \times 10^{-8}$
Jak-STAT signaling pathway				
STAT1	Signal transducer and activator of transcription 1	Up	1.62	$6.02 \times 10^{-5}$
CBL	E3 ubiquitin-protein ligase CBL	Up	2.28	0.007492469
SHP-1	Tyrosine-protein phosphatase non-receptor type 6	Down	-2.54	$5.79 \times 10^{-7}$
GRB2	Growth factor receptor-binding protein 2	Down	-1.74	$5.42 \times 10^{-5}$
EP300	E1A/CREB-binding protein	Up	1.67	0.000264763
CISH	Cytokine-inducible SH2-containing protein	Up	3.38	$4.58 \times 10^{-30}$
MYC	Myc proto-oncogene protein	Up	3.49	$9.21 \times 10^{-29}$
BCL2L1	Bcl-2-like 1 (apoptosis regulator Bcl-X)	Up	2.19	0.000500092
Toll-like receptor signaling pathway				
PIK3C	Phosphatidylinositol-4,5-bisphosphate 3-kinase	Up	3.73	$7.41 \times 10^{-32}$
MYD88	Myeloid differentiation primary response protein MyD88	Up	2.61	$1.96 \times 10^{-11}$
IRAK1	Interleukin-1 receptor-associated kinase 1	Up	2.49	$1.63 \times 10^{-22}$
NFKBIA	NF-kappa-B inhibitor alpha	Down	-3.11	$4.49 \times 10^{-17}$
MAP2K6	Mitogen-activated protein kinase 6	Down	-3.60	$1.29 \times 10^{-8}$
P38	p38 MAP kinase	Down	-3.04	$5.50 \times 10^{-10}$
FOS	Proto-oncogene protein c-fos	Up	2.30	$6.13 \times 10^{-6}$
Cytokine-cytokine receptor interaction				
CXCR4	C-X-C chemokine receptor type 4	Down	-2.93	$4.62 \times 10^{-7}$
CXCL14	C-X-C motif chemokine 14	Down	-1.65	$1.04 \times 10^{-12}$
IL11	Interleukin 11	Up	6.03	0.00595422
CNTFR	Ciliary neurotrophic factor receptor	Down	-1.86	0.001965213
LEP	Leptin	Down	-4.19	$1.23 \times 10^{-7}$
LEPR	Leptin receptor	Up	2.54	$1.02 \times 10^{-7}$
EGFR	Epidermal growth factor receptor	Up	2.22	$5.09 \times 10^{-5}$
TNFRSF11B	Tumor necrosis factor receptor superfamily member 11B	Up	4.86	$9.75 \times 10^{-40}$
TNFRSF9	Tumor necrosis factor receptor superfamily member 9	Up	4.37	$7.65 \times 10^{-15}$
Apoptosis				
TNFRSF10	Tumor necrosis factor receptor superfamily member 10	Up	2.26	0.000275944
MYD88	Myeloid differentiation primary response protein MyD88	Up	2.26	0.000275944
IRAK1	Interleukin-1 receptor-associated kinase 1	Up	2.49	$1.63 \times 10^{-22}$
MAP3K14	Mitogen-activated protein kinase 14	Up	3.13	$6.78 \times 10^{-8}$
PIK3C	Phosphatidylinositol-4,5-bisphosphate 3-kinase	UP	3.73	$7.41 \times 10^{-32}$
CAPN2	Calpain-2	Down	-1.60	$6.44 \times 10^{-9}$
CFLAR	CASP8 and FADD-like apoptosis regulator	Up	2.14	$2.40 \times 10^{-5}$
BCL2L1	Bcl-2-like 1 (apoptosis regulator Bcl-X)	Up	2.19	0.000500092
NFKBIA	NF-kappa-B inhibitor alpha	Down	-3.11	$4.49 \times 10^{-17}$
XIAP	E3 ubiquitin-protein ligase XIAP	Up	1.60	$3.86 \times 10^{-12}$
CASP3	Caspase 3	Down	-2.28	$1.94 \times 10^{-14}$
CASP9	Caspase 9	Up	2.20	$9.13 \times 10^{-15}$
mTOR signaling pathway				
PIK3R	Phosphoinositide-3-kinase, regulatory subunit	Up	3.73	$7.41 \times 10^{-32}$
PRKCA	Classical protein kinase C alpha type	Down	-1.69	$1.46 \times 10^{-7}$
AKT1S1	Proline-rich AKT1 substrate 1	Down	-1.63	$4.59 \times 10^{-7}$
AMPK	5'-AMP-activated protein kinase, catalytic alpha subunit	Down	-2.06	0.000638513



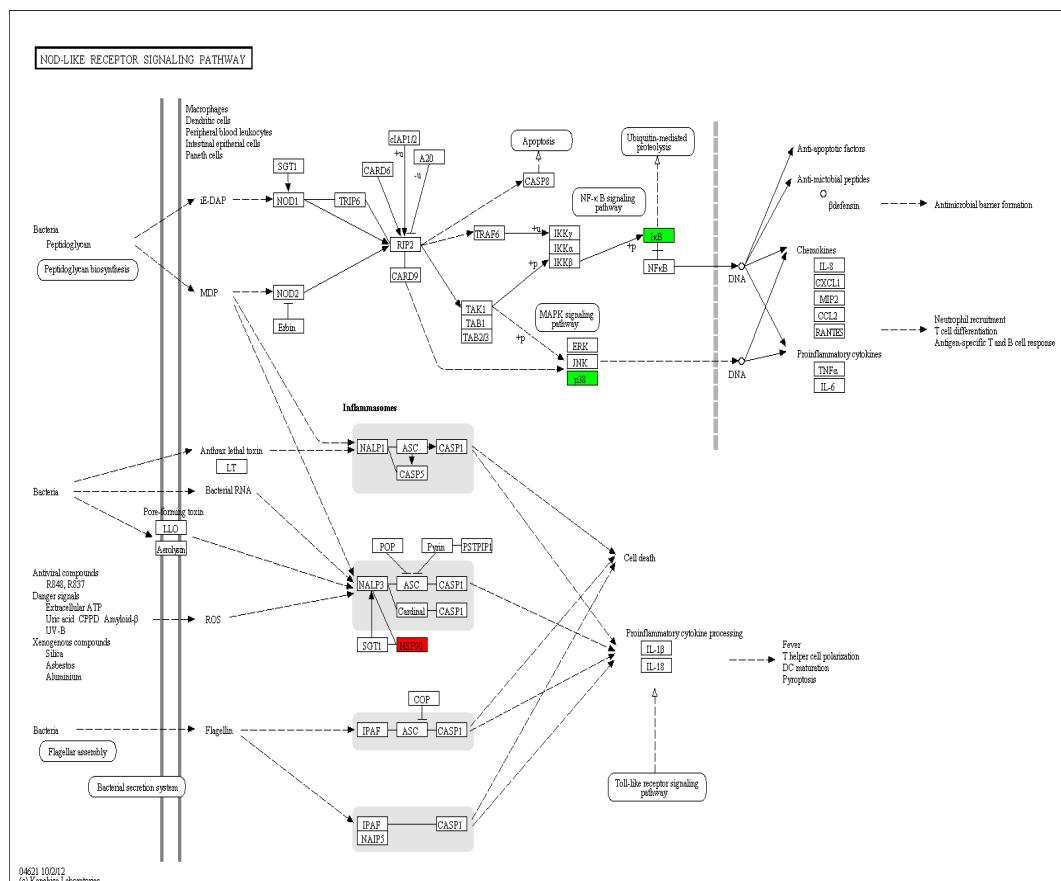
**Figure 9.** Significantly differentially expressed genes identified by KEGG that are involved in Toll-like receptor signaling pathway. Red boxes indicate upregulated DEGs; Green boxes represent downregulated DEGs; (For interpretation of colors used in this figure, please refer to the web version of this article).



**Figure 10.** Significantly differentially expressed genes identified by KEGG that are involved in the cytokine-cytokine receptor interaction signaling pathway. Red boxes indicate upregulated DEGs; Green boxes represent downregulated DEGs; Blue boxes show no changes in expression. (For interpretation of colors used in this figure, please refer to the web version of this article).



**Figure 11.** Significantly differentially expressed genes identified by KEGG that are involved in the Jak-STAT signaling pathway. Red boxes indicate upregulated DEGs; Green boxes represented downregulated DEGs; Blue boxes show no changes in expression. (For interpretation of colors used in this figure, please refer to the web version of this article).



**Figure 12.** Significantly differentially expressed genes identified by KEGG that are involved in the NOD-like receptor signaling pathway. Red boxes indicate upregulated DEGs; Green boxes represented downregulated DEGs; (For interpretation of colors used in this figure, please refer to the web version of this article).

## 5. Conclusions

In the present study, comparative transcriptome analysis of EPC cells infected with reovirus and uninfected EPC cells was performed to explore the molecular mechanisms of immunity and the development of molecular markers. EPC cells infected with reoviruses showed significant changes in the expression of a variety of immune-related pathways such as Toll-like receptor signaling, cytokine-cytokine receptor interactions, chemokine signaling pathways, and endocytosis. In addition, the identification of 21,985 SSR markers and 62,992 SNPs in the transcriptome of EPC cells facilitated the development of markers, the analysis of inheritance patterns, and establishment of QTLs. In summary, this study provides insights into the immunoprophylaxis of reovirus-infected cyprinids and provides a theoretical basis for the pathogenesis and immunological pathways of Cypriniform reoviruses, as well as providing useful information for fish farming.

**Acknowledgments:** This work was financially supported by The National Natural Science Foundation of China and Ministry of Agriculture 948 project. Thanks to He Biao for this study help.

**Author Contributions:** Yue-Hong Li and Cui-Yu Ba conceived and designed the experiments; Cui-Yu Ba performed the experiments; Xiao-Yan Du and Pei-Jun Zhang analyzed the data; Ping Chen and Ya-Nan Cai contributed reagents/materials/analysis tools and literature review; Cui-Yu Ba and Yue-Hong Li wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. He, Y.; Xu, H.; Yang, Q.; Xu, D.; Lu, L. The use of an in vitro microneutralization assay to evaluate the potential of recombinant VP5 protein as an antigen for vaccinating against Grass carp reovirus. *Virol. J.* **2011**, *8*, 132. [[CrossRef](#)] [[PubMed](#)]
2. Jian, J.C.; Wang, Y.; Yan, X.Y.; Ding, Y.; Wu, Z.H.; Lu, Y.S. Molecular cloning and prokaryotic expression of vp5 gene of grass carp reovirus strain GCRV096. *Virus Genes* **2013**, *47*, 483–489. [[CrossRef](#)] [[PubMed](#)]
3. Costa, V.; Angelini, C.; de Feis, I.; Ciccodicola, A. Uncovering the Complexity of Transcriptomes with RNA-Seq. *J. Biomed. Biotechnol.* **2010**, *2010*, 853916. [[CrossRef](#)] [[PubMed](#)]
4. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
5. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [[CrossRef](#)] [[PubMed](#)]
6. Cloonan, N.; Forrest, A.R.; Kolle, G.; Gardiner, B.B.; Faulkner, G.J.; Brown, M.K.; Taylor, D.F.; Steptoe, A.L.; Wani, S.; Bethel, G.; et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **2008**, *5*, 613–619. [[CrossRef](#)] [[PubMed](#)]
7. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* **2008**, *320*, 1344–1349. [[CrossRef](#)] [[PubMed](#)]
8. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
9. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **2008**, *456*, 470–476. [[CrossRef](#)] [[PubMed](#)]
10. Mitelman, F.; Johansson, B.; Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **2007**, *7*, 233–245. [[CrossRef](#)] [[PubMed](#)]
11. Ansorge, W.J. Next-generation DNA sequencing techniques. *N. Biotechnol.* **2009**, *25*, 195–203. [[CrossRef](#)] [[PubMed](#)]
12. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]

13. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]
14. Deng, Y.Y.; Li, J.Q.; Wu, S.F.; Zhu, Y.P.; Chen, Y.W.; He, F.C. Integrated nr Database in Protein Annotation System and Its Localization. *Comput. Eng.* **2006**, *32*, 71–72.
15. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, 115–119. [[CrossRef](#)] [[PubMed](#)]
16. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
17. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)] [[PubMed](#)]
18. Koonin, E.V.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Krylov, D.M.; Makarova, K.S.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S.; et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **2004**, *5*, R7. [[CrossRef](#)] [[PubMed](#)]
19. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2015**, *44*, D286–D293. [[CrossRef](#)] [[PubMed](#)]
20. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, 277–280. [[CrossRef](#)] [[PubMed](#)]
21. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
22. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [[CrossRef](#)] [[PubMed](#)]
23. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763. [[CrossRef](#)] [[PubMed](#)]
24. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, 222–230. [[CrossRef](#)] [[PubMed](#)]
25. Leng, N.; Dawson, J.A.; Thomson, J.A.; Ruotti, V.; Rissman, A.I.; Smits, B.M.; Haag, J.D.; Gould, M.N.; Stewart, R.M.; Kendziorski, C. EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **2013**, *29*, 1035–1043. [[CrossRef](#)] [[PubMed](#)]
26. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
27. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, X.; Wang, S.; Chen, S.; Chen, Y.; Liu, Y.; Shao, C.; Wang, Q.; Lu, Y.; Gong, G.; Ding, S.; et al. Transcriptome analysis revealed changes of multiple genes involved in immunity in *Cynoglossus semilaevis* during *Vibrio anguillarum* infection. *Fish Shellfish Immunol.* **2015**, *43*, 209–218. [[CrossRef](#)] [[PubMed](#)]
29. The PLoS ONE Staff. Correction: De Novo Characterization of the Spleen Transcriptome of the Large Yellow Croaker (*Pseudosciaena crocea*) and Analysis of the Immune Relevant Genes and Pathways Involved in the Antiviral Response. *PLoS ONE* **2014**, *9*, e97471. [[CrossRef](#)]
30. Wang, P.; Wang, J.; Su, Y.Q.; Mao, Y.; Zhang, J.S.; Wu, C.W.; Ke, Q.Z.; Han, K.H.; Zheng, W.Q.; Xu, N.D. Transcriptome analysis of the *Larimichthys crocea* liver in response to *Cryptocaryon irritans*. *Fish Shellfish Immunol.* **2015**, *48*, 1–11. [[CrossRef](#)] [[PubMed](#)]
31. Wang, L.; Feng, Z.; Wang, X.; Wang, X.; Zhang, X. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **2010**, *26*, 136–138. [[CrossRef](#)] [[PubMed](#)]
32. R Development Core Team. R: A language and environment for statistical computing. *Computing* **2014**, *1*, 12–21.



33. Lü, A.J.; Hu, X.C.; Wang, Y.; Zhu, A.H.; Shen, L.L.; Tian, J.; Feng, Z.Z.; Feng, Z.J. Skin immune response in the zebrafish, *Danio rerio* (Hamilton), to *Aeromonas hydrophila* infection: A transcriptional profiling approach. *J. Fish Dis.* **2014**, *38*, 137–150. [[CrossRef](#)] [[PubMed](#)]
34. Huang, Y.; Huang, X.; Yan, Y.; Cai, J.; Ouyang, Z.; Cui, H.; Wang, P.; Qin, Q. Transcriptome analysis of orange-spotted grouper (*Epinephelus coioides*) spleen in response to Singapore grouper iridovirus. *BMC Genom.* **2011**, *12*, 556. [[CrossRef](#)] [[PubMed](#)]
35. Tran, N.T.; Gao, Z.X.; Zhao, H.H.; Yi, S.K.; Chen, B.X.; Zhao, Y.H.; Lin, L.; Liu, X.Q.; Wang, W.M. Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish Shellfish Immunol.* **2015**, *45*, 72–82. [[CrossRef](#)] [[PubMed](#)]
36. Dawson, D.A.; Ball, A.D.; Spurgin, L.G.; Martín-Gálvez, D.; Stewart, I.R.; Horsburgh, G.J.; Potter, J.; Molina-Morales, M.; Bicknell, A.W.; Preston, S.A.; et al. High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC Genom.* **2013**, *14*, 176. [[CrossRef](#)] [[PubMed](#)]
37. Yu, H.; Xie, W.; Wang, J.; Xing, Y.; Xu, C.; Li, X.; Xiao, J.; Zhang, Q. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE* **2011**, *6*, e17595. [[CrossRef](#)]
38. Hubert, S.; Higgins, B.; Borza, T.; Bowman, S. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genom.* **2010**, *11*, 191. [[CrossRef](#)] [[PubMed](#)]
39. Van der Sar, A.M.; Stockhammer, O.W.; van der Laan, C.; Spaink, H.P.; Bitter, W.; Meijer, A.H. MyD88 innate immune function in a zebrafish embryo infection model. *Infect. Immun.* **2006**, *74*, 2436–2441. [[CrossRef](#)] [[PubMed](#)]
40. Rebl, A.; Goldammer, T.; Fischer, U.; Köllner, B.; Seyfert, H.M. Characterization of two key molecules of teleost innate immunity from rainbow trout (*Oncorhynchus mykiss*): MyD88 and SAA. *Vet. Immunol. Immunopathol.* **2009**, *131*, 122–126. [[CrossRef](#)] [[PubMed](#)]
41. Yu, Y.; Zhong, Q.W.; Zhang, Q.Q.; Wang, Z.G.; Li, C.M.; Yan, F.S.; Jiang, L.M. Full-length sequence and expression analysis of a myeloid differentiation factor 88 (MyD88) in half-smooth tongue sole *Cynoglossus semilaevis*. *Int. J. Immunogenet.* **2009**, *36*, 173–182. [[CrossRef](#)] [[PubMed](#)]
42. Takano, T.; Kondo, H.; Hirono, I.; Saito-Taki, T.; Endo, M.; Aoki, T. Identification and characterization of a myeloid differentiation factor 88 (MyD88) cDNA and gene in Japanese flounder, *Paralichthys olivaceus*. *Dev. Comp. Immunol.* **2006**, *30*, 807–816. [[CrossRef](#)] [[PubMed](#)]
43. Fu, D.; Chen, J.; Zhang, Y.; Yu, Z. Cloning and expression of a heat shock protein (HSP) 90 gene in the haemocytes of *Crassostrea hongkongensis* under osmotic stress and bacterial challenge. *Fish Shellfish Immunol.* **2011**, *31*, 118–125. [[CrossRef](#)] [[PubMed](#)]
44. Chen, Y.M.; Kuo, C.E.; Wang, T.Y.; Shie, P.S.; Wang, W.C.; Huang, S.L.; Tsai, T.J.; Chen, P.P.; Chen, J.C.; Chen, T.Y. Cloning of an orange-spotted grouper *Epinephelus coioides* heat shock protein90AB (HSP90AB) and characterization of its expression in response to nodavirus. *Fish Shellfish Immunol.* **2010**, *28*, 895–904. [[CrossRef](#)] [[PubMed](#)]

