



Article

Dynamic Transcriptional Landscape of Grass Carp (*Ctenopharyngodon idella*) Reveals Key Transcriptional Features Involved in Fish Development

You Duan ^{1,2} , Qiangxiang Zhang ^{1,2}, Yanxin Jiang ^{1,2}, Wanting Zhang ^{1,3}, Yingyin Cheng ^{1,3}, Mijuan Shi ^{1,3,*} and Xiao-Qin Xia ^{1,2,3,*}

¹ Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

² University of Chinese Academy of Sciences, Beijing 100049, China

³ The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

* Correspondence: shimijuan@ihb.ac.cn (M.S.); xqxia@ihb.ac.cn (X.-Q.X.); Tel./Fax: +86-27-6878-0915 (X.-Q.X.)

Abstract: A high-quality baseline transcriptome is a valuable resource for developmental research as well as a useful reference for other studies. We gathered 41 samples representing 11 tissues/organs from 22 important developmental time points within 197 days of fertilization of grass carp eggs in order to systematically examine the role of lncRNAs and alternative splicing in fish development. We created a high-quality grass carp baseline transcriptome with a completeness of up to 93.98 percent by combining strand-specific RNA sequencing and single-molecule real-time RNA sequencing technologies, and we obtained temporal expression profiles of 33,055 genes and 77,582 transcripts during development and tissue differentiation. A family of short interspersed elements was preferentially expressed at the early stage of zygotic activation in grass carp, and its possible regulatory components were discovered through analysis. Additionally, after thoroughly analyzing alternative splicing events, we discovered that retained intron (RI) alternative splicing events change significantly in both zygotic activation and tissue differentiation. During zygotic activation, we also revealed the precise regulatory characteristics of the underlying functional RI events.

Keywords: grass carp; development; zygotic activation; single-molecule real-time RNA sequencing; retained intron; lncRNA



Citation: Duan, Y.; Zhang, Q.; Jiang, Y.; Zhang, W.; Cheng, Y.; Shi, M.; Xia, X.-Q. Dynamic Transcriptional Landscape of Grass Carp (*Ctenopharyngodon idella*) Reveals Key Transcriptional Features Involved in Fish Development. *Int. J. Mol. Sci.* **2022**, *23*, 11547. <https://doi.org/10.3390/ijms231911547>

Academic Editor: Sarath Chandra Janga

Received: 31 August 2022

Accepted: 23 September 2022

Published: 30 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The absolute star of RNA study has been mRNA for a very long time. However, over the past 20 years, human understanding of RNA has substantially increased, and an increasing number of RNA kinds, functions, and regulatory mechanisms have been identified. The general functions of long non-coding RNAs (lncRNAs) and alternative splicing in biological processes are the most fascinating.

A family of RNA molecules known as lncRNAs are those that are longer than 200 bp and do not encode proteins. They can take part in a variety of life activities, including gene expression regulation, cell differentiation and development regulation, disease occurrence, and other biological processes [1–3]. Expressed retrotransposons, for example, are selectively active during the process of mammalian zygotic gene activation (the process of epigenetic reprogramming) [4–7].

Alternative splicing is the mechanism by which cells generate numerous mature mRNA products from a single pre-mRNA. It is a common mechanism in eukaryotes and is involved in a variety of biological processes, including development, cell and tissue differentiation, and disease incidence [8–10]. Retained intron (RI) is one sort of alternative splicing pattern in which alternative introns are not spliced out but are retained in mature mRNA.

RI is the most common type of alternative splicing event in plants and fungi, and it has been extensively researched [11,12]. It has also been gradually demonstrated to be functional in animals [13]. RI may regulate gene expression via a variety of ways, including nonsense-mediated mRNA decay (NMD), blocking mRNA export from the nucleus, and creating novel gene products. It is vital for plant development and stress resistance, and it has also been linked to a number of human disorders [13,14].

RNA-seq technology based on next-generation sequencing considerably simplifies transcriptome research. Because it does not rely on poly(A) tails for library construction, Illumina strand-specific RNA sequencing (ssRNA-seq) technology is frequently employed in systematic identification and expression profiling of lncRNAs [15–17]. Due to the fact that it provides specific benefits in the identification of transcript structure as a result of its lengthy sequencing reads, single-molecular real-time sequencing (SMRT)-based RNA-seq is frequently employed for full-length transcript assembly and alternative splicing analyses [18,19].

These sequencing technologies have made it much easier to construct a baseline transcriptome in a number of different species. As a result, our knowledge regarding the function of various RNAs and the alternative splicing events they participate in during the course of development has been significantly expanded [20–27]. Research in developmental biology can benefit from using fish embryo since it is simple to observe and manipulate and is a suitable material to study. If we carry out expression profiling research on a number of various time periods, it will unquestionably assist us in gaining a deeper understanding of the molecular control mechanism of ontogeny and differentiation at the transcriptional level.

Single-molecular real-time (SMRT)-based RNA-seq has been used a lot in fish studies to look at alternative splicing events. However, these studies have mostly focused on transcriptome assembly [28–31] and specific study subjects [32–38], and they have not looked at how alternatively spliced transcripts change in expression over time in a systematic way.

At the moment, most research on fish baseline expression is done on zebrafish, the most famous model fish. RNA microarray technology was used by Mathavan et al. in 2005 to look at the expression of 14,904 genes at 12 different times, from when the eggs were still unfertilized to when they were two days old [39]. Later, when RNA-seq technology advanced, the dynamics of gene expression throughout zebrafish development were investigated in more depth [40–44]. In comparison, studies on other fish species are not only scarce, but also of poor quality (with fewer time points). For example, only three samples encompassing three developmental time periods were gathered in a study of the Siberian sturgeon (*Acipenser baeri*) [45]. No more than five samples were collected per species in a comparative transcriptome investigation of six cichlid fishes [46]. Only six time points were explored in the bighead (*Hypophthalmichthys nobilis*) study, primarily including its early stage development [47]. In addition, 7 time points spanning 2 to 10 days post-fertilization (2–10 dpf) were sampled in the investigation of transcriptome dynamics of channel catfish (*Ictalurus punctatus*). However, because developmental morphology was not used as a reference during sampling, crucial time points in embryonic development were not gathered [48].

With an annual output of 5.7 million tons in 2018, accounting for 6.9 percent of the world's total farmed fish production, grass carp (*Ctenopharyngodon idella*) is the world's most productive farmed fish kind [49]. Currently, grass carp lncRNA research focuses mostly on the immunological response to infection, environmental stress response, feeding, and growth [50–53], with no systematic study to describe lncRNAs. The availability of the draft grass carp genome has tremendously aided genetics breeding research and other basic grass carp research [52,54–57]. However, the draft genome annotation is lacking in information on lncRNAs and alternatively spliced transcripts, making it challenging to meet the demands of increasingly diversified data processing. As a result, a thorough baseline expression investigation on grass carp is required to expand its gene functional

annotation while also gaining basic information of gene transcription dynamics in its development and differentiation.

In this study, we used SMRT-based RNA-seq and ssRNA-seq technology to systematically examine the essential transcriptional events in grass carp embryonic development to differentiated tissues. We collected 41 samples of 11 grass carp tissues at 14 time periods, built a transcriptome with a completeness of up to 93.98 percent (BUSCO), enhanced grass carp genome annotation, and recreated the entire grass carp development and differentiation landscape at the transcriptional level. We discovered a SINE family and its motif-related transcripts with unique expression patterns, and we detailed the specificity of RI alternative splicing events in embryonic development and tissue differentiation, as well as their possible regulatory elements.

2. Results

2.1. Transcriptome Sequencing and Assembly

Given the spatiotemporal specificity of gene expression, we sampled as many time points and tissue types as possible. We obtained 41 samples for transcriptome sequencing in a whole sibling population of grass carp, from fertilized egg to 197 days following fertilization, covering 22 time periods and 11 tissues (embryos are considered one tissue) (Table 1). These developmental stages are separated into four stages: zygotic genome activation (ZGA, T2–T8), gastrula to hatching out (GH, T9–T16), larvae (T17–T20), and juveniles.

Table 1. Sample information.

Sample ID	Time after Fertilization	Tissue	Characteristics
T2	0 hpf	ovum	ovum
T4	1.6 hpf	embryo	1, 2, 4 cells
T5	2.8 hpf	embryo	64 cells
T6	3.7 hpf	embryo	256 cells
T7	5.2 hpf	embryo	dome
T8	6.3 hpf	embryo	50% epiboly
T9	7.9 hpf	embryo	75% epiboly
T10	9.1 hpf	embryo	90% epiboly ~ bud
T11	11.1 hpf	embryo	4~6 somite
T12	12.9 hpf	embryo	15 somite
T13	14.8 hpf	embryo	18~19 somite
T14	18.0 hpf	embryo	25~26 somite, starting to shrink
T15	21.8 hpf	embryo	rotation
T16	31.8 hpf	full fish	hatching
T17	3 dpf	full fish	yolk exhausted, appearance of eyes, start to swim
T18	6 dpf	full fish	first feeding
T19	13 dpf	full fish	NA
T20	20 dpf	full fish	NA
B1, M1	39 dpf	B, M	NA
B2, M2, L1	69 dpf	B, M, L	NA
B3, M3, L2, G1, Sk1, Sp1, K1, H1, I1	134 dpf	B, M, L, G, Sk, Sp, K, H, I	NA
B4, M4, L3, G2, Sk2, Sp2, K2, H2, I2	197 dpf	B, M, L, G, Sk, Sp, K, H, I	NA

Note: hpf, hours after fertilization; dpf, days after fertilization. Abbreviations for tissues are B—brain, M—muscle, L—liver, G—gill, Sk—skin, Sp—spleen, K—kidney, H—heart, and I—intestine.

After sequencing, we were able to extract a total of 621 Gb of Illumina short reads and 87 Gb of PacBio long reads. Following assembly (shown in Figure 1A), we were able to get a total of 33,055 genes and 77,582 transcripts. In contrast to the annotation results of the grass carp genome (gc.v1) [54], the grass carp transcriptome (gc.v2) includes a greater number of alternative splicing events as well as more annotations for non-coding RNA (Figure 1B,C, Supplemental Table S1).

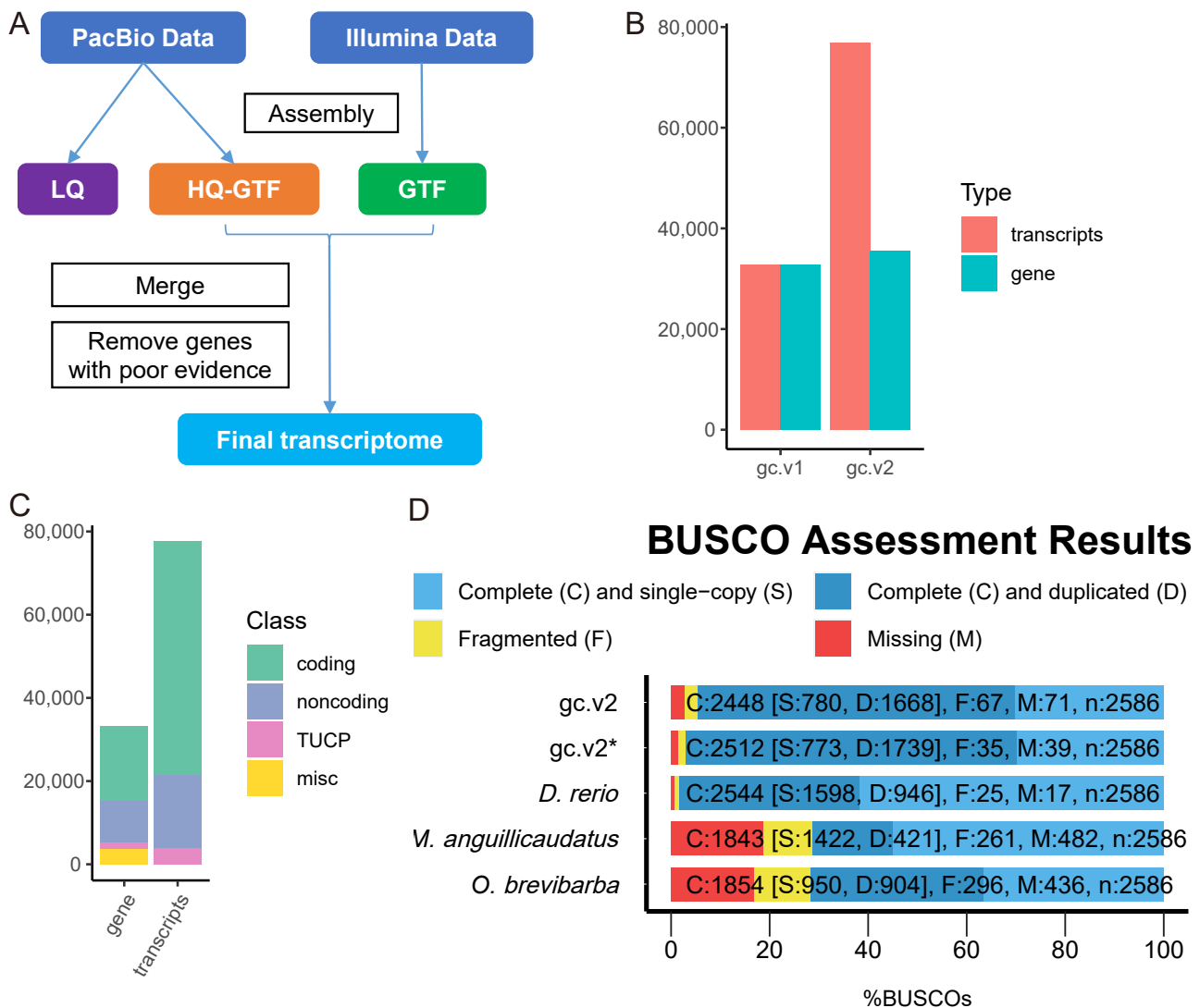


Figure 1. The process of constructing the whole transcriptome of grass carp. (A) A schematic representation of the assembly of the transcriptome. (B) The gc.v2 genome contains a greater number of transcripts and genes than the gc.v1 genome. The gc.v2 is the transcriptome that was assembled in this study, and the gc.v1 is the annotation information that was found in the draft grass carp genome [54]. (C) Statistics on the types of transcripts found in the gc.v2 transcriptome. (D) Analysis of the completeness of the gc.v2 transcriptome using BUSCO, which is based on a database of ray-finned fish. The result of calculating gc.v2* was done so under the assumption that all clusters produced by SMRT sequencing are able to generate transcripts based on the grass carp genome. This means that gc.v2* is equal to gc.v2 plus the transcripts that cannot be successfully aligned to the grass carp genome in the SMRT data analysis. *D. rerio* denotes the zebrafish transcriptome (GRCz11); *M. anguillicaudatus* represents the loach transcriptome based on SMRT sequencing [28]; *O. brevibarba* denotes the *Onychostoma abbrevibarba* transcriptome based on SMRT data [36].

Sequences in gc.v2 (3281 bp, median) are much longer than in gc.v1 (1642 bp) and are longer than the sequences both in loach (*Misgurnus anguillicaudatus*) transcriptome (3054) [28] and in *Onychostoma brevibarba* transcriptome (3054) [29], which were sequenced using SMRT technology. In addition to this, it was significantly longer than the RNAs that were found in the zebrafish transcriptome (2103) (GRCz11) (Supplemental Table S2).

The completeness of the gc.v2 transcriptome reaches 93.98 percent, which is better than that of the loach (*Misgurnus anguillicaudatus*) and the *Onychostoma brevibarba*, which are also based on SMRT sequencing; the only other transcriptome completeness that is lower is

that of the zebrafish (GRCz11) (Figure 1D). When the SMRT transcripts that could not be matched to the grass carp genome are taken into consideration, the level of completeness of our assembled transcriptome is remarkably similar to that of the zebrafish transcriptome: a difference of 1.72 percent, or 79 out of 4584 total transcripts.

2.2. Overview of Grass Carp Development and Tissue Differentiation

2.2.1. The Developmental Trajectory of Grass Carp

The expression levels were used in the principal component analysis (PCA) analysis of all 41 samples, and it was discovered that three-dimensional PCA had the ability to recreate the developmental trajectory and tissue differentiation scenarios of the samples (Figure 2). The principal component 1 (PC1) is able to effectively differentiate embryo from differentiated tissues, and the relationship between the larva fish (T17–T20) within one month of hatching is closer to differentiated tissue. On the other hand, the fish that is hatching (T16) is just located between the embryo and the juveniles.

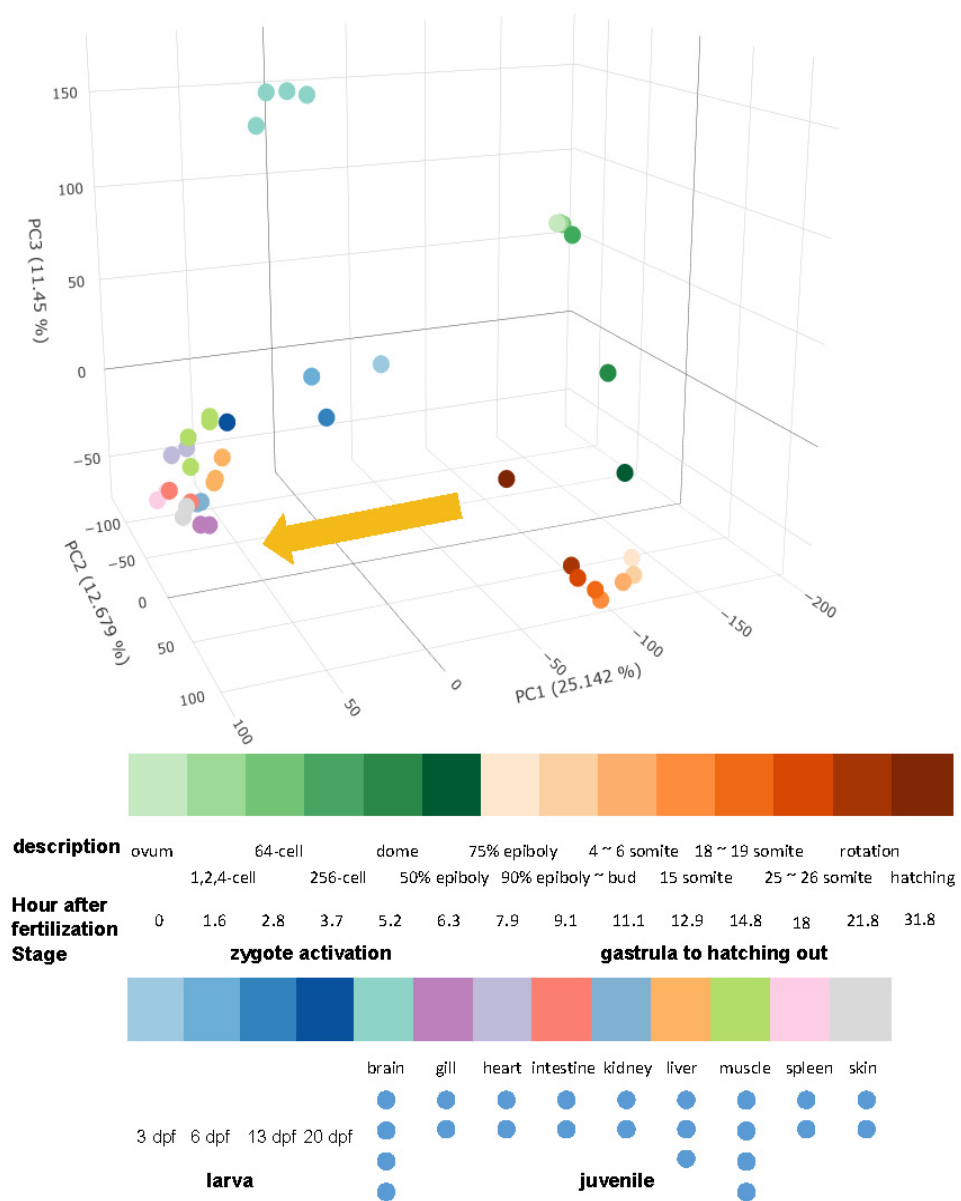


Figure 2. Principal compartment analysis (PCA) performed on all samples. (Top) Three-dimensional PCA diagram, in which the embryo development direction is indicated by the yellow arrow. (Bottom) The color panel and the development information for all samples.

There are a total of 4428 transcripts that are associated with PC1. According to the results of the GO enrichment analysis, they were primarily found in the nucleus (the cellular compartment, or CC), participated in the biological pathway (BP) that was associated with the metabolism of nucleic acids, and the nucleic acid binding was the most important molecular function that they possessed (Figure 3A). The majority of the considerably enriched KEGG pathways are those that are connected to immunological control and fundamental life regulation (Figure 3B). There are 4068 PC2-related transcripts, with more non-coding RNAs were found than expected ($p = 5.33 \times 10^{-38}$) (Supplemental Table S3). These transcripts were not enriched in any KEGG pathway; GO enrichment analysis shows that they were mainly development-related transcription factors (Supplemental Figure S1).

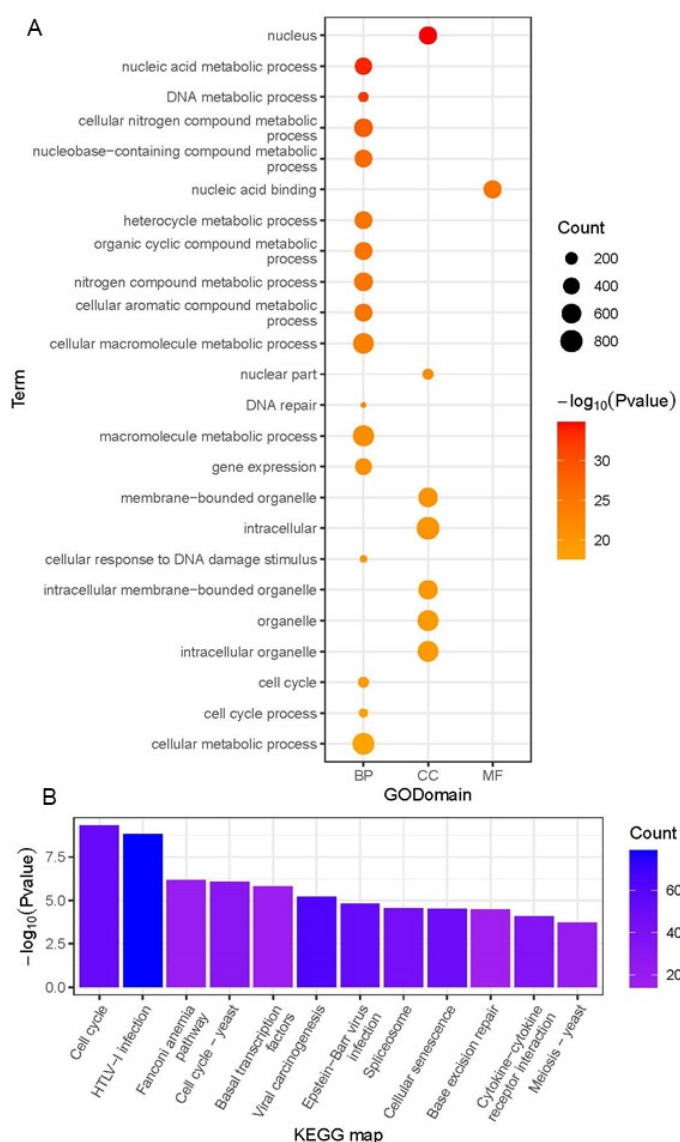


Figure 3. Functional enrichment results of PC1-related genes in all samples. (A) GO enrichment results (top 24); (B) KEGG enrichment results (top 12).

There are 4884 PC3-related transcripts, with more mRNAs found than expected ($p = 2.30 \times 10^{-9}$). GO enrichment analysis shows that they were mostly related to ion transport (Supplemental Figure S2). The most significant pathways in KEGG enrichment analysis were nicotine addiction, glutamatergic synapse, and neuroactive ligand–receptor interaction (Supplemental Figure S3).

2.2.2. Fish Brain Is Highly Differentiated

In order to explain changes in tissue composition in grass carp based on expression patterns, we carried out PCA analysis on all tissue samples (Figure 4A). It was discovered that PC1 based on tissue samples was highly comparable to PC3 based on all samples, and that PC2 based on tissue samples was highly similar to PC4 based on all samples. (Supplemental Figure S4). The differences between the various tissues of grass carp were, on average, substantially greater than those between the several phases of development present in the same tissue (Figure 4A). There was a significant difference between brain tissue and the other tissues, and the majority of this difference was explained by PC1 with an explanatory degree of up to 25 percent. PC1-related genes were mainly enriched in GO terms related to ion channels, ion transport, and synapses (Figure 4B), which clearly reflected the characteristics of metabolic activity in the brain; the most significant pathways of KEGG enrichment analysis were nicotine addiction, glutamatergic synapse, and neuroactive ligand–receptor interaction (Figure 4C).

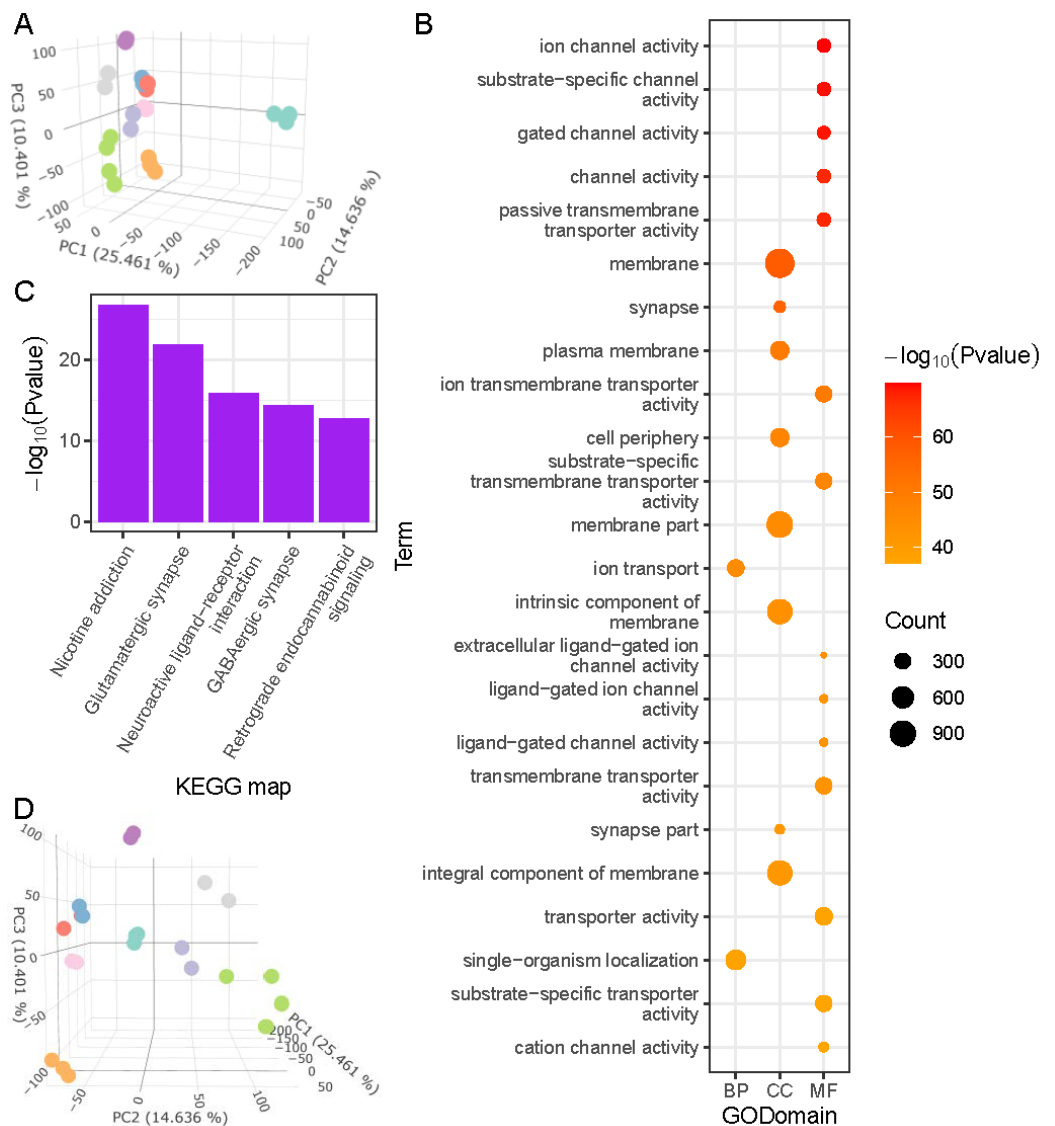


Figure 4. PCA results based on tissue samples. (A) Three-dimensional diagram of PCA results (sample colors are shown in Figure 2); (B) GO enrichment results of PC1-related genes (top 24); (C) KEGG enrichment results of PC1-related genes (top 5); (D) PC2 and PC3 (sample colors are shown in Figure 2).

2.2.3. Muscle Development and Basic Metabolic Regulation Describe Major Differences among Fish Organs

Except for the brain, PC2 and PC3 in tissue PCA analysis described the origin information of tissues jointly (Figure 4D). Muscle (HM) and heart (HH) were grouped together; liver (HL), spleen (HP), gut (HI), and kidney (HK) were grouped together to the left of PC2; and skin and gills were grouped together to the top right.

Significant GO terms in PC2 were largely connected to muscle, while significant KEGG pathways were mostly related to cardiac disease, according to enrichment analysis (Supplemental Figures S5 and S6). PC3-related genes were enriched in GO terms related to enzymes, amino acid metabolism, and organ development, while the significant KEGG pathways were mostly related to immunity (complement and coagulation cascades), regulation (ribosome), and amino acid metabolism (glycine, serine, and threonine metabolism) (Supplemental Figures S7 and S8). Coding RNAs were substantially more abundant in PC2-related genes, whereas non-coding RNAs were more abundant in PC3-related genes (Supplemental Table S4).

2.3. Expression Analysis of Grass Carp Transposons

2.3.1. Expression Scenery of Grass Carp Transposon

According to the annotation of the grass carp genome, transposable elements (TEs) are widely distributed in the grass carp genome, accounting for 38.06 percent of total genome sequences (Supplemental Table S5). Moreover, based on the gc.v2 transcriptome, we discovered that the majority of transcripts are connected with repeat elements (52,686 TE-related transcripts), yet only 64 of these transcripts were classified as transposons (see Section 4). When compared to their distribution in the genome, the discovered expressed transposons were imbalanced (Table 2, Supplemental Table S5). For example, the widely dispersed DNA transposons DNA/hAT (23.53 percent) were not found to be expressed, although DNA/TcMar (6.29 percent) were. Long terminal repeats (LTR) were a very prevalent retrotransposon (5.22 percent) in the grass carp genome, but only two of them were found to be expressed. The two most often expressed retrotransposons, long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE), belonged to non-LTR that were relatively infrequent (4.8 percent) in grass carp total TEs. Unclassified SINE, on the other hand, accounted for 64.06 percent (41/64) of all expressed transposons, despite accounting for only 0.6 percent of total TEs. Even more intriguing, all of the discovered expressed SINE transposons are members of rnd-3_family-293.

Table 2. Statistics of expressed transposable elements.

Class	Count	Ratio in Expressed TE	Count in Genome	Fraction in Total Transposon Elements
unclassified LTR	2	3.13%	14,757	0.88%
DNA/TcMar	8	12.50%	106,014	6.29%
LINE/L2	13	20.31%	44,915	2.66%
unclassified SINE	41	64.06%	10,123	0.60%

Note: In the repeat element annotation of grass carp genome, “unclassified LTR” and “unclassified SINE” belong to long terminal repeats (LTRs) and short interspersed nuclear element (SINE), respectively. They were not assigned with a clear sub-class [54].

2.3.2. SINE Element rnd-3_family-293 Is Specifically Expressed during Embryonic Development

The embryonic transcriptome of grass carp largely maintained the content of maternal RNA before the activation of the zygotic genome (T2–T6), and the amount of various RNA molecules was reasonably steady (Figure 5A). Transposons’ expression was rigorously restricted at the T2–T5 (ovum to 64-cell) stage compared to other genes, but was clearly expressed at the T6 stage (256-cell, Figure 5B). At the moment, all transposon el-

ements expressed belonged to *rnd-3_family-293*. These transcripts were only expressed during embryonic development (T6–T16, Figure 5C), and the patterns of expression were strikingly consistent.

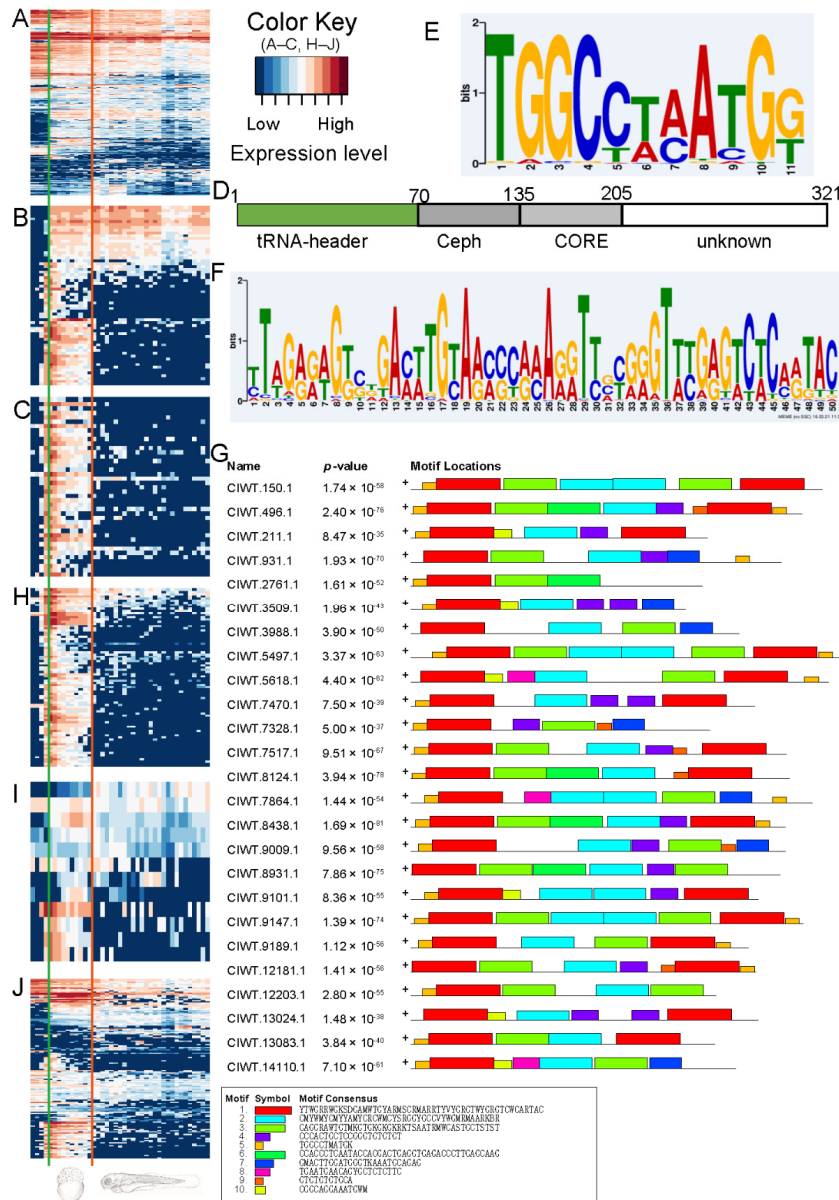


Figure 5. The grass carp transposons that have been activated. (A) Expression of 1000 genes chosen at random. The genes are listed along the rows, while the samples are listed down the columns. The samples are laid here according to the chronological order of their development. The samples labeled T6 (256 cells, left) and T7 (Dome, right) are found on both sides of the green line, whereas the samples labeled T16 (hatching, left) and T17 are found on both sides of the red line (3 days after fertilization, right). The samples that were obtained before zygotic genome activation are found to the left of the green line, and the samples that were collected between hatching and the end of the study are found to the right of the red line (including juvenile fish and 9 tissues). The order of the samples in (B), (C), (H), and (I) and (J) is the same as the order of the samples in (A); (B) Expression of all transposons; (C) Expression of transposons of *rnd-3_family-293*; (D) The diagram displays the structural pattern of *rnd-3_family-293*; (E) The motif 5; (F) The motif 1; (G) The structure pattern displays the motifs annotated by MEME for the transposons of *rnd-3_family-293* (25 were shown); (H) The expression of genes containing both motif 1 and motif 5; (I) The expression of genes containing only motif 1; (J) The expression of genes containing only motif 5.

2.3.3. Motif Analysis for Family rnd-3_family-293

The rnd-3_family-293 transposons have a tRNA-derived head sequence, a body with a 65 bp Ceph domain [58], a 70 bp CORE domain [59], and a 116 bp tail designated as unknown domain (Figure 5D). We examined the motifs of the 41 expressed transposons in rnd-3_family-293, and the large majority of these sequences contained motif 1 (41/41) and motif 5 (30/41) (Figure 5E,F), with the two motifs positioned in the upstream region of the tRNA header (Figure 5G).

2.3.4. Sequences Possessing Both Motif 1 and Motif 5 Were Specifically Expressed during Embryonic Development

We analyzed the transcriptome for transcripts related with motif 1 and motif 5 to see if they could be regulatory sequences. There was a strong correlation between the expression patterns of rnd-3_family-293 and those of motif-related transcripts (transcripts containing both motifs) (Figure 5C,H); transcripts containing just motif 1 or only motif 5 did not show this expression pattern (Figure 5I,J). We performed qPCR on eight motif-related genes and eight rnd-3_family-293 genes. CIWT.15163.1 and CIWT.15487 were predicted as tRNAs using Rfam for two of the eight motif-related genes, all of which are lncRNAs (Supplemental Table S6). As predicted by high-throughput sequencing, the qPCR data showed that motif-related transcripts and the rnd-3_family-293 transposon had different expression patterns from T6 to T8, with the maximum level of expression occurring at T7 and very low levels at other stages (T2, T5, T20) (Supplemental Figure S9).

2.4. Alternative Splicing Events in Grass Carp

2.4.1. Distributions of Grass Carp Alternative Splicing Events

In alternative splicing analysis, only transcripts supported by SMRT sequencing data were included to ensure the correctness of the results. A total of 5339 alternative splicing events were discovered in our study, belonging to 7 types: alternative 3' splice site (A3), alternative 5' splice site (A5), alternative first exon (AF), alternative last exon (AL), mutually exclusive exons (MX), retained intron (RI), and skipping exons (SE) (Figure 6A). Among them, SE (855, 24.72%), AF (501, 14.48%), A5 (664, 19.20%), RI (729, 21.08%), and A3 (604, 17.46%) were the most common type of alternative splicing (Figure 6A,B, Supplemental Table S7).

The composition ratio of alternative splicing types of grass carp samples from gastrulation to juvenile stage (GH, larva, juvenile) is close to the overall ratio (all) (Figure 6B). While the composition ratio in the ZA stage samples was distinct, the proportion of RI was much lower than in other stages of development (Figure 6B). However, the compositional ratio of ZA resembled that of ZA.zf (zebrafish ZGA stage) (Figure 6B).

2.4.2. RI Events Changes Dramatically during T6–T7

The Ψ value was used to describe the expression preference for splicing events. When $\Psi = 1$, all expressed transcripts are form 1; when $\Psi = 0$, all expressed transcripts are form 2; when $\Psi = 0.5$, two types of transcripts are equally expressed (Figure 6A, see Methods).

The distribution of Ψ values for AL and AF varied widely across all developmental time series samples (T2–T20), but these difference were not statistically significant due to the small number (<40) of detected alternative splicing events for both events (Supplemental Figure S10, Supplemental Table S7). The distribution of the value of splicing patterns A3, A5, AF, and SE altered from flat to narrow from before to after zygotic genome activation (T2–T6), indicating that the two kinds of alternatively spliced transcripts were expressed more uniformly in the embryos after zygotic activation. The most substantial changes in RI occurred during embryonic development, demonstrating a considerable rise in the mean value of during zygotic activation (Figure 6C). According to the *p*-value distribution of RI, its variation in embryonic phases was also substantially greater than that of other splicing types (Figure 6D). The most significant changes of RI events occurred

during T6–T7 ($p = 5.56 \times 10^{-37}$, the largest outlier in Figure 6D), i.e., from the 256-cell stage to the Dome stage, which corresponds exactly to the full activation of the zygotic genome.

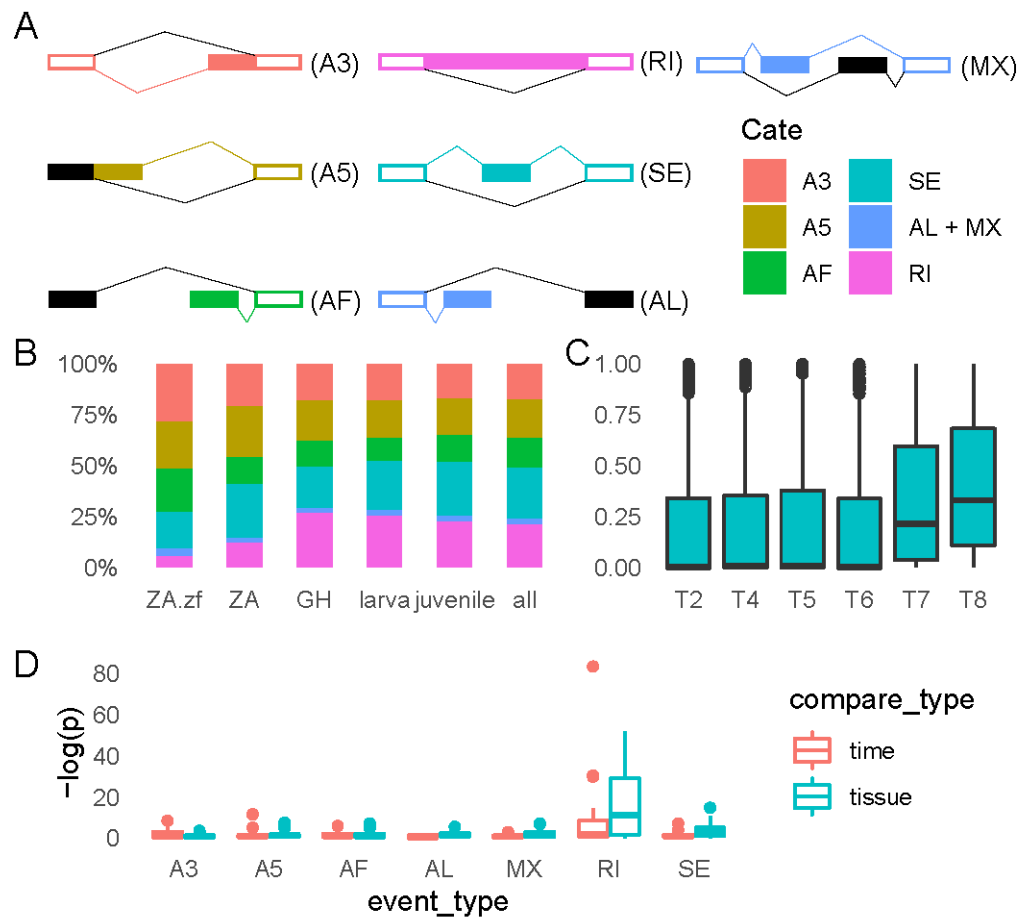


Figure 6. An examination of the various possibilities for splicing. (A) An example of a splicing pattern that has seven alternative splicing events that are fundamentally distinct from one another. In the diagram, the two distinct methods of splicing are indicated by the colors color and black, respectively. These colors represent the two possible outcomes of the process. The exons that are shared by both of the possible splicing types are denoted by the colored boxes that are otherwise empty. When the Ψ value is calculated, the colored splicing forms represent form 1; (B) The ratio of various alternative splicing events expressed at different stages. The ZA.zf indicates the ZGA stage of zebrafish [32]. Samples of grass carp were collected when the fish were in the ZGA stage (T2–T8), the GH stage (T9–T16), and the larva stage (T17–T20), respectively, as indicated by the letters ZA, GH, and larva, which are denoted by those letters. The statistical analysis was carried out on each and every sample of grass carp; (C) The variations in the values Ψ of alternative splicing events of the RI type that occurred during the ZA period; (D) A description of the distribution of the p -value. Classification of the comparison: tissue shows that the comparisons were carried out between 9 tissue samples pairwise and time indicates that the comparisons were carried out between adjacent time intervals in the time series samples from T2–T20.

2.4.3. Characteristics of RI Events during T6–T7

Differences calculated by subtracting the Ψ value in T6 (256 cells) from the Ψ value in T7 (dome) for the same splicing event reflects the change in expression preference of splicing events during development. The distribution of differences for RI alternative splicing events was significantly right-skewed compared to other events: that is, in the T7 period, more intron-retained (RI) transcripts were expressed. Based on this distribution, we defined alternative splicing events with a Ψ difference greater than 0.07 as “over” events, and others were named with “normal” events (Figure 7A).

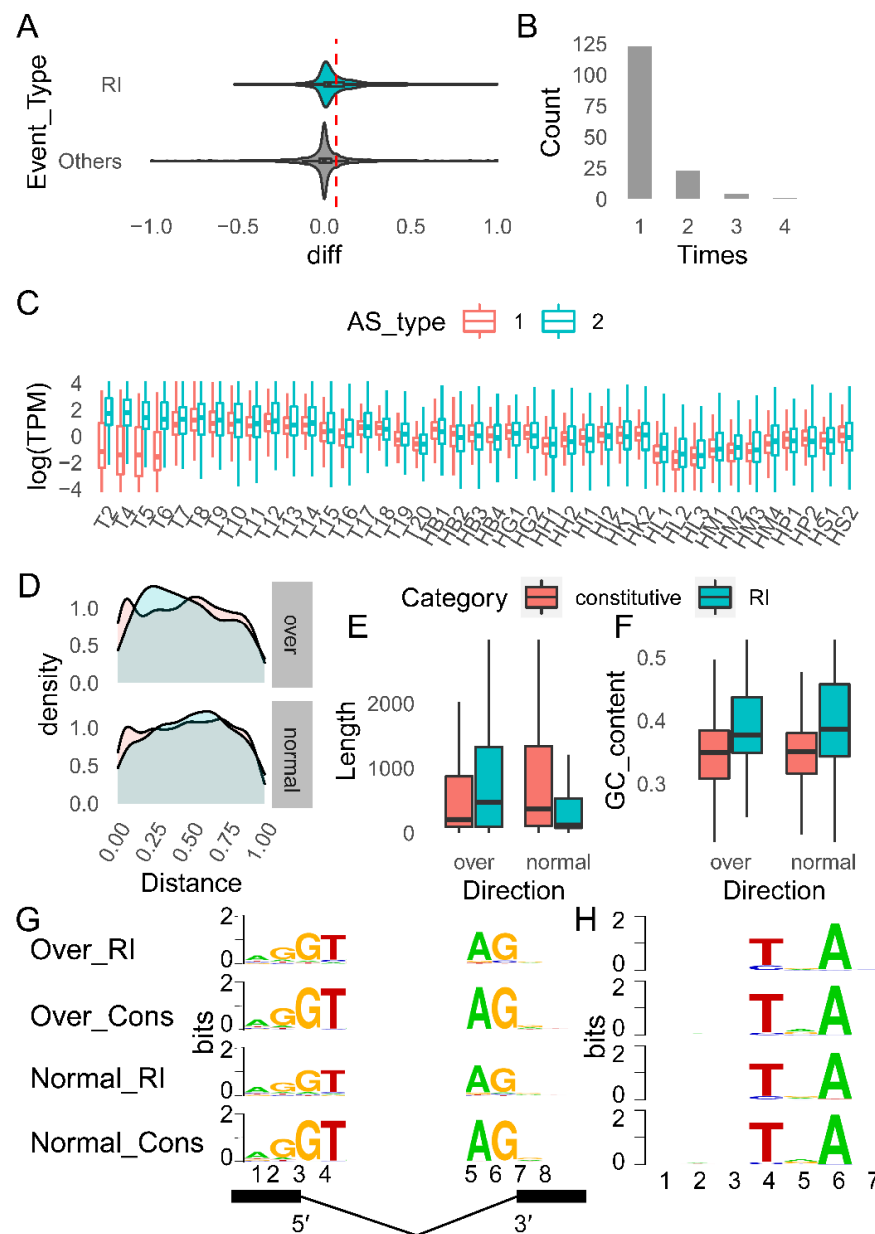


Figure 7. The distinguishing features of the transition from T6 (Dome) to T7 in the RI alternative splicing (256 cells). **(A)** The difference between the Ψ values of RI events ($\Psi(T7) - \Psi(T6)$) exhibited an obvious right-skewed distribution. The right side of the red dotted line (0.07), which was classified as an “over” event (which will be referred to as “over” later), and the left side of the line, which was defined as a “normal” event (which will be referred to as “normal” hereafter); **(B)** The total number of RI events that have taken place in a gene; **(C)** The expression ratio of the two different splicing forms that have taken place in RI events. The transcripts with the intron kept are denoted by the form “1,” while the transcripts with the intron spliced out are denoted by the form “2.”; **(D)** The position distribution of alternative introns (red) and constitutive introns (blue) on the gene (relative to the transcriptional start site); **(E)** The differences in GC content between alternative introns and constitutive introns; **(F)** The differences in length between alternative introns and constitutive introns; **(G)** The motif of the splice site. **(H)** Motifs of branch-point sequences are also referred to (BPS). Alternative introns of “over” transcripts are referred to as “Over_RI,” while constitutive introns of “over” transcripts are referred to as “Over_Cons.” Similarly, alternative introns of “normal” transcripts are referred to as “Normal_RI,” while constitutive introns of “normal” transcripts are referred to as “Normal_Cons”.

Among the “over” events, 151 genes encountered 184 RI events, with the majority of genes experiencing only one RI event and only 28 experiencing RI events greater than or equal to two times (Figure 7B). Among these genes, intron-retained isoforms tended to have modest levels of expression during T2–T6, but their expression levels began to grow after T7. In contrast, the expression of transcripts with introns spliced out did not vary much (Figure 7C). The three GO terms, nucleic acid binding, cellular metabolic process, and gene expression, were significantly enriched in the “over” event-related transcripts (hence referred to as “over” transcripts), indicating that these three GO terms were the specific functions of the “over” transcripts (Supplemental Tables S8 and S9).

In terms of intron length, GC content, branch-point sequence (BPS sequence), and other factors that may affect intron splicing, we compared alternative introns (introns engaged in RI events) to constitutive introns (introns that are always spliced out). Alternative introns in “over” transcripts were shown to be more likely to be located at the 5' position of the gene, whereas alternative introns in “regular” event-related transcripts (hereinafter referred to as “normal” transcripts) were found to be more likely to be located in the center of the gene (Figure 7D). However, the positions of the two types of introns were not significantly different (Supplemental Table S10). More specifically, the alternative introns of the “over” transcripts were longer than the constitutive introns of the “over” transcripts ($p = 0.03$) and were also longer than the alternative introns of the “normal” transcripts ($p = 3.04 \times 10^{-9}$), while the alternative introns of the “normal” transcripts were shorter than the constitutive introns of the “normal” transcripts ($p = 1.30 \times 10^{-34}$) (Figure 7E, Supplemental Table S10).

Constitutive introns showed a similar GC content distributions for both “over” and “normal” transcripts (Figure 7F, $p = 0.17$, Supplemental Table S10). The GC content of alternative introns was higher than that of constitutive introns in both types of transcripts, among which the GC content of alternative introns of “normal” transcripts was the highest and was more dispersed (over: $p = 2.56 \times 10^{-16}$; normal: $p = 6.56 \times 10^{-39}$). However, the distribution of GC content of alternative introns was not significantly different between “normal” and “over” transcripts ($p = 0.05$, Figure 7F, Supplemental Table S10).

The likelihood of a splice site being used can be characterized by splice junction strengths [60]. The splice junction strengths of alternative introns were significant weaker than of constitutive introns (5' splice site, 5'ss: $p = 2.04 \times 10^{-44}$, 3' splice site, 3'ss: $p = 2.49 \times 10^{-29}$, Figure 7G, Supplemental Figures S11 and S12, Supplemental Table S10). In the comparison of the splice site strength of introns between “over” and “normal” transcripts, no difference was found in constitutive introns (5'ss: $p = 0.11$, 3'ss: $p = 0.14$), while significant differences were found in alternative introns (5'ss: $p = 5.47 \times 10^{-3}$, 3'ss: $p = 0.04$). In particular, the splice junction strengths of alternative introns in “normal” transcripts were weaker (Figure 7G, Supplemental Figures S11 and S12, Supplemental Table S10).

The BPP software predicted that the BPS sequence of grass carp is NNNYNAN (Y denotes pyrimidine, N denotes any base). The BPS of alternative introns was lower than that of constitutive introns ($p = 0.05$). However, the BPS scores of alternative introns did not differ between “over” and “normal” transcripts (Figure 7H, Supplemental Figure S13, Supplemental Table S10).

2.4.4. Distribution of Splicing Isoforms in Nine Tissues and the Tissue Specificity of RI Alternative Splicing

The number of observed alternative splicing events in the 9 sequenced tissues of juvenile fish was positively correlated with the number of expressed genes (Supplemental Figure S14). The most expressed genes and alternative splicing events were found in the brain and gills, while the least were found in the liver and muscle. The ratio of alternative splicing events expressed in each tissue was relatively stable; the four categories of most expressed alternative splicing events were A3, A5, RI, and SE (Supplemental Figure S15, Supplemental Table S7).

We conducted pairwise comparisons of the distribution of Ψ values for alternative splicing across juvenile tissues, and the results differed slightly from those in developmental time series samples, as stated above. The results of pairwise tissue compar-

isons revealed that both RI and SE have highly tissue-specific expression (Figure 6D). Muscle was the most distinct among them in terms of RI alternative splicing events (Supplemental Figure S16A,B). The expression tendency of RI alternative splicing events differs most significantly between muscle and skin tissues ($p = 2.95 \times 10^{-23}$, Supplemental Figure S16C). Compared with in skin, the “upregulated” RI genes in muscle, which tended to express more intron-preserved isoforms in T7 than in T6 ($\Psi(T7) - \Psi(T6) > 0.07$, 94/813), were functional related to metabolism, especially energy metabolism. While the “downregulated” RI genes in muscle, which tended to express more intron-excluded isoforms in T7 than in T6 ($\Psi(T7) - \Psi(T6) < 0.07$, 262/813), were functional related to the regulation of mRNA synthesis. Spleen tissues showed the second-largest difference with muscle in terms of expression tendency of RI alternative splicing genes ($p = 1.32 \times 10^{-21}$, Supplemental Figure S16C). Compared with in spleen, the “upregulated” RI genes in muscle (79/716) were related to lipid metabolism, while the “downregulated” RI genes in muscle (251/716) were associated with the regulation of amino acids and nucleic acids (Supplemental Table S11).

3. Discussion

The transcriptome can be evaluated in three ways: (a) comprehensiveness—the fraction of all gene loci that are included; (b) exhaustiveness—the fraction of all transcripts from each locus that are known; and (c) completeness—the fraction of transcript models that cover the entire length of the physical RNA molecule, from start to end [61]. The transcriptome (gc.v2) assembled in this study is the most comprehensive (93.98 percent) compared to other non-model fish transcriptomes [28,32]. When compared to the original grass carp annotation (gc.v1), gc.v2 has a larger transcript/gene ratio (>2), indicating a better annotation of alternative splicing events and a higher level of transcriptome exhaustiveness. We anticipate obtaining a more complete molecular sequence due to the use of SMRT sequencing with longer reads [19,62]. In fact, if the median length of transcripts is used as a metric of transcriptome completeness, gc.v2 (3281 bp) outperforms gc.v1 (1642 bp) (Supplemental Table S2).

We created a dynamic picture of the grass carp transcriptome from a larger field of view for the first time under the condition that the transcriptome’s quality is ensured. This is a valuable resource for gaining a better understanding of fish development and tissue differentiation. The differences between samples in developmental time series are primarily represented in changes in the cell cycle, which is the most significant KEGG pathway enriched in the PC1-related transcripts defined in PCA analysis based on all samples. The basic fact related with this discovery is that early embryonic development of fish undergoes fast cell division, with a very short cell cycle, but late development embryos and adult cells have a longer cell cycle [63]. Second, in the same PCA analysis, mid-developmental transition samples differed the most from differentiated tissues and cleavage-stage embryos in terms of PC2 (Figure 2). This aligns nicely with the hourglass model or the ground zero concept: the embryo experiences a process of de-differentiation similar to the induction of somatic cells into pluripotent stem cells from early to middle embryonic development. The expressed genes are more conserved in the mid-embryonic stage, the telomere length is longer, and the entropy of the embryonic structure is lower, indicating that the stemness of the cell is stronger at this time, which is more like the starting point of the organism’s intergenerational rejuvenation [26,64,65].

Although fish are regarded to be lower vertebrates, the brain is highly differentiated and distinct from other organs in terms of tissue differentiation. The linkages between tissue and its origin germ layer are almost evident from the standpoint of development [66], which is reflected in the overall gene expression profiling. Except for highly differentiated brain tissue, tissues developing from the same layers clustered together in the PCA analysis based on all tissue samples. It suggests that tissues coming from the same germ layer may include a large number of genes/transcripts that express similarly: the aggregated muscle tissue (HM) and heart tissue (HH) both contain a large amount of muscle components

and are derived primarily from mesoderm; the liver (HL), spleen (HP), intestine (HI), and kidney (HK) are derived primarily from endoderm; and the skin in the upper right is derived from ectoderm. The origin of the gills in fish has been questioned. The gills of jawless vertebrates are thought to originate from the endoderm, while those of jawed vertebrates (cartilaginous and bony fishes) are thought to come from the ectoderm [67]. Using in situ hybridization and destiny mapping techniques, a recent study found that gill filaments in cartilaginous fish arise from the endoderm [68]. Our findings reveal that the gill is placed in the center of the endoderm and ectoderm, highlighting the complexity and specificity of this tissue's origin.

In vertebrate genomes, transposons and transposon-derived sequences are common [69–71]. Transposon elements, which are a major driving force in the production of new genes, are frequently found in expressed genes in the form of repeated sequences [72]. However, because of the potential harm to genome integrity, transposons' "transposition ability" is normally blocked [73]. This can be seen in the enormous disparities between the genome's broadly distributed fixed transposons and the modest number of expressed transposons. Transposon DNA/hAT, for example, is thought to be dormant in all species [74], yet it accounts for 12.77 percent of repetitive elements in the grass carp genome and has not been discovered for transcription. However, transposon transcription is not always suppressed. Recent research has revealed that retrotransposons are preferentially activated during mammalian embryonic development, particularly during the process of zygotic gene activation (epigenetic reprogramming) [4–7,75]. Transposon elements were found to be involved in ZGA events in a study of mouse single-cell RNA sequencing, with LTR associated with the minor wave of zygotic activation and SINE associated with the major wave of zygotic activation, where a large number of essential genes begin to be transcribed [76].

The expressed transposons in grass carp were family-specific and had nothing to do with their abundance on the genome. The transposons of an unclassified SINE family (rnd-3_family-293) were found to be preferentially expressed during embryonic development. More intriguingly, they were activated simultaneously during the 256-cell phase, when the majority of zygotic genes are not yet activated. This shows that fish retrotransposons, like those of mammals, may function during zygotic stimulation of embryonic development. Furthermore, we discovered that the majority of these transposons contained a combination of two patterns (motif 1 and motif 5) in their 5' tRNA header, and the genes containing the two motifs displayed an expression pattern similar to that of these transposons. Given that SINE transposons are transcribed by RNA polymerase III and the binding site for promoter recognition is located within the sequence [77], we believe that this unique expression pattern of rnd-3_family-293 is regulated by a distinct mechanism, which may initiate expression via RNA polymerase III recognition of motif 1 and motif 5.

Gene splicing is involved in many biological processes, including development, cell and tissue differentiation, and disease occurrence [9,78]. As a result, systematic mining is very important. We found a total of 5339 alternative splicing events using SMRT RNA-seq technology, with SE (24.72 percent) and RI (21.08 percent) having the largest proportions, which was consistent with the findings of two SMRT-based fish transcriptome studies [31,37]. The percentage of RI events fluctuates substantially during development and is quite low during the grass carp zygotic activation (ZA) stage. This finding corresponds to the transcriptome during zygotic activation in zebrafish [32]. RI splicing events occur far more frequently during other developmental stages (GH, larva, and juvenile). To our knowledge, no other fish species has had such a significant fluctuation in RI events during development. As RI events are demonstrated to be functional, interest in this topic in animals is growing rapidly [13,14,79,80].

In our research, we discovered that RI may be important in both embryonic development and tissue specificity maintenance in fish. The modifications were most pronounced during the maternal to zygotic transition (MZT, 256-cells-Dome), when the number of RI alternative splicing events rose considerably and more isoforms with intron retention were produced. The over genes, the genes with an upregulated Ψ value ($\Psi(T7) - \Psi(T6) > 0.07$,

genes expressing more intron-retained isoforms) during the zygotic activation stage exhibited some interesting features. Compared with the normal genes, the molecular functions of GO terms of over genes were significantly enriched for nucleic acid binding. In general, nucleic acid binding acts as a regulator in life activities.

We are particularly interested in the regulatory aspects of RI events, in addition to their function. In general, alternative introns have lower splicing signals to prevent spliceosome formation and hence remain in the mature RNA molecule. The placement at the 3' end of the gene, shorter intron length (shorter GT-AG spacing), higher GC content, weaker splice site strength, and lesser BPS sequence strength were the typical weak splicing signals (Table 3). However, the alternative introns we uncovered have some unique characteristics in terms of intron length and splice site strength. In general, alternative introns are short, which makes intron skipping easier during spliceosome formation [80,81]. In a study of chicken embryos, for example, alternative introns were shorter in RI occurrences [79]. Alternative introns in normal events meet this description in our study, whereas introns in over events are longer than constitutive introns. Furthermore, in our investigation, the splicing strength of the alternative introns of the over transcripts was greater than that of the normal transcripts. These distinguishing characteristics suggest that RI alternative splicing events have distinct purposes and meanings during the maternal zygotic transition, as well as a distinct regulatory mechanism.

Table 3. Characteristics of alternative introns in RI (compared with constitutive introns).

Category	Over Events	Normal Events	Classic Features [80]
Relative position	5' (not significant)	3' (not significant)	3'
Length	Long	Short	Short
GC content	High	High	High
Strength of splice site	Weak	Weak	Weak
Strength of BPS	Weak	Weak	Weak

4. Materials and Methods

4.1. Sample Collection

Each sample was taken from a population consisting entirely of the offspring of a single set of grass carp parents (bred in Xingfu village, Huanggang, China in 2017). During the process of collecting the samples, the information regarding the samples, such as the amount of time that has passed after fertilization, the type of tissue, and the morphophysiological properties, was recorded (Table 1). The fertilized eggs were continually viewed using a dissecting microscope (OLYMPUS, Tokyo, Japan) after artificial insemination, and the embryonic developmental phases were recognized by reference to the zebrafish development diagram [82].

For the samples taken between T2 and T19, a total of 20 fertilized eggs, embryos, and juveniles were obtained at each time point. The T20 sample consisted of a single fish that was replicated twice for biological analysis. One month after fertilization, sampling was continued every 30 to 60 days. During the sampling process, the grass carp were euthanized with a high dosage of eugenol. This was followed by a quick dissection and the collection of tissue samples, with two to three biological replicates at each time point.

A TRIzol (Invitrogen, Carlsbad, CA, USA) solution was added to the samples, and they were completely crushed in an OSE-Y40 homogenizer (Tiangen, Beijing, China) to release the RNases, which were then inactivated by further contact with a TRIzol solution. The ethics committee gave their blessing to the entire experiment.

4.2. Preparation of Sample for Library Construction

The manufacturer's instructions for utilizing TRIzol reagent were followed when obtaining total RNA. The Qubit[®] RNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, Gaithersburg, MD, USA) was used to measure the RNA concentration. Using the Agilent 2100 technology, RNA integrity was determined (Agilent Technologies, Santa Clara,

CA, USA). It was decided to use samples with RINs greater than 9 for the SMRT-based libraries and greater than 7 for the ssRNA-seq libraries in order to get the most accurate results. Those samples that met the requirements were kept in a -80°C refrigerator.

4.3. Library Construction and Sequencing

4.3.1. ssRNA-Seq

For each tissue sample, a library was generated and sequenced. For each sample point, the qualified RNA from biological replicates were mixed in equal amounts. Three micrograms of RNA was used for each sample in the library construction.

For the first step, the RibozeroTM rRNA Removal Kit (Epicentre, San Antonio, TX, USA) was employed. An Illumina NEBNext UltraTM Directional RNA Library Prep Kit was used to build the sequencing library, and the manufacturer's instructions were followed throughout the process. Finally, the Illumina HiSeq 4000 PE System was used to sequence libraries of sufficient quality.

With the help of IlluQC [83], adapter sequences and low-quality sequences that met the Q30 criteria were eliminated. When either the amount of nitrogen (N) in any read exceeded 10 percent of the total number of bases in the read or the amount of bases with a quality rating of 5 or lower exceeded 50 percent, the paired reads were discarded. FastQC and MultiQC were utilized in order to carry out quality information statistics on the aforementioned data [84].

4.3.2. SMRT-Based RNA-Seq

Before being utilized to generate sequencing libraries, the qualifying RNAs that were extracted from embryonic development samples were mixed together in equal amounts. In order to construct the sequencing library using the samples collected after hatching, the RNA samples obtained from the same tissue at various time points were combined in equal amounts to create the tissue samples, and then the tissue samples themselves were combined in equal amounts to create the sequencing library.

The SMARTerTM PCR cDNA Synthesis Kit (PacBio, Menlo Park, CA, USA) was used to do reverse transcription on a mixed RNA sample that was approximately 1 microgram in weight. The PCR product was then utilized in the construction of the SMRTbell library through the utilization of PacBio's SMRTbell template pre kit 1.0. The SMRTbell template was annealed to the sequencing primer, bound to polymerase, and then sequenced using PacBio's Sequel Binding Kit 3.0 on the PacBio Sequel platform.

4.4. Transcriptome Assembly

4.4.1. Generation Transcriptome for ssRNA-Seq Data and SMRT-Based RNA-Seq Data Separately

We processed the ssRNA-seq data with the classical process: aligned using Hisat2 (2.0.4) before assembled by StringTie (1.3.1c) [85]. The expression level was counted with Salmon (1.4.0) [86].

SMRT sequencing data was analysis with SMRT-link package (5.1.0) and Iso-Seq (Isoform Sequencing) procedure (<https://www.pacb.com/support/software-downloads/> accessed on 17 September 2018). First, the command "ccs" in the SMRT-link toolkit was used to obtain circular consensus sequences (CCS); then, "pbtranscript classify" was used to obtain full-length sequences (FLNCs); finally, "pbtranscript cluster" was used to cluster FLNCs to obtain high-quality transcripts. The software gmap (10 June 2019) was used to align high-quality transcripts back to genome, and only results with a mapping rate of over 85% were retained [87]. The "collapse_isoforms_by_sam.py" script in the cDNA_Cupcake package was used to generate a GFF file from the alignment results, with the parameter set as `-fq -dun-merge-5-shorter -c 0.95 -i 0.9` (https://github.com/Magdoll/cDNA_Cupcake, accessed on 18 September 2018). Transcripts with full-length CCS support numbers less than 2 and/or degraded from the 5' side were removed from the generated GFF file.

4.4.2. Merge ssRNA-Seq Transcriptome and SMRT Transcriptome

The ssRNA-seq transcriptome and the SMRT transcriptome were combined, and then gffread (gffread -M -K) was used to eliminate any duplicated transcripts that were found [88]. Using the program “gffcompare”, we compared the merged transcriptome with the ssRNA-seq transcriptome and the SMRT transcriptome in their own right. This allowed us to acquire the supporting information, which indicated whether or not the transcript was detected by both technologies. Salmon (version 1.4.0) was utilized to determine the amount of expression, and transcripts whose values were lower than 1 TPM were eliminated [86]. Two different filtering techniques were carried out to improve the quality of the assembled transcripts even further:

1. Eliminate any transcripts that might have been fragmented. It is quite simple to confuse fragmented mRNA transcripts with newly assembled lncRNAs. In order to locate these transcripts, we first determined the distance between the transcript and the border of the scaffold and then checked to see if the transcript was supported by SMRT data. More than 95 percent of introns were less than 7875 base pairs, as shown in the length distribution of introns in all transcripts (Supplemental Table S12). It was decided that the transcripts whose distance from the scaffold edge was less than this value and which were not supported by SMRT data were fragmented transcripts. As a result, these transcripts were removed from the analysis.
2. Remove possible precursor mRNAs. It is anticipated that the RNA-seq technology will discover a high number of mRNA fragments that have not been spliced or are in the process of being spliced. Because these fragments, unlike mature mRNA, exactly correspond to the full continuous region of the genome, it is common practice to confuse them with lengthy single-exon transcripts; nonetheless, they are eventually discovered to be long non-coding RNA. In order to get rid of the negative effects of such transcripts, we made the assumption that if a gene contains multi-exon transcripts as well as single-exon transcripts, the single-exon transcripts are very certainly precursors and should be removed as well.

4.5. Annotation of Transcriptome

Two tools, CPAT and CPC2, were utilized to predict the RNA coding ability in order to improve the accuracy of lncRNA prediction. In particular, the CPAT zebrafish model was used for prediction [89]. When both prediction findings from two software were non-coding, a transcript was designated as lncRNA; mRNA when both predictions were coding; and TUCP (transcripts of uncertain coding potential) when the predictions were inconsistent.

If the class of all of a gene’s transcripts is the same, the gene will be designated as the class of its transcripts; otherwise, the gene will be labeled “misc”.

A gene’s functional annotation is assigned based on the annotation result of its longest transcript. The GO function was annotated using blast2go (2.5.0), and the BLAST database was chosen from the nr library [90]. KO annotation was performed using the KAAS system [91] was utilized for KO annotation, and the species used for alignment were human (has), mouse (mmu), zebrafish (dre), common carp (ccar), medaka (ola), and Atlantic salmon (sasa). In KAAS annotation, the bidirectional best hit (BBH) approach was utilized. The RNA categorization (family) was annotated using cmscan (1.1.2) to search the Rfam database (14.1) [92,93].

4.6. Completeness of Transcriptome

The loach SMRT-based transcriptome [28] was obtained from the links mentioned in their paper. The SMRT-based transcriptome of *Onychostoma brevibarba* was obtained from the publication’s supplementary material [36]. The completeness of all transcriptomes was determined using BUSCO software (3.1.0) and OrthoDB’s actinopterygii (ray-finned fish) as analytic database [94].

4.7. Principal Component Analysis

The “prcomp” function in R was used to perform principal component analysis (PCA). The factoextra package was used to calculate the “variable contributions” of the transcripts to each principal component (PC). The PC-related transcripts were those that were associated with the PC with the highest contribution value. The plotly program was used to create interactive plots.

4.8. Annotation and Analysis of Transposons

4.8.1. Identification, Annotation, and Relationship Analysis of Transposons

Assembled transcripts and transposable elements (TEs) in the grass carp genome (Supplemental Figure S17) were used to calculate the overlap information [54]. Expressed transposons are transcripts with overlaps of more than 90 percent of repeat elements and 90 percent of the transcript itself. TE-associate transcripts are those whose repeating element overlap exceeds 90% but does not exceed 90% of the transcript’s own overlap. Transposons having N sequences within 20 bp upstream and downstream were specifically targeted for deletion. In order to annotate the structure of short interspersed elements (SINEs), we consulted the SINE Base database [95]. EBI clustal omega [96] was used to analyze the relationships between different sequences.

4.8.2. Motif Analysis

Motif was predicted de novo using MEME (5.3.3) [97]. FIMO (5.3.3) was used to search for motifs in all transcripts with MEME-1 starting coordinates less than 50 and MEME-5 starting coordinates less than 40 [98]. MAST (5.3.3) was used to search and visualize motif combinations [99].

4.9. Validation Using qPCR

The NEB E. Coli Poly(A) (#M0276S) kit was used to add Poly(A) tail. Genomic DNA removal and first strand synthesis were performed using the HiScript[®] III RT SuperMix kit for qPCR (+gDNAWIper) (Vazyme, China). ChamQ SYBR qPCR Master Mix (Vazyme, China) kit and CFX96 Real-Time PCR Detection System (Bio-Rad, USA) were used to perform RT-qPCR. The RPS27 gene (small subunit ribosomal protein S27e) was selected as the internal reference [100,101]. Annealing conditions of internal reference and target gene primers were recorded (Supplemental Table S6).

4.10. Alternative Splicing Analysis

The suppa software [102] was used to identify alternative splicing events and calculate Ψ values. The transcripts participating in the event must have a minimum total expression of more than 1 TPM. The following is the formula for determining the value:

$$\Psi = \sum_{k \in F_1} TPM_k / \sum_{j \in F_1 \cup F_2} TPM_j \quad (1)$$

F_1 is the set of transcripts spliced following the form 1 in a splicing event, and F_2 is the set of transcripts spliced following the form 2 in the same event. Because only simple splicing events were analyzed (no splicing event combinations were included), there were only two splicing forms for a splicing event. The Wilcoxon signed-rank test was used to compare the distributions of adjacent time points and tissues. All time points from T2 through T20 are considered adjacent. Tissue comparisons were performed using samples from the latest two sampling time points (134 dpf and 197 dpf).

The weblogo online application [103] was used to view the patterns of splice sites. The Maximum Entropy Model (MaxEnt) [104] was used to forecast the splice site strength. The software BPP [105] was used to forecast the branch-point sequence (BPS). The Mann–Whitney U test was used to assess distribution differences in splice site strength, BPS scores, intron length, and GC content.

4.11. Other Analysis

Fisher's test was used for KEGG enrichment analysis, while the R package "GOStats" was used for GO enrichment analysis [106]. Other data analysis was performed on the Linux (CentOS 7) platform using R (4.0.2), Python (3.6.11), and other software. The ggplot2 package was used for data visualization. Unless otherwise specified, the default parameters of the bioinformatics program used in this work were used during data processing.

5. Conclusions

Through continuous, high-frequency sampling, we obtained a high-quality reference transcriptome and dynamic transcriptional map of grass carp, an important resource for fish developmental study. On this basis, we systematically analyzed the changes of lncRNAs and alternative splicing events in grass carp development.

The zygotic genome activation (ZGA) is an extremely important research topic in fish developmental biology, but its mechanism has not been fully elucidated yet. Based on the reference transcriptome, we discovered a novel family of short interspersed nuclear element and its potential regulatory motifs associated with fish ZGA. Meanwhile, it was found that retained intron (RI) alternative splicing events dramatically change during ZGA, and this type of RI events are different with classical RI events in their regulatory characteristics. In addition, two types of alternative splicing events, RI and skipping exons (SE), are associated with tissue specificity maintenance in fish.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms231911547/s1>.

Author Contributions: Conceptualization, Y.D., M.S. and X.-Q.X.; software, Y.D. and W.Z.; validation, Q.Z.; formal analysis, Y.C.; investigation, Y.D. and Y.J.; resources, Y.D., Y.J. and Y.C.; data curation, W.Z.; writing—original draft preparation, Y.D. and Q.Z.; writing—review and editing, M.S. and X.-Q.X.; visualization, Y.D. and Q.Z.; supervision, X.-Q.X.; project administration, M.S. and X.-Q.X.; funding acquisition, M.S. and X.-Q.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (Grant No. 2021YFD1200804), the National Natural Science Foundation of China (Grant No. 31801055), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA24010206).

Institutional Review Board Statement: This study was ethically approved (permission No. IHBL2017035) and also approved by the local Animal Ethical and Welfare Committee (approval date 11 January 2017).

Data Availability Statement: The sample information and sequencing data have been submitted to the National Genomics Data Center (NGDC, <https://ngdc.cncb.ac.cn/>), accessed on 10 June 2022) with the accession number PRJCA010000.

Acknowledgments: We would like to thank the Wuhan Branch, Supercomputing Center, Chinese Academy of Sciences, for the support of data computing in this research work.

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

References

1. Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21. [[CrossRef](#)]
2. Beermann, J.; Piccoli, M.T.; Viereck, J.; Thum, T. Non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. *Physiol. Rev.* **2016**, *96*, 1297–1325. [[CrossRef](#)] [[PubMed](#)]
3. Ransohoff, J.D.; Wei, Y.; Khavari, P.A. The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 143–157. [[CrossRef](#)]
4. Gerdes, P.; Richardson, S.R.; Mager, D.L.; Faulkner, G.J. Transposable elements in the mammalian embryo: Pioneers surviving through stealth and service. *Genome Biol.* **2016**, *17*, 100. [[CrossRef](#)] [[PubMed](#)]

5. Fadloun, A.; Le Gras, S.; Jost, B.; Ziegler-Birling, C.; Takahashi, H.; Gorab, E.; Carninci, P.; Torres-Padilla, M.E. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **2013**, *20*, 332–338. [[CrossRef](#)] [[PubMed](#)]
6. Fu, B.; Ma, H.; Liu, D. Endogenous Retroviruses Function as Gene Expression Regulatory Elements During Mammalian Pre-implantation Embryo Development. *Int. J. Mol. Sci.* **2019**, *20*, 790. [[CrossRef](#)] [[PubMed](#)]
7. Bui, L.C.; Evsikov, A.V.; Khan, D.R.; Archilla, C.; Peynot, N.; Henaut, A.; Le Bourhis, D.; Vignon, X.; Renard, J.P.; Duranthon, V. Retrotransposon expression as a defining event of genome reprogramming in fertilized and cloned bovine embryos. *Reproduction* **2009**, *138*, 289–299. [[CrossRef](#)] [[PubMed](#)]
8. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415. [[CrossRef](#)]
9. Kim, H.K.; Pham, M.H.C.; Ko, K.S.; Rhee, B.D.; Han, J. Alternative splicing isoforms in health and disease. *Pflug. Arch.* **2018**, *470*, 995–1016. [[CrossRef](#)] [[PubMed](#)]
10. Bonnal, S.C.; Lopez-Oreja, I.; Valcarcel, J. Roles and mechanisms of alternative splicing in cancer—Implications for care. *Nat. Rev. Clin. Oncol.* **2020**, *17*, 457–474. [[CrossRef](#)]
11. McGuire, A.M.; Pearson, M.D.; Neafsey, D.E.; Galagan, J.E. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol.* **2008**, *9*, R50. [[CrossRef](#)] [[PubMed](#)]
12. Marquez, Y.; Brown, J.W.; Simpson, C.; Barta, A.; Kalyana, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* **2012**, *22*, 1184–1195. [[CrossRef](#)] [[PubMed](#)]
13. Wong, J.J.; Au, A.Y.; Ritchie, W.; Rasko, J.E. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *Bioessays* **2016**, *38*, 41–49. [[CrossRef](#)] [[PubMed](#)]
14. Braunschweig, U.; Barbosa-Morais, N.L.; Pan, Q.; Nachman, E.N.; Alipanahi, B.; Gonatopoulos-Pournatzis, T.; Frey, B.; Irimia, M.; Blencowe, B.J. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **2014**, *24*, 1774–1786. [[CrossRef](#)] [[PubMed](#)]
15. Ding, Z.; Tie, W.; Fu, L.; Yan, Y.; Liu, G.; Yan, W.; Li, Y.; Wu, C.; Zhang, J.; Hu, W. Strand-specific RNA-seq based identification and functional prediction of drought-responsive lncRNAs in cassava. *BMC Genom.* **2019**, *20*, 214. [[CrossRef](#)]
16. Yu, Y.; Zhou, Y.F.; Feng, Y.Z.; He, H.; Lian, J.P.; Yang, Y.W.; Lei, M.Q.; Zhang, Y.C.; Chen, Y.Q. Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. *Plant Biotechnol. J.* **2020**, *18*, 679–690. [[CrossRef](#)]
17. Zhou, Q.; Wan, Q.; Jiang, Y.; Liu, J.; Qiang, L.; Sun, L. A Landscape of Murine Long Non-Coding RNAs Reveals the Leading Transcriptome Alterations in Adipose Tissue during Aging. *Cell Rep.* **2020**, *31*, 107694. [[CrossRef](#)]
18. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)]
19. Wang, B.; Kumar, V.; Olson, A.; Ware, D. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Front. Genet.* **2019**, *10*, 384. [[CrossRef](#)]
20. Graveley, B.R.; Brooks, A.N.; Carlson, J.W.; Duff, M.O.; Landolin, J.M.; Yang, L.; Artieri, C.G.; van Baren, M.J.; Boley, N.; Booth, B.W.; et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **2011**, *471*, 473–479. [[CrossRef](#)]
21. Tan, M.H.; Au, K.F.; Yablonoitch, A.L.; Wills, A.E.; Chuang, J.; Baker, J.C.; Wong, W.H.; Li, J.B. RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome Res.* **2013**, *23*, 201–216. [[CrossRef](#)] [[PubMed](#)]
22. Brown, J.B.; Boley, N.; Eisman, R.; May, G.E.; Stoiber, M.H.; Duff, M.O.; Booth, B.W.; Wen, J.; Park, S.; Suzuki, A.M.; et al. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **2014**, *512*, 393–399. [[CrossRef](#)]
23. Gerstein, M.B.; Rozowsky, J.; Yan, K.K.; Wang, D.; Cheng, C.; Brown, J.B.; Davis, C.A.; Hillier, L.; Sisu, C.; Li, J.J.; et al. Comparative analysis of the transcriptome across distant species. *Nature* **2014**, *512*, 445–448. [[CrossRef](#)]
24. Hashimshony, T.; Feder, M.; Levin, M.; Hall, B.K.; Yanai, I. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* **2015**, *519*, 219–222. [[CrossRef](#)] [[PubMed](#)]
25. Boeck, M.E.; Huynh, C.; Gevirtzman, L.; Thompson, O.A.; Wang, G.; Kasper, D.M.; Reinke, V.; Hillier, L.W.; Waterston, R.H. The time-resolved transcriptome of *C. elegans*. *Genome Res.* **2016**, *26*, 1441–1450. [[CrossRef](#)] [[PubMed](#)]
26. Levin, M.; Anavy, L.; Cole, A.G.; Winter, E.; Mostov, N.; Khair, S.; Senderovich, N.; Kovalev, E.; Silver, D.H.; Feder, M.; et al. The mid-developmental transition and the evolution of animal body plans. *Nature* **2016**, *531*, 637–641. [[CrossRef](#)] [[PubMed](#)]
27. Owens, N.D.L.; Blitz, I.L.; Lane, M.A.; Patrushev, I.; Overton, J.D.; Gilchrist, M.J.; Cho, K.W.Y.; Khokha, M.K. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep.* **2016**, *14*, 632–647. [[CrossRef](#)]
28. Yi, S.; Zhou, X.; Li, J.; Zhang, M.; Luo, S. Full-length transcriptome of *Misgurnus anguillicaudatus* provides insights into evolution of genus *Misgurnus*. *Sci. Rep.* **2018**, *8*, 11699. [[CrossRef](#)]
29. Hu, F.; Fan, J.; Wu, C.; Zhu, M.; Zhou, Y.; Wang, S.; Zhang, C.; Tao, M.; Zhao, R.; Tang, C.; et al. Analysis of Chromosomal Numbers, Mitochondrial Genome, and Full-Length Transcriptome of *Onychostoma brevibarba*. *Mar. Biotechnol.* **2019**, *21*, 515–525. [[CrossRef](#)]
30. Luo, H.; Liu, H.; Zhang, J.; Hu, B.; Zhou, C.; Xiang, M.; Yang, Y.; Zhou, M.; Jing, T.; Li, Z.; et al. Full-length transcript sequencing accelerates the transcriptome research of *Gymnocypris namensis*, an iconic fish of the Tibetan Plateau. *Sci. Rep.* **2020**, *10*, 9668. [[CrossRef](#)]

31. Ren, L.; Yan, X.; Gao, X.; Cui, J.; Yan, P.; Wu, C.; Li, W.; Liu, S. Maternal effects shape the alternative splicing of parental alleles in reciprocal cross hybrids of *Megalobrama amblycephala* × *Culter alburnus*. *BMC Genom.* **2020**, *21*, 457. [[CrossRef](#)] [[PubMed](#)]
32. Nudelman, G.; Frasca, A.; Kent, B.; Sadler, K.C.; Sealfon, S.C.; Walsh, M.J.; Zaslavsky, E. High resolution annotation of zebrafish transcriptome using long-read sequencing. *Genome Res.* **2018**, *28*, 1415–1425. [[CrossRef](#)]
33. Mehjabin, R.; Xiong, L.; Huang, R.; Yang, C.; Chen, G.; He, L.; Liao, L.; Zhu, Z.; Wang, Y. Full-Length Transcriptome Sequencing and the Discovery of New Transcripts in the Unfertilized Eggs of Zebrafish (*Danio rerio*). *G3* **2019**, *9*, 1831–1838. [[CrossRef](#)] [[PubMed](#)]
34. Mohindra, V.; Dangi, T.; Chowdhury, L.M.; Jena, J.K. Tissue specific alpha-2-Macroglobulin (A2M) splice isoform diversity in Hilsa shad, *Tenualosa ilisha* (Hamilton, 1822). *PLoS ONE* **2019**, *14*, e0216144. [[CrossRef](#)]
35. Diao, J.; Yu, X.; Wang, X.; Fan, Y.; Wang, S.; Li, L.; Wang, Y.; Gai, C.; Ye, H.; Liu, H.; et al. Full-length transcriptome sequencing combined with RNA-seq analysis revealed the immune response of fat greenling (*Hexagrammos otakii*) to *Vibrio harveyi* in early infection. *Microb. Pathog.* **2020**, *149*, 104527. [[CrossRef](#)]
36. Zhang, X.; Zhou, J.; Li, L.; Huang, W.; Ahmad, H.I.; Li, H.; Jiang, H.; Chen, J. Full-length transcriptome sequencing and comparative transcriptomic analysis to uncover genes involved in early gametogenesis in the gonads of Amur sturgeon (*Acipenser schrenckii*). *Front. Zool.* **2020**, *17*, 11. [[CrossRef](#)]
37. Tian, Y.; Wen, H.; Qi, X.; Zhang, X.; Liu, S.; Li, B.; Sun, Y.; Li, J.; He, F.; Yang, W.; et al. Characterization of Full-Length Transcriptome Sequences and Splice Variants of *Lateolabrax maculatus* by Single-Molecule Long-Read Sequencing and Their Involvement in Salinity Regulation. *Front. Genet.* **2019**, *10*, 1126. [[PubMed](#)]
38. Feng, X.; Jia, Y.; Zhu, R.; Chen, K.; Chen, Y. Characterization and analysis of the transcriptome in *Gymnocypris selincuoensis* on the Qinghai-Tibetan Plateau using single-molecule long-read sequencing and RNA-seq. *DNA Res.* **2019**, *26*, 353–363. [[CrossRef](#)]
39. Mathavan, S.; Lee, S.G.; Mak, A.; Miller, L.D.; Murthy, K.R.; Govindarajan, K.R.; Tong, Y.; Wu, Y.L.; Lam, S.H.; Yang, H.; et al. Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet.* **2005**, *1*, 260–276. [[CrossRef](#)] [[PubMed](#)]
40. Rauwerda, H.; Pagano, J.F.; de Leeuw, W.C.; Ensink, W.; Nehrdich, U.; de Jong, M.; Jonker, M.; Spaink, H.P.; Breit, T.M. Transcriptome dynamics in early zebrafish embryogenesis determined by high-resolution time course analysis of 180 successive, individual zebrafish embryos. *BMC Genom.* **2017**, *18*, 287. [[CrossRef](#)] [[PubMed](#)]
41. Nepal, C.; Hadzhiev, Y.; Previti, C.; Haberle, V.; Li, N.; Takahashi, H.; Suzuki, A.M.; Sheng, Y.; Abdelhamid, R.F.; Anand, S.; et al. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Res.* **2013**, *23*, 1938–1950. [[CrossRef](#)] [[PubMed](#)]
42. Harvey, S.A.; Sealy, I.; Kettleborough, R.; Fenyes, F.; White, R.; Stemple, D.; Smith, J.C. Identification of the zebrafish maternal and paternal transcriptomes. *Development* **2013**, *140*, 2703–2710. [[CrossRef](#)] [[PubMed](#)]
43. Aanes, H.; Winata, C.L.; Lin, C.H.; Chen, J.P.; Srinivasan, K.G.; Lee, S.G.; Lim, A.Y.; Hajan, H.S.; Collas, P.; Bourque, G.; et al. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **2011**, *21*, 1328–1338. [[CrossRef](#)] [[PubMed](#)]
44. White, R.J.; Collins, J.E.; Sealy, I.M.; Wali, N.; Dooley, C.M.; Digby, Z.; Stemple, D.L.; Murphy, D.N.; Billis, K.; Hourlier, T.; et al. A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife* **2017**, *6*, e30860. [[CrossRef](#)] [[PubMed](#)]
45. Song, W.; Jiang, K.; Zhang, F.; Lin, Y.; Ma, L. RNA-sequencing of the sturgeon *Acipenser baeri* provides insights into expression dynamics of morphogenic differentiation and developmental regulatory genes in early versus late developmental stages. *BMC Genom.* **2016**, *17*, 564. [[CrossRef](#)]
46. Singh, P.; Borger, C.; More, H.; Sturmabauer, C. The Role of Alternative Splicing and Differential Gene Expression in Cichlid Adaptive Radiation. *Genome Biol. Evol.* **2017**, *9*, 2764–2781. [[CrossRef](#)]
47. Fu, J.; Zhu, W.; Wang, L.; Luo, M.; Song, F.; Dong, Z. Dynamic transcriptome sequencing and analysis during early development in the bighead carp (*Hypophthalmichthys nobilis*). *BMC Genom.* **2019**, *20*, 781. [[CrossRef](#)] [[PubMed](#)]
48. Ma, X.; Su, B.; Tian, Y.; Backenstose, N.J.C.; Ye, Z.; Moss, A.; Duong, T.Y.; Wang, X.; Dunham, R.A. Deep Transcriptomic Analysis Reveals the Dynamic Developmental Progression during Early Development of Channel Catfish (*Ictalurus punctatus*). *Int. J. Mol. Sci.* **2020**, *21*, 5535. [[CrossRef](#)]
49. FAO. *FAO Yearbook. Fishery and Aquaculture Statistics 2018*; Food and Agricultural Organization (FAO): Rome, Italy, 2020.
50. Gan, L.; Wang, Y.-z.; Chen, S.-j.; Lin, Z.-h.; Sun, J.-j.; He, Y.-h.; Tang, H.-j.; Peng, J.; Guo, H.-h. Identification and characterization of long non-coding RNAs in muscle sclerosis of grass carp, *Ctenopharyngodon idellus* fed with faba bean meal. *Aquaculture* **2020**, *516*, 734521. [[CrossRef](#)]
51. Fan, K.; Shen, Y.; Xu, X.; Tao, L.; Bao, T.; Li, J. LncRNA-WAS and lncRNA-C8807 interact with miR-142a-3p to regulate the inflammatory response in grass carp. *Fish Shellfish. Immunol.* **2021**, *111*, 201–207. [[CrossRef](#)]
52. He, Y.; Yu, H.; Zhao, H.; Zhu, H.; Zhang, Q.; Wang, A.; Shen, Y.; Xu, X.; Li, J. Transcriptomic analysis to elucidate the effects of high stocking density on grass carp (*Ctenopharyngodon idella*). *BMC Genom.* **2021**, *22*, 620. [[CrossRef](#)]
53. Li, L.; Jia, X.; Liu, Y.; He, Y.; Pang, Y.; Shen, Y.; Xu, X.; Li, J. lncRNA-SUMO3 and lncRNA-HDMO13 modulate the inflammatory response by binding miR-21 and miR-142a-3p in grass carp. *Dev. Comp. Immunol.* **2021**, *121*, 104082. [[CrossRef](#)] [[PubMed](#)]
54. Wang, Y.; Lu, Y.; Zhang, Y.; Ning, Z.; Li, Y.; Zhao, Q.; Lu, H.; Huang, R.; Xia, X.; Feng, Q.; et al. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat. Genet.* **2015**, *47*, 625–631. [[CrossRef](#)] [[PubMed](#)]

55. Hu, Y.; Wei, X.; Liao, Z.; Gao, Y.; Liu, X.; Su, J.; Yuan, G. Transcriptome Analysis Provides Insights into the Markers of Resting and LPS-Activated Macrophages in Grass Carp (*Ctenopharyngodon idella*). *Int. J. Mol. Sci.* **2018**, *19*, 3562. [[CrossRef](#)]
56. Huang, X.; Jiang, Y.; Zhang, W.; Cheng, Y.; Wang, Y.; Ma, X.; Duan, Y.; Xia, L.; Chen, Y.; Wu, N.; et al. Construction of a high-density genetic map and mapping of growth related QTLs in the grass carp (*Ctenopharyngodon idellus*). *BMC Genom.* **2020**, *21*, 313. [[CrossRef](#)] [[PubMed](#)]
57. Lu, X.; Chen, H.-M.; Qian, X.-Q.; Gui, J.-F. Transcriptome analysis of grass carp (*Ctenopharyngodon idella*) between fast- and slow-growing fish. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2020**, *35*, 100688. [[CrossRef](#)]
58. Akasaki, T.; Nikaido, M.; Nishihara, H.; Tsuchiya, K.; Segawa, S.; Okada, N. Characterization of a novel SINE superfamily from invertebrates: “Ceph-SINEs” from the genomes of squids and cuttlefish. *Gene* **2010**, *454*, 8–19. [[CrossRef](#)] [[PubMed](#)]
59. Gilbert, N.; Labuda, D. Evolutionary inventions and continuity of CORE-SINEs in mammals. *J. Mol. Biol.* **2000**, *298*, 365–377. [[CrossRef](#)]
60. Eng, L.; Coutinho, G.; Nahas, S.; Yeo, G.; Tanouye, R.; Babaei, M.; Dörk, T.; Burge, C.; Gatti, R.A. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: Maximum entropy estimates of splice junction strengths. *Hum. Mutat.* **2004**, *23*, 67–76. [[CrossRef](#)]
61. Uszczyńska-Ratajczak, B.; Lagarde, J.; Frankish, A.; Guigo, R.; Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **2018**, *19*, 535–548. [[CrossRef](#)]
62. Stark, R.; Grzelak, M.; Hadfield, J. RNA sequencing: The teenage years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [[CrossRef](#)] [[PubMed](#)]
63. Tadros, W.; Lipshitz, H.D. The maternal-to-zygotic transition: A play in two acts. *Development* **2009**, *136*, 3033–3042. [[CrossRef](#)]
64. Irie, N.; Kuratani, S. The developmental hourglass model: A predictor of the basic body plan? *Development* **2014**, *141*, 4649–4655. [[CrossRef](#)] [[PubMed](#)]
65. Gladyshev, V.N. The Ground Zero of Organismal Life and Aging. *Trends Mol. Med.* **2021**, *27*, 11–19. [[CrossRef](#)] [[PubMed](#)]
66. Rugh, R. *The Frog Its Reproduction and Development*; Blakiston: Philadelphia, PA, USA, 1951.
67. Kellicott, W.E. *The Development of the Vascular and Respiratory Systems of Ceratodus*; New York Academy of Sciences: New York, NY, USA, 1905; Volume 2.
68. Gillis, J.A.; Tidswell, O.R. The Origin of Vertebrate Gills. *Curr. Biol.* **2017**, *27*, 729–732. [[CrossRef](#)] [[PubMed](#)]
69. de Koning, A.P.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **2011**, *7*, e1002384. [[CrossRef](#)]
70. Rebollo, R.; Romanish, M.T.; Mager, D.L. Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **2012**, *46*, 21–42. [[CrossRef](#)]
71. Alexander, R.P.; Fang, G.; Rozowsky, J.; Snyder, M.; Gerstein, M.B. Annotating non-coding regions of the genome. *Nat. Rev. Genet.* **2010**, *11*, 559–571. [[CrossRef](#)]
72. Kapusta, A.; Kronenberg, Z.; Lynch, V.J.; Zhuo, X.; Ramsay, L.; Bourque, G.; Yandell, M.; Feschotte, C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **2013**, *9*, e1003470. [[CrossRef](#)] [[PubMed](#)]
73. Hollister, J.D.; Gaut, B.S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **2009**, *19*, 1419–1428. [[CrossRef](#)] [[PubMed](#)]
74. Arensburger, P.; Hice, R.H.; Zhou, L.; Smith, R.C.; Tom, A.C.; Wright, J.A.; Knapp, J.; O’Brochta, D.A.; Craig, N.L.; Atkinson, P.W. Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* **2011**, *188*, 45–57. [[CrossRef](#)] [[PubMed](#)]
75. Xu, R.; Li, C.; Liu, X.; Gao, S. Insights into epigenetic patterns in mammalian early embryos. *Protein Cell* **2020**, *12*, 7–28. [[CrossRef](#)] [[PubMed](#)]
76. Ge, S.X. Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. *BMC Genom.* **2017**, *18*, 200. [[CrossRef](#)]
77. Elbarbary, R.A.; Lucas, B.A.; Maquat, L.E. Retrotransposons as regulators of gene expression. *Science* **2016**, *351*, aac7247. [[CrossRef](#)] [[PubMed](#)]
78. Lee, Y.; Rio, D.C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **2015**, *84*, 291–323. [[CrossRef](#)] [[PubMed](#)]
79. Ren, J.; Sun, C.; Clinton, M.; Yang, N. Dynamic Transcriptional Landscape of the Early Chick Embryo. *Front. Cell Dev. Biol.* **2019**, *7*, 196. [[CrossRef](#)] [[PubMed](#)]
80. Monteuis, G.; Wong, J.J.L.; Bailey, C.G.; Schmitz, U.; Rasko, J.E.J. The changing paradigm of intron retention: Regulation, ramifications and recipes. *Nucleic Acids Res.* **2019**, *47*, 11497–11513. [[CrossRef](#)]
81. Li, X.; Liu, S.; Zhang, L.; Issaian, A.; Hill, R.C.; Espinosa, S.; Shi, S.; Cui, Y.; Kappel, K.; Das, R.; et al. A unified mechanism for intron and exon definition and back-splicing. *Nature* **2019**, *573*, 375–380. [[CrossRef](#)]
82. Kimmel, C.B.; Ballard, W.W.; Kimmel, S.R.; Ullmann, B.; Schilling, T.F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **1995**, *203*, 253–310. [[CrossRef](#)]
83. Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **2012**, *7*, e30619. [[CrossRef](#)]
84. Ewels, P.; Magnusson, M.; Lundin, S.; Kaller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)] [[PubMed](#)]

85. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650. [[CrossRef](#)] [[PubMed](#)]
86. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)] [[PubMed](#)]
87. Wu, T.D.; Watanabe, C.K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **2005**, *21*, 1859–1875. [[CrossRef](#)] [[PubMed](#)]
88. Pertea, G.; Pertea, M. GFF Utilities: GffRead and GffCompare [version 1; peer review: 3 approved]. *F1000Research* **2020**, *9*, 304. [[CrossRef](#)]
89. Duan, Y.; Zhang, W.; Cheng, Y.; Shi, M.; Xia, X.-Q. A systematic evaluation of bioinformatics tools for identification of long noncoding RNAs. *Rna* **2021**, *27*, 80–98. [[CrossRef](#)]
90. Conesa, A.; Gotz, S.; Garcia-Gomez, J.M.; Terol, J.; Talon, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)]
91. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [[CrossRef](#)]
92. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)]
93. Kalvari, I.; Nawrocki, E.P.; Ontiveros-Palacios, N.; Argasinska, J.; Lamkiewicz, K.; Marz, M.; Griffiths-Jones, S.; Toffano-Nioche, C.; Gautheret, D.; Weinberg, Z.; et al. Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **2021**, *49*, D192–D200. [[CrossRef](#)]
94. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
95. Vassetzky, N.S.; Kramerov, D.A. SINEBase: A database and tool for SINE analysis. *Nucleic Acids Res.* **2013**, *41*, D83–D89. [[CrossRef](#)] [[PubMed](#)]
96. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [[CrossRef](#)] [[PubMed](#)]
97. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1994**, *2*, 28–36.
98. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018. [[CrossRef](#)]
99. Bailey, T.L.; Gribskov, M. Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics* **1998**, *14*, 48–54. [[CrossRef](#)]
100. Zhong, Z.; Ao, L.; Zhao, L.; Zhang, Z.; Jiang, Y. Screening and validation of reference genes for qPCR analysis in gonads and embryos of Takifugu bimaculatus. *Aquac. Fish.* **2020**, *7*, 278–286. [[CrossRef](#)]
101. Caracausi, M.; Piovesan, A.; Antonaros, F.; Strippoli, P.; Vitale, L.; Pelleri, M.C. Systematic identification of human housekeeping genes possibly useful as references in gene expression studies. *Mol. Med. Rep.* **2017**, *16*, 2397–2410. [[CrossRef](#)]
102. Trincado, J.L.; Entizne, J.C.; Hysenaj, G.; Singh, B.; Skalic, M.; Elliott, D.J.; Eyras, E. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **2018**, *19*, 40. [[CrossRef](#)]
103. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]
104. Yeo, G.; Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **2004**, *11*, 377–394. [[CrossRef](#)] [[PubMed](#)]
105. Zhang, Q.; Fan, X.; Wang, Y.; Sun, M.A.; Shao, J.; Guo, D. BPP: A sequence-based algorithm for branch point prediction. *Bioinformatics* **2017**, *33*, 3166–3172. [[CrossRef](#)] [[PubMed](#)]
106. Beißbarth, T.; Speed, T.P. GOstat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **2004**, *20*, 1464–1465. [[CrossRef](#)] [[PubMed](#)]