Preprints.org

Article

Comparative Tissue Identification and Characterization of Long Non-Coding RNAs in the Globally Distributed Blue Shark *Prionace glauca*

Scarleth Bravo , Patricia Zarate , Ilia Cari , Ljubitza Clavijo , Ignacio Lopez , Nicole M Phillips , Rodrigo Vidal [*]

Article

Comparative Tissue Identification and Characterization of Long Non-Coding RNAs in the Globally Distributed Blue Shark *Prionace glauca*

Scarleth Bravo [1], Patricia Zarate [2], Ilia Cari [2], Ljubitza Clavijo [2], Ignacio Lopez [1], Nicole M Phillips [3] and Rodrigo Vidal [1,*]

[1] Laboratory of Genomics, Molecular Ecology and Evolutionary Studies, Department of Biology, Universidad de Santiago de Chile, Santiago, Chile; scarleth.bravo@usach.cl; ruben.vidal@usach.cl; ignacio.lopez.r@usach.cl

[2] Departamento de Oceanografía y Medio Ambiente, División de Investigación Pesquera, Instituto de Fomento Pesquero, Valparaíso, Chile; patricia.zarate@ifop.cl, ljubitza.clavijo@ifop.cl, ilia.cari@ifop.cl.

[3] School of Biological, Environmental, and Earth Sciences The University of Southern Mississippi Hattiesburg Mississippi USA; N.Phillips@usm.edu

[*] Correspondence: ruben.vidal@usach.cl

Abstract: Long non-coding RNAs (lncRNAs) are involved in numerous biological processes and serve crucial regulatory functions in both animals and plants. Nevertheless, there is limited understanding of lncRNAs and their patterns of expression and roles in sharks. In the current study, we systematically identified and characterized lncRNAs in the blue shark Prionace glauca from four tissues (liver, spleen, muscle, and kidney) using high-throughput sequencing and bioinformatics tools. A total of 21,932 high confidence lncRNAs were identified with 8,984 and 3,067 stably and tissue-specific expressed lncRNAs, respectively. In addition, a total of 45,007 differentially expressed (DE) lncRNAs were obtained among tissues, with kidney versus muscle having the largest numbers across tissues. DE lncRNAs trans target protein-coding genes were predicted, and functional gene ontology enrichment of these genes showed GO terms such as muscle system processes, cellular/metabolic processes, and stress and immune responses, all of which correspond with the specific biological functions of each tissue analyzed. These results advance our knowledge of lncRNAs in sharks and present novel data on tissue-specific lncRNAs, providing key information to support future functional shark investigations.

Keywords: elasmobranch; transcriptome; ncRNA; lncRNA; differential expression

1. Introduction

The sharks (Superorder Selachii) are an ancient group of cartilaginous fish with origins dating back to the Silurian to Ordovician periods (~400-450 million years before present, MYBP), with modern forms arising during the Jurassic period (195 MYBP) [1]. During this extensive evolutionary time, sharks survived at least five mass extinction events [2], which in some cases were followed by rapid diversification of surviving lineages [3]. For this reason, sharks may be considered a resilient group with potentially significant genomic characteristics that have allowed them to overcome challenges throughout their evolutionary history, such as adaptation to a changing environment and/or resistance to disease.

Following the most recent mass extinction at the end of the Cretaceous, and still today, the most diverse shark order is the Carcharhiniformes, comprising 337 of the 536 described extant species. The blue shark Prionace glauca (Linnaeus 1758) is a wide-ranging pelagic carcharhinid found at tropical and temperate latitudes in all oceans [4]. This species was assessed globally as Near Threatened on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species due to population declines resulting from mortalities in fisheries [5]. Blue sharks are harvested for fins and

meat via direct fisheries and when caught as bycatch in both commercial and small-scale fisheries [6,7]. As apex predators, blue sharks are critical to healthy ecosystem function and structure through increasing biodiversity and buffering against invasive species, transmission of disease, and climate change [8]. Additionally, as upper tropic level predators, blue sharks can be considered sentinel species of marine pollution as they are subject to bioaccumulation and biomagnification of contaminants in their tissues, including heavy metals and chemicals [9].

One early response of marine organisms to adapt to environmental changes and stress is alterations in the dynamics of transcript levels of particular genes [10]. It is widely accepted that RNA-sequencing (RNA-seq) is the principal sequencing technology used to understand the composition of transcriptomes and gene expression profiles in non-model species, such as blue sharks. Various studies have reported novel information into the transcriptomes of sharks [11–14], however these studies have focused primarily on the coding parts of the transcriptome. However, several studies on the complexity of eukaryotic transcriptomes have revealed that a sizeable portion of the genome gives birth to non-coding RNA (ncRNA) genes [15]. A class of ncRNA that has received considerable attention in recent years are the long non-coding RNAs (lncRNAs). LncRNAs correspond to DNA sequences that encode transcripts larger than 200 nucleotides with low to null protein-coding potential. They may interact with molecules such DNA, RNA, and proteins, regulating crucial molecular processes such as chromatin modelling, transcription, post-transcriptional, translational, and post-translational [16]. Several studies have reported lncRNAs are broadly distributed in diverse regions of the vertebrate genomes including retrotransposons in protein-coding genes, pseudogenes, introns, natural antisense transcripts, and intergenic regions [17]. Accumulating evidence shows that lncRNAs perform a significant role as gene expression regulators, during stages of development of organs, and in response to environmental changes, among others [18]. In this context, the regulatory effectiveness of lncRNAs is linked to its expression. Many lncRNAs display a tissue-specific profile of expression, generating meaningful clues about their specific activities within tissues. However, there are also stably expressed lncRNAs which are fundamental to the preservation of elementary cellular activities, regardless of tissue origin [19].

Due to the regulatory role of lncRNAs in several biological processes of metazoans, there are numerous studies devoted to the identification and characterization of lncRNAs in animals, plants and teleost fishes [20,21], but such works in sharks are scarce [22,23]. Although lncRNAs and coding-protein genes share several molecular characteristics [24], they exhibit different sequence attributes that have been exploited to build new statistical deep learning models to evaluate the coding potential of any given transcript in organisms lacking reference genomes, obviating the constraints of database-based annotation approaches [25]. Here, we identify and characterize (TS) and stably expressed lncRNAs across four tissues of blue shark using high-throughput RNA sequencing technology. Our results provide valuable information for understanding the regulation and function of lncRNAs in sharks.

2. Materials and Methods

Tissue samples from blue sharks were obtained opportunistically by on-board fishery observers from IFOP (Instituto Fomento Pesquero, Chile) during Chilean pelagic longline fisheries in the eastern Pacific Ocean in July 2022 (Exempt Resolution No. E-2022-071). Tissue samples from kidney, spleen, liver, and muscle from each of four juvenile blue sharks (two males and two females; mean size of 98.6 cm fork length ± 34.3 SD) were collected and immediately preserved in Allprotect Tissue Reagent (Qiagen, USA).

Total RNA was isolated from each tissue using Trizol (Invitrogen, USA) and RNeasy Mini KIT (Qiagen, USA) and genomic DNA residue was removed with TURBO DNAse treatment (Invitrogen, USA), according to the manufacturer's instructions. The qualities, concentrations, and integrities of RNA extracts were evaluated with a Qubit 3.0 Fluorometer (Life Technologies, USA; Qubit RNA HS kit) and Bioanalyzer 2100 (Agilent Technologies, USA; RNA Nano 6000 Assay Kit). A total of 1.7 µg RNA per sample was used for rRNA elimination employing the Ribo-Zero rRNA Removal Kit (Epicentre, USA). The NEBNext Ultra™ II Directional RNA Library Kit for Illumina (New England

Biolabs, USA) was used for the construction of 16 libraries (i.e., four tissues from each of four individuals), which were sequenced on an Illumina HiSeq 2500 sequencer (PE150). The qualities of the raw reads of each library were checked using FastQC v-0.12.0 (Andrews, 2010) software. High-quality clean reads were collected with Trimmomatic v-0.38 software (Bolger et al., 2014) by eliminating reads with adapters (or poly-N) and reads that were low-quality or too short (<97 bp). To remove potential exogenous contamination in the dataset, Kraken 2 [26] was used with a custom database built with all sequences of plasmids, viruses, fungi, bacteria, protozoa, and archaea in RefSeq (release 212). The clean reads of the different tissues were combined and assembled, without genome guidance, with Trinity v-2.8.4 separately (no_normalize_reads- and the rest of parameters default) [27]. The assembled transcripts for the four tissues were clustered into superTranscripts (Trinity "gene"), obtaining the final assembled blue shark transcriptome. To evaluate the assembly quality, we used BUSCO v-5 (Benchmarking Universal Single-Copy Orthologs) (eukaryota_odb10 lineage) [28] and rnaQUAST v-2.2.3 [29]. To identify high confidence blue shark lncRNAs, we applied several filters. First, we used Transdecoder v-5.7.0 (default parameters) (https://github.com/TransDecoder/TransDecoder) to search for open reading frames (ORFs). Second, all superTranscripts and ORFs were searched against NCBI/NR and UniProt/UniRef90 protein database (threshold= 1e-6) and sequences matching with proteins were eliminated. Third, all potential non-coding RNAs were searched against the RFAM database v-14.9 [30] by Infernal v-1.1.4 (default parameters) [31] and any significant match reported was removed. To assess the coding potential of the filtered dataset, two novel deep learning packages, LncADeep v-1.0 [32] and RNAsamba [33], were used. Deep learning models outperform widely used approaches to identify lncRNAs such as logistic regression (e.g., CPAT) and support vector machines (e.g., CPC2) [25]. If a superTranscritp was assigned as coding by at least one of these two packages, it was treated as a possible protein sequence and consequently excluded from the dataset. The remaining superTranscritps assigned by both packages as non-coding were considered as candidate lncRNAs. Finally, to determine high confidence lncRNAs in blue sharks, the clean reads of the candidate lncRNAs were mapped to the superTranscriptome with STAR v-2.7.10b [34] and gene expression levels were quantified (as transcripts per million, TPM) in Salmon v-1.10.1 [35]. Transcripts with TPM $\geq$ 1 were defined as high confidence lncRNAs. To identify putative orthologues of the predicted high confidence lncRNAs in blue sharks, ncRNA sequences for five other shark species spanning three orders were downloaded from NCBI. These included another Carcharhiniformes, the small-spotted catshark Scyliorhinus canicula (RefSeq accession: GCF_902713615.1); one Lamniformes, the white shark Carcharodon carcharias (RefSeq accession: GCF_017639515.1); and three species of Orectolobiformes, the whitespotted bambooshark Chiloscyllium plagiosum (RefSeq accession: GCF_017639515.1), whale shark Rhincodon typus (RefSeq accession: GCF_021869965.1), and zebra shark Stegostoma fasciatum (RefSeq accession (GCF_022316705.1). The predicted high confidence lncRNAs from blue sharks were aligned with those from these additional five species using the CRB-Blast package [36], and paired lncRNAs with a $\geq$ 60% identity and a cut-off value $\leq$ 1e-6 were considered as putative orthologue lncRNAs. To identify lncRNAs that may act as putative miRNA precursors in sharks, the high confidence lncRNAs identified in this study were aligned against known miRNA precursor sequences extracted from mirBase (Release 22.1, http://www.mirbase.org/) using BLASTn. The matched sequences with >90% identity and e-value < 1e-1 were considered as plausible miRNA precursors. Tissue specificity of lncRNAs was determined using the Preferential Expression Measure (PEM) index, calculated according to the TPM expression values by transcripts/tissues [37]. PEM is a single scalar metric that indicates the ubiquity or specificity of any given gene per tissue. PEM values range from 0 to 1, where genes that are tissue-specific have values of approximately 1. In this study, lncRNAs with PEM values $\geq$ 0.8 were considered tissue-specific, while those with TPMs greater than 0.5 and a TPM coefficient of variance less than 1 in all four tissues examined were identified as stably expressed lncRNAs. Clean reads from each tissue type were mapped to the assembled superTranscritptome with STAR v-2.7.10b [34] and raw read counts were obtained with Salmon v-1.10.1 [35]. The DESeq2 R package was used to identify lncRNAs with differential expression among the four different tissues [38]. LncRNAs with corrected P values

(Benjamini–Hochberg false discovery, FDR) < 0.05 and absolute value of log 2-fold change ≥ 1.5 were considered differentially expressed (DE). Given that numerous studies have reported a relevant role of lncRNAs in the regulation of protein-coding genes [39], we evaluated the potential interaction among mRNA targets and DE lncRNAs in blue sharks. There are two main approaches for predicting coding target genes of lncRNAs, cis or trans-target. Since a blue shark reference genome is not available, we selected the trans-target option. For this, the BLASTn algorithm was used to select protein-coding target genes complementary to DE lncRNAs (≤ 1e-6 and identity ≥ 95%). The programme LncTar [40] was used to confirm the relative stability of the interactions detected among lncRNAs and protein-coding genes, with a cut-off value of normalized free energy of −0.13. Four DE lncRNAs expressed in all the tissues were randomly chosen for RT-qPCR analysis (TRINITY_DN433_c0_g5, TRINITY_DN81282_c0_g1,TRINITY_DN4558_c1_g1 and TRINITY_DN8728_C1_G1). Reverse transcription was performed with a blend of random primers and oligo(dT) and iScript reverse transcriptase (iScript cDNA synthesis kit, Bio-Rad, USA) according to the manufacturer's instructions. LncRNA-specific primers for each of the selected lncRNAs were designed with Primer3 Plus (https://www.primer3plus.com/index.html) (Supplementary Table S1). Quantitative real-time PCR was conducted using RT² SYBR Green ROX qPCR Mastermix (Qiagen, USA) and Rotor-Gene Q (Qiagen). qPCR cycling consisted of 95 °C for 4 min, followed by 40 cycles of: 95°C for 6 s and 60°C for 45 s, with a final melting curve analysis at 95°C for 12 s, 60°C for 60 s, and 95°C for 15 s. All RT-qPCRs reactions were performed in triplicate and beta-actin (GenBank: AY455921.1) and TRINITY_DN16185_c1_g2 genes were utilized as normalizers. TRINITY_DN16185_c1_g2 was selected from a subgroup of eight lncRNA genes that exhibited no differential expression in the RNA-Seq data among the four tissue types. Cycle threshold values from the qPCR expression data for these eight lncRNA genes and beta-actin across the four tissues were analyzed with NormFinder [42] and BestKeeper [43] to determine the most reliable reference gene for RT-qPCR. The fold changes in the expression of the selected lncRNAs in distinct tissues were estimated based on the 2−ΔΔCt method [44]. PCR efficiencies of each pair of primers were calculated with the LinRegPCR V. 2021.2 [45] package, based on raw fluorescence data. Functional assessment of the identified DE lncRNAs was performed relative to the functional annotation of their protein-coding genes target. GO (Gene Ontology) biological process terms were determined with Cluster Profile v- 4.2.2 [46] and GO terms were considered as significant with a P adjusted-value (BH) < 0.05 and a Fold Enrichment > 2.

3. Results

3.1. De Novo Assembly Overview

Sixteen cDNA libraries from four tissues of blue sharks (Spleen 1-4, Liver 1-4, Muscle 1-4, and Kidney 1-4) were sequenced to obtain information of the identification and gene expression patterns of sharks lncRNA (NCBI BioProject accession PRJNA1139720). A total of 772,534,672 base pair (bp) PE raw reads were obtained, and after read trimming and exogenous decontamination, 755,635,033 clean reads remained, with a GC content and Q30 average for the samples of 46.51% (± 0.77) and 93.73% (±0.72), respectively (Supplementary Table S2). A total of 218,543 assembled transcripts (encoded by 166,479 trinity "genes") were obtained with an average length of 731.97 bp (range of 201 — 35,984 bp), a N50 value of 1,221 bp, and 45,534 ORFs predicted (≥100 aa). Assessment of the transcriptome completeness by BUSCO showed that nearly 99% of core genes were recognized: complete genes found (C), 98.7%; fragmented genes found (F), 0.9%; duplicated genes found (D), 22.4%, and missing genes (M), 0.4%.

3.2. Identification and Characterization of lncRNAs in Blue Sharks

A total of 21,932 high confidence lncRNAs were identified in blue sharks, with a length ranging from 201 to 9,311 bp (median and average lengths of 611 and 885 bp, respectively) (Supplementary Table S3). This number of lncRNAs is substantially higher than those previously reported for the whitespotted bambooshark [23] and Australian ghostshark Callorhinchus milii [22], although the

distributions of sequence lengths were comparable. The comparison with available ncRNA sequences of several shark species (i.e., white shark, whitespotted bambooshark, whale shark, small-spotted catshark, and zebra shark) revealed that the majority of lncRNAs in blue shark (N= 21,920) are novel and have not previously been documented in any of these other shark species (Supplementary Table S4). The lengths of the lncRNAs sequences identified in blue sharks were shorter than the annotated protein-coding genes (median and average lengths of 754 and 987 bp, respectively), witha slightly higher GC content for protein- coding genes (46.34%) compared to lncRNAs (42.58%). Notably, lncRNAs in blue sharks demonstrated lower expression levels than protein-coding genes in all the tissues analyzed (Figure 1). Further, two lncRNAs in blue sharks were identified as potential precursors of two know miRNAs (TRINITY_DN10569_c4_g1:Oryzias latipes-mir-181b-1 and TRINITY_DN11922_c0_g4: Homo sapiens-miR-1244-4).
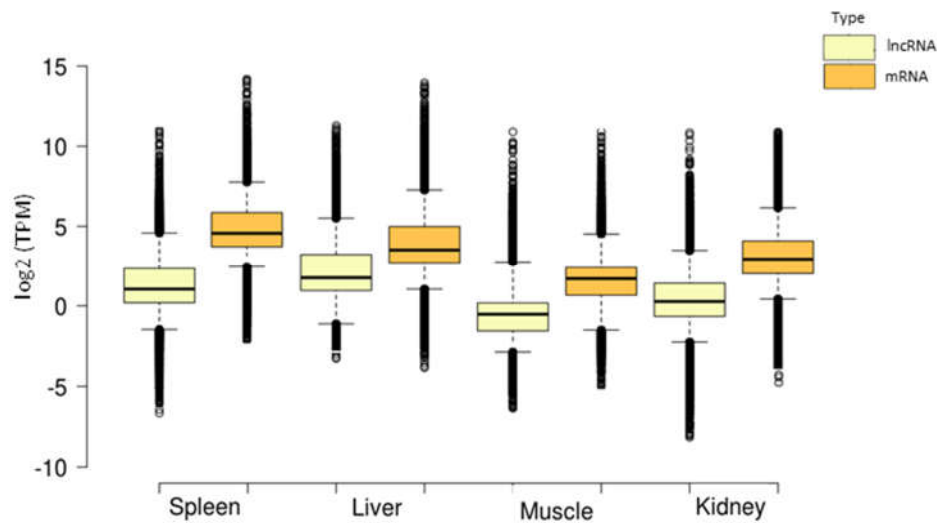

Figure 1. Gene expression (transcripts per million; TPM) distribution boxplots of long non-coding RNAs (lncRNA) and mRNAs per tissue type in the blue shark Prionace glauca.

3.3. Identification of lncRNAs Among Tissues

To categorize the global expression patterns of the lncRNAs identified among tissues, a principle component analysis (PCA) (RNA seq data normalized to relative log estimate) and heatmap (RNA seq data normalized to TPM) analyses were utilized. The four tissue types from blue sharks were distinctly clustered in the heatmap analysis and in the first and second Principal Component (PC1 and PC2), explaining 88.4 % of the variance (Figure 2).
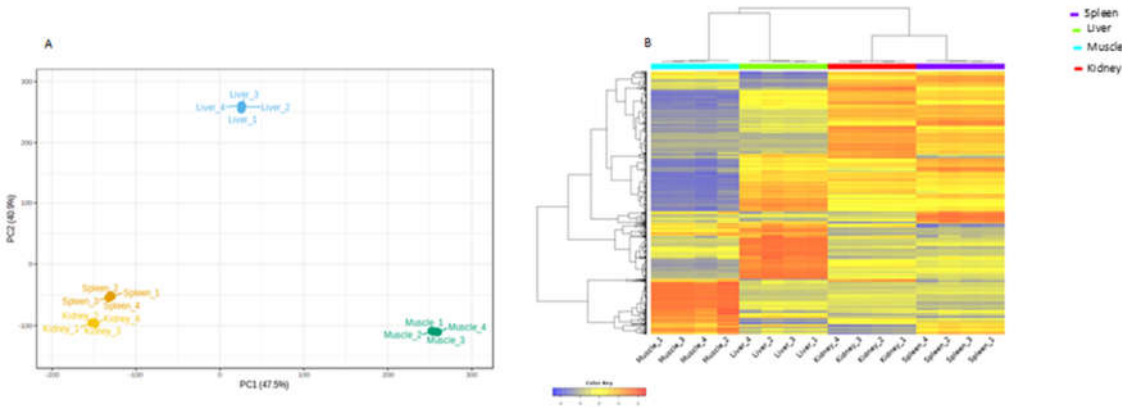
Figure 2. Principal component analysis (A) and heatmap (B) of high confidence long non-coding RNAs (lncRNAs) identified from blue shark Prionace glauca tissues.

A mean number of 14,821 lncRNAs were observed per tissue type, with only 6,226 shared among all tissues (Figure 3). Most of the lncRNAs identified were expressed (TPM ≥1) in the liver (14,768), spleen (16,176), and kidney (19,068), while muscle showed the smallest number (9,272). Kidney and spleen, key lymphoid organs in sharks, share a close pattern of lncRNAs expression. A total of 45,007 DE lncRNAs were obtained between pairwise tissue type comparisons, with the greatest and least number of DEs found among kidney vs muscle and kidney vs spleen, respectively (Figure 4) (Supplementary Table S5.
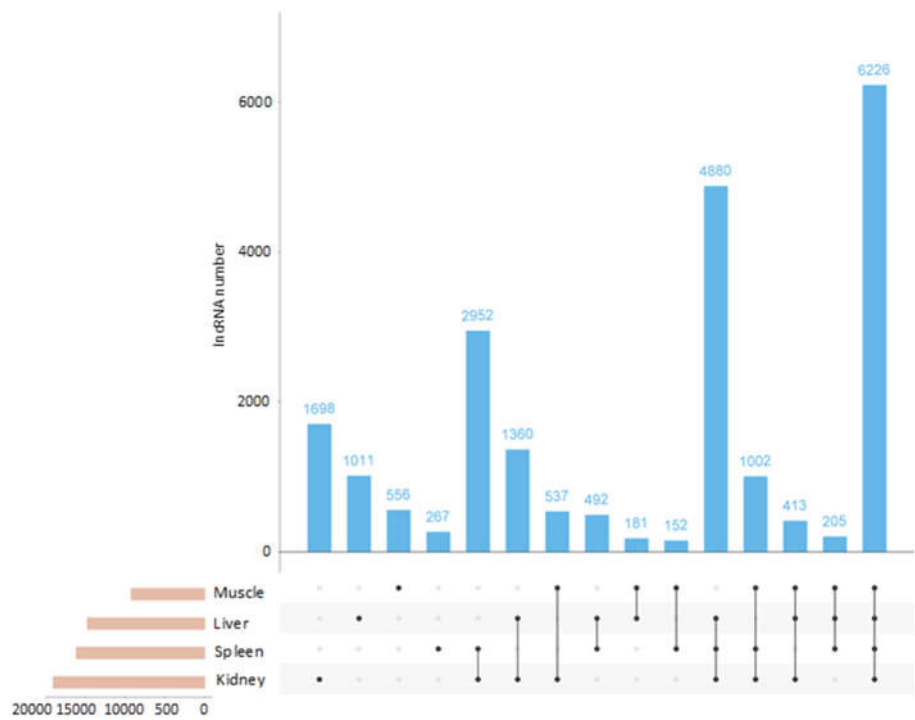

Figure 3. Upset plot showing the number of shared long non-coding RNAs (lncRNAs) expressed across blue shark Prionace glauca tissues.
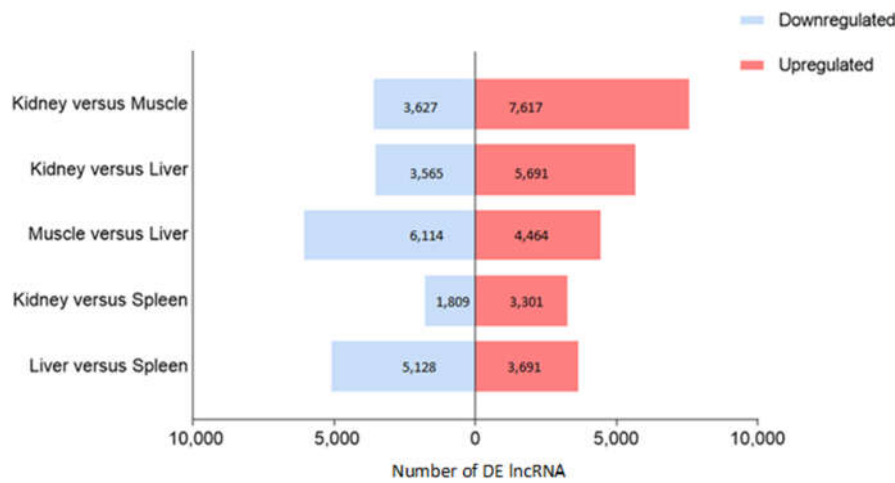

Figure 4. Total number of differentially expressed (DE) long non-coding RNA (lncRNA) in pairwise comparisons of blue shark Prionace glauca tissues.

An analysis of the protein-coding target genes of the total DE lncRNAs detected 454 protein-coding genes as trans-regulated targeted by 1,001 lncRNAs. GO enrichment analysis of these target genes across tissues showed diverse relevant biological process terms, such as muscle contraction (GO:0006936) and skeletal muscle fiber development (GO:0048741) in muscle tissues; phagocytosis (GO:0006909) and immune response (GO:0006955) in kidney tissues; glycogen metabolic process (GO:0005977) and signal transduction (GO:0007165) in liver tissues and defense response (GO:0006952) and RNA binding (GO:0003723) in spleen tissues (Figure 5).
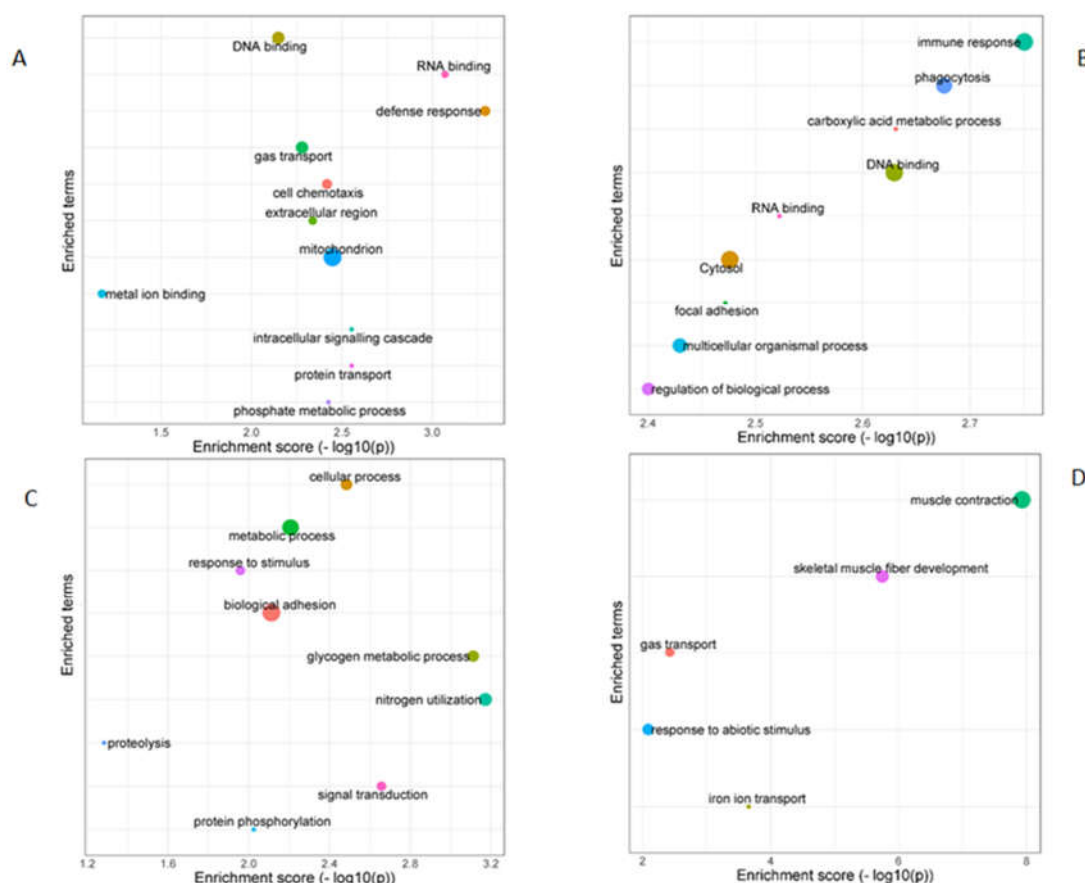

Figure 5. GO enrichment analysis of differentially expressed (DE) long non-coding RNA (lncRNA) protein target genes in blue shark *Prionace glauca* spleen (A), kidney (B), liver (C) and muscle (C) tissues.

3.4. Tissue-Specific lncRNA Expression

Using the PEM index, we found that liver (1,147) and muscle (1,028) had the greatest numbers of TS lncRNAs, while kidney (593) and spleen (299) had the least. Further, a density plot of maximum-normalized PEM values shows similar patterns among liver and muscle, as well as spleen and kidney tissues (Figure 6). In addition, a total of 8,984 stably expressed lncRNAs were identified across all the tissues (Supplementary Table S6).
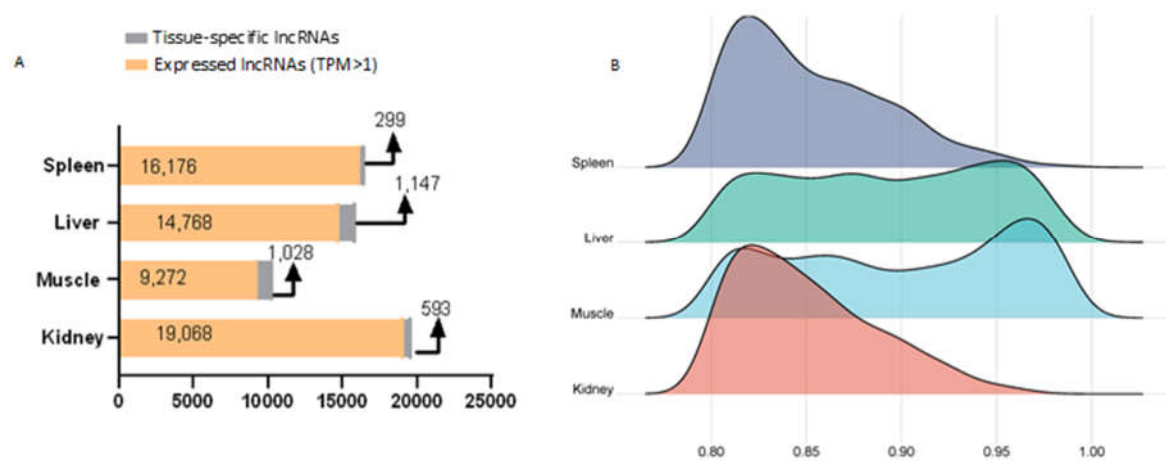
8

Figure 6. Number of tissue-specific (TS) and expressed lncRNAs across blue shark *Prionace glauca* tissues (A). Density plot of maximum-normalized Preferential Expression Measure (PEM) values by tissue (B).

3.5. Experimental RT-qPCR Validation of lncRNAs

All of the primer pairs designed for the four DE lncRNAs chosen for qPCR validation had an efficiency between 94.57 and 98.02%, within the ideal range for qPCR analysis. The relative expression of the tissues DE lncRNAs was in agreement with RNA-seq results (Figure 7).
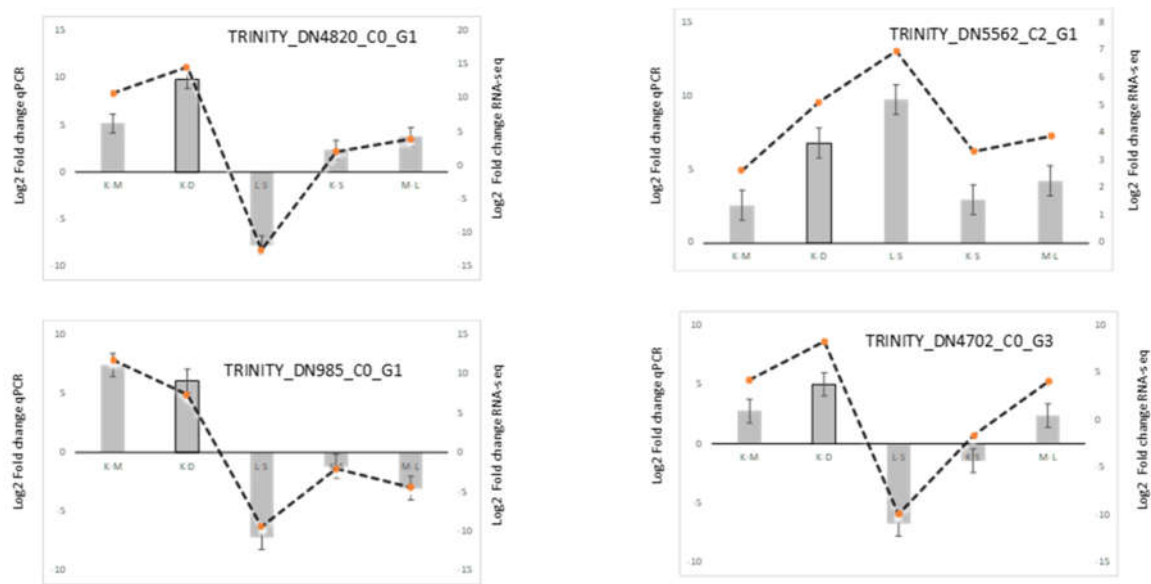

Figure 7. Validation of four lncRNAs in blue shark Prionace glauca using RT-qPCR. CT-qPCR raw data were analyzed by the 2 -ΔΔCt method using beta-actin and TRINITY_DN16185_c1_g2 as reference genes. Each column represents the mean of qPCR technical replicates ±SE and broken lines represent RNA-seq fold change tissue values.

4. Discussion

While a sizable amount of lncRNA research has been conducted on plant and animal model organisms, there is comparatively little information available on lncRNAs in non-model species, such as sharks. Numerous studies have shown that lncRNAs can modulate gene expression and participate in several physiological processes [39]. Studies on tissue-specific characterization and lncRNAs in sharks is very scarce compared to teleost fishes. Prior to this study, transcriptomics investigations on sharks focused more on identifying functional coding genes [11–14] rather than

ncRNA genes, including lncRNA. In this research, we used ribosomal-RNA depleted RNA-seq on an Illumina HiSeq 2500 platform and a highly stringent data processing pipeline to discard transcripts with protein-coding potential, minimizing the inclusion of false positive lncRNAs in blue sharks. A total of 21,932 lncRNAs were identified across kidney, spleen, liver and muscle tissues of blue sharks. Compared to protein-coding genes, these lncRNAs were shorter with lower expression levels and GC content, which is concordant with reports for other vertebrate lncRNAs studies [47]. The lncRNAs identified in the blue shark in this study demonstrates a high degree of evolutionary novelty, with little conservation of lncRNA across shark species. Only 5% of the lncRNAs identified in blue sharks were also reported in other cartilaginous species (i.e., white shark, whitespotted bambooshark, whale shark, small-spotted catshark, and zebra shark). Similarly, other studies have also shown low levels of sequence conservation of lncRNA across a wide taxonomic range [48]. Hara et al. (2018) [23] suggested a higher degree of lncRNA conservation within the cartilaginous fishes compared to teleost fishes. However, this hypothesis was based on a comparative blast approach including several teleost species with well- established genome annotation and only two shark species (brownbanded bamboo shark Chiloscyllium punctatum and the cloudy catshark Scyliorhinus torazame) with partial genome sequencing. Therefore, the greater conservation of lncRNAs in sharks in Hara et al. (2018) [23] may be due to an underestimation of the lncRNA diversity in the shark species involved in that study. Comparative lncRNA expression analysis showed that functionally similar tissues (e.g., spleen and kidney) clustered closely together in terms of their global transcriptomic patterns. Both spleen and kidney are recognized immune organs in teleost and cartilaginous fishes [49,50]. Since each organ has its own transcriptional pattern, several DE lncRNAs genes in functionally distinct tissues were expected. In fact, we found the largest number of differentially expressed lncRNAs gene among kidney vs. muscle. The main function of the muscles is movement while that of the kidney is the regulation of physiological homeostasis and immune response. Furthermore, the regulatory efficacy of lncRNAs is dependent on their own expression. Numerous lncRNAs exhibited tissue-specific expression signatures, offering crucial hints about the unique roles that these molecules perform within cells/tissues. Moreover, there were stably expressed lncRNAs widely expressed in all tissue/cell types, establishing the fundamental transcriptome conservation for elementary biological functions. PEM index analysis revealed that the number of TS lncRNAs differ considerably among tissues and appears to be unrelated to the quantity of expressed lncRNAs in each type of tissue. The liver and muscle showed the highest number of TS lncRNAs, perhaps driven by the need for more varied lncRNA repertoires or the existence of multiplex cell composition in these tissues. Conversely, the spleen and kidney tissues had the least number of TS lncRNAs, which could be associated with the specialized functional roles of these tissues. The higher number of stably expressed lncRNAs (8,984) than TS lncRNAs (3,067) across all tissues might be indicative of the former participating in the regulatory response of a wide range of tissue functions and cell signals. GO enrichment analysis of DE lncRNA protein-coding targets across tissues revealed that different biological process GO terms enriched and associated with the specific biological functions of each tissue. In muscle, DE lncRNAs targeted protein-coding genes involved in muscle system processes, and gas and iron transport. Blue sharks have long-distance dispersal across and between ocean basins [51], therefore lncRNAs are likely pivotal components in the modulation of muscle activities and muscle regeneration and differentiation necessary to undertake these migrations. In the kidney and spleen, the main GO terms involved responses to stress, cellular/metabolic processes, binding, and immune response. Both organs have been recognized as playing important roles in the immune and stress responses in sharks [50,52], and is consistent with the relevance and roles reported in several teleost species [53]. Additionally, the liver DE lncRNA protein-coding targets showed as outstanding metabolic and homeostasis cellular process GO terms. The liver is a large, multi-lobed organ in sharks with key catabolic and anabolic functions [54], therefore lncRNAs could represent relevant contributors to the biological functions of the liver.

This research provides novel information on lncRNAs from multiple tissues in a shark for the first time, filling a gap in our knowledge and characterization of sharks ncRNA. These data will provide a key resource to support future studies on functional characterization of lncRNAs in sharks

and serving as an important first step in understanding adaptation of a sentinel marine species to environmental changes and stress.

References

1   Turner.; S.; Miller.; R.M. New Ideas About Old Sharks. Am. Sci. 2005.; 93.

2   Guinot.; G; Condamine.; F.L. Global impact and selectivity of the Cretaceous-Paleogene mass extinction among sharks.; skates.; and rays. Science. 2023.; 379.; 802-806.

3   Slater.; T.S.; Ashbrook.; K.; Kriwet.; J. Evolutionary relationships among bullhead sharks (Chondrichthyes.; Heterodontiformes). Pap. Palaeontol. 2020, 6.; 425–437.

4   Ebert, D.A.; Dando, M.; Fowler, S. Sharks of the world. A Complete Guide.; Princeton University Press: Princeton, UK, 2021; pp 324-385.

5   Rigby.; C.L.;Barreto.; R.; Carlson.; J.; Fernando.; D.; Fordham.; S.; Francis.; M.P.; Herman.; K.; Jabado.; R.W.; Liu.; K.M.; Marshall.; A.; Pacoureau.; N.; Romanov.; E.; Sherley.; R.B.; Winker.; H. 2019. Prionace glauca. The IUCN Red List of Threatened Species 2019: e.T39381A2915850.

6   Camhi, M.D., Pikitch, E.K.; Babcock, E.A. Sharks of the Open Ocean: Biology, Fisheries and Conservation.; John Wiley & Sons, Chichester, 2009; 140-151.

7   Martínez-Ortiz.; J.; Aires-da-Silva.; AM.; Lennert-Cody.; CE.; Maunder.; MN. The Ecuadorian Artisanal Fishery for Large Pelagics: Species Composition and Spatio-Temporal Dynamics. PLoS One. 2015, 10.; e0135136.

8   Vidal.; A.; Cardador.; L.; Garcia-Barcelona.; S.; Macias.; D.; Druon.; J-N.; Coll.; M.; Navarro.; J.The relative importance of biological and environmental factors on the trophodynamics of a pelagic marine predator.; the blue shark (Prionace glauca). Mar. Environ. Res. 2023, 105808.

9   Alves.; L.F.; Lemos.; M.F.L.; Moutinho.; A.B.; Ceia.; F.R.; Munoz-Arananz.; J.; Jimenez.; B.; Cabral.; H.; Novaise.; S.C. (2023). Assessment of contaminants in blue sharks from the Northeast Atlantic: profiles.; accumulation dynamics.; and risks for human consumers. Environ. Pollut. 2023, 316.; 120467.

10   Page.; T.M.; Lawley.; J.W. The Next Generation Is Here: A Review of Transcriptomic Approaches in Marine Ecology. Front. Mar. Sci. 2022, 9.; 757921.

11   Onimaru.; K.; Tatsumi.; K.; Shibagaki.; K.; Kuraku.; S. A de novo transcriptome assembly of the zebra bullhead shark.; Heterodontus zebra. Sci. Data 2018, 5.; 180197.

12   Goshima.; M.; Sekiguchi.; R.; Matsushita.; M.; Nonaka.; M. The complement system of elasmobranches revealed by liver transcriptome analysis of a hammerhead shark.; Sphyrna zygaena. Dev Comp Immunol. 2016, 61.; 13-24.

13   Seixas.; M.J.; Domingues.; R.R.; Antunes.; A. Decoding the Transcriptome of Sharks.; Rays.; and Chimaeras: Insights into Their Physiology.; Morphology.; Evolution.; and Biomedical Applications. Fishes. 2023, 8.; 271.

14   Machado.; A.M.; Almeida.; T.; Mucientes.; G.; Esteves.; P.J.; Verissimo.; A.; Castro.; L.F.C. De novo assembly of the kidney and spleen transcriptomes of the cosmopolitan blue shark.; Prionace glauca. Mar. Genomics. 2018, 37.; 50-53.

15   Li.; J.; Liu.; C. Coding or Noncoding.; the Converging Concepts of RNAs. Front. Genet. 2019, 10.; 496.

16   Mattick.; J. S.; Amaral.; P. P.; Carninci.; P.; Carpenter.; S.; Chang.; H. Y.; Chen.; L. L.; Chen.; R.; Dean.; C.; Dinger.; M. E.; Fitzgerald.; K. A.; Gingeras.; T. R.; Guttman.; M.; Hirose.; T.; Huarte.; M.; Johnson.; R.; Kanduri.; C.; Kapranov.; P.; Lawrence.; J. B.; Lee.; J. T.; Mendell.; J. T.; Wu.; M. Long non-coding RNAs: definitions.; functions.; challenges and recommendations. Nat. Rev. Mol. Cell Biol. 2023, 24.; 430–447.

17   Ulitsky.; I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat Rev Genet. 2016, 17.; 601-614.

18   Zhang.; X.; Wang.; W.; Zhu.; W.; Dong.; J.; Cheng.; Y.; Yin.; Z.; Shen.; F. Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. Int. J. Mol. Sci. 2019, 20.; 5573.

19   Seifuddin.; F.; Singh.; K.; Suresh.; A.; Judy.; J. T.; Chen.; Y. C.; Chaitankar.; V.; Tunc.; I.; Ruan.; X.; Li.; P.; Chen.; Y.; Cao.; H.; Lee.; R. S.; Goes.; F. S.; Zandi.; P. P.; Jafri.; M. S.; Pirooznia.; M. lncRNAKB.; a knowledgebase of tissue-specific functional annotation and trait association of long noncoding RNA. Sci. Data. 202.; 7.; 326.

20   Kosinska-Selbi.; B.; Mielczarek.; M.; Szyda.; J. Review: Long non-coding RNA in livestock. Animal. 2020.; 14.; 2003–2013.

21   Domínguez-Rosas.; E.; Hernández-Oñate.; M. Á.; Fernandez-Valverde.; S. L.; Tiznado-Hernández.; M. E. Plant long non-coding RNAs: identification and analysis to unveil their physiological functions. Front. Plant Sci. 2023.; 14.; 1275399.

22   Hezroni.; H.; Koppstein.; D.; Schwartz.; M. G.; Avrutin.; A.; Bartel.; D. P.; Ulitsky.; I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. 2015.; 11.; 1110–1122.

23   Hara.; Y.; Yamaguchi.; K.; Onimaru.; K.; Kadota.; M.; Koyanagi.; M.; Keeley.; S. D.; Tatsumi.; K.; Tanaka.; K.; Motone.; F.; Kageyama.; Y.; Nozu.; R.; Adachi.; N.; Nishimura.; O.; Nakagawa.; R.; Tanegashima.; C.; Kiyatake.; I.; Matsumoto.; R.; Murakumo.; K.; Nishida.; K.; Terakita.; A.; Kuraku.; S. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nat. Ecol. Evol. 2018.; 1761–1771.

24   Quinn.; J. J.; Chang.; H. Y. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet.2016.; 17.; 47–62.

25   Ammunét.; T.; Wang.; N.; Khan.; S.; Elo.; L. L. Deep learning tools are top performers in long non-coding RNA prediction. Brief Funct. Genomics. 2022.; 21.; 230–241.

26   Breitwieser.; F. P.; Baker.; D. N.; Salzberg.; S. L. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biol. 2018.; 19.; 198.

27   Grabherr.; M. G.; Haas.; B. J.; Yassour.; M.; Levin.; J. Z.; Thompson.; D. A.; Amit.; I.; Adiconis.; X.; Fan.; L.; Raychowdhury.; R.; Zeng.; Q.; Chen.; Z.; Mauceli.; E.; Hacohen.; N.; Gnirke.; A.; Rhind.; N.; di Palma.; F.; Birren.; B. W.; Nusbaum.; C.; Lindblad-Toh.; K.; Friedman.; N.; Regev.; A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 2011.; 29.; 644–652.

28   Simão.; 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.

29   Bushmanova.; E.; Antipov.; D.; Lapidus.; A.; Suvorov.; V.; Prjibelski.; A. D. rnaQUAST: a quality assessment tool for de novo transcriptome assemblies. Bioinformatics. 2016.; 32.; 2210–2212.

30   Kalvari.; I.; Nawrocki.; E. P.; Ontiveros-Palacios.; N.; Argasinska.; J.; Lamkiewicz.; K.; Marz.; M.; Griffiths-Jones.; S.; Toffano-Nioche.; C.; Gautheret.; D.; Weinberg.; Z.; Rivas.; E.; Eddy.; S. R.; Finn.; R. D.; Bateman.; A.; Petrov.; A. I. (2021). Rfam 14: expanded coverage of metagenomic.; viral and microRNA families. Nucleic Acids Res. 2021.; 49(D1).

31   Nawrocki.; E. P.; Eddy.; S. R. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013.; 29.; 2933–2935.

32   Yang.; C.; Yang.; L.; Zhou.; M.; Xie.; H.; Zhang.; C.; Wang.; M. D.; Zhu.; H. LncADeep: an ab initio lncRNA identification and functional annotation tool based on deep learning. Bioinformatics. 2018.; 34.; 3825–3834.

33   Camargo.; A. P.; Sourkov.; V.; Pereira.; G. A. G.; Carazzolle.; M. F. RNAsamba: neural network-based assessment of the protein-coding potential of RNA sequences. NAR genom. bioinform. 2020.; 2.; lqz024.

34   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.

35   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.

36   Aubry.; S.; Kelly.; S.; Kümpers.; B. M.; Smith-Unna.; R. D.; Hibberd.; J. M. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. PLoS Genet. 2014, 10, e1004365.

37   Kryuchkova-Mostacci.; N.; Robinson-Rechavi.; M. A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform. 2017, 18, 205–214.

38   Love.; M. I.; Huber.; W.;Anders.; S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014, 15, 550.

39   Statello.; L.; Guo.; C. J.; Chen.; L. L.; Huarte.; M. Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. 2021.; 22.; 96–118.

40   Li.; J.; Ma.; W.; Zeng.; P.; Wang.; J.; Geng.; B.; Yang.; J.; Cui.; Q. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Brief. Bioinform. 2015.; 16.; 806–812.

41   Rao X.; Huang X.; Zhou Z.; Lin X. An improvement of the 2ˆ(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat Bioinforma Biomath. 2013 Aug;3(3):71-85.

42   Andersen.; C.L.; Jensen.; J.L.; Orntoft.; T.F. Normalization of real-time quantitative reverse transcription-PCR data: a model- based variance estimation approach to identify genes suited for normalization.; applied to bladder and colon cancer data sets. Cancer Res. 2004. 64.; 5245–5250.

43   Pfaffl.; M.W.; Tichopad.; A.; Prgomet.; C.; Neuvians.; T.P. Determination of stable housekeeping genes.; differentially regulated target genes and sample integrity: bestKeeper–Excel- based tool using pair-wise correlations. Biotechnol. Lett. 2004. 26.; 509–515

44   Livak.; K. J.; Schmittgen.; T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. 2001.; 25, 402–408.

45   Ruijter, J. M.; Ramakers, C.; Hoogaars, W. M.; Karlen, Y.; Bakker, O.; van den Hoff, M. J.; Moorman, A. F. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. 2009, 37, e45.

46   Yu, G.; Wang, L.; Han, Y.; He, Q. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012, 16, 284-287.

47   Mallory, A. C.; Shkumatava, A. LncRNAs in vertebrates: advances and challenges. Biochimie. 2015, 117, 3–14.

48   Szcześniak, M. W.; Kubiak, M. R.; Wanowska, E.; Makałowska, I. Comparative genomics in the search for conserved long noncoding RNAs. Essays Biochem.. 2021, 65, 741–749.

49   Press, C.McL.; Evensen, Ø. The morphology of the immune system in teleost fishes. Fish Shellfish Immunol. 1999, 9, 309-318.

50   Criscitiello M. F. What the shark immune system can and cannot provide for the expanding design landscape of immunotherapy. Expert Opin Drug Discov. 2014, 9, 725–739.

51   Veríssimo, A.; Sampaio, Í.; McDowell, J.R.; Alexandrino. P.; Mucientes, G.; Queiroz N.; da Silva, C.; Jones, C.S.; Noble, L.R.; World without borders-genetic population structure of a highly migratory marine predator, the blue shark (Prionace glauca). Ecol Evol. 2017, 7:4768-4781.

52   Ott, J. A.; Ohta, Y.; Flajnik, M. F.; Criscitiello, M. F. Lost structural and functional inter-relationships between Ig and TCR loci in mammals revealed in sharks. Immunogenetics. 2021, 73, 17–33.

53   Zhao, T.; Zou, Y.; Yan, H.; Chang, Y.; Zhan, Y. (2023). Non-coding RNAs targeting NF-κB pathways in aquatic animals: A review. Front Immunol. 2023, 14, 1091607.

54   de Melo, L.F.; Cabrera, M.L.; Rodrigues, A.C.B.; Turquetti, A.d.O.M.; Ruivo, L.P.; Bruno, C.E.M.; Rici, R.E.G. Morphological Description of Blue Shark Liver, Prionace glauca (Linnaeus, 1758), Elasmobranchii, Carcharhiniformes. Int. J. Adv. Eng. Res. Sci. 2019, 6, 286–290.