Comparative eye and liver differentially expressed genes reveal monochromatic vision and cancer resistance in the shortfin mako shark (*Isurus oxyrinchus*)

Running title: Transcriptome of *Isurus oxyrinchus*

Rodrigo R Domingues[1,6,7*], Vito Antonio Mastrochirico-Filho[2], Natalia J Mendes[2], Diogo T Hashimoto[2], Rui Coelho[3,4], Vanessa Paes da Cruz[5], Agostinho Antunes[6,7], Fausto Foresti[5], Fernando F Mendonça[1]

[1]Departamento de Ciências do Mar, Instituto do Mar, Universidade Federal de São Paulo – UNIFESP, Rua Carvalho de Mendonça, 144, Santos, São Paulo, Brasil, 11070-100 e-mail: domingues.pesca@gmail.com

[2]Centro de Aquicultura da UNESP, Universidade Estadual Paulista – UNESP, Jaboticabal, Via de Acesso Prof. Paulo Donato Castellane, s/n, Jaboticabal, São Paulo, Brasil, 14884-900

[3]Instituto Português do Mar e da Atmosfera (IPMA, I.P.), Av. 5 Outubro, 8700-305 Olhão, Portugal

[4]Centro de Ciências do Mar (CCMAR). Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

[5]Laboratório de Biologia e Genética de Peixes, Instituto de Biociências de Botucatu, Universidade Estadual Paulista – UNESP, Botucatu, São Paulo, Brasil, Rua Prof. Antônio Celso Wagner Zanin, 250, Botucatu, São Paulo, Brasil, 18618-689

[6]CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Av. General Norton de Matos, s/n, 4450-208 Porto, Portugal

[7]Departmento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal

Abstract

The shortfin mako, *Isurus oxyrinchus* is an oceanic pelagic shark found worldwide in tropical and subtropical waters. However, the understanding of its biology at molecular level is still incipient. We sequenced the messenger RNA isolated from eye and liver tissues. *De novo* transcriptome yielded a total of 705,940 transcripts. A total of 3,774 genes were differentially expressed (DEGs), with 1,612 in the eye and 2,162 in the liver. Most DEGs in the eye were related to structural and signaling functions, including nonocular and ocular opsin genes, whereas nine out of ten most overexpressed genes in the liver were related to tumor suppression, wound healing, and human diseases. Furthermore, DEGs findings provide insights on the monochromatic shark vision and a repertory of cancer-related genes, which may be insightful to elucidate shark resistance to cancer. Therefore, our results provide valuable sequence resources for future functional and population studies.

Keywords: Elasmobranch, gene ontology, photoreception, transcriptomics, tumor suppression.

1. Introduction

Chondrichthyes (sharks, rays and chimaeras) are known by their long evolutionary history with the earliest fossil evidence for sharks dating to 450 million years ago, during the Late Ordovician Period [1]. During their evolutionary history, sharks experienced at least five big mass extinction, and currently consist of approximately 1,160 living species, representing a small fraction of modern fish fauna [2,3]. Such resilience suggests that sharks have unique genetic traits that support their evolutionary history [3].

Although Chondrichthyes are of great evolutionary, ecologic and economic significance, they are still largely underrepresented in omics studies [3]. Indeed, only a few Chondrichthyes species had their genome sequenced and transcripts annotated so far (*e.g*., [4] for *Callorhinchus milii*, 0.937 Gbp / 18,872 protein-coding genes; [5] for *Rhincodon typus*, 3.44 Gbp / 19,384 protein-coding genes; [6] for *Chiloscyllium punctatum*, 4.7 Gbp and *Scyliorhinus torazame*, 6.7 Gbp; [7] for *Carcharodon carcharias*, 4.63 Gbp / 24,520 protein-coding genes). Despite few genome and transcriptome studies available for elasmobranchs, these are shedding light on evolutionary pathways of this taxon, and are contributing to medical application on human health [7]. For example, transcriptome and genome sequences of the great white shark, revealed important results on their improved capacity of wound healing and DNA repair, which may have large relevance for humans applications [7]. However, despite the recent advances in omics studies, most species with potential ecology, economic and health importance remain poorly studied at molecular level [8].

The shortfin mako shark *Isurus oxyrinchus*, is a large-bodied, commercially exploited, coastal and oceanic epipelagic shark species found worldwide in tropical and temperate waters [9]. As a consequence, the shortfin mako shark is currently listed as

globally Endangered by the International Union for Conservation of Nature (IUCN) [10]. This shark species is known to be the world's fastest shark, and it has evolved the capacity to elevate the temperature of several core body tissues (*e.g*., viscera, muscle, eye and brain) to enhance sustainable swimming performance [11]. As a result, this shark species can maintain relatively stable tissue temperatures even in different ambient water temperatures [11]. Some previous studies have been conducted on this shark biology and ecology [12,13], but very few have been conducted at molecular level. Only recently its whole mitochondrial genome was sequenced, and its heart transcriptome was annotated and compared with bony fishes [14,15].

Advances in molecular techniques to obtain large-scale sequences using next generation sequence (NGS) have expanded transcriptome information of several animals [16]. Furthermore, transcriptome sequences have been providing other information, such as gene discovery, gene expression, decipher of the complete protein coding sequence (CDs), and single nucleotide polymorphism (SNP) identification, which have applications in fields such as physiology, genetics and immunology ([8,17]. In addition, over the last years transcriptome was used in approximately 600 ecological studies and addressed questions on interactions between gene expression and population structure, gene expression and environment, and local adaptation [16]. Applying transcriptome assessment in wildlife is important to make progress in understanding how genome actually functions to give rise to complex traits and adapts to complex environments gradient. This is even more important for Chondrichthyes species, which face population threats due to overfishing and habitat loss [3,16]. Furthermore, the advances in the omics science are pivotal to shed light on unique molecular features of the wildlife associated to human health.

As such, the main goal of this study is to provide a *de novo* reference transcriptome and differentially expressed genes for the eye and liver of the shortfin mako shark. Specifically, deep RNA sequencing (RNA-seq) and *de novo* assembly were performed on those tissues (eye and liver) in order to obtain a repertory of genes and their functions. These sequences can now provide valuable molecular resources for future phylogenetic, ecological and biological studies on sharks.

2. Material and Methods

2.1. Sampling and RNA isolation

Tissue samples were opportunistically collected by onboard fishery observers from IPMA (*Instituto Português do Mar e da Atmosfera*) during Portuguese pelagic longline fisheries in the eastern Atlantic Ocean, and took place between October and November 2016. The onboard observers took samples exclusively from specimens that were dead at-haulback as part of their regular data and biological sampling work done on those commercial fisheries. Specifically, tissues samples from different organs (i.e., heart, kidneys, muscle and stomach) from four juveniles shortfin mako shark in total were collected and preserved in RNA-*later*® (ThermoFisher), and cold storage in order to avoid RNA degradation. The individuals comprised two males and two females, with fork length (FL) ranging between 150 and 177 cm. Once in the laboratory, the tissues were stored at -80°C until RNA isolation. No specific ethical approval or permit for animal capture was required for this work, as the specimens were captured under regular and licensed fishing operations with fisheries observers onboard, and were not specifically captured for this study.

The total RNA was isolated using the RNeasy Mini Kit (Qiagen Inc., CA, USA), according to the manufacturer protocol. All RNA samples were treated with DNAse I to

remove possible contaminants. Total RNA concentration was quantified using a Qubit™ spectrofluorometer (Thermo Fisher Scientific, Wilmington, USA). Each RNA sample was checked for RNA degradation using an Agilent 2100 BioAnalyzer, with RIN (RNA integrity number) value accepted above of 7. After quality control, only three tissues from the eye and four from the liver showed good quality for library build of RNA-seq paired-end.

2.2. Library and Illumina sequencing

Seven libraries for sequencing were prepared using RNA Truseq® (Illumina, San Diego, USA), according to the manufacturer protocol. In general, enrichment of mRNA was isolated from total RNA samples using oligo (*dT*) 25 magnetic beads (Dynabeads: Invitrogen, USA) for posterior random fragmentation by adding fragmentation buffer. Posteriorly, mRNA was used as a template for first-strand cDNA synthesis by adding random hexamers primers. Finally, fragments were end-repaired by incubation in the presence of T4 DNA polymerase and Klenow polymerase. The total mRNA extracted by poly-A tail of each tissue was then used to prepare Illumina TruSeq RNA sequencing libraries of cDNA according to the manufacturer instructions. High-throughput sequencing was performed on an Illumina HiSeq 2500 platform to generate paired-end 150 bp reads at Novogene Co., Ltd. (California, USA).

2.3. *De novo* assembly and assessment completeness

The quality of sequencing for each library containing the raw cDNA sequencing reads was assessed using the software FastQC [18]. Initial quality filtering (Q phred < 20) and removal of residual adapters from the Illumina sequencing were conducted by the Trimmomatic v.0.30 [19]. Therefore, sequencing that had ambiguous nucleotides,

low quality sequencing (Q phred score < 20), and less than 36 bp were removed. The clean reads from the eye and liver tissues were then assembled into contigs using Trinity v.2.8.4, using default parameters [20]. Mitochondrial and ribosomal RNA mitogenome from the shortfin mako shark [14], were retrieved from NCBI (GenBank accession Number KF361861), and then mapped against the *de novo* transcriptome, in order to remove contaminants, using Bowtie2 program [21]. Additionally, the large 45S and small 5S subunits of rRNA genes from the fish *Cyprinus carpio* (GenBank JN628435.1 and AB015590.1, respectively) were used as reference to remove any possible rRNA contamination. Redundant sequences were removed using a clustering program, namely CD-hit [22]. Subsequently, quality control and quantitative measures of assemblies were performed using TransRate [23]. In addition, to evaluate quality and completeness of *de novo* eye and liver assemblies, and to determine the proportion of paralogs (i.e., duplicate single-copy orthlogs) the program BUSCO v3 (Benchmarking Universal Single-Copy Orthlogs) pipeline was used [24], selecting the Core Vertebrate Gene (CVG) set. The accuracy of sequencing was also tested by percentage of mapped reads against different Chondrichthyes genomes available such as *C. carcharias* [7] (GenBank accession Number QUOW00000000.1), *C. milii* [4] (GenBank accession Number AAVX00000000.2), *C. plagiosum* (unpublished) (GenBank accession Number QPFF00000000.1), *C. punctatum* [6] (GenBank accession Number BEZZ00000000.1), *Leucoraja erinacea* [25] (GenBank accession Number AESE00000000.1), *R. typus* [5] (GenBank accession Number LVEK00000000.2), and *S. torazame* [6] (GenBank accession Number BFAA00000000.1).

2.4. Annotation and function assignment

Candidate protein-coding regions with a minimum cut-off of 100 amino acids

long and single best open reading frames (ORFs) from the *de novo* reference, eye and liver were identified using the TransDecoder v.5.5 program (26). Then, in order to provide comprehensive functional annotation, candidate peptides identified by TransDecoder were compared against different databases, such as the UniProt database [27], (nr) NCBI non-redundant database (nr) (only vertebrate) [28] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [29], using BLASTp (cut-off $1e^{-5}$) [30]. In addition, conserved protein domains were identified against the Pfam-A database using HHMMER v.3.2.1 program [hmmer.org]. All Gene ontology (GO) terms were classified into three categories: Biological process, Molecular functions, and cell component, using the PANTHER (protein annotation through evolutionary relationship) classification tool [31]. Conserved protein domains identified were classified using HMMER2GO v0.17 [https://github.com/sestaton/HMMER2GO].

2.5. Gene expression

Three-trimmed paired-end reads of each sample from eye and liver was aligned based on *de novo* transcriptome using Salmon program [32], in order to quantify transcript abundance. First, transcripts with very low counts across all libraries were filtered out. Then, the reads counts were normalized using the trimmed mean of *M*-values normalization (TMM) using the EdgeR package [33], which uses the likelihood ratio and quasi-likelihood F tests, in R v.3.5 program (http://www.r-project.org/). Once negative binomial models were fitted and dispersion estimates were obtained, the exact test based on quantile-adjusted conditional maximum likelihood (qCML) was used to identify differential expression genes (DEG). In order to know which of the expression levels changed by a certain amount, log fold-change (logFC > 2 < -2) and false discovery rate (FDR) threshold were used. Differentially expressed genes (FDR < 0.01)

were visualized in a heatmap plot using mixOmics package [34] in R program v.3.5. Finally, differentially expressed genes sequences were retrieved from *de novo* transcriptome reference using Better FASTA grep (BFG) program (https://gitlab.com/fethalen/bfg), and then annotated against the NCBI nr vertebrate database using BLASTp (cut-off $1e^{-5}$) and KEGG in order to know metabolic pathways. Finally, the PANTHER program was used to GO. The overall workflow of the transcriptome analysis is presented in Fig 1.

3. Results

3.1. Transcriptome sequence and assembly

A total of 42,784,010 reads, 89,95 Gb billion total bases, and 12,83 Gb of data by library were obtained by Illumina paired-end sequencing technology (HiSeq 4000). After the processing trimming steps, the sequences set were reduced to 41,749,858 (> 96%) high quality reads (Q 20), with an average length of 133 bp (Table 1). As there is no assembled genome of the shortfin mako shark available, a *de novo* transcriptome combining eye and liver sequences was assembled. *De novo* transcriptome yielded a total of 705,940 transcripts. Specifically eye assemblies yielded a total of 426,043 transcripts, whereas liver yielded a total of 302,189 transcripts. The mean length of transcripts of *de novo*, eye, and liver was 631 bp (N50 = 968 bp), 700 bp (N50 = 1,275 bp), and 756 bp (N50 = 1,391 bp), respectively (Table S1).

The comparison of the *de novo*, eye and liver assemblies using core vertebrate set from BUSCO, revealed that more than 90% of queried genes of each assembly were complete. Specifically, out of 2,586 queried genes from *de novo* assembly, 2,386 (92.3%) were complete (single: 1089, duplicate: 1297), 148 were partial and only 53 were missing. Likewise, eye assembly presented 2459 complete genes (single: 1403, duplicate: 1056), 63 were fragmented and 64 were missing, whereas liver presented

2374 complete genes (single: 1468, duplicate: 906), 79 were fragmented and 133 were missing (Fig. S1). The *de novo,* eye, and liver reads mapped against different Chondrichthyes genome yielded higher similarity to *Carcharodon carcharias*, with 148 million (51%), 70 million (56%), and 78 million (48%), respectively. On the other hand, the lowest similarity of reads was to *Callorhinchus milii*, with 19 million (6%), 11 million (9%), and 8 million (5%), respectively (Table 2).

3.2. Annotation and function assignment

The total number of predicted coding region (CDs) longer than 100 amino acids were 358,112 (*de novo*), 256,694 (eye), and 211,418 (liver) (Table S1). Overall, 120,873 (33,75%) contigs *de novo* were annotated against to UniProt, 163,397 (45,63%) contigs to nr NCBI (vertebrate), 104,030 (29,05%) contigs to PFAM, and 199,614 contigs (55,74%) to KEGG (Table S2).

In terms of the Gene Ontology annotations from UniProt, the functional annotation for *de novo*, eye, and liver did not differ in the distribution terms (Fig. 2A, S2). A total of 14,355 contigs were associated with Molecular Function, 19,662 with Biological Process, and 13,681 with Cellular Component (Fig. 2A). These categories were sub-divided into 36 groups, and most frequent terms were: catalytic activity 5,163 and binding 5,039 for Biological Process; cell process 6,332 and cell metabolic process 4,376 for Molecular Function; cell 5,766 and organelle 4,022 for Cellular Component (Fig. 2A). In addition, 360 genes were found associated to the immune system, 1,090 to growth, 520 to behaviour, and 184 to reproduction. In addition, a total of 421 KEGG pathways were identified to the *de novo* transcriptome. The most represented KEGG pathways were Signal Transduction (2,625 transcripts matched), followed by Global and Overview Maps (2,242), Infectious Disease: Viral (1,349), Endocrine System (1,294), and Immune System (1,058). The KEGG pathways classification are shown in

Fig. 2B. The eye and liver presented 471 and 415 KEGG pathways, respectively, and also did not differ in the distribution terms (Fig. S3).

3.3. Gene expression

Based on 637,657 contigs from seven libraries from eye and liver, 25,658 genes were identified in both tissues. Of these 3,774 genes were differently expressed (DE, log2FC and $p < 0.001$). Hierarchical clustering using only significant differently expressed transcripts, revealed distinctive expression profiles for each tissue (Fig. 3A & B). Tissue specific genes cluster were identified as those exhibiting high expression in one tissue versus another one (eye x liver). Liver comprised 2,162 differently expressed genes, whereas eye 1,612 genes. Overexpressed genes in the eye were mainly related to structural and signaling functions. It is of note that nine out of 10 DEGs found in the liver were related with tumor suppression and human diseases. The top 10 DEGs in the eye and liver are presented on the Table 3. The Gene Ontology functional annotation from UniProt for 3,774 DEGs indicated the most common terms as follows: binding (145 transcripts matched) and catalytic activity (132) for Molecular Function; cell process (241) and biological regulation (142) for Molecular Function; cell (233) and cell part (233) for Cellular Component (Fig. 4A). Regarding to KEGG pathways, a total of 364 KEGG pathways were identified to for 3,774 DEGs. The most represented KEGG pathways were Signal Transduction (412 transcripts matched), followed by Global and Overview Maps (388), and Endocrine System (229) (Fig. 4B).

4. Discussion

Over the last few years, omics research in aquaculture and fisheries species have attracted great interest, and a high number of genomic resources have been developed. Herein, we performed a comprehensive eye and liver tissue transcriptome sequencing of

shortfin mako sharks from seven mixed cDNA libraries. The functional annotation revealed that most of transcripts are involved with biological process and molecular function, and signal transduction and global and overview maps pathways In addition, DEGs analysis revealed that the most expressed genes in the eye are involved to structural and signaling functions, whereas most overexpressed genes in the liver are related to tumor suppression and other human diseases.

De novo assembly quality of the eye and liver are congruent with previous bony fish transcriptome studies using *de novo* assembly approach strategies, which reported a queried genes recovery rate of more than 80% [35]. According to [26] more than 80% read mapping is considered to be indicative of a good quality assembly. When compared to other completeness BUSCO analysis for elasmobranchs transcriptome, which ranged between 61% and 81.6% [7,36-38], our results reported higher recovery of CVG. Therefore, the *de novo* transcriptome provided here (> 90% completeness) is a useful resource for future omics studies of this ecologically and economically important shark species, which to date still did not have the complete genome available. In addition, mapping of reads against different complete Chondrichthyes genome species followed an evolutionary relationship similarity, being closer to *C. carcharias* and far to *C. milli*. The most GO common terms were those related to biological process, followed by molecular functions. Similar to our results, cellular and metabolic process for biological process and binding and catalytic activity for molecular functions were the most common term found to liver transcriptome of *Scyliorhinus canicula* [38]. The only study about eye transcriptome also revealed most genes ontology terms for biological process [39].

The eye expression profile differs from those obtained in liver tissues. Most of top 10 ranked DEGs was related with structural (*CRYBA1*, *VIT*, *COL6A6*, *CRB1*,

CHADLH, *SPOCK3*) and signaling (*TRPM3*, *SLC6A15*) functions. The genes related to structural function were involved in encoding βA1- and βA3-crystallins (*CRYBA1*), collagen (*COL6A6*, *CHADLH*), photoreceptor (*CRB1*), and proteoglycan (*SPOCK3*). Photoreceptor disc membrane and structural constituent of eye lens also were the most significant terms for *Salvelinus alpinus* eye transcriptome [40]. Regarding signaling function genes they were involved in calcium channel mediating constitutive calcium ion entry (*TRPM3*), and sodium-dependent neutral amino acid transporter and also may have a role as transporter for neurotransmitter precursors into neurons (*SLC6A15*). The genes (top 10 ranked DEGs) found in the eye were different from those commonly found in the of other fishes [41,42], except *CRYBA1*. The gene *CRYBA1* was present and also overexpressed in other transcriptome-eye shark [39]. The overexpressed photoreceptor gene (*CRB1*) is an indicative that the shortfin mako shark can use the vision as one of the primary senses. In fact, many vertebrates use vision for a variety of fundamental tasks such as navigation, foraging, predator avoidance, and mate choice [42]. The only gene found in the eye that plays a role in tumor suppression was *SPOCK3* that may interfere with tumor invasion [43].

Besides the 10 most DEGs, the genes encephalopsin (*OPSIN3*) and neuropsin (*OPSIN5*) were differentially expressed in nonocular (liver) and ocular (eye) tissues, respectively. *OPSIN3* is a nonocular opsin gene found main in the skin, heart, brain and liver in both vertebrates and invertebrates, and it is associated to nonvisual dependent light-dependent mechanism such as circadian rhythms, seasonal reproduction, and neural development in vertebrates [44,45]. This gene was also not found in the retina of the lantern shark *Etmopterus spinax,* but it was expressed in the skin [39]. On the other hand, *OPSIN5* is a photoreceptor gene found main in the eye and neural tissue [46]. The presence of few genes associated with photoreception expressed in eye indicates a

monochromatic vision of the shortfin mako, as in all other sharks studied to date [47]. Similarly, the presence of two opsin genes were also found in wobbegong sharks *Orectolobus maulatus* and *O. ornatus* [48], and only one in lantern shark [39], further confirming a monochromatic vision on the shortfin mako shark.

Nine out of 10 DEGs in the liver were associated with different types of cancer and other human diseases. Most of 10 DEGs in the liver play a crucial role in tumor suppression and are associated with cancer in humans. Although there are previous shark [36,37,49] and bony fish [50-52] liver transcriptome, the identification of overexpressed genes related to tumor suppression in this organ is presented in this study for the first time. Low cancer incidence has been attributed to sharks but without clear scientific evidence [53]. Indeed, even in low rates, elasmobranchs can develop both benign and malignant neoplasm, and the tumors reported until now are analogous to their counterparts in other organisms, including bony fishes, rodents, and even humans [54]. However, only recently the molecular mechanism related to resistance to cancers has been uncovering [7,15].

Two main reasons have been proposed to explain shark resistance to cancers. The first is the presence of immunity genes in the shark's genome. For example, recently two immunity genes, Legumain (*LGMN*) and BCL2 Associated Athanogene 1 (*BAG1*), were positively selected in different elasmobranchs [15]. The second is the shark genome stability due to its capacity on DNA damage and repair [7]. However, the knowledge of which genes are involved with cancers resistance is still incipient. Our findings revealed a subset of selected (9/top10) overexpressed genes involved in cancer resistance and also, when mutated, related to other human diseases.

Overexpressed genes such as *HABP2* can reduce colony formation and cellular migration of tumor cell related to thyroid cancer [55]. Thyroid neoplasm was reported

only one time in sharks, specifically in one spiny dogfish (*Squalus sucklis*) analyzed from the Pacific coast of Canada [54]. The rarity of this type de cancer in sharks is probably due to overexpression of *HABP2* that act as a cancer suppressor, and also due to high capacity of maintenance of genome stability in sharks [7,55]. Another gene found in liver, *PON2*, is a gene member of the paraoxonases gene family (*PON1*, *PON2* and *PON3*), which is expressed in different tissues, such as the liver. This gene encodes intracellular enzyme that may act as a cellular antioxidant, protecting cells from oxidative stress, which is also a significant factor for carcinoma [56]. Furthermore, *PONs* genes play important roles in cardiovascular diseases and other oxidative stress-related diseases, modulate susceptibility to infection, and especially *PON2* may provide neuroprotection [56].

The genes *NIT2*, *RMC1*, and *FGFRK1* are genes expressed in multiple types of tissues and are potential tumor suppressors. In humans, these genes are associated with suppressing the oncogenic properties of colon cells (*NIT2, RMC1*) and have a negative effect on cell proliferation, which plays an important role in prostate cancer (PCa) initiation and progression (*FGFRK1*) [57-60]. *SERPIND1* is a gene related to processes including inflammation, blood clotting, and cancer metastasis. Moreover, allelic variation in this gene is associated with heparin cofactor II deficiency, leading to increased thrombin generation and hypercoagulable state [61]. This gene is induced by dermatan sulphate, a glycosaminoglycan present main in the skin and connective tissue [62]. Dermatan sulphate accelerates the inhibitory potential of *SERPIND1* 8-fold higher in ray (*Raja radula*) than others mammals [63]. Thus, this gene plays an important role in the wound healing in sharks caused from natural and anthropogenic sources. Despite anecdotal accounts of rapid healing in elasmobranchs, only recently the time and rate of healing [64] as well as its molecular basis have been described [7]. The *METTL27*

belongs to methyltransferase family and it is deleted in Williams syndrome (GeneCards). In addition, the haploinsufficiency of this gene may be cause of certain cardiovascular and musculoskeletal abnormalities observed in the disease (GeneCards).

Another overexpressed gene in the liver, *ITIH3*, has molecular and biological function and is one of constituents of plasma serine protease inhibitors, and is also involved in proinflammatory processes [65]. This gene may stabilize the extracellular matrix through its ability to bind hyaluronic acid and interacts with inflammatory cells [65]. In addition, polymorphism in this gene may be associated with some human diseases involved in inflammatory diseases such as rheumatoid arthritis and inflammatory bowel diseases [66]. Moreover, polymorphism in this gene is also related to schizophrenia, major depressive disorder, and myocardial infarction [67,68]. Furthermore, *ITIH3* is highly expressed in plasma of gastric cancer patients [69].

The genes *CBS* and *A1CF* are related to molecular function as Hydro-lyase catalyzing pathway and posttranscriptional editing of a CAA codon to a UAA codon for stop in APOB mRNA, respectively [70,71]. These genes act as an oncogene related to malignant tumor progression promoting the proliferation of colon, ovarian and breast (*CBS*) and glioma cancer (*A1CF*) [72]. On the contrary, *CBS* has tumor-suppressive effects in glioma [73].

Those findings now reported can be used as a baseline for future development of bioactive compounds from shark-liver for different therapeutic applications. However, we strongly highlight that the shortfin mako shark is a pelagic shark already impacted and largely overfished by commercial fisheries, mainly as a food resource. Therefore, increased captures and further population declines for this additional purpose, is entirely discouraged. On the other hand, it might be possible to combine research activities in order to optimize the sampling of this endangered shark species, especially in cases

when scientific observers are onboard the commercial vessels and can continue to take samples from some specimens that are already dead when captured. Finally, our findings can be used to future studies such as molecular markers uncovering, population genomics, comparative genomics, and biomedical applications.

5. Data accessibility. The raw reads data of the eye and liver transcriptome of the *Isurus oxyrinchus* used in this study are archived in GenBank (SRA accession: PRJNA602859). Code: https://github.com/RRDomingues/DEG_shortfinmako_shark.

Ethics. No ethical approval or permit for animal experimentation was required, as the individuals were not sacrificed specifically for this study. All individuals were sampled by scientific onboard observers during the course of commercial fisheries from Portuguese longline fisheries.

Author contributions. R.C conduced and coordinated fieldwork. R.C, D. H, R.R.D., F.M, F.F. provided resources. N.M and D.H performed labwork and sequencing work. R.R.D and V.M conducted bioinformatics analysis. R.R.D wrote the manuscript. V.P.C, R.C, R.R.D, D.H. F. M, A.A. reviewed the manuscript. All authors approved the final version.

Conflict of interest. The authors declare that there are no conflicts of interests.

Acknowledgements. We would like to thank for Portuguese fishery observer program at IPMA (PNAB – EU/DCF) for the collection of tissue samples. We thank Alex Anoop for bioinformatics support. This work was funded by a Project from Fundação de Amparo à Pesquisa do Estado de São Paulo (Brasil) (Project Ref.: FAPESP

2014/19740-7) and Fundação para a Ciência e Tecnologia (FCT, Portugal). Fellowship for R.R.D. was provided by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2017/02420-8; 2018/21319-9). A.A. is partially supported by the Strategic Funding UID/Multi/04423/2019 through national funds provided by FCT and the European Regional Development Fund - program PT2020, by the European Structural and Investment Funds - Competitiveness and Internationalization Operational Program-COMPETE 2020, and by National Funds through the FCT under the project PTDC/CTA-AMB/31774/ 2017 (POCI-01-0145-FEDER/031774/2017).

References

[1] Andreev PS, Coates MI, Shelton RM, Cooper PR, Smith PM, Sansom IJ. (2015) Upper Ordovician chondrichthyan-like scales from North America. *Palaeontology* ***58***, 691-704.

[2] Weigmann S. (2016) Annotated checklist of the living sharks, batoids and chimaeras (Chondrichthyes) of the world, with focus on biogeographical diversity. *J. Fish. Biol.* ***88***, 837-1037.

[3] Domingues RR, Hilsdorf AWS, Gadig OBF. (2018) The importance of considering genetic diversity in shark and ray conservation policies. *Conserv. Genet.* ***19***, 501-525.

[4] Venkatesh B, Lee AP, Ravi V, Maurya AK, lian MM, et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* ***505***, 174-179.

[5] Read TD, Petit RA, Joseph SJ, Alam MT, Weil MR, et al. (2017) Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. *BMC Genomics* ***18***, 532.

[6] Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, et al. (2018) Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat. Ecol. Evol.* ***2***, 1761-1771.

[7] Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, et al. (2019) White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proc. Natl. Acad. Sci. USA*. ***116***, 4446-445.

[8] Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH. (2017) DNA sequencing at 40: past, present and future. *Nature* ***550***, 345-353.

[9] Ebert DA, Fowler S, Compagno L. (2013) Sharks of the world: a fully illustrated guide. Wild Nature Press, Plymouth.

[10] Rigby CL, Barreto R, Carlson J, Fernando D, Fordham S, et al. (2019). *Isurus oxyrinchus*. The IUCN Red List of Threatened Species 2019: e.T39341A2903170. https://dx.doi.org/10.2305/IUCN.UK.2019-1.RLTS.T39341A2903170.en. Downloaded on 03 August 2020.

[11] Dickson KA, Graham JB. (2004) Evolution of endothermy in fishes. *Physiol Biochem Zool.* ***77***, 998-1018.

[12] Campana SE, Marks L, Joyce W. (2005) The biology and fishery of shortfin mako sharks (*Isurus oxyrinchus*) in Atlantic Canadian waters. *Fish. Res*. ***73***, 341-352.

[13] Byrne ME, Cortés E, Vaudo JJ, Harvey GCMN, Sampson M, Wetherbee BM, Shivji M. (2017) Satellite telemetry reveals higher fishing mortality rates than previously estimated, suggesting overfishing of an apex marine predator. *Proc. R. Soc. B.* ***284***, 20170658.

[14] Chang CH, Shao KT, Lin YS, Tsai AY, Su PX, Ho HC. (2013) The complete mitochondrial genome of the shortfin mako, *Isurus oxyrinchus* (Chondrichthyes, Lamnidae). *Mitochondr. DNA* ***26***, 475-476.

[15] Marra NJ, Richards VP, Early A, Bogdanowicz SM, Bitar PDP, Stanhope MJ, Shivji MS. (2017) Comparative transcriptomics of elasmobranchs and teleosts highlight important processes in adaptive immunity and regional endothermy. *BMC Genomics* ***18***, 87.

[16] Alvarez M, Schrey AW, Richards CL. (2015) Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Mol. Ecol.* ***24***, 710-725.

[17] Wang B, Kumar V, Olson A and Ware D. (2019) Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Front. Genet.* ***10***, 384.

[18] Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

[19] Bolger AM, Lohse M, Usadel B. (2014) Trimmomatic: a flexible trimmer for

Illumina sequence data. *Bioinformatics* ***30***, 2114-2120.

[20] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA et al. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* ***29***, 644-652.

[21] Langmead B, Salzberg SL. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* ***9***, 357-359.

[22] Li W, Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* ***22***, 1658-1659.

[23] Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. (2016) TransRate:reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* ***26***,1134-1144.

[24] Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* ***31***, 3210-3212.

[25] King BL, Gillis JA, Carlisle HR, Dahn RD. (2011) A natural deletion of the *HoxC* cluster in elasmobranch fishes. *Science* ***334****, 1517*

[26] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, et al. (2014) De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* ***8***, 1494-512.

[27] The UniProt Consortium. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic. Acids Res.* ***47****,* 506-515.

[28] Pruitt KD, Tatusova T, Brown GR, Maglott DR. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic. Acids Res.* ***40***, 130-135.

[29] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic. Acids Res*. ***40***, 109-114.

[30] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol*. ***215***, 403-410.

[31] Mi H, Muruganujan A, Casagrande JT, Thomas PD. (2013) Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc*. ***8***, 1551-66.

[32] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. ***14***, 417-419.

[33] Robinson MD, McCarthy DJ, Smyth GK. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* ***26***, 139-140.

[34] Cao KL, Rohart F, Gonzalez I, Dejean S, Gautir B, Bartolo F, Monget P, Coquery J, Yao Z, Liquet B. (2016) mixOmics: Omics data integration project. R package version 6.1.1. https://CRAN.R-project.org/package=mixOmics.

[35] Carruthers M, Yurchenko AA, Augley JJ, Adams CE, Herzyk P, Elmer K. (2018) De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics* ***19***, 32.

[36] Mulley JF, Hargreaves AD, Hegarty MJ, Heller SR, Swain MT. (2014) Transcriptomic analysis of the lesser spotted catshark (*Scyliorhinus canicula*) pancreas, liver and brain reveals molecular level conservation of vertebrate pancreas function. *BMC Genetics* ***15***:1074.

[37] Chana-Munoz A, Jendroszek A, Sønnichsen M, Kristiansen R, Jensen JK, Andreasen PA, et al. (2017) Multi-tissue RNA-seq and transcriptome characterization of the spiny dogfish shark (*Squalus acanthias*) provides a molecular tool for biological research and reveals new genes involved in osmoregulation. *PLoS ONE* ***12***, e0182756.

[38] Machado AM, Almeida T, Mucientes G, Esteves PJ, Veríssimo A, Castro LF. (2018) De novo assembly of the kidney and spleen transcriptome of the cosmopolitan blue shark, *Prionace glauca*. *Mar. Genomics* ***37***, 50-53.

[39] Delroisse J, Duchatelet L, Flammang P, Mallefet J. (2018) De novo transcriptome analyses provide insights into opsin-based photoreception in the lanternshark *Etmopterus spinax*. *PLoS ONE* ***13***, e0209767.

[40] Christensen KA, Rondeau EB, Minkley DR, Leong JS, Nugent CM, Danzmann RG, et al. (2018) The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS ONE* ***13***, e0204076.

[41] McGowan KL, Passow CN, Arias-Rodriguez L, Tobler M, Kelley JL. (2019) Expression analyses of cave mollies (*Poecilia mexicana*) reveal key genes involved in the early evolution of eye regression. *Biol. Lett.* ***15***, 20190554.

[42] Musilova Z, Cortesi F, Matschineri M, Davies WIL, Patel JS, et al. (2019). Vision using multiple distinct rod opsins in deep-sea fishes. *Science* ***384***, 588-592.

[43] Nakada M, Yamada A, Takino T, Miyamori H, Takahashi T, Yamashita J, Sato H. (2001) Suppression of Membrane-type 1 matrix metalloproteinase (MMP)-

mediated MMP2 activation and tumor invasion by testican 3 and its splicing variant gene product, N-Tes 1. *Cancer Res*. ***61***, 8896-8902

[44] Koyanagi M, Takada E, Nagata T, Tsukamoto H, Terakita A. (2013) Homologs of vertebrate Opn3 potentially serve as a light sensor in nonphotoreceptive tissue. *Proc. Natl. Acad. Sci. USA*. ***110***, 4998-5003.

[45] Beaudry FEG, Iwanicki TW, Mariluz BRZ, Darnet S, Brinkmann H, Schneider P, Taylor JS. (2017) The non-visual opsins: eighteen in the ancestor of vertebrates, astonishing increase in ray-finned fish, and loss in amniotes. *J. Exp. Zool*. ***328***, 685-696.

[46] Kumbalasiri T, Provencio I. (2005) Melanopsin and other novel mammalian opsins. *Exp. Eye Res*. ***81***, 368-375.

[47] Hart NS. (2020) Vision in sharks and rays: Opsin diversity and colour vision. *Semin. Cell Dev. Biol*., doi.org/10.1016/j.semcdb.2020.03.012

[48] Theiss SM, Davies WIL, Collin SP, Hunt DM, Hart NS. (2012) Cone monochromacy and visual pigment spectral tuning in wobbegong sharks. *Biol. Lett.* ***8,*** 1019-1022.

[49] Goshima M, Sekiguchi R, Matsushita M, Nonaka M. (2016) The complement system of elasmobranches revealed by liver transcriptome analysis of a hammerhead shark, *Sphyrna zygaena*. *Dev. Comp. Immunol.* ***61***, 13-24.

[50] Chini V, Cattaneo AG, Rossi F, Bernardini G, Terova G, Saroglia M, Gornati R. (2008) Genes expressed in Blue Fin Tuna (*Thunnus thynnus*) liver and gonads. *Genes* ***410***, 207-213.

[51] Sun Z, Tan X, Liu Q, Ye H, Zou C, Ye C. (2019) Liver transcriptome analysis and de novo annotation of the orange –spotted groupers (*Epinephelus coiodes*) under cold stress. *Comp. Biochem. Physiol. – Part D Genomics Proteomics* ***29***, 264-273.

[52] Watanabe L, Gomes F, Vianez J, Nunes M, Cardoso J, Lima C, et al. (2018) De novo transcriptome based on next-generation sequencing reveals candidate genes with sexspecific expression in *Arapaima gigas* (Schinz, 1822), an ancient Amazonian freshwater fish. *PLoS ONE* ***13***, e0206379.

[53] Ostrander GK, Cheng CK, Wolf JC, Wolfe MJ. (2004) Shark cartilage, cancer and the growing threat of pseudoscience. *Cancer Res*. ***64***, 8485-7491.

[54] Cameron AT, Vincent S. (1915) Note on an enlarged thyroid occurring in an elasmobranch fish (*Squalus sucklii*). *J. Med. Res.* ***27***, 251-55.

[55] Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, et al. (2015) Germline HABP2 mutation causing familial nonmedullary thyroid cancer. *N. Engl. J. Med.* ***373***, 448-455.

[56] Witte I, Altenhöfer S, Wilgenbus P, Amort J, Clement AM, Pautz A, et al. (2011) Beyond reduction of atherosclerosis: PON2 provides apoptosis resistance and stabilizes tumor cells. *Cell Death Dis.* 13:e112.

[57] Costa LG, Cole TB, Garrick JM, Marsillach J, Furlong CE. (2017) Metals and paraoxonases. *Adv. Neurobiol.* ***18***, 85-111.

[58] Lin CH, Almeida T, Mucientes G, Esteves PJ, Veríssimo A, Castro LF. (2018) Growth inhibitory effect of the human NIT2 gene and its allelic imbalance in cancers. *FEBS. J*. ***274***, 2946-2956.

[59] Wan X, Corn PG, Yang J, Palanisamy N, Starbuck MW, et al. (2014) Prostate cancer cell-stromal cell crosstalk via FGFR1 mediates antitumor activity of dovitinic in bone metastases. *Sci. Transl. Med.* ***3***, 252ra122.

[60] Zheng B, Chai R, Yu X. (2015) Downregulation of NIT2 inhibits colon cancer cell proliferation and induces cell cycle arrest through the caspase-3 and PARP pathways. *Int. J. Mol. Med.* ***35***, 1317-1322.

[61] Heit C, Jackson BC, Andrews M, Wright MW, Thompson DC, Silverman GA, Nebert DW, Vasiliou V. (2013) Update of the human and mouse SERPIN gene superfamily. *Hum. Genomics* ***7***, 22.

[62] Rebl A, Goldammer T. (2018) Under control: The innate immunity of fish from the inhibitors'perspective. *Fish Shellfish Immun.* 77: 328-349.

[63] Mansour BM, Dhahri M, Vénisse L, Jandrot-Perrus M, Chaubet F, Maarouf RM. (2009) Mechanism of thrombin inhibition by heparin cofactor II and antithrombin in the presence of the ray (*Raja radula*) skin dermatan sulfate. *Thromb. Res.* 123: 902-908.

[64] Chin A, Mourier J, Rummer J. (2015) Blacktip reef sharks (*Carcharhinus melanopterus*) show high capacity for wound healing and recovery following injury. *Conserv. Physiol.* 3: cov062

[65] Fries E, Kaczmarczyk A. (2003) Inter-a-inhibitor, hyaluronan and inflammation. *Acta Biochim. Pol.* ***50***, 735-742.

[66] Zhuo L, Hascall VC, Kimata K. (2004) Inter-a-trypsin inhitibor, a covalent protein-glycosaminoglycan-protein complex. *J. Biol. Chem*. ***279***, 38079-38082.

[67] Ebana Y, Ozaki K, Inue J, Sato H, Iida A, et al. (2007) A functional SNP in ITIH3 is associated with susceptibility to myocardial infarction. *J. Hum. Genet.* ***52***, 220-229.

[68] Hamshere ML, Walters JTR, Smith R, Richards Al, Green E, et al. (2013) Genome-wide significant associations in schizophrenia to *ITIH3/4*, *CACNA1C* and *SDCCAG8*, and extensive replication of associations reported by the Schizophrenia PGC. *Mol. Psychiatry* ***18***, 708-712.

[69] Chong PK, Lee H, Zhou J, Liu SC, Loh MCS, Wang TT, Chan SP, Smoot DT, Ashktorab H, So JBY, Lim KH, Yeoh KG, Lim YP. (2010) ITIH3 is a potential biomarker for early detection of gastric cancer. *J. Proteome Res.* ***97***, 3671-3679.

[70] Kraus JP, Oliveriusová J, Sokolová J, Kraus E, Vlček C, Franchis R, Maclean KN, Bao L, Bukovská, Patterson D, Pačes V, Ansorge W, Kožich V. (1998) The human Cystathionene β-synthase (CBS) gene: complete sequence, alternative splicing, and polymorphisms. *Genomics* ***52***, 312-324.

[71] Lan L, Ui A, Nakajima S, Hatakeyama K, Hoshi M, Watanabe R, Janicki SM, Ogiwara H, Kohno T, Kanno S, Yasui A. (2010) The ACF1 complex is required for DNA double-strand break repair in human cells. *Mol. Cell* ***40***, 976-987.

[72] Zhu H, Blake S, Chan K, Pearson RB, Kang J. (2018) Cystathionine β-Synthase in Physiology and Cancer. *BioMed. Res. Int*. 3205125.

[73] Song Y, Shao L, Xue Y, Ruan X, Liu X, Yang C, Zheng J, Shen S, Chen J, Li Z, Liu Y. (2019) Inhibition of the aberrant A1CF-FAM224miR-590-3p-ZNF143 positive feedback loop attenuated malignant biological behaviors of glioma cells. *J. Exp Clin. Cancer Res.* ***38***, 248.

Tables

Table 1. Summary of the shortfin mako (*Isurus oxyrinchus*) sequencing quality.

Sample	Tissue	Raw reads	Filtered reads	Effective rate (%)	Q20 (%)	GC content (%)
T6	eye	46,304,612	44,710,905	98.16	96.47	46.19
T12	eye	41,041,779	39,598,192	98.01	96.65	46.03
T9	eye	42,497,353	40,870,743	97.79	96.59	45.68
T1	liver	43,847,078	41,757,990	96.84	96.69	45.03
T4	liver	36,390,692	34,663,509	97.31	96.26	45.15
T10	liver	49,027,688	47,169,940	97.88	96.57	45.11
T13	liver	40,378,929	38,496,955	96.96	96.67	45.26

Sample: sample name; Raw reads: total amount of reads of raw data; Filtered reads: total amount of reads of clean data; Effective Rate (%): (Filtered reads/Raw reads)*100%; Q20: (Base count of Phred value > 20) / (Total base count); GC content: (G & C base count) / (Total base count).

Table 2. Alignment summary from the *de novo*, eye, and liver transcriptome assemblies mapping of shortfin mako (*Isurus oxyrinchus*).

Assembly	**Mapped reads**		
	de novo	**eye**	**liver**
Carcharodon carcharias	148,576,955 (51.72%)	70,455,462 (56.28%)	78,121,493 (48.20%)
Chiloscyllium punctatum	24,671,897 (8.59%)	14,153,932 (11.3%)	10,517,965 (6.49%)
Chiloscyllium plagiosum	31,851,357 (11.08%)	18,108,516 (14.47%)	13,742,841 (8.48%)
Callorhinchus milii	19,288,507 (6.72%)	10,917,621 (8.72%)	8,370,886 (5.16%)
Leucoraja erinacea	24,115,727 (8.39%)	13,669,522 (10.92%)	10,446,205 (6.45%)
Rhincodon typus	34,753,571 (12.1%)	19,758,725 (15.79%)	14,994,846 (9.25%)
Scyliorhinus torazane	32,300,689 (11.24%)	17,817,165 (14.23%)	14,483,524 (8.93%)

Table 3. Top differentially expressed genes (top 10 ranked) in liver and eye transcriptome of the shortfin mako shark (*Isurus oxyrinchus*). The liver and eye value represents the scaled mean expression values (log2FC, $p < 0.01$).

	Transcript	UniProt	Gene	Protein name	Liver	Eye
1	TRINITY_DN3478_c0_g1_i7.p1	sp|P07317|CRBA1	Cryba1	Beta-crystallin A1	-5.3281	7.8058
2	TRINITY_DN132_c0_g1_i4.p1	sp|Q6UXI7|VITRN	VIT	Vitrin	-3.3665	7.7120
3	TRINITY_DN1568_c0_g1_i7	sp|Q8C6K9|CO6A6	Col6a6	Collagen alpha-6 (VI) chain	-3.6095	7.1011
4	TRINITY_DN19788_c0_g1_i1.p1	sp|Q9H2J7|S6A15	SLC6A15	Sodium-dependent neutral amino acid transporter B (0)AT2	-4.1715	6.4464
5	TRINITY_DN12239_c0_g1_i3.p1	sp|Q93YS4|AB22G	ABCG22	ABC transporter G family member 22	-4.0880	6.3514
6	TRINITY_DN13219_c0_g1_i5.p1	sp|P82279|CRUM1	CRB1	Protein crumbs homolog 1	-4.2842	5.9887
7	TRINITY_DN260_c0_g2_i1.p1	sp|Q9HCF6|TRPM3	TRPM3	Transient receptor potential cation channel subfamily M member 3	-5.3281	5.8574
8	TRINITY_DN398_c1_g1_i5.p1	sp|Q6UXK5|LRRN1	LRRN1	Leucine-rich repeat neuronal protein 1	-5.3281	5.8389
9	TRINITY_DN4946_c0_g3_i1.p1	sp|Q6NUI6|CHADL	CHADL	hondroadherin-like protein	-5.3281	5.8246
10	TRINITY_DN6338_c0_g1_i1.p1	sp|Q8BKV0|TICN3	*Spock3	Testican-3	-5.3281	5.8040
11	TRINITY_DN3019_c1_g1_i4.p1	sp|Q90952|PON2	*PON2	Serum paraoxonase/arylesterase 2	10.0371	-3.5295
12	TRINITY_DN304_c0_g1_i4.p1	sp|Q61704|ITIH3	*Itih3	Inter-alpha-trypsin inhibitor heavy chain H3	3.8265	-5.3281
13	TRINITY_DN5568_c2_g1_i11.p1	sp|Q4VBV9|NIT2	*nit2	Omega-amidase NIT2	7.6441	-4.2162
14	TRINITY_DN1350_c0_g1_i9.p1	sp|Q5R9H4|A1CF	*A1CF	APOBEC1 complementation factor	6.3413	-5.3281
15	TRINITY_DN11059_c0_g1_i1.p1	sp|Q8N6F8|MET27	*METTL27	Methyltransferase-like protein 27	8.2832	-5.3281
16	TRINITY_DN1202_c0_g2_i4.p2	sp|Q96DM3|RMC1	*RMC1	Regulator of MON1-CCZ1 complex	6.7105	-5.3281
17	TRINITY_DN4357_c0_g2_i1.p1	sp|P35520|CBS	CBS	Cystathionine beta-synthase	4.0142	-2.5915
18	TRINITY_DN1872_c1_g2_i3.p1	sp|P49182|HEP2	*Serpind1	Heparin cofactor 2	6.2890	-5.3281
19	TRINITY_DN3384_c0_g1_i4.p1	sp|Q7T2H2|FGRL1	*FGFRL1	Fibroblast growth factor receptor-like 1	8.5071	-3.1488
20	TRINITY_DN5252_c0_g1_i3.p1	sp|Q14520|HABP2	*HABP2	Hyaluronan-binding protein 2	7.3640	-4.0491

* means genes related to tumor suppression and human diseases.

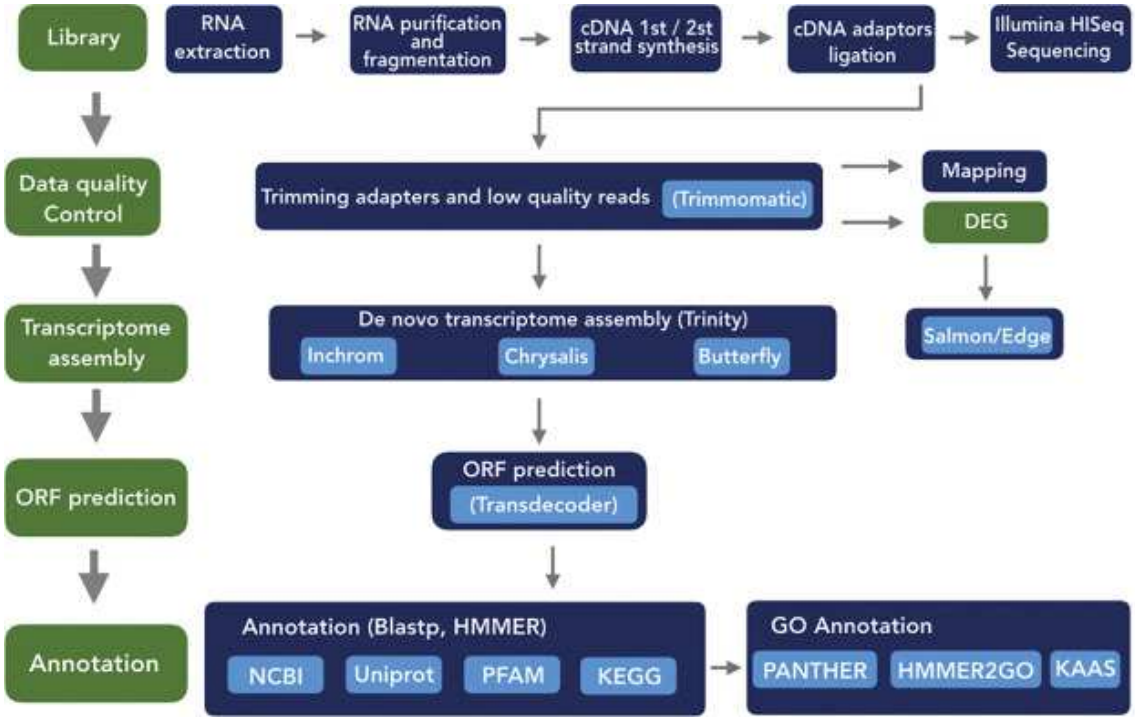

Fig. 1. Overview of transcriptome analysis performed. Green: main steps since library preparation until annotation. Dark blue: details about each step. Light blue: programs and/or database used.

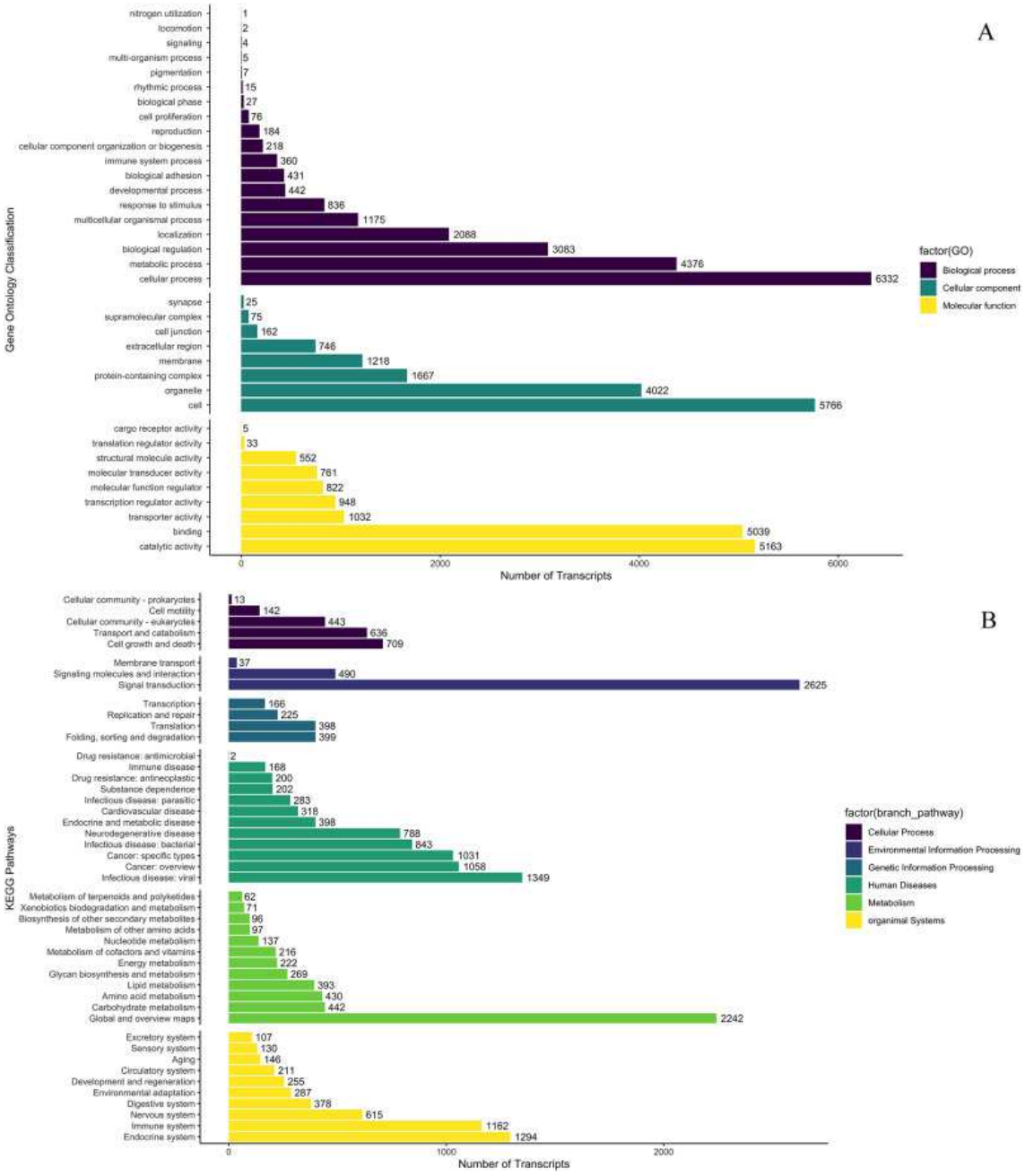

Fig. 2. (A) Gene ontology classification of functional annotation of *de novo* transcriptome from the shortfin mako shark (*Isurus oxyrinchus*). The results are summarized in three categories: Molecular function, biological process, and cellular component. (B) Pathway classification of *de novo* transcriptome according to the definition in KEGG. The pathways were clustered into cellular process, environmental Information processing, genetic information processing, metabolism, and organismal systems.

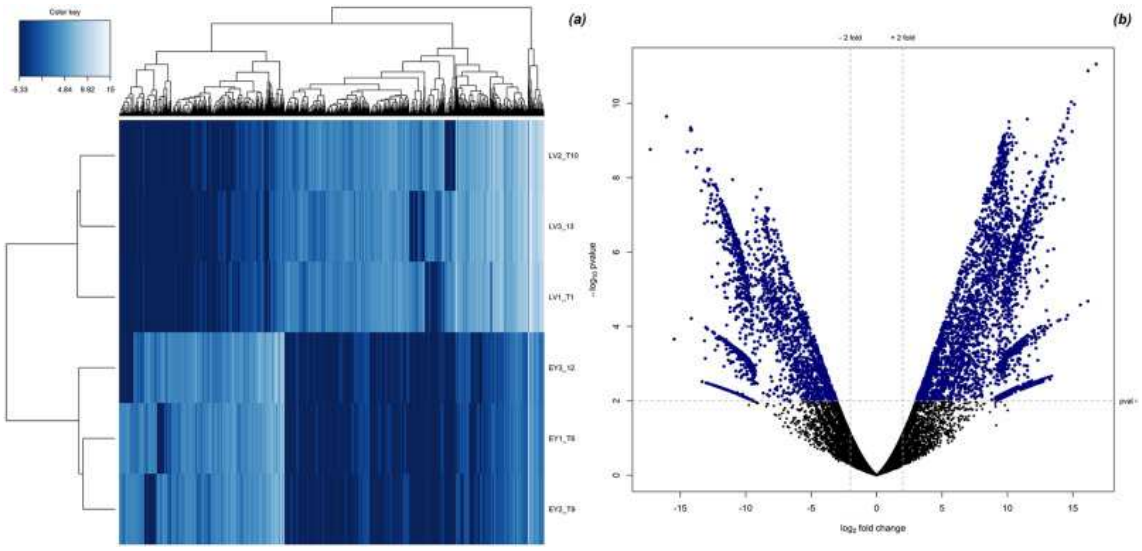

Fig. 3. (*a*) Heatmap of hierarchically clustered differentially expressed genes profile in the liver (up) and eye (down) of the shortfin mako shark (*Isurus oxyrinchus*). Genes are displayed as horizontal lines across samples (columns). Genes with more similar expression cluster together. Overexpressed genes in a sample are colored light blue, while low expressed genes are displayed in dark blue. (*b*) Volcano plot of differentially expressed genes (DEGs) showing the standardized mean difference and the adjusted p-value for the 3,774 genes. The blue dots represent DEG p <0.05 in the liver *vs* eye.

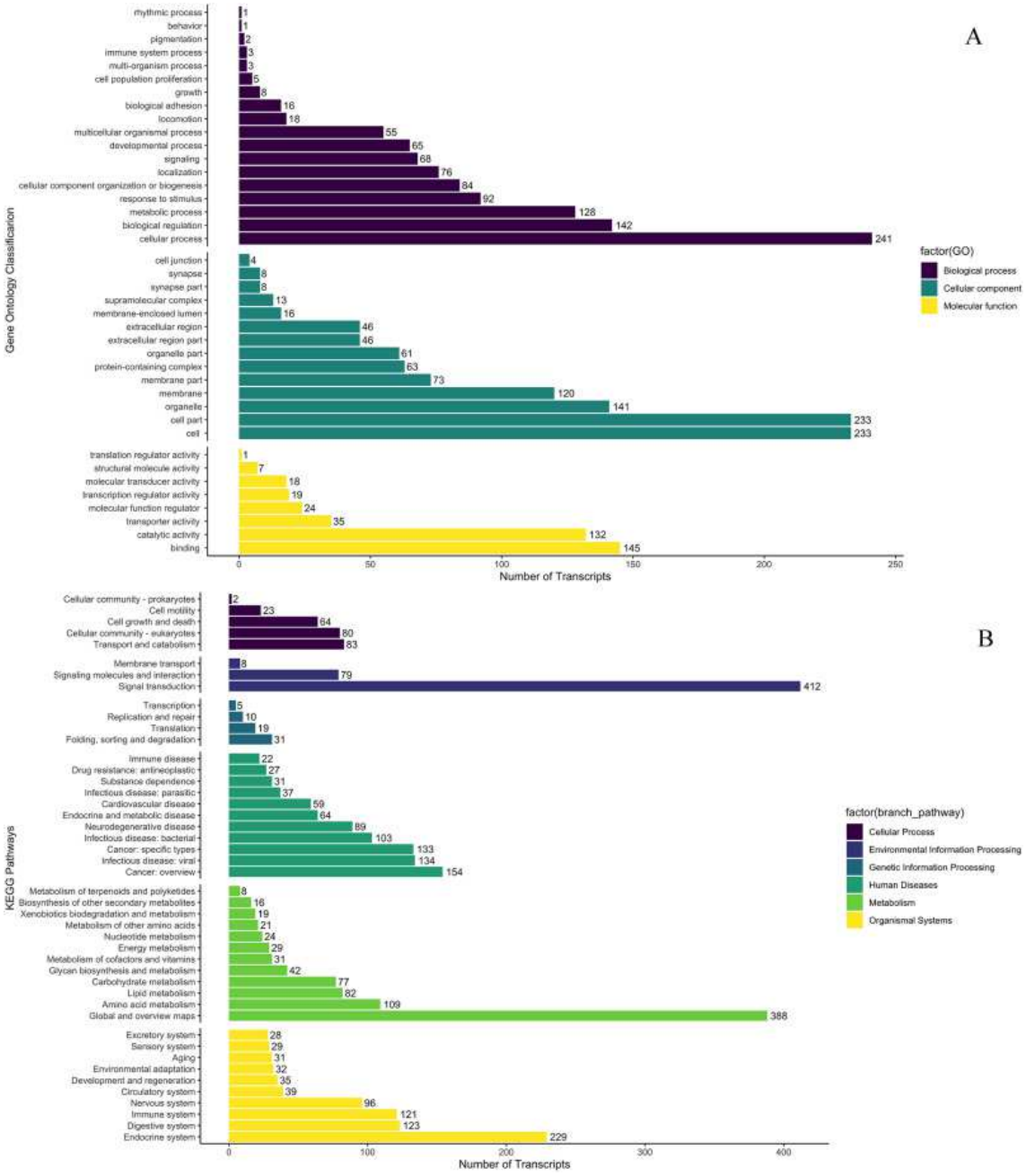

Fig. 4. (A) Gene ontology classification of functional annotation of Differentially Expressed Genes (DEGs) transcriptome from the shortfin mako shark (*Isurus oxyrinchus*). The results are summarized in three categories: Molecular function, biological process, and cellular component. (B) Pathway classification of *de novo* transcriptome according to the definition in KEGG. The pathways were clustered into cellular process, environmental Information processing, genetic information processing, metabolism, and organismal systems.