


Article

Nanopore-Based Full-Length Transcriptome Sequencing: In-Depth Exploration of Green Sea Turtle (*Chelonia mydas*) Genome

Qi Huang ^{1,*}, Yongjun Sun ², Linlin Zhao ^{3,*}, Wenbo Zhu ⁴, Fei Shao ⁵, Jin Xu ⁵  and Yongjian Qin ⁵¹ School of Food and Biological Engineering, Yantai Institute of Technology, Yantai 264000, China² Shandong Homey Aquatic Development Co., Ltd., Weihai 264305, China; sun29889@163.com³ First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266000, China⁴ Shandong Marine Resource and Environment Research Institute, Yantai 264006, China; zhuwenbo@shandong.cn⁵ Shandong Forestry Protection and Development Service Center, Jinan 250014, China; shaofei@shandong.cn (F.S.); xujincaf@163.com (J.X.); qinyongjian529@shandong.cn (Y.Q.)

* Correspondence: 18553336462@163.com (Q.H.); yxq19880209@126.com (L.Z.)

Abstract

The green sea turtle (*Chelonia mydas*), a widely distributed species, plays a crucial role in maintaining the marine ecosystem. However, studies on *C. mydas* require accurate and comprehensive genome annotation information. Long-read direct transcriptome data of *C. mydas* were obtained using direct RNA sequencing on the Oxford Nanopore Technologies (ONT) platform from blood tissue of a single captive individual. A total of 4061 novel transcripts were obtained by comparing long-read direct transcripts with genome annotation of *C. mydas*. We also predicted 2402 CDSs on the novel transcripts. Among them, 1208 (50.29%) had functional annotation information in the databases. In addition, we predicted and analyzed AS events, fusion transcripts, methylation sites, poly(A)s, and lncRNAs in the *C. mydas* long-read direct transcriptome. Overall, our study provides the a long-read direct blood transcriptome for *C. mydas* to complement and improve its genome annotation. This valuable resource will contribute to future research on *C. mydas*. Additionally, the analyses of transcriptome structure mentioned above may provide new insights and ideas for the study of *C. mydas*.

Keywords: *Chelonia mydas*; full-length transcriptome; ONT sequencing; novel transcript; functional annotation

Key Contribution: Considering that the green sea turtle plays a crucial role in maintaining the marine ecosystem, the present study generated a long-read direct blood transcriptome dataset that complements existing genomic resources for *C. mydas*.



Academic Editor: Jose Martin Pujolar

Received: 16 March 2026

Revised: 27 April 2026

Accepted: 29 April 2026

Published: 30 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The green sea turtle *Chelonia mydas* is a species of the genus *Chelonia* Brongniart, which belongs to the family Cheloniidae under the order Testudines. In general, *C. mydas* is distributed in subtropical, temperate, and tropical seas from 40° N to 40° S latitude [1]. It has a wide distribution with complex migration routes and numerous habitats [2,3]. *C. mydas* suffers hypothermia, stunning in seawater below 10 °C [4]. In addition to affecting seagrass ecosystems by grazing and contributing to biotic seed dispersal [5,6], *C. mydas*

also has positive effects on the resilience of coral reef ecosystems [7]. Therefore, as a widely distributed herbivore, *Chelonia mydas* is crucial to marine ecology.

C. mydas faces many threats such as climate change, marine debris, and fisheries by-catch [8–10]. This species has been included in “The IUCN Red List of Threatened Species” since 2004 [11]. Understanding the genetic diversity of *C. mydas* is critical to developing conservation strategies [12]. Meanwhile, *C. mydas* has a complex life history, mainly characterized by migration to different habitats at various life stages [13]. However, a complete understanding of the turtle’s life history during the marine stage still presents several knowledge gaps [14]. The longest documented lifespan of *C. mydas* is 75 years [15–18]. Some anecdotal evidence suggests that certain individuals may live over 100 years [15]. However, the longevity model and related mechanisms of *C. mydas* have not been fully elucidated. Moreover, as an ancient species, *C. mydas* also has great potential for studying the history and process of biological evolution [19].

High-quality and blood transcriptome data can support research on the above questions related to *C. mydas*. Wang et al. (2013) first assembled the genome of *C. mydas* using next-generation sequencing (NGS) technology, and the results suggested that turtles have existed for over 200 million years [20]. The finding is consistent with fossil evidence on turtles [21]. Unfortunately, the draft genomes contain inaccuracies in gene count [22]. This makes it difficult to accurately interpret the genetic information. RNA-seq has been used to improve the genomes of several organisms [23–25]. It has also been used in studies related to FP tumor infection in *C. mydas* [26,27]. However, these studies used only short transcripts obtained through NGS technology. Although short-read sequencing offers advantages over long-read sequencing in terms of accuracy and point mutation identification [28,29], it still has limitations in reconstructing and quantifying transcript isoforms [30–32]. In contrast, the long-read sequencing can improve the detection of transcript isoforms [33,34]. This is because long-read sequencing can sequence long-read direct transcripts and accurately identify splice sites, poly(A) sites, and transcription start sites [35]. So far, long-read direct transcripts obtained through long-read sequencing have successfully improved genome annotation in several species, such as *Rhincodon typus*, *Cydia pomonella* L., and *Danio rerio* [36–38]. Even in organisms without reference genomes, long-read direct blood transcriptomes can serve as substitute reference databases [39]. Therefore, long-read direct transcripts can be used to improve the *C. mydas* genome annotation. The Oxford Nanopore Technologies (ONT) platform, one of the two representative long-read sequencing technologies [40–42], was used in this study to generate long-read direct transcripts of *C. mydas*. The maximum read length can exceed 2 Mb [43]. Direct RNA sequencing using this platform eliminates amplification bias and accurately identifies transcript boundaries and structures [44,45]. It can also detect poly(A) lengths and methylation (m5C and m6A) sites [45,46]. The former is related to mRNA translation efficiency and stability, while the latter participates in multiple biological processes, including RNA metabolism [47–49].

In this study, the ONT platform was used to generate a long-read direct blood transcriptome of *C. mydas* from one individual. Blood was selected as the sampling tissue primarily based on the following key considerations: Firstly, blood samples can be easily obtained through non-lethal methods that cause minimal harm to individual animals, meeting ethical requirements. Blood has been widely utilized in transcriptomic studies of wild animals. Secondly, as a core component of the circulatory system, blood provides a comprehensive reflection of an organism’s overall physiological state and systemic metabolic changes, thereby demonstrating high representativeness. However, blood transcriptome profiles cannot fully represent the transcriptomic landscape of other tissues or the entire species. Due to read length characteristics (raw read N50 ~600 bp), the term long-read direct transcript refers to consensus sequences containing identifiable 5’ and 3’ ends with

poly(A) tails, not necessarily complete coverage of entire coding regions. The ONT platform was used to generate an efficient long-read direct blood transcriptome of *C. mydas* from a single individual, and these results yielded a rich and complete dataset, laying the foundation for discovering novel transcripts and CDSs. This can greatly improve the genome annotation of the *C. mydas*. Additionally, analyses of transcriptome structure can further supplement genome annotation information. The primary purpose of this study is to provide more comprehensive blood transcriptome annotation information for future research on *C. mydas*. In addition, transcriptome structure information may provide novel insights into the biological study of *C. mydas*.

2. Materials and Methods

2.1. Sample Collection and RNA Extraction

Blood was collected from a *C. mydas* at Haichang Ocean Park, Yantai, China. After rapid freezing with liquid nitrogen, the blood samples were stored in an ultra-low temperature freezer for subsequent RNA extraction. The TRNzol Universal Total RNA Extraction Kit (DP424; Tiangen Biotech Co., Ltd., Beijing, China) was used to extract total RNA from blood, strictly following the manufacturer's instructions. Total RNA was treated with DNase I to remove residual genomic DNA. RNA concentration and integrity were assessed using a NanoDrop (Thermo Fisher Scientific, Wilmington, DE, USA), Agilent 2100 (Agilent Technologies, Inc., Santa Clara, CA, USA), Agilent RNA 6000 Nano Kit (Agilent Technologies, Inc., Santa Clara, CA, USA), and 1% agarose gel electrophoresis. The RNA Integrity Number (RIN) of the sample was 7.8. After meeting the sequencing requirements for integrity and concentration, the RNA was used for subsequent studies.

2.2. Library Preparation and Oxford Nanopore PromethION Sequencing

RNA purification Kit (DP412; Tiagen Biotech Co., Ltd., Beijing, China) was used to purify approximately 1.0 µg of poly(A)-tailed mRNA. Subsequently, sequencing adapters and motor proteins were sequentially added to the purified product for library preparation. The constructed library was loaded onto the ONT sequencing platform using a flow cell. PromethION sequencer and PromethION Flow Cells (version 9.4) were then used for sequencing.

2.3. Sequencing Data Processing

The raw reads obtained from the ONT platform were in "fast5" format. The data format was converted from "fast5" to "fastq" through base calling using GUPPY software (version 3.2.6; <http://guppy-pe.sourceforge.net/>, accessed on 10 May 2025) with `rna_r9.4.1.1_70bps_hac`. The "fastq" file contains sequence information and the corresponding quality values. NanoFilt software (version 2.6.0; <https://github.com/wdecoster/nanofilt>, accessed on 10 May 2025), with parameters `-q 7 -l 50`, was used to filter short fragments and low-quality reads (length < 50 bp, Qscore < 7) from the raw "fastq" data and to remove the adaptors, producing clean reads for later analysis. Furthermore, Consensus sequences of the direct RNA error-corrected data were obtained using Flair software (version: 1.4.0; <https://bioconductor.org/packages/oldstats/bioc/flair/>, accessed on 10 May 2025). Meanwhile, GMAP was used to assemble these highly similar clean reads into consensus sequences [50]. Using GMAP [50], the consensus sequences were obtained and compared with genome annotation of *C. mydas* [20]. Based on the comparison results, the stringTie program (version 2.1.2; <http://ccb.jhu.edu/software/stringtie/index.shtml>, accessed on 10 May 2025) was used to remove redundancy with the following parameters: `--conservative -L -R`, resulting in non-redundant long-read direct transcripts of *C. mydas*. In this study, long-read direct transcripts denote consensus sequences with identifiable 5' and

3' ends and poly(A) tails; they do not guarantee complete coverage of full coding sequences. Minimap2 software (version 2.17-r941; <https://github.com/lh3/minimap2>, accessed on 10 May 2025) was used to compare the raw reads in “fast5” format to the *C. mydas* reference genome (version 2; https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_015237465.2/, accessed on 10 May 2025) with the obtained long-read direct transcripts using the following parameters: -ax splice-uf-k14 [51]. The number of mapped reads was tallied, and the mapping rate was calculated. To mitigate the impact of ONT sequencing errors on the detection of structural variations (AS events, fusions, methylation sites), we required that all identified structural features be supported by at least 3 high-quality aligned reads.

2.4. Prediction of Novel Transcripts

Due to software or data limitations, reference genome annotations may contain certain errors. Therefore, optimizing the structure of originally annotated transcripts is necessary to improve the accuracy of *C. mydas* genome annotations. Gffcompare software (version 0.11.2; <http://ccb.jhu.edu/software/stringtie/gffcompare.shtml>, accessed on 10 May 2025) was used to compare the known genome transcripts with the long-read direct transcripts, using the following parameters: -R-C-K-M. Novel transcripts were explicitly defined as sequences with gffcompare class codes O, J, I, X, or U that were absent from the reference annotation. J (novel isoforms of known genes) and U (intergenic novel transcripts) were treated separately in downstream analyses owing to differences in reliability and interpretation. When mapped reads supported regions beyond known gene boundaries, UTRs were expanded upstream and downstream to adjust transcript boundaries. In general, novel transcripts fall into five categories: other regions overlapping with reference exons on the same strand (O); one multi-exon matching at least (J); exonic overlaps on the opposite strand (X); transcripts fully contained within introns of reference genes (I); and unknown novel transcripts (U). Additionally, the distribution of known and novel transcripts on the chromosomes was analyzed and plotted.

2.5. CDS Prediction and Annotation

TransDecoder (version: 5.5.0; <http://transdecoder.sourceforge.net/>, accessed on 10 May 2025) was used to predict CDSs in the novel transcripts. The parameters used were: -m 50, -single_best_only. Functional annotation of CDSs refers to assigning metabolic pathways and functions based on existing databases. To obtain comprehensive annotation information, function annotations using Nr, COG, GO, Uniprot, and KEGG databases were performed using DIAMOND software (version: 2.6.0; [52]). The parameter used was “e-value 1e−5”. Domain annotation using the Pfam database was performed using HMMER (hmmscan, version 3.1; <http://hmmer.org/>). The parameter used was “e-value 0.01.” The number of annotated CDSs identified from the above databases was visualized using an UpSet Plot in TBtools software (version 1.087; [53]).

2.6. Structure Analyses of Long-Read Transcriptome

2.6.1. Analysis of Alternative Splicing (AS) Events and Fusion Transcripts

Alternative splicing (AS) refers to post-transcriptional mRNA processing. Precursor mRNAs transcribed from a single gene can undergo diverse splicing patterns, through which distinct exons are selectively retained or excluded to generate multiple mature mRNA isoforms. These varied transcripts are further translated into different protein variants, thereby driving the diversification of biological functions. In this study, putative alternative splicing events were predicted from the long-read direct RNA transcriptome of *C. mydas* using SUPPA2 (<https://github.com/comprna/SUPPA>, accessed on 10 May 2025; [54]) with parameters -f ioe-e SE SS MX RI FL. Fusion transcripts are chimeric RNAs formed by the end-to-end ligation of coding regions derived from at least two independent genes and

potentially regulated by shared regulatory elements. Here, only transcripts supported by a minimum of 10 high-quality aligned reads were retained for downstream fusion identification to reduce technical bias and improve analytical reproducibility. Fusion transcripts were detected using Tofu software (v13.0.0; [55]) under default parameters. The stringent screening criteria were defined as follows: (a) a transcript must uniquely map to a minimum of two distinct gene loci in the *C. mydas* genome annotation; (b) each anchored gene locus covers no less than 10% of the full transcript length; (c) the overall mapped coverage against the *C. mydas* reference genome exceeds 99%; (d) the genomic distance between two matched gene loci is no less than 100 kb [37]. Notably, all identified AS events and fusion transcripts in the present study are defined as candidate molecular features. Further biological replicates and experimental validation are required to confirm their authenticity and functional relevance, and no FDR-based multiple testing correction was applied in this analysis.

2.6.2. Association Analysis of Poly(A) Length and Transcript Expression

The poly(A) tail is located downstream of the mRNA 3' untranslated region and contributes to mRNA stabilization. It also facilitates the transport of mRNA into the cytoplasm. The efficiency of mRNA translation is regulated by poly(A) tail length. Analyzing poly(A) tail length helps investigate the dynamic regulation process of mRNA degradation and translation. Therefore, it is important to assess poly(A) tail length. Poly(A) tail length was calculated using Nanopolish software (version 0.12.5; <https://github.com/jts/nanopolish>, accessed on 10 May 2025). The parameter used was “poly(A)”. Additionally, the median poly(A) tail length of each transcript type was correlated with its expression levels to analyze the relationship.

2.6.3. LncRNA Analysis

LncRNAs are RNA molecules longer than 200 nucleotides that do not encode proteins. In this study, CPC2 (version 0.1; [56]), CNCI (version 2.0; <http://www.iccnci.org/>; [57]), and the Pfam-A (version 33.1; [58]) database were used to screen for lncRNAs. All aforementioned software were run with default parameters. CPC2 evaluated transcript coding potential using the CVM model, trained on four intrinsic features: ORF length, ORF integrity, isoelectric point, Fichett TESTCODE score [56]. By analyzing the spectrum of ANT, CNCI can effectively distinguish encoding and non-encoding sequences independent of known annotations [57]. The Pfam-A database contains high-quality structural domains of most known proteins.

2.7. Analysis of RNA Methylation

Direct RNA sequencing can detect base modifications while reading RNA base sequences. In this study, Tombo software (version 1.5; <https://github.com/nanoporetech/tombo>, accessed on 10 May 2025) was used to identify m5C methylation sites using “alternative_model” mode. The “de novo” mode of Tombo was used to identify m6A methylation sites. The “MINES” process (<https://github.com/YeoLab/MINES>, accessed on 10 May 2025) was added upstream for methylation site calculation. The m6A and m5C methylation sites were mapped onto the reference *C. mydas* genome. The corresponding site information was then used to plot the methylation site distribution on the reference genome. Enzymes related to RNA methylation recognize and bind to specific motifs to carry out methylation, thereby affecting gene expression. Five-base motifs were obtained by expanding two bases upstream and downstream from the m5C/m6A methylation sites. Motif sequence characteristics were obtained by using MEME software (version 5.5.3; <http://meme-suite.org/index.html>, accessed on 10 May 2025).

3. Results

3.1. *C. mydas* Long-Read Blood Transcriptome Sequencing

Nanopore PromethION platform was used to perform the direct RNA sequencing of *C. mydas*. The platform totally generated 8,541,562 raw reads, corresponding with 4,209,632,522 bp. The mean length, the N50, and the maximum length of all raw reads was respectively 492.8 bp, 612 bp, and 262,309 bp. 7,200,345 clean reads with the size of 3,952,961,215 bp were obtained by the platform after filtering. The mean length, the N50, and the maximum length of all clean reads was respectively 549 bp, 617 bp and 10,873 bp. By clustering the clean reads, 25,598 consensus sequences were obtained. The mean length, the N50 and the maximum length was respectively 1052 bp, 1941 bp and 10,856 bp. The redundancy of consensus sequences was removed after comparing with *C. mydas* reference genome.

3.2. Comparison with Reference Genome of *C. mydas*

After comparison with the reference genome, a total of 5,706,030 reads were mapped, with a mapping rate of 79.25%. Additionally, 4061 novel transcripts were identified by comparison with known transcripts in the reference genome. Among them, the numbers of “O”, “J”, “I”, “X”, and “U” types were 17, 734, 1364, 304, and 1642, respectively. Figure 1 shows the quantitative distribution of the five novel transcript types, while Figure 2 shows the distribution of known and novel transcripts across the chromosomes. It is worth noting that these distributions are derived from a single individual and cannot be generalized to the broader *C. mydas* population.

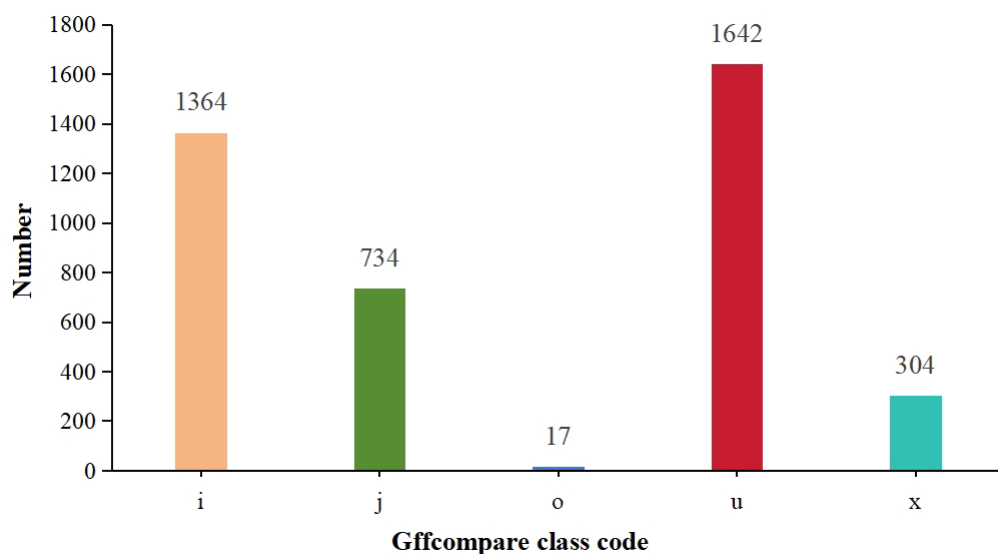


Figure 1. Quantitative distribution of five types of novel transcripts identified in the blood of a single *C. mydas* individual. Note: The horizontal coordinate indicates the type of novel transcript, and the vertical coordinate indicates the number of each type of novel transcript.

3.3. CDS Information

A total of 2402 CDSs were predicted from the novel transcripts. The predicted N50 of CDSs was 835, and their length distribution was plotted. To improve the genome annotation of *C. mydas*, CDSs needed to be annotated. The CDSs predicted from the novel transcripts were annotated using the aforementioned databases. Of the 2402 CDSs, 1208 had annotation information in at least one database. The quantitative CDS distribution in the Nr database reflected the sequence similarity between *C. mydas* and other species. The results showed that 1163 CDSs were annotated in the Nr database. Among them, 69.82% of the

annotated CDSs closely resembled *C. mydas* sequences, while 10.99% matched *Dermochelys coriacea* (Figure 3A). The relatively low match rate to the existing *C. mydas* genome likely reflects incomplete reference genome annotation rather than species divergence, supporting the value of long-read data for improving gene models. The top three most frequently annotated in the Pfam database were “DDE-TNP-4”, “KH-2”, and “RRM-1” (Figure 3B). GO classification provides a comprehensive description of gene and gene product properties within an organism. The results indicated that 790 CDSs had annotation information in the GO database. “Translation”, “Integral Component of Membrane”, and “Metal Ion Binding” were dominant in “cell composition”, “molecular function”, and “biological process”, respectively (Figure 3C). CDSs annotated in the COG database were classified into orthologous groups. The results showed that the CDSs were divided into 26 groups, with the three largest being “J”, “O”, and “R”. They represented “Translation, Ribosomal Structure, and Biogenesis”, “Posttranslational Modification, Protein Turnover, Chaperones”, and “General Function Prediction Only”, respectively (Figure 3D). Annotation in KEGG database allows analysis of gene product metabolic pathways. The results showed that “Signal transduction”, “Translation”, “Global and Overview map”, “Immune System”, and “Cell growth and death” were respectively dominant in “Environment Information Processing”, “Genetic Information Processing”, “Metabolism”, “Organismal Systems”, and “Cellular Processes” (Figure 3E). Figure 4 presents the Upset Plot showing CDSs annotated across six databases, with 105 CDSs annotated in all databases.

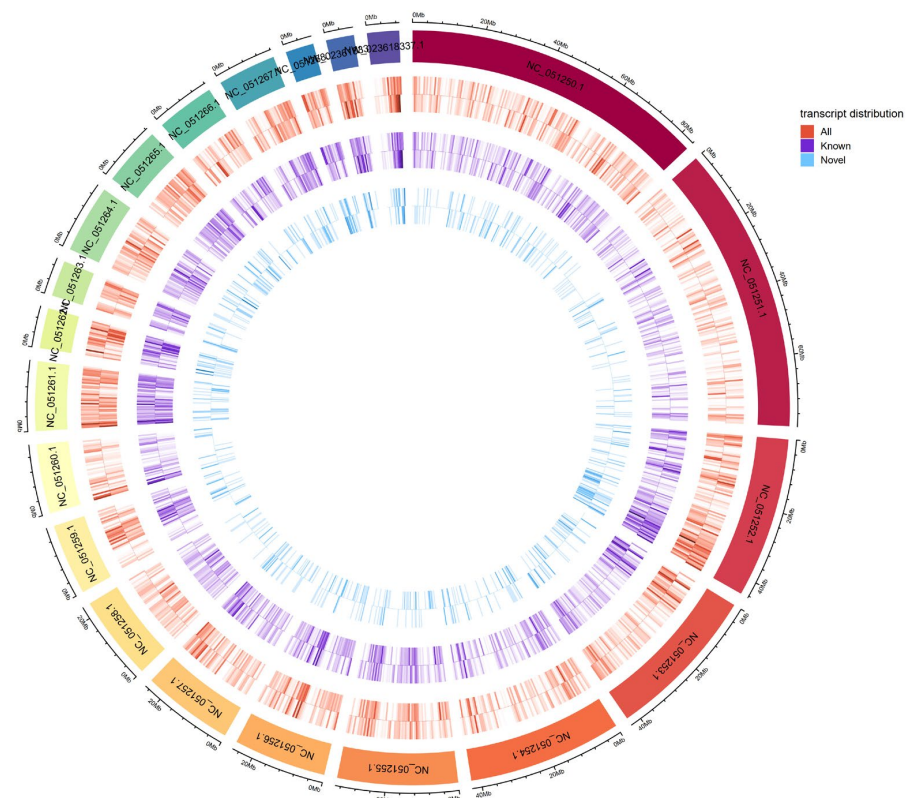


Figure 2. Chromosomal distribution of total, known, and novel transcripts in *C. mydas*. Note: The outermost ring shows the ID and length of chromosome. The other three rings represent the distribution of all transcripts, known transcripts, and novel transcripts on the chromosomes from outside to inside, respectively. Meanwhile, each of these three rings is divided into two parts, with the outer representing the distribution of positive chains and the inner representing the distribution of negative chains.

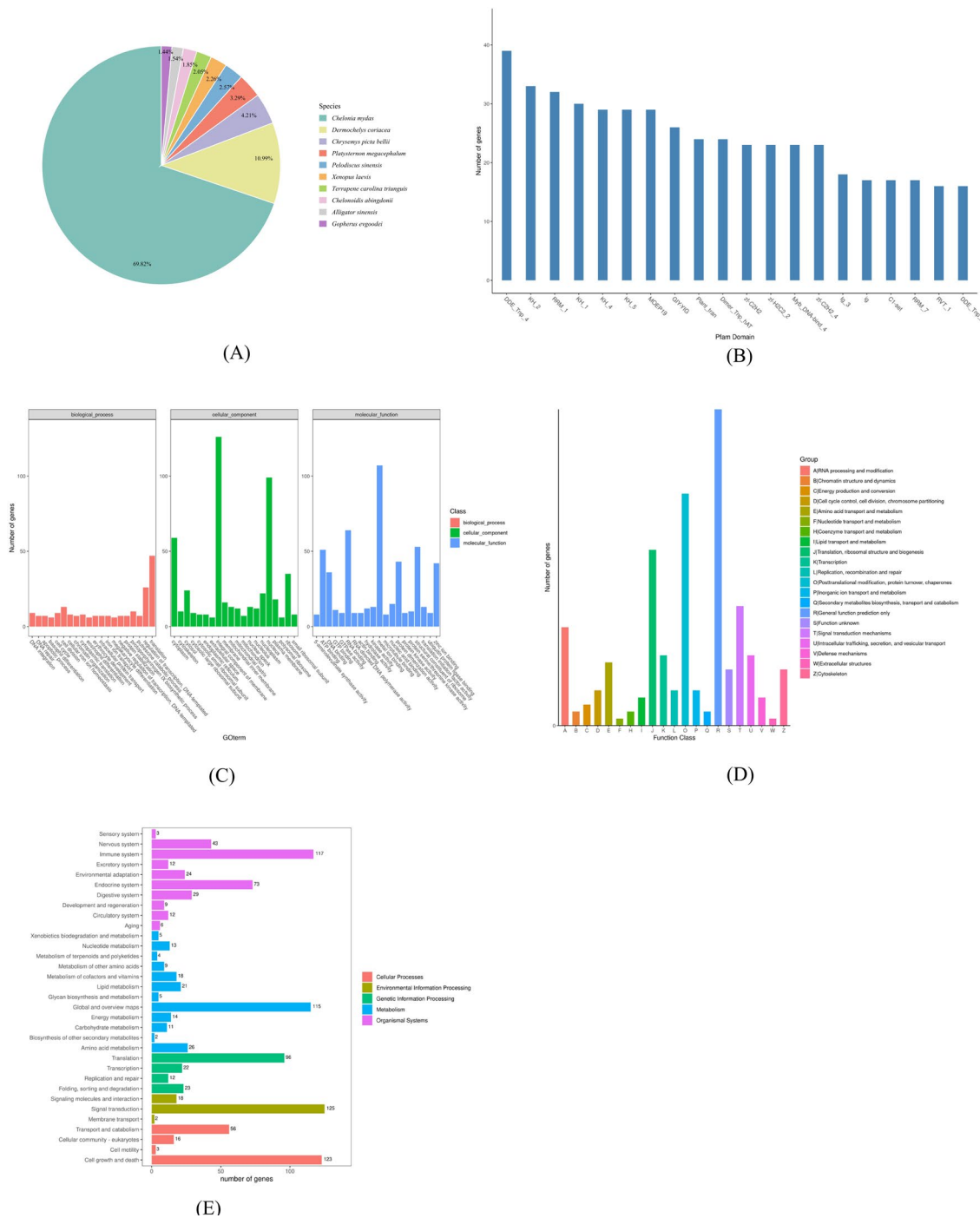


Figure 3. The annotation information of CDSs in Nr, Pfam, GO, COG, and KEGG database. Note: (A) The annotation information of CDSs in Nr. (B) The annotation information of CDSs in Pfam. (C) The annotation information of CDSs in GO. (D) The annotation information of CDSs in COG. (E) The annotation information of CDSs in KEGG.

3.4. Long-Read Direct Blood Transcriptome Structure Analysis

3.4.1. AS Events and Fusion Transcripts

A total of 2294 AS events were identified. The AS event types included: (I) Alternative 3' splice site (A3), (II) Alternative 5' splice site (A5), (III) Alternative First exon (AF), (IV) alternative last exon (AL), (V) Mutually exclusive exon (MX), (VII) Retained intron (RI), and (VII) Skipping exon (SE). The quantitative distribution of each AS event type is shown in Figure 5. Among all AS event types, alternative first exon (AF) was the most common

(26.6%), while mutually exclusive exon was the least common (1.0%). Additionally, analysis of the fusion transcripts revealed a total of 376 fusion transcripts.

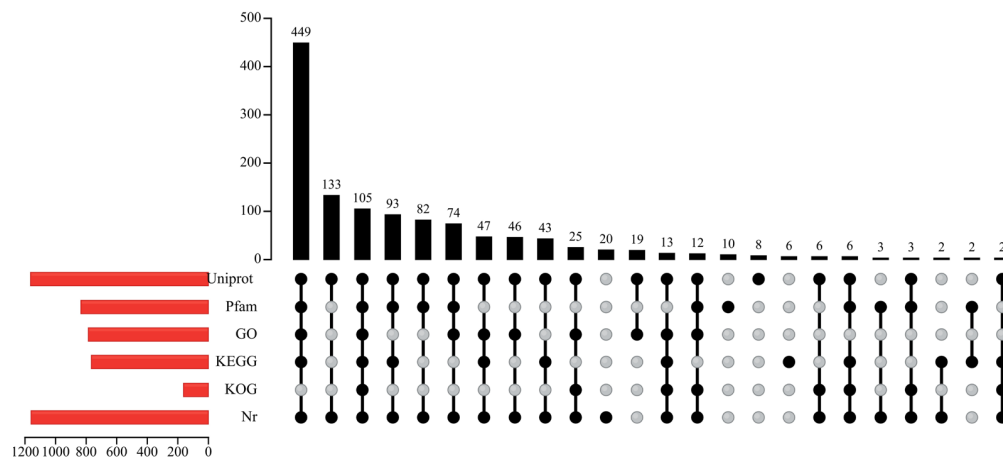


Figure 4. The UpSet plot of annotated genes in each database. Note: The bar chart on the left counts the number of CDS annotated in each database. The database of CDS annotation is represented by the points in the lower part, and different occupation and connection of points are categorized. The number of each category is counted by the bar chart in the upper part.

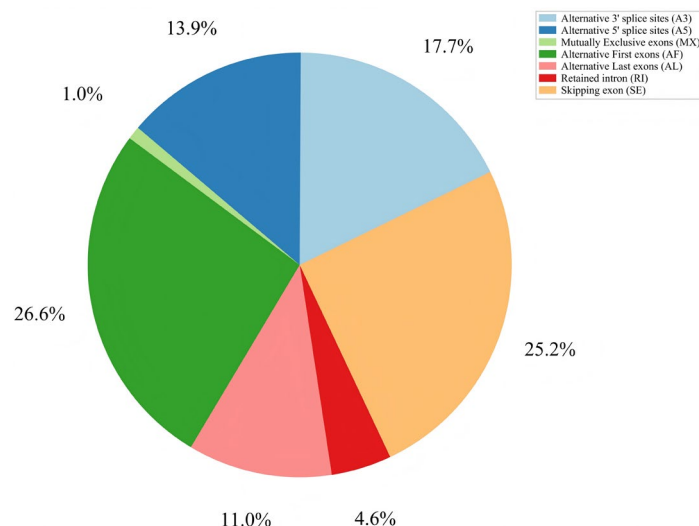


Figure 5. The types and the percentage of AS events in the long-read direct blood transcriptome of *C. mydas*.

3.4.2. Poly(A)s

Statistical analysis showed that the mean, Q25, and Q50 lengths of poly(A)s were 67.33, 43.79, and 58.32 bp, respectively. The length distribution of poly(A)s is shown in Figure 6A. In addition, we analyzed the correlation between the median poly(A)s length of each transcript type and transcript expression levels. The results showed a negative correlation between poly(A) length and transcript expression (Figure 6B). It is worth noting that a negative trend between median poly(A) length and transcript expression was observed, but this result is based on a single individual and lacks statistical replication support; the correlation and *p*-value are for descriptive illustration only.

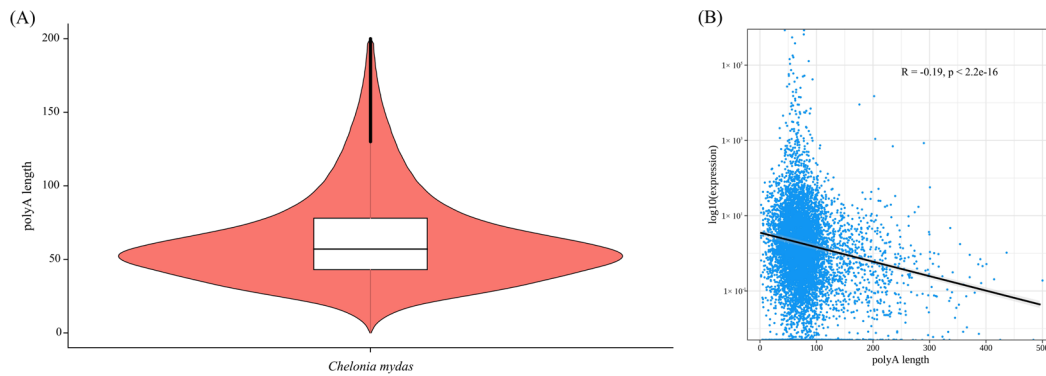


Figure 6. The analysis diagram of poly(A) length. Note: (A) The violin plot shows the distribution of the length of poly(A). (B) The chart shows the correlation between the length of poly(A) and the expression level of long-read direct transcripts. The horizontal coordinate represents the length of poly(A), and the vertical coordinate represents the expression of long-read direct transcripts. Trend line is for descriptive illustration only; no statistical significance is claimed due to the lack of biological replicates.

3.4.3. LncRNAs

Three different methods were used to screen for lncRNAs, identifying 3289, 3283, and 3153 lncRNAs, respectively. In total, 3507 lncRNAs were predicted, with 2982 lncRNAs commonly predicted by all three methods. The Venn diagram is shown in Figure 7. The identified lncRNAs exhibited canonical structural characteristics, with a markedly shorter median transcript length and fewer exons compared with mRNAs.

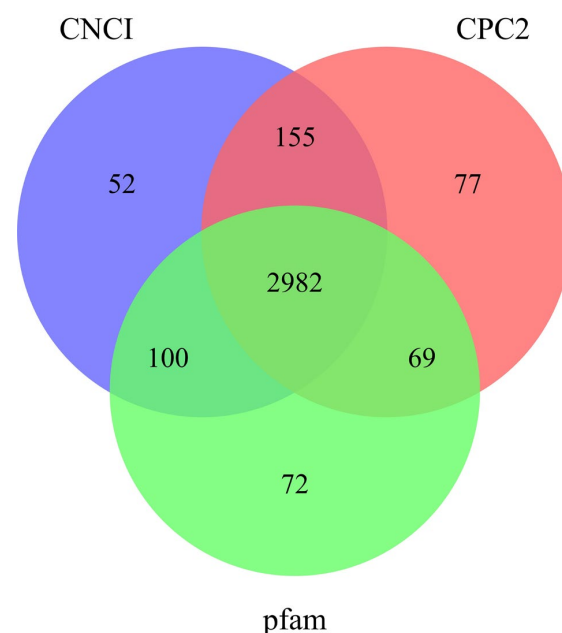


Figure 7. The Venn diagram of predicted lncRNAs by the Pfam database, CPC2 software and CNCI software. Note: The intersection of the three circles represents the number of lncRNAs predicted by the three tools together. The part where the two circles meet together represents the number of lncRNAs predicted by both tools. The part that has no meeting place represents the number of lncRNAs predicted by each tool alone.

3.5. RNA Methylation

m6A and m5C methylation sites were mapped to the reference genome. Corresponding site information was extracted to plot the distribution of methylation sites on the genome. Figure 8 shows the distribution of m5C (Figure 8A) and m6A (Figure 8B) methyla-

tion sites on the reference genome. Additionally, motifs of m5C and m6A methylation sites were identified. The results indicated that m5C motifs were irregular, whereas m6A motifs followed the pattern R (A or G) GACH (A, T, or C). The motif characteristics of m6A sites are shown in Figure 9.

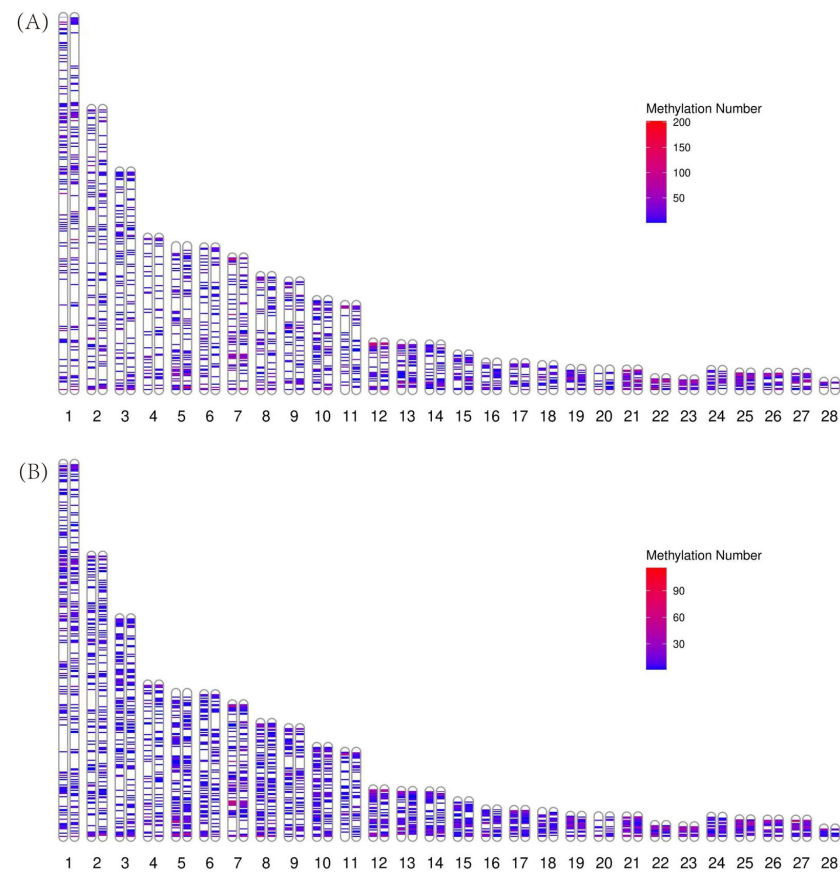


Figure 8. The distribution of methylation sites. Note: (A) The distribution of m5C methylation sites. The colors from blue to red represent methylation number from low to high. (B) The distribution of m6A methylation sites. The colors from blue to red represent methylation number from low to high.

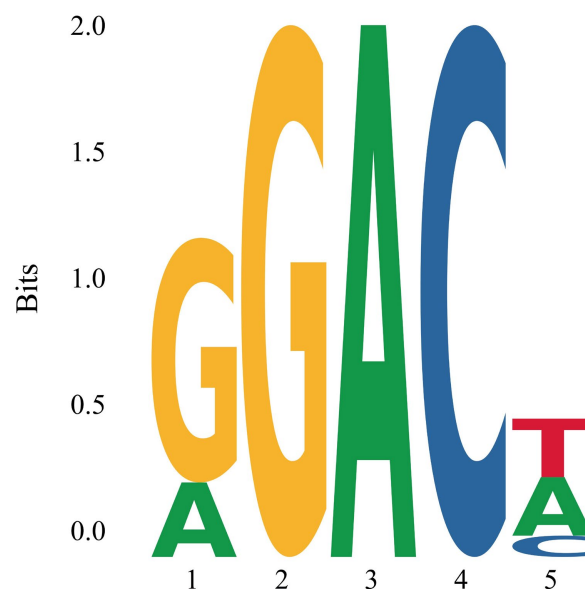


Figure 9. The motif of the m6A methylation sites. Note: The height represents the relative fundamental frequency of each site.

4. Discussion

The primary premise of studying gene function is to detect sequences efficiently and accurately. Although current short-read sequencing is relatively mature, its disadvantages are also quite apparent. Due to the short read length, short-read sequencing has difficulty identifying transcript isoforms and structural variants [59]. These issues affect the integrity and accuracy of the genomic annotation function information. As third-generation sequencing technology advances, it helps mitigate the shortcomings of NGS technology, enabling the acquisition of more complete long-read direct transcripts during the sequencing process [60]. In this study, the ONT sequencing platform was used to sequence the long-read direct transcriptome of *C. mydas*. Data analysis after sequencing showed that the longest length of the consensus sequences reached 10,856 bp, far beyond the read length range of NGS [61]. We identified 4061 novel transcripts by comparing them to *C. mydas* genome. Meanwhile, 2402 potential coding sequences were identified in the above novel transcripts. Among them, 1208 had annotation information in existing databases. This can complement and improve the genome annotation of *C. mydas*. However, we do not deny that errors associated with ONT technology and limited tissue sample collection may lead to an inflated number of predicted ORFs. Therefore, functional validation is required before designating these predicted genes as novel genes.

Pre-mRNA undergoes alternative splicing, leading to the generation of multiple transcript isoforms and enhancing transcriptome complexity. Therefore, a single gene can generate multiple proteins with distinct functionalities, giving rise to a wide array of proteins and traits [62,63]. Meanwhile, AS events have been reported to be involved in various biological processes, including immune activation, apoptosis, and germ cell production [64–66]. However, due to the limited read length generated, short-read sequencing struggles to fully reconstruct full-length transcripts and identify transcription isoforms, thereby impacting the comprehensive identification of AS events [30,31]. While long-read sequencing improves transcript isoform detection [34], the 2294 AS events identified in this study enriched the reference annotation, and this number of AS events is similar to that in other vertebrate transcriptomes. However, since they come from a single specimen, these events may not represent the complete AS repertoire of the species. AS events are involved in the responses and tolerance of organisms to changes in the external environment [67–69], making them useful in understanding the physiological changes in *C. mydas* under various threats. In addition, the differences in AS events were hypothesized to be an important factor affecting longevity [70]. Therefore, we speculate that the detected AS events in the present may provide insights for uncovering the longevity mechanisms of *C. mydas*. However, further functional experiments will still need to be conducted in the future to verify our hypothesis.

Fusion transcripts are produced by fusion DNA or trans-splicing events [71,72]. The formation of fusion transcripts may serve as a mechanism for augmenting the functional genome and enhancing genomic information content [73,74]. In this study, only 376 fusion transcripts were found. The quantity is substantially lower compared to that of humans, mice, and *Drosophila* [75]. This may be due to the limited accuracy of third-generation sequencing technology [29]. As an important mechanism of generating novel genes, gene fusion plays a critical role in the genetic and phenotypic evolution of organisms [76]. Therefore, the fusion transcripts produced by fusion genes may provide valuable insights for future research on the evolution of *C. mydas*. Meanwhile, the fusion transcripts generated by trans-splicing events expand the functional information database of the genome annotation of *C. mydas*.

The poly(A) tail is important for mRNA transport, stability, and translation [77]. Multiple polyadenylation sites can be present in one gene, thus producing different transcript

isoforms [78]. Long-read sequencing can identify poly(A) sites and the corresponding transcript isoforms [35]. Meanwhile, direct RNA sequencing eliminates amplification-induced bias and generates long reads, enabling the accurate measurement of long poly(A) [79]. The length of the poly(A) tail changes throughout the mRNA lifetime [79]; therefore, the median length of poly(A)s of each type of transcript was calculated for the correlation analysis. In this study, the median length of poly(A)s for each transcript type exhibited a negative correlation with transcript expression levels, aligning with the previously reported conclusion that short poly(A)s is generally a characteristic of highly expressed genes [80]. Therefore, the length distribution of poly(A)s may reflect the expression levels of various genes in *C. mydas* to a certain extent. The mean length of poly(A)s obtained was 67.33 bp. Previous studies have shown that organisms select different polyadenylation sites on genes under abiotic stress to produce different transcript isoforms [81]. These results provide a baseline description of poly(A) tail lengths in blood transcripts, which could be further explored in future studies investigating responses to environmental stress. However, what we do not deny is that the negative trend observed here should be treated as preliminary and descriptive only, statistical significance and biological validation require additional biological replicates.

LncRNAs refer to RNAs longer than 200 nucleotides that do not encode proteins [82]. LncRNAs play an important role in various essential biological processes, including transcriptional regulation, epigenetic regulation, cellular transport, organ or tissue development, and metabolic processes [83–85]. Three methods were used to predict lncRNAs, and the results were encouraging. On the one hand, the number of lncRNAs predicted by the three methods, both individually and jointly, exceeded 3000. On the other hand, the number of lncRNAs predicted by each method accounted for more than 90% of the total predicted quantity. This indicates that the predicted lncRNAs have high credibility. These predicted lncRNAs can complement and improve the genomic information of *C. mydas*. Meanwhile, although lncRNAs are generally moderately conserved [86,87], some lncRNAs are strongly conserved [88,89]. This may be valuable for studying the evolutionary history of *C. mydas*.

RNA methylation involves the specific addition of methyl groups to RNA methyladenine, facilitated by the catalytic activity of methyltransferase. RNA methylation is crucial to various biological processes, including RNA metabolism [48,49]. In this study, we plotted the distribution of m5C and m6A methylation sites on the genome annotation of *C. mydas*, enriching the genomic information to a certain extent. In recent years, FP benign tumor has become a focus of biological research on *C. mydas*. Although the pathogenesis of FP tumors is not fully understood, a herpes virus is considered the likely etiological agent [90]. Previous research suggested that RNA m6A methylation participates in viral infection regulation [91]. However, some studies indicated that both m6A and m5C RNA methylations are implicated in numerous cancers [92]. Based on these findings, we propose that future studies should investigate the potential involvement of RNA methylation in the regulatory mechanisms underlying FP tumors in *C. mydas*. This hypothesis, though currently speculative, offers a promising avenue for further research. Meanwhile, we identified a highly conserved m6A motif (RGACH), which has also been reported in vertebrate (such as zebrafish), and the RGACH was speculated to be involved in regulating key physiological processes such as heart development and the cell cycle progression of hematopoietic stem and progenitor cells [93,94]. Given the current limitations of per-read modification detection from ONT data and the absence of biological replicates, the methylation calls reported here should be viewed as candidate sites requiring further validation. Meanwhile, while we did not examine infected or tumor tissues, the catalog of candidate RNA methylation

sites established here may serve as a starting point for future work exploring the role of m6A and m5C in viral infection and tumor biology in *C. mydas*.

However, several limitations of this study should be acknowledged. First, the study uses a single biological sample without replicates, which restricts statistical inference and population-level generalization. Second, only blood tissue was analyzed, which cannot represent transcriptomic profiles of other organs. Third, all structural features would benefit from additional experimental validation.

5. Conclusions

In this study, we generated a long-read direct transcriptome resource for *C. mydas* using ONT. The data reveal some novel transcripts and structural features that complement the current genome annotation. However, because these findings are based on a single sample, they should be considered exploratory. Studies with biological replications and experimental validation are needed to confirm the biological relevance of these results.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/fishes11050269/s1>, the supplementary materials file contains the analysis process of the long-read direct transcriptome data, including expression, enrich, methylation, poly(A), and structure analyses.

Author Contributions: Q.H., W.Z., F.S., J.X. and Y.Q.: formal analysis, methodology, conceptualization, visualization, writing—review and editing, writing—original draft preparation, and data curation. L.Z.: software, methodology, project administration, supervision and validation. Y.S.: resources. The published version of the manuscript has been reviewed and approved by all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved by the Yantai University Laboratory Animal Ethics Committee (protocol code 2025/01/22 and date of approval 22 January 2025).

Data Availability Statement: All the data analysis results are presented in the Supplementary Materials file.

Conflicts of Interest: Author Yongjun Sun was employed by the company Shandong Homey Aquatic Development Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Hirth, H.F. *Synopsis of the Biological Data on the Green Turtle, Chelonia mydas (Linnaeus 1758)*; US Department of the Interior: Washington, DC, USA, 1997. Available online: <https://digitalmedia.fws.gov/digital/collection/document/id/1776> (accessed on 10 May 2025).
- Avise, J.C.; Bowen, B.W. Investigating sea turtle migration using DNA markers. *Curr. Opin. Genet. Dev.* **1994**, *4*, 882–886. [[CrossRef](#)] [[PubMed](#)]
- Jensen, M.P.; FitzSimmons, N.N.; Dutton, P.H.; Michael, P. Molecular genetics of sea turtles. In *The Biology of Sea Turtles*; Wyneken, J., Lohmann, K.J., Musick, J.A., Eds.; CRC Press: Boca Raton, FL, USA, 2013; Volume III, pp. 135–161. [[CrossRef](#)]
- George, R.H. Health problems and diseases of sea turtles. In *The Biology of Sea Turtles*; Lutz, P.L., Musick, J.A., Eds.; CRC Press: Boca Raton, FL, USA, 2017; pp. 363–385.
- Lal, A.; Arthur, R.; Marbà, N.; Lill, A.W.; Alcoverro, T. Implications of conserving an ecosystem modifier: Increasing green turtle (*Chelonia mydas*) densities substantially alters seagrass meadows. *Biol. Conserv.* **2010**, *143*, 2730–2738. [[CrossRef](#)]
- Tol, S.J.; Jarvis, J.C.; York, P.H.; Grech, A.; Congdon, B.C.; Coles, R.G. Long distance biotic dispersal of tropical seagrass seeds by marine mega-herbivores. *Sci. Rep.* **2017**, *7*, 4458. [[CrossRef](#)] [[PubMed](#)]
- Wabnitz, C.C.; Balazs, G.H.; Beavers, S.C.; Bjorndal, K.A.; Bolten, A.B.; Christensen, V.; Hargrove, S.K.; Pauly, D. Ecosystem structure and processes at Kaloko Honoko-hau, focusing on the role of herbivores, including the green sea turtle *Chelonia mydas*, in reef resilience. *Mar. Ecol. Prog. Ser.* **2010**, *420*, 27–44. [[CrossRef](#)]

8. Hawkes, L.; Broderick, A.; Godfrey, M.; Godley, B. Climate change and marine turtles. *Endanger. Species Res.* **2009**, *7*, 137–154. [[CrossRef](#)]
9. Schuyler, Q.; Harbesty, B.D.; Wilcox, C.; Townsend, K. Global analysis of anthropogenic debris ingestion by sea turtles. *Conserv. Biol.* **2013**, *28*, 129–139. [[CrossRef](#)]
10. Lewison, R.L.; Crowder, L.B.; Wallace, B.P.; Moore, J.E.; Cox, T.; Zydalis, R.; McDonald, S.; DiMatteo, A.; Dunn, D.C.; Kot, C.Y.; et al. Global patterns of marine mammal, seabird, and sea turtle bycatch reveal taxa-specific and cumulative megafauna hotspots. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5271–5276. [[CrossRef](#)]
11. Seminoff, J.A. *Chelonia mydas*. In *The IUCN Red List of Threatened Species*; International Union for Conservation of Nature: Gland, Switzerland, 2004; p. e.T4615A11037468. [[CrossRef](#)]
12. Li, M.; Zhang, T.; Liu, Y.; Li, Y.; Fong, J.J.; Yu, Y.; Wang, J.; Shi, H.; Lin, L. Revisiting the genetic diversity and population structure of the endangered Green Sea Turtle (*Chelonia mydas*) breeding populations in the Xisha (Paracel) Islands, South China Sea. *PeerJ* **2023**, *11*, e15115. [[CrossRef](#)]
13. Bowen, B.W.; Karl, S.A. Population genetics and phylogeography of sea turtles. *Mol. Ecol.* **2007**, *16*, 4886–4907. [[CrossRef](#)]
14. Hamann, M.; Godfrey, M.H.; Seminoff, J.A.; Arthur, K.; Barata, P.C.R.; Bjørndal, K.A.; Bolten, A.B.; Broderick, A.C.; Campbell, L.M.; Carreras, C.; et al. Global research priorities for sea turtles: Informing management and conservation in the 21st century. *Endanger. Species Res.* **2010**, *11*, 245–269. [[CrossRef](#)]
15. Behler, J.L.; King, F.W. *National Audubon Society Field Guide to North American Reptiles and Amphibians*; Chanticleer Press: New York, NY, USA, 1979.
16. Conant, R.; Collins, J.T. *A Field Guide to Reptiles & Amphibians: Eastern and Central North America*; Houghton Mifflin: Boston, MA, USA, 1998.
17. Tacutu, R.; Thornton, D.; Johnson, E.; Budovsky, A.; Barardo, D.; Craig, T.; Diana, E.; Lehmann, G.; Toren, D.; Wang, J.; et al. Human Ageing Genomic Resources: New and updated databases. *Nucleic Acids Res.* **2018**, *46*, D1083–D1090. [[CrossRef](#)]
18. Reinke, B.A.; Cayuela, H.; Janzen, F.J.; Lemaître, J.; Gaillard, J.; Lawing, A.M.; Iverson, J.B.; Christiansen, D.G.; Martínez-Solano, I.; Sánchez-Montes, G.; et al. Diverse aging rates in ectothermic tetrapods provide insights for the evolution of aging and longevity. *Science* **2022**, *376*, 1459–1466. [[CrossRef](#)]
19. Thomson, R.C.; Spinks, P.Q.; Shaffer, H.B. A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012215118. [[CrossRef](#)] [[PubMed](#)]
20. Wang, Z.; Pascual-Anaya, J.; Zadissa, A.; Li, W.Q.; Niimura, Y.; Huang, Z.Y.; Li, C.; White, S.; Xiong, Z.; Fang, D.; et al. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat. Genet.* **2013**, *45*, 701–706. [[CrossRef](#)] [[PubMed](#)]
21. Li, C.; Wu, X.C.; Rieppel, O.; Wang, L.T.; Zhao, L.J. An ancestral turtle from the Late Triassic of southwestern China. *Nature* **2008**, *456*, 497–501. [[CrossRef](#)]
22. Denton, J.F.; Lugo-Martinez, J.; Tucker, A.E.; Schrider, D.R.; Warren, W.C.; Hahn, M.W. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput. Biol.* **2014**, *10*, e1003998. [[CrossRef](#)]
23. Denoeud, F.; Aury, J.M.; Da Silva, C.; Noel, B.; Rogier, O.; Delledonne, M.; Morgante, M.; Valle, G.; Wincker, P.; Scarpelli, C.; et al. Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* **2008**, *9*, R175. [[CrossRef](#)]
24. Li, Z.; Zhang, Z.; Yan, P.; Huang, S.; Fei, Z.; Lin, K. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genom.* **2011**, *12*, 540. [[CrossRef](#)]
25. Elsik, C.G.; Worley, K.C.; Bennett, A.K.; Beye, M.; Camara, F.; Childers, C.P.; de Graaf, D.C.; Debyser, G.; Deng, J.; Devreese, B.; et al. Finding the missing honey bee genes: Lessons learned from a genome upgrade. *BMC Genom.* **2014**, *15*, 86. [[CrossRef](#)]
26. Blackburn, N.B.; Leandro, A.C.; Nahvi, N.; Devlin, M.A.; Leandro, M.; Martinez Escobedo, I.; Peralta, J.M.; George, J.; Stacy, B.A.; deMaar, T.W.; et al. Transcriptomic profiling of Fibropapillomatosis in green sea turtles (*Chelonia mydas*) from South Texas. *Front. Immunol.* **2021**, *12*, 630988. [[CrossRef](#)] [[PubMed](#)]
27. Kane, R.A.; Christodoulides, N.; Jensen, I.M.; Becker, D.J.; Mansfield, K.L.; Savage, A.E. Gene expression changes with tumor disease and leech parasitism in the juvenile green sea turtle skin transcriptome. *Gene* **2021**, *800*, 145800. [[CrossRef](#)] [[PubMed](#)]
28. Heather, J.M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107*, 1–8. [[CrossRef](#)]
29. Athanasopoulou, K.; Boti, M.A.; Adamopoulos, P.G.; Skourou, P.C.; Scorilas, A. Third-generation sequencing: The spearhead towards the radical transformation of modern genomics. *Life* **2021**, *12*, 30. [[CrossRef](#)] [[PubMed](#)]
30. Engström, P.G.; Steijger, T.; Sipos, B.; Grant, G.R.; Kahles, A.; The RGASP Consortium; Rättsch, G.; Goldman, N.; Hubbard, T.J.; Harrow, J.; et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **2013**, *10*, 1185–1191. [[CrossRef](#)]
31. Steijger, T.; Abril, J.F.; Engström, P.G.; Kokocinski, F.; Consortium, R.; Hubbard, T.J.; Guigó, R.; Harrow, J.; Bertone, P. Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* **2013**, *10*, 1177–1184. [[CrossRef](#)]
32. Leshkowitz, D.; Feldmesser, E.; Friedlander, G.; Jona, G.; Ainbinder, E.; Parmet, Y.; Horn-Saban, S. Using synthetic mouse spike-in transcripts to evaluate RNA-Seq analysis tools. *PLoS ONE* **2016**, *11*, e0153782. [[CrossRef](#)]

33. Wang, B.; Tseng, E.; Regulski, M.; Clark, T.A.; Hon, T.; Jiao, Y.; Lu, Z.; Olson, A.; Stein, J.C.; Ware, D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **2016**, *7*, 1170. [[CrossRef](#)]
34. Amarasinghe, S.L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M.E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **2020**, *21*, 30. [[CrossRef](#)]
35. Byrne, A.; Cole, C.; Volden, R.; Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, 20190097. [[CrossRef](#)]
36. Nudelman, G.; Frasca, A.; Kent, B.; Sadler, K.C.; Sealfon, S.C.; Walsh, M.J.; Zaslavsky, E. High resolution annotation of zebrafish transcriptome using long-read sequencing. *Genome Res.* **2018**, *28*, 1415–1425. [[CrossRef](#)]
37. Lou, F.R.; Wang, L.; Wang, Z.Y.; Wang, L.; Zhao, L.L.; Zhou, Q.J.; Lu, Z.; Tang, Y. Full-Length transcriptome of the Whale Shark (*Rhincodon typus*) facilitates the genome information. *Front. Mar. Sci.* **2022**, *8*, 821253. [[CrossRef](#)]
38. Xing, L.S.; Wu, Q.; Xi, Y.; Huang, C.; Liu, W.X.; Wan, F.H.; Qian, W. Full-length coding moth transcriptome atlas revealed by single-molecule real-time sequencing. *Genomics* **2022**, *114*, 110299. [[CrossRef](#)] [[PubMed](#)]
39. Hoang, N.V.; Furtado, A.; Mason, P.J.; Marquardt, A.; Kasirajan, L.; Thirugnanasambandam, P.P.; Botha, F.C.; Henry, R.J. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genom.* **2017**, *18*, 395. [[CrossRef](#)]
40. Zhou, T.; Chen, G.; Chen, M.; Wang, Y.; Zou, G.; Liang, H. Direct Full-Length RNA sequencing reveals an important role of epigenetics during sexual reversal in Chinese soft-shelled turtle. *Front. Cell Dev. Biol.* **2022**, *10*, 876045. [[CrossRef](#)] [[PubMed](#)]
41. Pardo-Palacios, F.J.; Wang, D.; Reese, F.; Diekhans, M.; Carbonell-Sala, S.; Williams, B.; Loveland, J.E.; De María, M.; Adams, M.S.; Balderrama-Gutierrez, G.; et al. Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat. Methods* **2024**, *21*, 1349–1363. [[CrossRef](#)]
42. Chen, Y.; Davidson, N.M.; Wan, Y.K.; Yao, F.; Su, Y.; Gamaarachchi, H.; Sim, A.; Patel, H.; Low, H.M.; Hendra, C.; et al. A systematic benchmark of Nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. *Nat. Methods* **2025**, *22*, 801–812. [[CrossRef](#)]
43. Payne, A.; Holmes, N.; Rakyar, V.; Loose, M. BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* **2019**, *35*, 2193–2198. [[CrossRef](#)]
44. Garalde, D.R.; Snell, E.A.; Jachimowicz, D.; Sipos, B.; Lloyd, J.H.; Bruce, M.; Pantic, N.; Admassu, T.; James, P.; Warland, A.; et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **2018**, *15*, 201–206. [[CrossRef](#)]
45. Jenjaroenpun, P.; Wongsurawat, T.; Pereira, R.; Patumcharoenpol, P.; Ussery, D.W.; Nielsen, J.; Nookaew, I. Complete genomic and transcriptional landscape analysis using third-generation sequencing: A case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Res.* **2018**, *46*, e38. [[CrossRef](#)] [[PubMed](#)]
46. Workman, R.E.; Tang, A.D.; Tang, P.S.; Jain, M.; Tyson, J.R.; Razaghi, R.; Zuzarte, P.C.; Gilpatrick, T.; Payne, A.; Quick, J.; et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **2019**, *16*, 1297–1305. [[CrossRef](#)]
47. Eckmann, C.R.; Rammelt, C.; Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 348–361. [[CrossRef](#)]
48. Zhao, B.S.; Roundtree, I.A.; He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 31–42. [[CrossRef](#)] [[PubMed](#)]
49. Xue, C.; Zhao, Y.; Li, L. Advances in RNA cytosine-5 methylation: Detection, regulatory mechanisms, biological functions and links to cancer. *Biomark. Res.* **2020**, *8*, 43. [[CrossRef](#)]
50. Wu, T.D.; Watanabe, C.K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **2005**, *21*, 1859–1875. [[CrossRef](#)]
51. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
52. Buchfink, B.; Reuter, K.; Drost, H.G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **2021**, *18*, 366–368. [[CrossRef](#)]
53. Chen, C.J.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.H.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)] [[PubMed](#)]
54. Trincado, J.L.; Entizne, J.C.; Hysenaj, G.; Singh, B.; Skalic, M.; Elliott, D.J.; Eyra, E. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **2018**, *19*, 40. [[CrossRef](#)] [[PubMed](#)]
55. Gordon, S.P.; Tseng, E.; Salamov, A.; Zhang, J.; Meng, X.; Zhao, Z.; Kang, D.; Underwood, J.; Grigoriev, I.V.; Figueroa, M.; et al. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS ONE* **2015**, *10*, e0132628. [[CrossRef](#)]
56. Kang, Y.J.; Yang, D.C.; Kong, L.; Hou, M.; Meng, Y.Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [[CrossRef](#)]
57. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [[CrossRef](#)]
58. Punta, M.; Coghill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2012**, *40*, D290–D301. [[CrossRef](#)]

59. Hu, T.; Chitnis, N.; Monos, D.; Dinh, A. Next-generation sequencing technologies: An overview. *Hum. Immunol.* **2021**, *82*, 801–811. [[CrossRef](#)]
60. Laver, T.; Harrison, J.; O'Neill, P.A.; Moore, K.; Farbos, A.; Paszkiewicz, K.; Studholme, D.J. Assessing the performance of the Oxford Nanopore technologies MinION. *Biomol. Detect. Quantif.* **2015**, *3*, 1–8. [[CrossRef](#)]
61. Weirather, J.L.; Mariateresa, D.C.; Wang, Y.; Paolo, P.; Vittorio, S.; Wang, X.J.; Buck, D.; Au, K.F. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* **2017**, *6*, 100. [[CrossRef](#)] [[PubMed](#)]
62. Nilsen, T.W.; Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457–463. [[CrossRef](#)] [[PubMed](#)]
63. Reddy, A.S.; Marquez, Y.; Kalyna, M.; Barta, A. Complexity of the alternative splicing landscape in plants. *Plant Cell* **2013**, *25*, 3657–3683. [[CrossRef](#)]
64. Boise, L.H.; González-García, M.; Postema, C.E.; Ding, L.; Lindsten, T.; Turka, L.A.; Mao, X.; Nuñez, G.; Thompson, C.B. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **1993**, *74*, 597–608. [[CrossRef](#)] [[PubMed](#)]
65. Lynch, K.W. Consequences of regulated pre-mRNA splicing in the immune system. *Nat. Rev. Immunol.* **2004**, *4*, 931–940. [[CrossRef](#)]
66. Lei, W.L.; Li, Y.Y.; Du, Z.; Su, R.; Meng, T.G.; Ning, Y.; Hou, G.M.; Schatten, H.; Wang, Z.B.; Han, Z.M.; et al. SRSF1-mediated alternative splicing is required for spermatogenesis. *Int. J. Biol. Sci.* **2023**, *19*, 4883–4897. [[CrossRef](#)]
67. Laloum, T.; Martín, G.; Duque, P. Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci.* **2018**, *23*, 140–150. [[CrossRef](#)]
68. Huang, S.; Dou, J.; Li, Z.; Hu, L.; Yu, Y.; Wang, Y. Analysis of Genomic Alternative Splicing Patterns in Rat under Heat Stress Based on RNA-Seq Data. *Genes* **2022**, *13*, 358. [[CrossRef](#)]
69. Zhang, X.; Zhang, X.; Yuan, J.; Li, F. The responses of alternative splicing during heat stress in the Pacific white shrimp *Litopenaeus vannamei*. *Genes* **2023**, *14*, 1473. [[CrossRef](#)]
70. Lee, B.P.; Pilling, L.C.; Emond, F.; Flurkey, K.; Harrison, D.E.; Yuan, R.; Peter, L.L.; Kuchel, G.A.; Ferrucci, L.; Melzer, D.; et al. Changes in the expression of splicing factor transcripts and variations in alternative splicing are associated with lifespan in mice and humans. *Aging Cell* **2016**, *15*, 903–913. [[CrossRef](#)]
71. Li, H.; Wang, J.; Mor, G.; Sklar, J. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* **2008**, *321*, 1357–1361. [[CrossRef](#)]
72. Liu, S.; Tsai, W.H.; Ding, Y.; Chen, R.; Fang, Z.; Huo, Z.; Kim, S.; Ma, T.; Chang, T.Y.; Priedigkeit, N.M.; et al. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res.* **2016**, *44*, e47. [[CrossRef](#)] [[PubMed](#)]
73. Gingeras, T.R. Implications of chimaeric non-co-linear transcripts. *Nature* **2009**, *461*, 206–211. [[CrossRef](#)]
74. Babiceanu, M.; Qin, F.; Xie, Z.; Jia, Y.; Lopez, K.; Janus, N.; Facemire, L.; Kumar, S.; Pang, Y.; Qi, Y.; et al. Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* **2016**, *44*, 2859–2872. [[CrossRef](#)] [[PubMed](#)]
75. Li, X.; Zhao, L.; Jiang, H.; Wang, W. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.* **2009**, *68*, 56–65. [[CrossRef](#)] [[PubMed](#)]
76. Zhou, Y.; Zhang, C.; Zhang, L.; Ye, Q.; Liu, N.; Wang, M.; Long, G.; Fan, W.; Long, M.; Wing, R.A. Gene fusion as an important mechanism to generate new genes in the genus *Oryza*. *Genome Biol.* **2022**, *23*, 130. [[CrossRef](#)]
77. Passmore, L.A.; Collier, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 93–106. [[CrossRef](#)]
78. Tian, B.; Hu, J.; Zhang, H.; Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **2005**, *33*, 201–212. [[CrossRef](#)] [[PubMed](#)]
79. Brouze, A.; Krawczyk, P.S.; Dziembowski, A.; Mroczek, S. Measuring the tail: Methods for poly(A) tail profiling. *Wiley Interdiscip. Rev. RNA* **2023**, *14*, e1737. [[CrossRef](#)]
80. Lima, S.A.; Chipman, L.B.; Nicholson, A.L.; Chen, Y.H.; Yee, B.A.; Yeo, G.W.; Collier, J.; Pasquinelli, A.E. Short poly(A) tails are a conserved feature of highly expressed genes. *Nat. Struct. Mol. Biol.* **2017**, *24*, 1057–1063. [[CrossRef](#)]
81. Yan, C.; Wang, Y.; Lyu, T.; Hu, Z.; Ye, N.; Liu, W.; Li, J.; Yao, X.; Yin, H. Alternative Polyadenylation in response to temperature stress contributes to gene regulation in *Populus trichocarpa*. *BMC Genom.* **2021**, *22*, 53. [[CrossRef](#)]
82. Spizzo, R.; Almeida, M.I.; Colombatti, A.; Calin, G.A. Long non-coding RNAs and cancer: A new frontier of translational research? *Oncogene* **2012**, *31*, 4577–4587. [[CrossRef](#)]
83. Wilusz, J.E.; Sunwoo, H.; Spector, D.L. Long noncoding RNAs: Functional surprises from the RNA world. *Genes Dev.* **2009**, *23*, 1494–1504. [[CrossRef](#)]
84. Qureshi, I.A.; Mattick, J.S.; Mehler, M.F. Long non-coding RNAs in nervous system function and disease. *Brain Res.* **2010**, *1338*, 20–35. [[CrossRef](#)] [[PubMed](#)]

85. Liao, Q.; Liu, C.N.; Yuan, X.Y.; Kang, S.L.; Miao, R.Y.; Xiao, H.; Zhao, G.G.; Luo, H.T.; Bu, D.C.; Zhao, H.T.; et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.* **2011**, *39*, 3864–3878. [[CrossRef](#)] [[PubMed](#)]
86. Marques, A.C.; Ponting, C.P. Catalogues of mammalian long noncoding RNAs: Modest conservation and incompleteness. *Genome Biol.* **2009**, *10*, R124. [[CrossRef](#)] [[PubMed](#)]
87. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [[CrossRef](#)]
88. Chodroff, R.A.; Goodstadt, L.; Sirey, T.M.; Oliver, P.L.; Davies, K.E.; Green, E.D.; Molnár, Z.; Ponting, C.P. Long noncoding RNA genes: Conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **2010**, *11*, R72. [[CrossRef](#)] [[PubMed](#)]
89. Necșulea, A.; Soumillon, M.; Warnefors, M.; Liechti, A.; Daish, T.; Zeller, U.; Baker, J.C.; Grützner, F.; Kaessmann, H. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **2014**, *50*, 635–640. [[CrossRef](#)]
90. Jones, K.; Ariel, E.; Burgess, G.; Read, M. A review of fibropapillomatosis in green turtles (*Chelonia mydas*). *Vet. J.* **2016**, *212*, 48–57. [[CrossRef](#)]
91. Zhou, Y.J.; Kong, Y.; Fan, W.G.; Tao, T.; Xiao, Q.; Li, N.; Zhu, X. Principles of RNA methylation and their implications for biology and medicine. *Biomed. Pharmacother.* **2020**, *131*, 110731. [[CrossRef](#)]
92. Yang, B.C.; Wang, J.Q.; Tan, Y.; Yuan, R.Z.; Chen, Z.S.; Zou, C. RNA methylation and cancer treatment. *Pharmacol. Res.* **2021**, *174*, 105937. [[CrossRef](#)] [[PubMed](#)]
93. Paramasivam, A.; Priyadharsini, J.V. m6A RNA methylation in heart development, regeneration and disease. *Hypertens. Res.* **2021**, *44*, 1236–1237. [[CrossRef](#)] [[PubMed](#)]
94. Han, Y.; Sun, K.; Yu, S.; Qin, Y.; Zhang, Z.; Luo, J.; Hu, H.; Dai, L.; Cui, M.; Jiang, C.; et al. A Mettl16/m6A/mybl2b/Igf2bp1 axis ensures cell cycle progression of embryonic hematopoietic stem and progenitor cells. *EMBO J.* **2024**, *43*, 1990–2014. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.