Computational and Structural Biotechnology Journal 23 (2024) 3186–3198

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Unveiling axolotl transcriptome for tissue regeneration with high-resolution annotation via long-read sequencing

Tian Qin [a,c,d,1], Jie Han [b,c,1], Chunmei Fan [b,c,d,e,1], Heng Sun [f], Naveed Rauf [b,c,d,g], Tingzhang Wang [h], Zi Yin [a,c,d,*], Xiao Chen [b,c,d,*]

[a] Department of Orthopedic Surgery of Sir Run Run Shaw Hospital, and Liangzhu Laboratory, Zhejiang University School of Medicine, Hangzhou, China
[b] Department of Sports Medicine & Orthopedic Surgery, The Second Affiliated Hospital, and Liangzhu Laboratory, Zhejiang University School of Medicine, Hangzhou, China
[c] Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cells and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China
[d] China Orthopedic Regenerative Medicine Group (CORMed), Hangzhou, China
[e] Key Laboratory of Novel Targets and Drug Study for Neural Repair of Zhejiang Province, Department of Clinical Medicine, School of Medicine, Hangzhou City University, Hangzhou, Zhejiang, China
[f] Medical 3D Printing Center, Orthopedic Institute, Department of Orthopedic Surgery, The First Affiliated Hospital, School of Basic Medical Sciences, Suzhou Medical College, Soochow University, Suzhou, Jiangsu 215000, China
[g] Key Laboratory of Tissue Engineering and Regenerative Medicine of Zhejiang Province, Zhejiang University School of Medicine, Hangzhou, China
[h] Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, China

ARTICLE INFO

Keywords:
Axolotl
Long-read sequencing
Regeneration
Novel transcribed regions
Alternative splicing

ABSTRACT

Axolotls are known for their remarkable regeneration ability. Exploring their transcriptome provides insight into regenerative mechanisms. However, the current annotation of the axolotl transcriptome is limited, leaving the role of unannotated transcripts in regeneration unknown. To discourse this challenge, we exploited long-read sequencing technology, which enables direct observation of full-length RNA transcripts, greatly enhancing the coverage and accuracy of axolotl transcriptome annotation. By utilizing this method, we identified 222 novel gene loci and 4775 novel transcripts, which were quantified using short-read sequencing data. Through the inclusive analysis, we discovered novel homologs, potential functional proteins, noncoding RNAs, and alternative splicing events in key regeneration pathways. In particular, we identified novel transcripts with high protein-coding potential implicated in cell cycle regulation and musculoskeletal development, and regeneration were identified. Interestingly, alternative splice variants were also detected across diverse pathways critical to regeneration. This specifies that these novel transcripts potentially play vital roles underpinning the robust regenerative capacities of axolotls. Single-cell transcriptomic analysis further revealed these isoforms to predominantly exist in axolotl limb chondrocytes and mature tissue cell populations. Overall, the findings significantly advanced consideration of the axolotl transcriptome and provided a new perspective for understanding the mechanisms of regenerative abilities of axolotls.

1. Introduction

The limited regenerative abilities of the human body present a major challenge in clinical settings. However, axolotls, a species of salamander, exhibit extraordinary regenerative capabilities. They can regenerate a wide range of organs and tissues, including limbs, spinal cords, hearts, and even parts of their brain. Therefore, axolotls, widely utilized as a model organism, play a significant role in studying organ regeneration. Despite substantial efforts made by the axolotl research community to advance the understanding of this process, our knowledge of the precise molecular mechanisms driving axolotl regeneration remains confined.

The analysis of the transcriptome provides valuable information about the functional, physiological, and biosynthetic aspects of cells, along with the underlying molecular mechanisms. For example,

* Corresponding authors at: Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cells and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China.
E-mail addresses: yinzi@zju.edu.cn (Z. Yin), Chenxiao-610@zju.edu.cn (X. Chen).
[1] These authors contributed equally.

https://doi.org/10.1016/j.csbj.2024.08.014
Received 27 May 2024; Received in revised form 8 August 2024; Accepted 14 August 2024
Available online 21 August 2024

transcriptome analysis revealed the expression of oncogenes such as *Cirbp* and *Kazald1* during the early stage of blastema formation in axolotls. These oncogenes are crucial for tissue regeneration in axolotls [1,2]. Single-cell RNA sequencing (scRNA-Seq) as well as single-nuclei RNA sequencing (snRNA-Seq) technologies further explore and resolve the cellular heterogeneity and provide comprehensive information about cell subtype-specific gene expression within multiple lineages in homeostatic and regenerating limbs [3–6]. However, the effectiveness of high-throughput short-read RNA sequencing (RNA-seq) methods in studying the transcriptome is hindered by the limitations of reference sequence annotation, which impact the accuracy and inclusiveness of the results. Therefore, the full potential of RNA-seq in advancing transcriptome research remains constrained despite its innovative impact.

Short-read sequencing is fundamentally limited in accurately capturing the complex transcriptional patterns generated by diverse mechanisms in eukaryotes, such as alternative splicing, alternative transcription initiation, and alternative transcription termination sites [7,8]. Likewise, when analyzing and annotating axolotl transcriptomes using short-read data, researchers face challenges due to the whole-genome duplication event. This duplication event may lead to the generation of sets of transcripts that are highly similar to each other, making it difficult to differentiate them solely based on short-read sequences. To accurately distinguish these transcripts, a thorough examination of their full-length sequences is required.

Long-read sequencing has emerged as a powerful solution to overcome these challenges by enabling the direct observation of complete transcripts [9,10]. This technology revolutionized the annotation of transcriptomes and genomes in various organisms. Its ability to accurately define novel transcribed regions (NTRs) and distinguish between closely related isoforms of annotated genes significantly expanded our understanding of the intricate transcriptional landscape. Long-read sequencing has been proven to be an invaluable tool in unraveling the complexities of gene expression regulation [11–14]. Moreover, it has the potential to enhance the detection/identification of cell type-specific genes in both bulk and single-cell RNA-seq datasets, leading to improved resolution and accuracy in cell clustering [15]. In this study, we utilized the state-of-the-art SMRT (Single Molecule Real-Time) platform developed by Pacific Biosciences (PacBio). The SMRT platform is renowned for its exceptional sequencing accuracy and not being limited by the transcript length [9,16]. With this technology, we improved the annotation of the axolotl transcriptome and discovered the involvement of novel transcribed regions (NTRs) in axolotl regeneration.

2. Materials and methods

2.1. Axolotl husbandry, surgery

This study followed essential ethical guidelines and received ethical approval from the Zhejiang University Ethics Committee (ZJU11602) for conducting animal experimentation. The axolotl colony (*Ambystoma mexicanum*) in this study was sourced from the Guilin Axolotl Breeding Base. Adult axolotls (average body length 18 cm, neotenic) were randomly selected as participants in the experiments. Prior to any surgical procedures, the animals were anesthetized using 0.03 % ethyl 3-aminobenzoate to ensure the welfare of the animals. The anesthesia was administered until the animals no longer responded to a tail pinch stimulus, ensuring they were in a fully anesthetized state during the surgical procedures. This approach aimed to minimize any potential discomfort or distress experienced by the animals during the experimental procedures. After anesthesia, the tails were amputated at the 1 cm site from the tip. The axolotls were sacrificed 3 days after surgery. We collected organs from 3 different adult axolotls, including testes, limb segments, blood vessels (aortas), livers, spleens, kidneys, spinal cords, brains, hearts, gill filaments, tails, and the regenerated tissues 3 days after tail amputation. We separately extracted RNA from tissues to make the total RNA of each sample identical, and then mixed these RNA to build the library.

2.2. Full-length library construction and sequencing

Full-length cDNA was synthesized using the SMARTer™ PCR cDNA Synthesis Kit. Subsequently, full-length cDNA fragments were size-selected using BluePippin and utilized to construct a sequencing cDNA library. After passing library quality control, the full-length transcriptome sequencing was performed using PacBio RS II.

2.3. Long-read mapping, redundancy removal, and similarity analysis

In this study, the full-length transcriptome sequencing generated 42.62 Gb of clean data (the data without adapter sequences, low-quality sequences, and redundant sequences). The sequencing data contained a total of 566,919 circular consensus (CCS) reads, of which 469,945 were full-length non-chimeric (FLNC) sequences. After clustering and polishing the FLNC sequences, 16,923 non-redundant transcript sequences were obtained. Genomic Mapping and Alignment Program (GMAP, version 2023–03-24) [17] was employed for PacBio long-read sequence mapping using the axolotl genome (AmexG_v6.0-DD.fa, available at https://www.axolotl-omics.org/assemblies). Initially, the gmap_build tool was used to build an index of the genome. To deal with the large genome of axolotl, an equivalent program called gmapl was utilized. The process included cross-species comparison, prohibition of short insertions/deletions, and alignment of full-length transcripts to the reference genome on the positive strand. The SAM results were subsequently processed using the collapse_isoforms_by_sam.py tool from the cDNA_Cupcake (version 29.0.0) to filter out sequences with identity under 90 % and coverage below 85 %, and merge with differences in exons only at the 5' ends. The makeblastdb (version 2.12.0) tool from BLAST was then used for similarity analysis. Sequences with less than 80 % identity and overlaps over 50 bp were excluded.

2.4. Long-read sequencing transcript annotation

The long-read transcript data were analyzed and compared with the axolotl genome annotation file (AmexT_v47-AmexG_v6.0-DD.gtf) using gffcompare (version 0.12.6) [18]. The transcripts were marked as follows: "u" represents potentially novel genes, "e" and "j" stand for potentially novel isoforms or inaccurate references, "= " indicates an exact match to the annotation, "c" represents a sequential subset of exons contained within the annotation (reference-annotated exonic subsets of transcripts with missing exons), and "i," "o," "p," "r," "s," and "x" are grouped under other transcripts. Next, we categorize the results with class codes "u", "e", and "j" as novel transcript regions (NTRs), as demonstrated in our dataset (GSA: CRA013464).

2.5. Short-read RNA sequencing

For short-read RNA-seq, RNA was extracted from sample sets containing more than three adult axolotls. The quality of the RNA was evaluated using an Agilent Bioanalyzer, and samples with a RIN (RNA Integrity Number) score of 9 or higher were selected. The RNA samples were treated with the Ribo-Zero rRNA removal kit from Illumina. TruSeq cDNA libraries were then prepared and sequenced on the Illumina HiSeq 2500 platform, generating paired-end reads of 100 base pairs. The resulting library contained more than 25 million paired reads. To analyze short-read data, hisat2 [19] (version 2.2.1) was used for alignment and raw reads convert against a reference genome to generate SAM files. Following alignment, samtools [20] (version 1.7) was utilized to convert the SAM file to BAM format. We applied featureCounts [21] (version 2.0.1) to count features from the aligned reads in the annotation file, which integrated reference transcriptome (Am_34-AmexG_v3.gtf) and NTRs by cuffmerge [22] (version 2.2.1). The counts obtained then

were transformed into TPM values, and transcripts were categorized into different TPM ranges (>100, 10–100, 1–10, 0.1–1, and <0.1). The number of transcripts falling into each TPM range was then counted, and a bar plot visualization was generated using ggplot2.

2.6. Characterization of NTRs with high protein-coding potential

To assess the protein-coding potential (CP) of the identified NTRs, we employed CPAT (v1.2.2) (http://rna-cpat.sourceforge.net/#cpat-py) for their classification. First, NTRs were extracted from the AmexG_v6.0-DD.fa file using gffread and utilized to create an index by samtools. Next, the make_hexamer_tab.py from CPAT was used to calculate the frequency of intra-frame hexamers based on the.fasta file, while excluding UTR mRNA sequences. The make_logitModel.py was employed to construct a species-specific model. Following this, CPAT.py was used to predict the coding potential of the sequences. Visualization of the results, including the determination of the coding probability (CP) cut-off value was achieved using the 10Fold_CrossValidation.r (an R package available on CPAT's website). The CP cut-off value is 0.493. This value ensures a comprehensive analysis of the coding potential within the NTRs based on sequence-specific features. NTRs with high CP were further examined for sequence conservation with known proteins and functional conservation with known protein domains. For the extraction of open reading frames (ORFs) from the high CP transcripts, we utilized TRANSDECODER [23] (v3.0.0) to ensure the inclusion of more than 300 bp ORFs and reduce false discovery rates in subsequent analyses, albeit that it potentially excluded short ORFs. The obtained ORFs were cross-verified using Blastx and then subjected to Blastp (version 2.2.26) against the UniProt database. Additionally, they were scored against PFAM-A using HMMER [24] (version 3.1b1).

2.7. Characterization of putative noncoding NTRs

We investigated the evolutionary conservation of putative noncoding NTRs using the PHAST [25] software suite (version 1.5, available at http://compgen.cshl.edu/phast/resources.php). The investigation was initiated by constructing indexes of the genomes for *X. tropicalis* (xenTro10), the African clawed frog (xenLae2), and the Tibetan frog (nanPar1) using Lastdb. Low CP NTRs were then aligned to these genomes using lastal and last-split, generating MAF files. These files were subsequently integrated, and a model was built using phyloFit based on these alignments. Conservation was predicted using both the phastCons and phyloP algorithms. PhastCons generated and output conservation scores to phastCons.scores.wig, while phyloP applied the likelihood ratio test to identify evolutionary changes, with results saved in phyloP.scores.wig. Both algorithms set cutoff values at the median of the scores within the 95th percentile for base-wise conservation analysis (phyloP) and contiguous window conservation analysis (phastCons).

2.8. Unraveling variable splicing events with long-read sequencing

The SUPPA2 [26] (version 2.3) provides a comprehensive toolkit for identifying alternative splicing events from full-length transcript.gtf files, which includes skipping exon (SE), alternative 5' or 3' splice sites (SS), mutually exclusive exons (MX), retained intron (RI), and alternative first and last exons (FL). The psiPerEvent function was utilized to calculate the percentage spliced in (PSI) values for these events, with results formatted in.psi files. The diffSplice function was further utilized to analyze differences in splicing between two samples, and output the results in.dpsi files. Additionally, cluster analysis was performed using clusterEvents function on the variable splicing events, setting a significance threshold of 0.05, a minimum cluster separation of 0.11, and a density threshold of 0.2. The analyses culminated in the visualization of the data through a pie chart of the splicing patterns.

2.9. Gene ontology (GO) enrichment analysis

Metascape (https://www.metascape.org) was used to perform GO analysis [27]. The R package ggplot2 was used to make the bar plots.

2.10. Single-cell RNA sequencing data analysis

The single-cell RNA sequencing data were sourced from the study by Qin et al. [4], which profiled a total of 938 individual cells. These cells were isolated from adult axolotl limbs at various time points during the regeneration process, including the unamputated forelimb (0 dpa), as well as the regenerating limb tips at 3-, 7-, and 21-days post-amputation (dpa). These time points represent the early, middle, and late stages of limb regeneration, respectively.

The single-cell RNA sequencing (scRNA-seq) expression data were analyzed using the Seurat v4.3.0.1 computational toolkit [28]. Following standard Seurat preprocessing, 15,608 highly variable genes across the dataset were identified. Principal component analysis (PCA) was then performed on this set of variable genes, which formed the basis for cell clustering and t-SNE visualization.

2.11. RNA fluorescence in situ hybridization (FISH) assay

Fluorescence-conjugated PB.4786.31-*Col1a2*, PB.1032.3-*Fbn1*, and PB.10723.1-*Fgfr2* probes for RNA FISH were generated in line with the protocols of Focobio Technologies. All RNA hybridization experiments were performed as described in the manuals of Focobio Technologies. Treated samples were visualized by confocal microscopy (FV1000, Olympus) (n = 3, 3 times replication in the laboratory).

2.12. Data access

All data needed to evaluate the conclusions in the paper are available in the Genome Sequence Archive (GSA) (CRA013464) maintained by the Beijing Institute of Genomics (BIG) Data Center [29,30]. Additional data related to this paper may be requested from the authors.

3. Results

3.1. High-resolution annotation of axolotl transcriptome

To explore axolotl transcripts and deepen our understanding of tissue regeneration, we utilized a combined experimental and computational methodology as depicted in Fig. 1A. Total RNA was extracted from multiple regenerating tissues of adult axolotls, including testes, limb segments, blood vessels (aortas), livers, kidneys, spinal cords, brains, hearts, gill filaments, tails, and 3-day-post-amputation regenerative tails, with the goal to collect a diverse array of the transcripts. Subsequently, cDNA libraries were constructed, and sequencing was performed on the PacBio SMRT platform. Similarly, a short-read RNA-seq dataset from the same tissue was generated for comparison (Fig. 1A). The Iso-Seq pipeline [31] was employed to cluster and assemble the raw high-quality long reads, resulting in the generation of final transcripts (see Methods). Through the full-length transcriptome sequencing, a total of 42.62 Gb of clean data was obtained. The sequencing data comprised 566,919 circular consensus (CCS) reads, with 469,945 of those being full-length non-chimeric (FLNC) sequences. In the initial assessment of the axolotl annotation illustration in the long-read transcriptome dataset, we aligned it with the latest version of the axolotl reference genome (AmexG_v6.0-DD) (https://www.axolotl-omics.org/) using GMAP, a gapped alignment mapper [32]. The transcripts with a %_identity lower than 80 and with more than 50 bp overlap were filtered out. As a result, 93,826 long-read transcripts were successfully mapped to the reference genome, with an overlap rate of 51.6 % to the 181,985 RefSeq transcripts (Fig. 1B). The high percentage of mapping to the reference sequence indicates the high quality of this dataset, supporting its utility

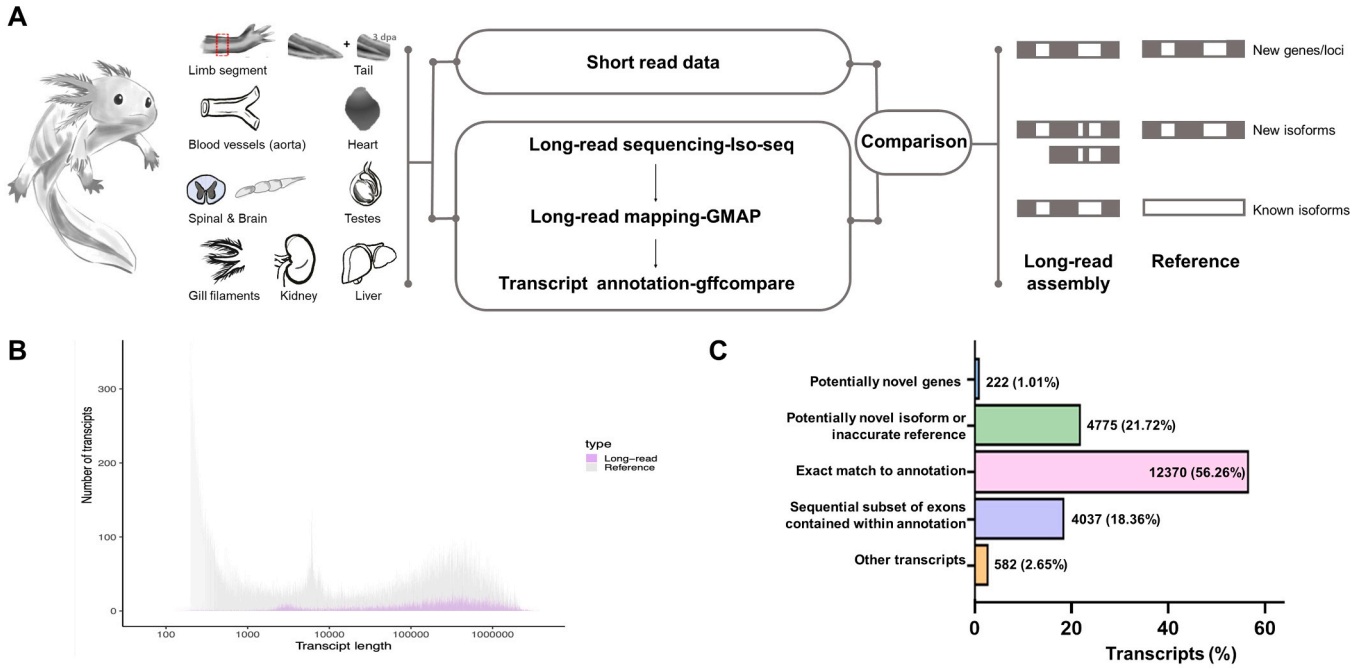


Fig. 1. Overview of the axolotl long-read transcriptome analysis. (A) Schematic showing the step-by-step process in the long-read transcriptome reconstruction pipeline. (B) Coverage of the reference axolotl transcriptome reference by the SMRT long-read RNA-seq. The gray histogram represents the RefSeq annotated transcript number across various transcript lengths. The light purple histogram represents the long-read transcripts that overlap with RefSeq annotations to any extent. (C) Potential novelty in the long-read transcriptome compared with the reference annotation.

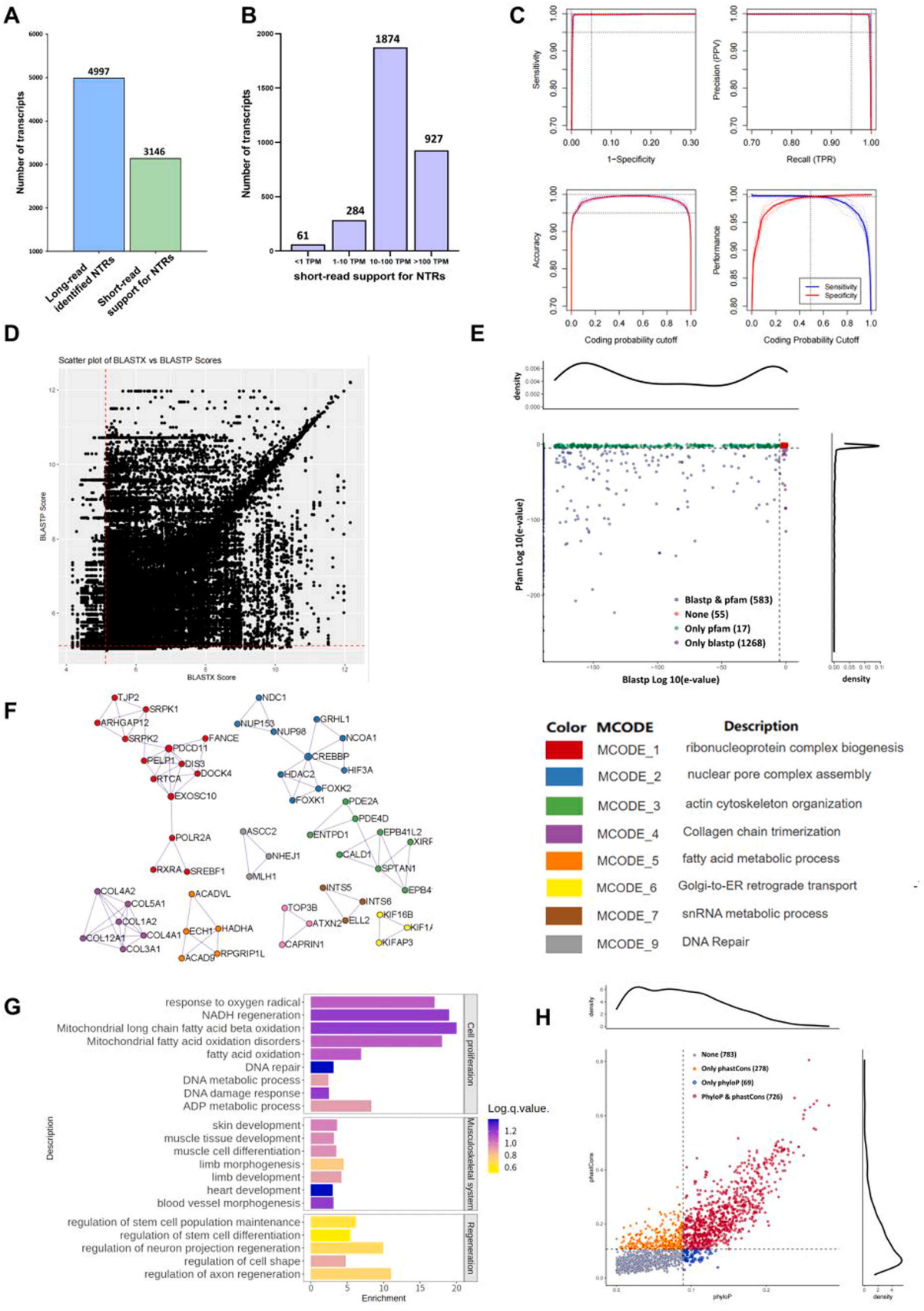
for the identification of novel transcripts.

In order to assess the presence of novel information within the long-read data, we analyzed the exon-intron structures of the long-read transcripts for similarity with the RefSeq transcriptome (Fig. 1C). To achieve this, we utilized the gffcompare tool from the IsoQuant suite to identify and classify the most closely matching reference transcript for each long-read transcript [33]. Our analysis revealed that out of 21,986 non-redundant exon-intron transcripts analyzed, only 12,370 transcripts (56.26 %) within the long-read data matched the annotation exactly (AmexT_v47-AmexG_v6.0-DD.gtf), indicating the presence of previously undiscovered transcriptional diversity. Among the identified transcripts, 4997 are considered promising novel discoveries, including 222 potentially novel genes. These 222 transcripts (1.01 %) signified potential NTRs that did not overlap with the reference annotation, while other 4775 transcripts (21.72 %) were classified as potentially novel isoforms of previously annotated genes, which may stem from various canonical alternative splicing events such as skipped exons and retained introns [34]. These NTRs should be further investigated as they may carry information relevant to tissue regeneration. An additional 582 transcripts (2.65 %) were mapped to genomic regions that are typically deemed less conducive to producing authentic transcripts, likely artifacts from sample preparation or the sequencing procedure. This category encompasses transcripts aligned to repetitive sequences, regions where exonic overlaps occur on the opposite DNA strand, and potential pre-mRNA fragments. Furthermore, a significant portion of these transcripts are short in length, posing challenges for their validation through computational methods due to their inherent complexity.

3.2. Characterization of novel transcribed regions in axolotls

To validate the expression of the 4997 potentially novel transcripts identified through long-read RNA-seq, we investigated whether there were any indications of expression in the short-read data by incorporating the NTRs into the reference transcriptome, and subsequently aligned the short-read data to the long-read transcriptome using hisat2 software [19]. By employing the methodology, we were able to quantitatively assess the expression levels of the identified novel transcripts [35]. Among the 4997 NTRs discovered by long-read seq, only 3146 (62.96 %) NTRs were detected in the short-read data, indicating that long-read RNA-seq significantly supplemented the NTR discovery (Fig. 2A). In order to assess the support for long-read transcripts from short-read RNA-seq data, we conducted a quantitative analysis to measure the expression levels of the corresponding short-read sequencing data within the NTRs. The results highlighted that a significant proportion, approximately 98 % of novel transcripts, exhibited a transcripts per million (TPM) value exceeding 1 %, and 89 % of these transcripts demonstrated consistently high expression levels (TPM $>$ 10) (Fig. 2B). These findings underscore the robustness and correlation in expression between long-read and short-read sequencing data, validating the support of short-read RNA-seq data for the identified long-read transcripts. These results indicate that long-read sequencing is more advantageous in detecting a greater number of NTRs compared to short-read data.

To identify potential protein-coding genes among the NTRs, we conducted a thorough analysis focusing on their protein-coding potential, sequence conservation with known proteins, and the presence of identified protein domains. We utilized the axolotl genome sequence to predict the protein-coding gene models. Given the scarcity of annotation information and limited protein-coding data available for the axolotl genome, we opted to utilize the genome and annotation database (assembly_Xenopus laevis_V10.1) of African clawed frog (*Xenopus laevis*) [36], which is phylogenetically closely related to the axolotl, to predict the protein-coding potential within the axolotl genome. Specifically, we employed the Coding-Potential Assessment Tool (CPAT) [37] to evaluate the protein-coding potential of these NTRs. CPAT evaluates the likelihood of these regions encoding functional proteins by a combination of metrics including open reading frame (ORF) length, sequence coverage by ORFs, Fickett's score for positional nucleotide usage, and hexamer bias in coding sequences. Through the application of CPAT, we steered a comprehensive assessment of the coding potential from the 4997 NTRs. Among them, CPAT identified 1923 sequences with a high protein-coding potential using a coding probability cutoff value of 0.493 (Fig. 2C). To probe deeper into these high coding potential transcripts,



(caption on next page)

Fig. 2. Characterization of NTRs and their coding potential. (A) Comparison of NTRs in long-read and short-read data. (B) The expression of novel transcripts identified from long-read data quantified by short-read data. The bar plot presents the number of short-read transcripts classified by TPM. (C) The workflow of CPAT. The AUC (area under the curve) plot (upper left) and the precision-recall (PR) curve (upper right) showed the rounds of cross-validations (blue dashed lines) and the averaged performance (the solid red curve); the two-graph ROC curve indicates the sensitivity and specificity achieved if $p = 0.493$ cutoff was used. (D) Scatter plot of BLASTX and BLASTP cross-validation analysis. The NTRs with a high bitscore, which is a score expressing the quality of the alignment, are used for downstream analysis. (E) The analysis of sequence homology and functional domain. The color scheme indicates significance ($e < 1 \times 10^{-5}$) by both BLASTP and Pfam (blue), BLASTP only (purple), or Pfam only (green). The point densities along the x- and y-axes are represented by marginal rug plots. (F) The network of genes is densely connected by using the MCODE algorithm (left) and the GO terms of top3 MCODEs (right). (G) Bar plot of GO analysis on all BLASTP and Pfam NTRs. (H) Characterization of putative noncoding NTRs. The x-axis represents the base-wise transcript conservation levels, measured as the fraction of conserved bases (base-wise phyloP score > 0.89). The y-axis represents the maximal 200-bp window conservation levels, calculated using a sliding window average phastCons score along the transcript. The color scheme indicates significance ($e < 1 \times 10^{-3}$) by both PhyloP and phastCons (red), phyloP only (blue), or phastCons only (yellow). The point densities along the x- and y-axes are represented by marginal rug plots.

we proceeded with additional evaluations. Initially, we assessed their sequence similarity to known proteins using BLASTP [38]. Afterward, we investigated the presence of functional protein domains by performing hmmscan against the Pfam database [39,40], and combined both analyses to filter the transcripts. As depicted in Fig. 2E, we identified 583 sequences that encoded functional proteins based on BLASTP (x-axis) and Pfam (y-axis) assessments (Fig. 2E; Table S1) with a cross-check of blastX (Fig. 2D). To further understand the function of these high coding potential transcripts, MCODE algorithm was then applied to identify the networks of the densely connected proteins (Fig. 2F). The NTRs in MCODE_1 were significantly enriched for roles in fundamental cellular processes such as ribonucleoprotein complex biogenesis. MCODE_4 predominantly contained COL3A1, COL4A1, COL5A1, and other collagens, suggesting a pivotal role of these NTRs in collagen fibril assembly and additional extracellular matrix architectures involving multimeric complex formation. To further characterize the 583 identified NTRs, we conducted a GO enrichment analysis of their encoded genes. Interestingly, we found that beyond involvement in cell proliferation and DNA repair, these NTRs demonstrated profoundly significant roles contributing to limb regeneration (Fig. 2G). Additionally, the NTRs were enriched for functions related to the regulation of stem cell differentiation, axon regeneration, and neuron projection regeneration. These results suggest that these NTRs may play crucial roles in supporting axolotl regeneration. The GO terms also revealed mitochondrial-related functions such as mitochondrial fatty acid oxidation, which is consistent with our previous findings on the involvement of mitochondria in axolotl regeneration [4]. This further highlights the importance of mitochondria in axolotl regeneration. Taken together, these results collectively demonstrate that the NTRs with high protein-coding potential evaluated by CPAT are not only involved in cell cycling and DNA repair, but also contain multiple collagen-coding proteins involved in the development and regeneration of the musculoskeletal system, which may account for the unique regeneration ability of axolotls.

Previous studies have indicated that long noncoding RNAs (lncRNAs) demonstrate a moderate level of sequence conservation [41,42]. Based on this indication, we next focused on exploring the non-protein-coding NTRs that were not identified as protein-coding genes in our analysis, and examined whether they exhibited evolutionary conservation as well. For this purpose, we calculated the fraction of significantly conserved bases in non-coding NTRs using the phyloP algorithm [43], followed by phastCons algorithm that could determine the maximally conserved sliding window within a phylogenetic alignment of three vertebrate species (*X. tropicalis*, African clawed frog, and Tibetan frog) using the phastCons algorithm [44]. The dual algorithmic approach analysis encompasses the identification of both conserved positions within individual transcripts and contiguous regions displaying high levels of conservation. The result demonstrated 1073 transcripts (278 phastCons only, 69 phyloP only, and 726 phastCons & phyloP) with increased conservation levels among 1856 putative non-protein-coding NTR group (Fig. 2H), indicating that a significant proportion of the non-coding NTRs are similar to the reference organisms. Even though these organisms don't have limb regeneration ability as seen in axolotls, the regeneration-promotion non-coding transcripts may stem from the remaining 783 NTRs. These highly-conserved lncRNAs through the analysis are listed in Table S2.

To further validate the expression of NTRs in various organs of axolotls, we employed fluorescence in situ hybridization (FISH) to detect PB.4786.31-*Col1a2*-AMEX60DD022398 (transcript_id) and PB.1032.3-*Fbn1*-AMEX60DD003755 (transcript_id) within high coding potential transcripts. We examined the expression levels of these two NTRs in the limbs, tails, gills, spinal cords, spleens, livers, hearts, aortas, limb regenerating tissues (7- and 14-days post-amputation, dpa), and tail regenerating tissues (7 dpa), respectively. PB.4786.31-*Col1a2* exhibited lower expression in normal limbs compared to regenerating tissues. Sparse cellular expression was observed in the subcutaneous mesenchymal tissues of 7 dpa limbs, with more expressing cells present in the subcutaneous mesenchymal tissues at 14 dpa (Fig. 3A). PB.4786.31-*Col1a2* was found to be relatively uniformly expressed in the tail fin cells, but in the 7 dpa regenerating blastema, some highly expressing cells were detected in the subcutaneous tissues (Fig. 3B). In the gills, PB.4786.31-*Col1a2* was identified with multiple cellular expressions in connective tissues (Fig. 3C). PB.4786.31-*Col1a2* was also found to be highly expressed in the spleens and in individual cells within liver and heart tissues. However, PB.4786.31-*Col1a2* was rarely expressed in the aortas (Fig. 3C). PB.1032.3-*Fbn1*-AMEX60DD003755 (transcript_id) exhibited low expression in normal limbs, with a lower expression level observed in the regenerating blastema of 7 dpa limbs. However, an increased number of positive cells were detected in the subcutaneous mesenchymal tissues at 14 dpa. PB.1032.3-*Fbn1* was found to have low expression in normal tail fin cells. Partial positive cells were identified in the spinal cord. PB.1032.3-*Fbn1* expression was scarce in gills. Fewer PB.1032.3-*Fbn1* positive cells were also detected in the spleens and hearts. In contrast, a higher number of PB.1032.3-*Fbn1* highly expressing cells were found in the livers. Nevertheless, PB.1032.3-*Fbn1* expression was low in aortas (Fig. S1).

3.3. Characterization of alternative splicing in axolotls

Alternative splicing plays a vital role during a lot of biological processes. In the current study, we identified 4345 potentially novel alternative splicing, thus we characterized the potential alternative isoforms present in axolotls using long-read data. We employed SUPPA2 (v2.2.1) and calculated the occurrence of seven types of alternative splicing events, including alternative first and last exons (AF/AL), alternative 5' and 3' splice sites (A5/A3), retained introns (RI), skipping exons (SE), as well as mutually exclusive exons (MX). By quantifying these alternative splicing events among the splice junction-validated spliced isoforms, we gained insights into the diversity of alternative isoforms in axolotls (Fig. 4A). Among them, 1635 SE (37.6 %) events were identified, which was the most prevalent, followed by 1269 AF (29.2 %). A5 accounted for 11.4 %, A3 8.5 %, RI 8.1 %, AL 3.1 %, and MX 2 % (Fig. 4B). As SE was reported to be the most common alternative splicing event, and has been used as a therapeutic target [45], it is possible that SE could also play an important role in axolotl tissue homeostasis maintenance and regeneration. Therefore, we performed gene ontology (GO) analysis to explore

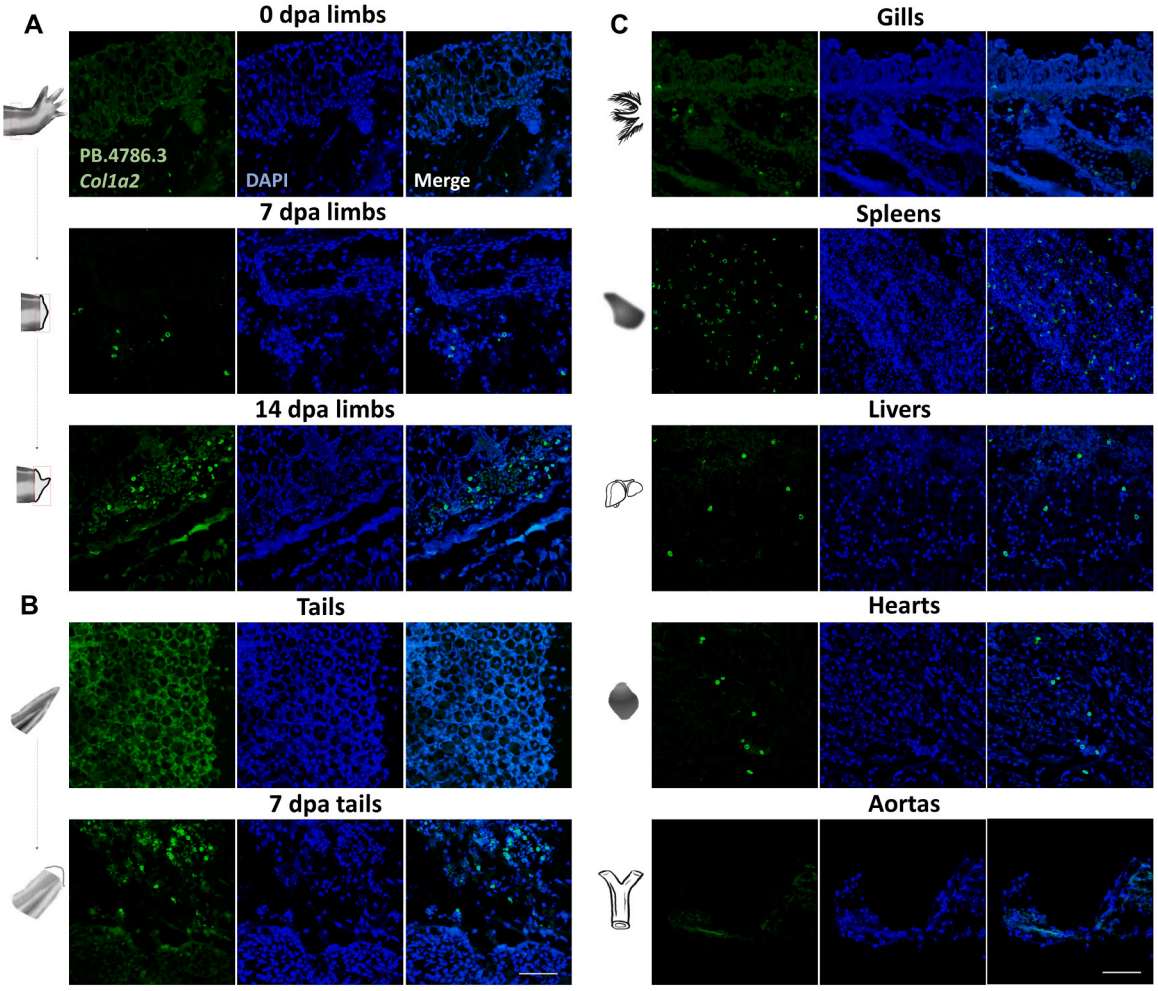


Fig. 3. The expression PB.4786.31-*Col1a2* in organs of axolotls. (A) FISH images of PB.4786.31-*Col1a2* FITC fluorescent probe staining in normal limbs, 7 dpa limbs, and 14 dpa limbs of axolotls. Left: PB.4786.31- *Col1a2* FITC fluorescent probe; middle: DAPI; right: merged image of FITC and DAPI. (B) FISH images of PB.4786.31-*Col1a2* FITC fluorescent probe staining in normal tails and 7 dpa tails of axolotls. (C) FISH images of PB.4786.31- *Col1a2* FITC fluorescent probe staining in gills, spleens, livers, hearts, and aortas. Bar= 50 μm.

the specific functions of these SE genes. The results depicted that SE genes function mainly contained viral infection pathways, cellular responses to stress, monocarboxylic acid metabolic process, motor proteins, Parkinson's disease, biological oxidations, extracellular matrix organization, and response to hormone (Fig. 4C). Among them, the most significant GO terms included cellular responses to stress, biological oxidations, and extracellular matrix organization (Fig. 4C1), indicating that the environmental change-response genes are likely to be regulated through SE. We also observed that the SE genes were significantly enriched in regeneration-related signaling pathways such as Wnt, FGF, TGF-β, Notch, and mTOR (Fig. 4D). This result supports our hypothesis that SE could be a critical event during tissue regeneration. All the alternative splicing (AS) genes are listed in Table S3, and all the SE GO terms are listed in Table S4.

Previous studies have reported the crucial roles of various signaling pathways including Wnt, TGF-β, FGF, BMP, Notch, and mTOR [46] in organ regeneration in axolotls. Given the enrichment of SE genes in these regeneration-related signaling pathways, we hypothesized that alternative splicing events may play a role in regulating these pathways and thereby influencing the regenerative process. Therefore, we proceeded to analyze the transcriptional alternative splicing of key genes involved in these critical regeneration signaling pathways (Fig. 5A, S2). Among them, the FGF signaling pathway was found to have the highest number of genes with alternative splicing events, including *Cttn*, *Fgfr2*, *Grb2*, *Mapk3*, *Ptpn11*, and *Ssh1*. This suggests a potential involvement of splicing events associated with these splicing factors in the axolotl regeneration process, possibly acting through the FGF signaling pathway. Other signaling pathways such as the mTOR, Notch, TGFβ, Wnt, and BMP also exhibit multiple alternative splicing events, indicating the involvement of multiple alternatively spliced transcripts in critical regeneration pathways. In order to gain a deeper understanding of the potential roles of the alternative splicing transcripts in regeneration, we examined the expression patterns of the corresponding pathways. To achieve this, a single-cell transcriptomic data from axolotl regeneration processes (0, 3, 7, and 21 dpa) were analyzed [4]. The identity of these clusters followed the research in 2021 [4], known as clusters 0–6, including identified as the general cluster, mitochondria (MT) cluster, connective tissue (CT) cluster, chondrocyte cluster, inflammation cluster, cycling cluster, and apical epithelium cap (AEC) cluster. The mitochondria cluster means the cluster highly expressed mitochondria-related genes (Fig. 5B). Key genes involved in these regeneration-related pathways, such as *Fgfr2*, *Ppp2ca*, *Smad2,* and *Bmr1b*, were predominantly expressed in cluster 3 (chondrocyte cluster) (Fig. 5C) and showed significant higher expression in mature tissues (0 dpa) and relatively mature tissues (21 dpa) (Fig. 5D). These findings indicate that alternative splicing transcripts of critical genes involved in regeneration pathways are primarily expressed in more mature tissues, suggesting a possible association between the maintenance of

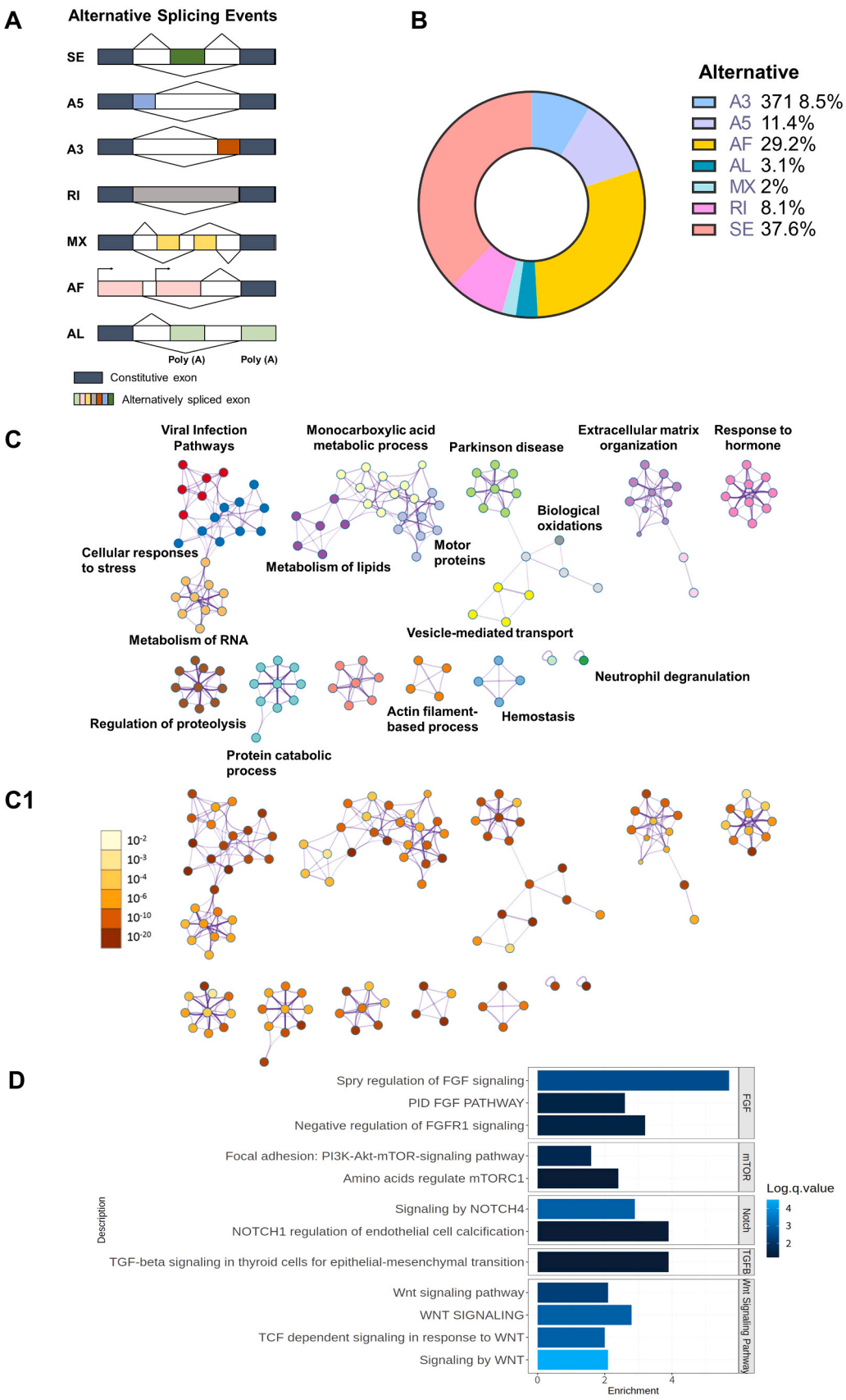


Fig. 4. Identification of alternative splicing events and differential splicing events in axolotls. (A) Schematic of the seven types of alternative splicing events. (B) The proportions of alternative splicing events in the seven types. (C) GO terms of SE genes represented by circle node network, and colored by GO cluster identity and p-value (C1). (D) Bar plot of SE genes enrichment in key signaling pathways for regeneration.

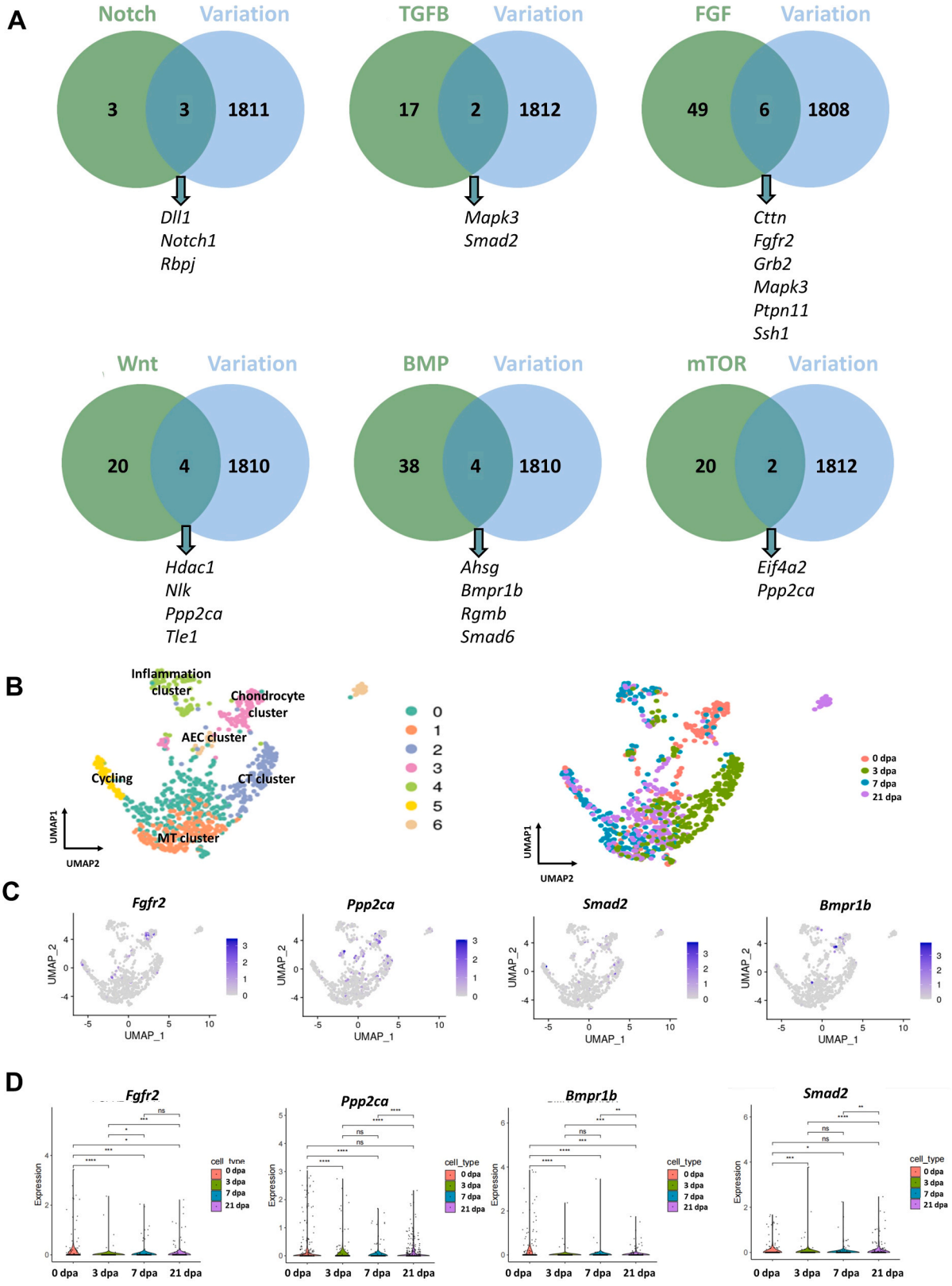


Fig. 5. Alternative splicing in key pathways for developmental regeneration in axolotls (A) Venn diagram of the overlapping parts of key pathway genes and alternative splicing genes for developmental regeneration in axolotl. (B) Uniform manifold approximation and projection (UMAP) visualizations identified using the computational pipeline colored by clusters (left) and time (right). (C) UMAP visualization of scRNA-seq data overlaid with expression of the alternatively spliced genes in regenerative signaling pathways. (D) Violin plots of the alternatively spliced genes in regenerative signaling pathways. One-way ANOVA was performed to assess statistically significant differences between groups. * $p < 0.05$, * * $p < 0.01$, * ** $p < 0.001$, * ** * $p < 0.0001$.

regenerative capacity in mature tissues and these alternative splicing events. Interestingly, *Notch1* expression was high in the early regenerative limb (3 dpa) and mature (0 dpa) tissues, but its expression significantly decreased at 7 dpa (Fig. S2). These data imply a dedifferentiation process during the middle stage of regeneration, which warrants further exploration.

We further validated the expression of alternative splicing PB.10723.1-*Fgfr2* in axolotls. It was detected in the subcutaneous region of normal limbs, with individual positive cells observed in the regenerating blastema of 7 dpa limbs (Fig. 6A). The number of positive cells increased at 14 dpa. PB.10723.1-*Fgfr2* was found to have a high expression in normal tail fin cells but exhibited lower expression in the regenerating blastema at 7 dpa (Fig. 6B). PB.10723.1-*Fgfr2* expression was low in gills. Fewer PB.10723.1-*Fgfr2* positive cells were also detected in the spleen and liver. Notably, a higher number of PB.10723.1-*Fgfr2* highly expressing cells were identified in heart valves. However, PB.10723.1-*Fgfr2* expression was low in aortas (Fig. 6C).

4. Discussion

Long-read sequencing technology is now being widely used to annotate novel transcripts and splicing events in both model organisms and cancer tissues. In this particular study, we successfully developed a cutting-edge analysis pipeline that capitalizes on the latest advancements in long-read sequencing technology. This pipeline significantly improved the precision and accuracy of the axolotl transcriptome annotation. By implementing this approach, we identified numerous presumptive protein-coding transcripts that were previously unknown. Additionally, we discovered thousands of novel isoforms of genes that were already annotated. It is important to note that our findings underwent meticulous and systematic validation to ensure their reliability and robustness. This study represents a significant advancement in understanding the axolotl transcriptome, shedding light on previously unexplored aspects of gene regulation and splicing events in this remarkable organism. Our results significantly expanded transcriptome annotation.

In the axolotl second-generation sequencing studies, diverse transcriptome data including both coding and non-coding RNAs have been presented. These studies provide valuable references and comparisons for our data [47]. In our study, we discovered 4997 novel isoforms or gene loci that were not previously annotated. These findings presented almost 40 % expansion of the axolotl RefSeq catalog, providing a more comprehensive understanding of the axolotl transcriptome. The NTRs were rigorously validated by short-read RNA-seq data, sequence homology, as well as functional domain conservation. Among the NTRs analyzed, we identified 583 candidate protein-coding genes involved in regeneration through DNA repair and regulation of the cell cycle, primarily focusing on the regulation of the musculoskeletal system and limb regeneration. This includes multiple collagen genes of NTRs. In recent years, multiple articles have reported that collagen plays a

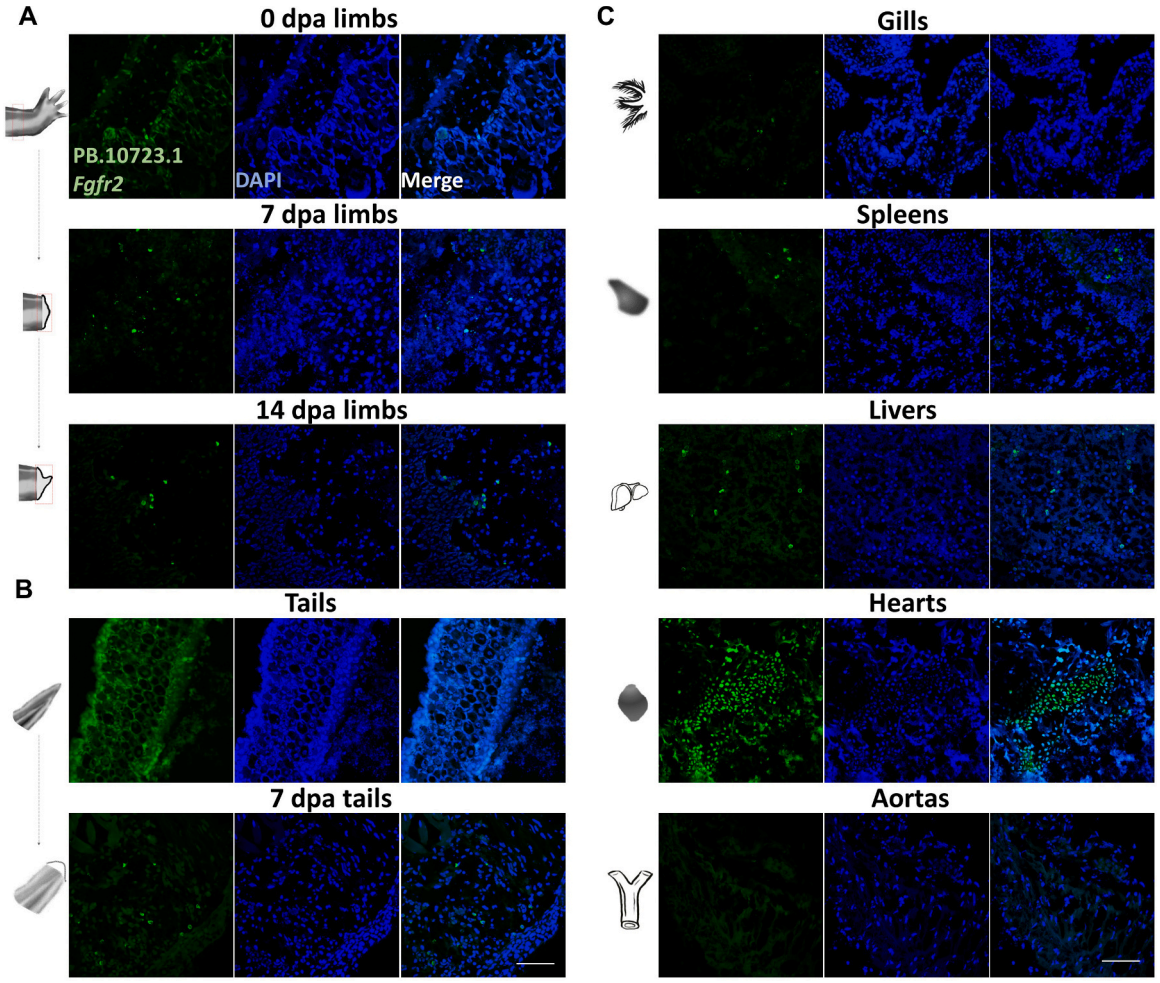


Fig. 6. The expression PB.10723.1-*Fgfr2* in organs of axolotls. (A) FISH images of PB.10723.1-*Fgfr2* FITC fluorescent probe staining in normal limbs, 7 dpa limbs, and 14 dpa limbs of axolotls. Left: PB.10723.1-*Fgfr2* FITC fluorescent probe; middle: DAPI; right: merged image of FITC and DAPI. (B) FISH images of PB.10723.1-*Fgfr2* FITC fluorescent probe staining in normal tails and 7 dpa tails of axolotls. (C) FISH images of PB.10723.1-*Fgfr2* FITC fluorescent probe staining in gills, spleens, livers, hearts, and aortas. Bar= 50 μm.

significant regulatory role in skin and musculoskeletal regeneration [48–50]. Previous single-cell studies on axolotl limb regeneration indicated distinct involvement of different collagens (*Col1a1* and *Col2a1*) in the process [4]. However, the complex expression and regulation of collagen genes in axolotl limb regeneration remain unclear. Here, using long-read sequencing technology, we predicted NTRs and identified multiple collagen gene transcripts, possibly uncovering the mystery of axolotl limb regeneration. In addition, we also discovered genes in NTRs that are associated with mitochondrial metabolism regulation (mitophagy, mitochondrial fatty acid oxidation). Mitochondria have been proven to play an important role in the regeneration of various tissues and organs, including the nervous system [51,52], liver [53], skin [54], and musculoskeletal system [55]. It has been reported to regulate tissue regeneration and aging through mechanisms such as dynamics [56], ROS (reactive oxygen species) regulation [53], metabolic reprogramming [51], as well as translocation and cell communication [57]. This finding aligns with our previous study on mitochondrial regulation of axolotl limb regeneration, further suggesting that mitochondrial metabolism is an important factor in axolotl regeneration, and providing crucial clues for enhancing organ regeneration ability in mammals. Further research and functional experimental validation of these NTRs may become a new breakthrough point in exploring the unique regenerative abilities of axolotls.

When identifying NTRs, we added extra validation steps including BLASTP and Pfam assessments to exclude the known genes or paralogs as possible. BLASTP infers structure and function by comparing the similarity of protein sequences, and it uses local sequence alignment methods (such as the Smith-Waterman algorithm). When the sequence similarity between paralogs is low, BLASTP can distinguish them more easily. The Pfam database classifies protein sequences based on Hidden Markov Models (HMMs), grouping proteins with common ancestors and similar functions into different families [58]. When paralogs belong to different Pfam families, their functional and structural differences are usually also significant, so they can be effectively distinguished by the prediction of BLASTP combined with Pfam. Combining with BLASTP and Pfam supposed to exclude most potential paralogs with low sequence similarity. However, since the axolotl gene annotation is limited, and we are unable to compare our data with more species in a limited study period, it cannot be thoroughly excluded that some NTRs could be paralogs of the known genes of other species. Therefore, more in depth investigation should be done in the future.

Alternative splicing isoforms are widely present in organisms and play vital roles in various biological processes, including cell differentiation, development, tissue-specific functions, and adaptation to environmental changes. Abnormal alternative splicing events are associated with various diseases, such as cancer, neurological disorders, and genetic diseases. In axolotl studies, three isoforms of hnRNP G with a differential binding affinity for a specific RNA probe were identified in the oocytes [59]. Thyroid hormones (THs) have been discovered to undergo alternative splicing, which affects gill regeneration in axolotls [60]. Our analysis using long-read data showed SE had the largest proportion in the identified isoforms. The functions of these alternatively spliced transcripts are mainly related to cellular responses to stress, biological oxidations, and extracellular matrix organization. The findings suggested that axolotl exhibits a higher prevalence of SE alternative splicing events in response to stress and biological oxidations. Through skipped exon (SE) type alternative splicing, axolotl may activate specific pathways involved in stress response, inflammation, and tissue repair. This enhanced ability to modulate gene expression and protein isoforms may contribute to the remarkable regenerative capacity observed in axolotls, consenting them to efficiently regenerate complex tissues and organs.

We also investigated the role of alternative splicing transcripts in axolotl regenerative pathways and found that multiple alternative splicing transcripts are present in the pathways, highlighting the importance of alternative splicing in axolotl regeneration. Through long-read sequencing, alternative splicing is observed in several critical transcripts of the Wnt pathway, including *Hdac1*, *Nlk*, *Ppp2ca*, and *Tle1*. The Wnt signaling pathway has been extensively validated for its crucial regulatory role in mammalian stem cell self-renewal and tissue regeneration [61]. In axolotl regeneration, the Wnt pathway was reported to regulate various functions during limb development and regeneration, such as limb bud initiation, growth, patterning, and skeletal differentiation, while controlling downstream factors like *Fgf10* and *Fgf8* [62]. Alternative splicing enhances gene expression diversity and functional complexity, regulating gene expression levels, enhancing protein functional diversity, and modulating cell development and adaptation to environmental changes [63]. The unique regenerative ability of axolotls may be associated with the alternative splicing of these key regeneration-related genes, opening up new avenues for exploring the mechanisms underlying axolotl regeneration. Combining single-cell transcriptomic sequencing analysis, we identified that these regenerative alternative splicing genes are primarily expressed in the axolotl chondrocyte subpopulation and mature tissues. The regenerative potential of mature axolotl tissues is a remarkable characteristic that sets them apart from most other vertebrate species most probably due to their unique cellular and molecular responses to injury, the presence of a permissive extracellular matrix environment, and the activation of specific regenerative pathways. In line with this, our previous studies demonstrated a significant relationship between axolotl regeneration and the dedifferentiation and redifferentiation of the chondrocyte cluster within mature tissues. Current study further revealed the existence of numerous alternative slicing events in critical regenerative pathway genes within the chondrocyte subpopulation. The observation highlights the exclusive contribution of chondrocytes in the process of regeneration again. This finding represents a potential breakthrough in understanding the mechanisms of axolotl regeneration, potentially paving the way for future research in this field.

Fundings

This work was supported by the National Key Research and Development Program of China (2022YFA1106800), NSFC grants (T2121004, 82222044, 32271406), and Zhejiang Provincial Natural Science Foundation of China (LZ22H060002) and “Leading Goose” Science and Technology Project of Zhejiang Province (2024SSYS0026).

CRediT authorship contribution statement

Tian Qin: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Data curation, Conceptualization. **Jie Han:** Writing – review & editing, Software, Resources, Methodology, Data curation. **Chunmei Fan:** Writing – review & editing, Software, Methodology, Data curation. **Heng Sun:** Writing – original draft, Investigation. **Naveed Rauf:** Writing – review & editing, Visualization. **Tingzhang Wang:** Supervision, Methodology, Conceptualization. **Zi Yin:** Resources, Project administration, Funding acquisition, Conceptualization. **Xiao Chen:** Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

Portions of the text in this paper have been modified using AI tools to enhance the fluency of the language. The authors have reviewed and edited the content, and take responsibility for the published material.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We thanked the Core Facilities, Zhejiang University School of Medicine for their technical support.

Data access

All data needed to evaluate the conclusions in the paper are available in the GSA (CRA013464) or present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.08.014.

References

[1] Bryant DM, Johnson K, DiTommaso T, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep 2017;18(3):762–76.
[2] Stewart R, Rascon CA, Tian S, et al. Comparative RNA-seq analysis in the unsequenced axolotl: the oncogene burst highlights early gene expression in the blastema. PLoS Comput Biol 2013;9(3):e1002936.
[3] Leigh ND, Dunlap GS, Johnson K, et al. Transcriptomic landscape of the blastema niche in regenerating adult axolotl limbs at single-cell resolution. Nat Commun 2018;9.
[4] Qin T, Fan CM, Wang TZ, et al. Single-cell RNA-seq reveals novel mitochondria-related musculoskeletal cell populations during adult axolotl limb regeneration process. Cell Death Differ 2021;28(3):1110–25.
[5] Gerber T, Murawala P, Knapp D, et al. Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. Science 2018;362(6413). 421-+.
[6] Lust K, Maynard A, Gomes T, et al. Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. Science 2022;377(6610). 1061-+.
[7] Florea LD, Salzberg SL. Genome-Guided Transcriptome Assembly in the Age of Next-Generation Sequencing. Ieee Acm T Comput Bi 2013;10(5):1234–40.
[8] Heyn P, Kircher M, Dahl A, et al. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep 2014;6(2):285–92.
[9] Rhoads A, Au KF. PacBio sequencing and its applications. Genom Prote Bioinf 2015;13(5):278–89.
[10] McGinn S, Gut IG. DNA sequencing - spanning the generations. N Biotechnol 2013;30(4):366–72.
[11] Gordon SP, Tseng E, Salamov A, et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. Plos One 2015;10(7).
[12] Thomas S, Underwood JG, Tseng E, Holloway AK, Informatics BBC. Long-Read Sequencing of Chicken Transcripts and Identification of New Transcript Isoforms. Plos One 2014;9(4).
[13] Kim KE, Peluso P, Babayan P, et al. Long-read, whole-genome shotgun sequence data for five model organisms. Sci Data 2014;1.
[14] Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. Nat Biotechnol 2013;31(11):1009.
[15] Lawson ND, Li R, Shin M, et al. An improved zebrafish transcriptome annotation for sensitive and comprehensive detection of cell type-specific genes. Elife 2020;9.
[16] Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol 2013;14(7).
[17] Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 2005;21(9):1859–75.
[18] Prjibelski AD, Mikheenko A, Joglekar A, et al. Accurate isoform discovery with IsoQuant using long reads. Nat Biotechnol 2023;41(7):915–8.
[19] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37(8):907–15.
[20] Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078–9.
[21] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30(7):923–30.
[22] Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28(5):511–5.
[23] Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 2013;8(8):1494–512.
[24] Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res 2018;46(W1):W200–w204.
[25] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39(Web Server issue):W347–52.
[26] Trincado JL, Entizne JC, Hysenaj G, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol 2018;19(1):40.
[27] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10(1):1523.
[28] Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184(13):3573–87. e3529.
[29] Wang Y, Song F, Zhu J, et al. GSA: genome sequence archive<sup/>. Genom Proteom Bioinforma 2017;15(1):14–8.
[30] Chen T, Chen X, Zhang S, et al. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. Genom Proteom Bioinforma 2021;19(4):578–83.
[31] O'Grady T, Wang X. zu Bentrup KH, Baddoo M, Concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. Nucleic Acids Res 2016;44(18).
[32] Wang J, Vesterlund L, Kere J, Jiao H. Identification of Novel Transcribed Regions in Zebrafish (Danio rerio) Using RNA-Sequencing. Plos One 2016;11(7):e0160197.
[33] Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012;7(3):562–78.
[34] Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 2008;40(12):1413–5.
[35] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016;34(5):525–7.
[36] Session AM, Uno Y, Kwon T, et al. Genome evolution in the allotetraploid frog Xenopus laevis. Nature 2016;538(7625):336–43.
[37] Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 2013;41(6).
[38] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389–402.
[39] Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016;44(D1):D279–85.
[40] Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42(Database issue):D222–30.
[41] Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res 2012;22(9):1775–89.
[42] Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Gene Dev 2011;25(18):1915–27.
[43] Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 2010;20(1):110–21.
[44] Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15(8):1034–50.
[45] Kim P, Yang M, Yiya K, Zhao W, Zhou X. ExonSkipDB: functional annotation of exon skipping event in human. Nucleic Acids Res 2020;48(D1). D896-d907.
[46] Zhulyn O, Rosenblatt HD, Shokat L, et al. Evolutionarily divergent mTOR remodels translatome for tissue regeneration. Nature 2023.
[47] Caballero-Pérez J, Espinal-Centeno A, Falcon F, et al. Transcriptional landscapes of Axolotl (Ambystoma mexicanum). Dev Biol 2018;433(2):227–39.
[48] Bretaud S, Nauroy P, Malbouyres M, Ruggiero F. Fishing for collagen function: About development, regeneration and disease. Semin Cell Dev Biol 2019;89:100–8.
[49] Ferreira AM, Gentile P, Chiono V, Ciardelli G. Collagen for bone tissue regeneration. Acta Biomater 2012;8(9):3191–200.
[50] Geahchan S, Baharlouei P, Rahman A. Marine Collagen: A Promising Biomaterial for Wound Healing, Skin Anti-Aging, and Bone Regeneration. Mar Drugs 2022;20(1).
[51] Cheng XT, Huang N, Sheng ZH. Programming axonal mitochondrial maintenance and bioenergetics in neurodegeneration and regeneration. Neuron 2022;110(12):1899–923.
[52] Han Q, Xie Y, Ordaz JD, et al. Restoring cellular energetics promotes axonal regeneration and functional recovery after spinal cord injury. Cell Metab 2020;31(3):623–41. e628.
[53] Lamanilao GG, Dogan M, Patel PS, et al. Key hepatoprotective roles of mitochondria in liver regeneration. Am J Physiol Gastrointest Liver Physiol 2023;324(3). G207-g218.
[54] Sreedhar A, Aguilera-Aguirre L, Singh KK. Mitochondria in skin health, aging, and disease. Cell Death Dis 2020;11(6):444.
[55] Levoux J, Prola A, Lafuste P, et al. Platelets Facilitate the Wound-Healing Capability of Mesenchymal Stem Cells by Mitochondrial Transfer and Metabolic Reprogramming. Cell Metab 2021;33(2):283–99. e289.
[56] Xu J, Su L, Han J, et al. Rapid and quantitative in vitro analysis of mitochondrial fusion and its interplay with apoptosis. Talanta 2021;222:121523.
[57] Watson DC, Bayik D, Storevik S, et al. GAP43-dependent mitochondria transfer from astrocytes enhances glioblastoma tumorigenicity. Nat Cancer 2023;4(5):648–64.
[58] Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic Acids Res 2021;49(D1). D412-d419.
[59] Kanhoush R, Praseuth D, Perrin C, Chardard D, Vinh J, Penrad-Mobayed M. Differential RNA-binding activity of the hnRNP G protein correlated with the sex genotype in the amphibian oocyte. Nucleic Acids Res 2011;39(10):4109–21.
[60] Lazcano I, Olvera A, Pech-Pool SM, Sachs L, Buisine N, Orozco A. Differential effects of 3,5-T2 and T3 on the gill regeneration and metamorphosis of the Ambystoma mexicanum (axolotl). Front Endocrinol (Lausanne) 2023;14:1208182.

[61] Clevers H, Loh KM, Nusse R. Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. Science 2014;346 (6205):1248012.

[62] Lovely AM, Duerr TJ, Qiu Q, Galvan S, Voss SR, Monaghan JR. Wnt signaling coordinates the expression of limb patterning genes during axolotl forelimb development and regeneration. Front Cell Dev Biol 2022;10:814250.

[63] Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 2009;10(11):741–54.