Research Article

Characterization and Functional Analysis of Polyadenylation Sites in Fast and Slow Muscles

Lulu Deng, Long Li, Cheng Zou, Chengchi Fang, and Changchun Li ⓘ

Key Lab of Agriculture Animal Genetics, Breeding, and Reproduction of Ministry of Education,
College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

Correspondence should be addressed to Changchun Li; lichangchun@mail.hzau.edu.cn

Many increasing documents have proved that alternative polyadenylation (APA) events with different polyadenylation sites (PAS) contribute to posttranscriptional regulation. However, little is known about the detailed molecular features of PASs and its role in porcine fast and slow skeletal muscles through microRNAs (miRNAs) and RNA binding proteins (RBPs). In this study, we combined single-molecule real-time sequencing and Illumina RNA-seq datasets to comprehensively analyze polyadenylation in pigs. We identified a total of 10,334 PASs, of which 8734 were characterized by reference genome annotation. 32.86% of PAS-associated genes were determined to have more than one PAS. Further analysis demonstrated that tissue-specific PASs between fast and slow muscles were enriched in skeletal muscle development pathways. In addition, we obtained 1407 target genes regulated by APA events through potential binding 69 miRNAs and 28 RBPs in variable 3′ UTR regions and some are involved in myofiber transformation. Furthermore, the de novo motif search confirmed that the most common usage of canonical motif AAUAAA and three types of PASs may be related to the strength of motifs. In summary, our results provide a useful annotation of PASs for pig transcriptome and suggest that APA may serve as a role in fast and slow muscle development under the regulation of miRNAs and RBPs.

1. Introduction

Polyadenylation involving in the cleavage of 3′ signaling region of precursor mRNA (pre-mRNA) and the addition of a poly(A) tail is the final core step of mRNA maturation [1, 2]. The upstream and downstream cis elements constantly regulate the cleavage of polyadenylation sites (PASs). These elements mainly consist of an upstream canonical poly(A) signal AAUAAA and a less defined downstream U/GU-rich region (typically GUGU) [3, 4]. Alternative polyadenylation (APA), a phenomenon that the same RNA molecule with distinct 3′ terminal, is widespread in eukaryotes as an important posttranscriptional regulatory mechanism. More than 70% of human genes produced mRNA isoforms by APA and about half in Arabidopsis, *Caenorhabditis elegans*, and zebrafish [5–8]. APA events can be classified roughly into two categories. In one case, PASs located in internal exons or introns named coding region-APA (CR-APA)

affect the length of coding regions. In another case, PASs located in the 3′ untranslated region (3′ UTR) (UTR-APA) result in the production of transcripts with diverse 3′ UTRs [9].

The development of skeletal muscles after being born is characterized by hypertrophy and type transformation of muscle fibers. Factors such as exercise, nutrition, and aging can lead to the transformation of four muscle fibers (I, IIA, IIX, and IIB) [10–13]. The fast and slow muscles contain four different muscle fibers in different proportions, resulting in differences in morphological, physiological, and biochemical characteristics [14, 15]. APA events can regulate muscle fiber transformation of fast and slow muscles and have gained more attention in recent years. Ocular pharyngeal muscular dystrophy (OPMD) is mainly expressed in myasthenia and amyotrophy of the eye and pharyngeal muscles [16]. de Klerk et al. found that an effect of poly(A) binding protein nuclear 1 (PABPN1) on APA may partly explain muscle

weakness in OPMD [17]. Subsequently, a further study observed that the distal PAS utilization rates of six genes (Atrogin-1, Arih2, Psme3, Psmd14, Rad23a, and Atg12) were significantly downregulated in PABPN1-DR mouse model. Reduced PABPN1 levels may cause Atrogin-1 dysregulation and also gave rise to myofiber atrophy, ECM thickening, and myofiber transitions [18]. In addition, the loss or gain of PASs in AMPKα1KO cells and AMPKα1KO mice resulted in changes in muscle area and enrichment of five important differentially expressed genes (Car3, Mylk4, Neb, Obscn, and Pfkm) and further regulated muscle development [19]. For example, Nebulin is a large actin-binding protein and the main cause of nemaline myopathy (NEM) [20].

Next-generation sequencing (NGS) technology has been widely used for transcriptome research on regulatory mechanisms and differential gene expression [21]. However, the length of sequencing reads limits the reconstruction of full-length splice variants [22]. And it is difficult to separate genuine PASs from internal priming events [23–25]. Based on single-molecule real-time (SMRT) sequencing, transcripts derived from PacBio platform without assembly are longer and even full length, which provide direct evidence for the annotation of isoforms of each gene. Especially, the Iso-seq library built with the oligo(dT) primer ensures the integrity of 3′ terminal. So far, the PacBio protocol Iso-Seq™ has been successfully applied to characterize the posttranscriptional APA events in various eukaryotes [26–30]. Our study used the PacBio dataset to annotate PASs of pigs and analyzed differential usages and functions of APA genes under the regulation of miRNAs and RBPs between fast and slow muscles. We also discussed the distribution of cis elements surrounding all identified PASs. Thus, this study updated pig genome annotation about PASs and supplied valuable resources for further transcriptome research.

2. Materials and Methods

2.1. Ethics Approval and Consent to Participate. The experimental white pigs were provided by the National Livestock Engineering Research Center of Huazhong Agricultural University. Animal breeding and slaughtering were carried out in accordance with the preapproval guidelines of the Standing Committee of the Hubei Provincial People's Congress No. 5. And all the experimental schemes were approved by the Scientific Ethics Committee of Huazhong Agricultural University.

2.2. Data Generation. The PacBio datasets in this paper were from our group to collect 38 tissues of Large White, including one organ (a 26-day-old whole embryo), 20 tissues (heart, kidney, tongue, stomach, liver, spleen, lung, longissimus muscle, duodenum, cecum, inguinal lymph nodes, precaval vein blood, inner ear, back subcutaneous fat (BSF), ovary, psoas muscle, endometrium (EN), corpus luteum, soleus muscle (SM), and extensor digitorum longus (EDL)) from an adult sow (210 days of age), and 17 tissues (heart,

tongue, stomach, uterus, liver, spleen, lung, longissimus muscle, duodenum, inguinal lymph nodes, precaval vein blood, BSF, psoas muscle, EN, SM, thymus, and EDL) from a one-day-old sow [31]. The total RNA of each tissue was extracted and pooled together.

Besides, RNA-seq datasets were also derived from data generated by paired-end sequencing of 8 tissues (BSF, SM, EDL, and EN from adult and one-day-old sows) on the illumine HiSeq2500 platform among 38 tissues described above [31].

2.3. Identification of PASs. By annotating loci and isoform, and removing redundant and false-positive gene structures, we constructed a gene structure annotation for FLNC reads. The specific method was performed as previously described [31]. Our Iso-Seq libraries were constructed using an oligo dT primer, and 3′ end of each isoform was retained. We used Transcriptome Analysis Pipeline for Isoform Sequencing (TAPIS) pipeline according to the default options to identify PASs. The site with maximum read depth was considered as a genuine PAS if the distance between multiple candidate sites was within 15 nt.

2.4. Motif Analysis. To assess the sequence composition surrounding cleavage sites, we calculated the proportion of each nucleotide at each position within the ±100 nt window. The upstream and downstream of PASs usually contain canonical recognition signals and other hexamer variants. We performed de novo motif searches using Signal Sleuth (v2) program with Motif Overlapping mode for hexamers that appeared in a 100 nt window upstream and downstream and a 50-nucleotide window upstream of identified PASs [32]. In addition, based on the single or multiple ends of PASs in the 3′ UTR region, these PASs were classified into three types: single end, single-end PASs that found in genes with single terminal 3′ UTR isoform; distal, distal PASs regarded as the last PASs of terminal 3′ UTR that undergo APA events; and proximal, proximal PASs defined as all PASs of terminal 3′ UTR that undergo APA events except the distal one. Nucleotide distributions of these three types were calculated and shown by Weblogo3 (v3.6) with the default color scheme [33]. Then, the Find Individual Motif Occurrences (FIMO, v5.0.2) method with p value <0.001 of MEME suite was utilized to calculate the frequency of the canonical polyadenylation signal AAUAAA and 12 most common variants (AUUAAA, AAUAUA, UAUAAA, AGUAAA, AAUACA, UUUAAA, AAAACA, CAUAAA, AAGAAA, GAUAAA, AAAAG, and AAUAGA) reported in the upstream 50-nucleotide window of three types of PASs [34, 35].

2.5. Tissue-Specific Analysis. After filtering out of adaptor sequences and discarding low-quality reads in RNA-seq raw data using Trimmomatic (v0.36), clean reads were mapped to the reference genome (ftp://ftp.ensembl.org/pub/release-94/fasta/sus_scrofa/dna/) of Sus scrofa with Hisat2 (v2.0.3) [36, 37]. Then, we used StringTie (v1.3.4) to reconstruct the

transcriptome for each sample on the basis of the structure annotation file above, and then all transcriptomes were merged into one nonredundant transcriptome using StringTie merge function [38]. We finally obtained PASs located in each sample by comparing positions of assembled transcripts with all identified PASs. The average length of $3'$ terminal in each tissue was replaced by the average of $3'$ UTR lengths of all genes. Moreover, the number of reads mapping to PAS-associated genes was counted using the htseq-count program (v0.6.1) and normalized as the gene expression [39]. Default parameters were used in all abovementioned programs. Because the lack of expression in PacBio transcriptome sequencing limits our quantitative and differential analysis of PASs coexpressed in each group, we define differential PASs as PASs expressed only in one group and not in other groups.

2.6. Prediction of Binding Sites about miRNAs and RBPs.
PASs located in genes with annotated START and STOP codons were selected from all $3'$ UTR PASs. In case where multiple STOP codons appeared, we determined coordinates of the codons that minimized the $3'$ UTR length as the only STOP codons. Extended $3'$ terminals of isoforms of PAS-associated genes were globally searched for the targeted binding sites of reported 457 miRNA mature seed sequences and 64 RBPs motif using Analysis of Motif Enrichment (AME, v5.0.3) method of the MEME suite [40]. Then, we selected differential miRNA and RBP binding sites based on APA genes. The motifs of these RBPs were obtained from the CisBP-RNA database (http://cisbp-rna.ccbr.utoronto.ca/bulk.php), and the seed sequences of the miRNAs were downloaded from the miRBase database (http://www.mirbase.org/cgi-bin/browse.pl).

2.7. Gene Ontology and Pathway Analysis.
We performed Database for Annotation, Visualization, and Integrated Discovery (DAVID) analysis to categorize the considerably enriched ($p < 0.05$) biological processes or pathways to assess the properties of genes with differential expression of PASs or variable presence of miRNAs and RBPs [41]. Because of the limitation of the Sus scrofa genome annotation in the DAVID database, we converted all genes into human homologous genes by BIOMART from Ensembl.

2.8. Validation of Variable $3'$ End by $3'$ RACE.
We randomly selected 6 APA genes in porcine skeletal muscle for experimental validation as described in other articles [28, 30]. The cDNA was prepared by reverse transcription of SMART Scribe TM Reverse Transcriptase (100 U). Next, 2.5μL of cDNA was amplified using the cDNA ends with gene-specific primers (GSPs) and Seq Amp DNA Polymerase. All GSPs for validation are listed in Table S1.

3. Results

3.1. Identification and Annotation of PASs.
Our research group has established an Iso-Seq database upon 38 Large White pig tissues [31]. Based on that, we used the TAPIS pipeline to identify PASs [30]. A total of 10,334 PASs were obtained, of which 5 identified on chromosome Y were removed. And other PASs were located in 6668 genes. Each gene was supported by an average of 37 poly(A) reads. Comparing with the Ensembl reference annotation, 84.52% of PASs were aligned to the annotated genes, and the remaining 1600 PASs were mapped to the intergenic regions (Figure 1(b) and Table S2), which might represent unknown genes or novel distal sites in the downstream $3'$ UTR regions of the known genes. These PASs which were located in annotated genes mainly corresponded to two modes based on the annotated genomic features: $3'$ most exon and upstream region (Figure 1(a)). As expected, major (83.52%, 7295) PASs were in the type of $3'$ UTR regions, and a small part of PASs was in introns (9.61%, 839), CDS (1.59%, 139), and $5'$ UTR regions (0.29%, 25). Considering that PASs in genomic regions only belonged to a small proportion of the overall polyadenylation events, we mainly concentrated on PASs that were located in $3'$ most exon. However, we found that a portion of genes might be misannotated for no match with our PacBio isoforms. In the first scenario, some adjacent annotated genes in reference annotation overlapped only a single gene in our structural annotation (Table S2), and another scenario was contrary to the first (Table S2). Thus, we removed PASs in those misannotated genes and some other PASs (138) located in overlapping genes in the following analysis.

3.2. Distribution of PASs in the Pig Genome.
All identified PASs were used to analyze the distribution of PASs in genes. Among 6668 PAS-associated genes, 2191 (32.86%) contained two or more polyadenylation events (APA) which resulted in variable $3'$ terminal (Figure 2(a)), and with an average of 1.55 PASs per gene. The distances between adjacent sites in each APA gene and their $3'$ UTR regions were calculated. We observed that most of them were relatively near in both cases, but presented two different modes obviously (Figure 2(b)). More than 75% and 69% of APA were shorter than 1 kb (Figure 2(b)). Then, our further analysis showed that the first two peaks of gene regions were similar to those of $3'$ UTR with a higher ratio, but the last peak that was mainly caused by the location of more PASs in the intron was only found in the gene regions (~15 kb) (Figure 2(c)). Then, we plotted the average position of PASs in $3'$ UTR regions according to the stop codon for numbers of PASs. Although the position distribution of PASs showed great changes, the distance at the same position indicated a slightly shortened trend as numbers of PASs in $3'$ UTR genes are increasing (Figure 2(d)). Finally, the length of APA genes was discovered to be related to the number of PASs as shown in Figure 2(e), which was consistent with previous studies in moso bamboo [28].

3.3. The Usage of Upstream and Downstream Elements.
It has been well documented that PASs were involved in the core polyadenylation upstream signal and the U/GU-rich downstream sequence element [3, 7, 35]. We studied the
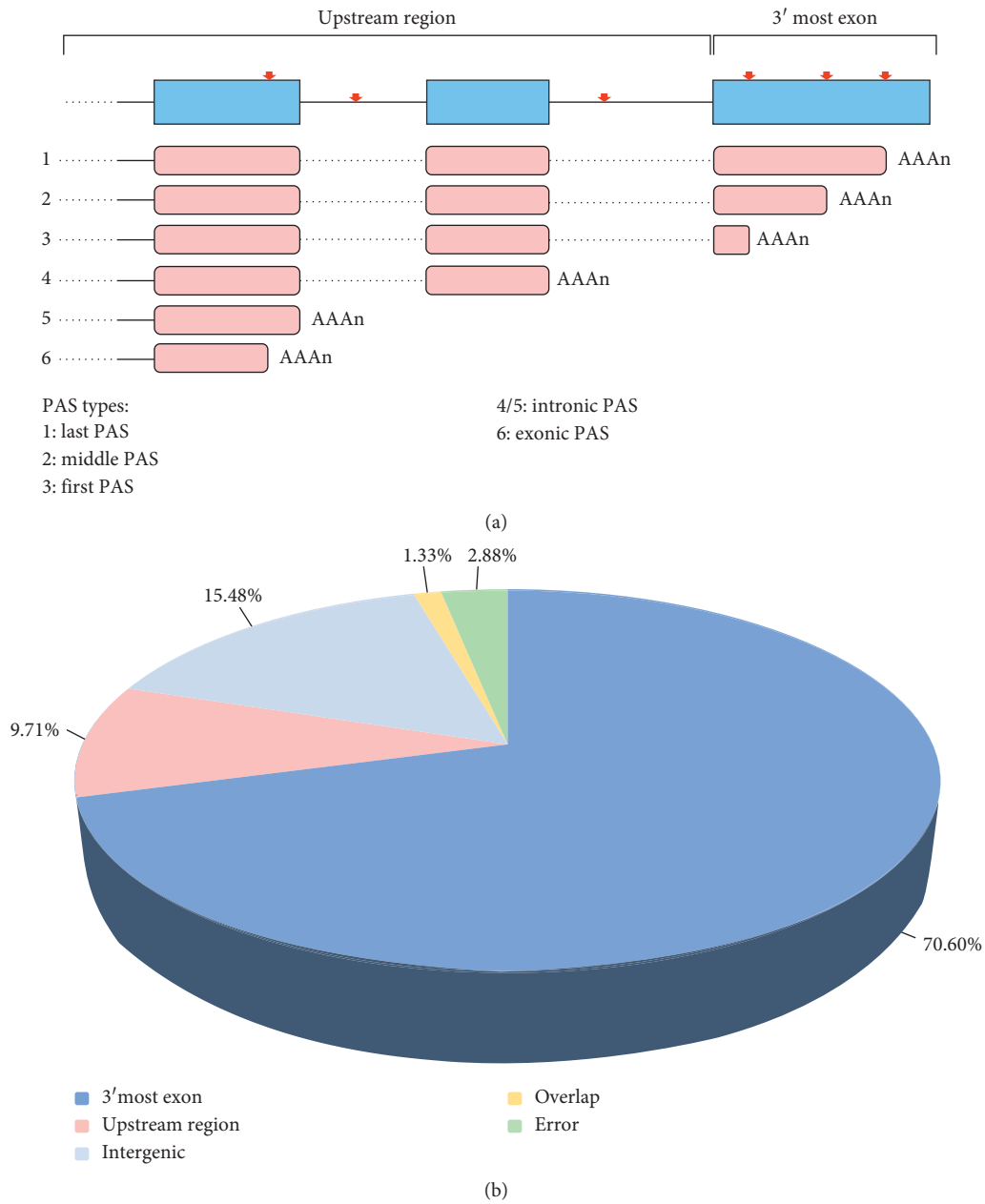
(a)

(b)

Figure 1: Schematic about PAS types. (a) Two PAS types across genomic regions including PASs in upstream region and 3′ most exon. The type number corresponded to isoform numbers shown in the figure. Blue and pink boxes, exons. Solid lines, introns. Red arrows, cleavage. Dashed lines, splicing. (b) Annotation and distribution of PASs from Ensembl. Intergenic, PASs in intergenic. Overlap, PASs in overlap genes. Error, PASs located in misannotated gene.

sequence composition surrounding all PASs and observed the distribution of these two elements in our data (Figure 3(a)). Then, we performed de novo motif searches by using the ±100 nt sequences upstream and downstream of all PASs and detected a strong enrichment of canonical motif AAUAAA and other variant signals peaking at 20 nt position (Figure 3(b)). Furthermore, the statistical results of the top ten motifs in the 50 nt window upstream of these PASs are shown in Figure 3(c). The canonical motif and variants accounted for the majority as expected, while we observed unusual A and U enriched hexamer.

The usage of upstream elements usually affects the cleavage and polyadenylation. Although we found a similar nucleotide distribution among the three types of PASs, AU enrichment level upstream of the proximal sites was relatively lower compared to the other two (Figure 3(d)). We further calculated the canonical signal and other 12 most common variants [35] and observed that AAUAAA accounted for a large proportion within single-end PASs (Figure 3(e)). Amongst PASs that underwent APA events, we also found a similar but lower enrichment level of canonical signal in the distal PAS (Fisher's exact test, 45%,
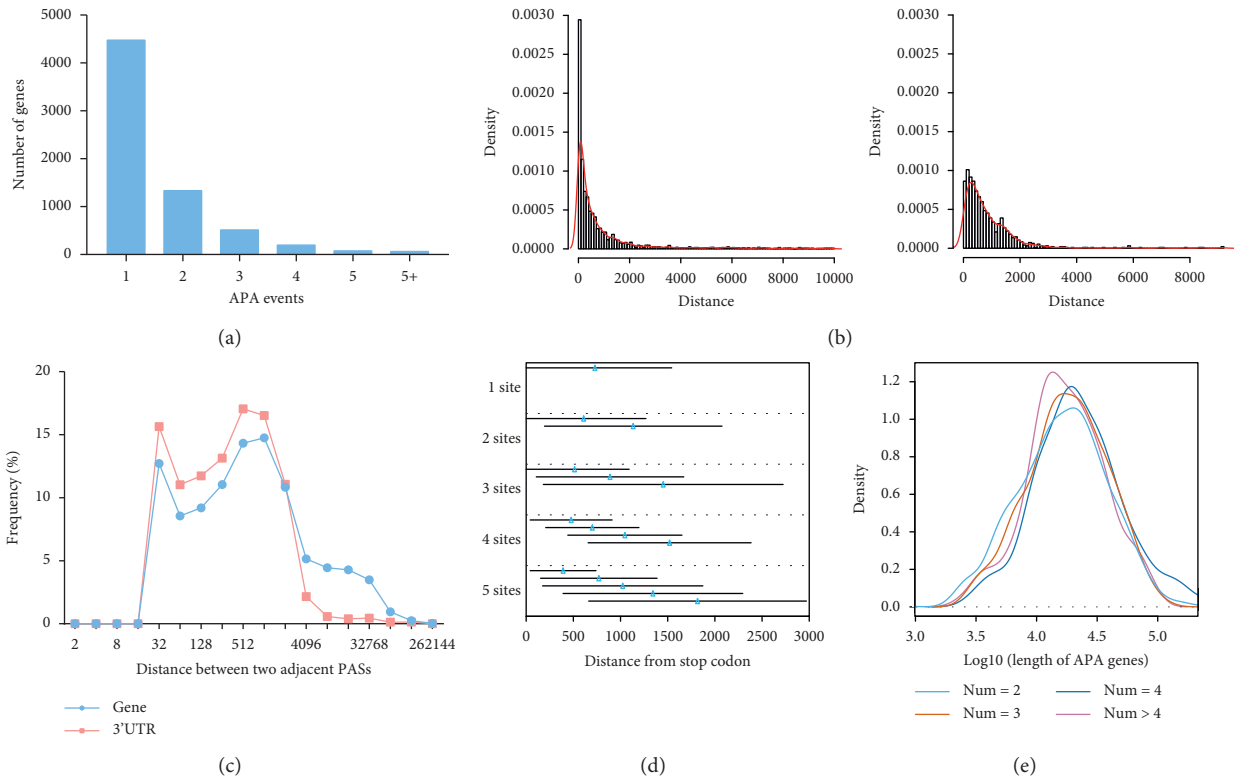
(a)

(b)

(c)

(d)

(e)

Figure 2: Distribution of PASs in APA genes and 3′ UTR regions. (a) Number of genes containing one or more PASs. (b) Histogram of the genomic distances between two adjacent PASs in APA genes and 3′ UTR regions. (c) Frequency of the genomic distances between two adjacent PASs in APA genes (pink) and 3′ UTR regions (powder blue). (d) Blue triangle, the average distance between the PAS and corresponding stop exon in 3′ UTR regions. Black horizontal solid line, standard deviation of the average distance. (e) Density of length for APA genes with different number of poly(A) sites.
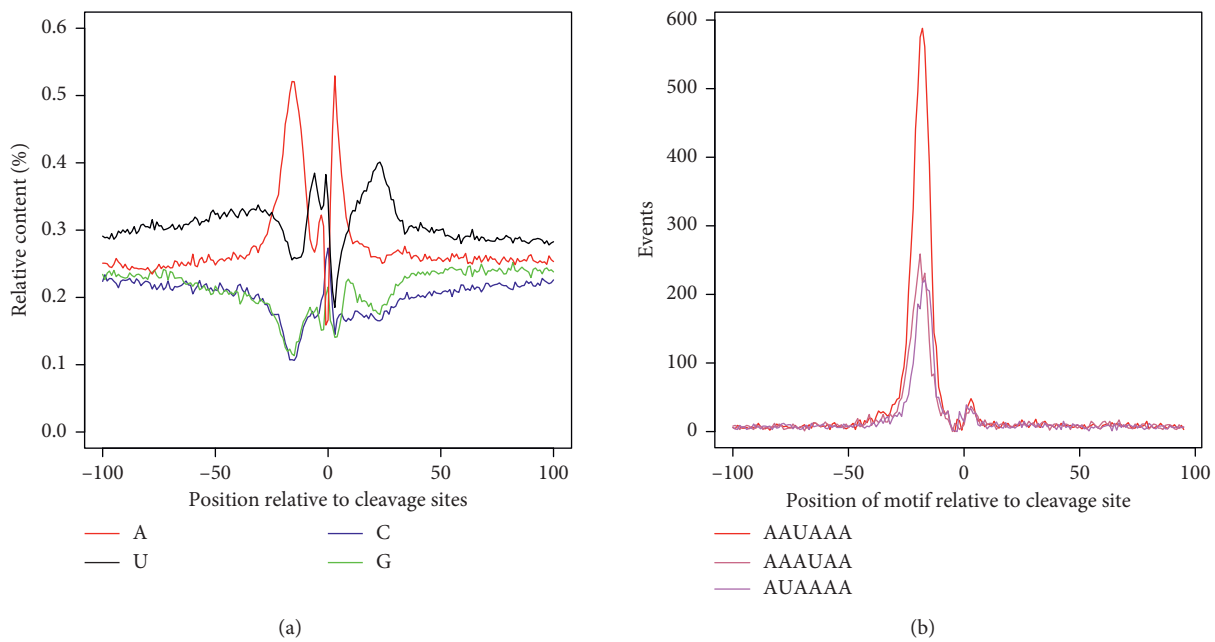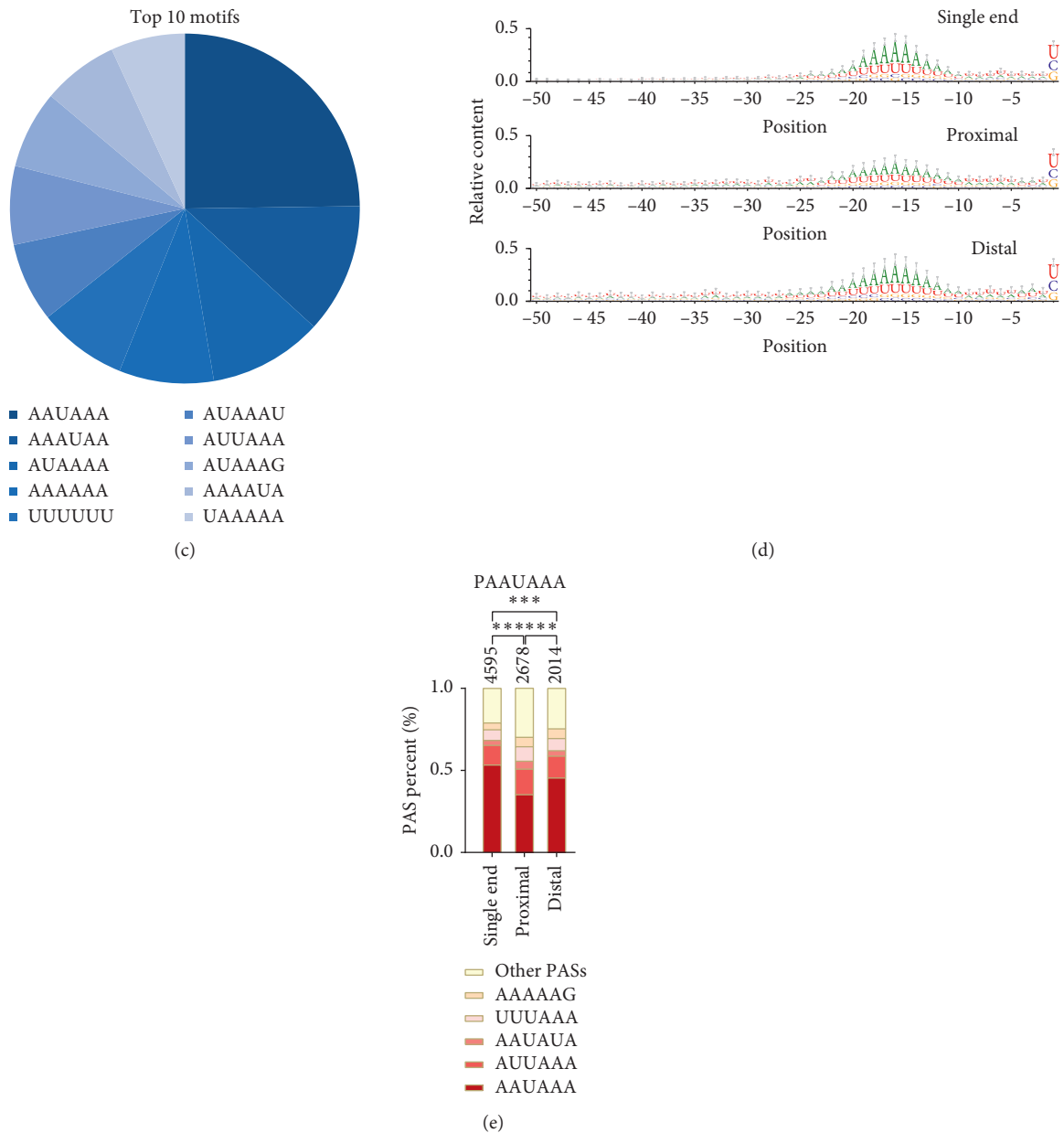

(a)

(b)

Figure 3: Continued.

(c)

(d)

(e)

FIGURE 3: Characteristics of elements and properties of different types of PASs. (a) Plotted the nucleotide distribution surrounding all identified PASs in a ±100 nt window. (b) Enrichment of the top three hexamers surrounding all identified PASs in a ±100 nt window. (c) The top 10 hexamers surrounding all identified PASs in a 50 nt window upstream. (d) Plotted the nucleotide distribution surrounding three types (as described in the Method section) of PASs in a 50 nt window upstream. The vertical axis showed the ratio of nucleotides. (e) Percentages of the top 10 hexamers in the 50 nt window upstream of different PASs.

$p = 3.59E − 09$). However, compared with distal PASs, proximal PASs obviously carried a significantly higher proportion of other weak variants with lower canonical signal (35%, $p = 1.92E − 12$).

3.4. Tissue Specificity Analysis of PASs. The tissue specificity of PASs has been confirmed in other organisms [8, 42, 43]. We analyzed PASs in four tissues at two ages by using RNA-seq datasets and the new modified structure annotation file. 6993, 7131, 7025, and 7319 PASs have be found in SM, EDL,

SBF, and EN of 1 d and 7046, 7006, 7042, and 7210 PASs in adult tissues, respectively (Figure 4(a) and Table S3). A total of 6423 and 6108 PASs had significant hits in both groups. And EN occupied the largest number of specific PASs that uniquely expressed. Moreover, we performed a further comparative analysis of tissue-specific PASs between SM and EDL in 1 d and adult periods (366 and 278), which were often used as representative muscles dominated by slow-twitch (type I) and fast-twitch (type II) fibers in animal experiments[44]. Gene ontology analysis revealed that these genes containing tissue-specific PASs ($p < 0.05$) were mainly
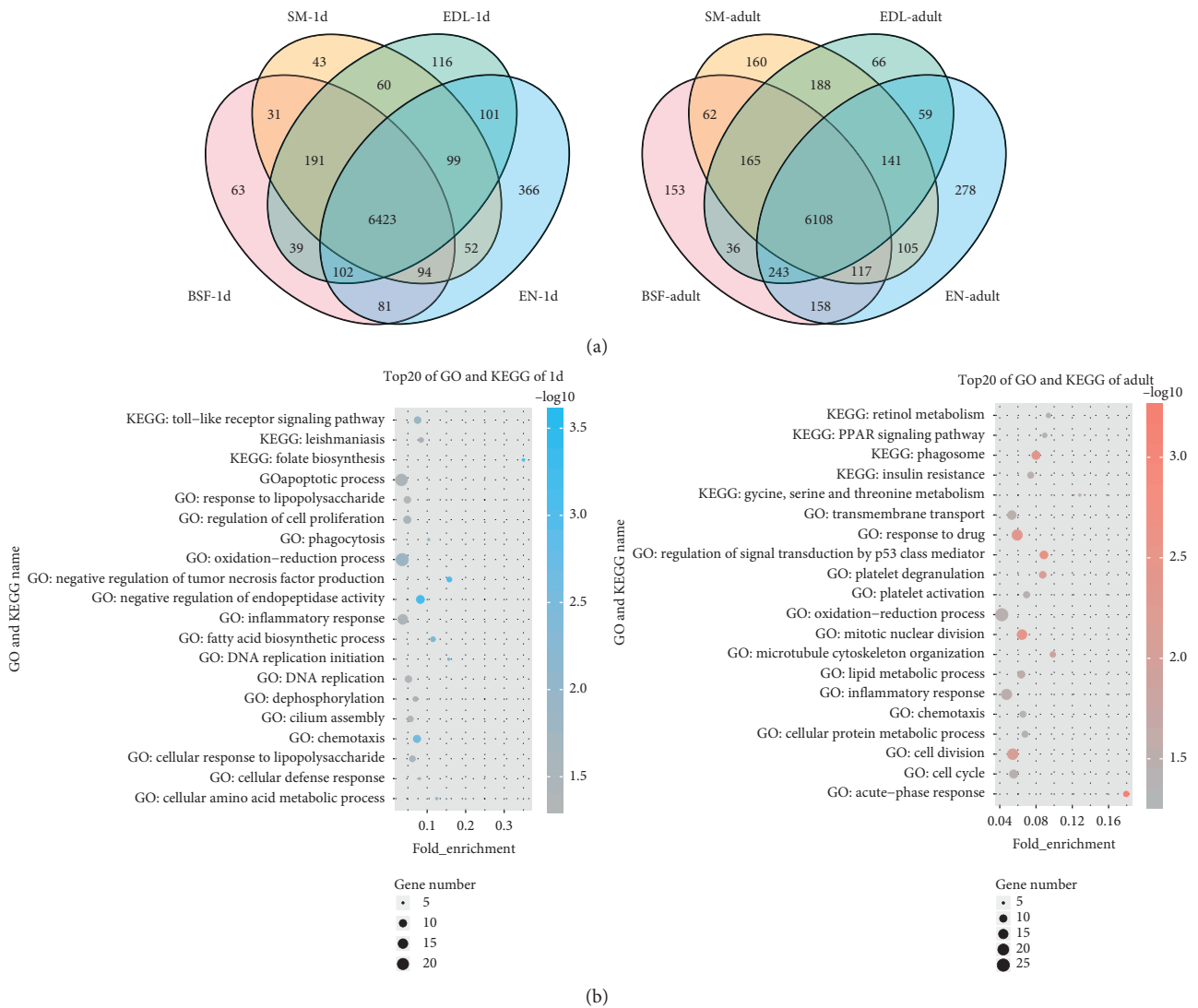
(a)

(b)

Figure 4: The usage of tissue-specific PASs in SM and EDL. (a) The number represented tissue-specific PASs identified in BSF, SM, EDL, and EN of 1 d and adult. (b) List of first 20 pathways and biological processes about the different usages of PAS-associated genes in two periods.

involved in oxidation-reduction process, cell division, response to drug, apoptotic process, and inflammatory response in terms of molecular function (Figure 4(b)). Moreover, we also found that part of these specific PAS-associated genes ($p < 0.05$) was significantly enriched in the pathways associated with energy metabolism, such as insulin resistance, retinol metabolism, glycine, serine and threonine metabolism, Toll-like receptor signaling pathway, leishmaniasis, and folate biosynthesis (Figure 4(b)). In the above four muscle tissues, the number of PASs and gene expression were only analyzed as weak positive correlation (Figure S1).

3.5. Annotation of miRNAs and RBPs. To detect possible RBPs and miRNAs involved in the regulation of gene expression and interacted with APA events, targeted binding sites for RBPs and miRNAs were searched within the 3′ UTR region. We downloaded 408 miRNA mature seed sequences from the miRBase database and 64 RBPs from CisBP-RNA

database. We used 3′ UTR regions of all identified PAS-related genes to comprehensively annotate the binding sites of RBPs and miRNAs and have predicted more than 80,000 events which are listed in Table S4. Based on APA for each target gene, we finally obtained 1149 3′ UTR regions that selectively bind to 69 miRNAs (Figure 5(a) and Table S5). In addition, 28 RBPs positioned in 1429 3′ UTR regions with varied lengths (Figure 5(b), Table S5). For the sake of further understanding the functions and associated pathways of these miRNAs and RBPs, we performed DAVID analysis by running queries against the DAVID database. Enrichment analysis results revealed that 1407 genes significantly ($p < 0.05$) participated in 133 biological processes (Figure 5(c)), including oxidation-reduction process, apoptotic process, protein transport, and mRNA splicing via spliceosome. 29 pathways were influenced by APA gene regulated by microRNA and RBP, respectively (Figure 5(c)). Some of these pathways and biological processes were muscle-development-related, such as muscle filament
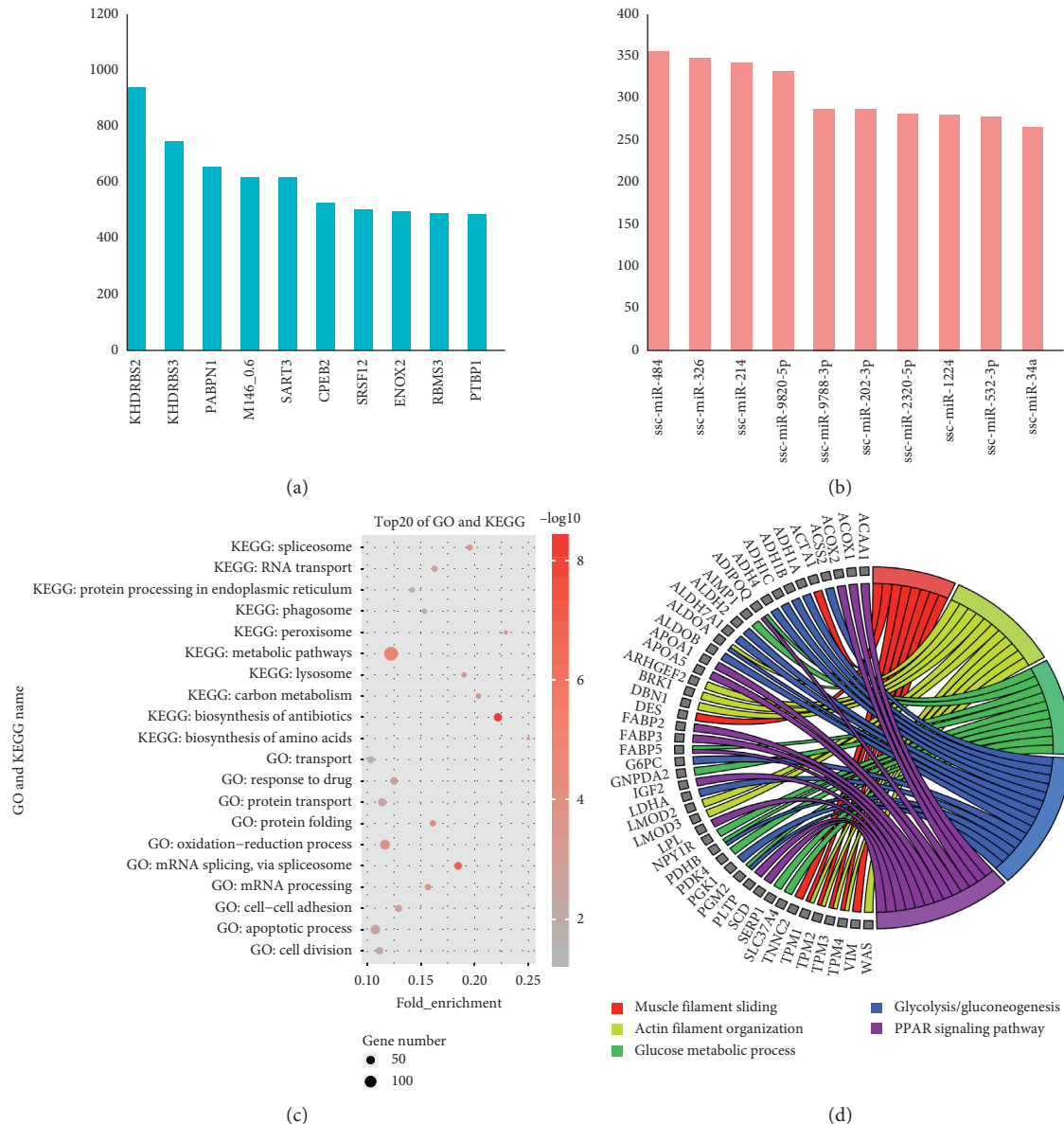
(a)

(b)

(c)

(d)

FIGURE 5: Effects of microRNAs and RNA binding proteins on APA genes. (a, b) The top 10 microRNAs and RNA binding proteins that differentially bound to the 3′ UTR ends. (c) List of first 20 pathways and biological processes about target genes of differentially microRNAs and RNA binding proteins. (d) List of 5 pathways and biological processes associated with skeletal muscle fiber transformation.

sliding, actin filament organization, glycolysis/gluconeogenesis, and PPAR signaling pathway (Figure 5(d)). Finally, we took the simultaneous existence of genes that are associated with tissue-specific PASs above and the predicted differential target genes of miRNA and RBPs. Some shared genes were discovered to be involved in muscle fiber transformation, such as HDAC (47–49), TRIB3 (50), and GYS1 [45].

3.6. Validation of Variable 3′ End by 3′ RACE. In order to verify the accuracy of the identification and analysis of PASs, we randomly selected 6 APA genes (all PASs identified in each gene are expressed in skeletal muscle) and validated using 3′ rapid amplification of cDNA ends (3′ RACE). The

results showed that the positions on the genome of all six genes were consistent with the band of 3′ RACE (Figure 6). Additional APA events were observed in the remaining two genes (ENSSSCG00000005021 and ENSSSCG00000011961), marked in red. The result of 3′ RACE can basically verify the accuracy of PASs identification and also implied that the current annotation was incomplete.

4. Discussion

Since the role of posttranscriptional modification has gradually gained attention, more and more studies have also focused on the polyadenylation of single gene in pig [46–48]. Many methods based on NGS technology currently emerged as the times require, including poly(A) site sequencing
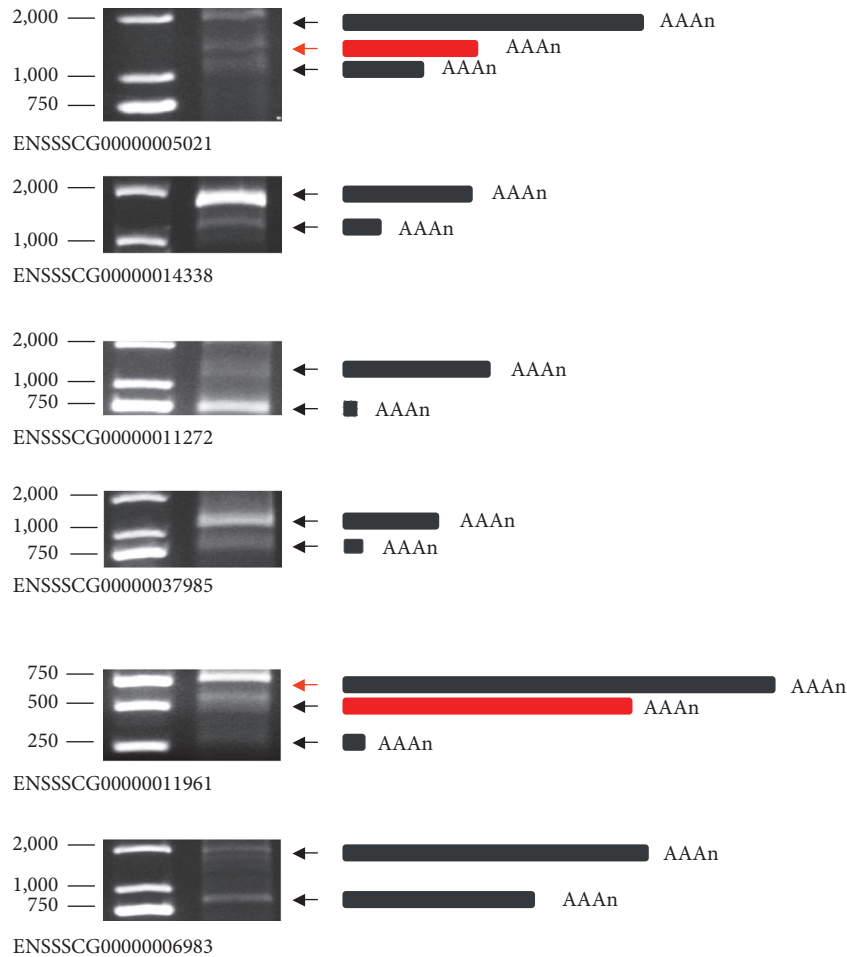
FIGURE 6: Validation of variable 3′ end by 3′ RACE. Upstream regions, solid lines. 3′ UTR, charcoal grey boxes. Unidentified PAS, red boxes.

(PAS-seq), direct RNA sequencing (DRS), poly(A)-test RNA-sequencing (PAT-seq), and whole-transcriptome termini site sequencing (WTTS-seq) [24, 49–51]. However, there are only limited articles comprehensively annotating porcine polyadenylation with RNA-seq datasets [43]. In general, short-read RNA sequencing techniques have the fact that alignments allow us to infer, but have limitations for directly understanding the complexity of potential isoforms [22]. This study used PacBio sequencing data from 38 pooled tissues of Large White sow to identify PASs. A more comprehensive atlas of porcine PASs without assembly was established using the reported TAPIS pipeline [30]. We obtained a total of 10339 PASs which correspond to 2191 APA genes. A large part of them were distributed in the 3′ UTR regions. Although the full-length transcriptome sequencing can guarantee the reliability of the identification results, due to the insufficient sequencing depth, number of samples, and funding in this study, the quality of the PacBio dataset has limited the identification of PASs to a certain extent. Further research may need to improve the sequencing depth and analysis methods [52].

After comparing the identified PASs to the Ensembl reference annotation, we not only found that a large number of PASs are located in the intergenic regions, but also

confirmed the existence of partial misannotated genes, which reflected the incompleteness of the current annotation. Single-molecule long-read sequencing technology is sufficient for the complete assembly of most known microbial genomes and has been applied to research studies of many species by now [26, 27, 30]. With the PacBio SMRT sequencing technology employment to whole-transcriptome profiling in rabbit, Chen et al. obtained 36,186 high-confidence transcripts and detected 24,797 alternative splicing (AS) and 11,184 APA events, indicating a significant improvement in the current annotation of the rabbit [26]. However, the inability of three-generation sequencing data to quantify genes makes it difficult for in-depth transcriptome study. Therefore, it is wise to combine the advantages of both sequencing methods. Gao et al. have developed a one-stop solution PRAPI that was used to perform differential expression analysis of APA through uniting PAS-seq and Iso-Seq libraries [53].

We obtained a total of 839 PASs that were mapped to the 700 intron regions of the reference genome and resulted in frame shift and subsequently produced shorter proteins. Previous studies have shown the prevalence of intron polyadenylation (IPA) in other biological processes. Singh et al. performed a comprehensive IPA analysis using 46 3′-seq

and RNA-seq immune cells and identified 4927 high-reliability IPA events [54]. In this study, we found some intron PASs in TPM1, TPM2, and TPM3 (members of the tropomyosin family). Research showed the high positive correlation between TPM3 MyHC-slow and suggested that both TPM1 and TPM3 were expressed with MyHC-2a [55]. Further studies have shown that the synergistic expression of TPM and MyHC is an important factor in determining the contractile properties of myofibers [56]. Besides, Daloii et al. show alterations in IPA utilization in OPMD mouse and cell models. IPA may have potential regulatory mechanisms for muscle fibers development and requires further investigation.

The comparative analysis of PASs in the fast and slow muscles confirmed the existence of tissue specificity again and indicated a potential APA-dependent regulatory function. When genes with tissue-specific APA sites were functionally enriched, we found that part of genes involving in PPAR signaling pathway and insulin resistance were related to the development of fast and slow muscles. In tissue-specific analysis result, genes encoding type II HDACs (histone deacetylases) were identified with loss of the distal PAS in the slow muscle of both periods, while genes encoding type I HDACs have lost the only identified PAS in the fast muscle of newborn pigs. Previous studies showed that inhibition of MEF2 activity by class II HDAC proteins led to suppression of chronic convulsions and oxidative muscle fiber formation [57]. Galmozzi et al. explored the functional mechanism of type I HDACs and found the pivotal role of class I HDAC activity in enhancing skeletal muscle oxidative metabolism and promoting energy expenditure by increasing the expression level of PGC-1α [58]. Another study pinpointed that class I HDACs act as key regulators of FoxO-induced muscle atrophy during skeletal muscle disuse [59]. Our analysis also showed that the distal PAS of TRIB3 was absent in adult fast muscle. Tribbles homolog 3 (TRIB3) encodes a protein kinase of the same name and plays an important role in a variety of cellular processes, including growth and differentiation. Overexpression of TRIB3 was found to reduce PPAR-α activity and also increase the expression of miR499/miR208b, which led to transformation toward a slow fiber type [60]. Liu et al. revealed a downstream mechanism for muscle fiber transformation, namely, that the myosin Myh7b-encoded miR-499 activated AMPK-PGC-1α signaling by direct inhibition to FNIP1 and thereby triggered a muscle mitochondrial oxidative metabolism program [61]. Moreover, the analysis demonstrated that PAS of CREB (cAMP response element-binding) only expressed in the adult fast muscle in our study. The CREB formed a complex with FLH3 to promote expression of a fast fiber gene and inhibited the slow fiber gene by reducing MyoD transcriptional activity [62]. On a patient with muscle-specific glycogen synthase deficiency, type I fibers predominated, glycogen deficiency and increased mitochondria, and identified homozygosity for a 2 bp deletion in the GYS1 gene [45]. The above results indicated that APA events may have an effect on muscle fiber transformation in fast and slow muscle tissues. However, some studies have shown that the difference in gene expression

between adult fast and slow muscles came from the differential expression of lncRNA, not APA or AS [63]. Therefore, systematic research on fast and slow muscle development is urgent to unify different opinions based on a combination of data analysis and experimental verification of different species and developmental stages.

Since miRNA and RBP binding sites are generally located in 3′ UTRs, UTR-APA can enable isoforms to evade miRNA regulation and have significant biological effects on their translation efficiencies, stability, and localization [9, 64]. In this study, the target genes of miRNAs and RBPs were predicted systematically in the variable 3′ UTR regions, hoping to further improve the porcine conserved site annotations. Genes that differentially carry miRNAs and RBPs are enriched in muscle filament slip, actin filament tissue, PPAR signaling pathways, and glycolysis/gluconeogenesis. Interestingly, the tissue-specific genes GYS1 and TRIB3 abovementioned also combined differentially expressed miRNAs and RBPs in adult pig muscle, while HDAC2 appeared in newborn one. We predicted 7 miRNA binding sites along the altered 3′ UTR region on TRIB3, including miR-532-3p/miR-885-3p. The expression of miR-532-3p proposed to target Slc2a4 was reduced in soleus muscle from diabetic rats [65]. Massively parallel sequencing analysis of Japanese black cattle showed that miR-885 was only expressed in the semitendinosus (slow type) [66]. Among the predicted RBPs and PABPN1 proteins are known cleavage and polyadenylation associated (CPA) factors [67]. We predicted a differential binding site for PABPN1 in 3′ UTR region of RAD23A (encode UV excision repair protein RAD23 homolog A). Decreased levels of PABPN1 resulted in muscle atrophy and ECM thickening [18].

As expected, the surrounding nucleotide distribution of all identified sites was consistent with other species observations, and the AAUAAA hexamer was also regarded as the most canonical PAS in Sus scrofa within the de novo motif observation [7, 68]. In addition, we have previously observed that the proximal and distal sites in the fast and slow muscles have different usages, so we attempted to globally explore the relationship between the proximal-distal PASs and the 10 most common recognition signals in humans. The results of the study showed that the distal PAS consistent with the single-end observation was associated with a strong classical recognition signal AAUAAA, while the proximal PASs clearly carried a higher proportion of weaker variants. This may be one of the reasons for different choices of PASs in fast and slow muscles. We need to more accurately understand the molecular basis of this recognition difference by combining the protein factors directed to cleavage sites [1, 9].

5. Conclusions

In conclusion, our study improved and complemented the annotation of PASs and APA events in pigs by long-read sequencing. And we also found some regular patterns of recognition signals of PASs in pigs. Combining RNA-seq data for the analysis about tissue specificity and miRNAs and RBPs of APA, our results revealed that the selectivity of

proximal and distal PASs possibly had potential regulation mechanism of skeletal muscle development and fiber transformation. These results not only indicated the universality of porcine polyadenylation events, provided a direction for the further study of posttranscriptional modification, and also had the positive significance for the study of fast and slow muscles in pigs.

Data Availability

The sequence datasets of this paper have been deposited in the Genome Sequence Archive (GSA; http://gsa.big.ac.cn/) of Beijing Institute of Genomics, Chinese Academy of Sciences, with accession number PRJCA000349.

Conflicts of Interest

The authors declare that they have conflicts of interest.

Authors' Contributions

CCL conceived the entire analysis process. LLD analyzed the main content of the data. LL completed the experimental verification section. LLD wrote this paper under the help of CCL, CZ, and CCF. All authors have read and approved the manuscript.

Acknowledgments

Supplementary Materials

Supplementary file 1: Table S1: this provided the sequence of gene-specific primers for 3′RACE. Supplementary file 2: Table S2: a total of 10,334 PASs were identified according to PacBio FLNC. This provided genomic features of PASs, including located chromosome and gene, aligned FLNC reads, and location and type of PAS. 3′ UTR, 3′ untranslated regions. 5′ UTR, 5′ untranslated regions. Overlap, the location of the PAS was annotated with two overlapping genes. Intron, intron region. Intergenic, PAS located in the intergenic region of the reference annotation, predicted as a gene by our analysis. CDS, sequence coding for amino acids in protein. Error, the reference annotation gene for PAS positioning is different from our analysis. Supplementary file 3: Table S3: PASs obtained in each of the SM, EDL, BSF, and EN tissues in two periods. Supplementary file 4: Table S4: binding sites for miRNAs and RBPs were predicted for all identified 3′ UTR regions. p values, the optimal enrichment

p value of the motif according to statistical tests. Adj_p value, p value was adjusted multiple times using Bonferroni correction. Here, the selection was based on the adj_p value <0.05. Supplementary file 5: Table S5: binding sites for miRNAs and RBPs were predicted for variable 3′UTR regions. p values, the optimal enrichment p value of the motif according to statistical tests. Adj_p value, p value was adjusted multiple times using Bonferroni correction. Here, the selection was based on the adj_p value <0.05. Supplementary file 6: Figure S1: Pearson correlation between gene expression and PAS number. (*Supplementary Materials*)

References

[1] R. Elkon, A. P. Ugalde, and R. Agami, "Alternative cleavage and polyadenylation: extent, regulation and function," *Nature Reviews Genetics*, vol. 14, no. 7, pp. 496–506, 2013.

[2] W. Chen, Q. Jia, Y. Song, H. Fu, G. Wei, and T. Ni, "Alternative polyadenylation: methods, findings, and impacts," *Genomics, Proteomics & Bioinformatics*, vol. 15, no. 5, pp. 287–300, 2017.

[3] B. Tian and J. L. Manley, "Alternative polyadenylation of mRNA precursors," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 1, pp. 18–30, 2017.

[4] H. S. Yeh and J. Yong, "Alternative polyadenylation of mRNAs: 3′-untranslated region matters in gene expression," *Molecules and Cells*, vol. 39, no. 39, pp. 281–285, 2016.

[5] Y. Shi, "Alternative polyadenylation: new insights from global analyses," *RNA*, vol. 18, no. 12, pp. 2105–2117, 2012.

[6] X. Wu, M. Liu, B. Downie et al., "Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation," *Proceedings of the National Academy of Sciences*, vol. 108, no. 30, pp. 12533–12538, 2011.

[7] C. H. Jan, R. C. Friedman, J. G. Ruby, and D. P. Bartel, "Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs," *Nature*, vol. 469, no. 7328, pp. 97–101, 2011.

[8] I. Ulitsky, A. Shkumatava, C. H. Jan et al., "Extensive alternative polyadenylation during zebrafish development," *Genome Research*, vol. 22, no. 10, pp. 2054–2066, 2012.

[9] D. C. Di Giammartino, K. Nishida, and J. L. Manley, "Mechanisms and consequences of alternative polyadenylation," *Molecular Cell*, vol. 43, no. 43, pp. 853–866, 2011.

[10] C. R. Ashmore and L. Doerr, "Comparative aspects of muscle fiber types in different species," *Experimental Neurology*, vol. 31, no. 3, pp. 408–418, 1971.

[11] L. Larsson, L. Edstrom, B. Lindegren, L. Gorza, and S. Schiaffino, "MHC composition and enzyme-histochemical and physiological properties of a novel fast-twitch motor unit type," *American Journal of Physiology-Cell Physiology*, vol. 261, no. 1, pp. C93–C101, 1991.

[12] K. Leisson, Ü. Jaakma, and T. Seene, "Adaptation of equine locomotor muscle fiber types to endurance and intensive high speed training," *Journal of Equine Veterinary Science*, vol. 28, no. 7, pp. 395–401, 2008.

[13] S. Ciciliot, A. C. Rossi, K. A. Dyar, B. Blaauw, and S. Schiaffino, "Muscle type and fiber type specificity in muscle wasting," *The International Journal of Biochemistry & Cell Biology*, vol. 45, no. 10, pp. 2191–2199, 2013.

[14] G. C. Sieck, W. Z. Zhan, Y. S. Prakash, M. J. Daood, and J. F. Watchko, "SDH and actomyosin ATPase activities of

different fiber types in rat diaphragm muscle," *Journal of Applied Physiology*, vol. 79, no. 5, pp. 1629–1639, 1985.

[15] M. Gil, M. À. Oliver, M. Gispert et al., "The relationship between pig genetics, myosin heavy chain I, biochemical traits and quality of *M. longissimus* thoracis," *Meat Science*, vol. 65, no. 3, pp. 1063–1070, 2003.

[16] B. Brais, J.-P. Bouchard, Y.-G. Xie et al., "Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy," *Nature Genetics*, vol. 18, no. 2, pp. 164–167, 1998.

[17] E. de Klerk, A. Venema, S. Y. Anvar et al., "Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation," *Nucleic Acids Research*, vol. 40, no. 18, pp. 9089–9101, 2012.

[18] M. Riaz, Y. Raz, M. van Putten et al., "PABPN1-Dependent mRNA processing induces muscle wasting," *PLoS Genet*, vol. 12, no. 5, Article ID e1006031, 2016.

[19] S. Zhang, Y. Zhang, X. Zhou et al., "Alternative polyadenylation drives genome-to-phenome information detours in the AMPKalpha1 and AMPKalpha2 knockout mice," *Scientific Reports*, vol. 8, no. 1, p. 6462, 2018.

[20] C. Wallgren-Pettersson, C. A. Sewry, K. J. Nowak, and N. G. Laing, "Nemaline myopathies," *Seminars in Pediatric Neurology*, vol. 18, no. 4, pp. 230–238, 2011.

[21] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[22] T. Steijger, The RGASP Consortium, J. F. Abril et al., "Assessment of transcript reconstruction methods for RNA-seq," *Nature Methods*, vol. 10, no. 12, pp. 1177–1184, 2013.

[23] E. Beaudoing, S. Freier, J. R. Wyatt, J. M. Claverie, and D. Gautheret, "Patterns of variant polyadenylation signal usage in human genes," *Genome Research*, vol. 10, no. 7, pp. 1001–1010, 2000.

[24] P. J. Shepard, E.-A. Choi, J. Lu, L. A. Flanagan, K. J. Hertel, and Y. Shi, "Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq," *RNA*, vol. 17, no. 4, pp. 761–772, 2011.

[25] A. Derti, P. Garrett-Engele, K. D. Macisaac et al., "A quantitative atlas of polyadenylation in five mammals," *Genome Research*, vol. 22, no. 6, pp. 1173–1183, 2012.

[26] S. Y. Chen, F. Deng, X. Jia, C. Li, and S. J. Lai, "A transcriptome atlas of rabbit revealed by PacBio single-molecule long-read sequencing," *Scientific Reports*, vol. 7, no. 1, p. 7648, 2017.

[27] Y. Chao, J. Yuan, S. Li, S. Jia, L. Han, and L. Xu, "Analysis of transcripts and splice isoforms in red clover (Trifolium pratense L.) by single-molecule long-read sequencing," *BMC Plant Biology*, vol. 18, no. 1, p. 300, 2018.

[28] T. Wang, H. Wang, D. Cai et al., "Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*)," *The Plant Journal*, vol. 91, no. 4, pp. 684–699, 2017.

[29] G. Zhang, M. Sun, J. Wang et al., "PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice," *The Plant Journal: For Cell and Molecular Biology*, vol. 97, no. 2, pp. 296–305, 2018.

[30] S. E. Abdel-Ghany, M. Hamilton, J. L. Jacobi et al., "A survey of the sorghum transcriptome using single-molecule long reads," *Nature Communications*, vol. 7, p. 11706, 2016.

[31] Y. Li, C. Fang, Y. Fu et al., "A survey of transcriptome complexity in *Sus scrofa* using single-molecule long-read sequencing," *DNA Research*, vol. 25, no. 4, pp. 421–437, 2018.

[32] J. C. Loke, E. A. Stahlberg, D. G. Strenski, B. J. Haas, P. C. Wood, and Q. Q. Li, "Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures," *Plant Physiology*, vol. 138, no. 3, pp. 1457–1468, 2005.

[33] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner, "WebLogo: a sequence logo generator," *Genome Research*, vol. 14, no. 6, pp. 1188–1190, 2004.

[34] C. E. Grant, T. L. Bailey, and W. S. Noble, "FIMO: scanning for occurrences of a given motif," *Bioinformatics*, vol. 27, no. 7, pp. 1017-1018, 2011.

[35] D. Retelska, C. Iseli, P. Bucher, C. V. Jongeneel, and F. Naef, "Similarities and differences of polyadenylation signals in human and fly," *BMC Genomics*, vol. 7, no. 1, p. 176, 2006.

[36] D. Kim, B. Langmead, and S. L. Salzberg, "HISAT: a fast spliced aligner with low memory requirements," *Nature Methods*, vol. 12, no. 4, pp. 357–360, 2015.

[37] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.

[38] M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg, "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown," *Nature Protocols*, vol. 11, no. 9, pp. 1650–1667, 2016.

[39] S. Anders, P. T. Pyl, and W. Huber, "HTSeq--a python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, no. 2, pp. 166–169, 2015.

[40] R. C. McLeay and T. L. Bailey, "Motif enrichment analysis: a unified framework and an evaluation on ChIP data," *BMC Bioinformatics*, vol. 11, p. 165, 2010.

[41] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.

[42] P. Sanfilippo, J. Wen, and E. C. Lai, "Landscape and evolution of tissue-specific alternative polyadenylation across *Drosophila* species," *Genome Biology*, vol. 18, no. 1, p. 229, 2017.

[43] H. Wang, R. Li, X. Zhou, L. Xue, X. Xu, and B. Liu, "Genome-Wide analysis and functional characterization of the polyadenylation site in pigs using RNAseq data," *Scientific Reports*, vol. 6, p. 36388, 2016.

[44] Y. Komiya, S. Sawano, D. Mashima et al., "Mouse soleus (slow) muscle shows greater intramyocellular lipid droplet accumulation than EDL (fast) muscle: fiber type-specific analysis," *Journal of Muscle Research and Cell Motility*, vol. 38, no. 2, pp. 163–173, 2017.

[45] J. M. Cameron, V. Levandovskiy, N. MacKay et al., "Identification of a novel mutation in GYS1 (muscle-specific glycogen synthase) resulting in sudden cardiac death, that is diagnosable from skin fibroblasts," *Molecular Genetics and Metabolism*, vol. 98, no. 4, pp. 378–382, 2009.

[46] A. Philippou, M. Maridaki, A. Halapas, and M. Koutsilieris, "The role of the insulin-like growth factor 1 (IGF-1) in skeletal muscle physiology," *In Vivo*, vol. 21, no. 1, pp. 45–54, 2007.

[47] S. Gol, J. Estany, L. J. Fraile, and R. N. Pena, "Expression profiling of the GBP1 gene as a candidate gene for porcine reproductive and respiratory syndrome resistance," *Animal Genetics*, vol. 46, no. 6, pp. 599–606, 2015.

[48] L. Teng, L. Hong, R. Liu, R. Chen, X. Li, and M. Yu, "Cellular localization and regulation of expression of the PLET1 gene in porcine placenta," *International Journal of Molecular Sciences*, vol. 17, no. 12, 2016.

[49] J. Mata, "Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation," *RNA Biology*, vol. 10, no. 8, pp. 1407–1414, 2013.

[50] P. F. Harrison, D. R. Powell, J. L. Clancy et al., "PAT-seq: a method to study the integration of 3′-UTR dynamics with gene expression in the eukaryotic transcriptome," *RNA*, vol. 21, no. 8, pp. 1502–1510, 2015.

[51] X. Zhou, R. Li, J. J. Michal et al., "Accurate profiling of gene expression and alternative polyadenylation with whole transcriptome termini site sequencing (WTTS-Seq)," *Genetics*, vol. 203, no. 2, pp. 683–697, 2016.

[52] S. Koren, G. P. Harhay, T. P. L. Smith et al., "Reducing assembly complexity of microbial genomes with single-molecule sequencing," *Genome Biology*, vol. 14, no. R1019, 2013.

[53] Y. Gao, H. Wang, H. Zhang, Y. Wang, J. Chen, and L. Gu, "PRAPI: post-transcriptional regulation analysis pipeline for Iso-Seq," *Bioinformatics*, vol. 34, no. 9, pp. 1580–1582, 2018.

[54] I. Singh, S. H. Lee, A. S. Sperling et al., "Widespread intronic polyadenylation diversifies immune cell transcriptomes," *Nature Communications*, vol. 9, no. 1, p. 1716, 2018.

[55] M. Oe, I. Nakajima, S. Muroya, M. Shibata, and K. Chikuni, "Relationships between tropomyosin and myosin heavy chain isoforms in bovine skeletal muscle," *Animal Science Journal*, vol. 80, no. 2, pp. 193–197, 2009.

[56] M. Oe, K. Ojima, I. Nakajima, K. Chikuni, M. Shibata, and S. Muroya, "Distribution of tropomyosin isoforms in different types of single fibers isolated from bovine skeletal muscles," *Meat Science*, vol. 118, pp. 129–132, 2016.

[57] M. J. Potthoff, H. Wu, M. A. Arnold et al., "Histone deacetylase degradation and MEF2 activation promote the formation of slow-twitch myofibers," *Journal of Clinical Investigation*, vol. 117, no. 9, pp. 2459–2467, 2007.

[58] A. Galmozzi, N. Mitro, A. Ferrari et al., "Inhibition of class I histone deacetylases unveils a mitochondrial signature and enhances oxidative metabolism in skeletal muscle and adipose tissue," *Diabetes*, vol. 62, no. 3, pp. 732–742, 2013.

[59] A. W. Beharry, P. B. Sandesara, B. M. Roberts, L. F. Ferreira, S. M. Senf, and A. R. Judge, "HDAC1 activates FoxO and is both sufficient and required for skeletal muscle atrophy," *Journal of Cell Science*, vol. 127, no. Pt 7, pp. 1441–1453, 2014.

[60] D. An, S. J. Lessard, T. Toyoda et al., "Overexpression of TRB3 in muscle alters muscle fiber type and improves exercise capacity in mice," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 306, no. 12, pp. R925–R933, 2014.

[61] J. Liu, X. Liang, D. Zhou et al., "Coupling of mitochondrial function and skeletal muscle fiber type by a miR-499/Fnip1/AMPK circuit," *EMBO Molecular Medicine*, vol. 8, no. 10, pp. 1212–1228, 2016.

[62] Y. Zhang, W. Li, M. Zhu, Y. Li, Z. Xu, and B. Zuo, "FHL3 differentially regulates the expression of MyHC isoforms through interactions with MyoD and pCREB," *Cellular Signalling*, vol. 28, no. 1, pp. 60–73, 2016.

[63] V. Raz, M. Riaz, Z. Tatum, S. M. Kielbasa, and P. A. C. 't Hoen, "The distinct transcriptomes of slow and fast adult muscles are delineated by noncoding RNAs," *The FASEB Journal*, vol. 32, no. 3, pp. 1579–1590, 2018.

[64] M. R. Fabian, N. Sonenberg, and W. Filipowicz, "Regulation of mRNA translation and stability by microRNAs," *Annual Review of Biochemistry*, vol. 79, no. 1, pp. 351–379, 2010.

[65] J. V. Esteves, C. Y. Yonamine, D. C. Pinto-Junior, F. Gerlinger-Romero, F. J. Enguita, and U. F. Machado, "Diabetes modulates MicroRNAs 29b-3p, 29c-3p, 199a-5p and 532-3p expression in muscle: possible role in GLUT4 and HK2 repression," *Frontiers in Endocrinology*, vol. 9, p. 536, 2018.

[66] S. Muroya, M. Taniguchi, M. Shibata et al., "Profiling of differentially expressed microRNA and the bioinformatic target gene analyses in bovine fast- and slow-type muscles by massively parallel sequencing," *Journal of Animal Science*, vol. 91, no. 1, pp. 90–103, 2013.

[67] Y. Shi, D. C. Di Giammartino, D. Taylor et al., "Molecular architecture of the human pre-mRNA 3′ processing complex," *Molecular Cell*, vol. 33, no. 3, pp. 365–376, 2009.

[68] M. S. Xiao, B. Zhang, Y. S. Li, Q. Gao, W. Sun, and W. Chen, "Global analysis of regulatory divergence in the evolution of mouse alternative polyadenylation," *Molecular Systems Biology*, vol. 12, no. 12, p. 890, 2016.