



Multitissue single-nucleus RNA-seq reveals cell type-specific regulatory patterns of alternative polyadenylation in pigs

Qiuhan Wen, Zhen Wang, Qi Bao, et al.

Genome Res. 2025 35: 2116-2129 originally published online June 16, 2025

Access the most recent version at doi:[10.1101/gr.280095.124](https://doi.org/10.1101/gr.280095.124)

References This article cites 64 articles, 8 of which can be accessed free at:
<http://genome.cshlp.org/content/35/9/2116.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white-bordered box containing the words "LEARN MORE". On the right, there is a photograph of a woman wearing a red superhero mask and cape, and the Cellecta logo, which consists of a cluster of green dots.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

© 2025 Wen et al.; Published by Cold Spring Harbor Laboratory Press

Multitissue single-nucleus RNA-seq reveals cell type-specific regulatory patterns of alternative polyadenylation in pigs

Qiuhan Wen,^{1,2} Zhen Wang,¹ Qi Bao,¹ Tianli Ding,¹ Haihan Zhang,³ Jianbo Li,⁴ Zhuang Liu,⁵ Jieping Huang,² and Guoqiang Yi^{1,6,7}

¹Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China; ²State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi Key Laboratory of Animal Breeding, Disease Control and Prevention, College of Animal Science and Technology, Guangxi University, Nanning 530004, China; ³College of Animal Science and Technology, Hunan Agricultural University, Changsha 410128, China; ⁴State Key Laboratory of Swine and Poultry Breeding Industry, Agro-Biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China; ⁵Dermatology and Venereology Division, Department of Medicine Solna, Center for Molecular Medicine, Karolinska Institute, 17176 Stockholm, Sweden; ⁶Kunpeng Institute of Modern Agriculture at Foshan, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Foshan 528226, China; ⁷Bama Yao Autonomous County Rural Revitalization Research Institute, Bama 547500, China

As an important posttranscriptional modification mechanism, alternative polyadenylation (APA) plays a crucial role in gene regulation and phenotypic diversity. Whereas extensive studies have explored the global APA landscape using bulk RNA-seq data, in-depth analyses of APA events at the single-cell level remain limited—particularly in farm animals. In this study, we construct a comprehensive APA atlas for 261 cell types across 19 porcine tissues based on single-nucleus RNA sequencing (snRNA-seq) data. This analysis reveals tissue- and cell type-specific patterns of APA. We find that many genes display a clear correlation between the average length of 3' untranslated regions (3' UTRs) and expression levels in various cell types, with most showing a negative correlation. Early cell types within the developmental lineage, such as spermatogonia and satellite cells, display longer 3' UTRs, especially for spermatogenesis, where 3' UTR lengths show significant decreasing trends along the differentiation trajectory. Notably, we find that variable 3' UTR lengths in the *CD47* and *GPD1* genes might be critical regulators during spermatogenesis and myogenesis, respectively, potentially through modulation of RNA-binding protein and miRNA binding sites. Furthermore, the SNP rs323354626, located in the 3' UTR of the *CD47* gene, significantly impacts gene splicing and is strongly associated with reproductive phenotypes. Additionally, we observe that neuronal cells generally possess longer 3' UTRs—a pattern conserved across humans, mice, fruit flies, and pigs. Together, these findings enrich the single-cell atlas of pigs by adding a layer of posttranscriptional regulation to the existing gene expression data, highlighting the significant role of cell type-specific 3' UTR lengths in cell commitment and complex trait regulation.

[Supplemental material is available for this article.]

Alternative polyadenylation (APA) refers to the process by which pre-mRNA adds poly(A) tails to the 3' end through the selection of different polyadenylation sites (PASs), generating diverse transcript isoforms with varying lengths of 3' untranslated regions (3' UTRs) (Reyes and Huber 2018). In general, the APA occurs in over 70% of protein-coding genes in mammals and manifests in two structurally distinct and prevalent forms: tandem 3' UTR (TUTR-APA) and alternative last exons (ALE-APA) (Derti et al. 2012; Baralle and Giudice 2017; Tian and Manley 2017; Wang et al. 2018; Mitschka and Mayr 2022). The TUTR-APA form arises when multiple PASs are situated within the same terminal exon, altering only the 3' UTR sequence while leaving the protein-coding sequence unchanged. Because the 3' UTR contains many essential regulatory elements—such as microRNA and RNA-bind-

ing protein (RBP) binding sites—variation in 3' UTR length can modulate target mRNA function, stability, and translation efficiency. In contrast, ALE-APA results from multiple PASs located in alternative terminal exons, which could lead to the production of truncated protein isoforms (Mariella et al. 2019; Goering et al. 2021). These alternatively polyadenylated mRNAs greatly expand the diversity of transcripts and proteins derived from a single gene and play important functions in the regulation, localization, and function of mRNAs and proteins, ultimately contributing to phenotypic diversity (Mariella et al. 2019; Li et al. 2021). Therefore, it is essential to comprehensively identify global APA events and explore their molecular regulatory mechanisms and biological impacts (Li et al. 2021).

Corresponding author: yiguoqiang@caas.cn

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280095.124>.

© 2025 Wen et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Since its first discovery in 1980, APA has been confirmed in various organisms and tissues (Wei et al. 2020; Lu et al. 2022), where it plays roles in diverse biological systems and processes, including cellular proliferation and differentiation (Ji et al. 2009; Shepard et al. 2011; Wei et al. 2020), organ development (Liu et al. 2024), and complex traits and diseases (Hong et al. 2020; Li et al. 2021). A previous study constructed a multitissue human 3' UTR APA quantitative trait loci (3'aQTLs) map, revealing that genetic variants within the 3' UTR region can regulate APA in candidate genes, thereby influencing various human complex traits and diseases (Li et al. 2021). Meanwhile, recent studies in pigs have highlighted the functional relevance of 3' UTRs (Wang et al. 2023; Zhao et al. 2023; Han et al. 2024). One study demonstrated that the global 3' UTR landscape changes dynamically with oocyte meiotic maturation, suggesting a significant regulatory role in the meiotic processes of oocytes (Zhao et al. 2023). Another study demonstrated that 3' UTRs regulate the response of porcine immature Sertoli cells to acute heat stress (Wang et al. 2023). As an important posttranscriptional regulatory mechanism, APA not only varies at the tissue level but also exhibits substantial cell-to-cell variability. However, the cell type-specific regulatory patterns and biological functions of APA remain poorly understood.

Single-cell/nucleus RNA sequencing (scRNA-seq and snRNA-seq) technologies have been widely used to explore gene expression profiles at the single-cell resolution and offer a unique opportunity to trace dynamic 3' UTR patterns across cell lineages. Several pioneering studies have systematically examined cell type-specific APA events using single-cell data (Göpferich et al. 2020; Agarwal et al. 2021; Lee et al. 2022). For instance, APA-mediated extension of 3' UTRs in neurons of autism spectrum disorder (ASD) patients affects high-confidence ASD risk genes related to neurodevelopment, thus playing a critical role in regulating neuronal differentiation and ASD pathology (Göpferich et al. 2020). To date, single-cell studies of APA in pigs are limited, despite pigs being valuable both as agricultural animals and as biomedical models (Lunney et al. 2021). Therefore, there is a pressing need to explore APA in pigs at the single-cell level to better understand its regulatory roles in traits of economic importance and further genetic breeding.

To expand the pig single-cell atlas with posttranscriptional regulatory insights, we aimed to construct a comprehensive APA atlas across major tissue types and investigate the cell type-specific patterns of APA regulation, focusing on their contributions to cellular differentiation, regulatory functions in key developmental processes, and implications for understanding complex trait biology in livestock breeding.

Results

Global profiling of alternative polyadenylation across pig tissues

To comprehensively profile alternative polyadenylation across pig tissues, we analyzed 10x Genomics single-nucleus RNA sequencing (10x snRNA-seq) data from our previous study as a preprint on bioRxiv (Chen et al. 2023), which identified 261 cell subtypes spanning 19 distinct pig tissues (Supplemental Table S1). To evaluate the feasibility of using 10x snRNA-seq data for detecting APA events, we first analyzed gene body coverage across different tissues and cell types. The results showed that 10x snRNA-seq reads were concentrated near the 3'-end regions of genes (Fig. 1A), enabling reliable detection of 3' UTR isoform expression levels and dynamic APA changes in each tissue and cell type. We identified a total of 40,748 PASs across 12,111 genes. Of these, 26.9% of

genes had a single PAS, whereas 73.1% exhibited multiple PASs (Fig. 1B). Classification based on genomic annotation revealed that 63.1% of PASs were within 3' UTRs, 21.6% in exons, 10.6% in introns, and 4.7% within 1 kb downstream of the last exon (Fig. 1C). Motif enrichment analysis around the PAS regions (± 90 bp from each selected site) identified the canonical hexamer [A/U]UAAA motif enriched at 19 nt upstream of PASs (Fig. 1D). This signal is known to be recognized by the cleavage and polyadenylation specificity factor (CPSF) complex (Mitschka and Mayr 2022), which is a key component of the 3'-end processing machinery. The presence of these classical motifs supports the reliability of the predicted PASs.

Based on the location of PASs, APA can be broadly classified into two main categories: tandem UTRs, where multiple PASs are found within the same terminal exon (TUTR-APA); and alternative last exons (ALE-APA), where multiple APA sites are found within different terminal exons (Fig. 1E). However, some genes contain both types of APAs, which we term as Mixed-APA genes. Across tissues, TUTR-APA was the most prevalent, whereas ALE-APA was less frequent. Tissues such as spleen, testis, and adipose exhibited the highest APA gene counts, consistent with previous findings (Hong et al. 2020; Li et al. 2021). In contrast, kidney, skeletal muscle, and colon had fewer APA events, indicating high tissue specificity (Fig. 1F).

Cell type-specific APA events in pigs

To investigate the expression patterns of 3' UTR isoforms across different cell types from 19 tissues, we utilized the LABRAT tool (Goering et al. 2021) to quantify the usage of alternative 3' UTR isoforms. We identified 4974 APA events, retaining 1231 events occurring in at least 100 cell types for downstream analysis. Given the heterogeneity in 3' UTR lengths across cell types, we classified 3' isoforms into two major categories in each cell type based on their length (Lee et al. 2022).

Hierarchical clustering revealed that skeletal muscle, ileum, cerebrum, and cerebellum favored longer 3' isoforms, whereas lymph, jejunum, testis, and spleen expressed predominantly shorter isoforms (Fig. 2A; Supplemental Fig. S1A). In order to better understand the tissue specificity of 3' isoforms, we performed an overlap analysis of genes that undergo 3' isoform lengthening or shortening. The cerebrum had the highest number of genes with tissue-specific lengthening of 3' isoforms, whereas the spleen had the most genes with tissue-specific shortening of 3' isoforms (Supplemental Fig. S1B,C). Interestingly, the lymph and spleen shared the highest number of genes with identical shortening APA events (Supplemental Fig. S1C). Genes shared between the cerebrum and cerebellum showed significant enrichment in pathways related to "Neuron projection morphogenesis," "Regulation of neuron projection development," and "Trans-synaptic signaling," highlighting the potential role of APA in influencing functional pathways critical for specific brain functions (Supplemental Fig. S1D).

In the inner ring, most cell types associated with the four aforementioned tissues, including tenocytes, astrocytes, excitatory neurons, and oligodendrocytes, showed strong 3' UTR lengthening (Fig. 2A). The proportion of cell types with lengthening 3' isoforms in skeletal muscle tissue was significantly higher than in other tissues. In contrast, all cell types corresponding to tissues previously identified with short 3' isoforms exclusively express these shortening forms, including lymphatic endothelial cells, immature enterocytes, elongating spermatids, and B cells.

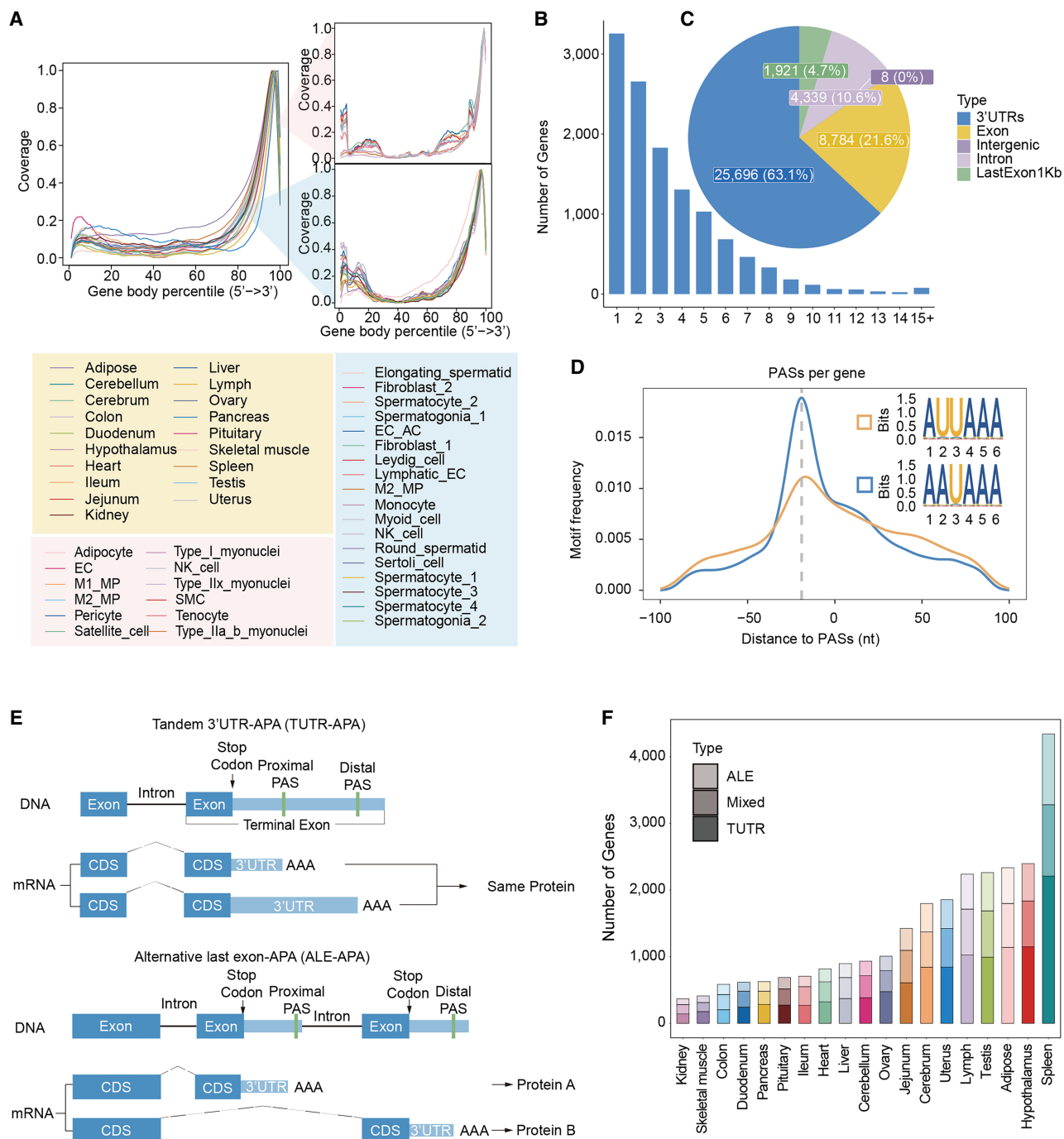


Figure 1. Identification of polyadenylation sites (PASs) using 10x snRNA-seq data. (A) Curve plot showing the coverage of each part of the gene body in 19 different tissues and in each cell type of two representative tissues (skeletal muscle tissue and testicular tissue). The plot on the *left* represents the percentage coverage of the genomes from 5' to 3' in different tissues, and the plot on the *right* is divided into two parts: the *upper right* part shows the gene body coverage of different cell types in skeletal muscle; and the *lower right* part shows the corresponding data in the testis. The *x-axis* is labeled with genome percentiles, and the *y-axis* is shown with coverage. (B) Distribution of the number of PASs per gene. (C) Genomic distribution of PASs. (D) Line plots illustrating the distribution of canonical PAS motifs (AAUAAA, AUUAAA), which are positioned from 90 nucleotides upstream to 90 nucleotides downstream of a random PAS selected on a transcript. (E) Two primary classes of 3' mRNA isoforms generated by APA. (F) Distribution of three different classes of APA genes across 19 tissues.

We conducted Spearman's correlation tests to evaluate the relationship between average expression level and PAS usage of APA genes among cell types. A total of 745 genes showed significant correlations between percent spliced in (PSI) and average expres-

sion level, 228 of which showed positive correlations and 517 of which showed negative correlations (Fig. 2B). For example, *EHBP1L1* and *SLC24A4* were positively correlated, whereas *TLR3* and *COPS3* were negatively correlated. *CD47* and *SLC38A10*

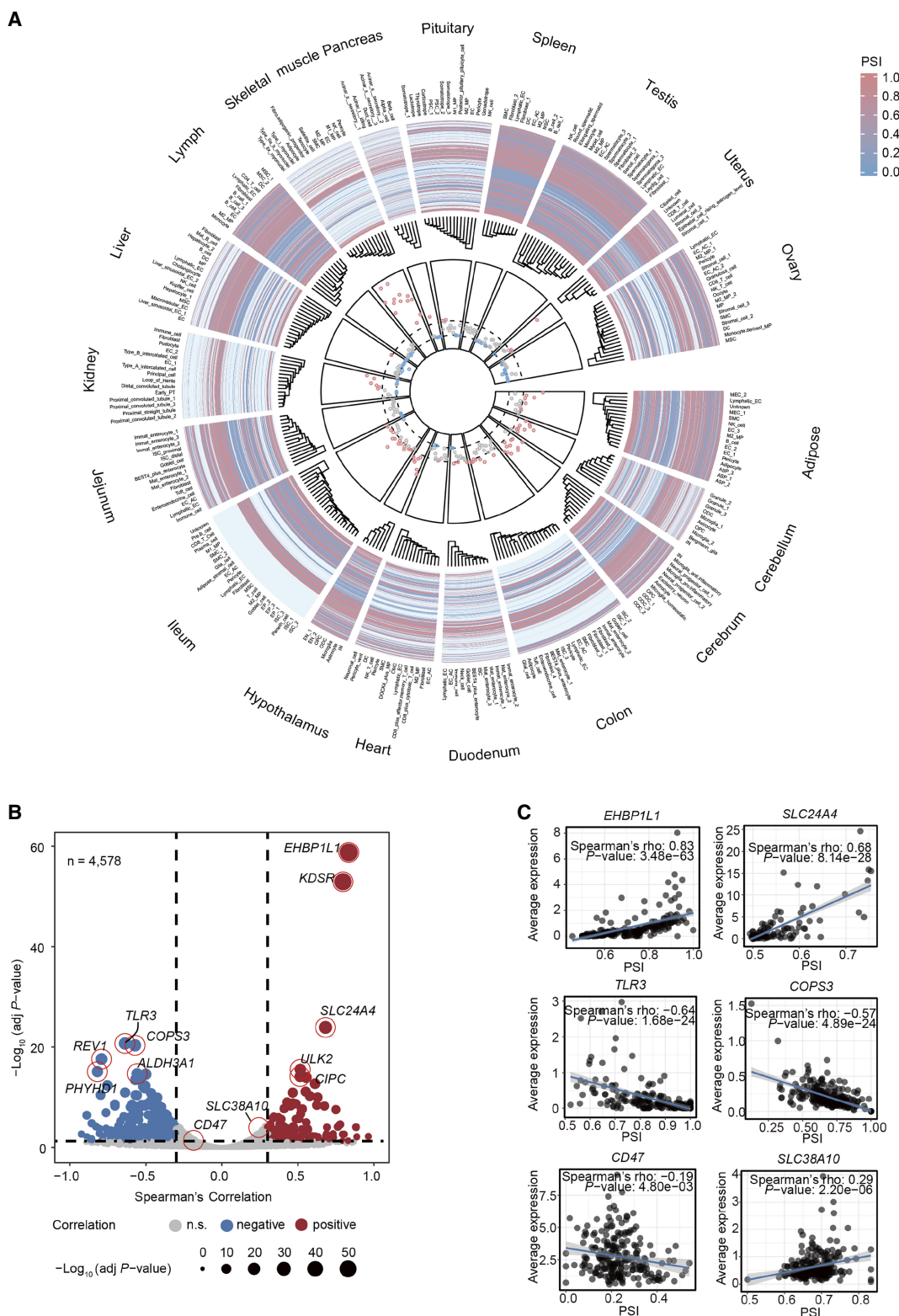


Figure 2. Landscape of global 3' UTR usage across 261 cell types. (A) The circular plot summarizing APA trends across 1231 genes. The outer heat map track illustrates the variation in 3' UTR length among genes for each tissue and cell type (red indicates higher PSI values, and blue indicates lower PSI values). In the inner ring, points represent the median PSI values of all genes for each cell type (261 cell types in total), with red points indicating high PSI (>0.5235), gray points indicating medium PSI (0.5000–0.5235), and blue points indicating low PSI (<0.5000). The dashed lines represent these thresholds. (B) Volcano plot showing Spearman's correlation between PSI and average expression level for each gene. Points are colored by correlation category: negative (blue, Spearman's ≤ -0.3 and adj P -value < 0.05), n.s. (gray), and positive (red, Spearman's ≥ 0.3 and adj P -value < 0.05). (C) Scatter plot of PSI and average expression level for *EHP1L1*, *SLC24A4*, *TLR3*, *COPS3*, *CD47*, and *SLC38A10*. Each dot represents one cell type.

showed no significant difference in expression between isoforms (Fig. 2C), suggesting functional diversification independent of expression level.

Dynamic APA usage in spermatogenesis

Global 3' UTR shortening, exhibiting cell type-specific patterns, has been reported in some model organisms, such as mice and *Drosophila* (Shulman and Elkon 2019; Lee et al. 2022). However, in pigs, the dynamics of these patterns, particularly at single-cell resolution, remain largely unexplored. Given that spermatogenesis is a highly organized and coordinated process of cell differentiation, we first extracted the germ cells from pig testis tissue and globally assessed the changes in mRNA isoforms. In total, we selected 12,977 germ cells, including spermatogonia, spermatocytes, and spermatids, to reannotate the finer cell subtypes involved in spermatogenesis. Based on the expression of canonical markers, we identified eight distinct subpopulations: Undifferentiated_Spermatogonia (*GFR1*), Differentiated_Spermatogonia (*KIT*), Leptotene_Spermatocytes (*STRA8*), Zygotene/Pachytene_Spermatocytes_1, Zygotene/Pachytene_Spermatocytes_2 (*SPAG6*), Spermatids_1 (*SPACA1*), Spermatids_2 (*FBXO24*), and Spermatids_3 (*RPM1*) (Fig. 3A,B). Meanwhile, in silico pseudotime analysis supported a continuous developmental trajectory from spermatogonia to spermatids (Fig. 3C). We next examined APA dynamics change during spermatogenesis at the single-cell level. Specifically, we calculated the mean relative usage of distal (RUD) score across all genes for each cell (Ye et al. 2021). The distribution of RUD scores among individual cells was consistent with a progressive 3' UTR shortening throughout spermatogenesis (Fig. 3C,D). To quantify this trend, we calculated the correlation between the estimated pseudotime and RUD scores. This analysis revealed a significant negative correlation between pseudotime and 3' UTR length (Spearman's correlation = -0.43 , P -value < 0.01) (Fig. 3E). These findings were further supported by published testis scRNA-seq data (Supplemental Fig. S2A–F), indicating a high degree of concordance despite differences between snRNA-seq and scRNA-seq platforms.

Our analysis revealed that 152 genes exhibited gradual 3' UTR shortening, whereas 82 genes showed 3' UTR lengthening during spermatogenesis (Fig. 3F). These genes with shortened 3' UTRs were enriched in pathways related to the regulation of chromosome organization, ribonucleoprotein complex biogenesis, and membrane organization. Conversely, genes with lengthened 3' UTRs were enriched in pathways such as regulation of mitotic cell cycle, phosphotransferase activity, and regulation of microtubule-based process (Fig. 3G).

Additionally, we applied SCAPE to estimate polyadenylation sites and quantify the weights of PASs for each gene in individual cells. For example, *CD47*, which encodes a cell-surface protein that suppresses phagocytosis (Polara et al. 2024), showed similar expression levels in all eight cell types but produced multiple mRNA variants with distinct 3' UTRs (Berkovits and Mayr 2015; Zhang et al. 2024). We identified three significantly differential PASs of the *CD47* gene. Among these, we focused on two PASs that showed the strongest statistical significance and most distinct usage patterns (Supplemental Table S2). Notably, the proximal PAS (Chr 13: 151,483,728:+) was exclusively expressed in late-stage spermatocytes and sperm cells, whereas the distal PAS (Chr 13: 151,488,515:+) of *CD47* exhibited preferential expression in immature sperm cell types (Fig. 3H). Considering that the differential usage of PASs might change the number of RBP binding sites, we

scanned the *CD47* differential 3' UTR using RBP motif position weight matrices. This analysis revealed that most RBP binding motifs within this region corresponded to ELAVL1 and ELAVL2 (Fig. 3I), which are known to promote cell proliferation and inhibit apoptosis (Kota et al. 2021; Yang et al. 2021; Lachiondo-Ortega et al. 2022).

Cell type-specific APA spectrum in skeletal muscle

Given the crucial role of the muscle differentiation in determining pork quantity and quality, we further investigated APA-regulated 3' UTR dynamics in skeletal muscle cells. Our analysis revealed that satellite cells—the stem cell population of skeletal muscle—exhibited a global trend toward relatively longer 3' UTRs during myogenesis, from satellite cells to mature muscle fibers. This observation parallels the trend identified in spermatogenesis. However, the differences in RUD score between satellite cells and other cell types was not significant, and this may be attributed to the limited number of cells analyzed (Fig. 4A,B). To validate these findings, we analyzed a larger skeletal muscle scRNA-seq data set comprising 21,992 cells (Xu et al. 2023). The results showed a significant negative correlation between RUD scores and pseudotime ($\rho = -0.37$, $P = 2.2 \times 10^{-16}$), confirming the gradual shortening trend of 3' UTRs during differentiation from satellite cells to mature muscle fibers (Supplemental Fig. S3A–E). In total, we identified 124 genes with shortened 3' UTRs and 173 genes with lengthened 3' UTR isoforms during myogenesis (Fig. 4C). GO and KEGG enrichment analyses revealed that the genes with shortened 3' UTRs are enriched in pathways involved in regulating RNA processing, muscle structural adaptations, and telomere maintenance (Fig. 4D). In contrast, the genes with lengthened 3' UTRs are overrepresented in pathways involving RNA metabolism, cellular energy metabolism, and organelle function (Fig. 4D).

Among these, *GPD1*—an enzyme involved in mitochondrial oxidation of cytosolic NADH and crucial for lipid metabolism (Oh et al. 2024)—showed significantly shorter 3' UTRs in satellite cells compared to Type_IIx_myonuclei. We identified two significantly differential PAS sites in the *GPD1* gene (Chr 5: 16,014,430:+ and Chr 5: 16,015,618:+) (Supplemental Table S3). RBP binding site screen analysis predicted that SRSF1 and PCBP4 might bind to the regions surrounding the differential polyadenylation sites (Fig. 4E,F).

Furthermore, we examined the expression patterns of key RBP genes involved in 3' UTR isoform regulation across different myogenic cell types, including *ELAVL1*, *MBNL1*, *MBNL2*, and *TNNC2*. Notably, *NOVA1*, *NOVA2*, *SCAF4*, *PABPN1*, *SRSF3*, *ELAVL1*, *RBFOX1*, and *MBNL2* were expressed at relatively higher levels in satellite cells, whereas *RBFOX2* and *MBNL1* were expressed at comparatively lower levels in these cells (Fig. 4G).

Enrichment of eQTLs and sQTLs in the 3' UTRs of different cell types

To explore the functional relationship between 3' UTR length, gene expression, and splice site regulation across various tissues and cell types, we first retrieved eQTL and sQTL data from the PigGTEx database (Teng et al. 2024). To ensure accurate enrichment analysis, we focused on tissues with abundant eQTL and sQTL data, including skeletal muscle, adipose, cerebellum, testis, brain, and liver. Given the minimal sQTL data for adipose, we used it as a benchmark and randomly sampled eQTLs and sQTLs from other tissues to ensure consistent sample sizes. This approach allowed for a more robust assessment of the effect of 3' UTR

Cell type-specific alternative polyadenylation in pigs

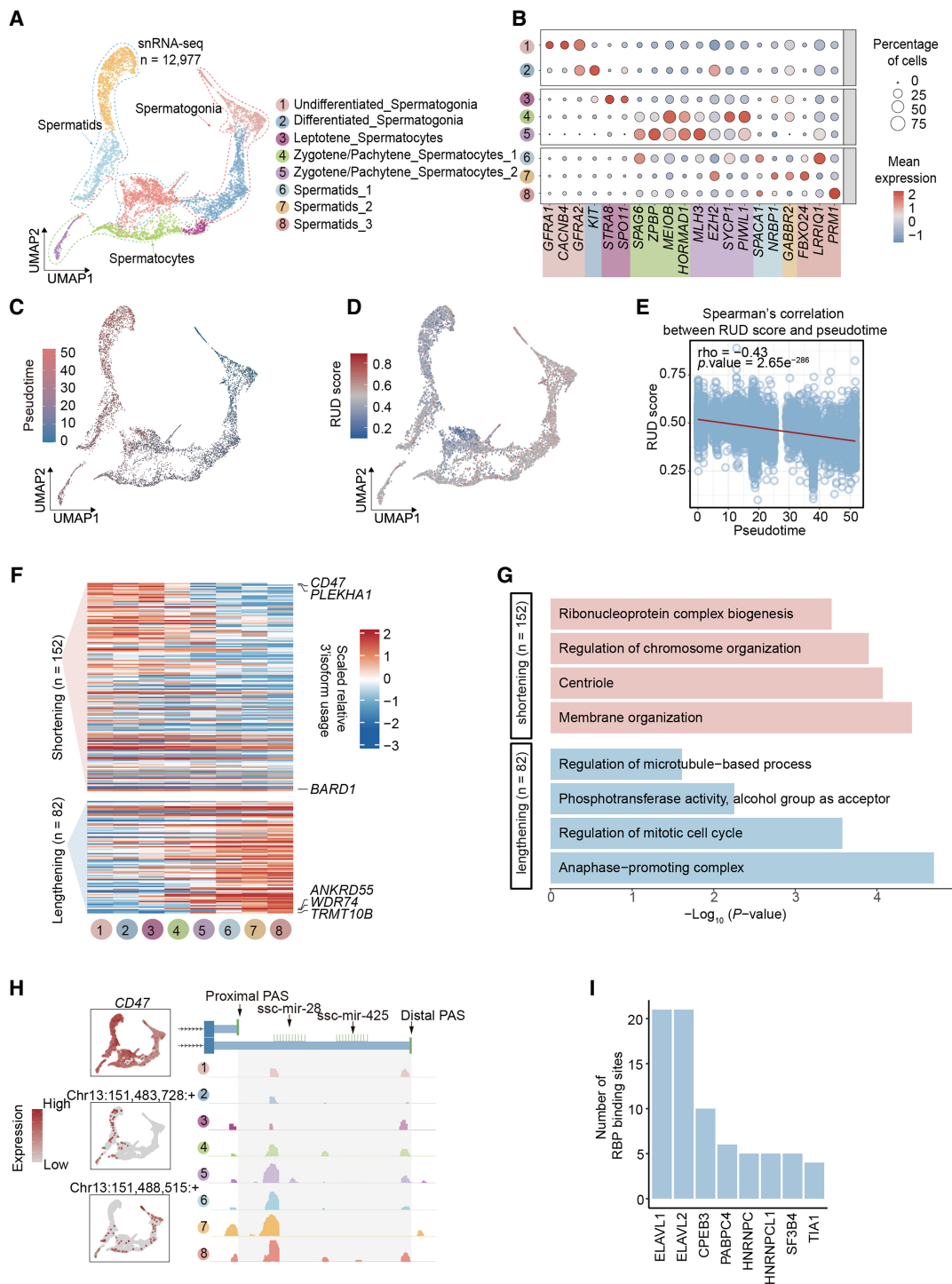


Figure 3. Global 3' UTR shortening in the testis germ lineage. (A) UMAP visualization showing eight distinct germline cell types during the spermatogenesis process. Each dot represents the gene expression profile of a single cell, with each cell type distinguished by a unique color. (B) Dot plot showing the expression levels of marker genes for each cell type. (C) Pseudotime trajectory plot depicting the dynamic process of spermatogenesis. Each point in the trajectory represents a single cell, colored according to its developmental stage within the spermatogenesis process. (D) UMAP displaying 3' isoform usage during spermatogenesis in individual cells. Each dot represents the relative length of the 3' UTR. (E) Spearman's correlation between the RUD score and pseudotime during spermatogenesis. Each point in the correlation plot represents a single cell, with the correlation analysis highlighting the relationship between 3' UTR length and the progression through pseudotime. (F) The scaled relative usage of 3' isoforms across different cell types during spermatogenesis. The upper heat map represents genes with shortened 3' UTRs, and the lower shows genes with lengthened 3' UTRs. (G) Significant biological process terms of genes with dynamic 3' isoform changes. (H) Expression levels, PAS usage, and miRNA binding sites around the CD47 gene. The left section features three UMAP plots. The first UMAP plot shows gene expression, and the subsequent two plots depict the expression of PASs. Each dot represents the expression level of PASs within individual cells. The right section presents an IGV plot. Colors in this plot represent different cell types. (I) Bar plots showing the RNA-binding protein around the differential 3' UTR between proximal PASs and distal PAS loci of CD47.

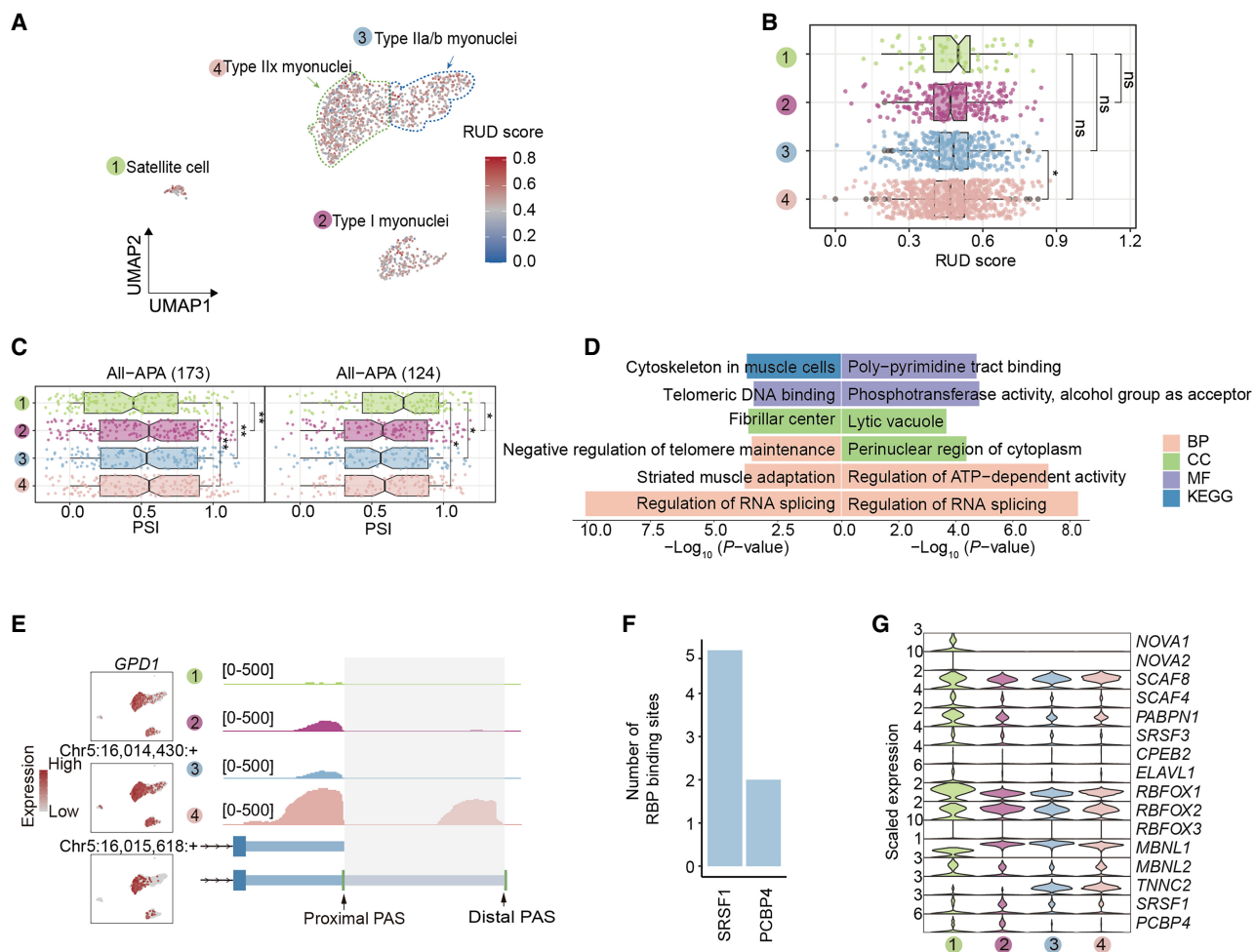


Figure 4. Differential 3' UTR length during myogenesis in skeletal muscle. (A) UMAP displaying 3' isoform usage during myogenesis in individual cells. Each dot represents the relative length of the 3' UTR. (B) Distribution of RUD for different types of muscle cells. Colors correspond to different cell types. The Wilcoxon rank-sum test was used for intergroup comparisons, and the difference between Type IIa/b myonuclei and Type IIx myonuclei was significant ($P < 0.05$), whereas other comparisons did not show significant differences (n.s.). (C) Expression levels, PAS usage, and genomic tracks around the *GPD1* gene. The left box plot shows genes with lengthened 3' UTRs, and the right box plot shows genes with shortened 3' UTRs. Colors correspond to different cell types. (D) Significant biological process terms of genes with dynamic 3' isoform changes. The left box plot shows genes with shortened 3' UTRs, whereas the right box plot shows genes with lengthened 3' UTRs. (E) Expression levels and PAS usage of *GPD1*. The left section features three UMAP plots, the first of which shows gene expression, whereas the subsequent two depict the expression of PASs, with each dot in these two plots representing the expression level of PASs within individual cells. The right section presents an IGV plot. Colors in this plot represent different cell types. (F) Bar plots showing the RNA-binding protein around the differential 3' UTR between proximal PASs and distal PAS loci of *GPD1*. (G) The violin plot showing the scaled expression levels of RBP genes.

length on gene expression and splice site regulation across different tissues.

Our analysis revealed significant differences in the number of eQTLs and sQTLs enriched within 3' UTRs across different tissues (Fig. 5A). Skeletal muscle exhibited the highest number of eQTLs and sQTLs enriched in 3' UTRs, whereas the liver showed the lowest. Notably, in the cerebellum and testis, the number of sQTLs enriched in 3' UTRs was higher than that of eQTLs (Fig. 5A), implying a heightened regulatory requirement at the splicing level in these tissues. This is consistent with the cerebellum's role in complex neural processing and the tightly regulated gene expression necessary for spermatogenesis in the testis.

The enrichment of 3' UTRs across different cell types within the same tissue also showed clear heterogeneity. For instance, pericytes in skeletal muscle, microvascular endothelial cell_2 (MEC_2)

in adipose tissue, microglia_1 in the cerebellum, and inhibitory neurons (INs) in the brain were enriched with more eQTLs and sQTLs, whereas neural progenitor cell_1, fibroblasts in the liver, and mature B cells (Mat_B_cell) were mainly enriched with sQTLs, reflecting cell-specific gene regulatory differences (Fig. 5A). Notably, during myogenesis, satellite cells and Type I myonuclei enriched more sQTLs and eQTLs in 3' UTRs than Type II myonuclei. During spermatogenesis, Leptotene_Spermatocytes enriched more eQTLs and sQTLs in 3' UTR (Fig. 5A).

We further investigated the impact of APA on gene function and how variations in the 3' UTR regions influence gene expression and complex traits in pigs. The *CD47* gene was prioritized as the top candidate to explore the functions in various cellular contexts, given its significantly varying 3' UTR length across different cell types. Initially, we analyzed the PSI values for the *CD47*

Cell type-specific alternative polyadenylation in pigs

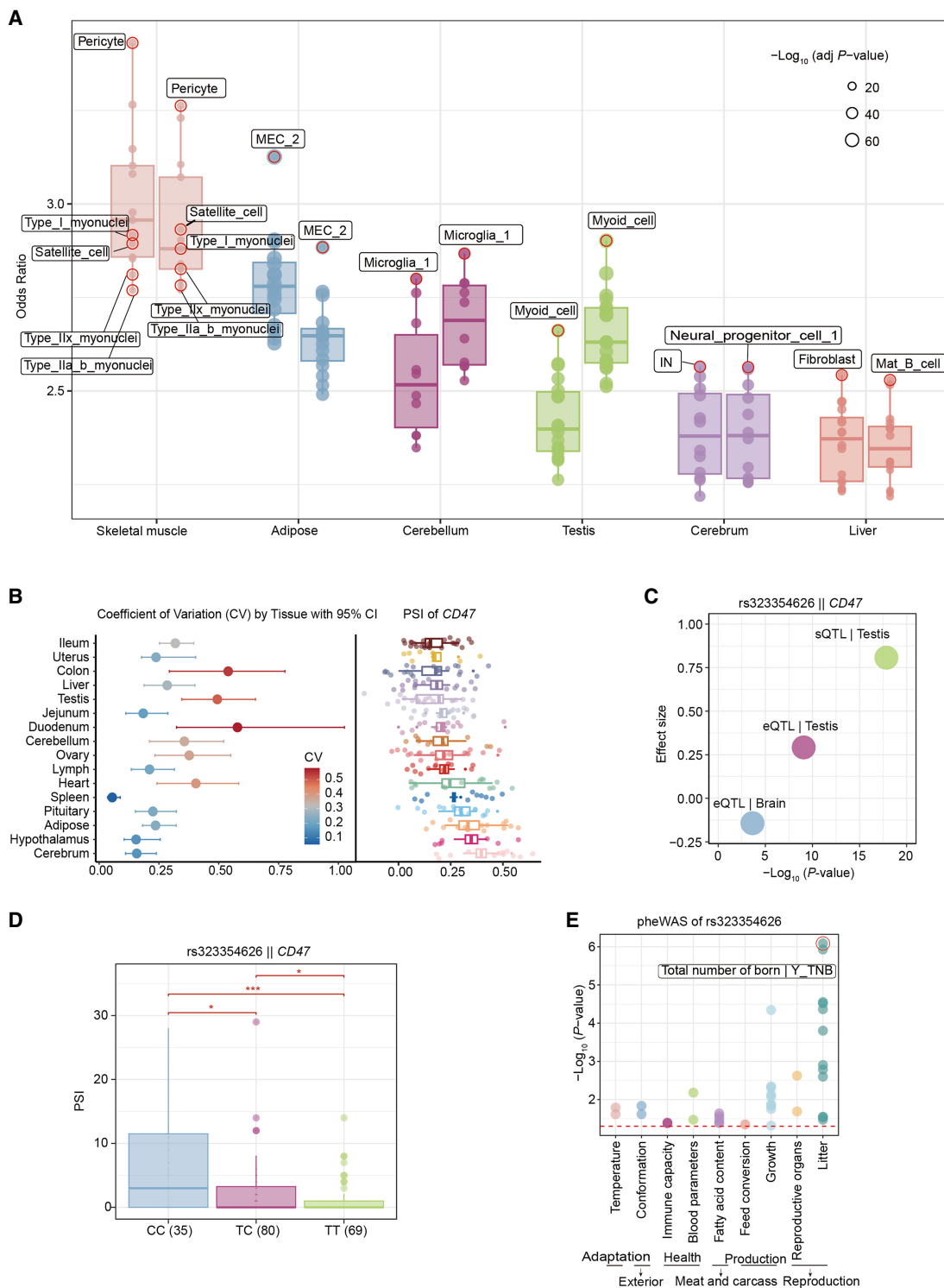


Figure 5. Analysis of the differences in 3' UTR among different cell types and its impacts on genotype and phenotype. (A) Enrichment of 3' UTR by eQTLs and sQTLs in different cell types. The box plot and dot plot on the *left* represent the data of eQTLs, whereas the data of sQTLs are on the *right*. The fill and border colors of each box plot represent different types of organization. The size of the point represents the significance. (B) Coefficient of variation and PSI analysis of *CD47* across different tissues. The *left* panel shows the coefficient of variation (CV) with its 95% confidence intervals for different tissues. The *right* panel displays the PSI values of the *CD47* gene in these tissues. (C) The scatter plot displaying the relationship between effect sizes and statistical significance for eQTLs and sQTLs related to *CD47* expression and splicing. (D) The PSI distribution of *CD47* gene under different genotypes (CC, CT, and TT). This chart includes point plots and box plots, where different colors are used to represent each genotype. (E) Dot plot presenting the pheWAS analysis results of SNP rs323354626 in different biological characteristic categories. Each point represents the strength of the association between a specific phenotype and rs323354626, represented by a negative logarithmic 10 transformation of *P*-values. The color of the dots distinguishes different subcategories, and there is also a red dashed line in the graph, indicating a significance threshold with a *P*-value of 0.05.

gene across cell types. The results showed that the *CD47* PSI values were significantly shorter in testicular tissue, with a higher coefficient of variation (CV), indicating substantial heterogeneity in the 3' UTR lengths within this tissue. Additionally, considerable variability was also observed in the duodenum and colon. In the cerebrum and hypothalamus, the 3' UTR of *CD47* is relatively longer, whereas it is shorter in the ileum and uterus (Fig. 5B).

To examine whether variation in the *CD47* 3' UTR influences gene regulation, we performed an overlap analysis between *CD47* 3' UTR variants and eQTL/sQTL loci. We found that a SNP locus rs323354626, located within the 3' UTR of the *CD47* gene, emerged as a good candidate. By integrating the eQTLs and sQTLs data from the PigGTE database, we revealed that this SNP had the strongest association with sQTLs in testis tissue, characterized by a large effect size and high statistical significance. It was also linked to a moderate eQTL effect in testis and a weaker, negative effect in brain tissue (Fig. 5C). These results indicate that the rs323354626 site is associated with *CD47* splicing in testis tissue. Furthermore, genotype-stratified PSI analysis revealed significant differences among CC, CT, and TT genotypes (Fig. 5D). To explore the potential impact of the rs323354626 site on phenotypes, we examined its association with all collected traits of economic importance in the PigBiobank database (Zeng et al. 2024). We found that the rs323354626 site exhibited the strongest association with reproductive-related phenotypes, particularly under the sub-phenotype "Litter" in the metric "Total number of born" (Fig. 5E). The significance test further supported these findings, suggesting a potential association between variations in the 3' UTRs of different isoforms of the *CD47* gene and reproductive functions.

Comparison of 3' UTR lengths across cell types in multiple species

Previous studies have reported that some RBPs regulate APA by increasing or suppressing the use of proximal or distal polyadenylation signal sites, thereby altering the length of the 3' UTRs (Agarwal et al. 2021; Mitschka and Mayr 2022). To further explore this mechanism, we analyzed the expression patterns of RBP genes across different cell types. Hierarchical clustering analysis revealed that RBPs associated with 3' UTR elongation are specifically upregulated in neuronal cells, distinctly segregating them from other somatic cell types (Fig. 6A). This specific expression pattern suggests that neurons may have longer 3' UTRs. Consistent with this, enhanced PSI values were observed in neurons from a pig model (Fig. 2A).

Given that neural cell types in pigs showed relatively longer 3' UTRs compared to other tissues, we sought to determine whether this pattern is conserved in other species. We obtained 3' UTR length information for different cell types in humans and mice from the scAPAdb database (Zhu et al. 2022) and quantified the dynamic changes in APA using the percentage of the proximal PAS usage index (PPUI). A smaller PPUI value indicates a relatively lengthened 3' UTR. We observed that 3' UTR length varies depending on the tissue context in mice, humans, and pigs (Figs. 2A, 6B, C). Brain tissues, particularly neuronal cell types, exhibit lower PPUI values, indicative of longer 3' UTRs, whereas testicular tissues are characterized by shorter 3' UTRs across these species. In testicular tissue, we specifically observed that 3' UTRs progressively shortened during spermatid maturation, transitioning from round to elongating spermatids. Similarly, early-stage cells in both testicular and muscle lineages consistently exhibited longer 3' UTRs. These trends are consistent with our previous findings (Figs. 2A, 6B,C). To further validate these findings, we expanded the scope

of our study to include nonvertebrates such as *Drosophila*. We obtained data from a previous study (Lee et al. 2022) and analyzed PSI values of APA genes across different cell types in *Drosophila* (Fig. 6D). We observed that brain tissue cells in *Drosophila* also showed relatively longer 3' UTRs compared to other cell types, which is consistent with the results from vertebrates. Collectively, these findings suggest that APA patterns and the regulation of 3' UTR length across different cell types are conserved across species.

In contrast to the conserved patterns observed in neural and testicular tissues, muscle tissue displayed a distinctive characteristic: whereas mouse muscle tissue showed medium-length 3' UTRs, pig skeletal muscle exhibited significantly longer ones (Figs. 2A, 6C). This species-specific difference may be attributed to the strong selective pressure from long-term artificial selection and domestication in pigs. To investigate this hypothesis, we analyzed skeletal muscle scRNA-seq data across three pig breeds: Duroc, Laiwu, and Wild boar (Xu et al. 2023). Correlation analysis demonstrated that Wild boar exhibited a stronger inverse relationship between APA dynamics and developmental progression ($\rho = -0.510$, $P < 0.001$) compared to domesticated breeds (Laiwu: $\rho = -0.210$, $P = 2.02 \times 10^{-72}$; Duroc: $\rho = -0.141$, $P = 5.08 \times 10^{-24}$) (Supplemental Fig. S4A). RUD score quantification across all breeds showed a consistent gradient across myogenic lineages, with satellite cells demonstrating significantly elevated values that progressively decreased during myogenic differentiation (Supplemental Fig. S4B). Statistical analyses of RUD scores revealed significant variations both between breeds and among cell types (Supplemental Fig. S4C,D). Between-breed comparisons demonstrated significant differences ($P < 0.0001$) in satellite cells and myonuclei cells, whereas myoblast populations remained largely comparable. Within each breed, satellite cells consistently exhibited the highest RUD scores, followed by a progressive decrease in differentiated cell types. Notably, the magnitude of these cell-type differences was most pronounced in Wild boar compared to the domesticated breeds (Duroc and Laiwu), suggesting that artificial selection in domesticated breeds may have attenuated APA heterogeneity during muscle development.

Discussion

The motivation behind this study was to investigate the posttranscriptional regulatory mechanism of APA, which is crucial for gene expression and cellular function regulation. Although numerous studies have cataloged APA using bulk RNA-seq, single-cell-level insights, especially in agricultural species, remain sparse. To address this gap, we leveraged snRNA-seq data from 19 porcine tissues to construct a cell type-resolved atlas of 3' UTR usage, thereby extending the pig single-cell compendium beyond gene-expression profiles and highlighting how cell type-specific 3' UTR architecture can influence complex traits.

The comprehensive understanding of the driving forces behind gene regulation in pigs remains an ongoing pursuit. At present, many studies have explored the regulatory mechanisms from multiple layers in pigs, like chromatin accessibility, histone modifications, and *cis*-regulatory variants (Pan et al. 2021; Jin et al. 2023; Quan et al. 2024; Teng et al. 2024). APA events are also one of the key drivers altering gene expression with clear tissue-type and cell-type specificity. Compared to tissue-level APA studies in pigs (Deng et al. 2020; Wang et al. 2023), our research further supplements the cell type-specific information on APA events and illustrates the cellular functions of APA changes in gene expression and complex traits. The predominance of

Cell type-specific alternative polyadenylation in pigs

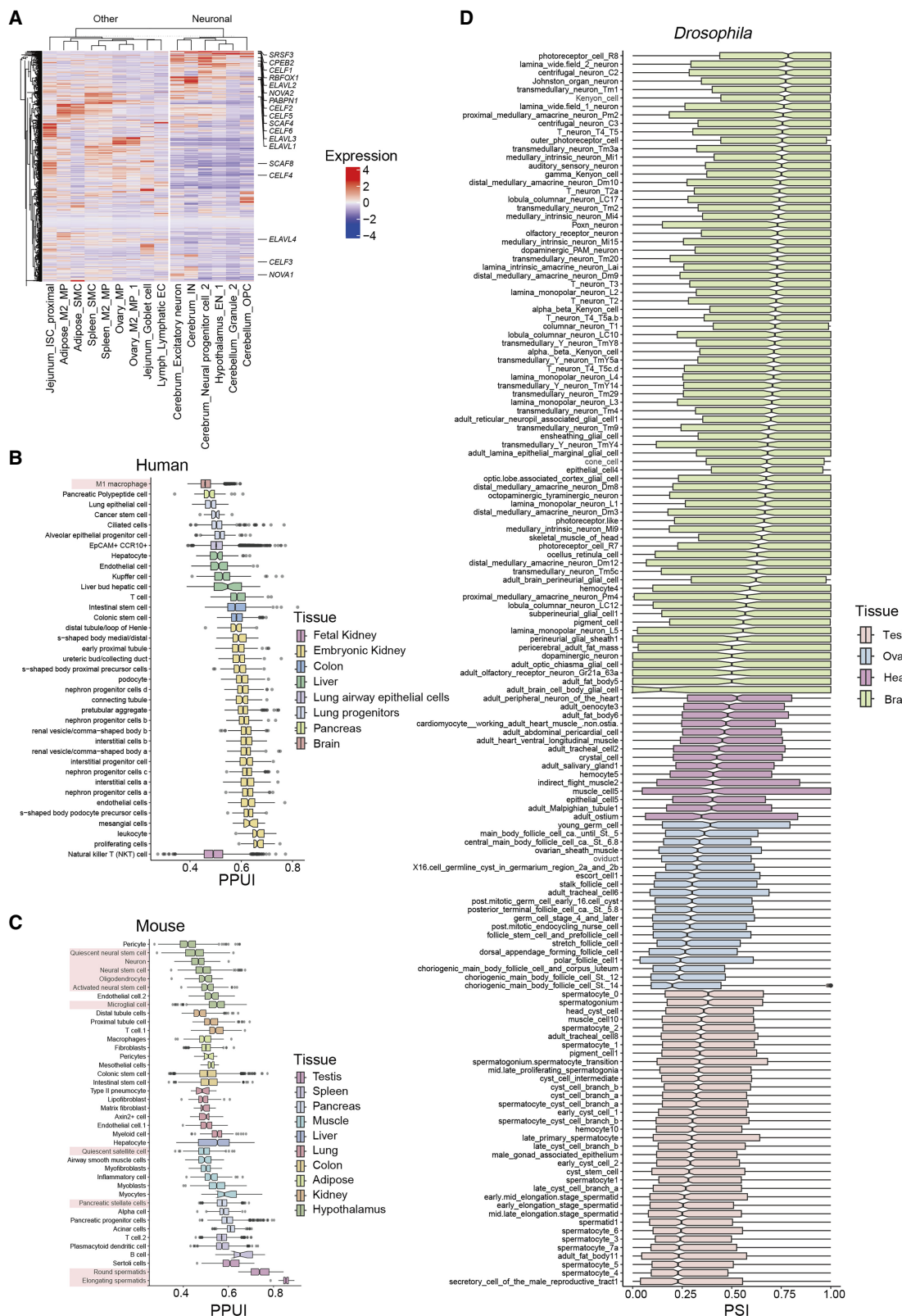


Figure 6. Comparative analysis of 3' UTR lengths across cell types in multiple species. (A) Heat map displaying the normalized expression levels of a set of 1304 RBPs across various cell types. (B) Box plot illustrating the variations in relative 3' UTR lengths across various human cell types. (C) Box plot illustrating the variations in relative 3' UTR lengths across various mouse cell types. (D) Box plot illustrating the variations in relative 3' UTR lengths across various *Drosophila* cell types.

TUTR-APA genes in various tissues enhances the diversity of gene expression (Mittleman et al. 2020). The proportion of ALE-APA was relatively lower, possibly because it can cause significant changes in protein sequence and function (Dubbury et al. 2018). Our tissue-specific 3' UTR patterns align with previous studies (Lee et al. 2022), suggesting that the longer 3' UTRs observed in neural tissues enhance neuronal plasticity, development, transcriptional complexity, and mRNA localization precision (Costessi et al. 2006; Miura et al. 2013; Sanfilippo et al. 2017; Lee et al. 2022; Zhang et al. 2023).

Our study found that cell types with strong differentiation potential, such as spermatogonia and satellite cells, generally have longer 3' UTRs than these mature cell types within the same differentiation path, consistent with prior observations (Shulman and Elkon 2019; Lee et al. 2022). In skeletal muscle, Type IIx myonuclei exhibited significantly shorter 3' UTRs compared to Type IIa/b myonuclei, which provides new clues for the role of posttranscriptional regulation in muscle fiber type determination. The 3' UTR contains many *cis*-regulatory elements, such as RBP binding sites and miRNA binding sites. Shortening of the 3' UTR may result in the deletion of these sites, thereby enabling the gene to adopt cell type-specific functions. For example, during spermatogenesis, the 3' UTR of the *CD47* gene progressively shortens, resulting in the loss of predicted binding sites for miRNAs (ssc-miR-28, ssc-miR-425) and RBPs (ELAVL1 and ELAVL2), which are known to promote proliferation and suppress apoptosis (Kota et al. 2021; Yang et al. 2021; Lachiondo-Ortega et al. 2022). We hypothesize that the differential usage of the *CD47* 3' UTR leads to distinct binding patterns of ELAVL1 and ELAVL2, which play a crucial role in spermatogonia by promoting cell proliferation and preventing apoptosis. Although our computational evidence supports this model, experimental validation using techniques such as RNA-EMSA, CLIP-seq, and cell type-specific perturbation is necessary. However, such approaches are currently constrained by technical challenges in porcine systems, like limited antibody availability and difficulty manipulating primary germ cells. Future studies employing these experimental techniques, once technically feasible, will be crucial for elucidating the precise molecular mechanisms by which *CD47* 3' UTR variations influence spermatogenesis.

The genetic regulation of APA, as revealed in our study, provides mechanistic insights into how posttranscriptional processes contribute to phenotypic variation in livestock. Whereas human studies have identified numerous 3' UTR variants affecting disease risk (Mariella et al. 2019; Mittleman et al. 2020; Li et al. 2021), the application of these data provides a novel opportunity to understand economically important genetic traits in livestock. We observed significant enrichment of eQTLs and sQTLs in the 3' UTRs across different tissues and cell types. Our investigations revealed that the SNP rs323354626, located within the 3' UTR of the *CD47* gene, markedly impacts gene expression, underscoring the pivotal role of 3' UTR variations in the regulatory mechanisms of genes. Additionally, this locus demonstrated significant associations with reproductive phenotypes, most notably within the "Litter" sub-phenotype of the "Total number of born," highlighting the critical role of 3' UTR variations in modulating reproductive functions. These findings suggest that cell type-specific 3' UTRs, arising from APA, modulate these elements, regulating gene expression and alternative splicing, thus impacting complex traits and phenotypes (Teng et al. 2024).

Our comparative analysis across species revealed both conserved and species-specific aspects of APA regulation. In neuronal

and germline cells, APA dynamics were remarkably conserved across evolution, from *Drosophila* to humans. Neurons consistently exhibited longer 3' UTRs, and spermatogenic cells showed progressive shortening during differentiation (Lee et al. 2022; Kang et al. 2023; Ulicevic et al. 2024). This evolutionary conservation underscores the essential and complementary role of APA in development and gene regulation. We observed distinct species-specific differences in muscle tissue, particularly between mice and pigs. Further analysis of different pig breeds revealed that domestication may have altered APA regulation patterns. This finding not only highlights the plasticity of APA regulation and its susceptibility to selective breeding pressures; this divergence suggests that artificial selection during pig domestication may have altered APA regulation patterns, highlighting both the plasticity of this mechanism and providing novel insights into how domestication can influence gene expression regulation.

Although this study marks significant advancements in understanding APA regulatory mechanisms, it faces several limitations that warrant attention for future research. First, despite the enrichment of 3' ends, the short-read lengths of snRNA-seq used to detect APA constrain the resolution of APA isoform detection. Future studies should consider employing long-read snRNA-seq technologies such as Oxford Nanopore Technologies (ONT) or Pacific Biosciences (PacBio) to overcome these limitations and provide more comprehensive insights. Second, our reliance on the CISBP-RNA database for RBP motifs introduces constraints due to potential incompleteness in the data set. Third, the use of snRNA-seq data primarily captures nuclear transcripts whereas most mature RNAs are processed in the cytoplasm, which might not fully reflect the final state of RNA processing. Fourth, cell number disparities across tissues may introduce biases, although our results aligned well with external scRNA-seq data sets. Finally, whereas our analysis shows enrichment of eQTL and sQTL in 3' UTR regions and suggests potential regulatory relationships between RBPs and APA events, these statistical associations do not necessarily imply direct causal relationships. Experimental validation through techniques such as RNA-EMSA, CLIP-seq, or cell type-specific perturbation studies would be required to establish the precise molecular mechanisms underlying these observations. Future research will require integration of more comprehensive multiomics data, advanced sequencing technologies, and functional validation experiments to better elucidate the molecular mechanisms and biological functions of APA regulation at single-cell resolution.

In summary, this study utilized snRNA-seq data from pigs to perform a comprehensive analysis of 3' UTR heterogeneity across 261 cell types, which led to the construction of a comprehensive single-cell APA atlas and the revelation of cell type-specific patterns of APA. Notably, we demonstrated that early immature cells within the same lineage tend to exhibit longer 3' UTRs, which serve as regulators by altering the number of RBP and miRNA binding sites. Furthermore, we identified potential sQTLs affecting the 3' UTR length of the *CD47* gene in sperm cells, which have a significant impact on reproductive traits. Cross-species analysis indicated that APA lengths are generally longer in brain tissues, indicating a degree of regulatory conservation. Conservation of APA in neural and germline cells across species emphasizes its fundamental biological role, and species- and breed-specific differences point to evolutionary plasticity. Together, our work highlights the crucial role of 3' UTR variation as a mechanism for cell-specific regulation and complex traits diversity.

Methods

Quantification of alternative polyadenylation in single-cell transcriptome data

Raw sequencing data were retrieved from the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE233285 (Chen et al. 2023). Briefly, the FASTQ files were aligned using CellRanger (Zheng et al. 2017) against an index created from the Ensembl database (<ftp://ftp.ensembl.org/pub/release-101/>), and cells were filtered if they contained <200 or >5000 genes, <500 or >15,000 UMIs, and >5% mitochondrial content. We used SAMtools (version 1.17) (Danecek et al. 2021) to filter out reads with a mapping quality <30 and split the BAM file into individual cell types based on the cell barcodes, which corresponded to 261 cell types across 19 tissues from previous study. Duplicate reads were then removed using UMI-tools (version 1.14) (Smith et al. 2017). The deduplicated BAM file was then converted to bigWig format using deepTools (version 3.5.1) (Ramírez et al. 2016) to enable visualization of the read coverage using Integrative Genomics Viewer (IGV) (Robinson et al. 2011). Using the deduplicated BAM file as input for “geneBody_coverage.py” from RSeQC, the coverage of reads over the gene body was calculated for different tissues and cell types.

Raw data of testicular and skeletal muscle scRNA-seq were obtained from the Genome Sequence Archive, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA; <https://ngdc.cncb.ac.cn/gsa/>): CRA014793 (Wang et al. 2024) and GSA: CRA011788 (Xu et al. 2023), respectively. FASTQ files were processed and aligned using CellRanger (Zheng et al. 2017) against an index created from the Ensembl database (<ftp://ftp.ensembl.org/pub/release-101/>). Single-cell downstream analyses were performed using Seurat (Stuart et al. 2019). Quality control metrics were applied following standard single-cell analysis protocols. For testis data, cells were retained if they had 500–11,000 genes, 1000–45,000 UMIs, <20% mitochondrial content, and >3% ribosomal content. For skeletal muscle data from three pig breeds (Duroc, Laiwu, and Wild boar), cells were filtered using thresholds of 200–5000 genes and <10% mitochondrial content. The filtered data were normalized using the LogNormalize method with a scale factor of 10,000, and variable features were identified using the “vst” method (nfeatures=3000). Principal component analysis was performed on the scaled data. Batch effects were removed using Harmony integration. Uniform Manifold Approximation and Projection (UMAP) was performed for dimensionality reduction. Cells were clustered using the Louvain algorithm with resolution=1.0. Cell types were annotated based on canonical marker genes.

We employed two strategies to assign 3' UTR profiles from snRNA-seq data. For the clustering-based quantification approach, we used LABRAT (version v0.3.0) (Goering et al. 2021) to quantify 3' UTR length variation at cell-type level by assigning PSI (also denoted as ψ) values to individual genes. PSI values range from 0 to 1, where 0 indicates exclusive usage of the most upstream PAS, and 1 indicates exclusive usage of the most downstream PAS. To classify 3' isoform lengths, we established thresholds based on the distribution of PSI values across all tissues and cell types. Specifically, for each tissue, we calculated the minimum and maximum PSI values among its cell types. The median of these minimum values (0.5000) and maximum values (0.5235) across all tissues were used as classification thresholds (Supplemental Fig. S5A,B). These thresholds, although numerically close, represent the natural boundaries in our data set that effectively distinguish known biological patterns, such as the characteristic longer 3' UTRs in neuro-

nal cells and shorter 3' UTRs in testicular cells. Cell types with median PSI values <0.5000 were categorized as having short 3' UTR usage, those with values >0.5235 as having long 3' UTR usage, and those between these thresholds as having medium-length 3' UTR usage. Based on the position of the PAS, LABRAT classified APA genes into two categories: TUTR-APA and ALE-APA.

For single-cell-level quantification of APA events, we used SCAPE (Zhou et al. 2022) to estimate the PASs of each gene and their corresponding weights. Only these PASs which expressed in more than five cells were kept. We used Seurat for downstream analyses (Stuart et al. 2019), using the script “DifferentialTest.R” to identify differential PASs (<https://github.com/LuChenLab/SCAPE/blob/main/SCAPE.R/R/DifferentialTest.R>). We then used movAPA (version v0.2.0) (Ye et al. 2021) to calculate the mean relative usage of the distal site score for each cell. A higher RUD score indicates higher usage of the distal PAS, whereas a lower RUD score means less usage of the distal PAS.

Pseudotime analysis

We performed in silico pseudotime trajectory analysis using Monocle3 (Cao et al. 2019) by ordering cells along a continuous developmental path from spermatogonia to spermatids. The Seurat object was converted to a Monocle3 cell_data_set for this analysis. Cell ordering was performed using the order_cells function, with the root set to undifferentiated spermatogonia cells.

Enrichment analysis of motifs

For each transcript, a random PAS was selected, and its 90-bp upstream and downstream sequences were extracted using BEDTools getfasta (Quinlan and Hall 2010), then analyzed for motifs using STREME from the MEME Suite (Bailey et al. 2015) (<https://meme-suite.org/>) with a motif width of six nucleotides and a *P*-value threshold of 0.05. The relative positions of motifs identified by STREME were then calculated with respect to the center of each sequence.

Analysis of RNA-binding proteins

We utilized the EuRBPDB (Liao et al. 2020), which contains 2205 genes that were annotated as an RBP gene in the *Sus scrofa* database. Among these, 1304 RBP genes were detected in the analyzed cell types and retained for subsequent analysis. We conducted hierarchical clustering analysis using the scaled expression data of these RBP genes. Additionally, we highlighted RBP genes that have been previously reported to be associated with 3' UTR lengthening (Agarwal et al. 2021; Lee et al. 2022; Mitschka and Mayr 2022).

RBP and miRNA binding site enrichment analysis

We used the FIMO program in MEME Suite (Grant et al. 2011; <https://meme-suite.org/>) to scan RBP sites within the different 3' UTRs of genes with the default parameter (Bailey et al. 2015). Position weight matrices (PWMs) for the human RBPs used were downloaded from CISBP-RNA (Lambert et al. 2018). miRNA binding sites were enriched using miRBase (<https://mirbase.org/>) by providing the target gene sequences (Kozomara et al. 2019).

Enrichment analysis of eQTL and sQTL in the 3' UTRs

We weighted the 3' UTR of each gene using PSI, obtaining the relative cell type-specific 3' UTRs. All eQTL and sQTL results were downloaded from PigGTEx (<https://piggtex.farmgtex.org/>). Subsequently, we employed LOLA (Sheffield and Bock 2016) software for enrichment analysis based on 3' UTRs. The region set was

defined by extending 11 base pairs upstream and downstream of each original position. Cell type-specific 3' UTRs were used as query regions. The universe regions not only comprised all potential genomic locations established by randomly sampling 100,000 protein-coding gene positions from a GTF file but also included the query regions themselves. These data were then inputted into LOLA for analysis. Enrichment results with q -values < 0.05 were deemed statistically significant and retained for further analysis.

Data access

Raw sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE233285.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Dr. Chet Loh (Cambridge, UK) for his meticulous editing and critical proofreading of the manuscript. This work was supported by the Shenzhen Science and Technology Program (KCXFZ20230731094302006), the National Key Research and Development Program of China (2021YFF1000600), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-NBSCA-202301 and CAAS-ZDRW202406), and Bama County Program for Talents in Science and Technology (Barenke20220028). We also thank the editor and anonymous reviewers for the constructive and meaningful discussions on the conceptualization and writing of the manuscript.

Author contributions: G.Y. and Q.W. conceptualized the study. Q.W., Z.W., Q.B., T.D., H.Z., J.L., Z.L., and J.H. conducted the investigation process. Q.B. and T.D. provided the study materials. Z.W. and Q.W. maintained the research data. G.Y. and Q.W. designed the methodology. Q.W. applied statistical and computational techniques to analyze the study data. Q.W. presented data visualization. G.Y. led the research. Q.W. wrote the initial draft. G.Y., H.Z., J.L., Z.L., and Z.W. reviewed and revised the manuscript. G.Y. acquired the financial support.

References

Agarwal V, Lopez-Darwin S, Kelley DR, Shendure J. 2021. The landscape of alternative polyadenylation in single cells of the developing mouse embryo. *Nat Commun* **12**: 5101. doi:10.1038/s41467-021-25388-8

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* **43**: W39–W49. doi:10.1093/nar/gkv416

Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* **18**: 437–451. doi:10.1038/nrm.2017.27

Berkovits BD, Mayr C. 2015. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**: 363–367. doi:10.1038/nature14321

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**: 496–502. doi:10.1038/s41586-019-0969-x

Chen L, Li H, Teng J, Wang Z, Qu X, Chen Z, Cai X, Zeng H, Bai Z, Li J, et al. 2023. Construction of a multi-tissue cell atlas reveals cell-type-specific regulation of molecular and complex phenotypes in pigs (preprint). bioRxiv doi:10.1101/2023.06.12.544530

Costessi L, Devescovi G, Baralle FE, Muro AF. 2006. Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. *Nucleic Acids Res* **34**: 243–253. doi:10.1093/nar/gkj425

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008. doi:10.1093/gigascience/giab008

Deng L, Li L, Zou C, Fang C, Li C. 2020. Characterization and functional analysis of polyadenylation sites in fast and slow muscles. *Biomed Res Int* **2020**: 2626584. doi:10.1155/2020/2626584

Derti A, Garrett-Engle P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**: 1173–1183. doi:10.1101/gr.132563.111

Dubbury SJ, Boutz PL, Sharp PA. 2018. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* **564**: 141–145. doi:10.1038/s41586-018-0758-y

Goering R, Engel KL, Gillen AE, Fong N, Bentley DL, Taliaferro JM. 2021. LABRAT reveals association of alternative polyadenylation with transcript localization, RNA binding protein expression, transcription speed, and cancer survival. *BMC Genomics* **22**: 476. doi:10.1186/s12864-021-07781-1

Göpferich M, George NO, Muelas AD, Bizyn A, Pascual R, Fijalkowska D, Kalamakis G, Müller U, Krijgsveld J, Mendez R et al. 2020. Single cell 3'UTR analysis identifies changes in alternative polyadenylation throughout neuronal differentiation and in autism. bioRxiv doi:10.1101/2020.08.12.24762

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018. doi:10.1093/bioinformatics/btr064

Han Y, Tu W, Zhang Y, Huang J, Meng X, Wu Q, Li S, Liu B, Michal JJ, Jiang Z, et al. 2024. Comprehensive analysis of single-nucleotide variants and alternative polyadenylation between inbred and outbred pigs. *Int J Biol Macromol* **278**: 134416. doi:10.1016/j.ijbiomac.2024.134416

Hong W, Ruan H, Zhang Z, Ye Y, Liu Y, Li S, Jing Y, Zhang H, Diao L, Liang H, et al. 2020. APAAtlas: decoding alternative polyadenylation across human tissues. *Nucleic Acids Res* **48**: D34–D39. doi:10.1093/nar/gkz876

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci* **106**: 7028–7033. doi:10.1073/pnas.0900028106

Jin L, Wang D, Zhang J, Liu P, Wang Y, Lin Y, Liu C, Han Z, Long K, Li D, et al. 2023. Dynamic chromatin architecture of the porcine adipose tissues with weight gain and loss. *Nat Commun* **14**: 3457. doi:10.1038/s41467-023-39191-0

Kang B, Yang Y, Hu K, Ruan X, Liu YL, Lee P, Lee J, Wang J, Zhang X. 2023. Infernap2 uncovers cell type-specific and spatially resolved alternative polyadenylation in the brain. *Genome Res* **33**: 1774–1787. doi:10.1101/gr.277864.123

Kota SK, Lim ZW, Kota SB. 2021. Elavl1 impacts osteogenic differentiation and mRNA levels of genes involved in ECM organization. *Front Cell Dev Biol* **9**: 606971. doi:10.3389/fcell.2021.606971

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**: D155–D162. doi:10.1093/nar/gky1141

Lachiondo-Ortega S, Delgado TC, Baños-Jaime B, Velázquez-Cruz A, Díaz-Moreno I, Martínez-Chantar ML. 2022. Hu antigen R (HuR) protein structure, function and regulation in hepatobiliary tumors. *Cancers (Basel)* **14**: 2666. doi:10.3390/cancers14112666

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. 2018. The human transcription factors. *Cell* **175**: 598–599. doi:10.1016/j.cell.2018.09.045

Lee S, Chen YC, Consortium FCA, Gillen AE, Taliaferro JM, Deplancke B, Li H, Lai EC. 2022. Diverse cell-specific patterns of alternative polyadenylation in *Drosophila*. *Nat Commun* **13**: 5372. doi:10.1038/s41467-022-32305-0

Li L, Huang KL, Gao Y, Cui Y, Wang G, Elrod ND, Li Y, Chen YE, Ji P, Peng F, et al. 2021. An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat Genet* **53**: 994–1005. doi:10.1038/s41588-021-00864-5

Liao JY, Yang B, Zhang YC, Wang XJ, Ye Y, Peng JW, Yang ZZ, He JH, Zhang Y, Hu K, et al. 2020. EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Res* **48**: D307–D313. doi:10.1093/nar/gkz823

Liu X, Shao Y, Han L, Zhu Y, Tu J, Ma J, Zhang R, Yang Z, Chen J. 2024. Microbiota affects mitochondria and immune cell infiltrations via alternative polyadenylation during postnatal heart development. *Front Cell Dev Biol* **11**: 1310409. doi:10.3389/fcell.2023.1310409

Lu P, Chen D, Qi Z, Wang H, Chen Y, Wang Q, Jiang C, Xu JR, Liu H. 2022. Landscape and regulation of alternative splicing and alternative polyadenylation in a plant pathogenic fungus. *New Phytol* **235**: 674–689. doi:10.1111/nph.18164

- Lunney JK, Van Goor A, Walker KE, Hailstock T, Franklin J, Dai C. 2021. Importance of the pig as a human biomedical model. *Sci Transl Med* **13**: eabd5758. doi:10.1126/scitranslmed.abd5758
- Mariella E, Marotta F, Grassi E, Gilotto S, Provero P. 2019. The length of the expressed 3' UTR is an intermediate molecular phenotype linking genetic variants to complex diseases. *Front Genet* **10**: 714. doi:10.3389/fgene.2019.00714
- Mitschka S, Mayr C. 2022. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat Rev Mol Cell Biol* **23**: 779–796. doi:10.1038/s41580-022-00507-5
- Mittleman BE, Pott S, Warland S, Zeng T, Mu Z, Kaur M, Gilad Y, Li Y. 2020. Alternative polyadenylation mediates genetic regulation of gene expression. *eLife* **9**: e57492. doi:10.7554/eLife.57492
- Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. 2013. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res* **23**: 812–825. doi:10.1101/gr.146886.112
- Oh S, Mai XL, Kim J, de Guzman ACV, Lee JY, Park S. 2024. Glycerol 3-phosphate dehydrogenases (1 and 2) in cancer and other diseases. *Exp Mol Med* **56**: 1066–1079. doi:10.1038/s12276-024-01222-1
- Pan Z, Yao Y, Yin H, Cai Z, Wang Y, Bai L, Kern C, Halstead M, Chanthavixay G, Trakooljul N, et al. 2021. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat Commun* **12**: 5848. doi:10.1038/s41467-021-26153-7
- Polara R, Ganesan R, Pitson SM, Robinson N. 2024. Cell autonomous functions of CD47 in regulating cellular plasticity and metabolic plasticity. *Cell Death Differ* **31**: 1255–1266. doi:10.1038/s41418-024-01347-w
- Quan J, Yang M, Wang X, Cai G, Ding R, Zhuang Z, Zhou S, Tan S, Ruan D, Wu J, et al. 2024. Multi-omic characterization of allele-specific regulatory variation in hybrid pigs. *Nat Commun* **15**: 5587. doi:10.1038/s41467-024-49923-5
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160–W165. doi:10.1093/nar/gkw257
- Reyes A, Huber W. 2018. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res* **46**: 582–592. doi:10.1093/nar/gkx1165
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26. doi:10.1038/nbt.1754
- Sanfilippo P, Wen J, Lai EC. 2017. Landscape and evolution of tissue-specific alternative polyadenylation across *Drosophila* species. *Genome Biol* **18**: 229. doi:10.1186/s13059-017-1358-0
- Sheffield NC, Bock C. 2016. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**: 587–589. doi:10.1093/bioinformatics/btv612
- Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-seq. *RNA* **17**: 761–772. doi:10.1261/rna.2581711
- Shulman ED, Elkon R. 2019. Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data. *Nucleic Acids Res* **47**: 10027–10039. doi:10.1093/nar/gkz781
- Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499. doi:10.1101/gr.209601.116
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177**: 1888–1902.e21. doi:10.1016/j.cell.2019.05.031
- Teng J, Gao Y, Yin H, Bai Z, Liu S, Zeng H, The PigGTEx Consortium, Bai L, Cai Z, Zhao B, et al. 2024. A compendium of genetic regulatory effects across pig tissues. *Nat Genet* **56**: 112–123. doi:10.1038/s41588-023-01585-7
- Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**: 18–30. doi:10.1038/nrm.2016.116
- Ulicevic J, Shao Z, Jasnovidova O, Bressin A, Gajos M, Ng AH, Annaldasula S, Meierhofer D, Church GM, Buskamp V, et al. 2024. Uncovering the dynamics and consequences of RNA isoform changes during neuronal differentiation. *Mol Syst Biol* **20**: 767–798. doi:10.1038/s44320-024-00039-4
- Wang R, Zheng D, Yehia G, Tian B. 2018. A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res* **28**: 1427–1441. doi:10.1101/gr.237826.118
- Wang Y, Wu ZW, Mou Q, Chen L, Fang T, Zhang YQ, Yin Z, Du ZQ, Yang CX. 2023. Global 3'-UTRome of porcine immature Sertoli cells altered by acute heat stress. *Theriogenology* **196**: 79–87. doi:10.1016/j.theriogenology.2022.11.014
- Wang X, Wang Y, Wang Y, Guo Y, Zong R, Hu S, Yue J, Yao J, Han C, Guo J, et al. 2024. Single-cell transcriptomic and cross-species comparison analyses reveal distinct molecular changes of porcine testes during puberty. *Commun Biol* **7**: 1478. doi:10.1038/s42003-024-07163-9
- Wei L, Lee S, Majumdar S, Zhang B, Sanfilippo P, Joseph B, Miura P, Soller M, Lai EC. 2020. Overlapping activities of ELAV/Hu family RNA binding proteins specify the extended neuronal 3' UTR landscape in *Drosophila*. *Mol Cell* **80**: 140–155.e6. doi:10.1016/j.molcel.2020.09.007
- Xu D, Wan B, Qiu K, Wang Y, Zhang X, Jiao N, Yan E, Wu J, Yu R, Gao S, et al. 2023. Single-cell RNA-sequencing provides insight into skeletal muscle evolution during the selection of muscle characteristics. *Adv Sci (Weinh)* **10**: e2305080. doi:10.1002/adv.202305080
- Yang C, Yao C, Ji Z, Zhao L, Chen H, Li P, Tian R, Zhi E, Huang Y, Han X, et al. 2021. RNA-binding protein ELAVL2 plays post-transcriptional roles in the regulation of spermatogonia proliferation and apoptosis. *Cell Prolif* **54**: e13098. doi:10.1111/cpr.13098
- Ye W, Liu T, Fu H, Ye C, Ji G, Wu X. 2021. movAPA: modeling and visualization of dynamics of alternative polyadenylation across biological samples. *Bioinformatics* **37**: 2470–2472. doi:10.1093/bioinformatics/btaa997
- Zeng H, Zhang W, Lin Q, Gao Y, Teng J, Xu Z, Cai X, Zhong Z, Wu J, Liu Y, et al. 2024. PigBiobank: a valuable resource for understanding genetic and biological mechanisms of diverse complex traits in pigs. *Nucleic Acids Res* **52**: D980–D989. doi:10.1093/nar/gkad1080
- Zhang Z, Bae B, Cuddleston WH, Miura P. 2023. Coordination of alternative splicing and alternative polyadenylation revealed by targeted long read sequencing. *Nat Commun* **14**: 5506. doi:10.1038/s41467-023-41207-8
- Zhang T, Wang F, Xu L, Yang YG. 2024. Structural-functional diversity of CD47 proteoforms. *Front Immunol* **15**: 1329562. doi:10.3389/fimmu.2024.1329562
- Zhao H, Wu ZW, Zhang R, Wang Y, Du ZQ, Yang CX. 2023. Dynamic changes of 3' UTR length during oocyte-to-zygote transition of in vitro pig embryos. *Reprod Domest Anim* **58**: 605–613. doi:10.1111/rda.14327
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**: 14049. doi:10.1038/ncomms14049
- Zhou R, Xiao X, He P, Zhao Y, Xu M, Zheng X, Yang R, Chen S, Zhou L, Zhang D, et al. 2022. SCAPE: a mixture model revealing single-cell polyadenylation diversity and cellular dynamics during cell differentiation and reprogramming. *Nucleic Acids Res* **50**: e66. doi:10.1093/nar/gkac167
- Zhu S, Lian Q, Ye W, Qin W, Wu Z, Ji G, Wu X. 2022. scAPAdb: a comprehensive database of alternative polyadenylation at single-cell resolution. *Nucleic Acids Res* **50**: D365–D370. doi:10.1093/nar/gkab795

Received October 12, 2024; accepted in revised form June 6, 2025.