

ORIGINAL ARTICLE

An integrated analysis of second- and third-generation transcriptome sequencing technologies reveals the *DAZAP1* function in pig testis

Xia Zhang^{1,2#} , Hailong Huo^{3#} , Honglin Li^{2#} , Yongqing Liu¹ , Fujie Qiao¹ , Changyao Li² , Jinlong Huo^{2*} 

¹Department of Biological and Food Engineering, Lyuliang University, Lvliang, Shanxi, China

²College of Animal Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, China

³Yunnan Open University, Kunming, Yunnan, China

How to cite: Zhang X, Huo H, Li H, Liu Y, Qiao F, Li C, Huo J. An integrated analysis of second- and third-generation transcriptome sequencing technologies reveals the *DAZAP1* function in pig testis. Anim Reprod. 2025;22(2):e20240141. <https://doi.org/10.1590/1984-3143-AR2024-012>

Abstract

The quality of pig sperm is one of the crucial determinants of reproductive ability, and sperm defects can shorten the reproductive life of boars and affect the production of offspring. During transcription and translation, the *DAZAP1* gene exerts regulatory control over alternative splicing, thereby exerting influence on vital cellular processes including cell growth, development, and spermatogenesis. In this study, we employed second- and third-generation transcriptome sequencing techniques to isolate and identify the *DAZAP1* gene and its transcripts using Banna mini-pig inbred line (BMI) testicular cDNA as a template. We identified three splice variants of the *DAZAP1* gene, including ENSSSCT00000023438.4 (*DAZAP1_X1*), ENSSSCT00000051975.3 (*DAZAP1_X2*), and ENSSSCT00000074738.2 (*DAZAP1_X3*). Furthermore, the transcript *DAZAP1_X2*, was subjected to comprehensive analysis. The *DAZAP1_X2* variant comprises 13 exons, with a coding sequence (CDS) length of 1254 bp (417 aa). Subsequently, enrichment analyses based on GO and KEGG pathways revealed that *DAZAP1_X2* primarily participated in pathways associated with spermatogenesis, movement of the 9+2 cilium structure, germ cell development, gamete generation, and sexual reproduction. The ceRNA analysis identified three miRNAs interacting with *DAZAP1_X2*: ssc-miR-107, ssc-miR-127, and ssc-miR-1343, which were primarily linked to spermatogenesis. Both the testis and urethral bulb had significant levels of *DAZAP1* expression, according to multi-tissue expression analysis. Subcellular localization indicated that the *DAZAP1* was mainly distributed in the cell nucleus. *DAZAP1* was critical for sperm formation and was essential for reproductive. These results shed light on the biological roles of *DAZAP1* in pigs.

Keywords: Banna mini-pig inbred line (BMI), *DAZAP1*, spermatogenesis, alternative splicing isoform, transcriptional regulation.

Introduction

Spermatogenesis is the procedure by which germ cells produce elongated sperm through mitosis, differentiation, and meiosis. This process involves significant chromatin condensation in the nucleus, which impacts sperm motility and fertilization (Cho et al., 2001; Hao et al., 2021; Hao et al., 2020). Spermatogenesis primarily occurs in the seminiferous tubules of the testes, where the somatic and sertoli cells are essential for the production and function of spermatozoa (Neto et al., 2016; Cheng and Mruk, 2010). After germ cell development, mature spermatids are

*Corresponding author: jinlonghuo973@163.com

Received: December 31, 2024 Accepted: May 09, 2025

Editor-in-Chief: Carlos Eduardo Ambrósio

Conflicts of interest: The authors have no conflict of interest to declare.

Xia Zhang, Hailong Huo and Honglin Li were co-first authors.



Copyright © The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

released from the sertoli cells into the tubule lumen and travel through the rete testis and efferent ducts into the epididymis. During this passage, they undergo biochemical changes that enable motility and fertilization capability (Leblond and Clermont 1952; You et al., 2021).

The *DAZ* gene family is primarily expressed in germ cells and encodes germ cell-specific RNA binding proteins characterized by a conserved DAZ domain (Tsui et al., 2000). In the human genome, this family includes three members: *BOULE*, *DAZL*, and *DAZ* (Reijo et al., 1995; Saxena et al., 1996). DAZ-associated protein 1 (DAZAP1) was identified in RAW macrophages as an RNA-binding protein that becomes phosphorylated upon LPS stimulation (Morton et al., 2006). DAZAP1 is ubiquitously present in germ cells during late testicular development (Vera et al., 2002). It interacts with the Y-chromosome-encoded DAZ protein and the autosomal DAZ-like protein DAZL, playing an essential role in normal development and spermatogenesis (Dai et al., 2001). DAZAP1 contains two RNA binding domains (RBDs) and a proline-rich C-terminal region, which regulates the RNA life cycle from early transcription to decay (Vera et al., 2002; Smith et al., 2011). Additionally, *DAZAP1* is involved in RNA transcription, splicing, processing, transport, and localization, playing a key role in post-transcriptional regulation (Wang et al., 2021; Sasaki et al., 2018). *DAZAP1* is vital for normal growth, development, and fertility in mice, with allelic knockout leading to severe defects in spermatogenesis and cell growth, underscoring its crucial role in both germ cell and somatic cell development (Dai et al., 2001).

The Banna mini-pig inbred line (BMI), derived from the indigenous Diannan Small-ear pigs (DSE), has subjected to 44 years of highly inbred breeding using full-sibling and parent-offspring mating, representing a large mammalian model, which offers a genetically highly homozygous and well-characterized experimental animal model for biomedical research, xenotransplantation, and transgenic animal studies (Wang et al., 2023; Huo et al., 2022). However, partial infertility in some inbred boars has led to discontinuities in lineages, hindering the population expansion. Currently, the studies on *DAZAP1* are primarily focused on mice, rats, and humans. Therefore, it is crucial to investigate the molecular functions and regulatory roles of *DAZAP1* in pigs. This study aims to identify alternative splicing isoforms of *DAZAP1* using combined second- and third-generation transcriptome sequencing techniques (Hu et al., 2021), and to elucidate the specific functions and molecular mechanisms of the key transcripts regulating the expression of genes associated with pig sperm production, maturation, and fertilization at the molecular and cellular levels. Overall, this study will lay the foundation for the continuation of breeding in boars, providing fundamental data for understanding the regulatory mechanisms of spermatogenesis in BMI testes.

Methods

Second- and third-generation transcriptome sequencing

Testicular samples were collected from four 12-month-old BMI males, and transcriptome profiling of *DAZAP1* in the testis was conducted using a combined approach of second- and third-generation sequencing technologies. *DAZAP1* transcripts were visualized using the Sashimi plot function in IGV. The reference genome index was then constructed using STAR-2.5.2 (Dobin et al., 2013). Finally, the expression of the *DAZAP1* was visualized with the R package Gviz.

Amplification and identification of different transcripts

The complete coding region of the *DAZAP1* mRNA sequence was obtained from the NCBI database and used as a template. Primers for gene coding region cloning were designed using Oligo7. Three primer pairs were designed based on different transcripts and labeled as F1R1 (F1: TGGGGCCAGTGACCTTCCG; R1: GCGTCCAAAGCCACTTCCG), F2R2 (F2: ACTGGAGCACCACCAAGAGAC; R2: TGACCGCGTCCAAAGCCACT), as well as F3R3 (F3:

CTTTCCCAGTATGGCGAGGT; R3: ACTTCAGTCCAGACACCCACA). Traditional PCR was employed for amplification, followed by electrophoresis detection.

Characterization of vital transcripts of *DAZAP1*

The nucleotide and protein sequences of *DAZAP1* obtained from sequencing were inspected and edited using the Lasergene 7.1 software package. Open reading frames (ORF) were obtained from the complete coding sequence of the gene. We conducted sequence alignment, secondary structure and tertiary structure analyses using Seqman, SPOMA (Sapay et al., 2006) and AlphaFold3 (Abramson et al., 2024). Additionally, we utilized ExPASy (Gasteiger et al., 2005) and NetPhos 3.1 (Blom et al., 2004) to analyze protein hydrophilicity and protein functional modification sites, respectively. BLAST and MegAlign were employed to analyze conserved domains and evolutionary relationships.

Protein-protein interaction analysis

Protein-protein interaction (PPI) network was constructed using String 11.5, and functional enrichment analysis, including GO and KEGG, was performed using the clusterProfiler with a significance threshold set at $P < 0.05$. Finally, the identified proteins were correlated with gene expression data to assess expression relationships.

Regulatory network analysis

To analyze molecular function, biological processes, and cellular components, we annotated the *DAZAP1* using the UniProt database. To identify miRNAs and lncRNAs regulating *DAZAP1*, transcriptome data were analyzed and the ceRNA transcriptional regulatory network was subsequently constructed.

Multi-tissue expression profiling

The cDNA templates from BMI tissues, including the heart, liver, spleen, lungs, kidneys, gonads, and accessory glands, were used to assess the expression levels of the *DAZAP1* gene across various tissues. The qPCR primers (F4: CCGAAGGAAGGATGGCAGAAA; R4: CACTTCTGTGACCACTCCAAAT) were designed at the exon-exon junction using Seqman and Oligo7 software. The *GAPDH* (F5: CCTTCATTGACCTCCACTACATGGT; R5: CCACAACATACGTAGCACCAGCATC) was used as a housekeeping gene. Data analysis followed the relative quantification $2^{-\Delta\Delta Ct}$ method (Rao et al., 2013).

Subcellular localization

Transfection was conducted when Swine Testis (ST) cell confluence reached 70-80% in each well. The cells were then incubated with 5% CO₂. Successful transfection was confirmed by observing green fluorescent protein expression under an inverted fluorescence microscope. Mitochondria were stained with Mito Tracker, and nuclei with Hoechst33342 to determine the intracellular localization of the target protein.

Results

Different alternative splicing isoforms of *DAZAP1*

The second- and third-generation sequencing results showed that there were three alternative splicing isoforms of *DAZAP1* in the BMI pig testes, namely ENSSSCT00000023438.4, ENSSSCT00000051975.3, and ENSSSCT00000074738.2, which were named here DAZAP1_X1,

DAZAP1_X2, and DAZAP1_X3, respectively (Figure 1). The results showed that our four samples matched the same transcript, which had the most exons and was consistent with the annotated transcript. DAZAP1_X1 variant consists of 11 exons, with a CDS length of 1197 bp (398 aa), including 23.31% A, 26.98% G, 31.00% C, and 18.71% T. The DAZAP1_X2 variant consists of 13 exons, with a CDS length of 1254 bp (417 aa), including 23.37% A, 17.38% T, 27.67% G, and 31.58% C. The DAZAP1_X3 variant comprises 12 exons, with a CDS length of 1221 bp (406 aa), including 23.26% A, 16.79% T, 28.17% G, and 31.78% C. Of note, DAZAP1_X2 had the longest CDS and the most exons, making it the major transcript.

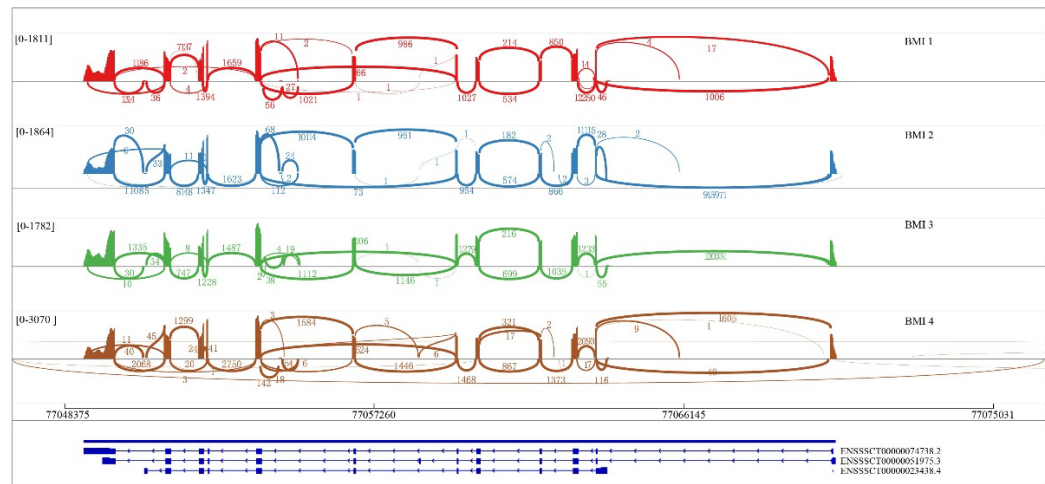


Figure 1. Gene structure of alternative splicing variants of porcine *DAZAP1* gene. The four different colors in the figure represent four distinct samples. The numbers above the arcs indicate expression levels, with higher numbers representing higher expression and lower numbers representing lower expression. The three tracks below represent three different transcripts (ENSSSCT00000023438.4, ENSSSCT00000051975.3, and ENSSSCT00000074738.2).

Sequence verification and identification of different *DAZAP1* transcripts

After ORF prediction and sequence alignment, it was determined that the three amplified sequences correspond to the three transcripts of the *DAZAP1* gene. Using cDNA from BMI testicular tissues as a template, three pairs of specific primers were designed to amplify three different splicing isoforms of the *DAZAP1* gene. Agarose gel electrophoresis confirmed that the amplified fragments matched the expected results, with lengths of 1275 bp, 1264 bp, and 1262 bp, respectively (Figure 2A-2B). PCR results indicated that the amplified fragments from the three primer pairs included the complete coding sequences (CDS) and the portion of the 5' and 3' untranslated regions (UTRs) (Figure 2C-2E). Consistent with the transcriptome data, DAZAP1_X2 emerged as the predominant transcript.

Sequence feature analysis and function prediction of DAZAP1_X2

The molecular weight of pig DAZAP1_X2 protein is 44.50 kDa, with a molecular formula of C₂₀₀₁H₂₉₆₇N₅₅₁O₅₈₆S₁₂, and an isoelectric point of 8.71. The secondary structure of the DAZAP1 protein, comprises 12.23% α -helix (51 aa), 5.52% β -turn (23 aa), 12.95% extended strand (54 aa), and 69.30% random coil (289 aa), respectively (Figure 3A). The tertiary structure of DAZAP1 protein is consistent with the secondary structure analysis results (Figure 3B). DAZAP1 contains phosphorylation sites for serine, threonine, and tyrosine residues (Figure 3C). The N-terminus of the DAZAP1 is hydrophobic, while the C-terminus is hydrophilic, with the maximum hydrophobicity value being 2.067 at the 139th amino acid position, and the minimum hydrophobicity value being -3.033 at the 94th amino acid position (Figure 3D).

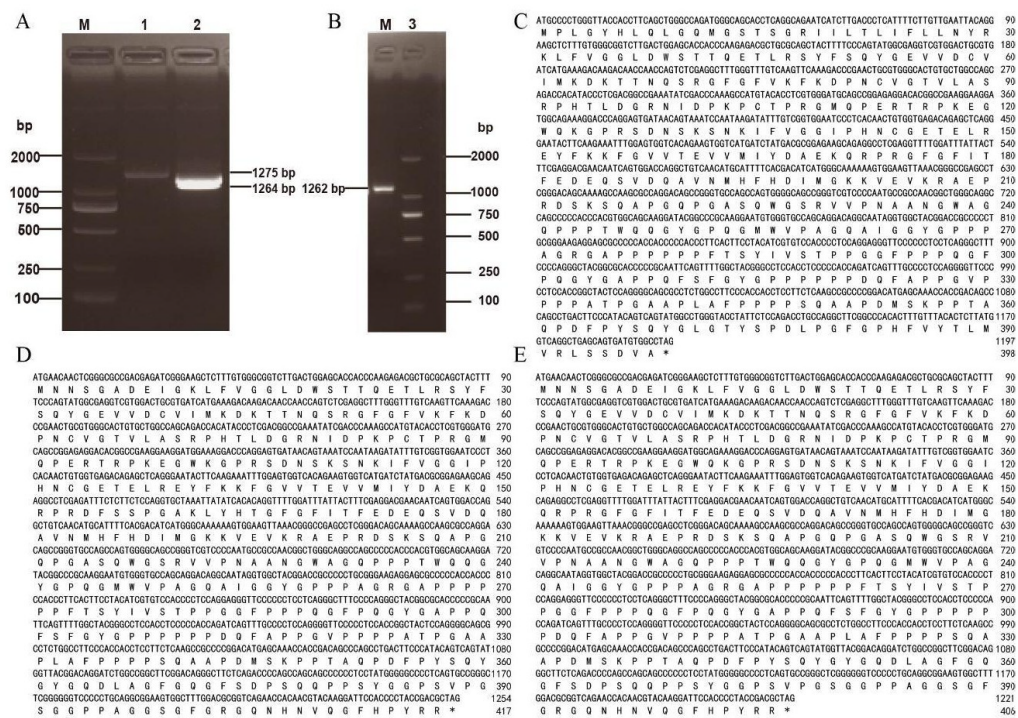


Figure 2. Identification of distinct transcripts of *DAZAP1*. (A-B) CDS sequence amplification of different transcripts of *DAZAP1*, M: DL2000 Marker; (C) Nucleic acid protein sequence of *DAZAP1_X1*; (D) Nucleic acid protein sequence of *DAZAP1_X2*; (E) Nucleic acid protein sequence of *DAZAP1_X3*. ATG: start codon, “*”: stop codon.

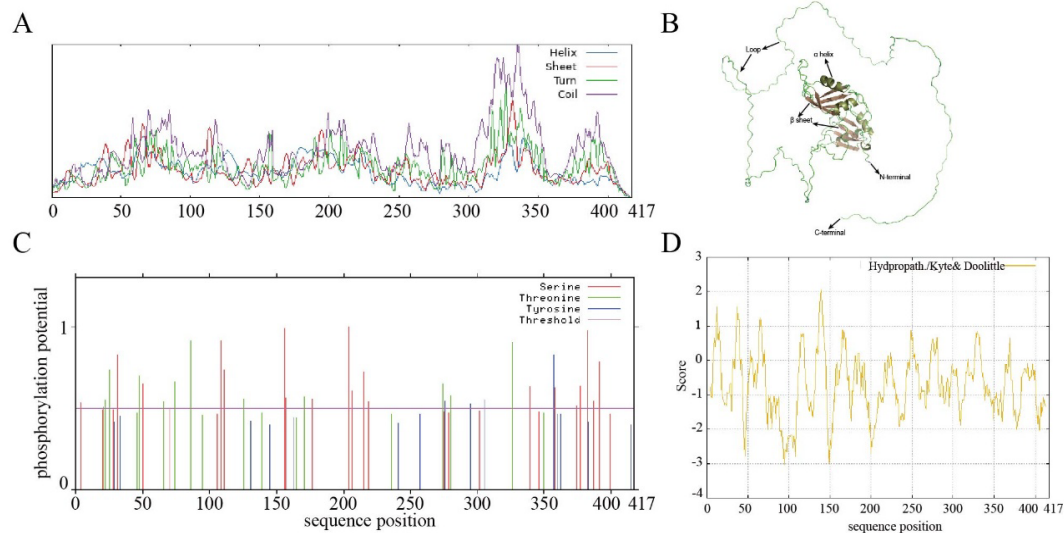


Figure 3. Structure of *DAZAP1* protein. (A) Secondary structure; (B) Tertiary structure; (C) Phosphorylation site; (D) Hydrophobicity.

Conservation and evolutionary relationship of *DAZAP1_X2* in multiple species

Comparison of amino acid sequences across 13 species classified pigs, along with goats, sheep, and cattle, as artiodactyls; cats, lions, and tigers as carnivores; mice and rats as rodents; humans, gorillas, and cynomolgus monkeys as primates; and horses and ungulates (Figure 4A). The similarity between the BMI pig and the sequences from cattle (*Bos taurus* XP_024850355), goat (*Capra hircus* XP_017905929), donkey (*Equus asinus* XP_044609611), horse

(*Equus caballus* XP_023502104), cat (*Felis catus* XP_023098767), human (*Homo sapiens* EAW69502), cynomolgus monkeys (*Macaca fascicularis* XP_045235029), mouse (*Mus musculus* NP_573451), lion (*Panthera leo* XP_042783741), tiger (*Panthera tigris* XP_042833498), chimpanzee (*Pan troglodytes* XP_054529711), and rat (*Rattus norvegicus* NP_001020913) was above 90% (Figure 4B), indicating a relatively conserved sequence. All branches conformed to the systematic classification standards. Analysis of 14 conserved domains of mammalian DAZAP1 using Weblogo revealed a high level of conservation across species (Figure 4C). In summary, the *DAZAP1* gene in BMI pigs exhibited significant conservation and sequence homology with other mammals throughout evolution.

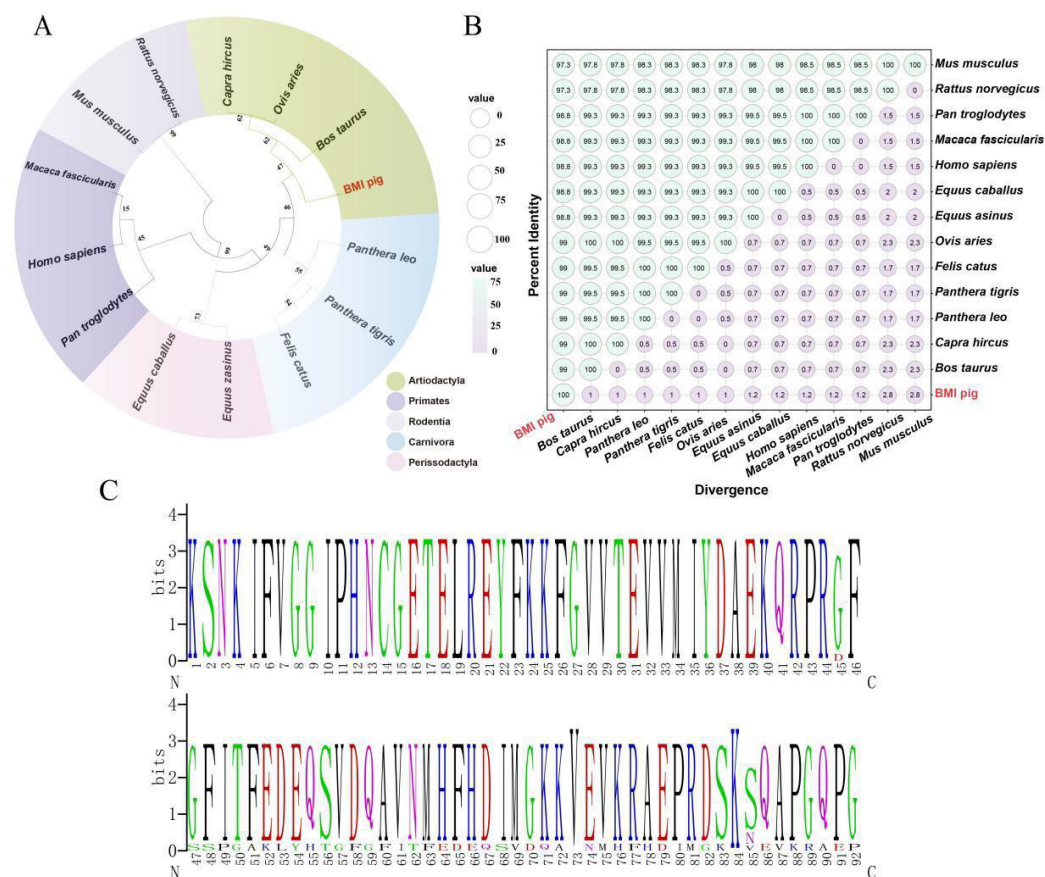


Figure 4. Comparison of the phylogenetic tree, similarity, and conserved domains of the amino acid sequences of DAZAP1. (A) Phylogenetic tree; (B) Similarity; (C) Conserved domain.

Protein-protein interactions of DAZAP1_X2

The pig DAZAP1 protein had interactions with various proteins including dickkopf-like acrosomal protein 1 (DKKL1), the outer dense fiber of sperm tails 4 (ODF4), RUN and SH3 domain-containing 1 (RUSC1), DEAD-box helicase 24 (DDX24), zinc finger CCCH-type containing 14 (ZC3H14), spermatogenesis associated 9 (SPATA9), cleavage stimulation factor subunit 2 (CSTF2), serine and arginine-rich splicing factor 1 (SRSF1), RNA binding motif protein 5 (RBM5), sperm acrosome-associated 5 (SPACA5), sperm as-sociated antigen 6 (SPAG6) and so on (Figure 5A). These proteins focused on various biological processes including protein secretion, mRNA translation, spermatogenesis, and cell growth and division. KEGG pathways enrichment analysis identified four significant pathways ($P < 0.05$) for these proteins, with notable enrichment in spliceosomes, nuclear transport, mRNA surveillance pathways, and monitoring proteins (Figure 5B). Moreover, GO enrichment analysis revealed that these proteins were predominantly associated with spermatogenesis, male gamete generation, sexual

reproduction, gamete generation, germ cell development, regulation of meiotic nuclear division I, 9+2 flagellar movement, sperm flagellum, RNA binding, and mRNA metabolic processes (Figure 5C). Correlation analysis using transcriptome sequencing data revealed a significant correlation ($P < 0.05$) between DAZAP1 and DKKL1, DNAH17, DNPEP, OAZ3, PCBP3, SPAZA19, CSTF2T, RBM5, TNPO2, CBY2, SRSF5, CETN (Figure 5D).

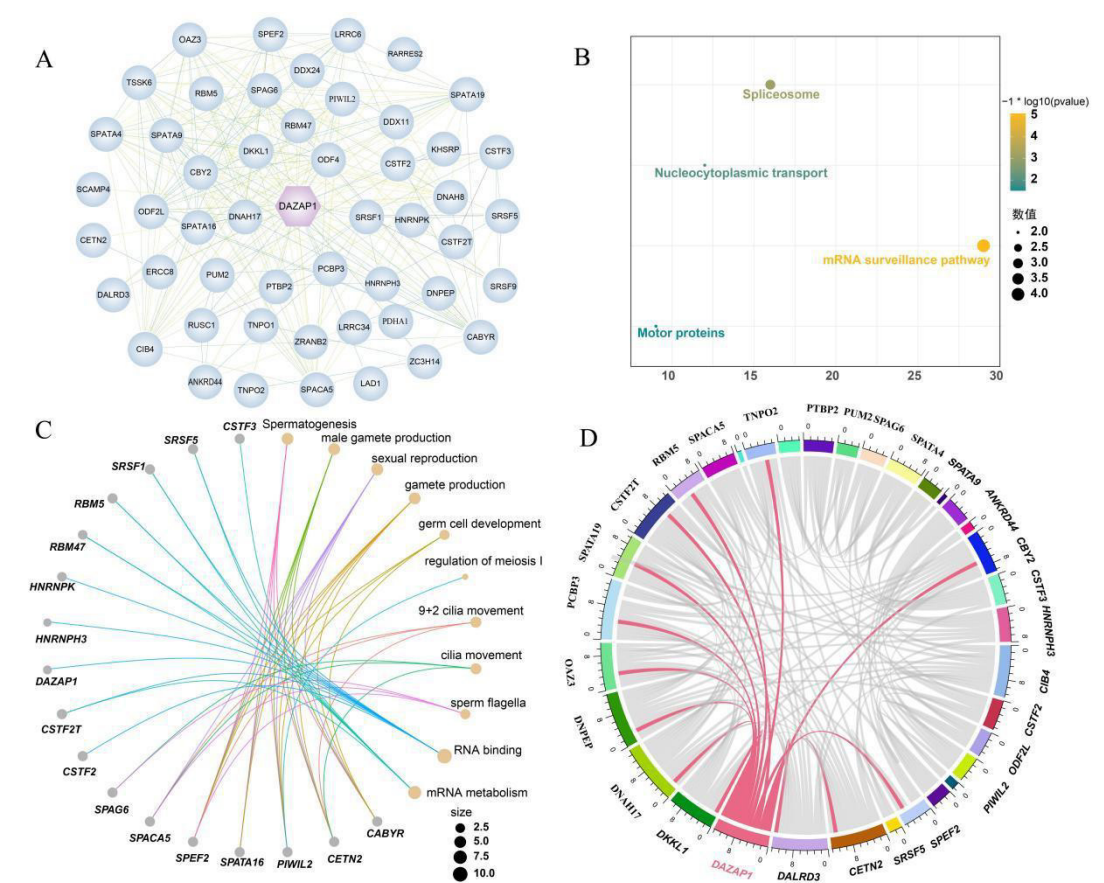


Figure 5. Protein interaction analysis. (A) protein- protein interacting protein network; (B) KEGG enrichment analysis diagram of interacting proteins; (C) GO enrichment analysis diagram of interacting proteins; (D) Correlation chord diagram.

Functional annotation and ceRNA regulatory network of DAZAP1_X2

Functional annotation of *DAZAP1* identified 18 Gene Ontology (GO) terms, including biological processes such as fibroblast proliferation, cell differentiation, spermatogenesis, and maternal placenta development. Cell components included the nucleoplasm, ribonucleoprotein complex, protein domain complex, cytoplasm, cell nucleus, and male germ cell nucleus. Molecular functions encompassed RNA stem-loop binding, poly(G) binding, poly(U) RNA binding, protein binding, mRNA 3' UTR binding, RNA binding, and nucleic acid binding. ceRNA regulatory network showed that *DAZAP1* was targeted regulation by ssc-miR-107, ssc-miR-127, and ssc-miR-1343. Furthermore, two lncRNAs (ENSSSCG00000042919.1 and ENSSSCG00000047764.1) competitively bind to ssc-miR-107 targeting *DAZAP1*, while one lncRNA (ENSSSCG00000042671.1) competitively binds to ssc-miR-127, and seven lncRNAs (ENSSSCG000000051212.1, ENSSSCG000000034567.2, ENSSSCG000000041056.1, ENSSSCG000000041612.1, ENSSSCG000000042507.1, ENSSSCG000000046512.1, ENSSSCG000000049720.1) competitively bind to ssc-miR-1343, affecting *DAZAP1* regulation (Figure 6).

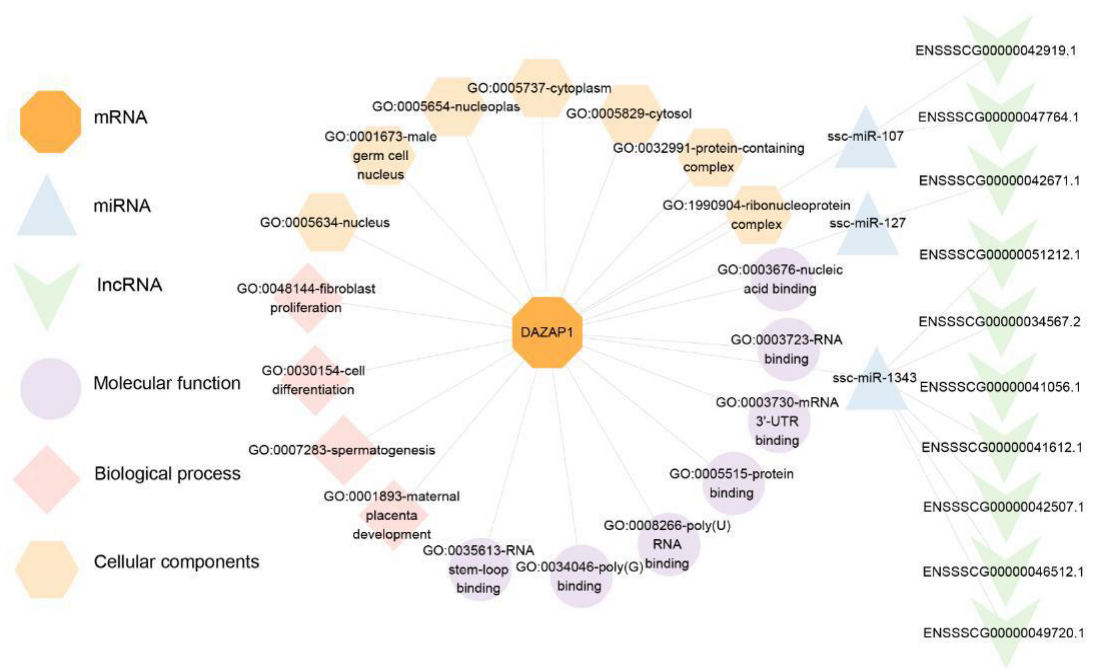


Figure 6. Functional annotation and ceRNA regulatory network of *DAZAP1*.

Expression profiles in multiple tissues analysis of *DAZAP1_X2*

To investigate the expression of the *DAZAP1_X2* across various tissues, relative quantification using qPCR was performed with *GAPDH* as the housekeeping gene. The $2^{-\Delta\Delta Ct}$ method was utilized for relative quantification. The results revealed varying degrees of expression across the 15 tissues examined. *DAZAP1* mRNA showed relatively higher expression in glandular organs such as the prostate, urethral bulb, testis, and epididymis. Moderate expression was observed in the duodenum, colon, heart, spleen, lungs, and kidneys, while relatively lower expression was detected in other tissues (Figure 7).

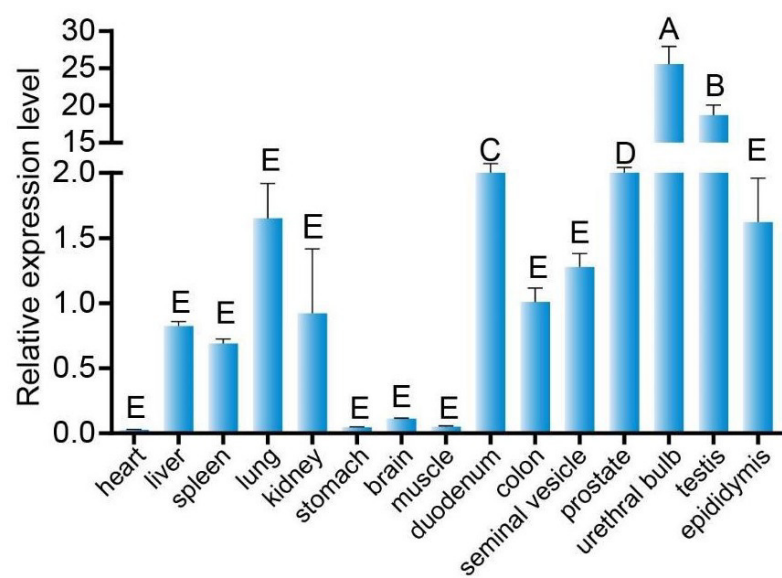


Figure 7. Multi-tissue expression profile of BMI *DAZAP1* gene. Different capital letters on the bars indicate significant differences ($P < 0.01$).

Subcellular localization analysis of DAZAP1_X2

The Swine Testis (ST) cells were transfected with pEGFP-C1-*DAZAP1* and pEGFP-C1 control plasmids. After 24 hours of expression, green fluorescence was observed. Subsequently, the cells were processed with MitoTracker Red (mitochondrial staining) and Hoechst33342 (nuclear staining), followed by observation under an inverted fluorescence microscope. In ST cells transfected with the pEGFP-C1-*DAZAP1* plasmid, green fluorescence was predominantly localized in the cell nucleus (Figure 8). To further analyze the *DAZAP1* protein localization, we merged the green, red, and blue fluorescence in the same field of view, suggesting that *DAZAP1* is mainly located in the cell nucleus.

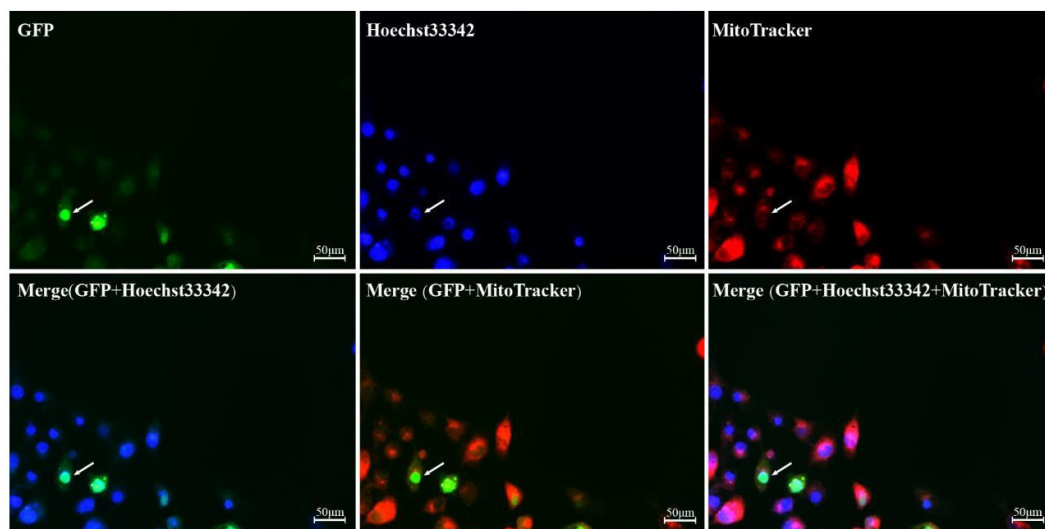


Figure 8. Subcellular localization of *DAZAP1* protein.

Discussion

Alternative splicing is the process by which exons of primary gene transcripts (pre-mRNA) are spliced in different ways, resulting in the production of mRNA and protein variants with different structures and functions (Blencowe, 2006). Proteins translated from selectively spliced mRNAs typically exhibit differences in amino acid sequences and biological functions. Selective splicing coordinates physiologically meaningful changes in the expression of protein isoforms, representing a key mechanism in the generation of complex proteomes in multicellular organisms (Lee and Rio, 2015). In this study, the testicular tissue of BMI pigs was sequenced using second- and third-generation transcriptome sequencing technologies. The coding sequences of three splice isoforms of the *DAZAP1* gene, namely *DAZAP1_X1*, *DAZAP1_X2*, and *DAZAP1_X3*, were obtained using molecular cloning techniques. Among these isoforms, *DAZAP1_X2* is the most complete, while *DAZAP1_X1* lacks the 7th exon and the 5' UTR region, and *DAZAP1_X3* lacks the 7th exon, indicating an exon skipping splicing pattern. Evolutionary analysis indicated that BMI *DAZAP1* shared a close evolutionary relationship with cattle, sheep, and goats. Its similarity to mammals such as rats, mice, horses, and donkeys exceeds 90%, suggesting a high conservation and sequence homology of BMI *DAZAP1* with other mammals in evolution. Protein domain analysis revealed that the *DAZAP1* protein contained an RNA recognition motif (RBM) domain of serine/arginine protein kinase expressed exclusively in the nuclei of male germ cells.

Protein interaction network analysis revealed that *DAZAP1* interacted with several proteins, including *DKKL1*, *ODF4*, *DDX24*, *ZC3H14*, *SPATA9*, *SPACA5*, and *SPAG6*. *DKKL1* has been independently identified as a distant homolog of the Dickkopf (*Dkk*) protein family, known to modulate wnt/ β -catenin signaling. It is specifically expressed in testicular tissue, primarily

localized in spermatogonia and round spermatids. In patients with azoospermia, the expression of *DKKL1* in testicular tissue is significantly reduced or absent compared to that in normal testicular tissue (Yan et al., 2012). The proper shape and movement of sperm tails are crucial for fertilization. Outer dense fiber 4 (*ODF4*) is specifically expressed in the testes. Knockout of *ODF4* in male mice results in curved and abnormally motile sperm tails, leading to male infertility (Ito et al., 2023). The Glu-Ala-Asp motif, characteristic of the DEAD-box family, represents an ATPase-active RNA helicase. *DDX24* plays a significant role in various biological processes, including embryonic development, spermatogenesis, cell growth, and tumorigenesis (Chen et al., 2022). Expression of *SPATA9* varies in patients with spermatogenic arrest, with not being expressed in patients arrested at the spermatogonia and primary spermatocyte stages (Cheng et al., 2003). Circular RNAs (circRNAs) are natural byproducts of eukaryotic transcription and RNA processing and have emerged as key regulatory factors in physiology and human diseases. *ZC3H14* binds to the spliceosome, interacting with the 3' and 5' exons-introns boundaries of circular exons, and forms dimers to promote circRNA biogenesis. Knockout mice lacking *ZC3H14* exhibit disrupted spermatogenesis and reduced levels of testicular circRNAs (Rha et al., 2017). *SPACA5*, also known as lysozyme-like protein 5 (*LYZL5*), encodes a putative protein of 159 amino acids, and is located on the X chromosome at p11.23. The *SPACA5* protein is observed only in the elongated spermatids of male spermatogenesis (Chen et al., 2006). The *SPAG6* encodes an axonemal protein maintaining the structure and function of the axoneme (Neilson et al., 1999). Variants of *SPAG6* are associated with multi-systemic disorders involving cilia and flagella (Wu et al., 2020). The biallelic mutations in *SPAG6* are associated with primary ciliary dyskinesia (PCD) (Neilson et al., 1999). Moreover, we found that these proteins were primarily involved in spermatogenesis, male gamete production, sexual reproduction, 9+2 flagellar movement, and meiosis. The discovery of these proteins potentially interacting with DAZAP1 provided new insights into understanding the function of this gene.

Competing endogenous RNAs (ceRNAs) are a class of short non-coding RNAs that regulate gene expression post-transcriptionally by inhibiting translation or promoting mRNA degradation (Cai et al., 2009). Among them, long non-coding RNAs (lncRNAs), mRNAs, and other RNAs function as natural miRNA sponges, inhibiting target mRNAs. lncRNAs regulate protein levels and influence cellular processes by competing with miRNAs (Poliseno et al., 2010). Functional annotation of DAZAP1 in this study revealed its involvement primarily in spermatogenesis and cell differentiation processes. Target gene prediction suggested that *DAZAP1* was regulated by three miRNAs: ssc-miR-107, ssc-miR-127, and ssc-miR-1343. Unlike most miRNAs that target complementary sequences in the 3'UTR region, certain members of the miR-107 family uniquely target the coding sequence of mRNA, thereby inhibiting translation without interference. MicroRNAs (miRNAs) play a crucial role in adipogenesis, with miR-127 promoting proliferation by enhancing the cell cycle and inhibiting differentiation, thus reducing lipid accumulation (Gao et al., 2019). miR-1343 increases the expression of pluripotency genes by inhibiting OTX2, thereby maintaining the pluripotency of porcine embryonic stem cells (Xie et al., 2019). Additionally, miR-1343 overexpressed in capacitated sperm which undergo critical physiological changes necessary for fertilization, suggesting that miR-1343 can serve as a therapeutic target to alleviate the impact of infertility (Sahoo et al., 2021). Therefore, the predicted association of miR-1343 with male infertility underscored its potential as a therapeutic target for early diagnosis and prognosis of male infertility.

DAZAP1 is a ubiquitously expressed hnRNP (heterogeneous nuclear ribonucleoprotein) involved in RNA transcription, splicing, and transport (Hsu et al., 2008; Yang et al., 2009). The DAZAP1 is expressed in various human and mouse tissues (Reijo et al., 2000), with high expression observed in the testes, and lower in other tissues (Dai et al., 2001). In BMI pigs, *DAZAP1* was highly expressed in the testes and seminal vesicles, while its expression was relatively low in other tissues, especially in the heart, stomach, brain, and muscles. In male reproduction, accessory glands including the prostate, seminal vesicles, and urethral bulb (Gilany et al., 2015). Seminal plasma is rich in tissue-specific proteins and peptides that bind directly to sperm (Manjunath et al., 2007). The transport of large molecules, such as proteins or RNA, between the nucleus and cytoplasm occurs through the nuclear pore complex (NPC).

The NPC is a selective and bidirectional transport channel for various cargo molecules, recognized by nuclear transport proteins via nuclear export signals (Kim et al., 2017). During spermatogenesis, DAZAP1 exhibits dynamic subcellular localization primarily in the nuclei of late spermatocytes and round spermatids (Lin and Yen 2006). In porcine testicular cells, DAZAP1 is primarily localized in the nuclear, with minimal presence in the cytoplasm. In summary, *DAZAP1* deletion or abnormalities can lead to spermatogenesis defects, highlighting its crucial role in spermatogenesis.

Conclusion

The integration of second- and third-generation transcriptome sequencing offered a detailed analysis of *DAZAP1*'s transcriptional regulation, identifying three distinct transcripts. Additionally, 50 interacting proteins were identified, primarily focusing on protein secretion, mRNA translation, spermatogenesis, and cell growth and division. Significant associations were found between DAZAP1 and genes such as *DKKL1*, *DNAH17*, *DNPEP*, *OAZ3*, *PCBP3*, *SPAZA19*, *CSTF2T*, *RBM5*, *TNPO2*, *CBY2*, *SRSF5*, and *CETN*. *DAZAP1* was associated with 18 GOs and was targeted by three miRNAs. *DAZAP1* was highly expressed in the gonads and accessory glands. These results enhance our insight into *DAZAP1*, contributing to the clarification of its role and mechanism in spermatogenesis within BMI pigs.

Acknowledgements

This work was supported by funds from the Basic Research Project of Shanxi Province (no. 202403021222371 to X.Z.), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (no. 2022L570 to X.Z.), the 2023 Lyuliang Development Zone's Plan to Introduce High-Level Scientific and Technological Talents (no. 2023RC26 to X.Z.), Agricultural Joint Project of the Yunnan Province Science and Technology Department (no. 202401BD070001-004 to J.L.H.), Basic Research Key Project of Yunnan Province (202501AS070041 to J.L.H.), the National Natural Science Foundation of China (no. 32060733 to J.L.H.).

Data availability statement

Research data is only available upon request.

References

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Židek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024;630(8016):493-500. <http://doi.org/10.1038/s41586-024-07487-w>. PMID:38718835.
- Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell*. 2006;126(1):37-47. <http://doi.org/10.1016/j.cell.2006.06.023>. PMID:16839875.
- Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4(6):1633-49. <http://doi.org/10.1002/pmic.200300771>. PMID:15174133.
- Cai Y, Yu X, Hu S, Yu J. A brief review on the mechanisms of miRNA regulation. *Genomics Proteomics Bioinformatics*. 2009;7(4):147-54. [http://doi.org/10.1016/S1672-0229\(08\)60044-3](http://doi.org/10.1016/S1672-0229(08)60044-3). PMID:20172487.
- Chen L, Lin X, Lei Y, Xu X, Zhou Q, Chen Y, Liu H, Jiang J, Yang Y, Zheng F, Wu B. Aerobic glycolysis enhances HBx-initiated hepatocellular carcinogenesis via NF-κBp65/HK2 signalling. *J Exp Clin Cancer Res*. 2022;41(1):329. <http://doi.org/10.1186/s13046-022-02531-x>. PMID:36411480.

- Chen YT, Iseli C, Venditti CA, Old LJ, Simpson AJ, Jongeneel CV. Identification of a new cancer/testis gene family, CT47, among expressed multicopy genes on the human X chromosome. *Genes Chromosomes Cancer*. 2006;45(4):392-400. <http://doi.org/10.1002/gcc.20298>. PMID:16382448.
- Cheng CY, Mruk DD. The biology of spermatogenesis: the past, present and future. *Philos Trans R Soc Lond B Biol Sci*. 2010;365(1546):1459-63. <http://doi.org/10.1098/rstb.2010.0024>. PMID:20403863.
- Cheng LJ, Li JM, Chen J, Ge YH, Yu ZR, Han DS, Zhou ZM, Sha JH. NYD-SP16, a novel gene associated with spermatogenesis of human testis. *Biol Reprod*. 2003;68(1):190-8. <http://doi.org/10.1095/biolreprod.102.004242>. PMID:12493713.
- Cho C, Willis WD, Goulding EH, Jung-Ha H, Choi YC, Hecht NB, Eddy EM. Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet*. 2001;28(1):82-6. <http://doi.org/10.1038/ng0501-82>. PMID:11326282.
- Dai T, Vera Y, Salido EC, Yen PH. Characterization of the mouse Dazap1 gene encoding an RNA-binding protein that interacts with infertility factors DAZ and DAZL. *BMC Genomics*. 2001;2(1):6. <http://doi.org/10.1186/1471-2164-2-6>. PMID:11604102.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. <http://doi.org/10.1093/bioinformatics/bts635>. PMID:23104886.
- Gao Y, Wang Y, Chen X, Peng Y, Chen F, He Y, Pang W, Yang G, Yu T. MiR-127 attenuates adipogenesis by targeting MAPK4 and HOXC6 in porcine adipocytes. *J Cell Physiol*. 2019;234(12):21838-50. <http://doi.org/10.1002/jcp.28660>. PMID:31206669.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the expasy server. In: John M. Walker: The Proteomics Protocols Handbook, Humana Press, 2005. pp. 571-607. <http://doi.org/10.1385/1-59259-890-0:571>.
- Gilany K, Minai-Tehrani A, Savadi-Shiraz E, Rezadoost H, Lakpour N. Exploring the human seminal plasma proteome: an unexplored gold mine of biomarker for male infertility and male reproduction disorder. *J Reprod Infertil*. 2015;16(2):61-71. PMID:25927022.
- Hao F, Kale S, Dimitrov S, Hayes JJ. Unraveling linker histone interactions in nucleosomes. *Curr Opin Struct Biol*. 2021;71:87-93. <http://doi.org/10.1016/j.sbi.2021.06.001>. PMID:34246862.
- Hao F, Murphy KJ, Kujirai T, Kamo N, Kato J, Koyama M, Okamoto A, Hayashi G, Kurumizaka H, Hayes JJ. Acetylation-modulated communication between the H3 N-terminal tail domain and the intrinsically disordered H1 C-terminal domain. *Nucleic Acids Res*. 2020;48(20):11510-20. <http://doi.org/10.1093/nar/gkaa949>. PMID:33125082.
- Hsu LC, Chen HY, Lin YW, Chu WC, Lin MJ, Yan YT, Yen PH. DAZAP1, an hnRNP protein, is required for normal growth and spermatogenesis in mice. *RNA*. 2008;14(9):1814-22. <http://doi.org/10.1261/rna.1152808>. PMID:18669443.
- Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. *Hum Immunol*. 2021;82(11):801-11. <http://doi.org/10.1016/j.humimm.2021.02.012>. PMID:33745759.
- Huo JL, Zhang LQ, Zhang X, Wu XW, Ye XH, Sun YH, Cheng WM, Yang K, Pan WR, Zeng YZ. Genome-wide single nucleotide polymorphism array and whole-genome sequencing reveal the inbreeding progression of Banna minipig inbred line. *Anim Genet*. 2022;53(1):146-51. <http://doi.org/10.1111/age.13149>. PMID:34658041.
- Ito C, Makino T, Mutoh T, Kikkawa M, Toshimori K. The association of ODF4 with AK1 and AK2 in mice is essential for fertility through its contribution to flagellar shape. *Sci Rep*. 2023;13(1):2969. <http://doi.org/10.1038/s41598-023-28177-z>. PMID:36804949.
- Kim YH, Han ME, Oh SO. The molecular mechanism for nuclear transport and its application. *Anat Cell Biol*. 2017;50(2):77-85. <http://doi.org/10.5115/acb.2017.50.2.77>. PMID:28713609.
- Leblond CP, Clermont Y. Spermiogenesis of rat, mouse, hamster and guinea pig as revealed by the periodic acid-fuchsin sulfurous acid technique. *Am J Anat*. 1952;90(2):167-215. <http://doi.org/10.1002/aja.1000900202>. PMID:14923625.
- Lee Y, Rio DC. Mechanisms and regulation of alternative Pre-mRNA splicing. *Annu Rev Biochem*. 2015;84(1):291-323. <http://doi.org/10.1146/annurev-biochem-060614-034316>. PMID:25784052.
- Lin YT, Yen PH. A novel nucleocytoplasmic shuttling sequence of DAZAP1, a testis-abundant RNA-binding protein. *RNA*. 2006;12(8):1486-93. <http://doi.org/10.1261/rna.42206>. PMID:16772659.
- Manjunath P, Bergeron A, Lefebvre J, Fan J. Seminal plasma proteins: functions and interaction with protective agents during semen preservation. *Soc Reprod Fertil Suppl*. 2007;65:217-28. PMID:17644964.

- Morton S, Yang HT, Moleleki N, Campbell DG, Cohen P, Rousseau S. Phosphorylation of the ARE-binding protein DAZAP1 by ERK2 induces its dissociation from DAZ. *Biochem J.* 2006;399(2):265-73. <http://doi.org/10.1042/BJ20060681>. PMID:16848763.
- Neilson LI, Schneider PA, Van Deerlin PG, Kiriakidou M, Driscoll DA, Pellegrini MC, Millinder S, Yamamoto KK, French CK, Strauss JF 3rd. cDNA cloning and characterization of a human sperm antigen (SPAG6) with homology to the product of the *Chlamydomonas* PF16 locus. *Genomics.* 1999;60(3):272-80. <http://doi.org/10.1006/geno.1999.5914>. PMID:10493827.
- Neto FT, Bach PV, Najari BB, Li PS, Goldstein M. Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol.* 2016;59:10-26. <http://doi.org/10.1016/j.semcdb.2016.04.009>. PMID:27143445.
- Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature.* 2010;465(7301):1033-8. <http://doi.org/10.1038/nature09144>. PMID:20577206.
- Rao X, Huang X, Zhou Z, Lin X. An improvement of the $2^{-(\Delta\Delta CT)}$ method for quantitative real-time polymerase chain reaction data analysis. *Biostat Bioinforma Biomath.* 2013;3(3):71-85. PMID:25558171.
- Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, de la Chapelle A, Silber S, Page DC. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet.* 1995;10(4):383-93. <http://doi.org/10.1038/ng0895-383>. PMID:7670487.
- Reijo RA, Dorfman DM, Slee R, Renshaw AA, Loughlin KR, Cooke H, Page DC. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol Reprod.* 2000;63(5):1490-6. <http://doi.org/10.1095/biolreprod63.5.1490>. PMID:11058556.
- Rha J, Jones SK, Fidler J, Banerjee A, Leung SW, Morris KJ, Wong JC, Inglis GAS, Shapiro L, Deng Q, Cutler AA, Hanif AM, Pardue MT, Schaffer A, Seyfried NT, Moberg KH, Bassell GJ, Escayg A, García PS, Corbett AH. The RNA-binding protein, ZC3H14, is required for proper poly(A) tail length control, expression of synaptic proteins, and brain function in mice. *Hum Mol Genet.* 2017;26(19):3663-81. <http://doi.org/10.1093/hmg/ddx248>. PMID:28666327.
- Sahoo B, Choudhary RK, Sharma P, Choudhary S, Gupta MK. Significance and Relevance of Spermatozoal RNAs to Male Fertility in Livestock. *Front Genet.* 2021;12:768196. <http://doi.org/10.3389/fgene.2021.768196>. PMID:34956322.
- Sapay N, Guermeur Y, Deléage G. Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics.* 2006;7(1):255. <http://doi.org/10.1186/1471-2105-7-255>. PMID:16704727.
- Sasaki K, Ono M, Takabe K, Suzuki A, Kurihara Y. Specific intron-dependent loading of DAZAP1 onto the *cox6c* transcript suppresses pre-mRNA splicing efficacy and induces cell growth retardation. *Gene.* 2018;657:1-8. <http://doi.org/10.1016/j.gene.2018.03.005>. PMID:29505834.
- Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, Rozen S, Dinulos MB, Disteche CM, Page DC. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet.* 1996;14(3):292-9. <http://doi.org/10.1038/ng1196-292>. PMID:8896558.
- Smith RW, Anderson RC, Smith JW, Brook M, Richardson WA, Gray NK. DAZAP1, an RNA-binding protein required for development and spermatogenesis, can regulate mRNA translation. *RNA.* 2011;17(7):1282-95. <http://doi.org/10.1261/rna.2717711>. PMID:21576381.
- Tsui S, Dai T, Roettger S, Schempp W, Salido EC, Yen PH. Identification of two novel proteins that interact with germ-cell-specific RNA-binding proteins DAZ and DAZL1. *Genomics.* 2000;65(3):266-73. <http://doi.org/10.1006/geno.2000.6169>. PMID:10857750.
- Vera Y, Dai T, Hikim AP, Lue Y, Salido EC, Swerdloff RS, Yen PH. Deleted in azoospermia associated protein 1 shuttles between nucleus and cytoplasm during normal germ cell maturation. *J Androl.* 2002;23(5):622-8. <http://doi.org/10.1002/j.1939-4640.2002.tb02303.x>. PMID:12185095.
- Wang P, Zhang X, Huo H, Li W, Liu Z, Wang L, Li L, Sun YH, Huo J. Transcriptomic analysis of testis and epididymis tissues from Banna mini-pig inbred line boars with single-molecule long-read sequencing. *Biol Reprod.* 2023;108(3):465-78. <http://doi.org/10.1093/biolre/ioac216>. PMID:36477198.
- Wang Q, Guo Y, Wang W, Liu B, Yang G, Xu Z, Li J, Liu Z. RNA binding protein DAZAP1 promotes HCC progression and regulates ferroptosis by interacting with SLC7A11 mRNA. *Exp Cell Res.* 2021;399(1):112453. <http://doi.org/10.1016/j.yexcr.2020.112453>. PMID:33358859.

- Wu H, Wang J, Cheng H, Gao Y, Liu W, Zhang Z, Jiang H, Li W, Zhu F, Lv M, Liu C, Tan Q, Zhang X, Wang C, Ni X, Chen Y, Song B, Zhou P, Wei Z, Zhang F, He X, Cao Y. Patients with severe asthenoteratospermia carrying SPAG6 or RSPH3 mutations have a positive pregnancy outcome following intracytoplasmic sperm injection. *J Assist Reprod Genet.* 2020;37(4):829-40. <http://doi.org/10.1007/s10815-020-01721-w>. PMID:32124190.
- Xie Y, Cao H, Zhang Z, Zhang S, Wang H. Molecular network of miR-1343 regulates the pluripotency of porcine pluripotent stem cells via repressing OTX2 expression. *RNA Biol.* 2019;16(1):82-92. <http://doi.org/10.1080/15476286.2018.1559688>. PMID:30567463.
- Yan Q, Wu X, Chen C, Diao R, Lai Y, Huang J, Chen J, Yu Z, Gui Y, Tang A, Cai Z. Developmental expression and function of DKKL1/Dkk1 in humans and mice. *Reprod Biol Endocrinol.* 2012;10(1):51. <http://doi.org/10.1186/1477-7827-10-51>. PMID:22817830.
- Yang HT, Peggie M, Cohen P, Rousseau S. DAZAP1 interacts via its RNA-recognition motifs with the C-termini of other RNA-binding proteins. *Biochem Biophys Res Commun.* 2009;380(3):705-9. <http://doi.org/10.1016/j.bbrc.2009.01.166>. PMID:19285026.
- You X, Chen Q, Yuan D, Zhang C, Zhao H. Common markers of testicular Sertoli cells. *Expert Rev Mol Diagn.* 2021;21(6):613-26. <http://doi.org/10.1080/14737159.2021.1924060>. PMID:33945376.

Author contributions

XZ: Formal analysis, Writing – original draft, Funding acquisition; HH: Writing – original draft; HL: Writing – original draft; YL: Methodology, Data curation; FQ: Data curation; CL: Methodology, Data curation; JH: Supervision, Writing – review & editing.