



OPEN

Deciphering the role of alternative splicing as a potential regulator in fat-tail development of sheep: a comprehensive RNA-seq based study

Mohammad Reza Bakhtiarizadeh

Although research on alternative splicing (AS) has been widely conducted in mammals, no study has investigated the splicing profiles of genes involved in fat-tail formation in sheep. Here, for the first time, a comprehensive study was designed to investigate the profile of AS events and their involvement in fat-tail development of sheep. In total, 45 RNA-Seq samples related to seven different studies, which have compared the fat-tailed vs thin-tailed sheep breeds, were analyzed. Two independent tools, rMATS and Whippet, along with a set of stringent filters were applied to identify differential AS (DAS) events between the breeds per each study. Only DAS events that were detected by both tools as well as in at least three datasets with the same Δ PSI trend (percent spliced in), were considered as the final high-confidence set of DAS genes. Final results revealed 130 DAS skipped exon events (69 negative and 61 positive Δ PSI) belonged to 124 genes. Functional enrichment analysis highlighted the importance of the genes in the underlying molecular mechanisms of fat metabolism. Moreover, protein–protein interaction network analysis revealed that DAS genes are significantly connected. Of DAS genes, five transcription factors were found that were enriched in the biological process associated with lipid metabolism like “Fat Cell Differentiation”. Further investigations of the findings along with a comprehensive literature review provided a reliable list of candidate genes that may potentially contribute to fat-tail formation including HSD11B1, SIRT2, STRN3 and TCF7L2. Based on the results, it can be stated that the AS patterns may have evolved, during the evolution of sheep breeds, as another layer of regulation to contribute to biological complexity by reprogramming the gene regulatory networks. This study provided the theoretical basis of the molecular mechanisms behind the sheep fat-tail development in terms of AS.

Sheep breeds can be divided into three major groups in terms of tail morphology, fat-tailed, thin-tailed and fat-rumped¹. Fat-tailed breeds constitute ~25% of the world's sheep breeds^{1,2}. It is believed that fat-tailed breeds have evolved through natural and artificial selection, after initial domestication of their wild ancestors, for adaptation to harsh and difficult environmental conditions (such as drought and periodic food shortages)³. However, fat-tail has lost its advantages in the current commercial production systems, as it influences mating, locomotion and carcass characteristics; requires more energy to produce and currently consumers prefer low-fat food^{3–6}.

Investigation of genetic factors affecting complex traits, such as fat metabolism, based on monitoring the underlying molecular mechanisms is of major interest among scientists. Understanding these mechanisms can dramatically accelerate genetic improvement. In this regard, numerous efforts have been made to better understand the genetic background of fat deposition in tail of sheep, such as^{6–10}, which can be applied to evolve new and improved sheep breeds. A literature review showed that most of these studies focused on comparative transcriptome/variant analysis between fat- and thin-tailed sheep breeds^{5–7,9–16}. In this context, our team previously conducted a comprehensive gene expression meta-analysis of six independent RNA-Seq datasets that compared transcriptome profiles of fat- and thin-tailed sheep breeds and a core set of differentially expressed genes (DEGs) as well as several pathways associated with lipid metabolism were reported¹⁷.

Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran. email: mrbakhtiar@ut.ac.ir

Based on the Ensemble database (release 106), transcript per gene ratio in sheep genome is ~ 1.7 (20,137 known protein coding genes and 34,569 known transcript), indicating an average of approximately two transcripts for each gene caused by alternative splicing (AS). AS is one of the fundamental regulatory features of eukaryotic cells that enables a gene to encode multiple transcripts and can lead to the enhanced diversity of transcriptome and proteome¹⁸. It is reported that AS can be more important than gene expression profiling in facilitating rapid adaptive divergence within short time periods¹⁹. Increasing evidence has highlighted the importance and the potential roles of AS in various tissue types, developmental stages and diseases²⁰. For example, it has been reported that 92–94% of the human genes are alternatively spliced, as most of them are tissue-specific²¹. Moreover, differential AS events have been identified between different breeds such as cattle²² and chicken²³. For instance, the identified differential AS (DAS) genes in muscle, brain and skin tissues of mules were significantly enriched in "muscle contraction" pathway²⁴. Miao et al., reported 356 DAS events in comparison between muscle tissue of small-tailed Han and Dorset sheep breeds, which were associated with variations in muscle development²⁵. Also, 411 DAS events were found in adipose tissue between small-tailed Han and Dorset sheep breeds²⁶. These reports reinforced the fact that AS events have diverged rapidly during vertebrate evolution²⁷. In this context, AS has been reported as an important source of species-specific divergence²⁸. These reports have resulted in a growing interest in understanding of AS events in various biological process as well as among different breeds. Although several studies have investigated the transcriptome differences between fat- and thin-tailed sheep breeds^{6,7,9–11,16}, comprehensive alterations of AS and their biological implications in fat deposition in tail of sheep has not been elucidated, yet. Fortunately, a large number of public RNA-Seq datasets provide an opportunity to investigate role of AS in gene regulation in fat-tail formation of sheep, which can be overlooked in standard transcriptome analysis. To fill this knowledge gap, a genome-wide analysis was performed to identify DAS events between fat-tail tissue of fat- and thin-tailed sheep breeds based on seven independent RNA-Seq studies. Furthermore, functional enrichment analysis was performed for a better understanding of the biological functions of the genes with identified DAS. This study tried to provide insights into the molecular basis that may contribute to fat-tail formation in sheep breeds.

Results

RNA-seq data analysis

In total, over 1.133 billion reads were analyzed, which were belonged to 22 fat- (totaling 556 million reads) and 23 thin-tailed (totaling 577 million reads) samples. On average, each sample contained more than 25 million reads, ranging from 7.5 to 50 million reads. After trimming, approximately two million reads were removed, indicating the high quality of the datasets. Out of a total of 1.131 billion clean reads, 1.036 billion reads (~ 92.6%) were successfully aligned to the genome. Of the aligned reads, 87.5% were uniquely mapped, while the remaining 12.5% were mapped to multiple locations in the genome. Alignment rates for individual samples ranged from 77.1 to 98.5%. The Supplementary File S1 contains a summary of the RNA-Seq datasets used in this study along with mapping results.

Identification of DAS events by rMATS

DAS events were identified by comparing the fat- and thin-tailed samples in each dataset. Using rMATS, 4,575 DAS were detected including 3773 skipping exon (SE), 667 mutually exclusive exon (MXE), 62 alternative 5' splice site (5SS), 51 alternative 3' splice site (3SS) and 22 intron retention (IR) events. The largest number of DAS events (1655) were identified in Study 7, followed by Study 4, Study 3, Study 6, Study 2, Study 1 and Study 5. The complete list of the identified DAS events by both tools is provided in Supplementary File S2. Figure 1 represents the number and percentage of different DAS events in the seven studies by the tools. The highest number of common DAS events was identified by rMATS between study 4 and 7 (150); and study 7 and 3 (144). Also, 69 DAS events were found to be common among studies 7, 4 and 3. In Fig. 2, UpSet plot represents the intersection between the DAS events from different studies based on the rMATS results.

After filtering out the DAS events that were common in less than three studies or showed ambiguous trend, 55 DAS (16 positive and 39 negative Δ PSI, as negative value indicates the exon is more skipped in fat-tailed group) corresponding to 49 genes were found. Of these, 50 DAS were assigned to SE and five to MXE. Moreover, out of 55 events, 30, 21 and four DAS were common in four, three and five studies. Totally, 14 and six significant biological processes and KEGG pathways were identified. Functional enrichment analysis revealed that the most of the enriched biological processes/pathways were related to mRNA splicing. Furthermore, some lipid metabolism-associated pathways were enriched including "Regulation of lipolysis in adipocytes" and "Lipid Catabolic Process", which indicates the importance of AS in fat deposition in tail of sheep. Two important genes involved in lipolysis were LIPE and PLIN1 that showed negative Δ PSI (exon is most skipped in fat-tailed vs thin-tailed breeds). A list of the common DAS events across different studies identified by rMATS and their functional enrichment analysis results are provided in Supplementary File S3.

Identification of DAS events by Whippet

On the other hand, 11,569 DAS were identified by Whippet tool including 10,118 SE, 607 MXE, 404 3SS, 248 5SS and 192 RI events. Study 5 included the largest number of DAS events (2199), followed by study 7, study 6, study 4, study 3, study 1 and study 2. SE and RI events were the most and least abundant DAS in both tools, respectively. The highest number of common DAS events was identified by Whippet between study 4 and 7 (146); and study 6 and 7 (131). Similar to the results of rMATS, the highest number of DAS events among three studies were found among studies 7, 4 and 3 (35) (Fig. 3).

In total, 215 DAS events (92 positive and 123 negative Δ PSI, as negative value indicates the exon is more skipped in fat-tailed group) met the stringent filtering criteria (be common in at least three studies with the same

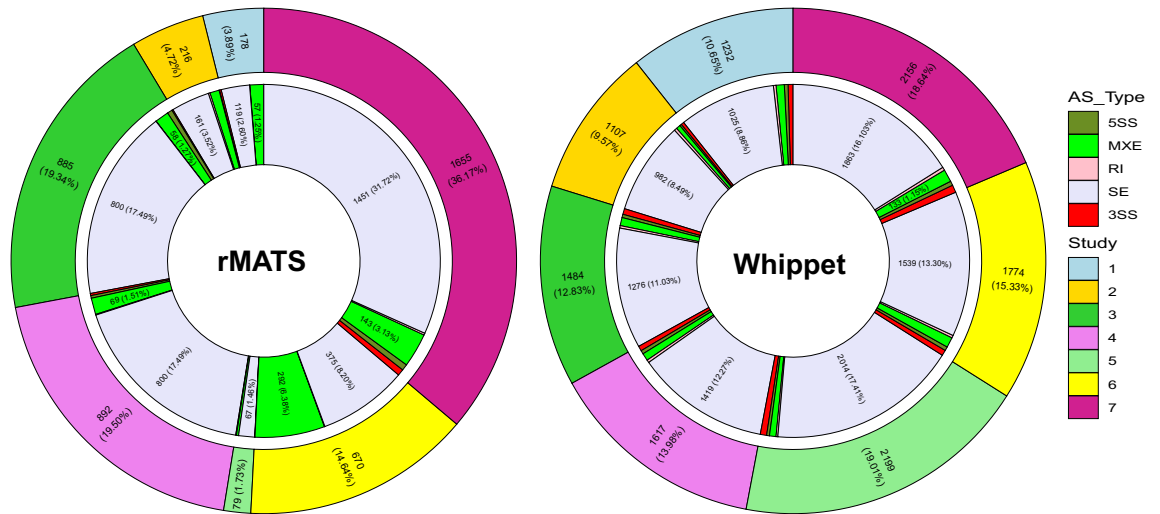


Figure 1. Number and percentage of the identified AS events by rMATS and Whippet tools in different studies. AS_Type indicates the different AS events including: skipped exon (SE), intron retention (IR), alternative 5' splice site (5SS), alternative 3' splice site (3SS) and mutually exclusive exon (MXE). Inner and outer circles indicate AS events and studies, respectively. Study 1 (PRJNA508203), Study 2 (PRJNA699984), Study 3 (PRJNA432669), Study 4 (PRJNA745517), Study 5 (PRJNA517348), Study 6 (PRJNA598581) and Study 7 (PRJNA792697).

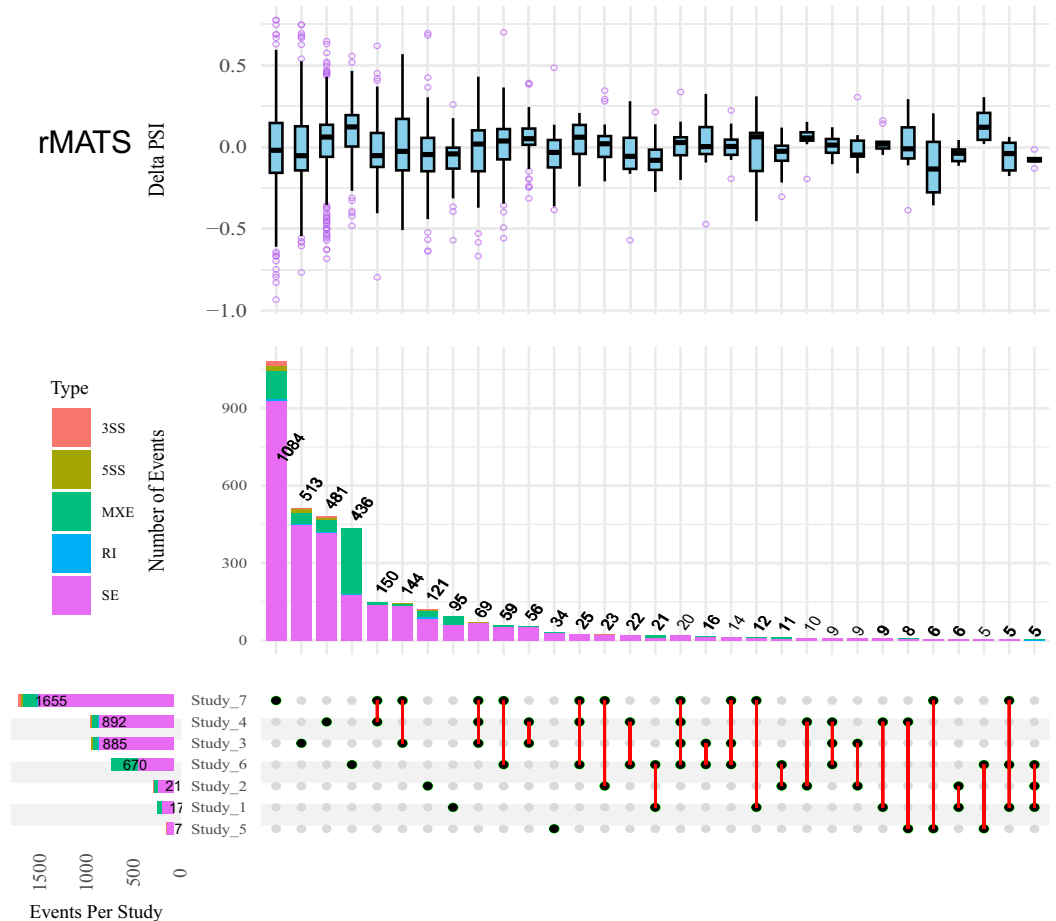


Figure 2. UpSet plot of the intersection of DAS events among the different studies identified by rMATS. The box plot in the top section indicates the distribution of Δ PSI in each group. The bar plot in the middle section indicates the number of event types shared by specific studies. The horizontal bar chart at the bottom left represents the total number of events found in each study. An observation threshold of five is used to exclude low-count subsets.

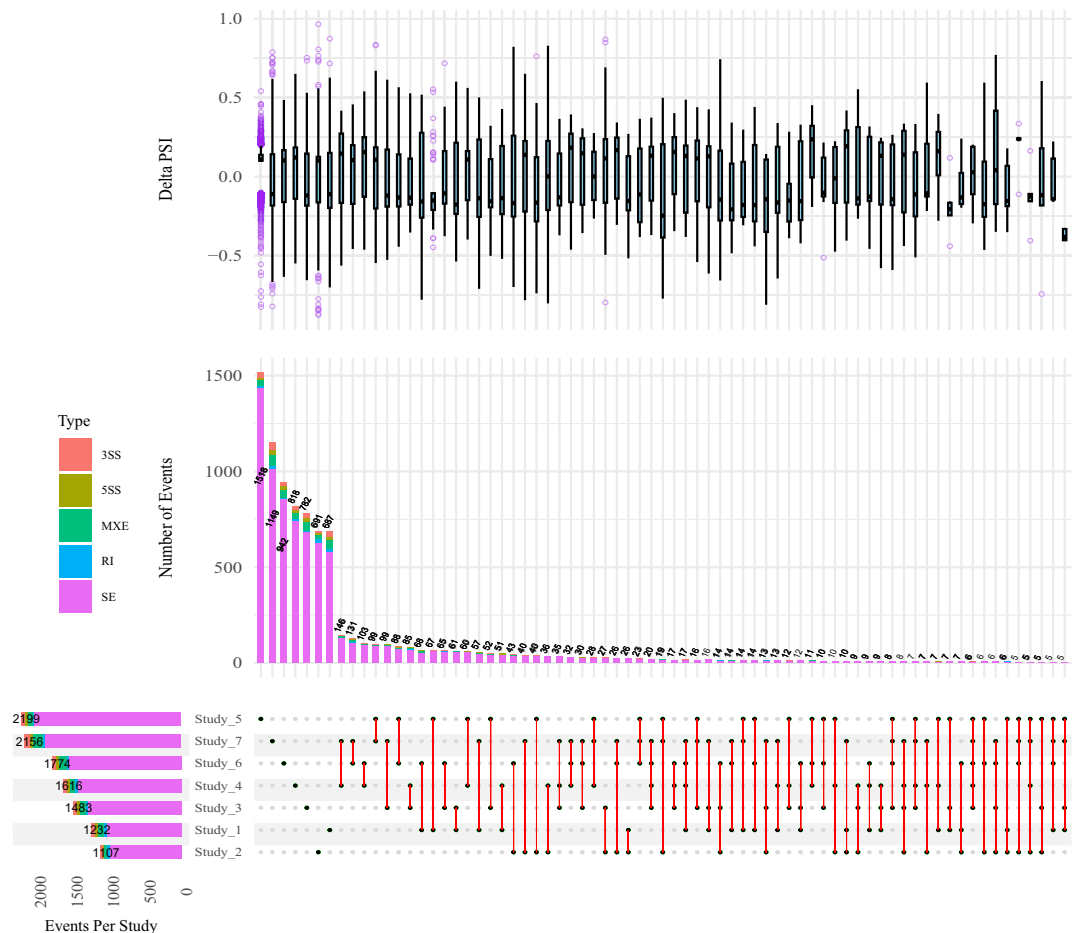


Figure 3. UpSet plot of the common DAS events among the different studies identified by Whippet. The box plot indicates the distribution of Δ PSI in each group. The bar plot in the middle section indicates the number of event types shared by specific studies. The horizontal bar chart at the bottom left represents the total number of events found in each study. An observation threshold of five is used to exclude low-count subsets.

Δ PSI trend), and were related to 189 genes. Most of DAS events were SE (176), followed by MXE (19), 3SS (11), RI (5) and 5SS (4). Out of 215 events, 5, 28, 75 and 107 DAS were common in six, five, four and three studies. The results of functional enrichment analysis showed 19 significant biological processes terms, which were related to “Sodium Ion Transport” as well as “mRNA splicing”. Complete results of the common DAS events across different studies identified by Whippet and their functional enrichment analysis can be found in Supplementary File S4.

Common DAS events between both tools

The DAS events that showed the same Δ PSI trend (up or down) in most of the datasets (at least three datasets) and detected by both tools, were finally considered as stable DAS events. Following these filtering criteria, 130 DAS events corresponding to 124 genes were detected, which all were related to SE class. Study 7 was characterized by the highest number of DAS events (110), followed by Study 6 (98), Study 3 (93), Study 4 (61), Study 5 (47), Study 2 (41) and Study 1 (32). Altogether, 61 statistically significant SE events with positive Δ PSI in fat-tailed breed and 69 SE events with decreased inclusion levels in thin-tailed breeds were identified by both tools and in at least in three datasets. For example, Δ PSI of SIRT2 gene for SE event is negative and indicates that the exon is more skipped in fat-tailed vs thin-tailed breeds (Supplementary File S5). A Circos plot of the genomic landscape of the common DAS along with heatmap related to their Δ PSI is presented in Fig. 4.

To have a comprehensive understanding of the biological functions of these genes, functional enrichment analysis was performed and 22 biological processes and two KEGG pathways were significantly enriched (Supplementary File S5). Of these, “mRNA Splicing” and “Spliceosome” were the most enriched biological process and pathways, respectively. Moreover, enrichment analysis indicated the involvement of the genes in “Lipid Catabolic Process”. Two important genes related to this term were HSD11B1 and SIRT2 (Fig. 5). The exons of the both genes were more skipped in the fat-tailed breeds, as SIRT2 showed a lower exon inclusion in the fat-tailed breeds (average Δ PSI in three studies = -0.25) in comparison to HSD11B1 (average Δ PSI in three studies = -0.12). Three important significant pathways, at the p-value level (<0.05), were found to be related to lipid metabolism including “AMPK signaling pathway”, “MAPK signaling pathway” and “Wnt signaling pathway”. Furthermore, several genes were found that were annotated to fat metabolism associated functions such as CEPT1, PPP2R5C,

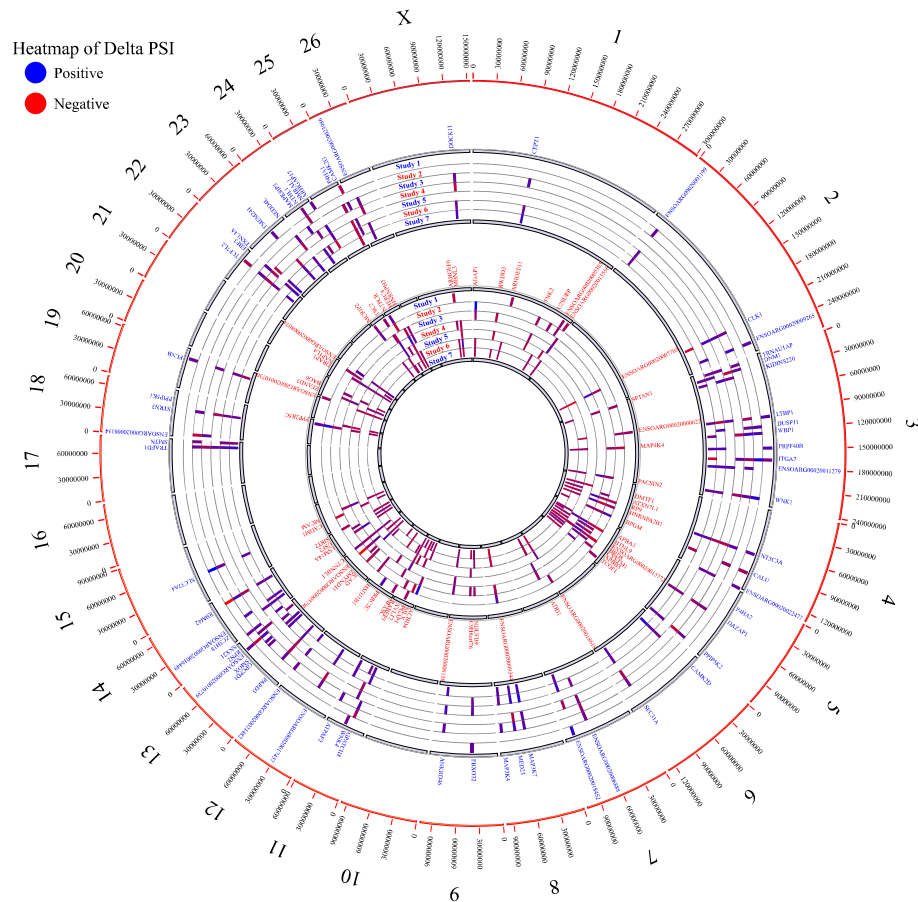


Figure 4. Circos plot of the genomic distributions of the final DAS, common DAS events between both tools. The outermost ring shows chromosome numbers. The position of the DAS with positive (blue) and negative Δ PSI (red) are shown in the first and third inner white rings, respectively. The heatmap in the second and fourth inner rings display the Δ PSI of each DAS with positive (blue) and negative Δ PSI (red) in different studies, respectively (studies from the outside: Study 1, Study 2, Study 3, Study 4, Study 5, Study 6 and Study 7). Positive Δ PSI indicates greater inclusion and negative Δ PSI equal decreased inclusion in fat-tailed breeds.

GPCPD1, FLNB, PGAP3, STRN3, TCF7L2 and ITGA7. The Sankey plot in Fig. 6 shows the associations of these genes with the relevant functional terms.

On the other hand, a significant PPI network (p -value < 0.001), of the DAS genes, contains 105 nodes and 54 edges was constructed using STRING database. Further annotating the DAS genes with AnimalTFDB database (version 3.0)²⁹ revealed that five of 124 genes were TFs including USF2, EBF3, MLX, TCF7L2 and DMTF1. GO analysis of these genes results in 43 significant terms including “Fat Cell Differentiation” (Supplementary File S5).

Discussion

Recent studies have shown that AS is more frequent and complex than initially thought, and AS is likely to be involved in shaping the variations among species and breeds. Previous studies have mainly focused on the transcriptional level to explore the genetic mechanisms behind the fat-tail formation in sheep^{6,7,9–11,16}, leaving the other important post-transcriptional regulation mechanisms, such as AS, that may also contribute to the evolution of biological complexity and phenotypic diversity in sheep fat-tail. Here, taking advantage of RNA-Seq method, a large number of samples related to seven studies, with the same purpose, were applied to elucidate the potential roles of AS regulation in forming the fat-tail of sheep.

In total, number of the identified DAS by Whippet (11,569 events) was more than two times than rMATS (4575 events), which is in line with the previous study that reported the same ratio³⁰. This finding can be attributed to the different algorithm behind the software. Whippet uses splice graph approach to represent each transcript and does not generate p -values, which can capture more DAS events³¹. On the other hand, rMATS uses a hierarchical framework to generate inclusion levels based on the number of read counts spanning the splice junctions and provides adjusted p -values for each DAS event³². The annotated sheep genome has 20,137 protein coding genes (ENSEMBL release 106), 18,221 was multi-exon genes, which are exposed to AS. Out of these, 2344 and 4545 genes were found to be DAS using rMATS and Whippet, respectively. This large number of DAS events suggests that AS can be considered as a major mechanism for regulating gene expression and protein diversity in fat-tail of sheep. Nonetheless considering each of these tools separately can be led to false positive results. Hence,

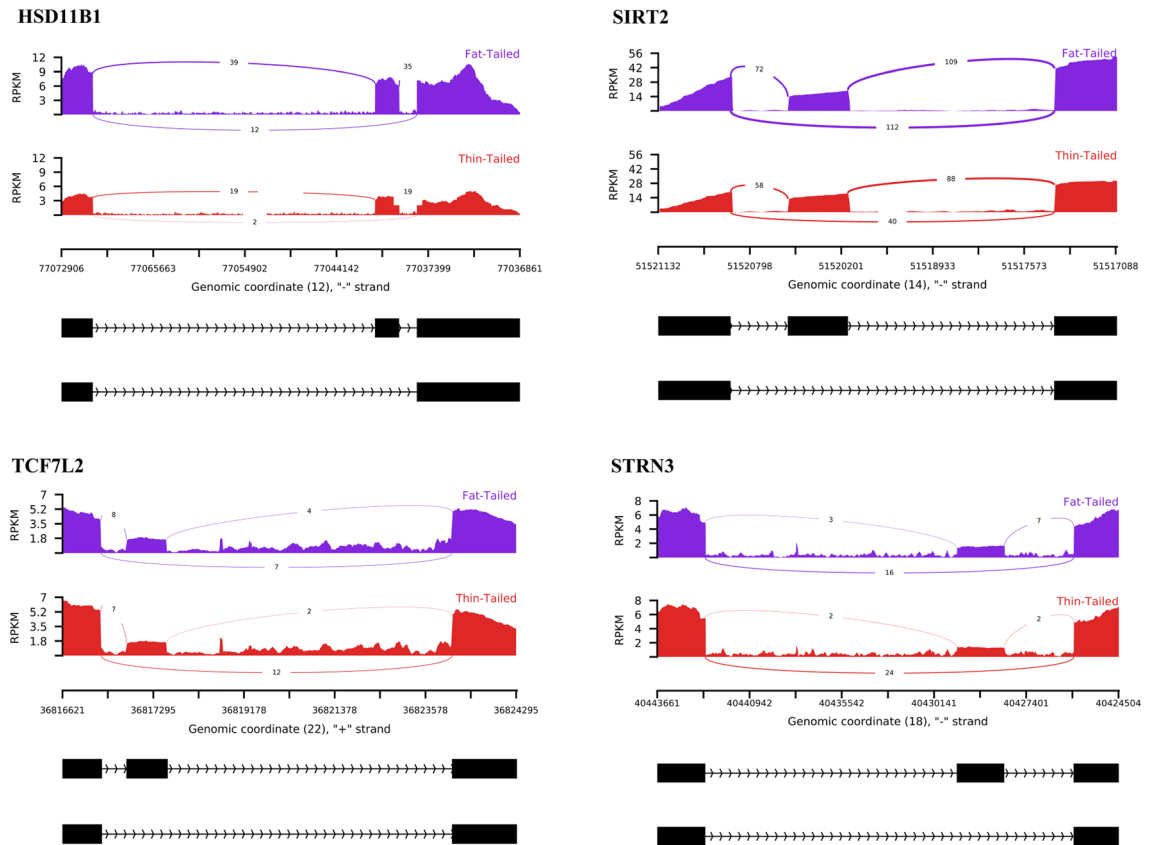


Figure 5. Sashimi plot of SE event within the HSD11B1, SIRT2, TCF7L2 and STRN3 genes. The blue and red tracks represent fat- and thin-tailed breeds, respectively. To provide the flexibility to compare the two groups of samples, biological replications of each group are averaged according to $-group$ parameter of `rmats2sashimiplot` tool. Numbers on curved lines indicate average counts of the biological replications of all the samples related to different studies. The average number of junction-spanning reads of each group is clearly higher in fat-tailed than thin-tailed breeds in HSD11B1 and SIRT2 genes and vice versa for the other two genes. The bottom black tracks show the genomic localization of splicing events.

multiple filters with stringent double verification (intersection of rMATS and Whippet) were implemented in the bioinformatics pipeline and permitted the identification of 130 DAS events (SE) between fat- and thin-tailed sheep breeds. This concurs well with the global AS pattern of mammals, as SE event is the most common type of AS (~40%)³³. For instance, SE was the most enriched event in muscle²⁵ and adipose²⁶ tissue of small-tailed Han than Dorset sheep breeds. However, the other AS events were not found, which is attributable to the applied stringent filtering approach. In other words, it is preferred to reduce the rate of false positive results to find the most reliable DAS genes, as may give rise to false negative findings. In a previous study, rMATS was used to investigate the splicing profiles of genes in adipose tissue of two breeds of sheep: polled Dorset and small tail Han²⁶. Totally, 411 DAS events related to 364 genes were identified, of which eight genes were common with the current study results. This low consistency, can be largely attributed to a high rate of the false positive results in Miao et al. study, as they applied one AS event detection tool on one RNA-Seq dataset. Furthermore, they used adipose tissue, as fat-tail tissue was used in the present study.

If AS plays an important role in regulating the fat deposition in the tail of sheep, it can be expected to find some DAS genes which are involved in lipid metabolism related processes. First, to compare the identified DAS genes with the reported DEGs in the previous cohort studies, a comprehensive literature review was performed. It is of interest to note that 50 DAS genes were reported as DEGs in these studies^{6,7,9–11,16}. These genes are listed in Supplementary File S5. Moreover, Functional analysis of the final DAS genes revealed DAS genes involvement in the relevant biological/pathways to fat metabolism including “Lipid Catabolic Process”, “Fat Cell Differentiation”, “AMPK signaling pathway”, “MAPK signaling pathway” and “Wnt signaling pathway”. Involvement of Wnt pathway³⁴ and MAPK pathway³⁵ in adipogenesis; and AMPK pathway³⁶ in lipid metabolism have been demonstrated. Of DAS genes, HSD11B1, SIRT2, STRN3 and TCF7L2 were annotated to be directly involved in lipid metabolism (Fig. 5). In contrast to HSD11B1 and SIRT2 genes, the exons of STRN3 and TCF7L2 were more skipped in the thin-tailed breeds. HSD11B1 as a member of family of short-chain dehydrogenases/reductases, encodes hydroxysteroid (11-beta) dehydrogenase type 1 that catalyze the conversion of inactive hormone cortisol into its active form cortisol³⁷. Role of HSD11B1 in the accumulation of abdominal fat has been demonstrated³⁸. It is reported that HSD11B1 can affect glucose uptake by adipocytes and leads to obesity in pig³⁹. Overexpression of this gene in adipocytes of mouse models is reported to be associated with the development

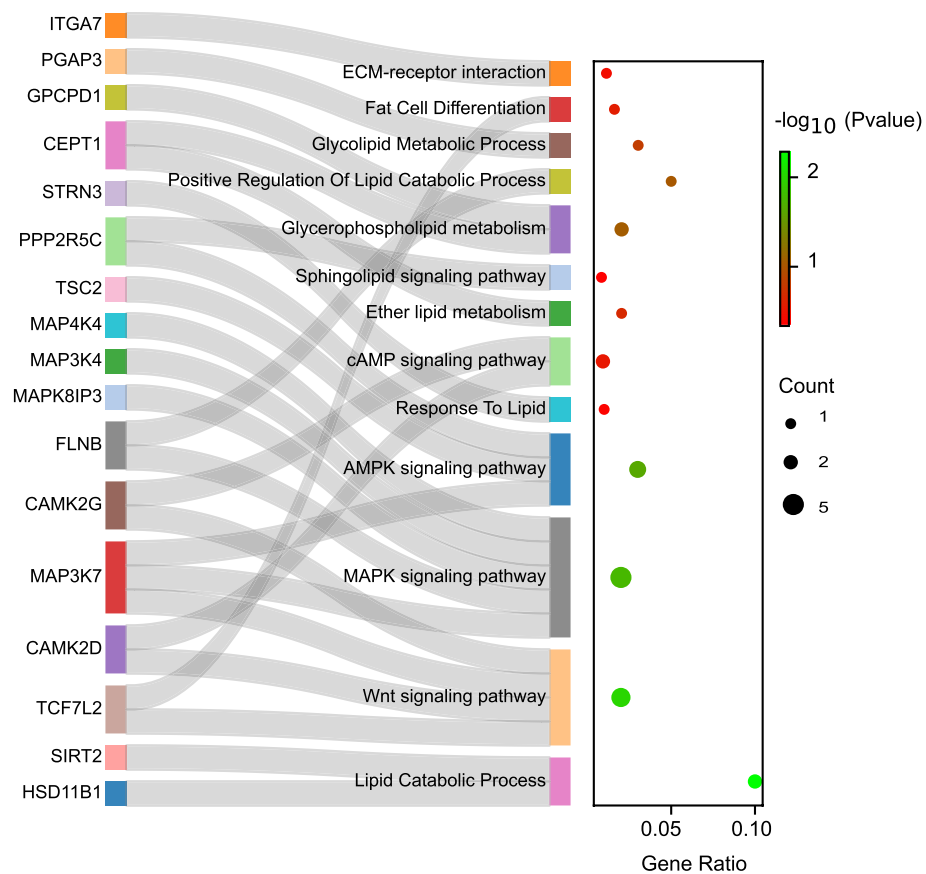


Figure 6. Sankey plot of the association between the DAS genes and the functional terms related to fat metabolism. The gene ratio indicates the number of DAS genes associated with the GO term or KEGG pathway divided by the total number of genes of genome that are annotated to the relevant functional terms. Count represents the number of DAS genes associated with the GO term or KEGG pathway.

of central obesity⁴⁰. In this regard, HSD11B1 knockout mice were protected against the development of obesity after high-fat diet exposure⁴¹. HSD11B1 was identified as differentially expressed gene according to my previous study on the same datasets⁶, which was up-regulated in fat-tailed breeds. Also, this gene was reported as DEG in the other previous studies that investigated the expression pattern of genes in fat-tail tissue of fat- and thin-tailed sheep breeds^{8,16}. Interestingly, DAS has been previously confirmed in adipose tissue of sheep for HSD11B1²⁶, which reinforce its importance in fat-tail tissue biology. More details of HSD11B1 gene and its important roles in development of obesity and body fat distribution has been comprehensively reviewed by Nascimento et al.³⁷. SIRT2 is a member of the family of sirtuins, a family of NAD⁺-dependent deacetylases, that regulate some important biological processes like lipid metabolism⁴². Potential function of this gene in fat deposition in pig is reported⁴³. It has been shown that SIRT2 is involved in the differentiation of rodent adipocyte precursors⁴⁴. Moreover, down regulation of SIRT2 in human stem cells promoted the expansion of visceral adipose tissue⁴⁵. In a recent study, SIRT2 was reported as a candidate that affect the lipid metabolism in cattle²². Recently, as a new idea for the regulation of lipid metabolism, sirtuins are discussed as regulators in lipid autophagy⁴². All these findings lead us to conclude that sirtuins family members, specially SIRT2, can be considered as novel candidates in modulating the gene expression in fa-tail of sheep. It is worthy of notice that SIRT2 was identified as DEG in the previous studies that compared the fat-tail tissue of fat- and thin-tailed sheep breeds^{10,16}.

The other important DAS gene in the present study was TCF7L2 that encodes a TF that plays a key role in the Wnt signaling pathway, which is involved in many biological processes such as lipid metabolism, regulating body mass and glucose metabolism⁴⁶. This gene is known to be an important regulator of triglyceride accumulation through Wnt signaling pathway³⁶. It is well known that TCF7L2 is involved in adipocyte differentiation and adipogenesis and its knockout promotes adipogenesis⁴⁷ and impairs adipocyte differentiation⁴⁸. Moreover, it is demonstrated that Wnt signal pathway plays important roles in mouse obesity⁴⁹. The STRN3 gene encodes a protein called striatin-3, which is a member of the striatin family of proteins that are involved in various cellular functions, such as signal transduction, cytoskeleton organization, cell motility⁵⁰ and lipid metabolism⁵¹. Since, these genes were identified as DAS in the present study and based on the evidence from the previous studies described above, it is possible to candidate them as potential regulators of sheep fat-tail development.

Looking specifically at the other functional terms of the DAS genes, “AMPK signaling pathway”, “MAPK signaling pathway”, “cAMP signaling pathway” and “Wnt signaling pathway”, revealed important genes involved in fat metabolism (Fig. 6). Two members of calcium/calmodulin-dependent protein kinases II (CAMK2) family including CAMK2D and KAMK2G were identified as DAS genes. A recent study has shown that CAMK2 regulates adipocyte lipolysis and its activity is increased in obese mice adipose tissue, which makes it a critical component of metabolic regulation in obesity⁵². Vital role of MAP3K4 (as an important member of the MAPK signaling pathway) in the hepatic lipogenesis of non-alcoholic fatty liver disease has been well described⁵³. The expression of other important member of MAPK in the final DAS genes, MAP4K4, is reported to be increased during obesity, which makes it an important negative regulator of adipose lipogenesis in metabolic disease⁵⁴. FLNB, as a member of the filamin family, is suggested to be strongly related to the differentiation of fat cells in Chickens⁵⁵. These findings along with the previous studies reinforced that these genes may be involved in morphological diversity of fat-tail among the sheep breeds. The results of the current study tempted us to speculate that AS (specially SE events) may be induced during the evolution of fat- and thin-tailed sheep breeds to increase the complexity of the transcriptome as well as the diversity of the proteome. Hence, understanding the patterns and functions of AS is crucial for revealing the molecular basis of sheep fat-tail development.

Conclusion

While understanding the molecular pathways of fat deposition in tail of sheep is crucial, little attention has been given to AS. The present study is the first comprehensive attempt to compare the patterns of AS in fat-tail of sheep breeds with different tail shape. I was particularly interested in genes that displayed significant changes at the AS event level, because such changes are often overlooked in gene expression analysis. Applying several miscellaneous methods allowed us to obtain stringent and robust results for the identification of splicing events. Of the final set of DAS events, 69 and 61 events showed negative and positive Δ PSI, respectively, which several genes were related to fat metabolism and were suggested as possible candidates involved in the sheep fat-tail development including HSD11B1, SIRT2, STRN3 and TCF7L2. Taken together, the results of the present study indicate that it is worth shifting our focus to post-transcriptional regulations mediated by various molecular mechanisms, such as AS, to reveal mechanisms underlying fat deposition. This study provides the first evidence that AS is an important mechanism contributing to regulation of gene expression of fat deposition in tail of sheep breeds.

Materials and methods

RNA-seq datasets

Based on a comprehensive literature review, seven independent RNA-seq datasets of the matched fat- and thin-tailed sheep breeds were obtained from the GEO database, which met the requirements for the analysis. All the samples were collected from fat-tail tissue of the male gender and were sequenced as paired-end (Table 1). A same bioinformatic pipeline was applied to analyze each RNA-Seq dataset, explained as follow.

Quality control, trimming and aligning of reads

Quality of the reads, before and after trimming, were assessed using FastQC (v0.11.5) tool⁵⁶. Trimming of poor-quality reads/bases and adapter sequences was performed using adaptive trimming algorithm of Trimmomatic (v0.38) software (TRAILING: 20 and MAXINFO: 80: 0.8)⁵⁷. Then, clean reads were aligned to the sheep genome (Oar Rambouillet v1.0) based on STAR software (v2.7.9a)⁵⁸. The STAR index was generated based on the Ensembl ovine GTF file (rambouillet_v1.0.106) with a sjdbOverhang of 100. Two-pass approach was applied to align the reads as follow; (1) In the first round, all the samples of each dataset were aligned to the genome and the splice junction outputs were retrieved. (2) All the junction files of the previous round were used as annotated junctions for the second round to improve the accuracy and sensitivity of read alignment and detect more spliced reads mapped to novel junctions.

Identification of AS events

Generally, five types of AS are commonly studied, including SE, IR, 5SS, 3SS and MXE. There are several software developed to identify DAS event based on these five events, as AS events abundance are compared across two or more groups. Benchmark studies suggested that to decrease the tool-specific false positives as well as to

Fat (Nu)*	Thin (Nu)*	Length of reads	Accession number
Lori (4)	Zel (4)	150	PRJNA508203 ⁶
Han-Fat (3)	Han-Thin (3)	75	PRJNA699984 ⁸
Lanzhou (3)	Han (3)	150	PRJNA432669 ¹⁰
DHH (3)	DDH (3)	150	PRJNA745517 ⁹
Hulun-Fat (3)	Hulun-Thin (3)	101	PRJNA517348 ¹⁶
Ghezel (3)	Zel (4)	150	PRJNA598581 ¹¹
Hu (3)	Tibetan (3)	150	PRJNA792697 ⁷

Table 1. List of the used RNA-seq datasets that compared fat- and thin-tailed sheep breeds. *Name of the fat- and thin-tailed breeds (Number of biological replications). Of the nine sheep breeds, Lori, Zel and Ghezel breeds were from Iran and the other sheep breeds were from China.

increases the validity of downstream analyses, it is better to overlap the results of two or more software⁵⁹. Here, to minimize the potential false positive results, two well-known tools with different background rMATS³² and Whippet³¹ were applied and overlapping results of both tools were taken into consideration. To this end, identification and quantification of AS events (SE, IR, 5SS, 3SS and MXE) as well as detection of DAS events between the two groups (fat- and thin-tailed breeds) were performed by both tools for each dataset, separately. Then, all the results of the seven datasets were integrated and DAS events that were observed in at least three datasets with the same Δ PSI trend (percent spliced in), up or down, were considered as reliable DAS, which may be potential candidate involved in fat-tail formation in sheep. It is worth emphasizing that a stable DAS have to be detected by both tools. For example, if a DAS was detected in three studies by rMATS with the same Δ PSI trend, it has to be identified in at least one study by Whippet. This level of stringency was employed to increase the robustness and reliability of the results.

rMATS (replicate multivariate analysis of transcript splicing)

Using a hierarchical framework, rMATS simultaneously model the variability among replicates and compute the inclusion level from two-groups RNA-seq data to identify the five classical AS events. rMATS quantify AS events as PSI, which is a ratio (the number of reads specific to exon inclusion isoform divided by the sum of reads specific to exon inclusion and exclusion isoforms) that indicates how efficiently reads of interest are spliced into transcripts. In this regard, Δ PSI is the difference between the PSI values of the two groups under study. Furthermore, P-value and false discovery rate (FDR) is calculated based on likelihood-ratio test to assess the statistical significance of the Δ PSI between two groups³². Here, the read alignments obtained from the above-mentioned two-pass approach was provided as input to rMATS (v4.1.2). Both exon–exon junction reads and the reads mapped to the exon body were used in the analyses (JCEC). In each dataset, AS events exhibiting a significant difference between fat- and thin-tailed samples with an FDR < 0.05 were considered as DAS.

Whippet

Based on gene annotation and genome files, Whippet builds contiguous splice graphs (CSGs) to represent each transcript. Then, to quantify splicing events, reads directly align to the CSGs. Whippet considers splicing events as nodes, as each node corresponds to an exonic region of the gene, and the incoming and outgoing edges to each node define the set of reads supporting its inclusion and exclusion, respectively³¹. Like rMATS, PSI is defined as a measure of the relative abundance of transcripts which contains the target exon over the relative abundance of all transcripts. This ratio can be calculated through number of the paths supporting the inclusion of the node divided by the total number of the paths supporting both the inclusion and exclusion of that node³¹. Here, Whippet v1.6.2 was applied to quantify read data and identify DAS between two breeds in each dataset. To create the splice graphs, sheep genome (Oar Rambouillet v1.0) and Ensembl ovine GTF file (rambouillet_v1.0.106) were used. Whippet was run with the “biascorrect” option to consider GC-content and 5’ sequence bias correction. To generate more accurate results, events with associated posterior probability > 0.8 and $|\Delta$ PSI| \geq 0.1 were considered as DAS.

Functional analysis

To identify the main biological processes or pathways that the genes with stable DAS may be involved in, functional enrichment analysis was performed. All these genes were used for gene ontology (biological process) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis using EnrichR tool⁶⁰. The terms with p-value < 0.01 and FDR < 0.2 were considered statistically significant. Moreover, these genes were subjected to protein–protein interaction (PPI) network analysis using STRING database to determine whether they are members of functional and significant interaction networks. Also, transcription factors (TFs) that were presented in the final DAS genes were identified according to AnimalTFDB database (version 3.0)²⁹.

Visualization of the results

Splicing events were visualized using *rmats2sashimiplot* (<https://github.com/Xinglab/rmats2sashimiplot>) R package based on the inclusion levels obtained from rMATS. *ComplexUpset* (<https://krassowski.github.io/complex-upset>, v1.3.3) and *interacCircos* (<https://github.com/mrcuizhe/interacCircos>, v1.2.0) R packages were applied to construct the upset and circos plots, respectively.

Data availability

The datasets analyzed during the current study are available in the GEO repository of NCBI (PRJNA508203, PRJNA699984, PRJNA432669, PRJNA745517, PRJNA517348, PRJNA598581, PRJNA792697).

Received: 30 October 2023; Accepted: 24 January 2024

Published online: 29 January 2024

References

- Gizaw, S., Van Arendonk, J. A. M., Komen, H., Windig, J. J. & Hanotte, O. Population structure, genetic variation and morphological diversity in indigenous sheep of Ethiopia. *Anim. Genet.* **38**, 621–628 (2007).
- Mohapatra, A. & Shinde, A. K. Fat-tailed sheep-an important sheep genetic resource for meat production in tropical countries: An overview. *Indian J. Small Rumin.* **24**, 1–17 (2018).
- Kalds, P. *et al.* Trends towards revealing the genetic architecture of sheep tail patterning: Promising genes and investigatory pathways. *Anim. Genet.* **52**, 799–812 (2021).
- Bakhtiarizadeh, M. R. & Salami, S. A. Identification and expression analysis of long noncoding RNAs in fat-tail of sheep breeds. *G3 Genes Genomes Genet.* **9**, 1263–1276 (2019).

5. Mohammadi, H., Shahrehabak, M. M. & Sadeghi, M. Association between single nucleotide polymorphism in the ovine DGAT1 gene and carcass traits in two Iranian sheep breeds. *Anim. Biotechnol.* **24**, 159–167 (2013).
6. Bakhtiarzadeh, M. R., Salehi, A., Alamouti, A. A., Abdollahi-Arpanahi, R. & Salami, S. A. Deep transcriptome analysis using RNA-seq suggests novel insights into molecular aspects of fat-tail metabolism in sheep. *Sci. Rep.* **9**, 1–14 (2019).
7. Jin, M. *et al.* Oar-miR-432 regulates fat differentiation and promotes the expression of BMP2 in ovine preadipocytes. *Front. Genet.* **13**, 1–8 (2022).
8. Guangli, Y. *et al.* Identification of the genetic basis for the large-tailed phenotypic trait in Han sheep through integrated mRNA and miRNA analysis of tail fat tissue samples. *Res. Sq.* **1**, 1–31 (2020).
9. Yuan, Z. *et al.* Integrative analysis of Iso-Seq and RNA-seq data reveals transcriptome complexity and differentially expressed transcripts in sheep tail fat. *PeerJ* **9**, 1–22 (2021).
10. Ma, L. *et al.* Comparative transcriptome profiling of mRNA and lncRNA related to tail adipose tissues of sheep. *Front. Genet.* **9**, 1–18 (2018).
11. Farhadi, S., Shodja Ghias, J., Hasanpur, K., Mohammadi, S. A. & Ebrahimie, E. Molecular mechanisms of fat deposition: IL-6 is a hub gene in fat lipolysis, comparing thin-tailed with fat-tailed sheep breeds. *Arch. Anim. Breed.* **64**, 53–68 (2021).
12. Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G. & McEwan, J. C. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.* **13**, 10 (2012).
13. Xu, Y.-X. *et al.* Whole-body adipose tissue multi-omic analyses in sheep reveal molecular mechanisms underlying local adaptation to extreme environments. *Commun. Biol.* **6**, 159 (2023).
14. Moradi, M. H. *et al.* Hitchhiking mapping of candidate regions associated with fat deposition in thin and fat tail sheep breeds suggests new insights into molecular aspects of fat tail selection. *Animals* **12**, 1–20 (2021).
15. Bakhtiarzadeh, M. R. & Alamouti, A. A. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. *Sci. Rep.* **10**, 1–13 (2020).
16. Fan, H. *et al.* A transcriptomic study of the tail fat deposition in two types of Hulun Buir sheep according to tail size and sex. *Animals* **9**, 1–17 (2019).
17. Hosseini, S. F., Bakhtiarzadeh, M. R. & Salehi, A. Meta-analysis of RNA-Seq datasets highlights novel genes/pathways involved in fat deposition in fat-tail of sheep. *Front. Vet. Sci.* **10**, 1–14 (2023).
18. Kornblihtt, A. R. *et al.* Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165 (2013).
19. Jacobs, A. & Elmer, K. R. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol. Ecol.* **30**, 4955–4969 (2021).
20. Lu, Y., Yue, D., Xie, J., Cheng, L. & Wang, X. Ontology specific alternative splicing changes in Alzheimer's disease. *Front. Genet.* **13**, 1–9 (2022).
21. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
22. Fang, X. *et al.* Comparative genome-wide alternative splicing analysis of longissimus dorsi muscles between Japanese Black (Wagyu) and Chinese Red Steppes Cattle. *Front. Vet. Sci.* **8**, 1–9 (2021).
23. Qi, X., Gu, H. & Qu, L. Transcriptome-wide analyses identify dominant as the predominantly non-conservative alternative splicing inheritance patterns in F1 chickens. *Front. Genet.* **12**, 1–14 (2021).
24. Gao, S. *et al.* Comparative transcriptome profiling analysis uncovers novel heterosis-related candidate genes associated with muscular endurance in mules. *Animals* **10**, 1–13 (2020).
25. Miao, X., Luo, Q., Zhao, H. & Qin, X. Comparative analysis of alternative splicing events in skeletal muscle of different sheep. *Heliyon* **9**, 1–10 (2023).
26. Miao, X., Luo, Q., Zhao, H. & Qin, X. Comparison of alternative splicing (AS) events in adipose tissue of polled dorset versus small tail han sheep. *Heliyon* **9**, 1–13 (2023).
27. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
28. Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21**, 2963–2975 (2007).
29. Hu, H. *et al.* AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, 33–38 (2019).
30. Lin, K. T. & Krainer, A. R. PSI-Sigma: A comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* **35**, 5048–5054 (2019).
31. Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. H. & Blencowe, B. J. Efficient and accurate quantitative profiling of alternative splicing patterns of any complexity on a laptop. *Mol. Cell* **72**, 187–200 (2018).
32. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Natl. Acad. Sci.* **111**, 5593–5601 (2014).
33. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: Diversification, exon definition and function. *Nat. Rev. Genet.* **11**, 345–355 (2010).
34. Christodoulides, C., Lagathu, C., Sethi, J. K. & Vidal-Puig, A. Adipogenesis and WNT signalling. *Trends Endocrinol. Metab.* **20**, 16–24. <https://doi.org/10.1016/j.tem.2008.09.002> (2009).
35. Zhang, Q. *et al.* Effects of maternal chromium restriction on the long-term programming in MAPK signaling pathway of lipid metabolism in mice. *Nutrients* **8**, 1–17 (2016).
36. Fei, X. *et al.* Transcriptome reveals key microRNAs involved in fat deposition between different tail sheep breeds. *PLoS ONE* **17**, 1–16 (2022).
37. do Nascimento, F. V. *et al.* Association of HSD11B1 polymorphic variants and adipose tissue gene expression with metabolic syndrome, obesity and type 2 diabetes mellitus: A systematic review. *Diabetol. Metab. Syndr.* **7**, 1–13 (2015).
38. Baudrand, R. *et al.* Overexpression of 11 β -hydroxysteroid dehydrogenase type 1 in hepatic and visceral adipose tissue is associated with metabolic disorders in morbidly obese patients. *Obes. Surg.* **20**, 77–83 (2010).
39. Liu, X. *et al.* Integrated analysis of the ceRNA network and M-7474 function in testosterone-mediated fat deposition in pigs. *Genes* **13**, 1–22 (2022).
40. Szidonya, L. & Hunyady, L. A transgenic model of visceral obesity and the metabolic syndrome. *Trends Endocrinol. Metab.* **13**, 1–10 (2002).
41. Kotelevtsev, Y. *et al.* 11 β -hydroxysteroid dehydrogenase type 1 knockout mice show attenuated glucocorticoid-inducible responses and resist hyperglycemia on obesity or stress. *Proc. Natl. Acad. Sci.* **94**, 14924–14929 (1997).
42. Sun, X., Yang, X., Gui, W., Liu, S. & Gui, Q. Sirtuins and autophagy in lipid metabolism. *Cell Biochem. Funct.* <https://doi.org/10.1002/cbf.3860> (2023).
43. Ding, R. *et al.* Identify known and novel candidate genes associated with backfat thickness in Duroc pigs by large-scale genome-wide association analysis. *J. Anim. Sci.* **100**, 012 (2022).
44. Lantier, L. *et al.* SIRT2 knockout exacerbates insulin resistance in high fat-fed mice. *PLoS ONE* **13**, e0208634 (2018).
45. Perrini, S. *et al.* Reduced SIRT1 and SIRT2 expression promotes adipogenesis of human visceral adipose stem cells and associates with accumulation of visceral fat in human obesity. *Int. J. Obes.* **44**, 307–319 (2020).
46. Long, J.-K., Dai, W., Zheng, Y.-W. & Zhao, S.-P. miR-122 promotes hepatic lipogenesis via inhibiting the LKB1/AMPK pathway by targeting Sirt1 in non-alcoholic fatty liver disease. *Mol. Med.* **25**, 1–13 (2019).

47. Shu, L. *et al.* Decreased TCF7L2 protein levels in type 2 diabetes mellitus correlate with downregulation of GIP-and GLP-1 receptors and impaired beta-cell function. *Hum. Mol. Genet.* **24**, 2388–2399 (2015).
48. Chen, X. *et al.* The diabetes gene and Wnt pathway effector TCF7L2 regulates adipocyte development and function. *Diabetes* **67**, 554–568 (2018).
49. Abou Ziki, M. D. & Mani, A. The interplay of canonical and noncanonical Wnt signaling in metabolic syndrome. *Nutr. Res.* **70**, 18–25 (2019).
50. Lu, Q. *et al.* Striatin assembles a membrane signaling complex necessary for rapid, nongenomic activation of endothelial NO synthase by estrogen receptor α . *Proc. Natl. Acad. Sci.* **101**, 17126–17131 (2004).
51. Zhang, Y. *et al.* STK25 enhances hepatocellular carcinoma progression through the STRN/AMPK/ACC1 pathway. *Cancer Cell Int.* **22**, 1–23 (2022).
52. Dai, W., Choubey, M., Patel, S., Singer, H. A. & Ozcan, L. Adipocyte CAMK2 deficiency improves obesity-associated glucose intolerance. *Mol. Metab.* **53**, 1–14 (2021).
53. He, Z. *et al.* Identification of MAP3K4 as a novel regulation factor of hepatic lipid metabolism in non-alcoholic fatty liver disease. *J. Transl. Med.* **20**, 1–13 (2022).
54. Danai, L. V. *et al.* Map4k4 suppresses Srebp-1 and adipocyte lipogenesis independent of JNK signaling. *J. Lipid Res.* **54**, 2697–2707 (2013).
55. Ma, X. *et al.* MiRNAs and mRNAs analysis during abdominal preadipocyte differentiation in chickens. *Animals* **10**, 1–15 (2020).
56. Andrews, S., Krueger, F., Seonds-Pichon, A., Biggins, F. & Wingett, S. FastQC. A quality control tool for high throughput sequence data. *Babraham Bioinform.* **1**, 1–10 (2015).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Olofsson, D., Preußner, M., Kowar, A., Heyd, F. & Neumann, A. One pipeline to predict them all? On the prediction of alternative splicing from RNA-Seq data. *Biochem. Biophys. Res. Commun.* **653**, 31–37 (2023).
60. Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, 90–97 (2016).

Author contributions

M.R.B. conceived the ideas, designed study, analyzed the data, interpreted the data and wrote the main manuscript text.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-52855-1>.

Correspondence and requests for materials should be addressed to M.R.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024