

SCIENTIFIC REPORTS



OPEN

Landscape of alternative splicing in *Capra_hircus*

Tieshan Xu¹, Feng Xu³, Lihong Gu², Guang Rong¹, Mao Li¹, Fei Qiao¹, Liguang Shi¹, Dingfa Wang¹, Wanliang Xia¹, Wenjuan Xun¹, Ting Cao¹, Yiming Liu¹, Zhemin Lin² & Hanlin Zhou¹

Received: 6 November 2017

Accepted: 20 September 2018

Published online: 11 October 2018

Alternative splicing (AS) is a fundamental regulatory process in all higher eukaryotes. However, AS landscapes for a number of animals, including goats, have not been explored to date. Here, we sequenced 60 samples representing 5 tissues from 4 developmental stages in triplicate using RNA-seq to elucidate the goat AS landscape. In total, 14,521 genes underwent AS (AS genes), accounting for 85.53% of intron-containing genes (16,697). Among these AS genes, 6,342 were differentially expressed in different tissues. Of the AS events identified, retained introns were most prevalent (37.04% of total AS events). Functional enrichment analysis of differential and specific AS genes indicated goat AS mainly involved in organ function and development. Particularly, AS genes identified in leg muscle were associated with the “regulation of skeletal muscle tissue development” GO term. Given genes were associated with this term, four of which (*NRG4*, *IP6K3*, *AMPD1*, and *DYSF*) might play crucial roles in skeletal muscle development. Further investigation indicated these five genes, harbored 13 ASs, spliced exclusively in leg muscle, likely played a role in goat leg muscle development. These results provide novel insights into goat AS landscapes and a valuable resource for investigation of goat transcriptome complexity and gene regulation.

The majority of eukaryotic genes are comprised of exons and introns. Their transcribed pre-mRNAs undergo RNA splicing where introns are excised and exons are joined together to form mature mRNA sequences. The exons and introns contained in pre-mRNAs can either be included or excluded from the mature mRNA through a process called alternative splicing (AS)¹. AS was first discovered in the infectious adenovirus cycle^{2,3}, it's a fundamental regulatory process in the endogenous genes⁴ of all higher eukaryotes⁵.

AS is a widespread mechanism which increases transcriptome and proteome diversity and controls many biological processes in eukaryotes. Alternatively spliced mRNA isoforms encode different protein variants with altered amino acid sequences and therefore increase proteome diversity. It has been suggested that AS is involved in many biological processes, including several diseases in mammals⁶ and regulation of stress responses in plants⁷. In some cases, alternatively spliced mRNA can generate truncated proteins, which may interact with their partners to interfere with the formation of alternative homo- or hetero-dimers^{8,9}. In addition, AS can regulate gene expression at the transcriptional and translational levels and can increase the complexity of microRNA-based gene regulation.

Previous studies found that ~35% of intron-containing genes in humans and ~5% of intron-containing genes in *Arabidopsis* underwent AS based on alignment of expressed sequence tag (EST) contigs to genomic DNA^{10,11}. With the advent of tiling arrays and high-throughput sequencing, researchers found that 95% of intron-containing genes in humans¹² and >60% of intron-containing plant genes undergo AS^{13,14}. With recent advances in RNA isolation techniques¹⁵, sequencing techniques, and analysis tools (such as rMAT software for replicate RNA-Seq data)¹⁶, NGS-based RNA-seq datasets provide a rich resource for uncovering novel AS events and AS regulatory mechanism in a number of biological processes for different organisms.

Goats (*Capra_hircus*) serve as an important source of meat, milk, fiber, and pelts, and have also fulfilled agricultural, economic, cultural, and even religious roles throughout human civilization¹⁷. In addition, goats are now used as animal models for biomedical research, providing insights into the genetic basis of complex traits and transgenic production of peptides for medical purposes^{18,19}, which greatly relies on our understanding of gene

¹Tropical Crop Genetic Resource Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, P. R. China. ²Institute of Animal Science & Veterinary, Hainan Academy of Agricultural Science, Haikou, P. R. China. ³National Maize Improvement Center, China Agricultural University, Beijing, P. R. China. Tieshan Xu, Feng Xu and Lihong Gu contributed equally. Correspondence and requests for materials should be addressed to Z.L. (email: linzhemin2017@163.com) or H.Z. (email: zhouhanlin2017@163.com)

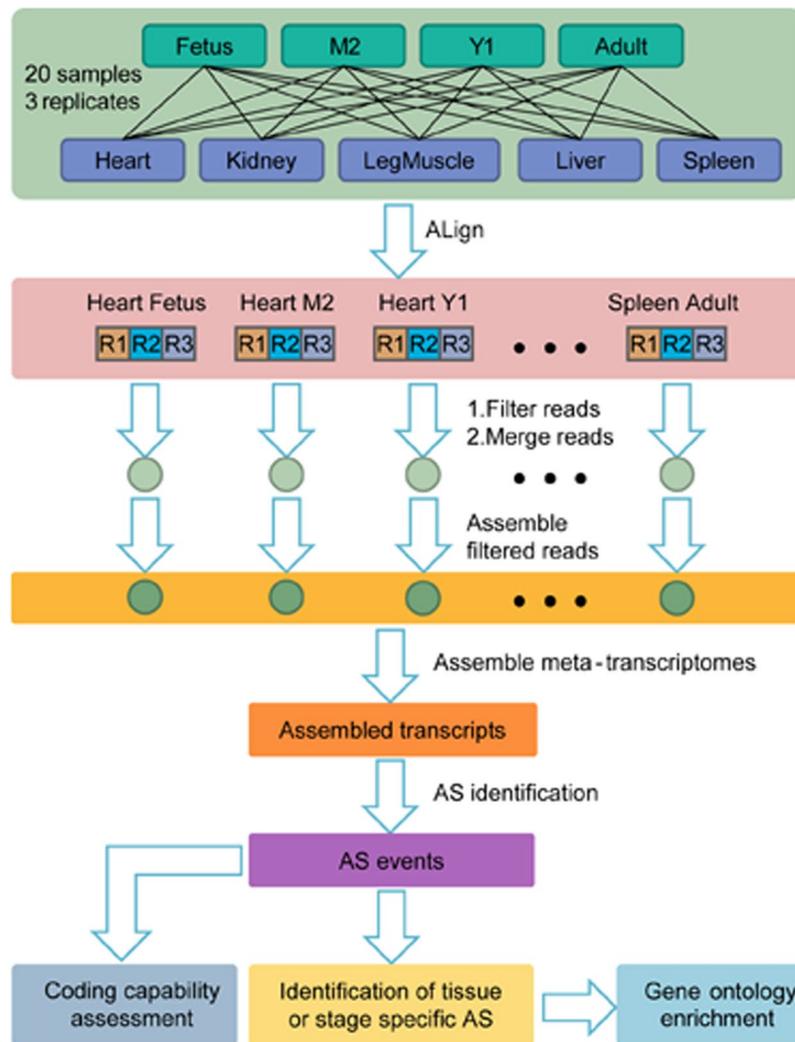


Figure 1. Pipeline for identifying goat AS events in this study. Fetus, M2, Y1, and Adult represent goats at embryonic stages beyond 135 d, two months of age, one year of age, and two years of age, respectively. Heart, Kidney, Leg muscle, Liver, and Spleen with three replicates (R1, R2, and R3) at each developmental stage were collected. After filtering and merging of raw reads, and assembling of filtered reads, assembled transcripts were obtained. After AS identification based on the assembled transcripts, coding capability assessment, and identification of tissues and developmental stage specific AS was performed. Gene ontology enrichment analysis for tissues and developmental stage specific AS were performed to identify the functional roles of AS genes.

regulatory mechanisms. Therefore, investigation of goat gene regulatory mechanisms, including AS regulatory mechanisms, is especially important. The AS characteristics for several genes in goats have been investigated, including *Izumo1*²⁰, *Lin28B*²¹, *GSK3 β* ²², and *NFIX*²³. The goat draft genome (ftp://ftp.ncbi.nlm.nih.gov/genomes/Capra_hircus/)²⁴ had been finished in 2013. The goat draft genome (CHIR 1.0) contains 2.52 G bases, 22,175 protein-coding genes and large number of ruminant-specific repeats, which comprise 42.2% of the goat genome. The goat draft genome provides an excellent platform for genome-wide AS detection. However, investigation of the goat AS landscape on a genome-wide level has not been performed.

Here, we performed genome-wide detection and characterization of the AS landscape in goats using poly (A)⁺ RNA-seq data from five tissues across four developmental stages of Hainan Black goat in triplicate. Sequencing and analysis of these 60 samples allowed for detection and characterization of intron features, AS events, AS types, differential AS and functional enrichment analysis of genes undergoing AS at the genomic level for the first time in goats. Our results provide comprehensive insights into the goat AS landscape and a basis in further investigation of the functional roles of AS in goat gene expression.

Results

Overview of RNA-seq Data. To investigate the AS landscape in goats, we carried out high-throughput RNA-seq for 60 samples spanning five tissue types (heart, kidney, leg muscle, liver, and spleen) across four developmental stages (fetus, M2, Y1 and adult) from Hainan Black goat (Supplementary Table S1), utilizing a stringent pipeline to identify the AS landscape (Fig. 1). In total, 1.38 billion raw reads (344.16 Gb) were obtained, with an

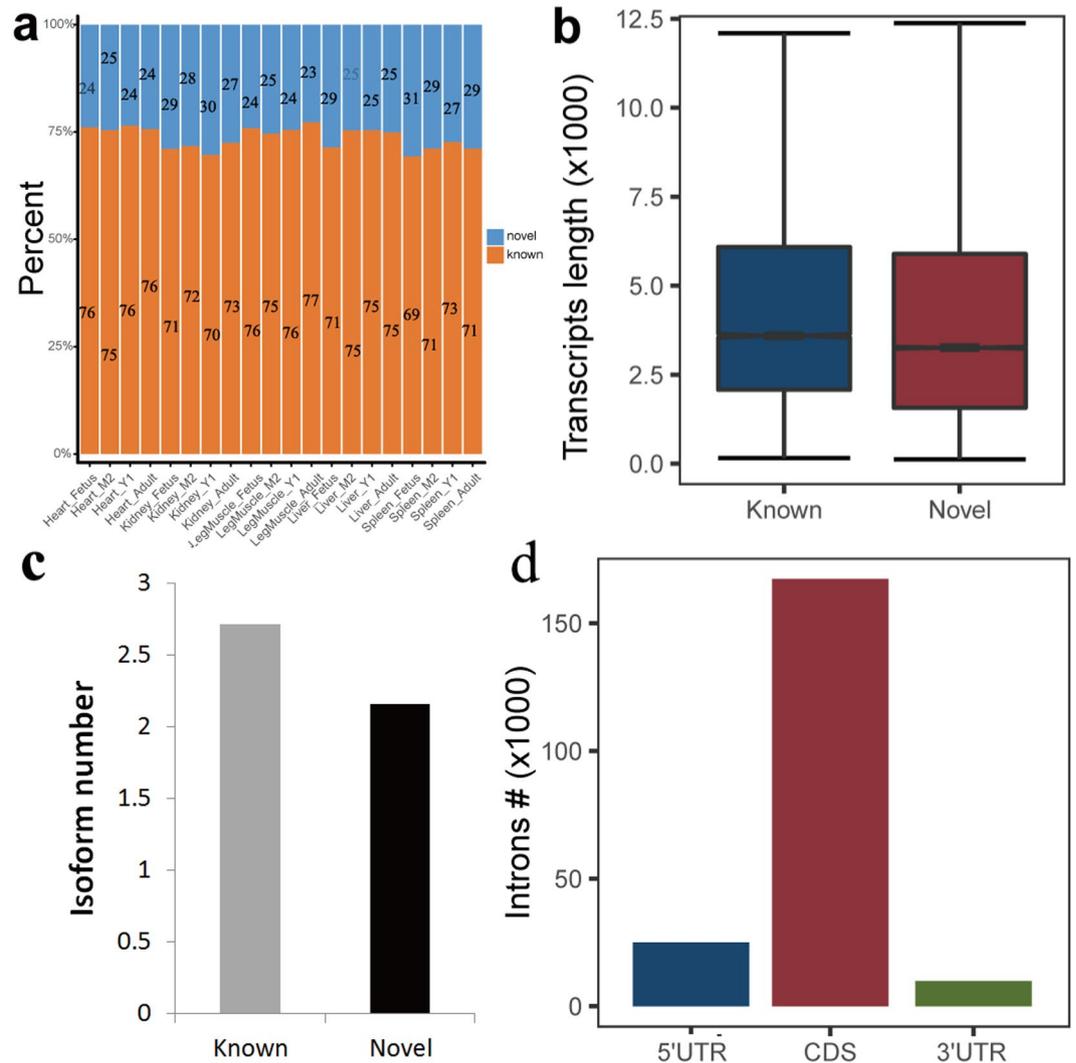


Figure 2. Characteristics of splice junctions (SJs) detected in this study. **(a)** The ratio of known (annotated in goat genome CHIR1.0) and novel (not annotated in goat genome CHIR1.0) SJs. **(b)** The comparison between known and novel goat transcript lengths. **(c)** Distribution of introns along annotated goat genes.

average of ~22.9 million raw reads (5.74 Gb) per sample. After filtering, a total of 1.35 billion high-quality reads (338.38 Gb) remained, representing an average of ~22.6 million (5.64 Gb) per sample (Supplementary Table S2). We then mapped the high-quality reads to the goat genome (CHIR1.0; ftp://ftp.ncbi.nlm.nih.gov/genomes/Capra_hircus/) using Tophat²⁵, which resulted in an alignment rate of 78.98%. Of the mapped reads, 97.14% mapped uniquely to one locus, while the remaining 2.86% mapped to multiple loci (Supplementary Table S3).

Detection of splice junctions (SJs) and transcript assembly. Transcript assembly is required for AS identification, and correct transcript assembly largely depends on accurate identification of SJs. Therefore, we performed SJ detection using TopHat²⁶. Initially obtained SJs were filtered according to two criteria described in detail in the methods. In total, we identified 680,911 SJs, including 479,361 novel SJs accounting for 70.4% of the total SJs. However, more than 70% of SJs were known SJs (annotated in goat genome) for each tissue (Fig. 2a). This is due to the majority of known SJs in a tissue always overlapped with that in another tissue, but the minority of novel SJs did not overlapped with that in another tissue. These results indicate the current SJs annotation in the goat genome (CHIR1.0) is largely incomplete.

To reduce the number of false positive assembled transcripts, we discarded transcripts that contained intronic reads >15% and displayed expression levels <10% of the major isoform from the same gene. Identified transcripts were then added to the goat gene annotation using Cuffcompare. In total, 55,035 genes (including 18,834 annotated genes and 36,201 novel genes) and 124,139 transcripts (including 83,489 annotated transcripts and 40,650 novel transcripts) were detected. In this study, we sequenced the cDNA fragment of transcribed RNAs with PolyA+, and tried to assemble them into full length transcripts. Because intergenic regions accounted for majority of goat genome and many long non-coding RNAs or other expression noise located in these regions, we detected 36,201 novel genes. However, almost all the current goat annotated genes were protein-coding genes in the paper of Dong *et al.*²⁴, resulting in 22,175 protein-coding genes annotated. Thus, huge gap of gene counts

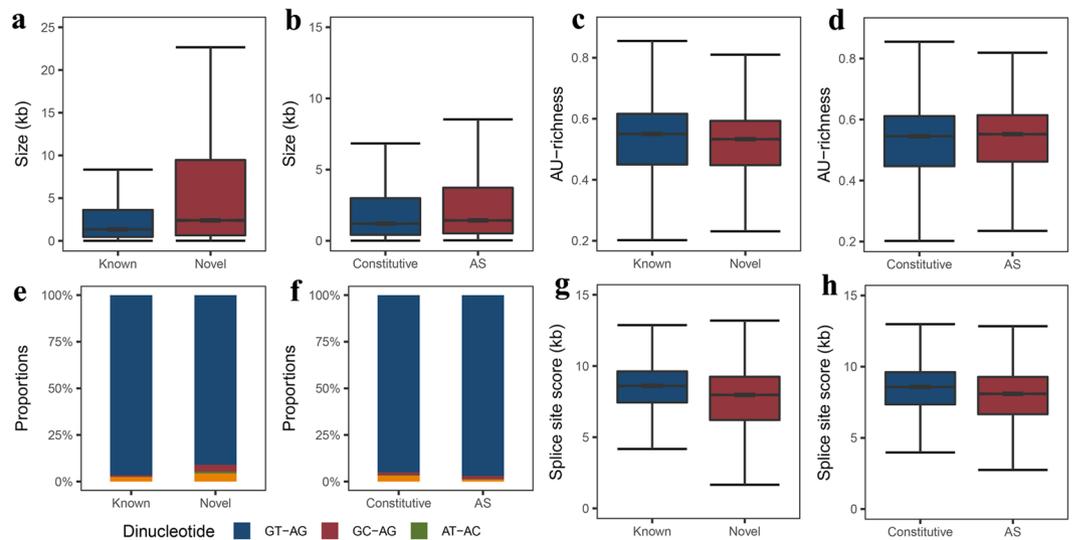


Figure 3. Characteristics of introns detected in this study. (**a** and **b**) Size comparison between known and novel introns, and between constitutive and alternatively spliced (AS) introns. (**c** and **d**) AU-richness comparison between known and novel introns, and between constitutive and AS introns. (**e** and **f**) Comparison of dinucleotide proportions between known and novel introns, and between constitutive and AS introns. (**g** and **h**) Comparison of splice site scores between known and novel introns, and between constitutive and AS introns.

existed between the current study and Dong *et al.*²⁴. The average length and average isoform number distributions for known and novel transcripts are presented in Fig. 2b,c.

In total, we identified 16,997 intron-containing genes, of which 14,521 (85.53%) underwent AS (AS genes). This result is higher than estimated by the current goat genome, in which 78.33% of annotated intron-containing genes were AS genes (16,829 out of 21,484). These results indicated that the goat AS landscape is more complex than indicated in the annotated goat genome (CHIR1.0).

Sequence characteristics of introns. In total, we detected 202,083 introns (Supplementary Table S4) for performance of downstream sequence analysis in order to investigate their sequence characteristics. The distribution of introns along goat annotated genes is presented in Fig. 2d. Overall, the majority of introns were spliced from coding sequence (CDS) genome regions.

Previous studies have demonstrated that several SJ characteristics affect the splice efficiency, including intron size, AU percentage, the dinucleotides at the intron borders, and the sequence of the 5' and 3' splice sites^{27,28}. We therefore investigated these sequence characteristics in the identified goat introns, the results of which are presented in Fig. 3a–h and Supplementary Table S4.

The average length of the predicted introns was 6,103.75 nt (median = 1,472 nt; Fig. 3a,b and Supplementary Table S4). The average length of introns identified in this study is much longer than in the goat genome annotation (average = 3,955.00 nt, median = 1,347.00 nt respectively) (wilcoxon rank sum test, P -value = 1.2×10^{-111})²⁴ and the results of Hawkins *et al.*²⁹. In addition, the intron sizes produced by novel SJs (mean = 10,720.00 nt, median = 2,399.00 nt) is longer than produced by known SJs (wilcoxon rank sum test, P -value = 0), which may indicate that introns in genes undergoing AS tend to be larger than introns in genes that do not undergo AS (Supplementary Table S4). Furthermore, we identified 4,515 (~2% of total predicted introns) enormous introns in our study with a length > 50 kb (Supplementary Table S4). In addition, we also calculated the intron sizes of other five species according to the GFF annotation files and genome sequences (Supplementary Table S5). The results indicated that the average intron size of goat identified in this study was much shorter than that of human and slightly shorter than that of mouse, but much longer than those of chicken, lizard, and frog.

The AU-richness of introns contributes to intron recognition and splicing efficiency²⁷. We therefore examined the AU-richness of identified introns (Fig. 3c,d and Supplementary Table S4). The AU-richness of introns in genes undergoing AS was slightly higher (53.40%) than introns in genes that do not undergo AS (51.96%) (wilcoxon rank sum test, P -value = 4.3×10^{-35}). The AU-richness of known introns (annotated in goat genome CHIR1.0) was 52.46%, which was slightly higher than that of novel introns (51.22%) (wilcoxon rank sum test, P -value = 3.72×10^{-153}). The AU-richness of introns in this study was consistent with that of human, mouse and chicken, but much lower than that of lizard and frog (Supplementary Table S5). The AU-richness of introns in this study was also much lower than the results of Marquez *et al.*¹³ and Yu *et al.*¹⁴ in *Arabidopsis*, who both reported intron AU-richness higher than 60%^{13,14}.

We then examined the dinucleotides at intron borders (Fig. 3e,f). The results indicated GT-AG sequences made up the majority of dinucleotides at the intron borders, which accounted for >95% of total dinucleotides. The percentages of GT-AG for known introns were slightly higher than novel introns. Our results were consistent with the results of Marquez *et al.* (2012) and Yu *et al.*¹⁴ in *Arabidopsis*^{13,14}.

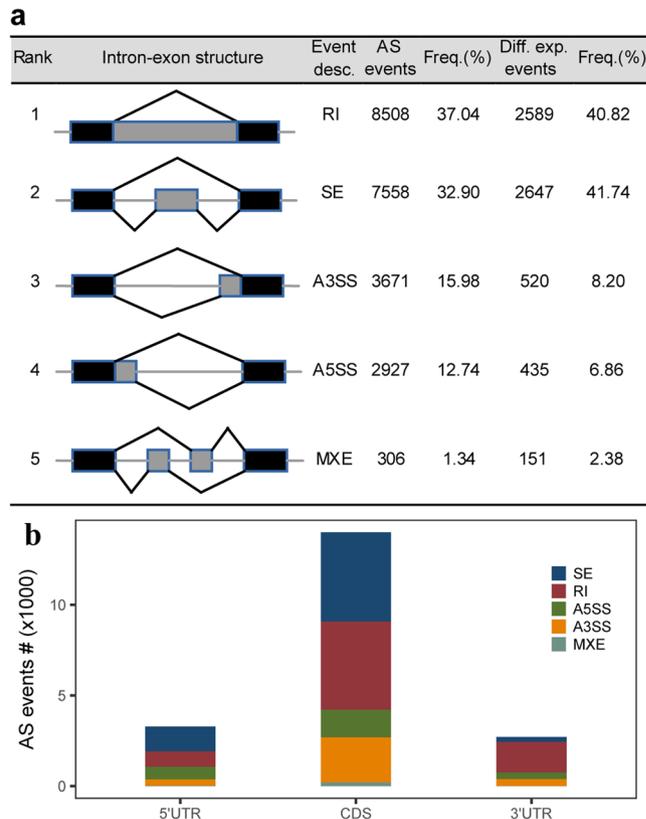


Figure 4. Types and distribution of five major AS events detected in this study. **(a)** Identified AS events and their frequencies, and differentially existed AS events and their frequencies. SE represents skipped exon AS events, RI represents retained intron AS events, A5SS represents alternative 5' splice site AS events, A3SS represents alternative 3' splice site AS events, and MXEs represents mutually exclusive exon AS events. **(b)** The distribution of AS events and types in different regions of goat annotated genes.

Finally, we compared the splice site scores (how similar the splice sites fits the consensus sequence) for constitutive introns, AS introns, known introns, and novel introns to explore differences in splicing power (Fig. 3g,h and Supplementary Table S4). The average splice site score for constitutive introns (mean = 8.09, median = 8.53) was slightly higher than that of AS introns (mean = 7.65, median = 8.10) (wilcoxon rank sum test, P -value < $1e-255$). Similarly, the average splice site score for known introns (mean = 8.30, median = 8.61) was higher than that of novel introns (mean = 6.86, median = 7.97) (wilcoxon rank sum test, P -value < $1e-255$), which may due to easier identification of introns with strong splicing power.

AS types and distribution. To account for the effect of biological variability, all the samples of a specific stage were collected from three different goats. The types and distribution of AS events in our dataset were determined using rMATS software¹⁶, which supports AS detection using RNA-seq data with biological replicates. In this study, we considered five major AS types described in¹², including skipped exons (SE), retained introns (RI), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS), and mutually exclusive exons (MXEs). We identified a total of 22,970 AS events across 8,460 genes belonging to one of the five AS types (Supplementary Table S6). RI was the most common AS event (8,508), accounting for 37.04% of the total AS events (Fig. 4a and Supplementary Table S6). SE was the second most prevalent AS event (32.90%). Our results also indicate that A3SS and A5SS account for a considerable amount of AS events (15.98% and 12.74% respectively), while MXE is a rare event (1.34%) (Fig. 4a). We also calculated the frequency of the five main AS types in humans, frogs, and lizards using the online human RNA-seq data from human (GSE45237), frog and lizard (GSE41338), and compared them to the results obtained in this study (Table 1). The most significant difference in frequency of AS type between goats and other species was that RI was the most commonly observed AS event in goats, but only accounted for a small portion in other species. In addition, SE was the most common AS type observed in humans, frogs, and lizards, while SE was the second most common AS type in goats. The results above indicate that the goat AS landscape is different from other species^{30,31}, further investigation is required to explore the mechanism underlying these differences.

Subsequently, we assessed the distribution of AS types along annotated genes. We presumed the longest isoform of a gene was the representative isoform, divided the gene into three different regions (5' UTR, CDS, and 3' UTR), and counted the number of AS events by type in each region. Results indicated that the majority of AS events (14,001) fell in CDS, among which SE and RI accounted for 35.09% and 34.75%, respectively, with the other three AS types accounting for ~30%. A considerable number of AS events occurred in 5' UTRs and 3' UTRs (3,292 and 2,715, respectively). The RI events occurred at a higher frequency in 3' UTRs (1,692; 62.32%)

AS Types					
Species	SE/%	A3SS/%	A5SS/%	RI/%	MXE/%
Human	47.04	22.48	14.87	14.59	1.02
Goat	32.90	15.98	12.74	37.04	1.33
Frog	51.82	17.28	12.48	14.50	3.92
Lizard	68.61	12.48	8.35	6.11	4.45

Table 1. Comparison of AS frequencies across four species for the five main AS types.

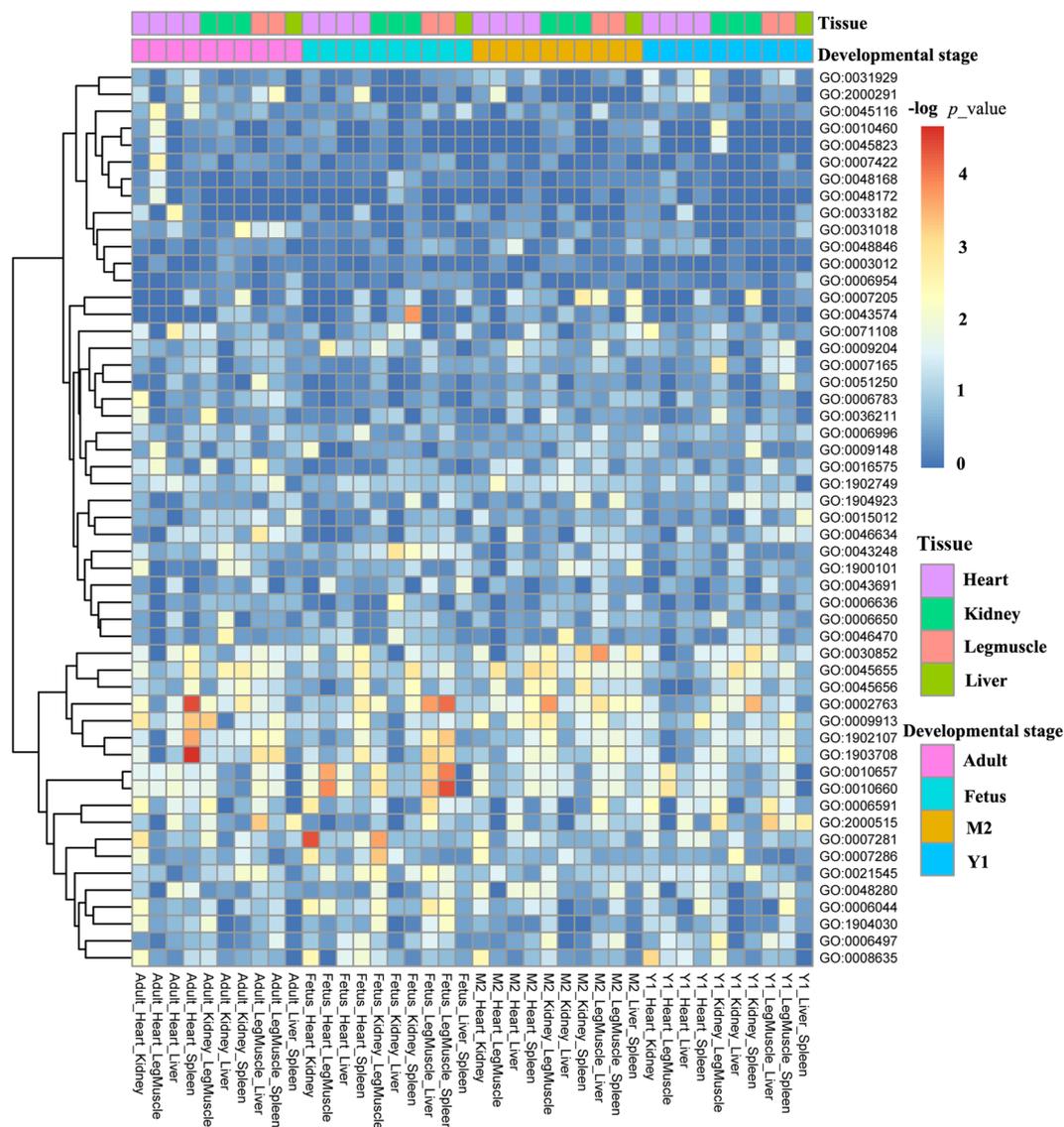


Figure 5. Differential alternative splicing events (DASE) across tissues at the same developmental stage.

than 5' UTRs (849; 25.79%), while SE events were more frequent in 5' UTRs than 3' UTRs (41.56% VS 9.65%, respectively) (Fig. 4b).

Differential AS events (DASE). We identified DASE across tissues at the same developmental stage, and across developmental stages for the same tissue using a likelihood-ratio test in the rMATS package¹⁶ followed by GO term functional enrichment analysis. DASE across tissues at the same developmental stage were found to be involved in the functional maintenance of organs (Fig. 5 and Supplementary Table S7). For example, DASE between heart and leg muscle at the adult timepoint were mainly enriched in positive regulation of heart rate or heart contraction. DASE between leg muscle and liver, and between leg muscle and spleen at the fetal timepoint were mainly enriched in regulation of muscle cell apoptosis or regulation of leukocyte differentiation. While DASE across tissues were related to organ maintenance, the DASE across developmental stages of the same tissue

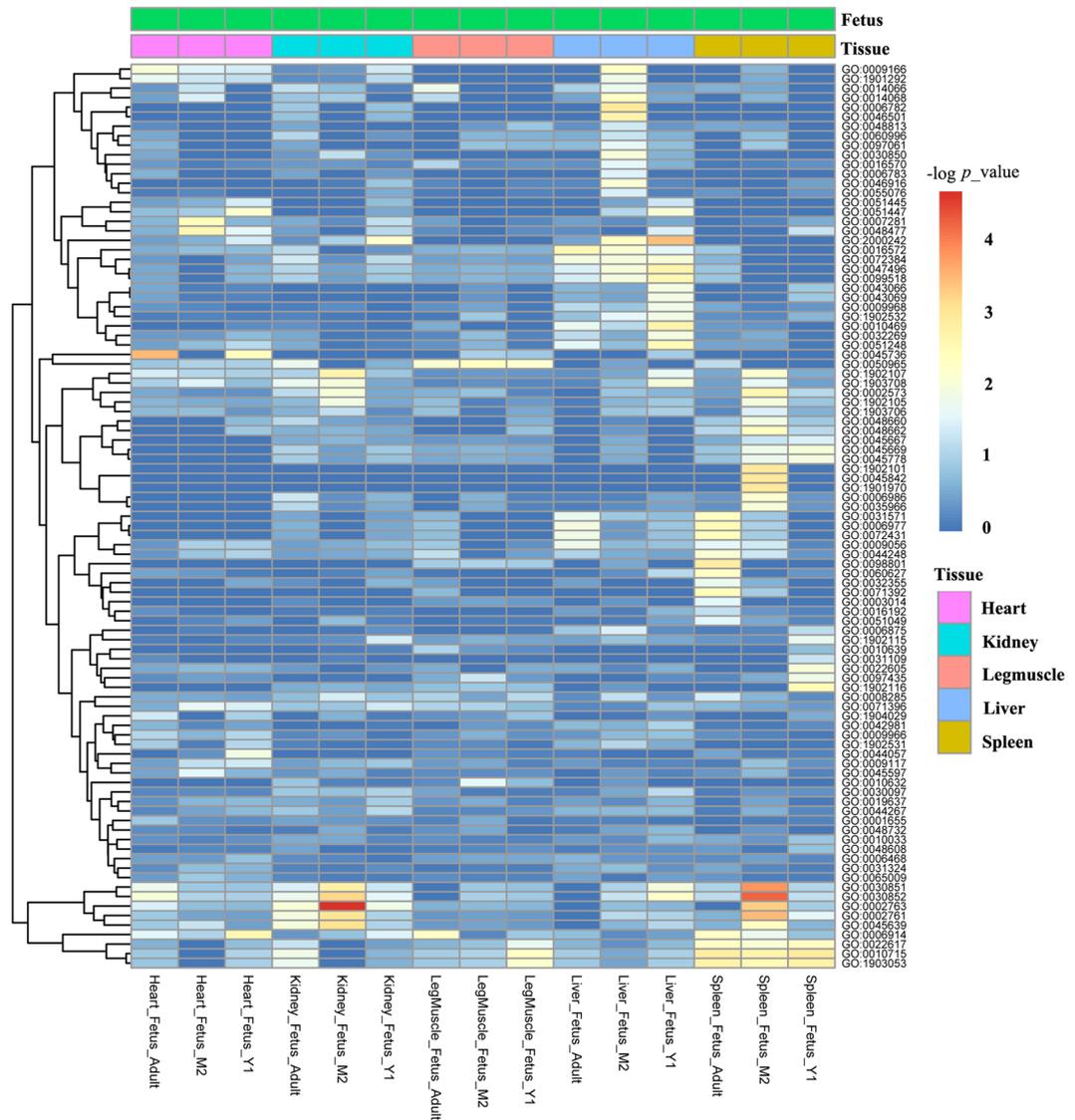


Figure 6. The DASE across developmental stages for the same tissue.

were related to different physiological stages (Fig. 6 and Supplementary Table S8). Looking at the DASE across developmental stages in spleen as an example, 11 of the 15 enriched GO terms between fetus and M2 are associated with the regulation of cell growth, differentiation, and proliferation, while DASE between fetus and Y1 were involved in the regulation of organelle assembly and organization.

Tissue- and developmental stage- specific AS. It has been shown that most AS events display strong specificity to a particular tissue or developmental stage^{5,32,33}. In this study, we explored the tissue- and developmental stage- specific AS events to assess the extent of regulation specific to tissues and developmental stages.

To investigate tissue-specific AS events, we first combined the samples from the same tissue at various developmental stages. We then compared the specificity using the Tau (τ) method³⁴. Of the 43,396 AS events, 9,463 spanning 3,580 genes were located exclusively in one tissue (Fig. 7A,B and Supplementary Data S1). The majority of tissue-specific AS events located in spleen and kidney, with 2,879 and 2,864 AS events, respectively (representing 1,081 and 1,023 AS genes respectively). Substantial tissue-specific AS events were observed in liver and leg muscle as well, corresponding to 1,487 and 1,319 AS events in 550 and 534 genes, respectively. The fewest number of tissue specific AS events were observed in the heart.

We further investigated developmental stage-specific AS events for each tissue to explore the regulatory potential of AS in development (Supplementary Data S2). The results indicated there are very few AS events specifically located in a development specific manner, with only 35, 177, 87, 249, and 130 AS events specifically located in various stages of heart, kidney, leg muscle, liver, and spleen development, respectively. This suggests that the majority of tissue-specific AS events are specifically existed across multiple developmental stages, and that the majority of AS events are tissue-specific as opposed to developmental stage-specific. In addition, more AS events were found to be specifically located in fetus tissues, with the exception of fetal heart, indicating fetal tissues appear different AS profiles compared to postnatal stages.

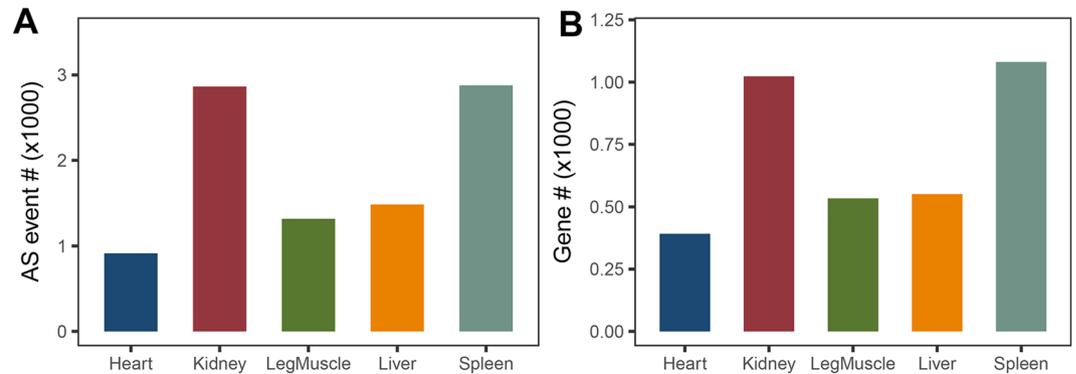


Figure 7. Tissue- specific AS events and AS genes. **(A)** Tissue- specific AS events in different tissues. **(B)** Tissue-specific AS genes in different tissues.

AS genes may be involved in the function and development of organs. Our DASE enrichment analysis identified GO terms involved in the functional maintenance of organs and to differences in physiological stages (Supplementary Data S3). We also found that AS genes might be involved in the regulation of organ function and development. For instance, many immune-related GO terms were found to be enriched for AS genes specifically expressed in spleen, such as immune system process, T cell differentiation involved in immune response, negative regulation of T cell differentiation, and immune response. This indicates AS genes specifically expressed in spleen may play important roles in spleen function. In addition, many GO terms related to material metabolism were enriched for AS genes specifically expressed in liver. Previous studies have shown that the liver is a crucial organ for material metabolism^{35,36}, which supports the findings of this study. In leg muscle, the GO term of “regulation of skeletal muscle tissue development” (GO: 0048641) was significantly enriched for AS genes specifically expressed in spleen. These genes included *BBS5*, *NRG4*, *IP6K3*, *AMPD1*, and *DYSF* (Fig. 8 and Supplementary Table S9). These results suggest that AS genes specifically spliced in leg muscle might play a major role in regulating leg muscle development.

To further investigate whether AS events might play key roles in leg muscle development, we analyzed the AS events covered by the five genes enriched in GO: 0048641 (Fig. 8 and Supplementary Table S9). In total, we detected 33 AS events in these five genes, of which 13 were specifically existed in leg muscle at significantly higher levels and one was existed in heart at significantly higher levels. The other 19 AS events were not significantly different in any tissues (Fig. 9 and Supplementary Table S9). Previous studies illustrated that four of these five genes played crucial roles in skeletal muscle development. *NRG4* (Neuregulin 4) has been shown to stimulate both the PI3K/AKT and STAT5 signaling pathway both *in vitro* and *in vivo*³⁷. PI3K/Akt signaling plays important roles during *IGF1* promoted myoblast proliferation and skeletal muscle growth in embryonic chickens³⁸. In addition, *STAT5*, which is required in GH (such as *IGF1*) actions³⁹, is involved in animal skeletal muscle development. *DYSF* (dysferlin gene) plays a key role in muscle development, as evidenced by the role *DYSF* mutations play in human muscular dystrophy⁴⁰ and that *DYSF* loss delays human muscle differentiation⁴¹. *AMPD1* (Adenosine Monophosphate Deaminase 1) has been identified as a candidate gene associated with meat production traits⁴², and a recent report revealed that *IP6K3* (inositol hexakisphosphate kinase 3) acts as an energy sensor⁴³ and is involved in apoptosis⁴⁴, and thus contributes to skeletal development. Taken together, we concluded that the 13 AS events differentially existed in leg muscle at significantly higher levels across these five genes are likely responsible for goat leg muscle development. However, further investigation is required to identify the mechanism through which these five genes and the 13 AS events regulate goat leg muscle development.

Prediction of transcripts of unknown coding potential (TUCP). To investigate the coding potential of transcripts containing at least one AS event (AS transcripts), we performed prediction of transcripts of unknown coding potential (TUCP) (Fig. 10). We identified 30,160 transcripts (48.98% of total AS transcripts) with open reading frames that can be translated into proteins. In addition, 19,027 AS transcripts (30.90% of total AS transcripts) were identified as long non-coding RNA (lncRNA) and 10,915 transcripts (17.72% of total AS transcripts) were TUCP, which likely function as regulators of gene expression or protein function. The remained transcripts were considered as other transcripts in this study.

Discussion

Since the AS phenomenon was first identified in the infectious adenovirus cycle^{2,3}, it has been demonstrated that AS represents a fundamental regulatory process in all higher eukaryotes⁵. Because of this, AS has become an important research focus in the field of eukaryotic gene regulation¹⁴. In this study, we performed RNA-seq analysis to unearth the goat AS landscape for the first time. Our results indicate that 85.53% of intron-containing goat genes undergo AS, with an average of 2.72 AS events per gene (22,970 AS events/8,460 AS genes). The percentage of AS genes identified in this study is much higher than that of soybean (63%)⁴⁵ but slightly lower than that in humans (95%)¹². These results represent an underestimation of goat AS genes due to the fact that only samples from goats reared under normal conditions without external stresses were utilized in this study. This is in line with previous studies indicating many AS events under stress treatment⁴⁶.

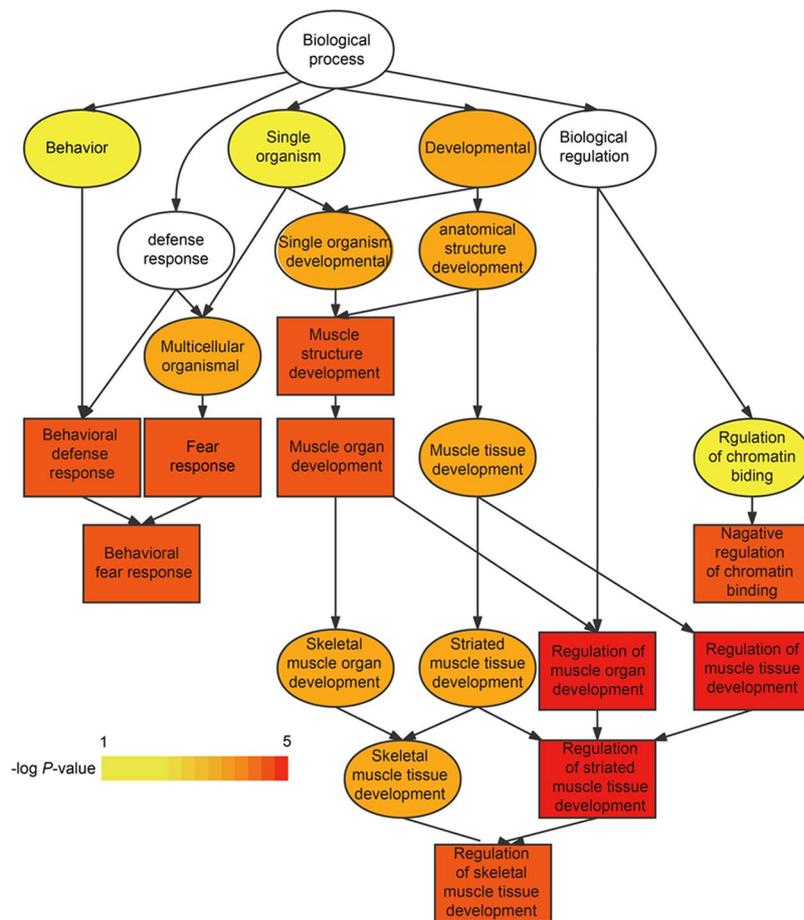


Figure 8. Functional cluster of top 10 GO terms enriched for AS genes specifically expressed in leg muscle. Colors from light to dark represent significance level of enriched GO terms (weak to strong).

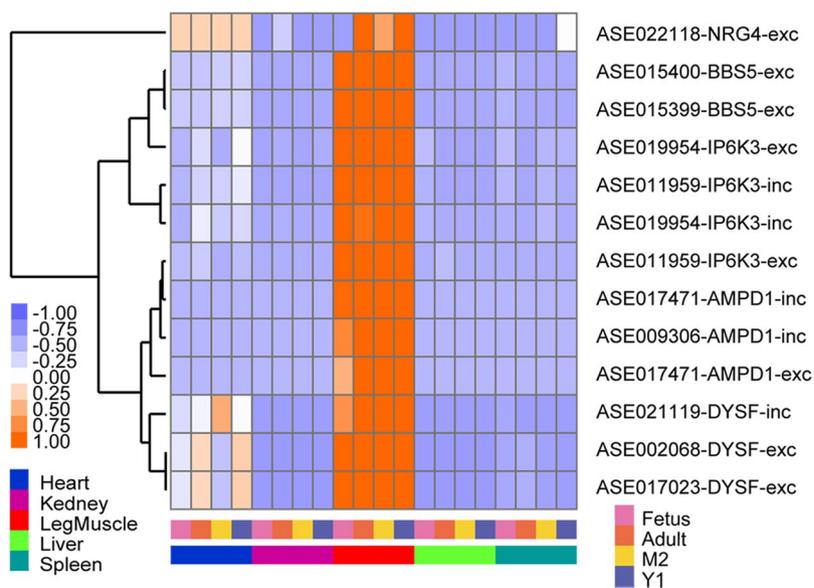


Figure 9. The 13 AS events, covered by *BBS5*, *NRG4*, *IP6K3*, *AMPD1*, and *DYSF* genes, were expressed higher in leg muscle than in other tissues.

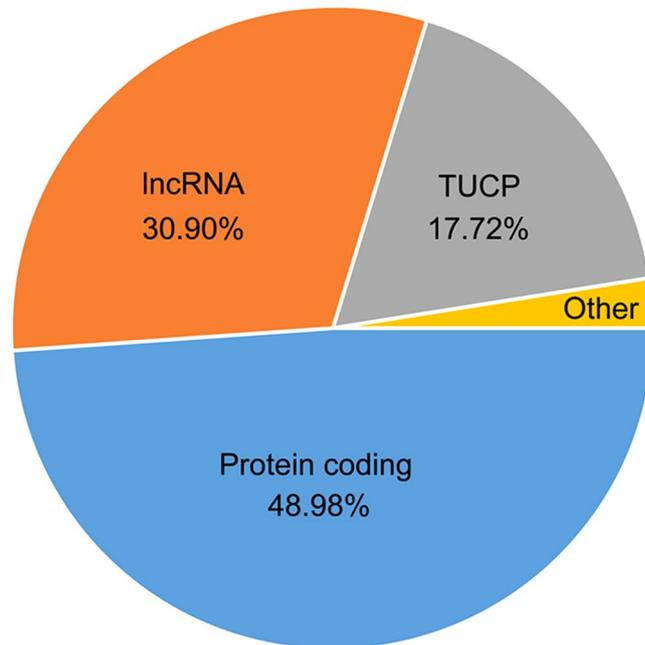


Figure 10. The prediction of transcripts of unknown coding potential (TUCP) for AS transcripts.

We found some outstanding features of goat introns compared to those of plants and other mammals. The average length of goat introns (6,103.75 nt) is much longer than other animals (chicken, lizard, and frog, average 3085.04 nt, 3974.93 nt, and 2161.58 nt, respectively) and plants, like *Arabidopsis* (average 298 nt)¹³. However, intron sizes of goat is much shorter than that of human. Surprisingly, we found 4,515 enormous introns (>50 kb) in our study (Supplementary Table S4), which is much more than previously observed in other mammals such as humans (3,473), mice (2,435), dogs (2,223) and chickens (853)⁴⁷. Shepard *et al.*⁴⁷ demonstrated that abundant amounts of repetitive elements (mainly SINEs and LINEs) in large introns can form stems with each other, and these stems with long loops within large introns allow intron splice sites to quickly identify one another, reducing the distance between donor and acceptor sites. Therefore, there should be more functional repetitive elements in the goat genome than other mammalian genomes, which will stimulate future research into goat intron splicing.

RI AS events were the most common AS type identified in this study (37.04%). This is contradictory to previous results in human^{48–50} where RI AS events have been found to be much less frequent. However, a recent study has indicated that RI AS events are far more frequent in mammals than previously predicted, and that ~53% and ~51% of all human and mouse introns have the potential to be retained in poly(A)⁺ transcripts, respectively⁵¹. Thus, the high RI percentage observed in the goat genome is well supported, and warrants further investigation in the future.

Previous observations have indicated many AS events are tissue-specific⁴⁵. Our study demonstrated that the number and frequency of AS events vary dramatically in the different tissues (Supplementary Data S2). The number of AS events identified was higher in functionally complex tissues, such as spleen and kidney. This result is consistent with previous reports in nervous system and brain^{33,52}. Recent studies have demonstrated that AS not only can increase proteome diversity, but also regulate gene expression^{14,53}. Therefore, the tissue-specific AS events obtained in this study will provide a strong basis in further investigation of the effects of AS on proteome diversity and gene expression.

Previous analysis of AS events in human tissues indicated that skeletal muscle is one of the tissues with the highest expression of tissue specific alternative exons^{50,54}. In this study, we identified 1,319 AS events specifically existed in goat leg muscle, many of which are involved in goat leg muscle development, including five genes (*BBS5*, *NRG4*, *IP6K3*, *AMPD1*, and *DYSF*) associated with the “regulation of skeletal muscle tissue development” GO term. 13 of the AS events harbored by these five genes were differentially existed in leg muscle at significantly higher levels. Four of the five genes (*NRG4*, *IP6K3*, *AMPD1*, and *DYSF*) have been previously reported to play crucial roles in skeletal muscle development. These results provide a basis in investigating the potential splicing-related regulatory mechanism of these five genes in goat leg muscle development.

In conclusion, we performed a comprehensive analysis of the goat AS landscape at the genome-wide level for the first time. These results provide a valuable resource for understanding gene expression and the biological function of AS in goats. Our data, which only contains the tissues in normal conditions, combined with stress-associated goat AS profiles will present tremendous resources for exploring the regulatory mechanisms underlying goat tissue development.

Materials and Methods

Animal management. The management of Hainan Black goats used in this study was identical with that described in our previous work⁵⁵. Briefly, animals were reared on cultivated grasses including king grass (*Pennisetum purpureum* K. Schumacher × *P. typhoideum* Rich), stylo (*Stylosanthes guianensis* SW.), *paspalum* (*Paspalum scrobiculatum* Linn.) and guinea grass (*Panicum maximum* Jacq.). The goats received routine vaccinations to general epidemic diseases yearly in spring and autumn. All kids stayed with their mothers up to weaning at 2 months of age. Pre-weaning kids had free access to the cultivated grasses ad libitum and were supplied with kids' concentrated supplement. The post-weaning kids were separated from their mothers and penned together. Because the body weight of Hainan Black goats continuously increases till two years old of age, goats > two years of age were considered adult goats.

Sample collection. Three healthy goats from each of four developmental stages were selected. The four developmental stages were embryonic late stage (the embryonic ages beyond 135 d, Fetus), two months of age (M2), one year of age (Y1), and adult (>two years old). Leg muscle, kidney, heart, liver, and spleen were collected from each goat at each developmental stage in sterile condition. All samples were collected within 15 min after exsanguination, immediately immersed in liquid nitrogen, and stored at -80°C .

RNA isolation, RNA-seq library preparation and sequencing. RNA isolation, RNA-seq library preparation and sequencing were performed as previously described⁵⁶. Briefly, total RNA was isolated from all samples using the RNAiso plus kit (Takara, Dalian, China) following the manufacturer's instructions. The RNA quality was analyzed by 1.0% agarose gel electrophoresis and spectrophotometric absorption at 260 nm in a Nanodrop ND-1000[®] Spectrophotometer. All RNA samples were treated with DNase I recombinant (Roche, Shanghai, China). The mRNA was separated from 6 mg of total RNA and fragmented into short fragments using fragmentation buffer. After first strand cDNA synthesis and purification, sequencing adapters were ligated to the 5' and 3' ends of the fragments, after which the products with 5' and 3' adapters were amplified purified. Finally, the libraries were sequenced on Illumina sequencing platform (HiSeq[™] 2500).

The evaluation of RNA-seq data. We evaluated the reliability of our RNA-seq data in identifying and analyzing goat AS profiles based on two aspects. First, we assessed read distributions along goat annotated genes. Wang *et al.* (2009) revealed RNA fragmentation provides more even coverage along the gene body, while reducing coverage at the 5' and 3' ends during the RNA-seq library construction⁵⁷. As this approach was used to construct our libraries, we assessed the read distribution along goat annotated genes. Overall, most reads mapped to the body of goat annotated genes for each sample (Supplementary Fig. S1a), which is consistent with Wang *et al.*⁵⁷. We then performed coverage analysis of each annotated transcript to assess the percentage of each transcript covered. The results indicated that more than 80% of the annotated transcripts were covered by at least four uniquely mapped reads in our RNA-seq data (Supplementary Fig. S1b). Taken together, the results above indicate that our RNA-seq data is of high enough quality to comprehensively evaluate the goat AS landscape.

Read alignment. First, we removed sequencing adaptors using Trim Galore version 0.3.7, which automatically identifies and removes adaptor sequences in paired-end reads. We also removed low quality sequences (reads where more than 30% of the bases had PHRED quality scores <20) and ambiguous bases using homemade Perl scripts. Then, we downloaded the goat genome sequence (build CHIR1.0) and gene annotation from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Capra_hircus/). Next, we mapped the high-quality reads to the goat reference genome using TopHat2 version 2.2.6²⁶, utilizing Bowtie2 version 2.1.0⁵⁸ for mapping, allowing a maximum total alignment gap length of 3 nt (`-read-gap-length 3`), no more than 2 mismatches (`-read-mismatches 2`), and setting the `-read-edit-dist` option to 3. The read realign edit distance was set as 0, and no discordant reads pair were reported. The uniquely mapped reads were used for further analysis.

Transcriptome assembly. To increase the transcriptome assembly accuracy, we first removed potential erroneous reads covering SJs or distributed in exonic regions. Specifically, we used TopHat2, which reduces the false discovery of introns in tandem repeats, to predict SJs following the method described in Marquez *et al.*¹³. To eliminate false positives resulting from erroneous alignments, we filtered out SJs predicted from reads with mismatches. We also filtered out SJs supported by one or more reads with mismatches that were within 10 nt of a junction supported by perfectly matched reads. Only SJs supported by at least 2 reads after removing PCR duplicates were considered in downstream analysis.

To investigate the transcription atlas of each tissue at each developmental stage, we merged the filtered reads of the three biological replicates and assembled a meta-transcriptome using Cufflinks2 version 2.2.1⁵⁹, with the following parameters (`-F 0.1 -j 0.15 -u -b`) to reduce misidentification of anti-sense transcripts and incorrect fusions of two or more transcripts with the genome reference¹⁴. We used the genome sequence for read fragment bias correction, and set a threshold of 0.1 to filter out low abundance isoforms that may not be reliably assembled⁶⁰. To reduce the false discovery of RI, we ignored incompletely spliced isoforms with a relative abundance cutoff of 0.15, which is calculated from the minimum coverage depth in the intronic region divided by the number of spliced reads⁵⁹. After producing the 20 meta-assemblies, we merged them using Cuffmerge version 2.2.1. We mapped assembled transcripts against all CHIR1.0 gene models using Cuffcompare version 2.2.1, which filters out redundant transcripts with the same intron chain but different transcription start or stop sites.

Expression estimation of assembled transcript. Expression levels of the assembled transcripts were determined using Cufflinks version 2.2.1⁵⁹. Normalized abundance estimates (FPKM) were computed for all assembled transcripts, by applying the geometric normalization method⁶¹.

Identification of DASE. By providing the assembled gene models, we identified the main five types of AS events using rMATS¹⁶. To minimal the false discovery of AS events, we removed events within intergenic regions, and those with different transcription directions. We defined the exon exclusion isoform (EEI) as the transcript with a larger intron, and exon inclusion isoform (EII) as the transcript with a shorter intron. Then, we quantified and normalized each isoform by counting the reads spanning the spliced region in each sample. Taking the abundance of AS derived transcripts into account, we assessed the relative abundance of EII and EEI by defining AS score as follows:

$$AS\ score = \frac{PSI - 0.5}{0.5} \quad (1)$$

$$PSI = \frac{Counts_{EII}}{Counts_{EII} + Counts_{EEI}} \quad (2)$$

AS-score is transformed from percent spliced in (PSI), which is defined by Burge lab⁶² and also employed by rMATS¹⁶. AS-score varies from -1 to 1 , with -1 only EEI expression, 1 indicating only EII expression, and 0 indicating equal expression of both isoforms. If the AS score ranged from 0 to 1 , it indicates the EII form is the major isoform.

The DASE were identified by comparing samples across different developmental stages within the same tissue, and samples across different tissues within the same developmental stage using the likelihood-ratio test with three biological replicates. AS events with a false discovery rate (FDR) < 0.05 were considered differentially AS events.

Identification of tissue- or developmental stage- specific AS events. To identify tissue or developmental stage- specific AS events, we investigated the expression- specificity of each AS transcript following the pipeline of Supplementary Fig. S2. Because one pre-mRNA could be alternative spliced into different isoforms, we used rMATS to quantify the normalized local read count of ESI and EII for each AS event instead of quantifying the global abundance of the whole transcript. To reduce the false discovery rate, we filtered out isoforms with low (read count < 3) coverage. Then, we used Tau (τ) to measure the specificity. Tau is one of the most robust methods³⁴, taking both expression abundance and the number of samples into consideration. It is defined as:

$$\tau = \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n - 1}; \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} (x_i)} \quad (3)$$

If one isoform was expressed in a single tissue or developmental stage, it was considered a tissue- or developmental stage- specific AS event, respectively.

Gene functional analysis. As many goat genes do not have GO annotations, we first performed diamond blastx⁶³ search all goat annotated cDNAs obtained from CHIR 1.0 against NCBI non-redundant protein database (nr; <ftp://ftp.ncbi.nih.gov/blast/db/FASTA/>) with a cutoff e-value of 10^{-5} . Sequences were further analyzed by Blast 2GO version 2.5⁶⁴ with the default parameters using updated databases for GO (gene ontology) mapping, inter-pro-scan, enzyme code. In addition, we used all the expressed genes detected in this study as the background to perform GO enrichment analysis of specific or differential expressed genes. During the GO analysis, we used hypergeometric test, implemented in topGO⁶⁵, and adjusted the P-values by FDR.

Prediction of transcripts of unknown coding potential (TUCP). We followed the methods described in Lyer *et al.*⁶⁶, which integrated predictions from the alignment-free Coding Potential Assessment Tool (CPAT)⁶⁷ and searched for Pfam 30.0⁶⁸ matches to assess the coding potential. CPAT uses a logistic regression model and takes four sequence features as parameters: open reading frame size, open reading frame coverage, Fickett TESTCODE statistic, and hexamer usage bias. To optimize the balanced accuracy metric, we randomly sampled 2,000 of the putative noncoding and protein-coding transcripts. Finally, we used a CPAT probability of 0.40 as the cutoff, as it achieved accurate discrimination of lncRNAs and protein-coding genes (sensitivity = 0.97, specificity = 0.97, Supplementary Fig. S3). As additional evidence of coding potential, we scanned all transcripts for Pfam A or B domains across the six reading frames. We designated putative noncoding transcripts with either a Pfam domain or a positive CPAT prediction as TUCPs.

Ethics statement. All sample collection and subsequent experiments were approved by, and all methods were performed in accordance with, the Ethical and Animal Welfare Committee of Beijing, China. Goats were slaughtered using the electric shock method followed by jugular vein bloodletting method within 30 seconds to ameliorate their suffering.

Data Availability

RNA-seq data used in this study has been uploaded to the Short Read Archive (SRA) under the accession number SRP109247. In addition, we have built a database to display and easily download our results (http://xufeng.ngrok.xiaomiqiu.cn/jbrowser/JBrowse-1.12.3/index.html?data=goat&loc=NC_005044.2%3A1206.1408&tracks=goat_AS_gtf%2CTotal.bed%2CGenes.bed&highlight=).

References

- Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514**, 1–30, <https://doi.org/10.1016/j.gene.2012.07.083> (2013).
- Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* **74**, 3171–3175 (1977).
- Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
- Leff, S. E., Rosenfeld, M. G. & Evans, R. M. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu Rev Biochem* **55**, 1091–1117, <https://doi.org/10.1146/annurev.bi.55.070186.005303> (1986).
- Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**, 153–165, <https://doi.org/10.1038/nrm3525> (2013).
- Zou, F. *et al.* Sex-dependent association of a common low-density lipoprotein receptor polymorphism with RNA splicing efficiency in the brain and Alzheimer's disease. *Hum Mol Genet* **17**, 929–935, <https://doi.org/10.1093/hmg/ddm365> (2008).
- Filichkin, S. A. *et al.* Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in *Arabidopsis*. *Mol Plant* **8**, 207–227, <https://doi.org/10.1016/j.molp.2014.10.011> (2015).
- Seo, P. J., Hong, S. Y., Kim, S. G. & Park, C. M. Competitive inhibition of transcription factors by small interfering peptides. *Trends Plant Sci* **16**, 541–549, <https://doi.org/10.1016/j.tplants.2011.06.001> (2011).
- Seo, P. J. *et al.* Targeted inactivation of transcription factors by overexpression of their truncated forms in plants. *Plant J* **72**, 162–172, <https://doi.org/10.1111/j.1365-313X.2012.05069.x> (2012).
- Mironov, A. A., Fickett, J. W. & Gelfand, M. S. Frequent alternative splicing of human genes. *Genome Res* **9**, 1288–1293 (1999).
- Brett, D., Pospisil, H., Valcarcel, J., Reich, J. & Bork, P. Alternative splicing and genome complexity. *Nat Genet* **30**, 29–30, <https://doi.org/10.1038/ng803> (2002).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**, 1413–1415, <https://doi.org/10.1038/ng.259> (2008).
- Marquez, Y., Brown, J. W., Simpson, C., Barta, A. & Kalyna, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res* **22**, 1184–1195, <https://doi.org/10.1101/gr.134106.111> (2012).
- Yu, H., Tian, C., Yu, Y. & Jiao, Y. Transcriptome survey of the contribution of alternative splicing to proteome diversity in *Arabidopsis thaliana*. *Mol Plant* **9**, 749–752, <https://doi.org/10.1016/j.molp.2015.12.018> (2016).
- Li, S., Yamada, M., Han, X., Ohler, U. & Benfey, P. N. High-resolution expression map of the arabidopsis root reveals alternative splicing and lincRNA regulation. *Dev Cell* **39**, 508–522, <https://doi.org/10.1016/j.devcel.2016.10.012> (2016).
- Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA* **111**, E5593–5601, <https://doi.org/10.1073/pnas.1419161111> (2014).
- MacHugh, D. E. & Bradley, D. G. Livestock genetic origins: goats buck the trend. *Proc Natl Acad Sci USA* **98**, 5382–5384, <https://doi.org/10.1073/pnas.111163198> (2001).
- Ebert, K. M. *et al.* Transgenic production of a variant of human tissue-type plasminogen activator in goat milk: generation of transgenic goats and analysis of expression. *Biotechnology (N Y)* **9**, 835–838 (1991).
- Ko, J. H. *et al.* Production of biologically active human granulocyte colony stimulating factor in the milk of transgenic goat. *Transgenic Res* **9**, 215–222 (2000).
- Xing, W. J. *et al.* Molecular cloning and characterization of Izumo1 gene from sheep and cashmere goat reveal alternative splicing. *Mol Biol Rep* **38**, 1995–2006, <https://doi.org/10.1007/s11033-010-0322-9> (2011).
- Cao, G. *et al.* Analysis on cDNA sequence, alternative splicing and polymorphisms associated with timing of puberty of Lin28B gene in goats. *Mol Biol Rep* **40**, 4675–4683, <https://doi.org/10.1007/s11033-013-2562-y> (2013).
- Hou, Y. *et al.* Multiple alternative splicing and differential expression pattern of the glycogen synthase kinase-3beta (GSK3beta) gene in goat (*Capra hircus*). *Plos One* **9**, e109555, <https://doi.org/10.1371/journal.pone.0109555> (2014).
- Zhang, X. *et al.* Novel alternative splice variants of NFIX and their diverse mRNA expression patterns in dairy goat. *Gene* **569**, 250–258, <https://doi.org/10.1016/j.gene.2015.05.062> (2015).
- Dong, Y. *et al.* Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat Biotechnol* **31**, 135–141, <https://doi.org/10.1038/nbt.2478> (2013).
- Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578, <https://doi.org/10.1038/nprot.2012.016> (2012).
- Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, <https://doi.org/10.1186/gb-2013-14-4-r36> (2013).
- Goodall, G. J. & Filipowicz, W. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58**, 473–483 (1989).
- Lorkovic, Z. J., Wiczyk Kirk, D. A., Lambermon, M. H. & Filipowicz, W. Pre-mRNA splicing in higher plants. *Trends Plant Sci* **5**, 160–167 (2000).
- Hawkins, J. D. A survey on intron and exon lengths. *Nucleic Acids Res* **16**, 9893–9908 (1988).
- Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960, <https://doi.org/10.1126/science.1160342> (2008).
- Zheng, C. L., Fu, X. D. & Gribskov, M. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA* **11**, 1777–1787, <https://doi.org/10.1261/rna.2660805> (2005).
- Wan, J. *et al.* Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs. *Nucleic Acids Res* **41**, 8503–8514, <https://doi.org/10.1093/nar/gkt652> (2013).
- Raj, B. & Blencowe, B. J. Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* **87**, 14–27, <https://doi.org/10.1016/j.neuron.2015.05.004> (2015).
- Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* **18**, 205–214, <https://doi.org/10.1093/bib/bbw008> (2017).
- van den Berghe, G. The role of the liver in metabolic homeostasis: implications for inborn errors of metabolism. *J Inherit Metab Dis* **14**, 407–420 (1991).
- Fabbrini, E. & Magkos, F. Hepatic Steatosis as a Marker of Metabolic Dysfunction. *Nutrients* **7**, 4995–5019, <https://doi.org/10.3390/nu7064995> (2015).
- Jakel, H. *et al.* The liver X receptor ligand T0901317 down-regulates APOA5 gene expression through activation of SREBP-1c. *J Biol Chem* **279**, 45462–45469, <https://doi.org/10.1074/jbc.M404744200> (2004).
- Yu, M. *et al.* Insulin-like growth factor-1 (IGF-1) promotes myoblast proliferation and skeletal muscle growth of embryonic chickens in the PI3K/Akt signalling pathway. *Cell Biol Int* **39**, 910–922, <https://doi.org/10.1002/cbin.10466> (2015).
- Klover, P. & Hennighausen, L. Postnatal body growth is dependent on the transcription factors signal transducers and activators of transcription 5a/b in muscle: a role for autocrine/paracrine insulin-like growth factor I. *Endocrinology* **148**, 1489–1497, <https://doi.org/10.1210/en.2006-1431> (2007).
- Roche, J. A. *et al.* Myofiber damage precedes macrophage infiltration after *in vivo* injury in dysferlin-deficient A/J mouse skeletal muscle. *Am J Pathol* **185**, 1686–1698, <https://doi.org/10.1016/j.ajpath.2015.02.020> (2015).
- de Luna, N. *et al.* Absence of dysferlin alters myogenin expression and delays human muscle differentiation “*in vitro*”. *J Biol Chem* **281**, 17092–17098, <https://doi.org/10.1074/jbc.M601885200> (2006).

42. Wang, L. *et al.* Molecular characterization and expression patterns of AMPdeaminase1 (AMPD1) in porcine skeletal muscle. *Comp Biochem Physiol B Biochem Mol Biol* **151**, 159–166, <https://doi.org/10.1016/j.cbpb.2008.06.009> (2008).
43. El-Said, K. S., Ali, E. M., Kanehira, K. & Taniguchi, A. Molecular mechanism of DNA damage induced by titanium dioxide nanoparticles in toll-like receptor 3 or 4 expressing human hepatocarcinoma cell lines. *J Nanobiotechnology* **12**, 48, <https://doi.org/10.1186/s12951-014-0048-2> (2014).
44. Crocco, P. *et al.* Contribution of polymorphic variation of inositol hexakisphosphate kinase 3 (IP6K3) gene promoter to the susceptibility to late onset Alzheimer's disease. *Biochim Biophys Acta* **1862**, 1766–1773, <https://doi.org/10.1016/j.bbdis.2016.06.014> (2016).
45. Shen, Y. *et al.* Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* **26**, 996–1008, <https://doi.org/10.1105/tpc.114.122739> (2014).
46. Staiger, D. & Brown, J. W. Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* **25**, 3640–3656, <https://doi.org/10.1105/tpc.113.113803> (2013).
47. Shepard, S., McCreary, M. & Fedorov, A. The peculiarities of large intron splicing in animals. *Plos One* **4**, e7853, <https://doi.org/10.1371/journal.pone.0007853> (2009).
48. Galante, P. A., Sakabe, N. J., Kirschbaum-Slager, N. & de Souza, S. J. Detection and evaluation of intron retention events in the human transcriptome. *RNA* **10**, 757–765 (2004).
49. Sakabe, N. J. & de Souza, S. J. Sequence features responsible for intron retention in human. *BMC Genomics* **8**, 59, <https://doi.org/10.1186/1471-2164-8-59> (2007).
50. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476, <https://doi.org/10.1038/nature07509> (2008).
51. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**, 1774–1786, <https://doi.org/10.1101/gr.177790.114> (2014).
52. Schreiner, D. *et al.* Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. *Neuron* **84**, 386–398, <https://doi.org/10.1016/j.neuron.2014.09.011> (2014).
53. Swarup, R., Crespi, M. & Bennett, M. J. One Gene, Many Proteins: Mapping Cell-Specific Alternative Splicing in Plants. *Dev Cell* **39**, 383–385, <https://doi.org/10.1016/j.devcel.2016.11.002> (2016).
54. Castle, J. C. *et al.* Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet* **40**, 1416–1425, <https://doi.org/10.1038/ng.264> (2008).
55. Xu, T. S. *et al.* Identification and characterization of genes related to the development of skeletal muscle in the Hainan black goat. *Biosci Biotechnol Biochem* **76**, 238–244, <https://doi.org/10.1271/bbb.110461> (2012).
56. Xu, T. *et al.* Identification of differentially expressed genes in breast muscle and skin fat of postnatal Pekin duck. *Plos One* **9**, e107574, <https://doi.org/10.1371/journal.pone.0107574> (2014).
57. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63, <https://doi.org/10.1038/nrg2484> (2009).
58. Langdon, W. B. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *Bio Data Min* **8**, 1, <https://doi.org/10.1186/s13040-014-0034-0> (2015).
59. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**, 46–53, <https://doi.org/10.1038/nbt.2450> (2013).
60. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329, <https://doi.org/10.1093/bioinformatics/btr355> (2011).
61. Li, X., Nair, A., Wang, S. & Wang, L. Quality control of RNA-seq experiments. *Methods Mol Biol* **1269**, 137–146, https://doi.org/10.1007/978-1-4939-2291-8_8 (2015).
62. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**, 1009–1015, <https://doi.org/10.1038/nmeth.1528> (2010).
63. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60, <https://doi.org/10.1038/nmeth.3176> (2015).
64. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676, <https://doi.org/10.1093/bioinformatics/bti610> (2005).
65. Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607, <https://doi.org/10.1093/bioinformatics/btl140> (2006).
66. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208, <https://doi.org/10.1038/ng.3192> (2015).
67. Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74, <https://doi.org/10.1093/nar/gkt006> (2013).
68. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–230, <https://doi.org/10.1093/nar/gkt1223> (2014).

Acknowledgements

We thank Haopeng Yu for technical assistance; Luli Zhou, Xiong Zhou, Yage Zhang, Yan Huang, and Xianzhou Huang for sample preparation; Kyle Schachtschneider for manuscript modification. This work was supported by Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences [Grant No. 1630032017034]; and Major Technology Project of Hainan province, China [Grant No. ZDKJ2016017-01 and ZDKJ2016017-03].

Author Contributions

T.X., F.X. and L.G. contributed equally to this work. T.X. wrote the manuscript and generated all the Figures. F.X. and L.G. performed the majority of bioinformatics analysis. G.R. organized Figures. M.L. performed verification for the results of bioinformatics analysis. F.Q. designed the experimental pipeline. L.S., D.W. and Y.L. carried out some bioinformatics analysis. W.X., W.X. and T.C. prepared mRNA. Y.L. constructed the library. Z.L. and H.Z. provided financial supports. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33078-7>.

Competing Interests: The authors declare no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018